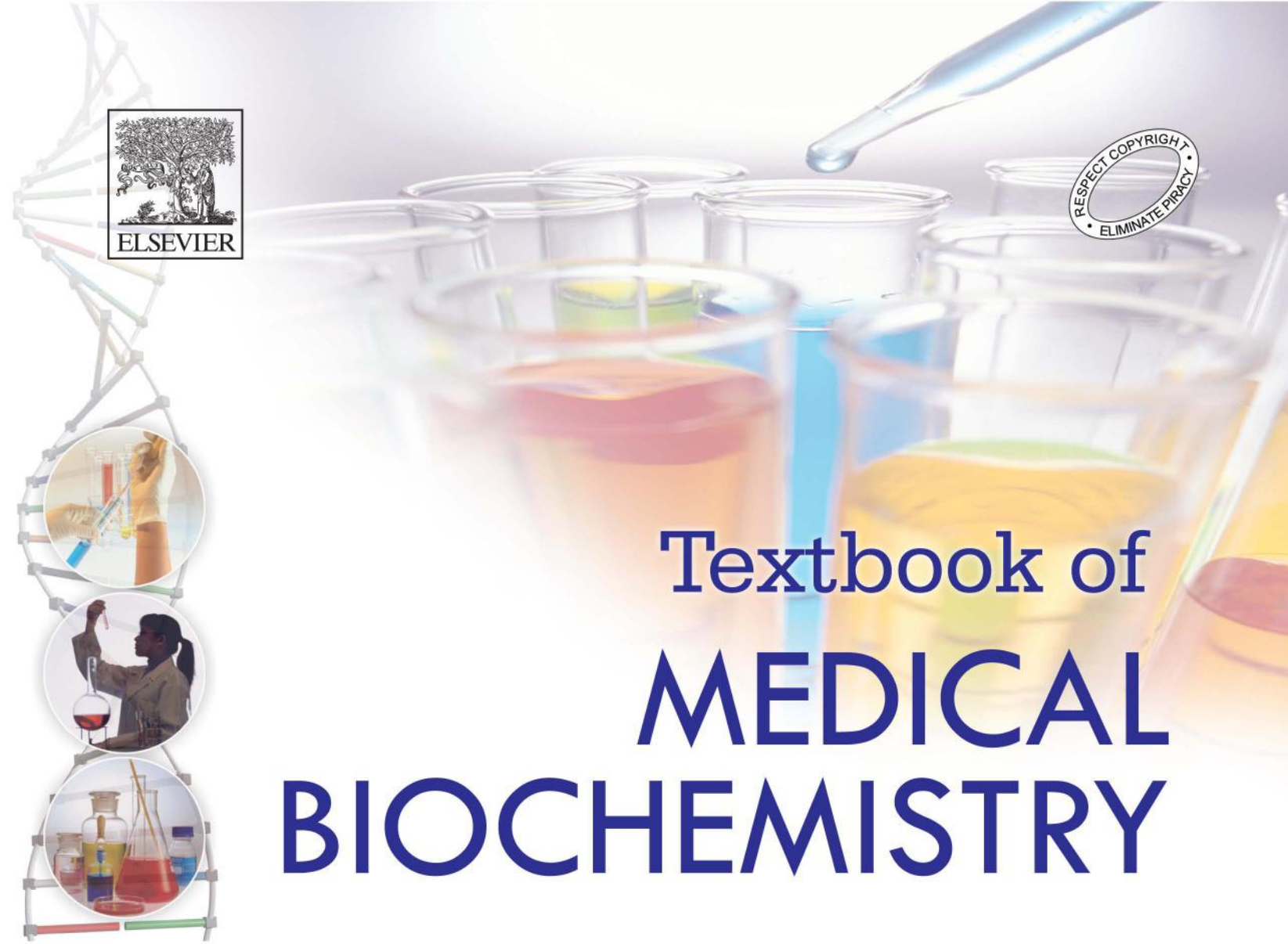


Textbook of MEDICAL BIOCHEMISTRY

Third Edition

Dinesh Puri

Textbook of

Medical Biochemistry

“This page intentionally left blank"

Textbook of

Medical Biochemistry

Third Edition

Dinesh Puri, MD, PhD
Professor and Head, Department of Biochemistry
University College of Medical Sciences and
Guru Teg Bahadur Hospital
Delhi

ELSEVIER
A division of
Reed Elsevier India Private Limited

Textbook of Medical Biochemistry, 3/e
Puri

ELSEVIER
A division of
Reed Elsevier India Private Limited

Mosby, Saunders, Churchill Livingstone, Butterworth Heinemann and Hanley & Belfus are the Health Science imprints of Elsevier.

First Edition 2002
Second Edition 2006
Third Edition 2011

ISBN 978-81-312-2312-3

Managing Editor: Shabina Nasim
Copy Editor: Goldy Bhatnagar
Manager – Publishing Operations: Sunil Kumar
Manager – Production: NC Pant

Published by Elsevier, a division of Reed Elsevier India Private Limited.

Registered Office: 622, Indraprakash Building, 21 Barakhamba Road, New Delhi 110 001.
Corporate Office: 14th Floor, Building No. 10B, DLF Cyber City, Phase II, Gurgaon 122 002, Haryana, India.

Typeset by Olympus Infotech Pvt. Ltd., Chennai, India.

Printed and bound at Sanat Printer, Kundli (Haryana), India.

Dedicated to

My Father and Brother

“This page intentionally left blank"

FOREWORD

This book has a flavour of its own both, in scope and content, distinctly different from a number of already existing textbooks on Medical Biochemistry. Author has successfully met the requirements of a student-centred and need-based, innovative curriculum and dovetailed this in right proportion with pertinent aspects of basic, applied and clinical concepts relevant to Health Sciences. While main concepts of chemical and molecular basis of diseases have not been lost sight of and the emerging frontiers of recombinant DNA research and genetic engineering have been given due importance, the student is not burdened with the details linked to and inherent in the present day avalanche of information. A rather balanced view of facts ranging from molecular to organismal levels has been undertaken. The first few chapters on basic concepts set the right pace in this direction for the students to follow.

Emphasis is given to the presentation and analysis of clinical cases so that the application of biochemical principles, including the aberrations in disease, is succinctly brought out. The approach appears to be to "understand biochemistry through health problems" or better perhaps to "understand health problems through biochemistry". At the end of each chapter, questions related to clinical cases which are of immense help to the student in the diagnostic analysis of medical problems are discussed. The "how and why" of the health problems not only seeks to unravel underlying basic mechanisms but asking questions and finding answers evoke yet another method of active learning rather than sitting with voluminous books.

I am certain that the book would be found useful as much by the medical undergraduates, as by those in clinical practice and those appearing for entrance examinations for higher courses.

Dr KN Sharma
Executive Director
Gokula Education Foundation (Health Sciences)
Bangalore
Formerly Dean and Principal
University College of Medical Sciences and
Guru Teg Bahadur Hospital, Delhi

“This page intentionally left blank"

PREFACE TO THE THIRD EDITION

It gives me a feeling of immense pleasure and satisfaction to present the third edition of *Textbook of Biochemistry*. It had earlier been well received by students and academics all over the country and abroad for its distinctive character. The book is primarily designed for undergraduate medical students and it caters to MCI recommendations. I am indebted to the large number of students and academics from whom I got useful feedback during our interactions personally and through letters. The feedback served as the most important primer for making several modifications that are incorporated in this edition.

The main emphasis in the third edition has been to make the book more reader friendly by adding a number of new figures, flowcharts and tables. Print has been enlarged and colour pattern changed to make it more soothing to eyes. Important points have been highlighted and boxed for a quick review of the core concepts. Though the book is primarily for undergraduate students, it serves as an authentic resource material for the postgraduate students. For that purpose, the higher-level learning topics have been given in separate boxes. The clinical cases have been particularly appreciated by students preparing for various competitive examinations for their utility in bridging the (artificial) gap between clinical medicine and basic medical sciences. In view of this, as a supplement to the book a website has been created. The case histories and related questions have been retained in the text, but the case discussions have been expanded and shifted to the website. The website also contains glossary and a question bank containing several thought provoking short questions, which are being further upgraded.

I sincerely hope that this new edition proves to be useful for the undergraduate medical students. I will appreciate and gratefully acknowledge any comments/suggestions for further improvement of the book. Suggestions and comments from teachers and students can be e-mailed at indiacontact@elsevier.com.

Dinesh Puri

PREFACE TO THE FIRST EDITION

Unprecedented advances in knowledge and understanding in the field of medical biochemistry, as in other basic sciences, continue to overwhelm. In view of rapidly expanding knowledge base and the accelerated information input, it has now become essential to re-examine what facts, ideas and attitudes are required by the students in the field of medical sciences. Educational objective of teaching of biochemistry to the students of health sciences is mainly two-fold: to provide them with knowledge of biochemical and molecular basis of the diseases, and to enable them to use this knowledge in the analysis of the health problems. The method of teaching followed at present in medical colleges, which dwells at length on chemical and structural details of biomolecules and the metabolic pathways, has often been criticized for its minimum applicability to the clinical realities. Moreover, knowledge gained by the traditional methods is seldom retained till the clinical teaching starts in the later years of the MBBS course. The students often find it difficult to use the biochemical principles and facts learnt earlier in the analysis of clinical cases. This book attempts to fill this gap and provides adequate interface and is designed to help the student not only to acquire the knowledge of fundamental concepts in biochemistry and molecular biology but to appreciate their applicability in clinical context.

Each chapter in this book is self-contained and serves a dual teaching function: to highlight the basic concepts, and to relate them to health problems. First few chapters cover chemistry and metabolism of major biomolecules, principles of thermodynamic and bioenergetics, electron transport and mitochondrial membrane transporters. Concepts in molecular biology are dealt with in five chapters which cover topics such as structure and function of gene and regulation of gene expression. Recent developments in recombinant DNA research and genetic engineering are also briefly discussed. The remaining chapters cover essential aspects of physiological chemistry, endocrinology and other relevant topics from a biochemical perspective. Clinical cases have been included in most of the chapters and questions related to these cases are discussed under "Case Discussion" at the end of the respective chapter. These cases are helpful in understanding the correlation between biochemical events and pathophysiological state of body.

This book is useful not only for the medical undergraduates and those appearing for various entrance examinations, but serves as a good resource in clinical practice also.

Dinesh Puri

ACKNOWLEDGEMENTS

This textbook is the outcome of dedicated efforts of many individuals and it is my proud privilege to show my gratitude to one and all. I am, in particular, indebted to Prof. KN Sharma whose constructive criticism helped in the planning and organization of the book. My special thanks to Prof. KM Prabhu for helping me in the improvement of the text and to Prof. AK Tripathi for preparation of sections on Immunological techniques and vaccines (in Chapter 33). I would also like to express gratitude to younger faculty members and residents of my department for going through the text and giving constructive criticism. I am thankful to my wife and children who graciously overlooked my absence at home while I spent long hours in libraries. It was a privilege and a rewarding experience to work with the publishing team at Elsevier, a Division of Reed Elsevier India Private Limited, Ms Shabina Nasim and Ms Goldy Bhatnagar under the able guidance of Mr Vidhu Goel.

Dinesh Puri

“This page intentionally left blank"

CONTENTS

“This page intentionally left blank"

CHAPTER

WATER, BUFFERS AND ACID-BASE CHEMISTRY

1

Water is the most abundant compound in the living world. It makes up 70% or more of the body weight in living systems. Water is required for various intra- and extracellular functions, such as transfer of chemical energy, and transport of nutrients. It is the medium in which various enzyme-catalyzed biochemical reactions take place.

The ability of water to carry out diverse functions is due to its unique properties. These properties, followed by acids, bases, buffers are described in this chapter. The basic concepts of acid-base chemistry are also dealt with.

After going through this chapter, the student will be able to understand:

- Electric dipolar nature of water and some of its unusual properties.
- Differences between (a) an acid and a base, and (b) a strong acid/base and a weak acid/base; and interpret titration curve of a weak acid.
- pH expression and the Henderson–Hasselbalch equation and the behaviour of buffers.
- Three-tier defense against imminent pH alterations and causes of various types of acid-base disorders.
- Compensatory mechanisms in each type of acid-base disorder, and the biochemical alterations in acid-base parameters in arterial blood in both the compensated and the un-compensated cases.

I. Water as Principal Biological Fluid

A. Ionic Properties of Water

Dissociability

Water molecule can dissociate to yield a proton (H^+) and a hydroxyl ion (OH^-). The disassociation is reversible.

$$H_2O \rightleftharpoons H^+ + OH^-$$

These two products influence properties of several important constituents of cells. For example, H^+ concentration determines pH of the medium, which in turn influences catalytic activities of various cellular enzymes.

Polarity

Though water molecule appears to be neutral, it is polar with distinct positive and negative poles, therefore it is referred to as an **electric dipole**. Polar nature of water accounts for its role as a transport medium for a variety of other polar molecules such as plasma proteins, and ions like K^+ and Na^+.

B. Electric Dipolar Nature of Water

The electric dipolar nature of water molecule is due to its structure which consists of two hydrogen atoms linked covalently to an oxygen atom (Fig. 1.1).

Each of the hydrogen atom shares a pair of electrons with the oxygen atom. Oxygen has a strong electron-withdrawing tendency, and therefore, these electrons tend to lie more towards oxygen and away from hydrogen. Consequently hydrogen atom acquires a localized partial positive charge. Oxygen has two more pairs of electrons which remain unshared. These electrons impart partial negative charge to oxygen.

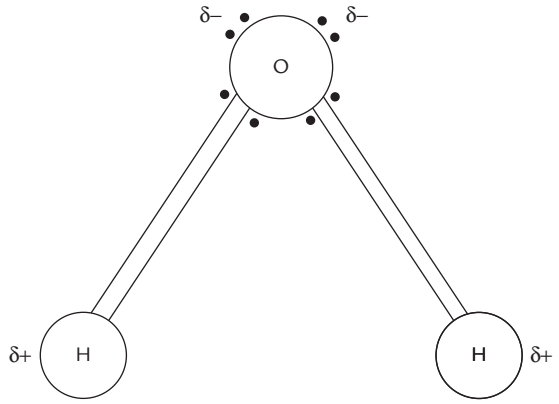


Fig. 1.1. Dipolar nature of water molecule. The two unshared electron pairs of the oxygen atom give it a localized partial negative charge, and the strong electron-withdrawing tendency of oxygen gives the two hydrogen atoms partial positive charges.

Water molecule forms a dipole because oxygen atom has high electron density (hence having partial negative charge) and hydrogen atoms are relatively electron deficient (hence having partial positive charges).

C. Hydrogen Bonds

Opposite poles of the adjacent water molecules attract each other, i.e. positive pole of one molecule attracts negative pole of the adjacent one and vice versa. This type of electrostatic attraction is called **hydrogen bond** (Fig. 1.2). As shown in Figure 1.2, arrangement of electrons around oxygen is nearly tetrahedral. Consequently, each water molecule can theoretically form hydrogen bonds with as many as four neighbouring water molecules. Since these molecules are in a stage of continuous motion, the hydrogen bonds are continuously being broken and reformed. At room temperature, each water molecule forms hydrogen bonds with three or four water molecules, at any given instant. In ice, however, this motion is restricted and four hydrogen bonds are possible, resulting in formation of crystal lattice.

In liquid water at room temperature, each water molecule engages in hydrogen bonding with three or four other water molecules. In ice, each water molecule is fixed in space and hydrogen bonds with the maximum of four other molecules.

Hydrogen bonds (bond energy = 4.5 kcal/mole) are much weaker than the covalent bond (bond energy = 110 kcal/mole). However, the hydrogen bonds are present

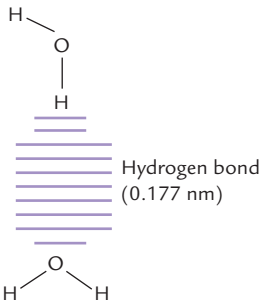


Fig. 1.2. Hydrogen bonding (designated by dashes) between the hydrogen atom of the upper water molecule and the oxygen atom of the lower water molecule.

in very large numbers and their collective strength confers great internal cohesion to the liquid water. Half-life of a hydrogen bond is very short—less than 1×10^{-9} sec. Consequently, liquid water is not viscous but very fluid in nature.

D. Properties of Water

On account of its polar nature and consequent hydrogen bonding, water possesses the following unusual properties.

Higher Melting and Boiling Points

Hydrogen bonding between adjacent water molecules cause strong intermolecular forces of attraction. Large amount of thermal energy is required to overcome these cohesive forces. As a result, melting and boiling points of water are relatively higher than other common liquids (Table 1.1). No such intermolecular attraction is present among molecules of non-polar liquids like benzene or hexane resulting in lower melting and boiling points.

Solvent Properties

Water is a better solvent than other common liquids. Most **crystalline salts** (e.g. sodium chloride), which are nearly insoluble in non-polar solvents, readily dissolve in water. Better solvent properties are because of dipolar character of water. For example, the crystal lattice structure

Table 1.1. Melting and boiling points of some common liquids

	Melting Point (°C)	Boiling point (°C)
Water	0	100
Propanol	−127	97
Benzene	6	80
Methanol	−98	65
Chloroform	−63	61
Acetone	−95	56

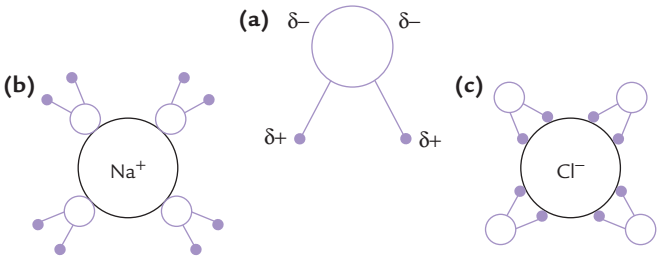


Fig. 1.3. High solubility of NaCl in water. Water hydrates Na^+ and Cl^- ions and pulls them away from the lattice. **(a)** Structure of water, **(b)** Hydration of Na^+, **(c)** Hydration of Cl^-.

of sodium chloride is held together by strong electrostatic attractions between alternating positive sodium and negative chloride ions. When this compound is exposed to water, the negative pole of the dipolar water is attracted towards Na^+ and the positive pole is attracted towards Cl^-. Thus, water hydrates both the component ions of the crystal and then pulls them away from the lattice (Fig. 1.3). Consequently, the crystal lattice is rapidly broken, which accounts for its high solubility in water.

Hydrogen bonds are individually weak, but they shape several physical properties of water, such as high boiling point and solvent properties.

Similarly, water tends to dissolve several other **organic compounds** with polar functional groups, such as sugars, alcohols, aldehydes and ketones. Solubility of these substances is also due to tendency of water molecules to form hydrogen bonds with the polar groups (hydroxyl, carbonyl or alcohol) of these substances.

In addition to crystalline salts and organic compounds, the third category of compounds with which water interacts in a characteristic way are the **amphipathic compounds.** These compounds have both hydrophobic (water repelling) and hydrophilic (water attracting) portions within the same molecule. An example is sodium salt of the long chain fatty acid, also called soap; its long hydrocarbon tail forms the hydrophobic portion and the negatively charged carboxyl group forms the hydrophilic portion (Fig. 1.4a). Soap molecules are readily dispersed by water to form aggregates called **micelles** (Fig. 1.4b). In a micelle, the hydrophilic carboxylate groups are exposed to the aqueous exterior and the hydrophobic hydrocarbon chains move toward the interior. Movement of the non-polar hydrophobic tails towards the interior of the micelle is because the surrounding water molecules tend to associate more with the hydrophilic carboxylate group than with the hydrophobic portions of the molecule. Towards the core, the hydrophobic tails tend to associate

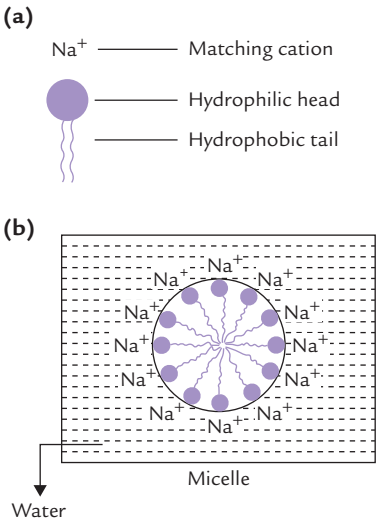


Fig. 1.4. **(a)** An amphipathic soap molecule with a hydrophilic head and hydrophobic tail, **(b)** Interaction of water with soap to form a micelle.

with one another; their association is known as **hydrophobic interactions.**

Thus, the driving forces for the formation and stability of micelles are: (a) Tendency of the surrounding water molecules to form hydrogen bonds with each other and with the negative carboxylate groups, and (b) hydrophobic interactions between the hydrocarbon tails.

Besides soap, micellar structures are also formed by other amphipathic molecules, e.g. phospholipids, nucleic acids and certain proteins. Micellar arrangement of these molecules forms the core of biological membranes.

All biochemical reactions occur in the aqueous environment, so properties of biomolecules are largely determined by their interactions with water. Hydrophilic substances (e.g. polar molecules and ions) readily dissolve in water but the non-polar substances have a tendency to minimize contact with water.

E. Ionization Behaviour of Water

Equilibrium Constant

Degree of ionization of a dissociable compound is indicated by its equilibrium constant. Let us consider a reversible reaction:

$$A + B \rightleftharpoons C + D$$

The equilibrium constant of this reaction can be derived by making use of the law of mass action which states that rate of a chemical reaction depends, among other factors, on concentrations of the substrate molecules.

Therefore, rate of the forward reaction (V_F) depends on concentration of A, B or both. Conversely, rate of the reverse reaction (V_R) depends on concentration of C, D or both

$$V_F \propto [A]\,[B] \quad (1)$$

$$V_R \propto [C]\,[D] \quad (2)$$

(or)

$$V_F = K_F\,[A]\,[B] \quad (3)$$

$$V_R = K_R\,[C]\,[D] \quad (4)$$

where K_F and K_R are the proportionality constants for the forward and the reverse reactions respectively (the square brackets indicate the molar concentrations).

Equilibrium is defined as a condition when rate of the forward reaction (V_F) equals the rate of the reverse reaction (V_R). Therefore, at equilibrium the following equality exists:

$$V_F = V_R$$

$$K_F\,[A][B] = K_R\,[C][D]$$

$$\frac{K_F}{K_R} = \frac{[C][D]}{[A][B]} \quad (5)$$

The ratio of the two constants, K_F/K_R, is also known as **equilibrium constant** (K'_{eq}) of the reaction. It indicates relative concentrations of the reactants and the products (i.e. the composition of the equilibrium mixture of a reaction). It has a characteristic value for a given chemical reaction at a given temperature. It does not depend on the starting concentrations of either the reactants or the products. Using the above principle, the equilibrium constant or the **dissociation constant** of water is expressed as:

$$K'_{eq} = \frac{[H^+][OH^-]}{[H_2O]} \quad (6)$$

Ion Product of Water

Dissociation of water proceeds to a limited extent only. At 25°C, only about 1 out of 10 million water molecules is in an ionized state. Therefore, the molar concentration of water at equilibrium is equal to the number of grams of water in one liter (1000 gm), divided by the gram molecular weight of water (i.e. 1000/18 = 55.5). The equation (6) can, therefore, be rewritten as:

$$K'_{eq} = \frac{[H^+][OH^-]}{55.5} \quad (7)$$

$$55.5\ K'_{eq} = [H^+]\,[OH^-]$$

The value of K'_{eq} has been determined by electrical conductivity measurement of water and found to be 1.8×10^{-16} mol/L at 25°C. Substitution of this value in equation (7) yields

$$1 \times 10^{-14} = [H^+]\,[OH^-] = K_w$$

The symbol K_W used to designate the product of H^+ and OH^- concentrations is called the **ion product** of water. Its value always remains the same, i.e. **10^{-14} M** in any aqueous solution. Thus, the concentration of H^+ can be calculated if concentration of OH^- is known or vice versa as shown below:

In a neutral solution,

$$[H^+] = [OH^-] = 10^{-7}\,M$$

In acidic solution, concentration of H^+ exceeds 10^{-7} M. Therefore, a corresponding decreases in the concentration of OH^- must occur, so that product of two remains 1×10^{-14}. Conversely, in the alkaline solutions concentration of OH^- is high ($> 10^{-7}$ M) and that of H^+ must be correspondingly low.

pH Scale or Measure of H^+ Concentration

The term **pH** is defined as negative logarithm of the hydrogen ion concentration, expressed in moles.

$$pH = -\log [H^+] \quad (8)$$

pH measurement is a convenient way of designating the actual concentration of H^+, and thus of OH^- in any aqueous solution. Mean pH values of body fluids are given in Table 1.2. The pH value of a neutral solution having an H^+ concentration of 10^{-7} M is 7 as calculated below:

$$pH = \log\frac{1}{10^{-7}} = \log\,(1 \times 10^7)$$

$$= \log 1 + \log 10^7$$

$$= 0 + 7$$

$$pH = 7$$

In acidic solutions, where H^+ concentration if greater than 10^{-7} M, the pH values becomes less than 7. Conversely, in the alkaline solutions pH is greater than 7 because concentration of H^+ is less than 10^{-7} M.

Water molecules can dissociate reversibly into H^+ and OH^-. The pH value of a solution is related to the concentration of H^+. In acidic solutions pH is < 7, in basic solutions, pH is > 7 and in neutral solutions pH = 7.

Table 1.2. Mean pH of some body fluids

Fluid	pH	Fluid	pH
Gastric juice	1.0	Blood	7.4
Urine	6.0	Bile	7.7
Human milk	6.7	Pancreatic juice	8.8
Tears	7.4		

Table 1.3. The pH scale

Concentration of H^+ (moles)	pH	Concentration of OH^- (moles)	pOH
1.0	0	10^{-14}	14
0.1	1	10^{-13}	13
0.01	2	10^{-12}	12
0.001	3	10^{-11}	11
0.0001	4	10^{-10}	10
0.00001	5	10^{-9}	9
10^{-6}	6	10^{-8}	8
10^{-7}	7	10^{-7}	7

Any two solutions, pH values of which differ by one unit, differ with respect to their H^+ concentrations by tenfolds. Thus, a solution with pH value of 5.0 has 10 times higher H^+ concentration compared to the one with a pH of 6.0. Another expression analogous to pH, though less commonly used, is **pOH**. It denotes concentration of the hydroxyl ions and is defined as the negative logarithm of the OH^- ion concentration, expressed in moles ($pOH = -\log [OH^-]$).

Since product of molar concentrations of H^+ and OH^- equals 1×10^{-14}, the expressions pH and pOH are inversely related to each other, so that their sum always equals 14. The pH scale is represented in Table 1.3.

pH measurement: pH of a solution is measured by **indicator dyes**, **indicator papers**, or more accurately, by the **pH meter**.

Indicator dyes: Chemically an indicator (designated as HI) is a weak acid.

The indicators most commonly used are phenol red or phenolphthalein. Action of an indicator is based on the principle that its colour in the dissociated from (I^-) is different from its colour in the undissociated form (HI). Which one of the two forms—dissociated or undissociated—predominates at a given instant depends on pH of the surrounding medium. Thus, colour of the indicator changes as pH of the surrounding medium changes.

In other words, at a given pH, colour of an indicator depends on relative proportions of the dissociated and the undissociated forms.

Note: A weak acid exists predominantly in an undissociated form at a pH value lower than its pK′ (pK′ is the negative logarithm of the equilibrium constant: $pK' = -\log K'$). When pH rises above the pK′ value, the dissociated from predominates; and both the forms are present in equal concentrations when pH equals pK′.

Indicator papers (pH papers): Principles of action of the indicator papers is same as that of the indicator dyes.

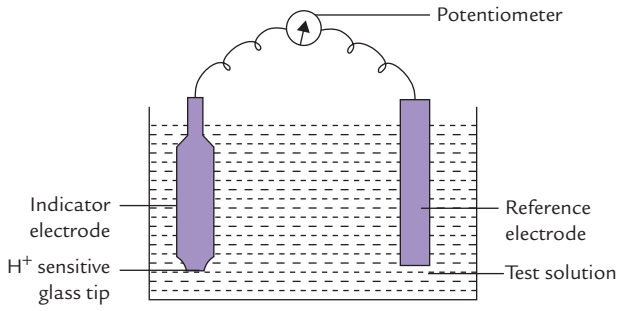


Fig. 1.5. pH meter. The reference electrode generates a stable electric potential and the indicator electrode generates a potential based on H^+ concentration of the test solution.

Like the indicator dyes, the indicator papers are useful in a specific pH range. This pH range, which occurs around its pK′ value, brings about a visible change in colour of the indicator.

pH meter: It is used for accurate measurement of pH. The basic components of a pH meter are shown in Figure 1.5.

- *Reference electrode:* It consists of a metal and its salt (in contact with a solution containing the same anion). Mercury/murcurous chloride or Ag/AgCl are most commonly used reference electrodes. The reference electrode generates a stable electrical potential.
- *Indicator electrode:* It consists of a silver wire coated with AgCl, immersed into solution of AgCl. These are placed into a tube containing a special *glass membrane tip*, which is sensitive only to hydrogen ions.
- *Potentiometer:* It measures electromotive forces.

When the indicator electrode is dipped into a solution of unknown pH, a potential difference is established across the glass membrane tip. The magnitude of potential difference varies with the hydrogen ion concentration (a measure of pH) of the test solution. The potential difference generates a signal (i.e. electromotive force) which is read against the fixed potential of the reference electrode.

Presently, pH meters which use single electrode are also available.

II. Acids, Bases and Conjugate Acid-Base Pairs

A. Acids

Acids are defined as substances that are ionized in dilute aqueous solutions to liberate protons (H^+). They are commonly referred to as **proton donors**.

$$HA \rightleftharpoons H^+ + A^- \qquad (9)$$

Table 1.4. Comparison of equilibrium constant (K'_{eq}) and pK′ values of a strong acid and a weak acid

	Degree of ionization	K'_{eq}	pK′
Strong acid	Higher	Higher	Lower
Weak acid	Lower	Lower	Higher

Acids such as hydrochloric acid (HCl), nitric acid (HNO_3) and sulphuric acid (H_2SO_4), which have a strong tendency to liberate protons, are called **strong acids.** They undergo near complete ionization in dilute aqueous solutions. Therefore, their equilibrium lies far towards the right. At the equilibrium point, concentration of the disassociated form far exceeds that of the undissociated form.

$$HCl \rightleftharpoons H^+ + Cl^- \quad (10)$$

$$H_2SO_4 \rightleftharpoons 2H^+ + SO_4^{-2} \quad (11)$$

Conversely, the **weak acids** are only partially ionized in water. Their equilibrium lies towards the left. At equilibrium point, concentration of the undissociated form exceeds that of the dissociated form. In living systems, behaviour of weak acids is more important than that of the strong acids. Acetic acid (CH_3COOH) is an example of weak acid; it mildly dissociates to yield an acetate anion and a proton.

$$CH_3COOH \rightleftharpoons CH_3COO^- + H^+ \quad (12)$$

Since degree of ionization is greater in case of strong acids, equilibrium constant of strong acids is higher than that of weak acids (Table 1.4).

pK′ value of an acid is defined as the negative logarithm of its equilibrium constant ($pK' = -\log K'_{eq}$). It is an inverse relationship. Hence, stronger acids having larger K'_{eq} values have relatively lower pK′ values. Conversely, weak acids have relatively low K'_{eq} values and high pK′ values.

B. Bases

Bases are defined as are substances that accept protons in aqueous solutions, and commonly known as **proton acceptors.** For example, Cl^-, SO_4^{-2}, CH_3COO^- accept a proton (reactions 10–12) and therefore, are bases.

Stronger bases have greater tendency to accept protons, whereas the weaker bases have weaker tendency for the same. For instance, equilibrium of the dissociation of acetic acid (reaction 12) lies towards the left, which implies that acetate ion has a strong tendency to accept proton. It is therefore, a **strong base.** Cl^- and SO_4^{-2} have weaker tendency to accept protons and are therefore weak bases.

Acids can donate protons and bases can accept protons. The strength of an acid is expressed as its pK′: it corresponds to the pH value at which the ionizable group is half-protonated.

C. Conjugate Acid-Base Pairs

In ionization reactions (discussed above), there is a proton donor-proton acceptor pair.

This pair is known as **conjugate acid-base pair**.

$$\underset{\text{(Proton donor)}}{HA} \rightleftharpoons \underset{\text{(Proton acceptor)}}{H^+ + A^-}$$

Conjugate base of a weak acid is always a strong base. For example, acetic acid is a weak acid because of its weak dissociability, while acetate, its conjugate base, has a strong tendency to accept a proton.

On the other hand, conjugate base of a strong acid is a weak base (reactions 10 and 11).

Titration Curve of Weak Acids

Titration is a procedure that is used to quantitatively determine the amount of acid in a given solution. In this procedure, an acid reacts with an alkali of known concentration till the point of neutralization (end point).

Titration curve is drawn by plotting the pH against the amount of alkali added. It indicates the pH changes of the acid that occur when alkali is added in small amounts.

Example: Titration curve of acetic acids vs sodium hydroxide of shown in Figure 1.6:

1. *At the start of the titration*, acetic acid is present predominantly in undissociated state (HA). As the alkali (sodium hydroxide; 0.1N) is added, it reacts with the acid molecules to form acetate ions (A^-). Thus, there is a progressive rise in the concentration of base (i.e. acetate) with a concomitant fall in the concentration of the acid.

   $$CH_3COOH + NaOH \rightarrow CH_3COO^- + Na^+ + HOH$$

Note: Acetate exists in the form of its sodium salt, sodium acetate (CH_3COONa), though the two have been written separately in the above equation.

2. *At the midpoint* of titration about half of the acetic acid loses its proton to form acetate. Thus, concentrations

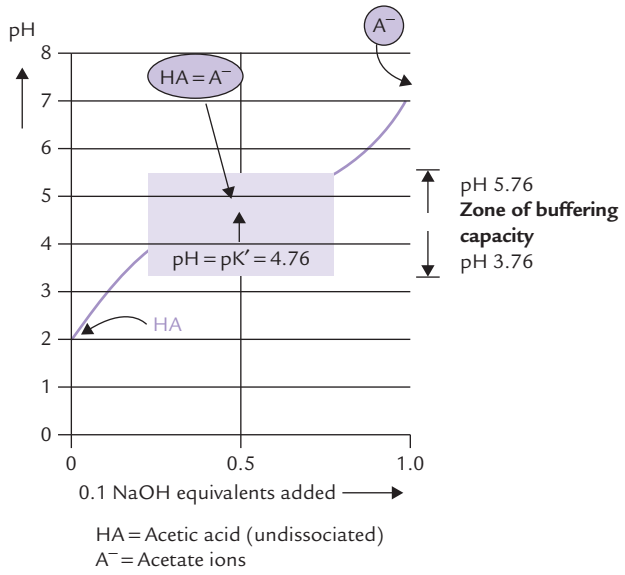


Fig. 1.6. Titration curve of acetic acid (HA), showing predominant ionic forms at various pH points. At midpoint of the titration, the concentration of acetic acid and acetate are equal, and pH is numerically equal to pH of acetic acid. The zone of buffering capacity is shaded.

of the acetic acid and its conjugate base (acetate) are equal at this point.

3. *Towards the end* of the titration, most of acetic acid loses its proton, so that the predominant ionic species present is acetate.

It may be observed that at the start and towards the end, the titration curve is relatively steep, indicating that addition of alkali causes significant alteration in pH. However, at the midpoint, the curve is relatively flat, which indicates that small additions of alkali do not elevate the pH significantly. Therefore, the flat zone represents the **buffering zone** because of the resistance it offers to any imminent alteration of pH with addition of alkali (or acid). The following important relationships exist at the midpoint of this zone:

- Acid and its conjugate base are present in equimolar concentrations.
- pK′ of the acid equals pH of the medium.

The latter relationship is obtained by using the **Henderson–Hasselbalch** equation:

$$pH = pK' + \log \frac{[Base]}{[Acid]}$$

$$\text{At the midpoint, } \log \frac{[Base]}{[Acid]} = \log 1 = 0$$

Hence, the Henderson–Hasselbalch equation can be rewritten as:

$$pH = pK' + 0 \quad \text{or} \quad pH = pK'.$$

To learn more about the Henderson–Hasselbalch equation, including derivation and significance, refer Box 1.1.

The Henderson-Hasselbalch equation relates the pH of a solution to the concentration of an acid and its conjugate base. It can be used to determine which proportion of an acid is in the protonated or deprotonated state at a given pH.

D. Buffers

***Buffer** is a system that resists any alteration in its pH when a small amount of acid or alkali is added to it.* It comprises two major components: a weak acid (HA) and its conjugate base (A^-). A buffer system is most effective when (a) these two components are present in equimolar concentrations, and (b) pH (of the medium) equals pK′ (of the acid-base pair).

A buffer remains effective when pH is within the range of pK′ ± 1.

Buffers resist changes in pH within about one pH unit of the pK′ of the buffering species (pK′ ± 1 pH unit).

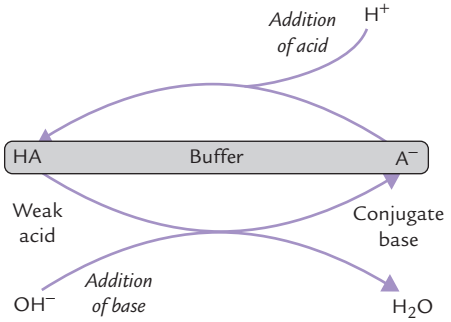
Mechanism of Action of a Buffer System

When a small amount of acid is added, it is taken up by the base component of the buffer (A^-), and any pH change is averted.

$$H^+ + A^- \rightarrow HA$$

Similarly, the acid component of the buffer system (HA) is capable of reacting with any OH^- that is added.

$$OH^- + HA \rightarrow A^- + H_2O$$



Thus, buffering action is the net result of capacity of the base component (of the acid-base pair) to neutralize the added acid, and of the acid component to neutralize the added base. During these reactions, the sum of the acid and the base does not change, only their ratio may change. For example, when H^+ is added, the buffer base (A^-) is consumed, but an equivalent amount of acid (HA)

is generated. Conversely, with addition of a base, the decrease in acid component of the buffer system is balanced by a corresponding increase in the base component.

Major Body Buffers

Cellular metabolism predominantly generates acids. Body buffers avert imminent pH change by these acids. Thus, buffers are first-line of defense against acid load. The important ones are *bicarbonate buffer, phosphate buffer, proteins*, etc.

Bicarbonate Buffer

It is the **principal extracellular buffer,** comprising carbonic acid (the proton donor) and bicarbonate (the proton acceptor). It functions in the same way as other conjugate acid-base pairs. However, there are important differences:

1. The base constituent, bicarbonate (HCO_3^-) is regulated by kidneys.
2. The acid component (H_2CO_3) is regulated by pulmonary ventilation.

Thus, bicarbonate buffer is subject to regulation by kidneys and lungs. It is part of *a three-tier defense*, described later, that also includes kidney and lungs.

The acid component of this buffer (also called respiratory component) is generated from dissolved carbon dioxide [CO_2(d)] and water, by the reaction shown below. The reaction is catalyzed by the enzyme *carbonic anhydrase*.

$$CO_2(d) + H_2O \underset{}{\overset{\text{Carbonic anhydrase}}{\rightleftharpoons}} H_2CO_3$$

The dissolved carbon dioxide in blood circulation is in equilibrium with the gaseous carbon dioxide [CO_2(g)] in the air space of the lungs. As a result, concentration of carbonic acid is ultimately dependent on the partial pressure of carbon dioxide in the gas phase. Carbonic acid can dissociate to yield bicarbonate.

$$H_2CO_3 \rightleftharpoons H^+ + HCO_3^-$$

Thus, reversible equilibria exist between the gaseous carbon dioxide in the lungs and the bicarbonate ions in blood plasma, as shown below:

$$HCO_3^- + H^+ \underset{\text{Step 1}}{\rightleftharpoons} H_2CO_3 \underset{\text{Step 2}}{\rightleftharpoons} CO_2(d) + H_2O$$

$$CO_2(d) + H_2O \underset{\text{Step 3}}{\rightleftharpoons} CO_2(g)$$

Action of the bicarbonate buffer involves these equilibria. For instance, when an acid is added to blood, concentration of H^+ rises. The latter is taken up by HCO_3^- resulting in the rise of concentration of carbonic acid (Step 1).

This causes the Step 2 to go forward, and the concentration of carbon dioxide (d) in the blood rises. This in turn results in an increase in the pressure of carbon dioxide in the gas phase in the lungs (Step 3), and the extra carbon dioxide is exhaled through increased rate of breathing.

Reverse series of reactions occur when an alkali (OH^-) is added. It is taken up by carbonic acid to form HCO_3^-.

$$H_2CO_3 + OH^- \rightarrow H_2O + HCO_3^-$$

Concentration of carbonic acid falls momentarily, but is quickly replenished from large pool of gaseous carbon dioxide. Rate of breathing decreases under these circumstances so that the carbon dioxide is retained and dissolved in water to form carbonic acid.

Bicarbonate buffer is highly effective: Bicarbonate buffer system is an effective physiological buffer because of its equilibration with a large reserve of gaseous carbon dioxide in the air space of the lungs. Since pK′ of carbonic acid is 6.1, the bicarbonate buffer should be most effective at or around pH of 6.1 (i.e. 6.1 ± 1) as a buffer is most effective when pH equals pK′. However, bicarbonate buffer is highly effective at the physiological pH of 7.4 also because of its equilibration with gaseous carbon dioxide.

Carbonic anhydrase is the principle enzyme that catalyzes generation of HCO_3^-. Decreased activity of this enzyme, therefore, results in decreased plasma bicarbonate concentration. Consequently, the ratio of bicarbonate to carbonic acid (normally 20) tends to fall, resulting in a fall of pH (see Henderson–Hasselbalch equation). *A state of primary bicarbonate deficit is therefore associated with the fall in physiological pH.*

Phosphate Buffer

It is the **major intracellular buffer**. Its pK′ value of 6.86 is near the intracellular pH of 7.0. Therefore, this buffer is very effective intracellularly. It consists of the following components:

1. $H_2PO_4^-$ as the proton donor (i.e. the acid component).
2. $H_2PO_4^-$ as the proton acceptor (i.e. the base component).

The two components are related to each other as below:

$$H_2PO_4^- \rightleftharpoons HPO_4^{-2} + H^+$$

Proteins

They may act as buffers because many of them have amino acids which behave like weak acids. For example, histidine acts as a buffer because of presence of imidazole group (Chapter 4). It has a pK′ value of 6.0 which is close to the physiological pH. Therefore, it is very effective in living systems.

BOX 1.1

Henderson–Hasselbalch Equation

The relationship between pH, pK′, concentrations of weak acid and conjugate base (or salt) is expressed by Handerson–Hasselbalch equation.

$$pH = pK' + \log \frac{[Base]}{[Acid]} \left(\text{or } pH = pK' + \log \frac{[Salt]}{[Acid]} \right)$$

Dissociation of a weak acid yields a conjugate acid-base pair

$$HA \rightleftharpoons H^+ + A^-$$

Weak acid — Proton — Conjugate base

By definition, the dissociation constant (K′) of the acid is

$$K' = \frac{[H^+][A^-]}{[HA]}$$

The above equation can be written as

$$[H^+] = K' \frac{[HA]}{[A^-]}$$

By taking logarithm of both sides, and multiplying by –1,

$$-\log[H^+] = -\log K' - \log \frac{[HA]}{[A^-]}$$

Substituting pH for –log H^+ and pK′ for –log K′

$$pH = pK' + \log \frac{[A^-]}{[HA]}$$

The equation can be used to calculate change in pH of a buffer on addition of a known quantity of acid and alkali. For example, when a strong base (NaOH) is added, acetic acid is a converted to its conjugate base.

$$NaOH + HA \rightleftharpoons A^- + H_2O + Na^+$$

Change in the ratio of base to acid $\left(\frac{[A^-]}{[HA^-]} \right)$ occurs, causing changes in pH of the medium, as can be seen from the equation.

Another important application of the equation is to find concentrations of acid and base by measuring pH. Thus, the equation has enormous practical application in clinical biochemistry in assessing the acid-base status, severity of acid-base disorder and predicting limits of compensation of body buffers. These aspects are covered in detail later in this chapter.

Most of the buffering capacity of blood is due to **haemoglobin** (Hb), which is present within RBCs. Hb is more effective than plasma proteins because of (a) its high content of histidine, and (b) its higher concentration (12–16 g/dL). As a result, buffering capacity of Hb is much higher than that of the plasma proteins.

The intracellular fluid (ICF) proteins serve as major buffers. Since ICF volume is around 60% of the total body fluid volume, proteins may be considered as most abundant buffers in the body. However, the interstitial fluid, which forms more than 70% of the extracellular fluid (ECF) lacks proteins.

The acetate buffer is ineffective in the body since its pK′ is 4.76, which is far removed from the physiological pH. Similarly ammonium buffer with pK′ of 9.25 is not effective at the physiological pH.

III. Acid-Base Balance: Applied Aspects

A. Overview

Various body fluids have nearly fixed pH values (Table 1.2), which lie within definite range. For example, pH

of blood is 7.4 and it remains within a range of **7.35–7.45**. Intracellular pH is taken to be around 7.0, although its value varies among different organelles. Deviations from normal values disrupt structural and functional integrity of body proteins, including enzymes. Properties of nucleic acids, cell membranes and other cellular constituents are also adversely affected by alteration of pH. These changes may have hazardous effects on the body, and are often incompatible with life. Therefore, it is important that the regulation of pH is given utmost priority. Failure of pH regulation occurs in a number of disorders, collectively known as acid-base disorders.

B. The Three-tier Defense

Normal body metabolism poses a constant threat to pH because it generates various products that can alter the blood pH. For example, carbon dioxide is the major acidic end product of normal metabolism that tends to lower the pH. In addition, certain abnormal metabolic states and disorders are also capable of causing alteration in pH: the pH may fall (i.e. acidosis) or rise (i.e. alkalosis). A three-tier defense system, comprising buffers, lungs, and kidneys therefore, remains constantly in operation to guard against any changes in pH. Various components may act at different times, but their overall effect is well synchronized.

(a) Buffers serve as the first-line of defense against acid load and, as mentioned earlier, bicarbonate buffer is most important for being regulated by pulmonary ventilation and renal adjustments.

(b) Pulmonary ventilation takes from a few minutes to few hours to become operational. It has a direct bearing on acid-base balance of the body because carbon dioxide is an acidic substance. Since carbon dioxide is exhaled during expiration, increase in respiratory activity reduces the acidity of body fluids.

The pulmonary ventilation system operates by *negative feedback*. For instance, fall in pH of ECF stimulates peripheral chemoreceptors and central chemoreceptors. The reflex response to these stimulations is an increase in the rate and depth of respiration. This removes the excessive carbon dioxide, thereby leading to a compensatory rise in pH. Thus, the decreased pH is brought back to normal by hyperventilation. Conversely, increased pH is compensated by hypoventilation.

(c) Renal adjustments take from several hours to few days to become effective, but provide a long-term solution by supplementing buffer action. For instance, the non-volatile acids (HA) produced endogenously are initially neutralized by the bicarbonate buffer.

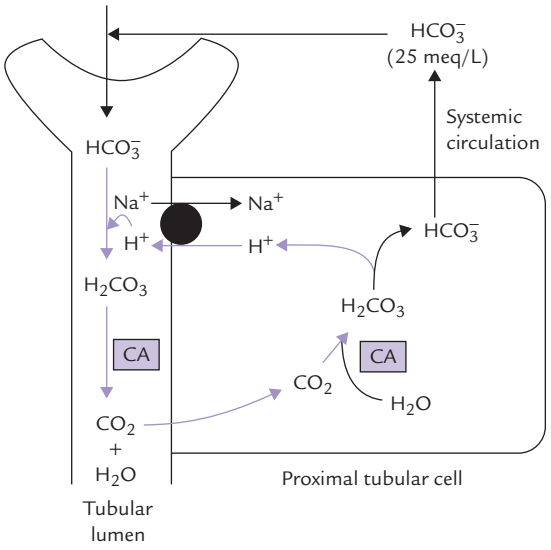


Fig. 1.7. Reclamation of bicarbonate by proximal tubular cell (CA = *carbonic anhydrase*, ● = $Na^+ - K^+$ antiport).

$$HCO_3^- + HA \longrightarrow H_2CO_3 + A^-$$

(The anions of these acids (A^-) pass on to Na^+, which is the predominant extracellular cation.)

Though this reaction averts immediate risk of a pH fall, it is accompanied by another imbalance, i.e. fall of bicarbonate ion concentration. The latter must be replenished in order to preserve the buffering capacity. Renal adjustments accomplish this by generating bicarbonate ions (see Fig. 1.7). Under normal circumstances, kidneys excrete about 50 mmol of H^+ in urine and simultaneously generate about 50 mmol of bicarbonate ions. However, in acidosis, kidneys can eliminate as much as 500 mmol of H^+ per day; in alkalosis excretion of H^+ falls markedly.

Ability of kidneys to excrete variable amounts of acids and to generate variable amounts of bicarbonate ions depending on body requirements, makes them act as **final defense** mechanism against any imminent pH alterations.

C. Mechanisms for Control of Bicarbonate

Renal tubular cells and erythrocytes possess a specific system, called *carbonate dehydratase* system, which enables them to use carbon dioxide for generating bicarbonate. Under normal circumstances, the erythrocyte mechanism merely makes fine adjustments in the plasma bicarbonate

concentration in response to specific body requirement while the kidneys play a major role in maintaining the bicarbonate concentration.

Note: This system is present in gastric parietal cells also.

Carbonate Dehydratase System

The principle enzyme of this system is *carbonic anhydrase,* which catalyzes the first reaction of the following chain:

$$CO_2 + H_2O \xrightleftharpoons{\text{Carbonic anhydrase}} H_2CO_3 \rightleftharpoons H^+ + HCO_3^-$$

Cells remove one of the final products of the chain, i.e. H^+. Consequently, both the reactions are pulled towards right to generate bicarbonate ions. For each H^+ removed from the cell, one bicarbonate ion is generated. Rise in intracellular concentration of H^+ and carbon dioxide accelerates these reactions. A fall in plasma concentration of bicarbonate also enhances the reaction sequence.

In a normal human being, at a plasma pCO_2 of 5.3 kPa (equivalent to CO_2 concentration of about 1.2 mmol/L), erythrocytes and renal tubular cells maintain the extracellular bicarbonate concentration at about 25 mmol/L. Thus, value of the extracellular $[HCO_3^-]/[CO_2]$ ratio (both in mmol/L) is 20 : 1. Taking pK′ value of carbonic acid to be 6.1, it can be calculated from the Henderson–Hasselbalch equation that this ratio represents a pH of about 7.4.

$$\underset{(7.4)}{pH} = \underset{(6.1)}{pK'} + \log \frac{[HCO_3^-](25\ mmol/L)}{[CO_2](1.2\ mmol/L)}$$

In practice, the *partial pressure of carbon dioxide (pCO_2)* is measured and its concentration in plasma is obtained by multiplying the pCO_2 with the solution constant for carbon dioxide.

$$\underset{(kPa)}{pCO_2} \times 0.23 = \underset{(mmol/L)}{[CO_2]}$$

where 0.23 is the **solution constant**. The equation can be rewritten as:

$$pH = pK' + \log \frac{[HCO_3^-]}{pCO_2 \times 0.23}$$

Role of Kidneys in Bicarbonate Homeostasis

Kidneys play an important role in bicarbonate homeostasis through the following mechanisms:

1. Reabsorption of the filtered bicarbonate, i.e. **bicarbonate reclamation.**
2. Generation of bicarbonate ions, termed **new bicarbonate generation.**

Both these actions depend on the *carbonate dehydratase* system. As indicated earlier, this system is operative in renal tubular cells and possesses considerable reserve capacity.

Bicarbonate Reclamation

The unmodified glomerular filtrate contains bicarbonate ions in the same concentration as plasma (i.e. 25 mmol/L). About 4000 mmol of ions are filtered by glomeruli each day. If all the filtered bicarbonate were lost in urine, severe bicarbonate deficit would soon develop. However, specific mechanisms exist in proximal tubular cells which permit conservation (reclamation) of all the filtered bicarbonate. The most important of these mechanisms is the Na^+–H^+ antiport system. This system operates in the renal tubular cells.

$Na^+ - H^+$ antiport system: Bicarbonate ions (HCO_3^-) and H^+ are produced by the *carbonate dehydratase* system, as shown in Figure 1.7. The H^+ ions are transported into the tubular lumen in exchange for Na^+ present in tubular fluid. The exchange occurs through mediation of the $Na^+ - H^+$ antiport. Within the tubular lumen, the H^+ reacts with the filtered bicarbonate to form carbon dioxide and water. This reaction needs *carbonic anhydrase*, located in the brush border of the proximal tubular cells.

The luminal membrane of the tubular cell allows the carbon dioxide to diffuse into the cell. At the same time, the bicarbonate, the other product of *carbonate dehydratase* system, moves across the *basolateral membrane* of the renal tubular cell to enter systemic circulation. Movement of bicarbonate in this direction is favoured by (a) increased intracellular concentration of bicarbonate, (b) electrical potential, and (c) impermeability of the luminal membrane to bicarbonate.

Through this process of reclamation the loss of the filtered bicarbonate ions in urine is prevented. These ions are reabsorbed, though indirectly, in the form of carbon dioxide; direct reabsorption is not possible since the luminal membrane is impermeable to bicarbonate ions. Further, this mechanism permits excretion of H^+ and prevents loss of Na^+ in urine.

Control: The *carbonate dehydratase* mechanism may be stimulated by a rise in pCO_2 or a fall of bicarbonate concentration within the tubular cells.

New Bicarbonate Generation

This mechanism operates in the later pat of distal convoluted tubules (DCT) and the collecting ducts. In these segments, H^+ secretion is associated with new bicarbonate generation (Fig. 1.8). The *carbonate dehydratase* system generates equimolar amounts of H^+ and bicarbonate. The hydrogen ions are secreted into the tubular lumen and combine with A^- (A^- represents a filtered anion of

an acid, e.g. lactate, acetoacetate or sulphate). HA thus formed is eliminated in urine.

The HCO_3^- ions, the other product of the *carbonate dehydratase* system, is absorbed into systemic circulation together with Na^+. Thus, this mechanism serves to increase alkali reserve of the body.

Urinary Buffers

Hydrogen ion excretion requires presence of suitable buffer systems in urine. The H^+ secreted into the tubular lumen causes acidification of urine. The minimum pH of urine is about 4.5 which is equivalent to 0.03 mEq/L of the H^+. At this H^+ concentration, sufficiently high concentration gradient is built up across the luminal cell membrane which prevents further secretion of H^+. Increase in the excretion of H^+ occurs only when the secreted H^+ is taken up by a luminal buffer system. The buffer takes up the secreted H^+, thus preventing the building of the concentration gradient. In normal urine, the most important buffer is **phosphate buffer** (Fig. 1.9a). **Ammonia,** produced in renal tubular cells, also acts as urinary buffer (Fig. 1.9b). Together, they permit excretion of 30–40 mmol of hydrogen ions every 24 hours.

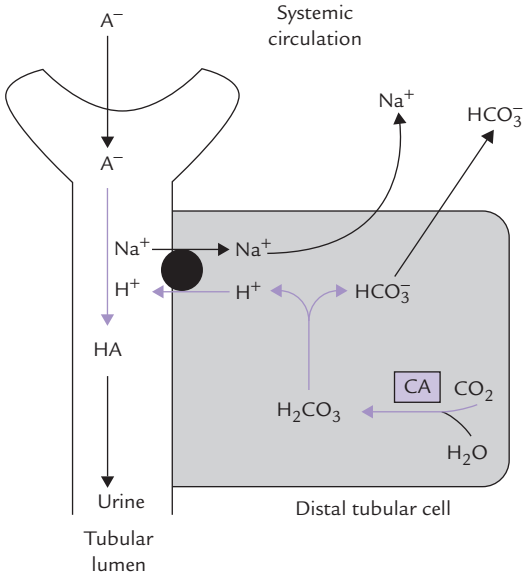


Fig. 1.8. New bicarbonate generation by distal tubular cell (● = $Na^+ - H^+$ antiport).

Phosphate is present in the glomerular filtrate; about 4/5th of it is in the form of HPO_4^{-2}. This combines with the secreted H^+ and is converted to the monovalent anion, $H_2PO_4^-$:

$$HPO_4^{-2} + H^+ \rightarrow H_2PO_4^-$$

Thus, HPO_4^{-2} and $H_2PO_4^-$, form a conjugate acid-base (i.e. buffer) pair.

$$pH = 6.8 + \log \frac{[HPO_4^{-2}]}{[H_2PO_4^-]}$$

where 6.8 is pK′ of this system.

As more H^+ is secreted, more and more of the divalent form is converted to monovalent form until, at pH below 5.5, most of this is in monovalent form.

Ammonia, the other important urinary buffer, is produced by deamination of glutamine in renal tubular cell (Fig. 1.9b).

$$\text{Glutamine} \xrightarrow{\textit{Glutaminase}} \text{Glutamate} + \text{Ammonia} + H^+$$

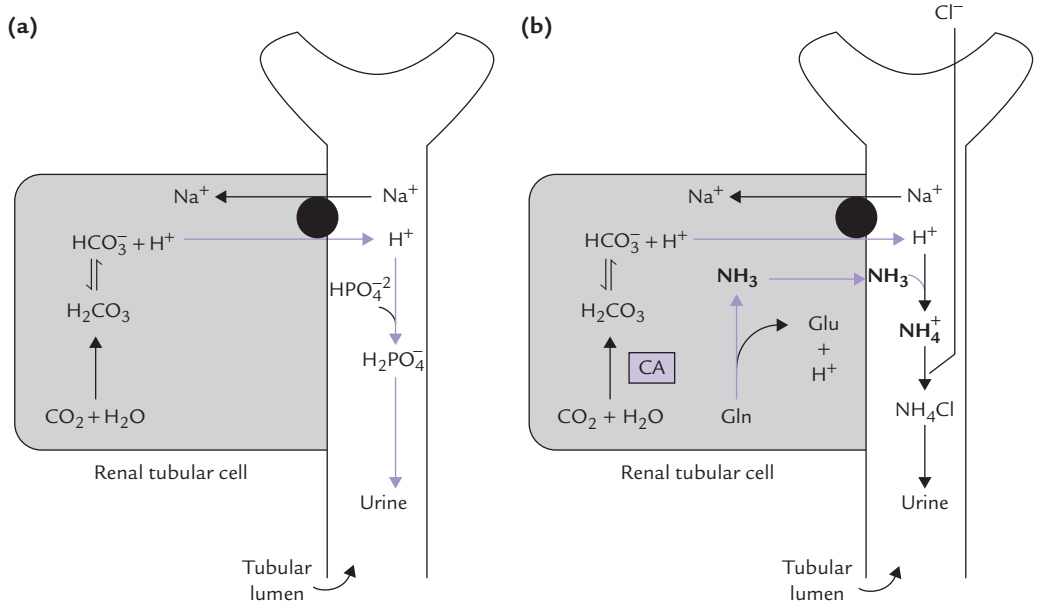


Fig. 1.9. Buffering of H^+ in urine. **(a)** Phosphate buffer, **(b)** Ammonia buffer.

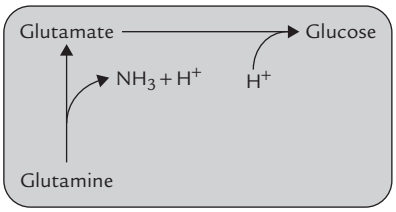


Fig. 1.10. Magnified view of a renal tubular cell showing that the H^+ produced in the cell during glutamine to glutamate conversion is incorporated in glucose.

Glutaminase, present in tubular cells is responsible for this conversion. Its activity is induced in chronic acidosis, thus permitting increased ammonia production, which can take up H^+ in tubular lumen. Ammonia and ammonium form a buffer pair

$$NH_3 + H^+ \rightarrow NH_4^+$$

$$pH = 9.8 + \log \frac{[NH_3]}{[NH_4^+]}$$

where 9.8 is pK′ of this system.

This buffer system allows increased hydrogen ion excretion via ammonium ions. Ammonia can readily diffuse across cell membranes, but ammonium ions cannot do so and are, therefore, eliminated in urine. As the urine becomes more acidic, it contains increasing amount of ammonium ions. The *ammonia-ammonium buffer pair becomes active only after the buffering power of phosphate is exhausted*.

Ammonia buffer produces an H^+ intracellularly during the conversion of glutamine to glutamate (Fig. 1.10). This apparently negates the effect of ammonia buffering system since an H^+ is generated in the cell for each H^+ that is buffered in the tubular lumen. Glutamate enters gluconeogenesis, a process requiring input of H^+. The H^+ generated earlier is probably incorporated in glucose during gluconeogenesis.

Role of Erythrocytes

Erythrocytes are also capable of generating bicarbonate ions, as shown in Figure 1.11.

Since erythrocytes lack aerobic pathways, they are incapable of generating substantial amount of carbon dioxide. The plasma carbon dioxide diffuses into the cell along a concentration gradient, where it generates bicarbonate ions through the *carbonate dehydratase* system. The other product, H^+ is buffered by haemoglobin. As the concentration of bicarbonate ions rises intracellularly, it diffuses into the extracelllular fluid along a concentration gradient. To maintain electroneutrality, diffusion of chloride ions occurs in the opposite direction, i.e. **chloride shift.**

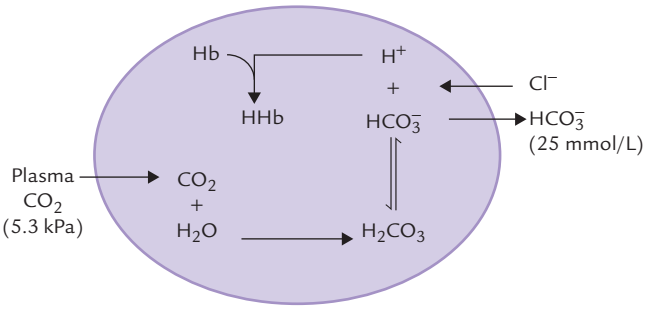


Fig. 1.11. Erythrocyte mechanism for bicarbonate generation (Hb = haemoglobin).

Note: Haemoglobin is an important blood buffer which works effectively in cooperation with the bicarbonate system.

$$pH = pK' + \log \frac{[Hb]}{[HHb]}$$

D. The Acid-Base Disorders

A number of acid-base disorders are known which are classified on the basis of changes in the components of **bicarbonate-carbonic acid buffer.** The three components (pH, HCO_3^- and H_2CO_3 are related as per the Henderson–Hasselbalch equation.

$$pH = pK' + \log \frac{[Base]}{[Acid]}$$

Base = bicarbonate ions; Acid = carbonic acid.
pK′ of carbonic acid = 6.1.

Considering the fact that plasma H_2CO_3 equilibrates with the dissolved as well as the gaseous carbon dioxide, the equation can be rewritten as:

$$pH = 6.1 + \log \frac{[HCO_3^-]}{[CO_2]}$$

$$= 6.1 + \log \frac{[HCO_3^-]}{[pCO_2 \times 0.23]}$$

where pCO_2 is the partial pressure of CO_2 and 0.23 is the constant. (Concentrations of HCO_3^- and CO_2 are measured in mmol/L, and pCO_2 is expressed in kPa.)

From this equation, it may be observed that decreased pH value (i.e. **acidosis**) can result mainly due to two reasons: (a) primary decrease in bicarbonate ion (HCO_3^-) concentration, and (b) primary increase in pCO_2. The primary decrease of bicarbonate ions $[HCO_3^-]$, is called *metabolic acidosis* and primary increase of pCO_2 is called *respiratory acidosis*. Conversely, increased pH, i.e. ***alkalosis*** may result either due to primary increase in HCO_3^- concentration (*metabolic alkalosis*), or to primary decrease of pCO_2 (i.e. *respiratory alkalosis*).

Disturbances in the blood buffer system can lead to conditions known as acidosis, with a pH as low as 7.1, or alkalosis, with a pH as high as 7.6.

Acidosis

Metabolic Acidosis

A disorder resulting from a *primary fall in bicarbonate level* (< 24 mmol/L) resulting in a decreased pH of arterial plasma (because of decreased ratio for bicarbonate concentration to pCO_2 in the Henderson–Hasselbalch equation) is metabolic acidosis

$$\mathbf{pH}\downarrow = 6.1 + \log \frac{[HCO_3^-]\downarrow}{pCO_2 \times 0.23}$$

Causes: Some common causes of bicarbonate depletion are given below:

1. *Excessive production of organic acids:* When an acid metabolite (HA) enters blood, it is taken up by the base component of the bicarbonate buffer (i.e. HCO_3^-). Consequently, the concentration of these ions falls, giving rise to a bicarbonate deficit, therefore metabolic acidosis occurs. A mild tendency of acidosis produced in this manner is almost a normal occurrence. This is because several body reactions continually produce non-volatile acids. Some examples include oxidation of the –SH group of cysteine to sulphate and hydrolysis of phosphates. However, these reactions do not cause any serious disorder in the body because the acids are neutralized by bicarbonate buffer. The mild bicarbonate deficit so produced is promptly reverted by the renal and the pulmonary mechanisms. In other words, the amount of acid generated is too small to override body's defenses. On the other hand, certain processes release acids into the blood in very large quantities which cannot be taken care of by body's defenses. As a result, **severe bicarbonate deficit** leading to a fall in pH occurs (Case 1.1).

   Most common processes that release acids are **lactic acidosis** and **ketoacidosis**.

   Lactic acidosis arises due to generation of excessive lactic acid from anaerobic glycolysis, e.g. during intense muscular exercise.

   Ketoacidosis results due to excessive production of ketone bodies (acetone, acetoacetate and β-hydroxybutyrate) in liver under certain abnormal metabolic states such as starvation and uncontrolled diabetes mellitus.

   In addition to acute acidosis causing conditions mentioned above, certain conditions lead to the development of **chronic acidosis**. These are: *phenylketonuria, hyperammonaemias* or *maple syrup urine disease*. But in these disorders acidosis is less important than other metabolic consequences.
2. *Excessive loss of HCO_3^- from the body:* Duodenal fluid, with a bicarbonate concentration about twice that of plasma, is alkaline. Loss of this fluid may occur through intestinal fistula. If the rate of loss exceeds that of the renal ability to regenerate bicarbonate ions, the plasma bicarbonate may fall enough to cause acidosis. For the same reason, excessive loss of bicarbonate ions in severe diarrhoea causes acidosis.
3. *Generalized renal tubular dysfunction:* It may cause loss of a mixture of ions in urine. Renal bicarbonate reclamation and generation are also impaired due to damaged Na^+–H^+ antiport.
4. *Poisoning or overdose:* Production of acid metabolites in poisoning and overdoses also lead to metabolic acidosis. For example, in methanol poisoning, formate is produced; and in ethylene glycol poisoning, production of oxalate occurs.
5. *Increase in chloride ion:* In the acidosis-causing conditions discussed so far, no change in the concentration of chloride ion occurs. However, in some conditions, such as renal tubular acidosis and acetazolamide therapy, fall of plasma bicarbonate concentration is accompanied by a simultaneous increase in chloride ion concentration. In short, bicarbonate is substituted by chloride and a state of **hyperchloraemic acidosis** results.

Overproduction of acidic metabolites, e.g. ketone bodies in uncontrolled diabetes mellitus or lactic acid during severe exercise, are the two commonest causes of metabolic acidosis.

Compensation of metabolic acidosis: In metabolic acidosis, the primary bicarbonate deficit results in the fall of bicarbonate to pCO_2 ratio. Thus, the compensatory mechanisms are aimed at elevation of this ratio (a normal ratio of 20 is crucial for maintaining the blood pH at 7.4). This is accomplished by **compensatory hyperventilation** which washes off carbon dioxide to decrease pCO_2. Consequently, the above-stated ratio is elevated.

Usually, the metabolic acidosis is because of excessive acid (HA) production. Acceleration of the following reaction occurs under such circumstances:

$$HCO_3^- + HA \rightleftharpoons H_2CO_3 + A^-$$

The H_2CO_3 so produced dissociates to water and carbon dioxide. Though carbon dioxide increases temporarily in this manner, it induces hyperventilation which

washes off the CO_2. Consequently, the pCO_2 (and hence $[CO_2]$) tends to fall, which reverts the ratio of HCO_3^- : pCO_2 to normal. In this manner, hyperventilation serves as the major compensatory process in metabolic acidosis. But the pulmonary compensation is said to be incomplete because (a) concentration of HCO_3^- gets reduced because there is no new generation of HCO_3^- consumed earlier, and (b) the anion of the acid metabolite (A^-) is still present in extracellular fluid.

These abnormalities must be corrected for complete restoration of the acid-base status. The bicarbonate concentration is self-correcting: decreased bicarbonate concentration causes stimulation of renal *carbonate dehydratase*, which in turn enhances bicarbonate reclamation and bicarbonate generation. Consequently, the depleted bicarbonate concentration slowly rises towards normal. The anion of the acid metabolite (A^-) is eliminated by renal excretion. In some cases, the acid metabolite (e.g. *lactate*) is slowly metabolized. With these processes, the compensation is complete and acid-base status returns to normal.

Plasma findings in metabolic acidosis: These are as follows:

1. **Bicarbonate concentration** is always low.
2. **pCO_2** is usually low (compensatory change).
3. **pH** is low in uncompensated (or partially compensated) cases and normal in fully compensated cases.
4. **Concentration of chloride** ions is normal in most cases. However, raised value of this ion is seen in renal tubular acidosis or acetazolamide therapy.

Anion gap: Cause of metabolic acidosis is usually apparent from clinical history of the patient. Knowledge of anion gap serves as an additional tool in delineating the cause. Anion gap is estimated by measuring the difference between the sums of the concentrations of principal cations (Na^+ and K^+) and principal anions (Cl^- and HCO_3^-).

$$([Na^+] + [K^+]) - ([Cl^-] + [HCO_3^-])$$

Average reference values of these ions are: Na^+, 140; K^+, 4; Cl^-, 100; and HCO_3^-, 25 (all values are in mmol/L). Therefore, in healthy individuals the anion gap has an average value of **19 mmol/L**.

The **anion gap** refers to the discrepancy between the measured electrolytes, but in reality no such gap exists. It merely represents unmeasured net negative charge on plasma proteins. In short, the negatively charged amino acid side chains on the proteins account for the apparent discrepancy between the measured electrolytes.

In metabolic acidosis, deviations from the above-stated normal values occur. The anion gap can increase or remain normal depending on the underlying cause of metabolic acidosis.

Increased anion gap: When excessive production of acids is the cause of metabolic acidosis, concentration of HCO_3^- decreases but that of Cl^- remains unaffected. Consequently, anion gap is increased.

Normal anion gap: In renal tubular acidosis or acetazolamide therapy, fall of bicarbonate is accompanied by increase in chloride ion concentration. Hence, anion gap does not change in these conditions, which are, therefore, called **normal anion gap acidosis** or **hyperchloraemic acidosis**.

Respiratory Acidosis

It results from a decrease in alveolar ventilation, causing decreased elimination of CO_2 by the lungs. The consequent **increase in pCO_2** is the **primary event** in the acid-base disturbance.

$$\mathbf{pH} \downarrow = 6.1 + \log \frac{[HCO_3^-]}{\uparrow \mathbf{pCO_2} \times 0.23}$$

Decreased alveolar ventilation, may result in a number of conditions (Table 1.5).

Compensation of respiratory acidosis: In respiratory acidosis, the ratio of HCO_3^- : pCO_2 is reduced, and so the role of compensatory mechanisms is primarily to retain more of bicarbonate ions. In contrast to metabolic acidosis, where respiratory compensation is important, in this type of acidosis the primary defect being in lungs, the physiological response of lungs cannot be expected to play any significant role. **Renal response** plays an important role in bringing the acid-base status back to normal. *Acceleration of the carbonate dehydratase mechanism* in erythrocytes is also important for the short-term compensation.

Renal response: In respiratory acidosis, the renal mechanism compensates for the disturbance by generating

Table 1.5. Causes of respiratory acidosis

Airway obstruction	Respiratory centre depression
Asthma	**Sedatives**
Chronic obstructive airway disease	General anaesthesia
Pulmonary diseases	**Neuromuscular diseases**
Respiratory distress syndrome	Tetanus
Severe pneumonia	Neurotoxins
Thoracic diseases	
Kypnoscoliosis	
Flail chest	

HCO_3^- ions and by eliminating H^+ (Figs 1.7 and 1.8). The *carbonate dehydratase* mechanism in renal tubular cells is accelerated by increased pCO_2. As a result, most (or all) of the filtered HCO_3^- is reabsorbed and more of new HCO_3^- is generated. The urine becomes acidic as more of H^+ is secreted. Renal response develops slowly. In cases of chronic carbon dioxide retention renal compensation is maximum.

The H^+ secreted during respiratory acidosis is exchanged with the luminal Na^+, thus preventing urinary loss of the latter.

Plasma findings in respiratory acidosis: The biochemical changes in acute and chronic respiratory acidosis differ from each other (Table 1.6). This is because of the fact that in acute cases there is insufficient time for the renal mechanisms to generate HCO_3^- and add them into plasma. Slight rise in plasma [HCO_3^-] that occurs in acute stages is mostly derived from erythrocytes. This mild degree of compensation is inadequate to prevent fall in pH. In chronic respiratory acidosis, on the other hand, the renal mechanisms are sufficiently speeded up to elevate the levels of HCO_3^- to such extent that pH fall is averted.

Alkalosis

The body's defenses against alkalosis are less effective than those against acidosis. This is because the biological systems have little capacity to buffer the OH^- ions, and the bicarbonate buffer is far less effective when the pH rises above 7.4. Fortunately, alkalosis is far less common than acidosis because the body metabolites are acidic, not alkaline.

Metabolic Alkalosis

It is characterized by an increased serum bicarbonate concentration as a result of either an excessive loss of acid or abnormal retention of bicarbonate. In accord with the Henderson–Hasselbalch equation, the primary rise in bicarbonate results in increase in pH.

$$\mathbf{pH} \uparrow = 6.1 + \log \frac{[HCO_3^-]\uparrow}{pCO_2 \times 0.23}$$

Causes: Some of the major causes of metabolic alkalosis are:

1. *Excessive renal generation of bicarbonate:* In distal part of the renal tubules (distal convoluted tubules), K^+ and H^+ ions compete with each other for secretion into the tubular lumen in exchange for Na^+ (Fig. 1.12). Because of this reciprocal relation, changes in potassium-balance influence the H^+ balance. If the plasma potassium level is low (i.e. hypokalaemia), less of this ion is available intracellularly for secretion. Hence, more of H^+ ions are secreted and the urine becomes acidic. However, H^+ secretion is accompanied by equimolar generation of HCO_3^-. The latter passes into ECF to raise the HCO_3^- concentration, and therefore, the extracellular pH rises. This type of alkalosis is termed **hypokalaemic alkalosis**.
2. *Loop diuretics*: These diuretics inhibit the pumping of sodium in the loop of Henle and, therefore, increase sodium load on DCT and collecting ducts. This results in increased Na^+ – K^+ exchange at these sites resulting in excessive K^+ loss and hence hypokalaemia. Thus, in this case *alkalosis is a consequence of hypokalaemia*.

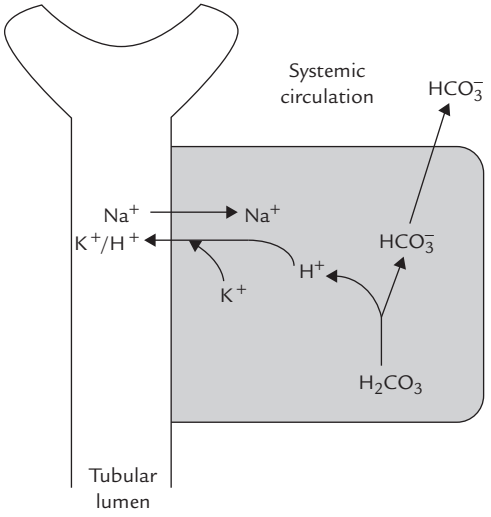


Fig. 1.12. Secretion of K^+ and H^+ in exchange for Na^+ in distal convoluted tubules.

Table 1.6. Summary of biochemical findings in arterial blood in various types of acidosis and alkalosis

Parameter	Acidosis			Alkalosis		
	Metabolic	Respiratory		Metabolic	Respiratory	
		Acute	Chronic		Acute	Chronic
pH	Low	Low	Normal or slightly low	High	High	Slightly high or normal
$\|HCO_3^-\|$	Low	Slightly high	High	High	Slightly low	Low
pCO_2	Low	High	High	High	Low	Low

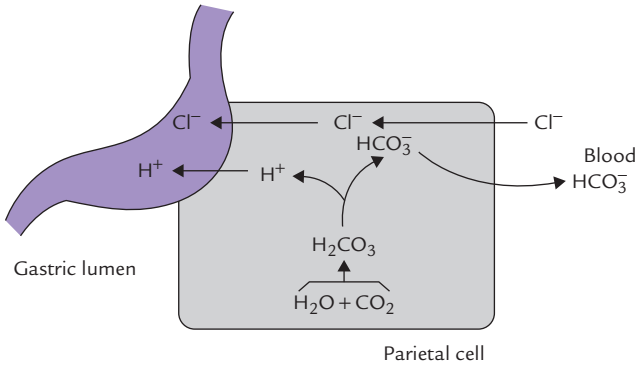


Fig. 1.13. Mechanism of secretion of H^+ and Cl^- by parietal cell into gastric lumen.

3. *Pyloric stenosis:* In this condition, excessive loss of H^+ and Cl^- occurs in vomitus. This necessitates increased secretion of H^+ into the gastric lumen by the parietal cells. As shown in Figure 1.13 secretion of H^+ and Cl^- is accompanied by the generation of equivalent amount of HCO_3^- by these cells. The HCO_3^- ions, thus generated, move into the extracellular fluid resulting in elevated pH.

   The narrowing at pylorus has further hazardous consequences: (a) it prevents loss of alkaline duodenal contents, which could have served as a safeguard against alkalosis, and (b) it causes volume depletion (hypovolumia), which is a known cause of alkalosis, as noted below.
4. *Hypovolumia:* It triggers a series of events that ultimately result in increased secretion of aldosterone from adrenal cortex. Since aldosterone causes K^+ elimination through urine, hypokalaemia results. Hypokalaemia in turn leads to alkalosis.
5. *Alkali ingestion:* Massive quantities of ingested alkali cause alkalosis. Inappropriate treatment of acidotic conditions and chronic alkali ingestion can lead to this state.

Compensation of metabolic alkalosis: The compensatory mechanism of metabolic alkalosis must bring down the ratio of HCO_3^- : pCO_2. This occurs by *respiratory response* which involved depression of the respiratory centre by elevated plasma HCO_3^- concentration. It leads to underbreathing; and thereby tends to build up the pCO_2 (Case 1.2). This depresses above-mentioned ratio, which in turn leads to decrease of the elevated pH towards normal. However, this response has limited capacity because underbreathing leads to oxygen deficit. Moreover, the alkaline state is not capable of sufficiently depressing the respiratory centre.

Respiratory Alkalosis

The primary abnormality in respiratory alkalosis is the **decrease of arterial pCO_2** and, therefore, rise of pH.

$$\mathbf{pH} \uparrow = 6.1 + \log \frac{[HCO_3^-]}{\downarrow \mathbf{pCO_2} \times 0.23}$$

Causes: Fall in the arterial pCO_2 is due to hyperventilation resulting from tissue hypoxia, high altitude, anaemia or bacterial sepsis. Hysterical overbreathing, assisted breathing, or certain nervous system disorders (encephalitis, head injury) may also lower pCO_2 and hence cause alkalosis.

Decreased ventilation in obstructive lung diseases prevents efficient expiration of carbon dioxide to cause respiratory acidosis, and hyperventilation accelerates the loss of carbon dioxide and causes respiratory alkalosis.

The **compensation** is aimed at lowering HCO_3^- concentration, which will revert the HCO_3^- to pCO_2 ratio (thereby causing correction of pH). The compensatory mechanism is triggered by decreased pCO_2 which slows down the *carbonate dehydratase* mechanism in renal tubular cells and RBCs. Consequently, decrease in the HCO_3^- reclamation and generation occurs. The compensation is slow to develop, like in the case of respiratory acidosis. If a steady low pCO_2 is maintained, maximal compensation develops within 36–72 hours.

The biochemical findings in arterial blood in various types of acidosis and alkalosis are summarized in Table 1.6.

Exercises

Essay type questions

1. What is role of *carbonic anhydrase* system and renal mechanisms in the maintenance of blood pH?
2. Describe the various mechanisms by which are responsible for maintenance of pH of the body fluids.
3. Classify the acid-base disorders. Discuss their causes and characteristic biochemical parameters.
4. Mention the biochemical determinants of metabolic and respiratory acidosis. Why is bicarbonate buffer system especially important as a blood buffer?
5. Describe how excess acid (H^+) is excreted chiefly in urine, highlighting role of urinary buffers.

Write short notes on

1. Chloride shift
2. Anion gap
3. Phosphate buffer
4. Respiratory acidosis
5. Titration curve of weak acid

CLINICAL CASES

CASE 1.1 A semiconscious man, bleeding profusely

A 28-year-old man was brought to the hospital emergency in a shocked state. He was well known to the department for frequent admissions related to ethanol abuse. He was bleeding profusely from arms and hands, and smelt alcohol. Apparently, he had a fall while trying to board a moving bus. Breathing was fast and shallow (respiratory rate 48 per minute) and the pulse was fast and irregular at 120 beats per minute. Examination showed fast and shallow pulse and there were cardiac arrhythmias. There was no evidence of head injury or any major organ damage.

Blood sample was sent for laboratory analysis. The following derangements in certain acid-base parameters were detected:

Investigations	Patient's report	Reference range
Blood pH	7.05	7.35–7.45
pCO_2	3.6 kPa	4.7–6.0 kPa (36–46 mmHg)
$[HCO_3^-]$	17 mmol/L	21–28 mmol/L
Serum lactate	6.4 mmol/L	0.4–1.4 mmol/L
Serum pyruvate	160 μmol/L	40–80 μmol/L

Q.1. What is the acid-base disturbance present in this patient?
Q.2. What could be the cause of acid-base derangement observed?
Q.3. What would be the differential diagnosis?
Q.4. Mention the cause of cardiac arrhythmias.

CASE 1.2 A 32-year-old man with persistent vomiting

A 32-year-old man reported in hospital emergency with complaints of persistent vomiting for one week. He had generalized muscular cramps. On examination he appeared dehydrated and had shallow respiration.

Blood sample was analyzed with the following results:

Investigations	Patient's report	Reference range
$[HCO_3^-]$	38 mmol/L	21–28 mmol/L
$[H^+]$	28 nmol/L	36–44 nmol/L
pCO_2	8.6 kPa	4.7–6.0 kPa (36–46 mm Hg)
Serum urea	64 mg/dl	15–45 mg/dl
Serum creatinine	1.8 mg/dl	0.6–1.4 mg/dl
Na^+	148 mmol/L	135–145 mmol/L
K^+	2.9 mmol/L	3.6–5.0 mmol/L

Q.1. Identify the nature of acid-base disturbance.
Q.2. What could be the cause of this acid-base disorder?
Q.3. Comment on the other biochemical test results.
Q.4. Give the reason for the development of muscle cramps.
Q.5. Suggest the line of treatment.

CHAPTER

2

CHEMISTRY OF CARBOHYDRATES

Carbohydrates are the most abundant organic molecules in nature and the major functional constituents of living cells. The primary source of energy in animal cells, carbohydrates are synthesized in green plants from carbon dioxide, water and solar energy. The term carbohydrate was coined by Karl Schmidt in the mid 19th century, which literally means a **hydrate of carbon:** a compound with an empirical formula $(CH_2O)_n$, where n is an integer of 3 or greater. Structurally carbohydrates have two important features:

1. more than one **hydroxyl group** (polyhydroxyl).
2. a **carbonyl group** as either an aldehyde group (–CHO–) or a ketone group (–CO–).

Presently it is known that carbohydrates may contain elements other than C, H, and O (e.g. N, S, and P). The hydroxyl groups may be free or substituted, and the carbonyl groups may be present in the free reducing form or in non-reducing form in glycosidic linkages. Moreover, a large number of compounds are classified as carbohydrates even though they do not have the above empirical formula (e.g. deoxyribose, $C_5H_{10}O_4$).

This chapter gives a brief account of general characteristics, chemistry, and functions of carbohydrates.

After going through this chapter, the student should be able to understand:

- Chemistry, properties, structures, biological significance and classification of carbohydrates; stereo-chemical and reducing properties; isomers, anomers and epimers, the common monosaccharide derivatives and their occurrence.
- Nature and type of glycosidic linkage that occur in common disaccharides and homopolysaccharides; and occurrence and role of these compounds.
- Structure and role of glycosaminoglycans and various types of mucopolysaccharidoses.

I. Biological Significance of Carbohydrates

Carbohydrates are widely distributed in nature, and they perform a vast range of functions. They play a primary role in energy metabolism. For example, glucose is the chief fuel molecule in all tissues and organs. Brain, spinal cord, peripheral nerves and erythrocytes are almost exclusively dependent on glucose. When in surplus, carbohydrates are stored as glycogen or starch.

Besides meeting the major energy requirements of the human body, carbohydrates perform other important roles also. They are components of nucleic acids and are covalently linked with lipids and proteins. Glycoproteins, hybrids of carbohydrates and proteins, form blood group substances and some hormones, and provide recognition elements on cell membranes. As mucopolysaccharides, carbohydrates provide the structural framework for the tissues and organs of the human body and serve as lubricants and support elements of connective tissue. Glycolipids (sphingosine derivatives) are important membrane constituents. Finally, carbohydrates form structural basis of some intracellular messengers, and play a role in intercellular communication.

Table 2.1. Structure and nomenclature of monosaccharides

Number of carbon atoms	Generic name	Molecular formula	Functional group	
			Aldehyde	Ketone
3	Triose	$C_3H_6O_3$	Glyceraldehyde	Dihydroxyacetone
4	Tetrose	$C_4H_8O_4$	Erythrose	Erythrulose
5	Pentose	$C_5H_{10}O_5$	Ribose	Ribulose
6	Hexose	$C_6H_{12}O_6$	Glucose	Fructose
7	Heptose	$C_7H_{14}O_7$	Glucoheptose	Sedoheptulose

II. Classification of Carbohydrates

Carbohydrates may be classified as monosaccharides, disaccharides, oligosaccharides and polysaccharides (the term saccharide is derived from Greek word for sugar).

A. Monosaccharides

The ***monosaccharides*** *are the simplest forms consisting of 3 to 9 carbon atoms, which serve as the building blocks of all carbohydrates.* They are sub-classified on the basis of (a) number of carbon atoms, e.g. trioses (C-3), tetroses (C-4), pentoses (C-5), and hexoses (C-6), (b) type of the functional group, which is aldehyde in aldoses and ketone in ketoses (Table 2.1, Fig. 2.1).

Monosaccharides are linked by covalent linkages (i.e. **glycosidic linkages**) to yield larger structures such as disaccharides, oligosaccharides and polysaccharides.

Monosaccharides of greatest biological importance are pentoses, which constitute a part of the nucleic acids (ribose and deoxyribose), and the hexoses, which serve a variety of physiological roles, discussed later.

B. Disaccharides

The ***disaccharides*** *consist of two monosaccharide units linked by a glycosidic bond.* They yield two monosaccharide units upon hydrolysis. For example, maltose, a disaccharide, can be cleaved into two glucose molecules. Similarly, lactose is hydrolyzed into galactose and glucose, and sucrose into glucose and fructose.

C. Oligosaccharides

Three to ten monosaccharide units, linked by glycosidic linkages, make a molecule of ***oligosaccharide***. For example, maltriose consists of three glucose molecules. A clinically important series of oligosaccharides are blood group substances A, B and H.

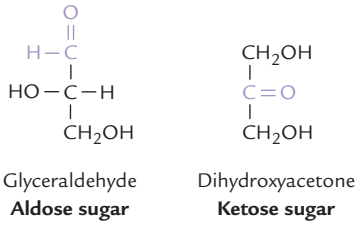


Fig. 2.1. Structures of the trioses. The carbonyl groups are in colour.

D. Polysaccharides

Usually the ***polysaccharides****, also known as glycans, contain hundreds of covalently linked monosaccharide units.* Carbohydrates are generally found in nature in the form of polysaccharides. Some examples are cellulose, starch, chitin, mucopolysaccharides, etc.

Monosaccharides, the simplest carbohydrates, are classified as aldoses or ketoses depending on the presence of aldehyde or keto group. Disaccharides contain two, oligosaccharides contain 3–10 and polysaccharides consist of more than 10 monosaccharides linked by glycosidic bonds.

III. Structural Properties

The structures of glucose and fructose shown in Figure 2.2 are referred to as **open chain** structures. The carbons are numbered, starting with the aldehyde carbon for aldoses, or with the terminal carbon closest to the keto carbon for ketoses. The open chain form accounts for some of the properties of the monosaccharides, as described below.

A. Isomerism

Different compounds that have the same molecular formula are called ***isomers*** *of one another,* e.g. glucose and fructose ($C_6H_{12}O_6$) are aldose-ketose isomers (Fig. 2.2). Similarly

glyceraldehyde and dihydroxyacetone ($C_3H_6O_3$) are aldose-ketose isomers (Fig. 2.1).

B. Asymmetric Carbon Atom

The carbon atom to which four different substituent groups are attached is called a **chiral** or **asymmetric carbon atom**. The second, third, fourth and fifth carbons of glucose are, therefore, asymmetric (Fig. 2.2). Presence of an asymmetric carbon imparts two important properties to the molecule: stereoisomerism and optical activity.

Stereoisomerism

The stereoisomers have same structural formula but arrangement of the substituent groups around the asymmetric carbon atoms is different. Number of the possible stereoisomers is given by 2^n, where n is the number of asymmetric carbon atoms. For glyceraldehyde, only two stereoisomers are possible, whereas for glucose, which contains four asymmetric carbon atoms (see Fig. 2.2), the number of stereoisomers is $2^4 = 16$.

Optical Activity

When a plane polarized light is passed through a sugar solution, it may undergo a rotation towards the right or the left direction. The sugars causing the rightward rotation are called *dextrorotatory* (+), whereas those causing the rotation towards the left are called *levorotatory* (–). These two forms are referred to as the **optical isomers**. **The optical isomers of a given sugar have identical chemical properties.**

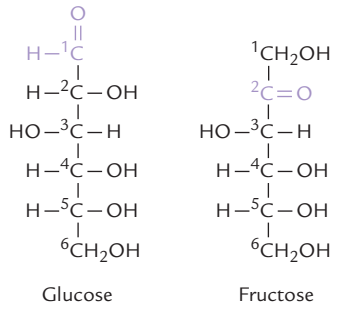


Fig. 2.2. Structures of common hexoses in D-configuration (open chain).

C. D and L Forms

A special type of *isomerism is found in the pairs of structures that are mirror images of each other. These mirror images are called* ***enantiomers****; the two members of a pair are designated as D and L sugars.* They are related to each other like the left hand and the right hand, and therefore, this phenomenon, which seems to offer handedness to a molecule, is called **chirality** (Greek: *"kheir"*, meaning hand). Enantiomers of glucose and fructose are shown in Figure 2.3. Physical and chemical properties of enantiomers of a given sugar are identical, except their optical rotations: +113° for D- and –113° for L-glucose.

Glyceraldehyde serves as a means of classifying all monosaccharides into the D- and L-series. If the penultimate (next to last) carbon of a sugar has the configuration of D-glyceraldehyde (with hydroxyl group on right-hand side; Fig. 2.3), then it belongs to the D-series. A similar relationship exists between L-glyceraldehyde (with hydroxyl group on left-hand side) and the L-series of monosaccharides. With a few exceptions, the *sugars appearing in human metabolism are of the D type.*

The D and L terminology refers to the structure based on the D and L forms of glyceraldehydes, not on how the molecule rotates polarized light. The D and L isomers are mirror images or enantiomers.

D. Ring Structures of Monosaccharides

Monosaccharide molecules of 4, 5 or 6 carbons are quite flexible, and this flexibility brings the aldehyde group (or the keto group) in close proximity to other hydroxyl groups on the same molecule. Reaction of aldehyde and

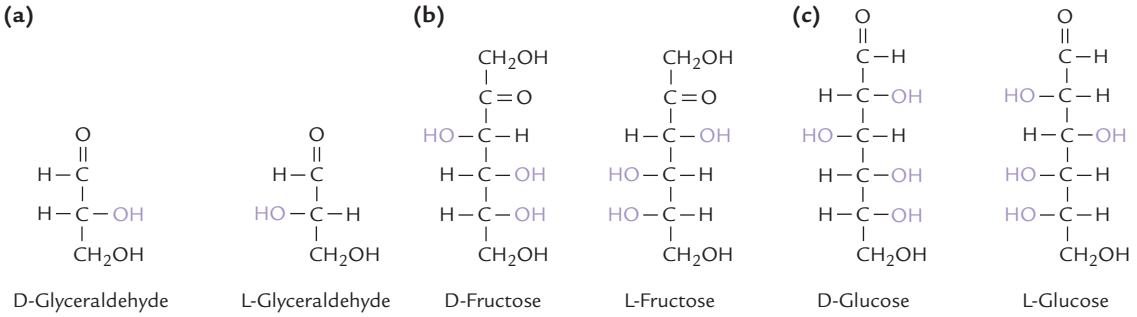
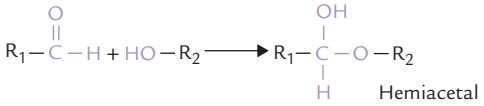
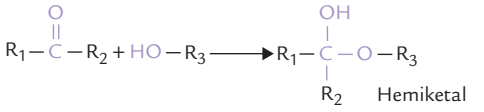


Fig. 2.3. Enantiomers (D- and L-isomers) of **(a)** glyceraldehyde, **(b)** fructose, and **(c)** glucose.

ketone with the hydroxyl group may then occur to form **hemiacetal** and **hemiketal** respectively. This results in cyclization of the linear forms to stable ring structures. The cyclization is a spontaneous process in solutions.





Pyranose and Furanose Ring Structures

If the ring structure formed by cyclization is six-membered (made of 5 carbons and 1 oxygen), it is called a **pyranose ring**; if it is five-sided (made of four carbons and one oxygen), it is called a **furanose ring** (Fig. 2.4). Fructose, a ketohexose, is usually found in the furanose form, called **fructofuranose**, whereas glucose is most stable in the pyranose form. More than 99.7% of the total glucose molecules in solution exist in pyranose form, and termed **glucopyranose**.

Tetroses or larger monosaccharides cyclize by reaction of the aldehyde or ketone group with a hydroxyl group on another carbon atom to form either a 6-membered pyranose ring or 5-membered furanose ring.

Anomers

Formation of ring structure results in creation of an additional asymmetric carbon, called **anomeric carbon** (C-1 of an aldose, C-2 of a ketose). For an aldohexose, such as glucose, the number of asymmetric carbon atoms increases to 5 and the number of possible isomers, therefore, increases to $2^5 = 32$. Isomerism as a result of the anomeric carbon in glucose is illustrated in Figure 2.5, where the glucose is shown to be in a pyranose ring form (D-glucopyranose). It may assume either the α- or β-configuration about the anomeric carbon. Note that these two forms differ in the arrangement of the substituents around the carbonyl carbon and are called **anomers**. Their systematic names are α-D-glucopyranose and β-D-glucopyranose respectively. Together they make up well over 99% of the total sugars present in solution, only about 0.0025% being in the open chain form. Both anomers are present in equilibrium with the open chain (non-cyclic aldehyde) form.

Most monosaccharides spontaneously produce cyclic hemiacetal or hemiketal forms, which have either α or β conformation at their anomeric carbon.

It is important to note that cyclic forms of aldohexoses, with their five asymmetric carbons have total of 32 (2^5) stereoisomers; and that each of the 16 isomers that belong to D- or L-series has two anomeric forms.

Mutarotation

The anomers are not stable under ordinary conditions and tend to interconvert constantly. The interconversion of the two anomeric forms is referred to as mutarotation (Fig. 2.5). It is caused by spontaneous opening and reclosure of the ring. In case of glucopyranose, the equilibrium between the α- and β-anomers favours the β form (63.6%) over the α form (36.4%). The equilibrium is

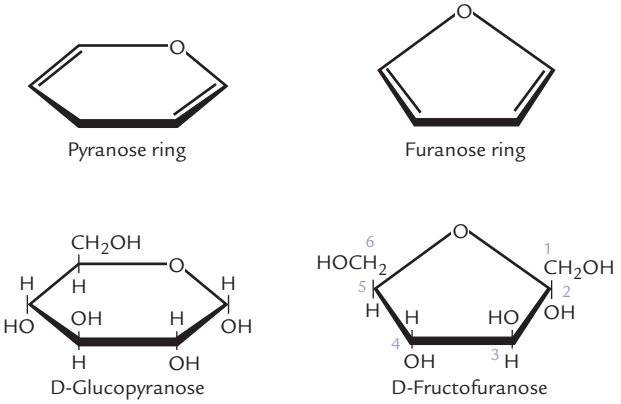


Fig. 2.4. Pyranose and furanose forms of glucose and fructose. Glucopyranose is formed by reaction of aldehyde at C-1 with hydroxyl group at C-5; and fructofuranose is formed by reaction of ketone at C-2 with hydroxyl group at C-5.

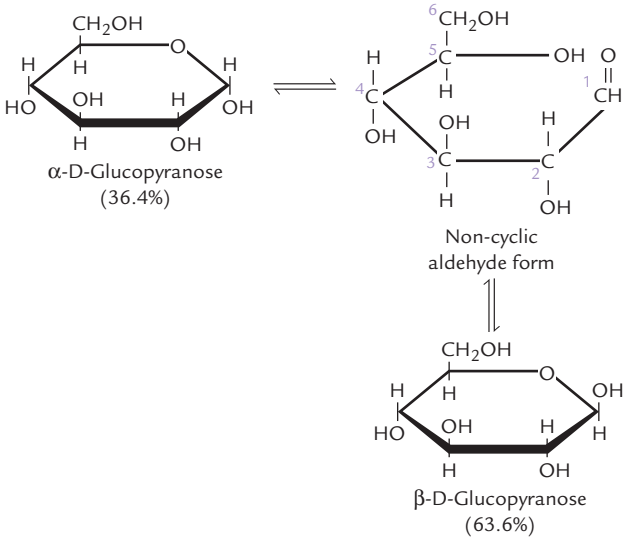


Fig. 2.5. Mutarotation: interconversion of α-D-glucopyranose and β-D-glucopyranose.

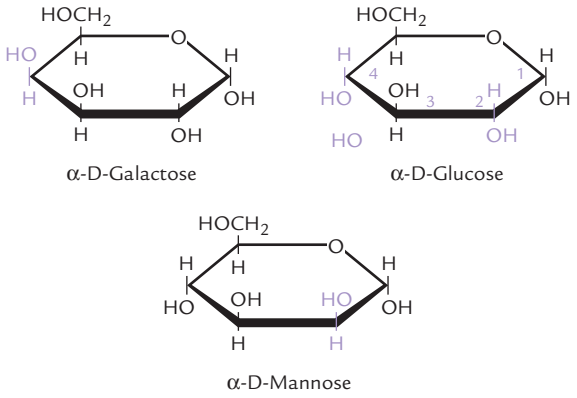


Fig. 2.6. Epimers of glucose: galactose (C-4 epimer) and mannose (C-2 epimer), shown as Howarth projections.

reached within several hours in neutral α- or β-solutions, but mutarotation is accelerated greatly in the presence of acids.

Mutarotation is followed in a *polarimeter* by measuring the optical rotation. The α-D-glucopyranose has a specific rotation of +112.2° and the α-D-glucopyranose of +18.7°. Both undergo mutarotation and over a period of a few hours, the specific rotation changes to attain a stable value of +52°. This change represents the interconversion of the two anomeric forms to yield the equilibrium mixture, consisting of about two-third of the α-form, one-third of the β-form, and a very small amount of the non-cyclic form.

α and β forms interconvert via the open chain form (mutarotation).

E. Diastereomers and Epimers

Figure 2.6 shows structures of three sugars, which differ in the arrangement of substituents at one or more (but not all) chiral carbons. They are **diastereomers**. Glucose and mannose differ in the orientation of substituents around only one of their asymmetric carbons, and are called **epimers**. Epimers have different physical and chemical properties. Just as mannose differs from glucose on the basis of carbon 2 configuration, another epimer of glucose is galactose, which differs on the basis of carbon 4 configuration. Thus mannose is termed C-2 epimer of glucose and galactose as a C-4 epimer of glucose.

Fischer and Howarth projections: A common way of depicting cyclic structures is shown in Figure 2.7 (these are referred to as **Fischer structures**). A closer look at them shows that some bonds are short but others are excessively long. But this cannot be the case in nature.

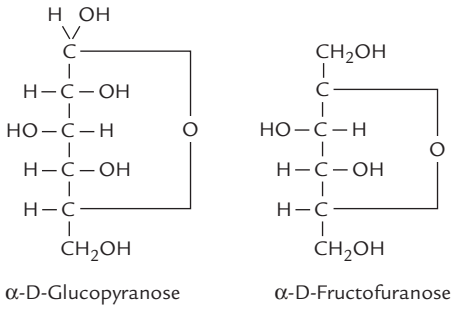


Fig. 2.7. Fischer projection of α-D-glucopyranose and α-D-fructofuranose.

Carbohydrate chemists, therefore, more often use the so called **Haworth projections**, which are fairly representative of reality. Haworth representations of some monosaccharides, including α-D glucose, are shown in Figure 2.6. Note that those groups that are to the right in the Fischer structures appear down in the Haworth structures and those appearing to the left in the Fischer formulas point up in the Haworth structures.

IV. Hexoses, Pentoses, and Related Disaccharides and Oligosaccharides

A. Hexoses and Pentoses

Among various types of monosaccharides present in human body, hexoses are physiologically the most important. Glucose, galactose, fructose and mannose are the most predominant hexoses. Glucose is important because it is a component of the storage sugars in plants (starch) and animals (glycogen); the blood sugar often mentioned is glucose, which is used as fuel by various tissues for their energy needs. Red blood cells and brain utilize glucose exclusively, although brain in prolonged starvation also utilizes ketone bodies. Roles of hexoses in body are given in Table 2.2.

Pentoses are integral constituents of nucleic acids. Ribose, an aldopentose, is one of the most important molecules in biochemistry, being present in RNA. DNA contains deoxyribose, the corresponding deoxy sugar.

B. Disaccharides

Covalent joining of two monosaccharides units by glycosidic bond forms disaccharide; *the glycosidic bond links anomeric carbon of a monosaccharide to a hydroxyl group of the other.* A number of disaccharides of biological significance contain glucose, fructose and galactose. *Maltose,*

Table 2.2 Sources and physiological importance of some hexoses

Sugar	Source	Importance
1. D-Glucose	Hydrolysis of starch, sugarcane and disaccharides like lactose and maltose. Most fruits contain glucose.	The principal sugar used by the tissues. Carried by the blood to these tissues.
2. D-Galactose	Hydrolysis of the milk sugar (i.e. lactose).	Occurs in glycolipids and glycoproteins. Changes to glucose in the liver and metabolized. Used in the synthesis of milk sugar in the mammary gland.
3. D-Fructose	Fruit juices, honey, sugarcane.	Changes to glucose in the liver (and intestine) and metabolized.
4. D-Mannose	Hydrolysis of plant mannose and gums.	Occurs in glycoproteins, a constituent of prosthetic group of oligosaccharides of albumin, globulins and mucoproteins.

the malt sugar, is a glycoside of two glucose molecules in an $\alpha(1 \rightarrow 4)$ linkage; *lactose*, present in milk includes galactose and glucose in $\beta(1 \rightarrow 4)$ glycosidic bond; and *sucrose*, the table sugar includes glucose and fructose in an $\alpha(1 \rightarrow 2)$ linkage (Fig. 2.8). *Cellobiose*, an isomer of maltose, which can be derived from a plant polysaccharide cellulose, involves $\beta(1 \rightarrow 4)$ linkage.

In sucrose, both anomeric carbon atoms are involved in the glycosidic bond. Therefore, the sucrose molecule, in contrast to lactose, maltose, and cellobiose, does not have a reducing end and is called a ***non-reducing*** sugar (the reducing end is the end with a free anomeric carbon: free implies not involved with the formation of a glycosidic bond). Benedict's test, one of the earliest tests to detect presence of sugar in urine, which is based on the reducing properties of the sugar, is therefore, negative with sucrose.

Glycosidic bonds are acetal or ketal bonds, involving the anomeric carbon (the aldehyde or keto carbon, respectively) of one of the participating monosaccharides. Once engaged in glycosidic bond formation, the aldehyde or keto carbon loses its reducing properties.

These disaccharides, with the exception of cellobiose, have food value for human beings, and the intestinal tract contains specific glycosidases that can cause their hydrolysis into their constituent monosaccharides. α-*Glucosidase (maltase)* causes the hydrolysis of maltose to glucose but is inactive with cellobiose. *Sucrase* catalyzes the hydrolysis of sucrose to glucose and fructose, and α-*galactosidase (lactase)* converts lactose to galactose and glucose.

Invert sugar: Sucrose is dextrorotatory (+66.5°). When it is hydrolyzed, one molecule of glucose (+52.5°) and one molecule of fructose (−92°) is formed. As the levorotatory effect of fructose is greater than the dextrorotatory effect of glucose, the hydrolysate is levorotatory. Because the optical rotation is inverted following hydrolysis by *sucrase*, the enzyme is termed *invertase* and substrate sucrose is termed invert sugar.

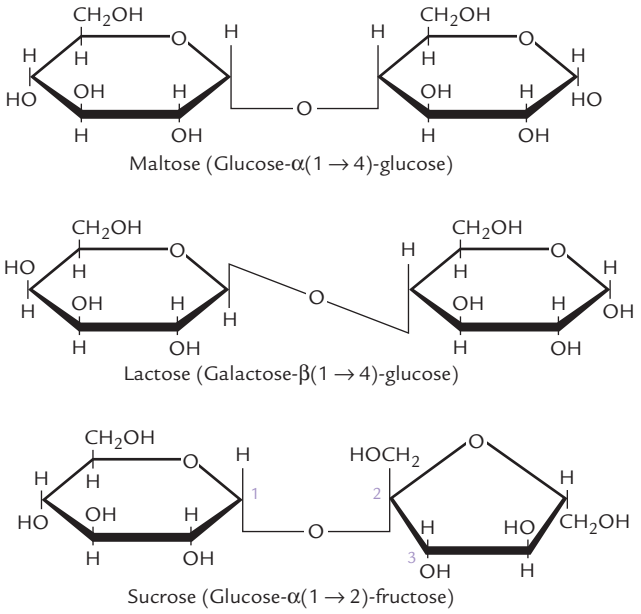


Fig. 2.8. Some common disaccharides.

Lactose intolerance, caused by absence of *lactase*, results in an inability to digest lactose. The undigested lactose moves through the digestive tract to colon, where bacterial fermentation generates large quantities of CO_2, H_2 and organic acids. These products cause diarrhoea, bloating and painful gastrointestinal upsets in such patients. Lactose intolerance may be the result of a genetic trait, in which case it is absent in infants, who must be given a non-milk artificial formula; or more commonly in adults, in whom *lactase* disappears after adolescence.

C. Oligosaccharides

Oligosaccharides vary from disaccharides to complex branched structures of glycosidically linked monosaccharide units. Considerable variation in the linkage exists: anomeric carbon (C-1 or C-2) of a monosaccharide unit may be linked to either C-1, C-2, C-3, C-4 or

C-6 of another, and moreover, this linkage may be in the α- or β-configuration. Each of these linkages forms a different molecule with distinct properties. Comparatively, the number of possible molecules formed by joining two identical amino acids, such as glycine, together is only one.

Such variation in linkage in oligosaccharides is the basis for the mechanism of cell-cell recognition. The number of variations of bonds suggests that it is much easier to create the enormous variability required for identifying different cells by using sugars rather than amino acids.

Glycoprotein

Attachment of oligosaccharide to a protein yields glycoprotein (more appropriately defined as proteins having oligosaccharide attachments). The attachment to the protein usually occurs at a serine or threonine residue (*O-linked*), but there are also attachments to the amide nitrogen of asparagine residues (*N-linked*). The remainder of the chain is built stepwise by sequential addition of other monosaccharide units. Monosaccharides like L-fucose and N-acetylglycosamines generally appear at the ends of the chain.

Functions: Glycoproteins are involved in a variety of functions including lubrication, many being secreted, and another large group involved with cell adhesion and cell recognition (for details regarding structure and function of glycoproteins see Chapter 5). The functions of the carbohydrate chains of the glycoproteins are diverse: they stabilize the protein against denaturation, protect it from proteolytic degradation, enhance its solubility, or serve as recognition signals to facilitate cell-cell interaction. Moreover, the antigenic sections of a number of glycoproteins, including blood group substances, consist of oligosaccharides.

V. Derived Sugars

The term derived sugars is applied to monosaccharides whose structure cannot be represented by the general formula $(CH_2O)_n$, or which have some unusual features. Some of the physiologically important derived sugars as described here.

A. Acid Sugars

The *acid derivatives of the sugars, produced by oxidation of the aldehyde carbon or the hydroxyl carbon or both, are called the* ***acid sugars***. For example, **D-glucuronic acid** is formed by carboxylate group substitution at the sixth carbon position (i.e. by oxidation of the hydroxyl group at C-6; Fig. 2.9a), and **D-gluconic acid** at the first carbon (i.e. by oxidation of the aldehyde group at C-1 (Fig. 2.9b). Likewise oxidation of the last carbon of galactose gives galacturonic acid. **Iduronic acid** is an epimer of glucuronic acid (Fig. 2.9c); both these acid sugars are important constituents of glycosaminoglycans, discussed later. When both the first and the sixth carbons of glucose are oxidized to carboxyl groups, the product is **saccharic acid**.

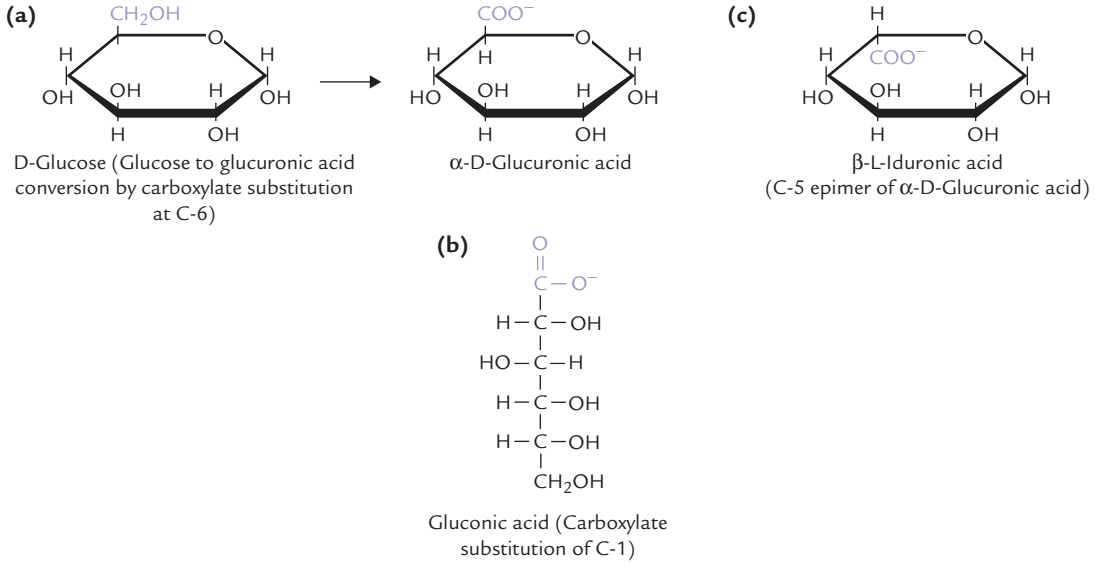


Fig. 2.9. The acid sugars – produced by oxidation of an aldehyde group/or hydroxyl group of a monosaccharide to carboxyl group.

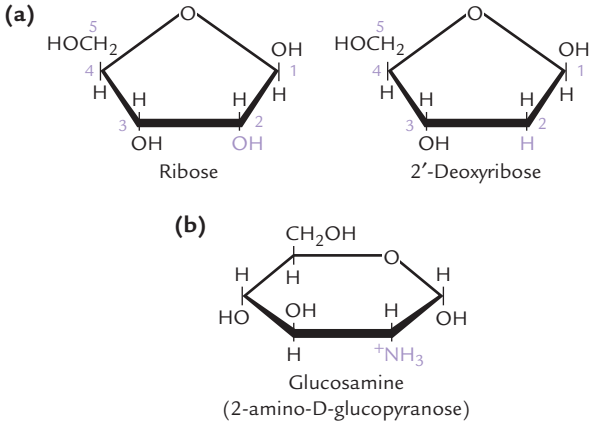


Fig. 2.10. Other monosaccharide derivatives.

B. Deoxy Sugars

The ***deoxy sugars*** *are the ones in which a hydroxyl group of the sugar is replaced by hydrogen atom*; for example, the deoxyribose is obtained from ribose (Fig. 2.10a). Deoxyribose is present in DNA molecules. Deoxyglucose is an important inhibitor of glucose metabolism.

C. Amino Sugars

A large group of derived monosaccharides are the ***amino sugars*** *where an amino group replaces the –OH residue on carbon 2 of the hexose (Fig. 2.10b), such as glucose, galactose and mannose.* The corresponding compounds are glucosamine, galactosamine and mannosamine. The amino group is usually acetylated. N-acetylglucosamine and N-acetylgalactosamine are components of connective tissue proteoglycans, glycoproteins, and complex lipids. An unusual amino sugar is N-acetylneuraminic acid (NANA), sometimes called **sialic acid**, which is derived from N-acetyl mannosamine and pyruvate. It occurs in glycoproteins and complex lipids.

Monosaccharide derivatives include acid sugars, deoxy sugars, reduced sugars, amino sugars, phosphoric acid esters and glycosides.

Several antibiotics (e.g. **erythromycin** and **carbomycin**) contain amino sugars. Presence of amino sugars is related to the bacteriostatic or the bactericidal activities of these drugs.

D. Phosphoric Acid Esters

They are produced by *attachment of a phosphate group with hydroxyl group of sugar.* The phosphorylated sugars, such as glucose 6-phosphate and glucose 1-phosphate, are thermodynamically favoured to serve as the metabolic intermediates.

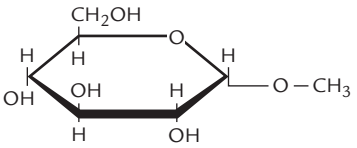
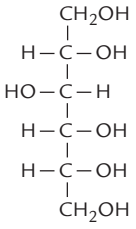


Fig. 2.11. Methyl α-D-glucose is a methyl glycoside formed by covalent linking of α-D-glucose with methyl alcohol (aglycone) by glycosidic linkage.

E. Sugar Alcohols

Monosaccharides are reduced at their carbonyl group to yield the corresponding polyhydroxy alcohols, which are intermediates of some secondary metabolic pathways. Both aldoses and ketoses can undergo reduction to form the sugar alcohols. For example, the reduced product of glucose (or fructose) is sorbitol. Similarly, mannose yields mannitol and galactose yields galactitol; the latter triggers cataractogenesis (formation of cataract) in galactosaemias.



F. Glycosides

When the *hemiacetal or hemiketal group of a monosaccharide is linked covalently with an alcohol, sterol or phenol group, the compound so formed is called a* ***glycoside*** (recall that reaction of aldehyde and ketone with the hydroxyl group forms hemiacetal and hemiketal respectively). It has two components: a carbohydrate and a non-carbohydrate (termed aglycone), and the linkage between the two is called *glycosidic linkage*. A simple example is a methyl glycoside (Fig. 2.11). The term glycosidic means that there is bond to a sugar and in this case the O-glycosidic bond is between a sugar (α-D-glucose) and methyl alcohol (Fig. 2.11).

Medical significance: Glycosides are found in many drugs. **Digoxin** is a cardiac glycoside, which is used to treat cardiac failure. The glycosidic bond is between a sugar

and a steroid (the latter is similar in structure to the oestrogens). Several other medications also have glycosidic bonds as main features of their structures. For example, **phlorhizin**, a glycoside of glucose and phloretin is used in renal damage. It is obtained from rose bark. **Digitonin**, a cardiac stimulant (obtained from leaves of foxglove) is formed from galactose and digitogenin. Various **antibiotics**, such as streptomycin, puromycin and erythromycin, have O-glycosidic bonds.

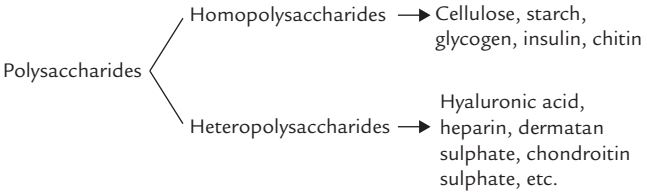
VI. Polysaccharides

Most carbohydrates occur in nature as complex, high molecular weight polysaccharides or glycans, which comprise several monosaccharides or monosaccharide-derivatives, linked by glycosidic bonds. They are of two major types: **homopolysaccharides**, which are made from one particular sugar unit; and **heteropolysaccharides**, which are made from different sugar units. In both cases there may be linear or branched polymers and the constituent sugar units are linked by glycosidic bonds; the latter link the anomeric carbon of a monosaccharide to the hydroxyl group of another.

Storage of energy: Polysaccharides such as starch and glycogen store the chemical energy intracellularly (i.e. storage polysaccharides). Starch is the principal storage polysaccharide in plants. Its counterpart in the animal cells, namely glycogen, is often referred to as animal starch. Both glycogen and starch are stored intracellularly in the form of cytoplasmic clusters or granules.

Structural elements: Polysaccharides are important constituents of various cell components (i.e. structural polysaccharides). For example, hyaluronic acid is an essential component of connective tissue, and cellulose serves as important structural component of the cell wall. Cellulose is the most abundant extracellular macromolecule.

Note: The most abundant intracellular molecules are proteins.



A. Homopolysaccharides

Glycogen is the most important homopolysaccharide in human body. Plant homopolysaccharides include **starch, cellulose, inulin**, etc. The tough exoskeletons of many lobsters, crabs and several insects consist largely of structural homopolysaccharide, **chitin**, which is a linear polymer of N-acetylglucosamine.

Starch

Starch is a plant *homopolymer of glucose*. It is a mixture of two distinct polysaccharide forms: amylose and amylopectin. **Amylose** is an unbranched polymer of glucose, linked by α(1→4) bonds, and occurs in helical coiled conformation. **Amylopectin**, on the other hand, is a branched form; most of the constituent glucose are joined in α(1→4) linkages but additional α(1→6) bonds occur every 25–30 residues, creating branch points (Fig. 2.12).

Like other branched polymers, starch has only one reducing end and multiple non-reducing ends. (The reducing end is the end with free C-1, i.e. the C-1 not involved in the formation of a glycosidic bond.)

Starch hydrolysis: Most plants have two hydrolyzing enzymes: *α-amylase* and *β-amylase*; both are specific for α(1→4) linkages in large polymers and have no action on maltose or maltriose. Cleavage by *β-amylase* is much more common in plants and is characterized by successive removal of maltose units beginning at a non-reducing end (*exoglycosidase*), until the action of enzyme is blocked at the α(1→6) linkage. Thus, action of *β-glycosidase* halts at branch points leaving a branched homopolymer called **limit dextrin**. The *α-amylase* cleaves internal glycoside bonds (*endoglycosidase*) randomly, as discussed below.

Starch digestion in humans: It is effected by *salivary amylase* and *pancreatic amylase*; both are *α-amylases*. There is only 1% difference in their composition but the *pancreatic amylase* is more important from quantitative viewpoint because food does not generally remain long enough in the mouth to be thoroughly digested by

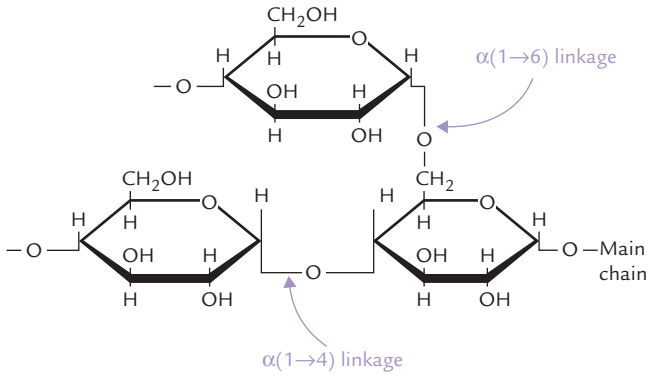


Fig. 2.12. α(1→4) and α(1→6) glycosidic linkages between glucose residues in starch and glycogen. The α(1→4) linkages occur in straight chain, and α(1→6) at branch points.

salivary amylase. The *pancreatic α-amylase*, on the other hand with a pH optimum of 6.9 degrades the starch molecule extensively within duodenum where food stays longer. It cleaves α(1→4) linkages that are located internally (*endoglycosidase*), but has no effect on α(1→6) linkages.

Products formed by action of *amylase* are α-maltose and maltriose. From amylopectin α-dextrins (oligosaccharides made up of 5–9 glucose units with one α(1→6) bond) and linear oligosaccharides are additionally produced. Under physiological conditions, starch is hydrolyzed to produce 40% maltose, 20% maltriose, 30% α-dextrin and 5% linear oligosaccharides.

In acute pancreatitis, the *pancreatic amylase* levels in serum are highly elevated and this serves as a reliable diagnostic tool.

Glycogen

Glycogen is the *storage form of glucose* in animals and commonly referred to as animal starch. Structurally, it resembles amylopectin but is more extensively branched: the α(1→6) linkages occur every 10–18, units, whereas in amylopectin they occur about every 25. Thus, it has more reducing ends than amylopectin and its α(1→4) linked oligomers are shorter (Fig. 2.13). Moreover, glycogen may have several hundred thousand glucose residues (its molecular weight is up to several millions), whereas amylopectin has only 300–6000 residues.

The storage polysaccharides are homopolysaccharides consisting of glucose residues linkaged by α(1→4) bonds with α(1→6) bonds at branch points. The major storage polysaccharides are glycogen (in animals), starch (in plants) and dextran (in yeast and bacteria).

Fig. 2.13. A diagrammatic representation of glycogen. Each circle represents a glucose residue (* = non-reducing ends, O–O = α(1→4) linkage, arrows indicate α(1→6) linkage (branch point), R = remainder of the glycogen molecule).

How does storage of glucose as a polymer (glycogen), rather than simply glucose, offer an advantage? If glucose were stored (as glucose) rather than a polymer, it would cause an enormous osmotic force within the cell. This would cause water to move into the cell and the cell would burst.

Glycogen is stored in cytoplasmic granules in the hepatocytes and the muscle cells. *Bulk amounts of glycogen are stored in liver and muscles*, although most other cells may contain smaller quantities. Highest concentration of glycogen is reached in the liver in the well-fed state (6–8 g/100 g wet weight approximately). The liver glycogen plays a vital role in glucose homeostasis (Chapter 9). Skeletal muscles have much lower glucose concentration (1–2 g/100 g wet weight approximately), but net amount of glycogen stored in muscles (200–300 g) is much more than that in liver (50–100 g). This is because the total muscle mass exceeds that of liver.

Roles of liver and muscle glycogen are different: The muscle glycogen stores provide fuel reserves for the synthesis of ATP during muscle contraction. On the other hand, the liver glycogen serves to maintain the blood glucose concentration within normal range, particularly in the early stages of fasting.

The cytoplasmic granules that store glycogen contain the enzymes of glycogenesis (glycogen synthesis) as well as those of glycogenolysis (glycogen degradation). During muscle contraction or in fasting state, the enzymes of glycogenolysis become active; and conversely, during well-fed state the enzymes of glycogenesis are activated (see Chapter 15).

Celluloses

They are the most abundant organic compounds on earth, being essential components of plants, comprising 20–45% of their cell wall mass. Wood is about 50% cellulose, while cotton is almost pure cellulose. Wood enables trees to attain towering heights and yet it is just a polymer of glucose.

The basic unit of cellulose is a chain of glucose residues linked by β(1→4) *glycosidic bonds*. The chains are in parallel alignment and held in place by interchain hydrogen bonds. Major variations are in the degree of polymerization which appears to be biphasic: either being less than 500 glucose units or between 2500 and 4500 glucose units per chain. These hydrogen bonded chains are further stabilized by the presence of other polysaccharides, such as hemicellulose, pectin, and lignin, which function as cementing materials.

Cellulose cannot be digested: The nutritional value of cellulose is summarized as: "Glucose everywhere but not a drop to eat". This is because it is indigestible as

humans lack the enzyme *cellulase* that can cleave the β(1→4) glycosidic bonds. *Cellulase* is found in snails, bacteria, fungi and insects. Ruminants can digest cellulose because of the presence of bacteria (rumens) in their stomach which hydrolyze the β(1→4) glycosidic bonds. Even with *cellulases* from bacteria, cows have to chew their cud, which means vomiting out half-digested cellulose and eating it again. (Rabbits are worse as they pass hard pellets initially and eat them to produce the soft pellets).

Despite its nutritional value being nil the indigestible polysaccharides, collectively called **dietary fibres** (include cellulose hemicellulose, lignin and pectin) are important in our digestive processes because they provide roughage which is believed to be important for keeping the contents of our intestines mobile and in lowering the possibility of developing bowel cancer (Chapter 26).

Cellulose has certain commercial uses as well. Cellulose derivatives are used for manufacturing emulsifiers, plastics and explosives.

Chitin

Composed of N-glucosamine residues linked by β(1→4) glycosidic bonds, chitin is a subtle variation of cellulose. The polymers are held together by hydrogen bonding. It is found in exoskeletons.

Celluloses and chitin are polysaccharides whose constituents are β glycosidically linked; the β-linkages cause them to adopt rigid and extended structures.

The plants and animals contain several other structural and storage homopolysaccharides, as shown in Table 2.3.

B. Heteropolysaccharides

Heteropolysaccharides *(also called heteroglycans) are the polymers made from more than one kind of monosaccharides or monosaccharide derivatives.* They may exist in the free-state or conjugated (joined) with lipids, peptides or proteins. Study of heteropolysaccharides is one of the most exciting areas of new knowledge in biochemistry and has created immense interest in recent years, much the same way genetics did a decade earlier.

Heteropolysaccharides are involved with viral and bacterial attachment to cells, the immune system, the migration of normal cells, the migration of cancerous cells, the fertilization of ova, and the changes in cells during tissue development. These functions are discussed in detail later in this chapter.

Glycosaminoglycans

Glycosaminoglycans (GAGs)/mucopolysaccharides are a group of high molecular weight, linear polysaccharides, with repeating disaccharide units. Each disaccharide unit comprises of an amino sugar and an acid sugar (uronic acid).

$(XY)_1 — (XY)_2 — (XY)_3 \ldots (XY)_n —$ Core protein

X = Amino sugar
Y = Uronic acid (mostly)

Table 2.3. Homopolysaccharides

Homopolysaccharide	Composition	Occurrence and role
Chitin	Polymer of N-acetyl D-glucosamine linked by β(1→4) glycosidic linkages	Present in exoskeleton of invertebrates such as crustaceae, insects and spiders. Integral component of cell walls of most fungi and many algae.
Inulin	A fructosan, consisting of D-fructose units, linked covalently	Occurs in tubers and roots of dahlias and dandelion. Used for determination of glomerular filtration rate and to measure the body water volume.
Agar	Polymer of galactose units which are sulphated	Occurs in sea weeds. Purified form of agar is used for laboratory culture of bacteria.
Dextrins	Consists of α-D-glucose residues linked by α(1→4) and α(1→6) glycosidic bonds	Formed in course of hydrolytic breakdown of starch (limit dextrins are formed first as the degradation reaches branch points). Found in various foodstuffs obtained from starch producing plants. Used as adhesives and binders.
Dextran	α(1→4), α(1→6) and α(1→3) linked D-glucose units, which are interconnected in such a way so as to form a network	Given intravenously it increases plasma volume (i.e. plasma expander), and therefore used in case of blood loss.

GAGs are the most important group of heteropolysaccharides in humans. They were first isolated from mucin and therefore also called **mucopolysaccharides**. These complex sugars are major components of the extracellular matrix of connective tissue, including bone and cartilage, synovial fluid of joints, vitreous humor of eye and secretions of mucus-producing cells. The GAGs found in connective tissue may attain extremely large size, with molecular weight of 1–80 millions.

Structure

GAGs are unbranched polysaccharides of disaccharide repeats; each repeat unit comprising a sugar unit linked to a glycosamine, hence called glycosaminoglycans.

- Glycosamine (amino sugar): It is either glucosamine ($GluNH_2$) or galactosamine ($GalNH_2$), both of which are present in their N-acetylated forms. The amino group may be sulphated rather than acetylated in some GAGs.
- Acid sugar unit: Besides glycosamine, the other sugar unit present in a disaccharide repeat is *mostly* an acid sugar (uronic acid). Usually it is D-glucuronic acid (GluUA), although in some cases it may be L-iduronic acid (IdUA).

Glycosaminoglycans are unbranched polysaccharides containing acid sugars and amino sugars that are often sulphated.

GAGs are covalently attached to a protein (termed core protein) to form hybrid molecules, termed **proteoglycans**; the sugar component constitutes up to 95% of the proteoglycan molecule. Linking of polysaccharide chains to the core protein occurs by a core trisaccharide, Gal-Gal-Xyl.

GAG — *Core trisaccharide* — Core protein

An exception is hyaluronic acid, the longest GAG, which is not attached to a core protein.

Proteoglycan-monomers and proteoglycan aggregates: The proteoglycan molecules tend to aggregate, forming complex structures: proteoglycan-monomers and proteoglycan-aggregates.

Proteoglycan monomer comprises linear, unbranched polysaccharide chains, covalently linked to their core protein. The chains, each of which may be composed of about 100 monosaccharide units, extend out from the core protein giving an appearance of "bottle brush" (Fig. 2.14a).

Several proteoglycan monomers associate together to form a **proteoglycan aggregate**. A molecule of hyaluronic acid is also involved in formation of these aggregates by providing attachment sites for the core proteins (Fig. 2.14b). The association between core proteins and hyaluronic acid is not covalent but occurs through ionic interactions in which a link protein participates. This overall aggregate (also known as **aggrecan**) is a large complex macromolecule consisting of a three-dimensional array of proteoglycans bound to hyaluronic acid. This creates a stiff matrix in which collagen and other components of the extracellular matrix are embedded.

The *proteoglycan aggregates are polyanionic* because of the many negative charges of the carboxyl groups of the uronic acids, and the sulphate groups attached to some of the sugars. Being highly hydrophilic, these groups attract water molecules to form a hydrated gel like matrix that forms body's ground substance.

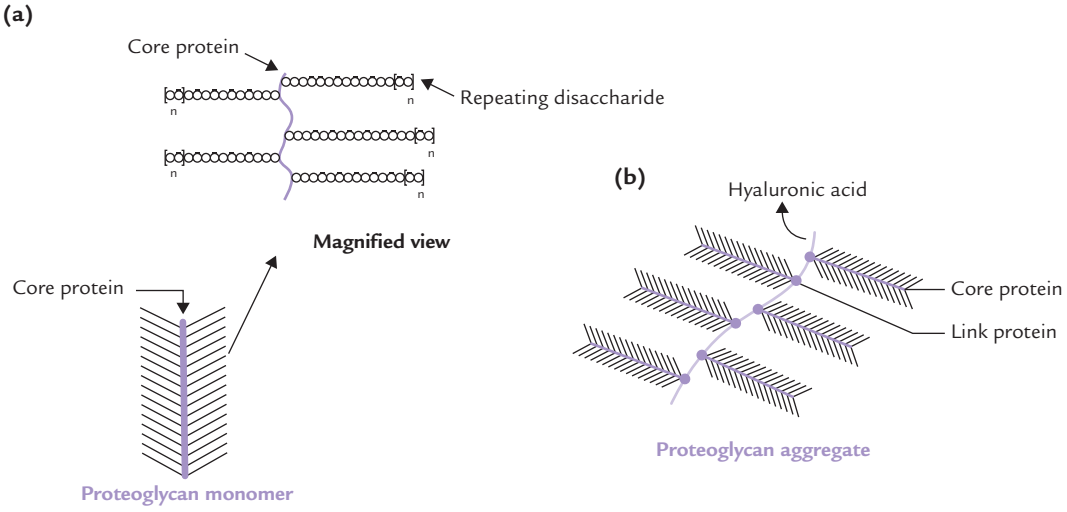


Fig. 2.14. The proteoglycan structure. **(a)** Bottle brush structure of a proteoglycan monomer: Linear, polysaccharide chains linked to core protein; a magnified segment is also shown, **(b)** Proteoglycan aggregate showing aggregation of several monomers on a central strand of hyaluronic acid with the help of link proteins (● = represents link protein).

Attachment of glycosaminoglycans to a protein molecule forms proteoglycans. Proteoglycans are enormous molecules consisting of central strand of hyaluronic acid with attached core proteins that bear numerous glycosaminoglycans and oligosaccharides. They bind water and minerals forming a hydrated gel which acts as a cushion against mechanical shock.

Proteoglycans versus glycoproteins: It is important to distinguish proteoglycans from glycoprotein. Glycoproteins (i) have short oligosaccharide chains (1–20 sugars in length), which are (ii) highly branched, (iii) generally do not have a repeating sequence and (iv) constitute 1–30% of the hybrid. On the other hand, sugar chains of proteoglycans are longer (100 or more sugars), linear and unbranched, with disaccharide repeats.

Relationship Between Structure and Function

The peculiar structural features of GAGs that explain their suitability to perform a number of specialized functions in the body are described below.

Gel-forming component of extracellular matrix (ECM): GAGs have polyanionic character, meaning numerous negative charges are present on a single molecule. Because of these charges, GAGs have two important properties:

- The heteropolysaccharide chains repel one another and, therefore, tend to exist in extended conformation in solutions.
- The anionic groups being strongly hydrophilic tend to associate with water.

The special ability of the chains to bind large amounts of water produces the gel-like matrix that forms the body's ground substance. Due to charge repulsion the chains tend to "*slip*" past each other, in a way two magnets of the same polarity do. This produces the "*slippery*" consistency of mucus secretions and synovial fluid.

Structural support to connective tissue: GAGs form a matrix or ground substance that stabilizes and supports the cellular and fibrous components of tissues. The character of connective tissue depends, to a large extent, on the relative proportions of ground substance and embedded fibrous proteins. For example, tendon (having high tensile strength) is composed primarily of fibres, whereas cartilage is rich in ground substance.

Others: GAGs play an important role in mediating cell-cell interactions. They are also structural components of the extracellular domains of several membrane proteins. Their slippery consistency, mentioned earlier, makes them suitable for a lubricant action in joints, tendon sheaths, and bursae.

Large aggregates of polysaccharide chains linked to their core proteins are non-covalently associated with hyaluronic acid via linked proteins forms a jelly-like matrix in which collagen fibres (and other components of extracellular matrix) are embedded. This arrangement provides both rigidity and stability to connective tissue and also allows for a degree of flexibility and compressibility.

Classification

The GAGs are classified in to six major types according to their monomeric compositions, type of glycosidic linkages, and degree and location of their sulphate units (Table 2.4). However, all types share some important features:

(a) Composition: They contain repeating disaccharide units of an amino sugar (glycosamine) and uronic

Table 2.4. Structure and distribution of the proteoglycans

Proteoglycan	Characteristic disaccharide	Sulphation	Tissue locations
1. Hyaluronic acid	D-Glucuronic acid, N-Acetylglucosamine	None	Joint and ocular fluids
2. Chondroitin sulphates	D-Glucuronic acid, N-Acetylgalactosamine 4-sulphate	GalNAc	Cartilage, tendons, bone
3. Keratan sulphates	D-Galactose, N-Acetylglucosamine 6-sulphate	GlcNAc	Cartilage, cornea
4. Dermatan sulphate	L-Iduronic acid, N-Acetylgalactosamine 4-sulphate	IdUA, GalNAc	Skin, valves, blood vessels
5. Heparin	Iduronic acid, N-Sulphoglucosamine	$GlcNH_2$, IdUA	Mast cells, liver
6. Heparan sulphate	Glucuronic acid, N-Sulphoglucosamine	GlcNc	Cell surfaces

GalNAc = N-acetylgalactosamine, $GlcNH_2$ = glucosamine, GlcUA = D-glucuronic acid, IdUA = L-iduronic acid.

acid. Keratan sulphate is an exception for it contains galactose in place of uronic acid.

(b) Linkage: With exception of heparin and heparin sulphate, where the linkage between the amino sugar and uronic acid is uniformly 1→4, in all other GAGs it is alternating 1→4/1→3.

(c) Core protein: GAGs are usually attached to a core protein by a core trisaccharide, Gal-Gal-Xyl. The latter is attached to a serine or threonine residue of the core protein by O-glycosidic linkage. Exceptions are hyaluronic acid and keratan sulphate. In case of keratan sulphate, the linkage to core protein is N-glycosidic (discussed in Chapter 5). Hyaluronic acid is a rather unusual GAG for it is not attached to a core protein.

Hyaluronic acid: The repeating disaccharide unit of hyaluronic acid comprises **D-glucuronic acid** and **N-acetylglucosamine,** joined by *β(1→3) linkage* (Fig. 2.15). Hyaluronic acid is different from other GAGs in lacking sulphate groups, not being covalently attached to protein and for not being limited to animal tissue (it is found in bacteria also). The *polysaccharide chain is longest* of all the GAGs, with molecular weight of 1×10^5 to 1×10^7 (250–25,000 repeating disaccharide units). It is a viscous jelly-like substance that fills the intercellular spaces of the animal tissues. It is found in synovial fluid of joints, vitreous humor of the eye, umbilical cord and loose connective tissue. It primarily serves as a lubricant and shock absorber. Some pathogenic bacteria secrete an enzyme *hyaluronidase*, which cleaves the glycosidic bonds of the hyaluronic acid. This renders the tissues more susceptible to invasion by the bacteria. *Hyaluronidase* also hydrolyzes the outer polysaccharide coat of the ovum and thereby makes the penetration by spermatozoa possible.

Chondroitin 4- and 6-sulphates (CS): These are the *most abundant GAGs in the body* that comprise **D-glucuronic acid** and **N-acetylgalactosamine units:** the latter are sulphated on either C-4 or C-6 (Fig. 2.15). Chondroitin sulphate is mostly present in the cartilage where it binds collagen and holds fibres in a tight, strong network. Tendons, ligaments, bones and aorta contain relatively smaller amount of this mucopolysaccharide.

Keratan sulphate (KS): It is the most unusual of all GAGs: the *only one that does not contain an acid sugar*. The repeating disaccharide unit consists of **N-acetylglucosamine** and **galactose**. It is linked to protein by rather unusual linkages (N-glycosidic), that are usually found in glycoproteins. *It helps to keep cornea transparent*.

Dermatan sulphate (DS): It was originally isolated from skin, but is also found in blood vessels and heart valves.

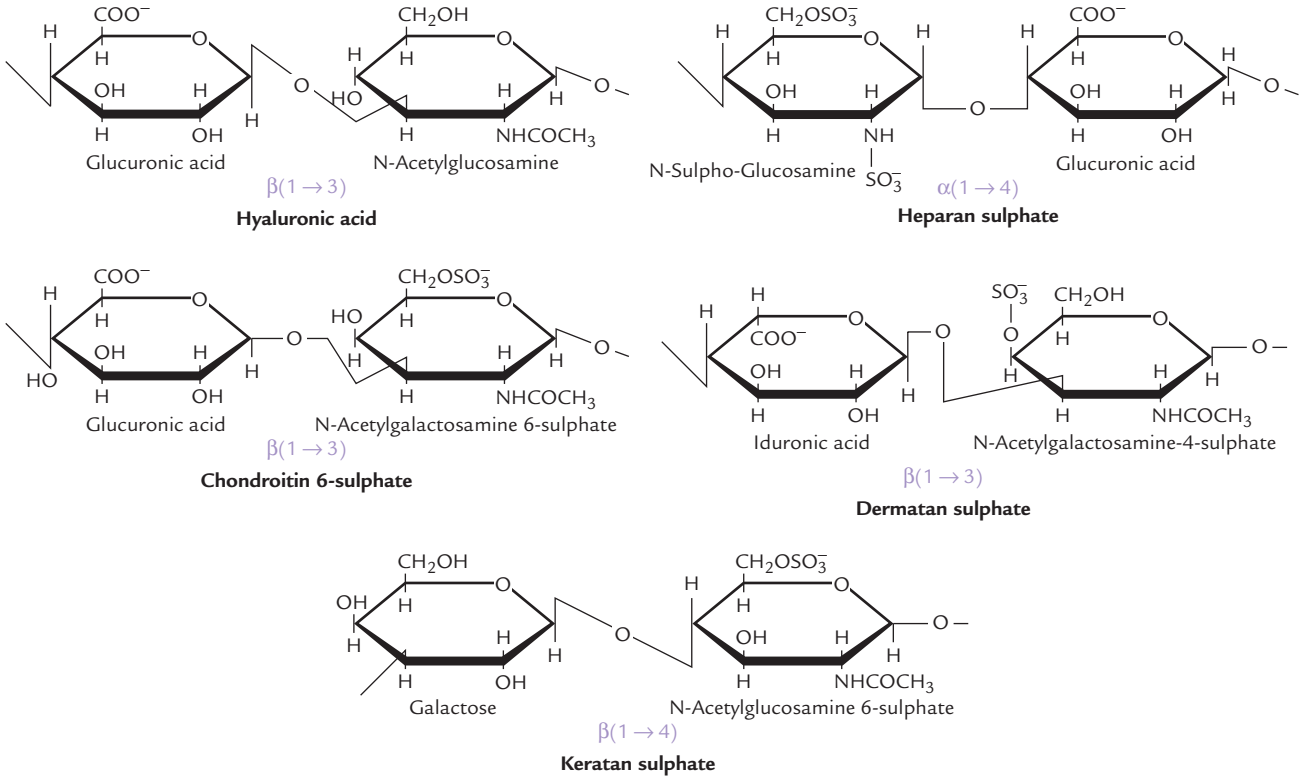


Fig. 2.15. Repeating disaccharide units of some glycosaminoglycans.

It helps maintain shape of these tissues. The predominant acid sugar is iduronic acid, though a variable amount of glucuronic acid is also present. The repeating disaccharide unit is made up of **iduronic acid** and **N-acetylgalactosamine 4-sulphate**. The L-iduronic acid, a C-5 epimer of D-glucuronic acid, is formed in an unusual reaction by epimerization of the latter, after it has been incorporated into the polymer.

Heparin: The repeating disaccharide unit comprising **glucosamine** and **glucuronic acid (or iduronic acid)**, is rich in sulphate groups. Almost all glucosamine residues are bound to sulphate group. An average of 2.5 sulphate groups per disaccharide unit is seen. The linkages between the amino sugar and the uronic acid is uniformly 1→4 in heparin (and heparin sulphate). (Alternating 1→4/1→3 linkages seen in other GAGs.) Unlike other GAGs (which are extracellular compounds), heparin is an *intracellular component* of mast cells that line arteries, especially in liver, lungs, and skin. Heparin is a potent anticoagulant that helps to prevent clotting of the circulating blood. It is therapeutically used to prevent clotting during intravenous therapy, and to inhibit clotting in various pathological conditions, such as following a heart attack.

Heparan sulphate (HS): It contains the same disaccharide units as heparin except that some of the glucosamine units are acetylated and there are fewer sulphate groups. It is an extracellular GAG, found in basement membrane and is an essential component of cell surfaces.

The structures of GAGs are quite variable. Thus, hyaluronic acid is a huge polysaccharide, chondroitin 4-sulphate has sulphur on C-4 rather than C-6 of the amino sugar; dermatan sulphate contains some glucuronic acid besides iduronic acid, and the sulphate of the amino sugar may be either on C-4 or on C-6; heparan sulphate contains some iduronic acid besides glucuronic acid; and heparian contains both glucuronic acid and iduronic acid.

Mucopolysaccharidoses

Mucopolysaccharidoses are a group of hereditary disorders of proteoglycan metabolism that are clinically progressive and characterized by excessive intralysosomal accumulation of GAGs in various tissues. Such accumulation accounts for clinical manifestations of mucopolysaccharidoses such as coarse facial features, thick skin, and corneal opacity. Defective cell function leads to mental retardation, growth deficiency, and skeletal dysplasia.

Biochemical defect: The underlying defect in mucopolysaccharidoses is impaired degradation of mucopolysaccharides by lysosomal enzymes. The lysosomal enzymes are hydrolytic in action and catalyze degradation of the GAGs at an optimum pH of 5.0. Such low optimum pH has a protective value since it prevents destruction of the cellular constituents in case the enzyme leaks out of the lysosomes. Half-life of the GAGs is relatively short (3–10 days), with exception of keratan sulphate (120 days). Therefore, with impaired degradation, the cellular concentration of the GAGs rapidly builds up.

Table 2.5. Mucopolysaccharidoses

Type	Syndrome	Enzymatic defect	Accumulated metabolite
I	Hurler's	*α-L-Iduronidase*	DS; HS
II	Hunter's	*Iduronate sulphatase*	DS; HS
III	Sanfilippo's	*Heparin sulphatase*	HS
IV	Morquio's	*Galactosamine sulphatase*	KS; CS
V	Scheie's	*L-Iduronidase*	DS
VI	Maroteaux-Lamy's	*N-Acetylgalactosamine 4-Sulphatase*	DS
VII	SLY's	*β-Glucuronidase*	DS; HS

CS = chondroitin sulphate, DS = dermatan sulplate, HS = heparan sulphate, KS = keratan sulphate.

Features of mucopolysaccharidoses: In mucopolysaccharidoses, the tissues that produce GAGs are most affected. In these tissues, degradation of GAGs is impaired. Lysosomal vesicles become swollen with incompletely degraded mucopolysaccharides. A characteristic finding of diagnostic significance is excessive elimination of the GAGs in urine. Diagnosis can be confirmed by measuring the concentration of *lysosomal hydrolases.*

The mucopolysaccharidoses are autosomal and recessively inherited, with an exception of Hunter's syndrome which is X-linked. Prenatal diagnosis of these disorders is possible. Children who are homozygous for these disorders are apparently normal at birth. The condition deteriorates gradually, but in severe cases the progression is rapid and death occurs in childhood. Unfortunately, no effective treatment exists at present.

Clinical types: Several types of mucopolysaccharidoses are recognized depending on the *lysosomal hydrolase* that is deficient (Table 2.5).

Exercises

Essay type questions

1. Define carbohydrates and give one example from each class/subclass. Explain why starch can be utilized in humans but not cellulose?

2. What are mucopolysaccharides? Name some and explain their biological significance.
3. How do physical properties of glycosaminoglycans relate to their biological roles?
4. Name some homopolysaccharides and explain their biological significance.
5. Compare and contrast structures and functions of cellulose, chitin, starch and glycogen.

Write short notes on

1. Mutarotation
2. Enantiomers
3. Chondroitin sulphate
4. Invert sugar
5. Cardiac glycoside
6. Anomers
7. Howarth and Fisher projections
8. Derived sugars

CHAPTER

CHEMISTRY OF LIPIDS

3

Lipids, carbohydrates and proteins form bulk of the organic matter in living systems. However, unlike the other two, lipids are not well defined chemically. In general, **lipids** *are a heterogeneous group of water-insoluble, oily or greasy substances that can be extracted with non-polar solvents (e.g. benzene or ether), but not with the polar solvents.* Because of their insolubility in aqueous solutions, body lipids are mostly found in isolated compartments. For example, droplets of triacylglycerols are present in adipocytes and some lipids are membrane bound. When present in blood, they form complexes with proteins so that they can be transported in the aqueous plasma. In contrast to other major biomolecules, *lipids do not form polymers.*

In this chapter, various types of lipids have been described, with a special emphasis on structure-function relationships. After going through this chapter, the student should be able to understand:

- Basic chemistry, nomenclature, properties and significance of fatty acids in the body.
- Chemistry, classification, properties and functions of various types of lipids and the related medical implications.

Lipids serve as fuel molecules, highly concentrated energy stores, signal molecules, components of cell membranes, and carriers of fat soluble vitamins, e.g. A, D, E and K. A layer of subcutaneous fat acts as thermal insulator and acts as a cushion, which protects several delicate organs by absorbing mechanical shocks. Both lipids and lipid derivatives serve as vitamins and hormones (Table 3.1).

The building blocks of most lipids are fatty acids. Some lipids such as cholesterol and terpenes lack fatty acids. However, these lipids are potentially related to fatty acids because they are synthesized from the catabolic end product of fatty acid degradation (i.e. acetyl CoA).

Lipids are a diverse group of molecules that are soluble in organic solvents, and in contrast to other major types of biomolecules, do not form polymers.

Table 3.1. Lipids and their major biological functions in the human body

Lipid	Biological function
Triacylglycerols	Energy storage, thermal insulation
Waxes	Keeps skin lubricated and waterproof
Phospholipids	Membrane components, detergents, surfactant, second messengers
Sphingolipids	Components of membranes, especially of nervous tissue and the myelin sheath
Sterols	Precursors of biologically useful compounds, such as bile acids, steroids, sex hormones, vitamin D, component of membranes

I. Fatty Acids

A. General Characteristics

The ***fatty acid*** *molecule consists of a long hydrocarbon chain with a polar carboxyl group at its end (Fig. 3.1).* Since the

pK value of the carboxylate group is around 4.85 (between 4.7 and 5.0), it rapidly ionizes at the physiological pH to form the carboxylate ion (COO^-). The carboxylate ion has polar characteristics, with high affinity for water (hydrophilic). However, the predominant portion of the fatty acid molecule is the long hydrocarbon chain in which carbon atoms are in their lowest oxidation state. Being non-polar in nature, hydrocarbon chain accounts for predominantly non-polar character of the fatty acid molecule. This in turn accounts for the oily or greasy nature of lipids.

Thus, a fatty acid molecule contains both polar (hydrophilic) and non-polar (hydrophobic) regions; such molecules are called **amphipathic molecules**. Fatty acids are simplest of all amphipathic substances in the body.

Naturally occurring fatty acids contain even number of carbon atoms (most contain 14 to 24). Of these, most abundant are the fatty acids containing 16 or 18 carbon atoms. The fatty acids with hydrocarbon chains containing one or more double bonds are called **unsaturated fatty acids**, whereas those lacking any double bonds are referred to as **saturated fatty acids**.

Fatty acids are carboxylic acids, whose chain length and degree of unsaturation varies. Most fatty acids have even number of carbons in an unbranched chain.

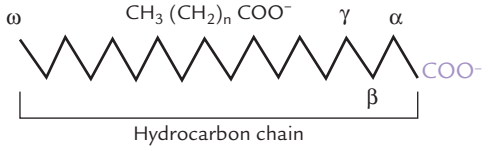


Fig. 3.1. Structure of a fatty acid.

Functions: Fatty acids perform following functions in the body:

1. **Membrane lipids:** They are the components of the more complex membrane lipids.
2. **Fuel molecules:** They are components of stored fat in the form of triacylglycerols.
3. **Hormones:** Derivatives of fatty acids serve as hormones (such as prostaglandins) and **intracellular second messengers** (such as IP3 and DAG).
4. Numerous proteins are covalently modified by fatty acids. Palmitic acid and myristic acid, for example, are directly attached to some proteins.

B. Nomenclature

The nomenclature of fatty acids is based on the following characteristics of the hydrocarbon chain:

- Chain length
- Presence of double bonds and their positions.

As shown in Table 3.2, the abbreviation 18;1 indicates an 18-carbon fatty acid having a single double bond. Likewise, 18;2 indicates an 18-C fatty acid with two double bonds. Δ^n indicates position of these bonds; for example, $\Delta^{9,12}$ indicates double bonds at C-9 and C-12, starting from the carboxyl end.

Thus, the 18-C saturated fatty acid (Fig. 3.2a) is called octadecanoic acid: octa (8) and deca (ten) imply chain length (18 carbons). Its common name is **stearic acid** and it is the most abundant fatty acid in nature. Introduction of a double bond at the C-9 position results in the formation of octadeca (mono) enoic acid (common name, **oleic acid**). Similarly, the 18-C fatty acid having two

Table 3.2. Some naturally occurring fatty acids

Carbon atoms	Systematic name	Common name	Melting point (°C)	Occurence % of total
Saturated				
12	n-Dodecanoic acid	Lauric acid	44.2	< 1
14	n-Tetradecanoic acid	Myristic acid	53.9	3
16	n-Hexadecanoic acid	Palmitic acid	63.1	23
18	n-Octadecanoic acid	Stearic acid	69.6	6
20	n-Eicosanoic acid	Arachidic acid	76.5	< 1
Unsaturated				
18;1–Δ^9	n-Octadeca-enoic acid	Oleic acid	13.4	50
18;2–$\Delta^{9,12}$	n-Octadeca-dienoic acid	Linoleic acid	–5	10
18;3–$\Delta^{9,12,15}$	n-Octadeca-trienoic acid	Linolenic acid	–11	< 1
20;4–$\Delta^{5,8,11,14}$	n-Eicosa-tetraenoic acid	Arachidonic acid	–49.5	< 1

18;2–$\Delta^{9,12}$ implies 18-C fatty acid with 2 double bonds at C-9 and C-12, 20;4–$\Delta^{5,8,11,14}$ implies 20-C fatty acid with 4 double bonds at C-5, C-8, C-11, and C-14.

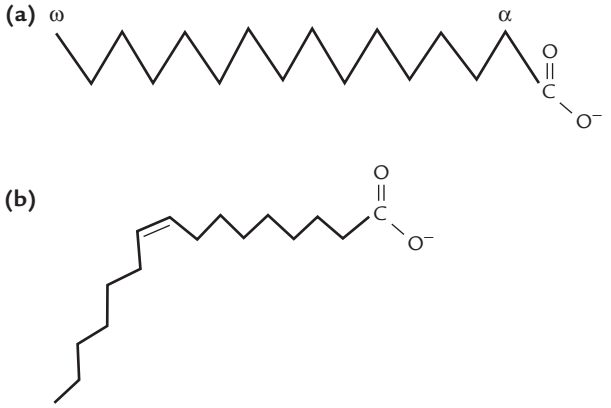


Fig. 3.2. Hydrocarbon chain. **(a)** Saturated, **(b)** Unsaturated with a *cis* double bond.

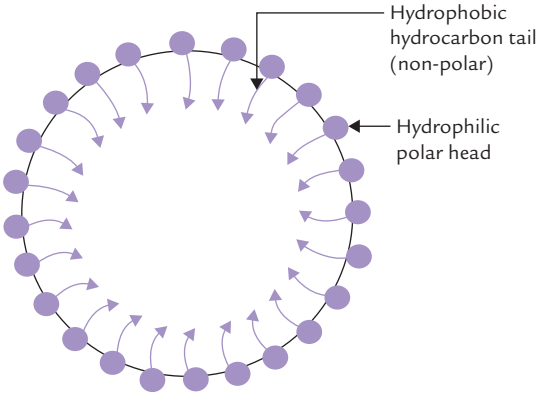


Fig. 3.3. Emulsification of fat by soap.

double bonds, at C-9, and C-12, is called octadeca-dienoic acid (common name **linoleic acid**), and that having three double bonds, at C-9, C-12 and C-15 positions, is called octadeca-trienoic acid (common name **linolenic acid**). The latter two, and all others that contain more than one double bond, are known as **polyunsaturated fatty acids (PUFA)**.

The double bonds are in *cis* geometrical configuration (Fig. 3.2b); this configuration produces a rigid bend in the aliphatic chain. They are *non-conjugated*, i.e. they are spaced at three-carbon intervals.

The carbon atom at the non-polar end of the aliphatic chain is called the **omega (ω) carbon**. Fatty acids can also be named by counting the carbon atoms from the ω carbon. Thus, the octadecatrienoic acid (linolenic acid) having double bonds at C-9, C-12 and C-15, can also be designated as a fatty acid of ω-3 series. This is because it contains a double bond attached to C-3 when counted from the non-polar (ω) end. Similarly octadeca-dienoic acid belongs to the ω-6 series.

C. Properties

Solubility

Fatty acids are predominantly non-polar in nature because of the **long hydrocarbon chain**. This accounts for the insolubility of lipids in water and other polar solvents, and solubility in the non-polar solvents.

Melting Point

Melting point of fatty acids is dependent on two factors:

1. **Level of unsaturation** of hydrocarbon chain: The melting point falls as the number of double bonds increases.
2. **Chain length** of fatty acid: Increase in chain length results in elevating the melting point. Saturated fatty acids of less than eight carbon atoms are liquid at physiological temperature, whereas those containing more than 10 are solids at this temperature.

In saturated fatty acids, hydrocarbons chains are in an extended conformation (Fig. 3.2a) and so can be packed together into a compact structure.

In unsaturated fatty acids, compact packing is prevented due to presence of *rigid bends* in their hydrocarbon chains; these bends are produced by the double bonds. Consequently, the unsaturated chains are loosely packed and are, therefore, more easily disturbed by thermal energy. This accounts for the lower melting point of the latter. For example, melting point of stearic acid (69.6 °C) is about fivefolds higher than that of oleic acid (13.4 °C).

Soap Formation

Fatty acids react with alkalies, such as NaOH or KOH, to produce the corresponding salts, called soaps.

$$\underset{\text{Fatty acid}}{\text{R–COOH}} + \text{NaOH} \longrightarrow \underset{\text{Soap}}{\text{R–COONa}} + \text{HOH}$$

Soaps are amphipathic molecules where the ionized carboxyl group constitutes the polar head, and the hydrocarbon tail forms the non-polar portion. *Because of their amphipathic nature, soaps can disperse oily or greasy substances into finer droplets; the process is called* ***emulsification***. The hydrophilic polar portion of the soap molecule remains at the surface, forming a stable cover, whereas the hydrocarbon tail extends into the interior to cause the dispersion (Fig. 3.3).

Calcium or magnesium soaps of fatty acids are highly insoluble, and hence do not emulsify oily substances. They precipitate as white, insoluble curds.

Hydrogenation

The double bonds of unsaturated fatty acids can be hydrogenated in the presence of catalysts, e.g. nickel to yield the corresponding saturated fatty acids.

Halogenation

Double bonds of unsaturated *fatty acids are capable of adding halogens*, the result of which are dihalogen fatty acids. This property is often used to assess the extent of unsaturation in fats and oils via reaction with iodine to obtain the so-called **iodine number** (number of gm of iodine required to saturate 100 gm of the lipid). The higher the iodine number, the more unsaturated the oil.

Oxidation

Unsaturated fatty acids can spontaneously react with atmospheric oxygen to form fatty acid peroxides, fatty acids epoxides and fatty acid aldehydes. Fat which undergoes this type of oxidation becomes **rancid**. Rancid fat is unpleasant in taste and smell and can be toxic. Rancidity is prevented by anti-oxidants, e.g. vitamin E. Degree of rancidity is measured by **acid number,** defined as the number of milligrams of KOH required to neutralize fatty acids in a gram of fat.

D. Essential Fatty Acids

The lipid biosynthetic capacity of the body can generate various fatty acids needed by the body. Key exceptions to these are the highly unsaturated fatty acids, e.g. **linoleic acid** and **linolenic acid,** containing unsaturated sites beyond carbon 9 (Table 3.2). Lack of endogenous synthesis makes their consumption in diet-essential, and hence they are called **essential fatty acids**.

It is important to note that **arachidonic acid,** a 20-C unsaturated fatty acid (Table 3.2) can be synthesized only from linoleic acid (Chapter 11). *Therefore, in deficiency of linoleic acid, arachidonic acid also becomes an essential fatty acid*.

Biomedical Importance

Essential fatty acids (PUFAs) are important for following reasons:

1. *Membrane components:* They are components of phospholipids and form biomembranes. Because the double bonds are in *cis* configuration, the PUFA molecules are sickle-shaped and so they increase fluidity of biomembranes. They are essential components of mitochondrial membranes, and therefore, their deficiency decreases efficiency of biological oxidation.
2. *Synthesis of eicosanoids:* The eicosanoids are biologically active lipids that are derived from polyunsaturated 20-carbon fatty acids, mainly arachidonic acid. They are short lived (half-life < 5 minutes), and act as **paracrine** or **autocrine messengers** within tissue of origin. Arachidonic acid can be processed to eicosanoids by two alternative pathways: the *cyclooxygenase pathway* which produces the prostaglandins, prostacyclin and thromboxane; and the *lipoxygenase pathway*, which produces, among other products, the leukotrienes (see Chapter 12 for further details).
3. *Effect on serum cholesterol*: Ingestion of PUFAs increases esterification and excretion of cholesterol, thereby lowering serum cholesterol level. Hence, EFAs have **anti-atherogenic effect**, which is augmented by their fibrinolytic activity (Chapter 12).
4. *Role of vision*: Dietary linolenic acid produces docosahexanoic acid in retinal photoreceptor membranes. It enhances the electrical response of the photoreceptors to illumination.
5. *Structural elements*: PUFAs are present in high concentrations in the lipids associated with structural elements of tissues.

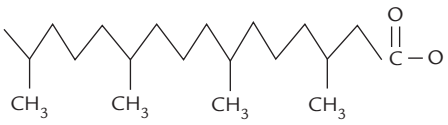
Deficiency of EFAs: Approximately 2–3% of daily calorie intake should be accounted by EFAs. Dietary deficiency is characterized by scaly dermatitis, poor wound healing and hair loss. However, the deficiency symptoms are rare due to wide distribution of EFAs. They are most commonly seen in patients suffering from severe fat malabsorption, in patients kept on parenteral nutrition for long, and in infants fed on low-fat milk formulas. The symptoms are readily cured by dietary linoleic acids.

A diet that supplies as little as 1% of the calories in from of EFAs is considered adequate. Interestingly, the attempts to induce a deficiency syndrome by selective omission of linolenic acid from diet have been unsuccessful, not only in humans but also in animals. Details regarding metabolism of EFAs, are given in Chapter 11.

E. Branched Chain Fatty Acids

Most naturally occurring fatty acids are of the straight chain variety, though lesser amounts of branched chain fatty acids (for example, phytanic acid) also occur in nature.

Significant amount of this fatty acid is present in dairy products. Defective catabolism of phytanic acid causes Refsum's disease **Refsum's disease,** which is characterized by accumulation of phytanic acid in plasma and tissues (Chapter 11).



II. Classification of Lipids

Lipids are classified into three groups—**simple lipids**, **compound lipids** and **derived lipids** (Table 3.3). Simple lipids are esters of fatty acids with alcohol. Compound lipids contain, besides fatty acids and alcohols, non-lipid polar groups like phosphate, carbohydrates, etc. Substances derived from the above groups of lipids by hydrolysis are termed derived lipids. These include fatty acids, monoglycerides, diacylglycerides, quinones and steroids.

A. Simple Lipids

The major simple lipids are the triacylglycerols and the waxes. They are neutral molecules, devoid of an electric charge. Moreover, they lack any polar group. This accounts for their hydrophobicity and water repellent properties. They serve as major fuel reserves in the animal and the plant cells.

Triacylglycerols

Triacylglycerols are the simplest and the most abundant lipids, also referred to as *fats, neutral fats* or *triglycerides.* They are fatty acid esters of glycerol and are the major storage fats in animal cells. Triacylglycerols stored in the adipose tissue in humans can meet the energy requirements of body for several weeks.

Table 3.3. Classification of lipids according to their chemical structure

A. Simple lipids
1. Triacylglycerols
2. Waxes

B. Compound lipids
1. Phospholipids
   - Phosphoglycerides
   - Sphingomyelins (or sphingophospholipids)
2. Glycolipids
   - Cerebrosides
   - Sulphatides
   - Globosides
   - Gangliosides

C. Derived lipids
Fatty acids (saturated and unsaturated)
Steroids and their fatty acid esters
Quinones, ketone bodies, lipoproteins
Polyisoprenoid compounds

Note: 1. Sphingomyelins and glycolipids may be grouped together as sphingolipids.
2. Cholesterol, cholesteryl esters and triacylglycerols are termed neutral lipids because they are uncharged.

Structure: A triacylglycerol molecule is an ester of three fatty acids and a trihydroxy-alcohol, called glycerol (Fig. 3.4). The three carbon atoms of glycerol are designated as α, β and α′, or more commonly as 1, 2, 3; the β carbon is chiral. Because about two-third of all fatty acids in the human organism are unsaturated, triacylglycerols reflect this usually by containing two molecules of unsaturated and one molecule of a saturated fatty acid. Such triacylglycerols are called mixed triacylglycerols. By contrast, the **simple triacylglycerol** have all three hydroxyl groups of the glycerol esterified to the same fatty acid. For example, three stearic acids bound to glycerol makes a simple triacylglycerol, called tristearin.

Oils, butter, and other food fats are complex mixtures of simple and mixed triacylglycerols containing fatty acids of varying chain lengths and unsaturation.

A triacylglycerol molecule lacks any free polar group because the polar hydroxyl groups of glycerol and the carboxyl groups of fatty acids are involved in ester linkage. This imparts the triacylglycerols a complete *non-polar character*.

The monoacylglycerols (monoglycerides) and diacylglycerols (diglycerides) have only one and two fatty acids, respectively, bound to glycerol.

Properties: Properties of a triacylglycerol depend on the nature of the constituent fatty acids.

1. *Melting point*: Triacylglycerols containing only saturated fatty acids, such as tristearin and tripalmitin, are solids at physiologic temperature (37°C) since melting point of the saturated fatty acids is relatively higher (Table 3.2). Conversely, the triacylglycerol that contain only unsaturated fatty acids, such as triolein, are liquids at room temperature. The latter can be solidified chemically by partial hydrogenation;

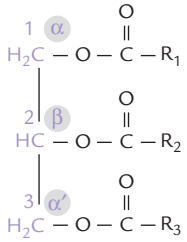


Fig. 3.4. Basic composition of a triacylglycerol. The glycerol backbone is coloured. The fatty acids esterified with the second carbon (R_2–COOH) and the third carbon (R_3–COOH) are usually unsaturated fatty acids, whereas (R_1–COOH) is a saturated fatty acid.

the process causes conversion of the unsaturated fatty acids to the saturated ones. Since shorter chain fatty acids also have lower melting point, butter, which is rich in such fatty acids, has a soft consistency.

2. *Solubility*: Lack of any polar group makes the triacylglycerols *insoluble in water*. However, they are *soluble in the non-polar solvents*, such as ether, benzene and chloroform.
3. *Specific gravity*: Triacylglycerols have a *lower specific gravity than water*. Therefore, when mixed with water, they float above to form the upper layer.
4. *Hydrolysis*: Triacylglycerols undergo hydrolysis when boiled with strong acids. *Lipases*, the gastrointestinal *esterases*, catalyze the hydrolysis at milder physiological conditions prevailing in the human organism. These enzymes are quite specific, and they do not necessarily remove all three fatty acid molecules from a triglyceride molecule. Thus, pancreatic *lipase*, the main lipid digestive enzyme, catalyzes the removal of fatty acids from positions 1 and 3 only.

   Hydrolysis of triacylglycerols by alkalis, such as NaOH and KOH, is called ***saponification*** (i.e. "soap formation"). It yields a mixture of sodium or potassium salts of the fatty acids (called soaps) and glycerol (Fig. 3.5).

Triacylglycerols (fats or triglycerides) consist of 3 fatty acid chains esterified to glycerol backbone. Simple triacylglycerols have three identical fatty acids and mixed triacylglycerols have two or three different fatty acids.

Saponification number: The number of milligrams of KOH required to saponify the free and combined fatty acids in one gram of a given fat is called saponification number. Fats having short chains with more carboxyl groups will have higher saponification number when compared to those having long chain fatty acids.

5. *Autoxidation*: When exposed to air, the triacylglycerols containing polyunsaturated fatty acids, are attacked by molecular oxygen at the double bonds of the hydrocarbon chains to yield complex products. This process is responsible for the spoiled taste of the **rancid fats.**

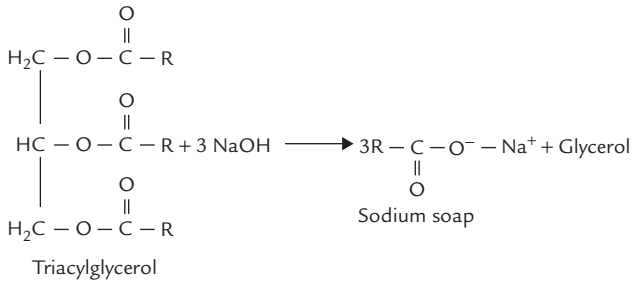


Fig. 3.5. Saponification (alkaline hydrolysis) of a simple triacylglycerol.

Functions: Triacylglycerols *store chemical energy,* and act as *thermal insulators and mechanical shock absorbers*.

1. **Storage of chemical energy:** Large triacylglycerol stores are present in the fat cells (adipocytes) of the adipose tissue. Subcutaneous tissues, viscera, and the abdominal cavity contain adipose tissue. Within cytoplasm of these cells, droplets of triacylglycerols coalesce to form a large globule which may fill most of the cell volume. The cells not only store triacylglycerols, but also mobilize them as fuel molecules that are transported to other tissues by blood.

Triacylglycerols store huge amounts of energy: Body can store unlimited amounts of triacylglycerols, which can sustain biological functions for several weeks. In a typical 70 kg man, the triacylglycerols constitute about 11 kg of his total body weight, which may provide about 100,000 kcal. If this amount of energy were stored in glycogen, at least 66 kg glycogen would have been required, and his total body weight would have been at least 55 kg greater. Evidently, triacylglycerols are more concentrated source of energy. This is accounted by their highly reduced and anhydrous nature.

- The reduced nature of the hydrocarbon chains means that carbons are present in lowest oxidation state. This accounts for a higher energy yield (9 kcal/g) upon complete oxidation of fatty acids, in contrast with about 4 kcal/g for carbohydrates and proteins.
- The anhydrous nature means that water molecules do not associate with triacylglycerols because of their hydrophobic character, and so space requirement is minimal. In contrast, the polar carbohydrates and proteins are highly hydrated: 1 g of dry glycogen binds about 2 g of water. Consequently, 1 g of nearly anhydrous fat stores more than six times as much energy as 1 g of hydrated glycogen.

These facts clearly explain why 11 kg of stored anhydrous triacylglycerols are equivalent to about 66 kg hydrated glycogen.

Triacylglycerols are the major energy store and the major dietary lipids in humans. Being insoluble in water, they are stored in specialized adipose (fat) cells.

2. In addition to serving as important energy store depots, the subcutaneous triacylglycerols act as **thermal insulators** against cold atmosphere.
3. A layer of fat is present around several delicate organs, which acts as a ***mechanical shock absorber***.

Waxes

Waxes are esters of long chain unsaturated fatty acids, having 14–36 carbon atoms, with long chain alcohols

(consisting of 16–22 carbons). In vertebrates, the waxes are secreted by skin glands. They keep the skin lubricated, water proof and pliable. Birds secrete large amount of waxes, which prevents any unwanted accumulation of water on their feathers because of the water repellent property of the waxes. This is especially important in flying birds whose body weight has to be kept light. Leaves of many plants also have a protective cover of waxes.

B. Compound Lipids

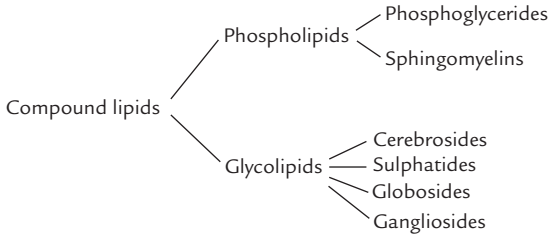
Compound lipids differ from simple lipids in having an additional non-lipid portion. Nearly 60% of lipids in the brain and most of the membrane lipids belong to this class. In contrast to the non-polar nature of the simple lipids, the compound lipids are amphipathic molecules. They can be subclassified into the following categories:

- Phospholipids which contain an additional phosphate group.
- Glycolipids which contain carbohydrate groups (Table 3.2).

Phospholipids

These are amphipathic molecules with non-polar aliphatic (hydrocarbon) *"tails"* and polar (phosphoryl-X) *"heads"*. They are major components of the biological membranes, accounting for more than half of the total lipids in most membranes. The polar heads of the phospholipid tend to extend to exterior, whereas the non-polar tails move towards the interior, where they associate with other non-polar constituents of the membrane including glycolipids, cholesterol and some proteins.

Phospholipids can be further subdivided into two groups: **phosphoglycerides** and **sphingomyelins.** Glycerol phosphate forms the backbone structure of the phosphoglycerides while an amino alcohol, sphingosine is present in the sphingomyelins.



Phosphoglycerides

Their basic structure is very similar to that of triacylglycerol (one fatty acid each is esterified to two of the carbon atoms of glycerol) except the third carbon that is esterified to phosphoric acid. The structure so formed is

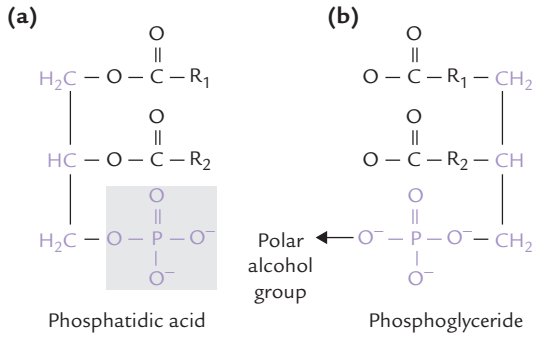


Fig. 3.6. (a) The structure of phosphatidic acid, the parent compound and biosynthetic precursor of phosphoglycerides, **(b)** A polar alcohol group is attached to the phosphate group to form a phosphoglyceride.

phosphatidic acid (diacylglycerol 3-phosphate) or **phosphatidate** molecule, as shown in Figure 3.6. Phosphatidate occurs only sparingly in the free form, and is not a major membrane lipid itself. However, it forms the baseline structure from which various phosphoglycerides or glycerophospholipids are derived.

The phosphoglycerides are derived from phosphatidic acid by attachment of a polar alcohol group to the phosphate group (of phosphatidic acid) by a phosphodiester bond. The common alcohol groups of phosphoglycerides are the ethanolamine, choline, the amino acid serine, myoinositol and glycerol (Table 3.4).

Plasmalogens and cardiolipins are less common types of phosphoglycerides. The phosphate group in these phosphoglycerides is negatively charged at physiologic pH values, and also, the polar alcohol (X-OH) bound to the phosphate is either charged or has a high hydrogen-bonding potential. Together with phosphate, it forms the hydrophilic head group of the molecule, and the two fatty acids form two hydrophobic tails.

Thus, phosphoglycerides are **amphipathic**, the hydrophilic and hydrophobic parts being combined in the same molecule. In common phosphoglycerides, the fatty acid in position 1 is usually saturated (either palmitic or stearic acid), whereas that in position 2 is unsaturated.

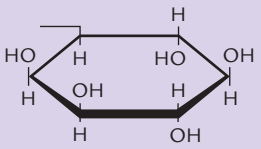
Two less common phosphoglycerides are **plasmalogens** (mainly in myocardium) and **cardiolipins**.

Plasmalogens: The widespread plasmalogens are distinguished from phosphoglycerides by the presence of an α-β-unsaturated fatty alcohol, rather than a fatty acid residue, in position 1. It is linked with carbon 1 of glycerol by an ether bond (Fig. 3.7).

Usually ethanolamine is present in the hydrophilic head groups of plasmalogens (ethanolamine plasmalogens).

Cardiolipins: Two molecules of phosphatidic acid esterified through their phosphate groups to a glycerol molecule yield a cardiolipin molecule (Fig. 3.8). It is a complex

Table 3.4. The polar alcohol groups (X-OH) in the major membrane phosphoglycerides

Name of X-OH	Formula of –X	Name of phosphoglyceride
Ethanolamine	$—CH_2CH_2NH_3^+$	Phosphatidylethanolamine
Choline	$—CH_2CH_2N(CH_3)_3^+$	Phosphatidylcholine (lecithin)
Serine	$—CH_2CH(NH_3^+)COO^-$	Phosphatidylserine
Myo-inositol		Phosphatidylinositol
Glycerol	$—CH_2CH(OH)CH_2OH$	Phosphatidylglycerol
Phosphatidylglycerol	See Fig. 3.8	Cardiolipin (diphosphatidylglycerol)

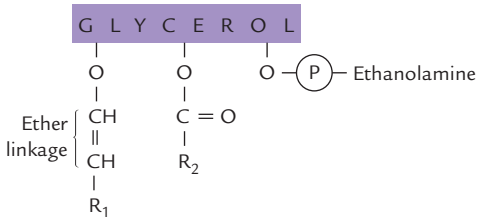


Fig. 3.7. Structure of plasmalogen. R_1 is hydrocarbon chains of fatty alcohol esterified to C-1 of glycerol, and R_2 is hydrocarbon chain of fatty acid esterified to C-2 of glycerol.

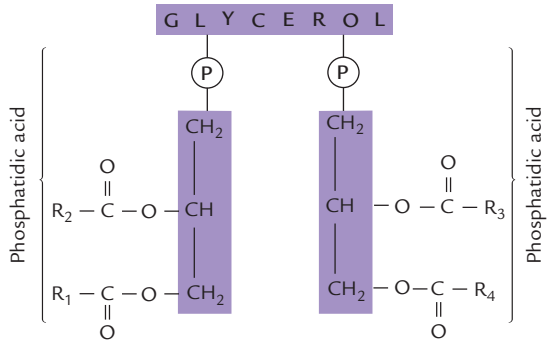


Fig. 3.8. Structure of cardiolipin (diphosphatidylglycerol); R_1, R_2, R_3 and R_4 represent hydrocarbon chains of fatty acids.

molecule of fatty acids, three glycerol molecules and two phosphate ions. Cardiolipin is an important component of the inner mitochondrial membrane and some bacterial membranes. *Cardiolipin is the only glycerophospholipid that is antigenic.*

Functions of phosphoglycerides

1. *Membrane lipids:* Phosphoglycerides are easily the most abundant membrane lipids, accounting for more than half of the total lipids (55–70%) in most membranes. They form the well known *lipid bilayer structures* because of their amphipathic nature. The bilayer structure forms a permeability barrier in which their hydrophilic head groups try to surround themselves with water, whereas hydrophobic tails avoid water, and so move towards the interior (Chapter 7). Cardiolipins and plasmalogens are relatively less common; the former is present in abundance only in the inner mitochondrial membrane, whereas plasmalogens are present in most tissues, especially in muscle and nervous tissue, where they account for up to 10% of the total phospholipids.

Phosphoglycerides, derived from phosphatidate, are amphipathic molecules that contain two fatty acid chains and a polar head group. In biomembranes, the polar head group lies outside and the hydrophobic tails of fatty acids are buried in interior.

2. *Surfactants:* Most phosphoglycerides function only as structural components of biological membranes, but one phosphoglyceride serves a special function as the major constituent of lung surfactant. It is *dipalmitoyl phosphatidylcholine* (dipalmitoyl lecithin), secreted by cells in the alveolar wall. It reduces surface tension by forming a thin bilayer film that lines the alveolar wall.

Dipalmitoyl-phosphatidylcholine is the most important component of lung surfactant (reduces surface tension of the thin fluid film that lines the alveolar walls).

Thus, dipalmitoyl-phosphatidylcholine helps to maintain shape of the alveoli and to prevent their collapse due to the high surface tension of the surrounding aqueous medium. Within the uterus, fetus is bathed in amniotic fluid, therefore, surfactant is not required. However, after the birth, surfactants are needed to prevent lung collapse. Normally, the fetus starts producing surfactants after the 30th week of

gestation, and adequate levels are reached at term. However, in premature infants, the adequate levels are not reached at the time of birth, which results in *respiratory distress* (**Case 3.1**). This condition accounts for 15–20% of neonatal deaths in western countries.

3. *Generation of second messengers:* Two intracellular signals (or second messengers), inositol triphosphate and diacylglycerol are generated from a membrane phospholipid, phosphatidyl inositol bisphosphate (Chapter 29).
4. *Anchoring certain proteins to cell membrane:* The phosphoglycerides, being amphipathic, can interact with non-polar as well as polar substances. Thus, they readily interact with the polar group of proteins as also with the cell membranes (which are predominantly non-polar). This makes possible anchoring of the proteins to biological membranes.
5. *Biologic detergents:* Phosphoglycerides are excellent biologic detergents; this action is especially important in the small intestine, where lipids must be emulsified for digestion purposes. The detergent property is also accounted by amphipathic nature of phosphoglycerides.
6. *Lipoprotein structure:* Lipids are transported as lipoproteins and phosphoglycerides are essential structural components of lipoproteins (Chapter 12).

Lysophospholipids: These are the compounds formed from phosphoglycerides by the removal of one of their fatty acid moiety. They are produced by action of the enzymes, ***phospholipases***. They account for 1–2% of total phospholipids in the living system. Some lysophospholipids possess potent haemolytic action. The haemolytic action of snake venom is because of its ability to generate lysophospholipids (since it is rich in *phospholipases*). However, toxic action of venoms cannot be solely attributed to this mechanism.

Sphingomyelins

Nervous tissue, especially myelins of nerves, and the erythrocyte membranes are rich in sphingomyelins, which are characterized by having an *18-C amino-alcohol*, known as **sphingosine** (Greek word, "*sphingein*" means to bind tightly). The hydroxyl groups in sphingosine are present at carbons 1 and 3, and amino group is present at carbon 2 (Fig. 3.9). A fatty acid gets attached at the amino group at C-2 to form a sphingosine derivative called **ceramide** (Fig. 3.10). Free forms of sphingosine and ceramide occur in nature in negligible amounts.

Of the two hydroxyl groups in ceramide, the one at C-3 remains unsubstituted, and the one at C-1 carries a variable substituent (a *phosphorylbase*, e.g. phosphocholine, phosphoethanolamine). The *phosphorylbase* forms the polar head group of the sphingomyelin.

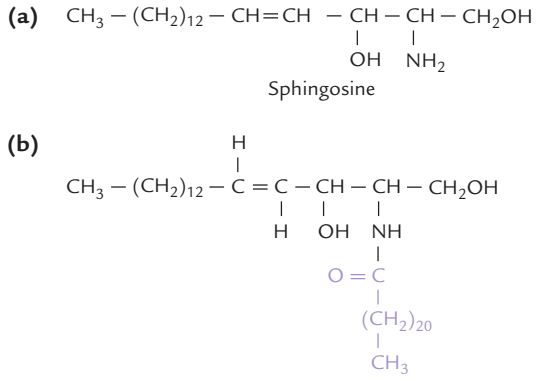


Fig. 3.9. Structure of sphingosine$_a$. **(a)** Sphingosine, an 18-carbon amino alcohol, with an amino group at C-2, and hydroxyl groups at C-1 and C-3, **(b)** Ceramide, formed by covalent attachment of fatty acid to the amino group at C-2.

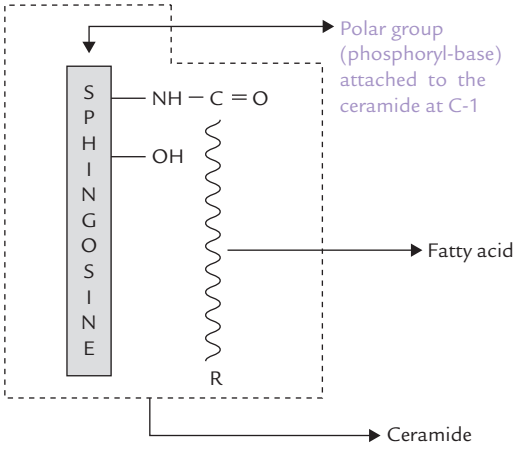


Fig. 3.10. Structure of a sphingomyelin: Ceramide + phosphoryl base (linked to C_1).

Sphingosine + Fatty acid = Ceramide
Ceramide + Phosphoryl base = Sphingomyelin

Like phosphoglycerides (and other sphingolipids, discussed below) the sphingomyelin molecule has two hydrophobic tails. Only one of these comes from a fatty acid; the other one is the hydrocarbon tail of sphingosine itself.

Core structure of all sphingolipids is ceramide, a long chain fatty acid derivative of sphingosine. Sphingomyelins are the only sphingo-lipids that contain a phosphate group.

Recall that **sphingomyelins can be grouped together with glycolipids as sphingolipids** (both contain sphingosine; Table 3.3). They serve as essential membrane components, as discussed in the next section.

Glycolipids

The glycolipids are formed by attachment of a carbohydrate component (mono- or oligosaccharide) to ceramide

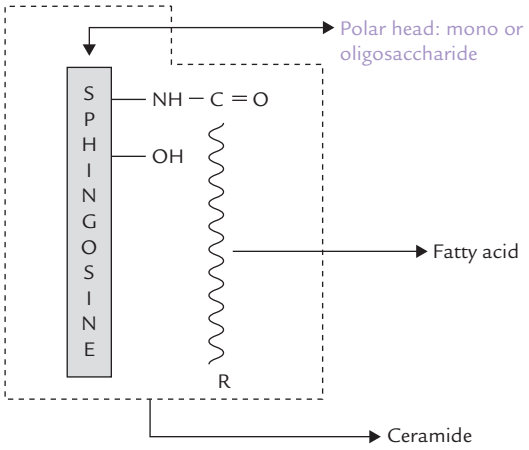


Fig. 3.11. Structure of a glycolipid.

(Fig. 3.11). They are commonly referred to as **glycosphingolipids** (Greek word, *glykos* means "sweet") because of presence of sphingosine backbone. Depending on the nature of the carbohydrate component attached, four types of glycolipids are recognized: **cerebrosides, sulphatides, globosides** and **gangliosides** (Table 3.3).

Cerebrosides: These are ceramide monosaccharides: a monosaccharide is attached to ceramide. The monosaccharide is commonly either galactose (**galactocerebroside**) or glucose (**glucocerebroside**). Galactocerebroside is the most common cerebroside found in membranes, whereas glucocerebroside serves primarily as an intermediate in the synthesis and degradation of the more complex glycosphingolipids.

Members of a group of glucocerebrosides (or galactocerebrosides) may differ from each other in the type of fatty acid attached to the sphingosine. The fatty acids commonly present are the C-24 fatty acids, such as lignoceric acid, nervonic acid, cerebronic acid and oxynervonic acids. The cerebrosides formed by these four fatty acids are kerasin, nervon, cerebron and oxynervon respectively.

Sulphatides: These are sulphuric acid esters of galactocerebroside, found predominantly in nervous tissue.

Globosides: They contain two or more sugars (oligosaccharide chain) attached to ceramide.

Gangliosides: They are similar to globosides except that they also contain **sialic acid**. They are found primarily in the ganglion cells of the central nervous system, particularly at the nerve endings. The oligosaccharide chains of gangliosides contain hexoses, hexosamines, and characteristically one or more molecules of a sialic acid, a C-9 sugar, also called **N-acetylneuraminic acid** (NANA). It is an acid sugar derivative that characteristically occupies terminal positions in the oligosaccharide chain.

Gangliosides can be classified on the basis of the number of component NANA molecules and by the length (and sequence) of their oligosaccharide units. GM means a ganglioside with a single (mono) NANA residue, whereas GD, GT and GQ would indicate two, three and four NANA residues in the molecule, respectively. The number after GM—for example, GM_1, GM_2 or GM_3—indicates the sequence of sugars:

1. represents the sequence Gal-GalNAc-Gal-Glc-ceramide,
2. means GalNAc-Gal-Glc-ceramide, and
3. is Gal-Glc-ceramide. The structures of GM_3 and GM_2 ganglioside, for example, are shown below:
   GM_3: NANA-Gal-Glc-ceramide
   GM_2: NANA-GalNAc-Gal-Glc ceramide

The numbers (GM_1, GM_2, GM_3) are derived from the relative mobility of the glycolipids on thin layer chromatography—GM_1 are the largest and therefore, migrate most slowly.

The complex structure of gangliosides is built up stepwise, one sugar residue at a time, in the Golgi apparatus. These complex sugars are degraded by a series of *exoglycosidases* in lysosomes. They are of medical interest because several lipid storage disorders involve the accumulation of NANA-containing glycosphingolipids in cells. These disorders lead to serious impairment of development of nervous system (Chapter 11).

Glycolipids, the sugar containing lipids, are derived from sphingosine. The amino group of sphingosine is acylated by a fatty acid, and its primary hydroxyl group attaches with the sugar unit(s).

Functions of Glycolipids

1. The glycolipids are essential components of **biological membranes** (though less abundant than phosphoglycerides), with highest concentration in plasma membrane. Maximum amount is found in nervous tissue: galactocerebrosides in sulphated form are important constituents of myelin, and gangliosides and galactocerebrosides are abundant in grey matter of the brain. They are typically located in the outer layer of the plasma membrane, where they interact with the extracellular environment.
2. Sphingolipids are known to play a role in **cell-cell interactions**, **growth**, and **development**. They are highly antigenic, and have been identified in several tumour antigens, blood group antigens, and the embryonic antigens specific for particular stages of fetal development.

3. The gangliosides act as **receptors** for toxic agents and some pathogens such as *Vibrio cholerae*, influenza virus, and tetanus toxin. In various malignant disorders, when the cells lose control of cell division and growth, there is a dramatic change in the glycolipid composition of the membrane.

C. Derived Lipids

Unsaponifiable compounds obtained by hydrolysis of simple or compound lipids are called derived lipids, e.g. steroids, ketone bodies, and quinones (Table 3.3).

Steroids

Steroids *are complex molecules consisting of four fused carbon rings.* The designation of rings and numbering of carbon atoms is shown in Figure 3.12. There is a phenanthrene nucleus made of three six-membered rings (A, B and C rings) and a cyclopentane which forms D ring. The steroid nucleus in a fully saturated form is known as cyclopentanopherhydrophenanthrene. *The alcohol derivatives of steroids, in which one or more* OH *groups are present in the steroid nucleus, are termed* ***sterols***. Cholesterol, ergosterol, coprosterol and sitosterol are some important sterols.

Unlike other membrane lipids that contain two wriggly hydrophobic tails, cholesterol (as also steroid hormones and vitamin D) has stiff carbo-cyclic ring structure–the steroid ring system with a solitary hydroxyl group.

In animal tissues, **cholesterol is the major sterol**. It has a single polar head group, i.e. the hydroxyl group at the C-3 position and the rest of the molecule is non-polar. It is present in two forms in serum: free and esterified. The latter is esterified with a fatty acid molecule at the third carbon hydroxyl group. The esterified cholesterol is an important component of plasma lipoproteins and the outer cell membrane.

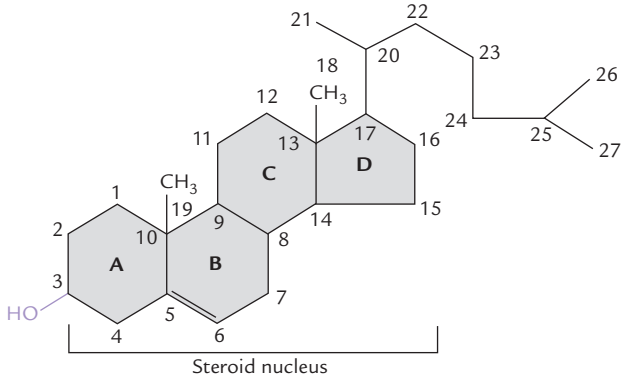


Fig. 3.12. Structure of cholesterol.

Phospholipids, glycolipids and cholesterol are the three major kinds of membrane lipids. All are amphipathic; the first two form bimolecular sheets in aqueous media, and cholesterol is most hydrophobic of the three.

Functionally, cholesterol is a very important molecule being precursor for three useful compounds: the **bile acids**, the **steroid hormones**, and **vitamin D**. Cholesterol is an essential component of all biological membranes. With its ring system and a solitary hydroxyl group, it is least water-soluble of all membrane lipids. Disorders of cholesterol metabolism play a role in aetiology of cardiovascular diseases and cholesterol is a major component of gall stones.

In plants and yeasts a different membrane steroid, phytosterol, is present, and most bacteria have no steroid at all.

Colour reactions of sterols: The chloroform solution of a sterol is dehydrated by strong dehydrating agents (e.g. acetic anhydride and strong H_2SO_4) to form coloured products. This forms the basis for detection of cholesterol by Salkowski and Liebermann Burchard reactions.

Other Lipid Derivatives

A detailed account of the other lipid derivatives such as lipoproteins, ketone bodies (Chapter 11) quinones polyisoprenoid compounds including tocopherol and carotenoids (Chapter 18) are discussed in appropriate chapters.

Exercises

Essay type questions

1. Name the phospholipids and state their biological importance. What products are released upon complete hydrolysis of a phospholipid molecule?
2. Name essential fatty acids. Which one of these is most important and why? Discuss biomedical importance of polyunsaturated fatty acids.
3. Give an account of lipid-classification giving examples from each class/subclass. Describe structure and functions of cholesterol in human body.

Write short notes on

1. Essential fatty acids
2. Sphingolipids
3. Respiratory distress syndrome

CLINICAL CASE

CASE 3.1 Shortness of breath in a newborn

A 25-year-old woman, with a history of insulin dependent diabetes mellitus for the last 15 years, was registered in the antenatal clinic. Following evidence of fetal distress at 34th week of gestation, caesarian section was performed and a male baby was delivered. At birth, the baby presented with tachypnoea (respiratory rate >70 per minute), expiratory grunting, flaring of alae-nasi, retraction of ribs and sternum, and cyanosis. Radiological examination showed prominence of bronchial air shadows and a generalized opacity (ground glass appearance). Biochemical analysis of blood sample did not show any abnormal results. The baby was shifted to the neonatology unit for further management.

Q.1. What is the most probable diagnosis? State the underlying biochemical lesion in this condition.
Q.2. How would you diagnose this condition in the antenatal period?
Q.3. Discuss the biochemical basis of the clinical features.
Q.4. Outline the treatment of the child.

AMINO ACIDS, PEPTIDES AND PROTEINS

CHAPTER

4

Proteins are the most abundant macromolecules in living cells. The name protein means "first" or "foremost". The body proteins constitute 50% or more of dry weight of the living cell and exhibit enormous diversity; hundreds of different types of proteins are usually found in a single cell. All proteins are made up of a set of building block molecules, called **amino acids**, which are covalently linked to one another. Peptides are the short amino acid sequences linked covalently.

After going through this Chapter the student should be able to understand:

- Basic structure, classification, general characteristics and acid-base properties of amino acids, application of these properties for separation of individual amino acids in a mixture.
- General structure and properties of peptides, formation and characteristics of peptide bond, and role of some biologically active peptides in the body.
- Classification, various levels of organization of proteins; primary, secondary, tertiary and quaternary structures; and methodologies for separation of proteins.

I. Amino Acids

Amino acids in free form are found in small amounts in the living systems, but the bulk exists as constituent units of proteins. More than 300 amino acids have been identified in nature, but all body proteins are constructed from a basic set of only 20 amino acids, which are specified by genetic code. These DNA coded amino acids, called **primary amino acids**, are regarded as alphabet of protein structure. It is amazing that all body proteins, which display high structural diversity, could be made up of such a limited number of amino acids.

Proteins and peptides are formed from a string of amino acids linked together by peptide bonds. The latter term refers to amide bond between two amino acids.

A. Composition

Amino acid are structures that have an amine and an acid group. Amino acids have a central carbon, the **α-carbon**, that has bonds to an amino group, a carboxylic acid, a hydrogen and a variable side chain designated as R group (Fig. 4.1). Because of their position on the α-carbon, the carboxyl and amino groups are called the α-carboxyl and α-amino groups, respectively. The side chain R defines chemical nature and structure of different amino acids; for example, it is $-CH_3$ in alanine, $-CH_2OH$ in serine, CH_2COOH in aspartic acid, and –H in glycine. Thus,

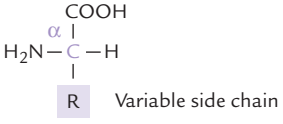


Fig. 4.1. Simple amino acid.

R group represents the variable portion which differs from one amino acid to another.

B. Stereochemistry

The α-carbon has four different groups attached to it and so is a **chiral** or **asymmetric carbon** Hence, there are two possible enantiomers, L and D, which represent non-superimposable mirror images and are optically active. Nearly all amino acids occurring in proteins are of the L-form. D-amino acids are rare in nature although they do occur in some bacterial products. This is opposite of sugars which nearly always occur as the D isomer. Note that the L designation has nothing to do with the way an amino acid rotates the polarized light; for example L-leucine rotates polarized light 10.4° to the left, whereas L-arginine rotates polarized light 12.5° to the right (the enantiomers of these compounds rotate polarized light to the same degree but in the opposite direction). The designation (D and L) is purely structural, being based on the structure of L-glyceraldehyde (Fig. 4.2).

The α-carbon of all amino acids is asymmetric, except for glycine (Fig. 4.3), which, therefore, lacks the mirror image pair of enantiomers and is not optically active.

Numbering of carbons in amino acids: The various carbon atoms are assigned sequential letters in Greek alphabets, beginning with the carbon next to the carboxyl group (Fig. 4.2). Standard numbering schemes are also used; for example the carbon atoms in Figure 4.2 are labelled 1–6 and α through ε.

Amino acids are chiral molecules. Only L-amino acids are found in proteins (D-forms occur in bacterial peptides).

C. Classification

The 20 amino acids in proteins encoded by DNA are shown in Table 4.1; each can be designated by a three-letter abbreviation and one-letter symbol. They are grouped as per the classification schemes based on:

(a) polarity and charge on R groups,
(b) structure of side chains,
(c) catabolic fate of the amino acid,
(d) body's ability to synthesize the amino acid.

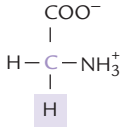


Fig. 4.3. Glycine (R = H). It's α-carbon atom is not asymmetric.

Table 4.1. The primary amino acids

Primary amino acid	Three-letter abbreviation	One-letter symbol
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

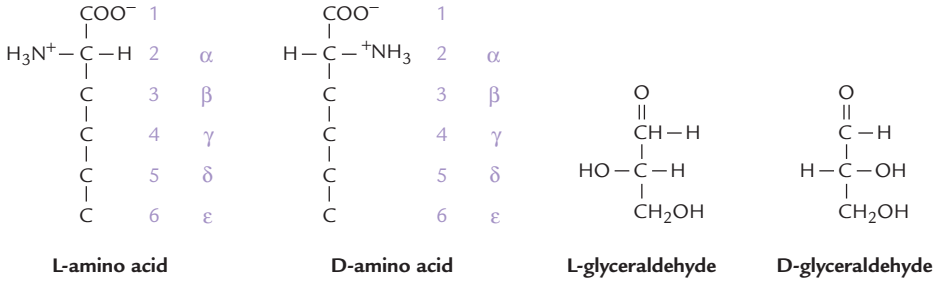


Fig. 4.2. Stereochemistry and numbering of carbons in amino acids.

Table 4.2. Amino acids according to polarity and charge of their R groups (side chains) at pH 7

Non-polar R groups	Polar but uncharged R groups	Charged polar R groups
Alanine	Serine	**Negatively charged**
Valine	Threonine	Aspartic acid
Leucine	Glycine	Glutamic acid
Isoleucine	Asparagine	**Positively charged**
Phenylalanine	Glutamine	Lysine
Tryptophan	Tyrosine	Arginine
Methionine	Cysteine	Histidine
Proline		

Based on Polarity and Charge on R Groups

The properties of an amino acid are, in large part, dependent on its side chain (–R), which uniquely defines each of the 20 amino acids. Based on polarity and charge on the side chain, the primary amino acids are allocated among three different subgroups (Table 4.2).

Non-polar R Groups

Eight amino acids are classified as having non-polar side chains (R group). In proteins these amino acid residues are usually buried in the hydrophobic interior of the biomolecule and are out of contact with water. Alanine, valine, leucine and isoleucine have **aliphatic hydrocarbon side chains** ranging in size from methyl group for alanine to isomeric butyl groups for leucine and isoleucine (Fig. 4.4). *Alanine is the most abundant amino acid in most proteins*. Phenylalanine with its phenyl moiety and tryptophan with its indole group contain **aromatic side chains**, and (together with the aliphatic amino acids) contribute to the internal hydrophobic interactions of the proteins. They are also responsible for the ultraviolet absorption of most proteins, discussed later. Methionine has a **thiol ether side chain.** Proline differs from other amino acids in that its side chain **pyrrolidine ring** includes both the α-carbon and the α-amino group. Chemically speaking proline is not an α-amino ($-NH_2$) acid but rather an α-imino (–NH) acid.

Uncharged Polar R Groups

Side chains of these amino acids are uncharged, and they have polar groups (–OH, –SH, –NH, C = O) that can hydrogen bond to water. Serine and threonine, for example, bear **hydroxylic groups** of different sizes and glycine has **hydrogen** for its R group. Asparagine and glutamine have **amide-bearing side chains** of different sizes. Tyrosine has a **phenolic group** (and like phenylalanine and tryptophan is aromatic). Cysteine has a **thiol group** that can form disulphide bond with another cysteine through oxidation of the two thiol groups to form cystine (Fig. 4.5).

Charged Polar R Groups

Five amino acids have charged side chains (Fig. 4.4), having acidic or basic groups. These groups can assume respectively, negative or positive charge at physiologic pH values.

1. The **basic amino acids** have a **positive charge** on their R group at physiological pH values (because the R groups are protonated). Lysine has a primary amino group attached to the terminal ε-carbon of the side chain (i.e. butylammonium side chain), which has pK′ of 11. Arginine is the most basic amino acid (pK′ = 13) and its **guanidine group** exists as a protonated guanidinium ion at pH 7.0. Histidine, which carries an **imidazoline ring** as the side chain functions as a general acid-base catalyst. This is because with its pK′ = 6.0, it ionizes within the physiological pH range. In contrast to lysine and arginine, which are fully charged at the physiological pH, histidine is only partially charged; its side chains being weakly basic.
2. The **acidic amino acids**, have **negative charge** at physiologic pH, e.g. glutamic acid and aspartic acid. Both contain carboxylic acids on their side chains and are ionized at pH 7.0, and as a result, carry negative charges on their β- and γ-carboxyl groups, respectively. In their ionized state, they are referred to as aspartate and glutamate (asparagine and glutamine are, respectively, the amides of aspartic acid and glutamic acid).

The allocation of the 20 amino acids among the three different subgroups is somewhat arbitrary. The following examples illustrate this point:

- Tryptophan with its heterocyclic aromatic R group may be thought of as uncharged polar amino acid (not uncharged non-polar amino acid as in Table 4.2).
- Glycine with its smallest R group might as well be classified as non-polar amino acid.
- Side chains of tyrosine and cysteine are ionizable, particularly at higher pH values, and so they might be classified as charged polar amino acids.

Based on Structure of Side Chain

Structures of side chain R groups of the 20 primary amino acids have been shown in Figure 4.4, which forms the basis of their classification into the following subgroups:

- **Aliphatic amino acids:**
  Branched chain amino acids: Valine, leucine, isoleucine
  Sulphur containing amino acids: Methionine, cysteine
  Amide group containing amino acids: Asparagine, glutamine
  Hydroxy amino acids: Serine, threonine, tyrosine
  Simple amino acids: Glycine and alanine

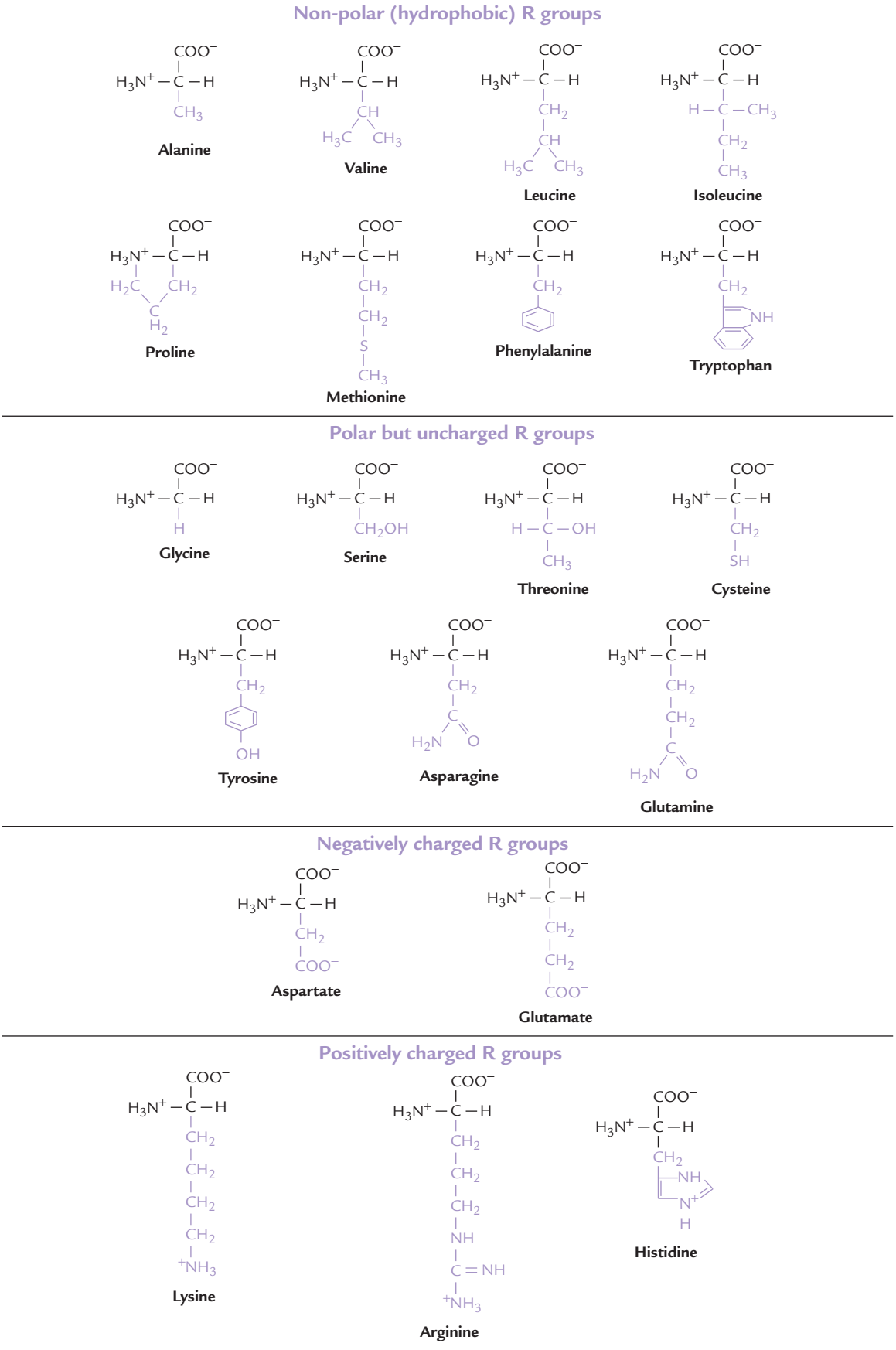


Fig. 4.4. The common amino acids. They are shown with their amino and carboxyl groups ionized as they occur at pH 7.0. Portions in black are common to all amino acids, remaining portions are their R groups.

- **Aromatic amino acids:** Phenylalanine, tyrosine, tryptophan
- **Imino acid:** Proline
- **Heterocyclic amino acids:** Histidine, tryptophan
- **Dicarboxylic amino acids:** Aspartic acid and glutamic acid
- **Dibasic amino acids:** Arginine, lysine

Tryptophan finds place in two subgroups (aromatic and heterocyclic), which again shows arbitrary nature of the classification schemes.

Based on Catabolic Fate of the Amino Acid

From the viewpoint of their catabolic fate, amino acids may be divided into three categories: those that can give rise to glucose and glycogen, those that can give rise to ketone bodies and those that can give rise to both. They are called the **glucogenic,** the **ketogenic,** and both **glucogenic** and **ketogenic** amino acids, respectively. Examples of amino acids of each category are given later (Chapter 13).

Based on Body's Ability to Synthesize the Amino Acid

Two major subclasses are recognized—essential and non-essential amino acids.

Essential amino acids cannot be endogenously synthesized, and therefore, their dietary intake is essential. Examples are leucine, isoleucine, methionine, phenylalanine, lysine, tryptophan, valine and threonine. The **non-essential amino acids**, on the other hand, can be endogenously synthesized in adequate quantities so as to meet the body's requirements. Examples include alanine, aspartate, asparagine, glutamate, glutamine, tyrosine, serine, proline, glycine, and cysteine.

Two amino acids, histidine and arginine, are called the **semi-essential amino acids** (Chapter 13).

Specific Roles of Some Side Chains

Properties of the amino acids are determined by the functional groups in their side chains and by the non-covalent interactions that the latter can form. In addition, the functional groups may perform some specific roles, thereby imparting special properties to proteins.

1. *The hydroxyl groups of serine and threonine* exist at the catalytic sites of certain kinds of enzyme, e.g. the enzymes that regulate energy metabolism and fuel storage in the body. They can be phosphorylated, hence their charge changes from neutral to negative, which changes the shape and function of the enzyme proteins. Thus, these (hydroxyl) groups are important in the regulation of activities of these enzymes (Chapter 5).

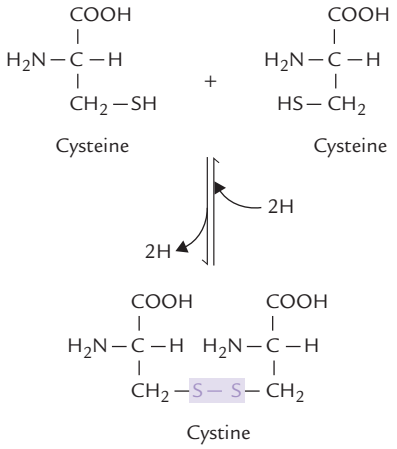


Fig. 4.5. Formation of cystine.

2. *Thiol group of a cysteine* can form a disulphide bond with another cysteine through the oxidation of the two thiol groups (Fig. 4.5). The dimeric compound so formed, called cystine, is important in cross-linking the adjacent polypeptide chains.
3. *Amide bearing side chain of asparagine* forms covalent linkage with oligosaccharide unit (called **N-glycosidic linkage**) in glycoproteins. The –OH groups of serine and threonine also similarly form linkage with reducing end of oligosaccharide, called **O-glycosidic linkage** (Chapter 5).
4. *Aromatic side chains* are responsible for the *ultraviolet absorption* of most proteins, which have absorption maxima between 275 nm and 285 nm. Tryptophan has greater absorption in this region than the other two aromatic amino acids. Since nearly all proteins contain aromatic amino acids, the amount of light absorbed at 280 nm by a protein is used as an indirect measure of protein concentration.
5. *Hydroxyl group of tyrosine* is subject to phosphorylation in some enzymes, which is important in the regulation of metabolic pathways.
6. *Imidazole group of histidine* is important in the buffering activities of proteins.
7. *Cyclic pyrrolidine group of proline* introduces bend in the peptide chain.
8. *Butylammonium side chain of lysine* binds with the co-enzymes pyridoxal phosphate and biotin.
9. *R group of glycine* provides little stearic hindrance because of its small size, so that proteins can bend or rotate easily wherever glycine forms part of their structure.

Chemically, proline is α-imino acid (not α-amino acid). Its side chain pyrrolidine ring includes both the α-amino group and α-carboxyl group. It forces a bend in a polypeptide chain.

D. Non-standard Amino Acids

In addition to the 20 primary amino acids found in protein, some "non-standard amino acids" are present in small amounts in specialized structures. They may also occur in free or combined states and independently play a variety of biological roles.

1. *Non-standard amino acids found in proteins*: These amino acids are produced by specific modification of a primary amino acid residue after the polypeptide chain has been synthesized. They are important, and in most cases, essential for the function of the protein. Hydroxylation, methylation, acetylation, carboxylation and phosphorylation are some common modifications, though more elaborate modifications are found in some amino acid residues. Some examples of modified amino acids are: *hydroxylysine* and *hydroxyproline* in collagen, *methylhistidine* in muscle proteins, *phosphoserine* in casein and certain enzymes, γ-*carboxyglutamic* acid in prothrombin and other calcium-binding proteins, and *pyroglutamic* acid as the N-terminal amino acid in thyrotropin and several proteins.
2. *Non-standard amino acids not found in proteins*: Termed non-protein amino acids, they occur in free state in cells, e.g. *ornithine* and *citrulline* that are intermediates of urea cycle. The most abundant amino acid in human organism, *taurine*, also occurs in a free state in bile and plays an important role in fat digestion and absorption.
3. *Biologically active amino acids*: These amino acids are used for functions other than protein synthesis. Examples: (a) as chemical messengers for communication between cells, e.g. glycine, dopamine (tyrosine derivative), γ-aminobutyric acid (glutamate derivative; refer to Box 4.1), (b) as local mediator of allergic reactions, e.g. histamine (histidine derivative), and (c) as a hormone, e.g. thyroxine (another tyrosine derivative).
4. *D-Amino acids*: They are components of bacterial polypeptides that are widely distributed as constituents of bacterial cell walls. They are also found in many bacterially produced peptide antibiotics, including valinomycin and gramicidin A.

E. Acid-Base Properties of Amino Acids

Amphoteric Nature

An amino acid is capable of acting as both an acid (i.e. proton donor) and a base (i.e. proton acceptor). Such substances are termed as ***amphoteric substances***. The amphoteric nature of amino acids is because of presence of the following ionizable groups:

1. *Carboxyl group*, which can donate a proton, therefore acting as an acidic group.
2. *Amino group*, which is capable of accepting a proton, thereby acting as a basic group.

Note: In case of acidic amino acids, an additional acidic group (i.e. carboxyl group) is present, and in basic amino acids, an additional basic group (i.e. amino group) is present.

Each Amino Acid can Exist in Three Charged States: Positive, Neutral, Negative

The three charged states of an amino acid are represented in Figure 4.6. These are the diprotonated form (positively charged, cationic), the deprotonated form (negatively charged, anionic) and the dipolar form (neutral, zwitterion). The form in which an amino acid exists at a given time is determined by relative values of the following two parameters:

1. Isoelectric pH of the amino acid.
2. pH of the surrounding medium.

Isoelectric pH (pI) *of an amino acid is defined as the pH value at which carries equal number of positive and negative charges, so that net charge on the molecule is nil*. Predominant form of an amino acid at its pI is **zwitterion** (in German Zwitter means "hermaphrodite").

For a simple amino acid such as glycine, the pI is halfway between the pK′ values of the two ionizable groups, and so calculated by averaging the two pK′ values.

$$\text{pI} = \frac{\text{pK}' \text{ of carboxylate group} + \text{pK}' \text{ of amino group}}{2}$$

The pK′ values for the carboxylate and the amino groups being 2.4 and 9.8 respectively, the pI is 6.1.

The pI of acidic amino acids is halfway between the pK′ of two acidic (carboxylate) groups, and the pI of basic amino acids is half-way between the pK′ of two basic (amino) groups.

Acidic amino acids thus have low pI values, basic amino acids have high pI values, and neutral amino acids have pI values near 6.0 (Table 4.3).

When the pH of the surrounding medium is lowered below the pI value of an amino acid (i.e. concentration of protons raised), its carboxyl group accepts a proton to produce the protonated form. This form of amino acid carries net positive charge.

$$\text{pH} < \text{pI} \rightarrow \text{Amino acid is positive}$$

Conversely, elevation of pH (i.e. low concentration of protons) results in loss of proton from the anionic form.

BOX 4.1

Amino Acids and their Derivatives as Neurotransmitters

Some amino acids and amino acid derivatives serve as neurotransmitters—the extracellular messengers that transmit a targeted message from a neuron to responding cells. A neurotransmitter is released by a neuron at a synapse to effect another post-synaptic neuron or organ (e.g. heart, lungs).

Amino acids:

1. *Glycine* is the major inhibitory neurotransmitter in the brainstem and spinal cord. When released onto a neuron, it hyperpolarizes the neuron and hence decreases its activity.
2. *Glutamic acid* is the major excitatory neurotransmitter. It is released by a large number of neurons and is very important in the region of the brain which influences memory.

Amino acid derivatives:

1. *Dopamine* is derived from tyrosine during catecholamine biosynthesis. It acts as a neurotransmitter in brain, where it has many functions, e.g. controlling blood pressure. Increase in dopamine activity causes aggressiveness, and may lead to Schizophrenia or Huntington's chorea.
2. *Norepinephrine*, derived from tyrosine, is an important hormone and neurotransmitter. As a hormone, it is a part of the fight or flight response. It is the main neurotransmitter of the branch of the autonomic nervous system called the sympathetic nervous system. It plays several roles as a neurotransmitter in the brain.
3. *Epinephrine*, also derived from tyrosine, is similar to norepinephrine in many of its functions. It is released as a hormone in the **fight-or-flight response** by the adrenal medulla. It is a major neurotransmitter in certain regions of the brain.
4. *Serotonin* (5-hydroxytryptamine) is a very important neurotransmitter in brain. It is involved with spinal reflexes, sleep-wake cycle, flow of sensor afferents and habituation.
5. *Gamma-amino butyric acid (GABA)*, derived by decarboxylation of glutamate, is the major **inhibitory neurotransmitter** in brain. Most neurons in the brain are continually receiving *GABA* input that keeps them quiet. Patients with anxiety are prescribed Valium (benzodiazepine), which enhances the action of *GABA* on the brain.

Interestingly, the major inhibitory neurotransmitter in the brain, *GABA* is formed from a major excitatory neurotransmitter, glutamic acid.

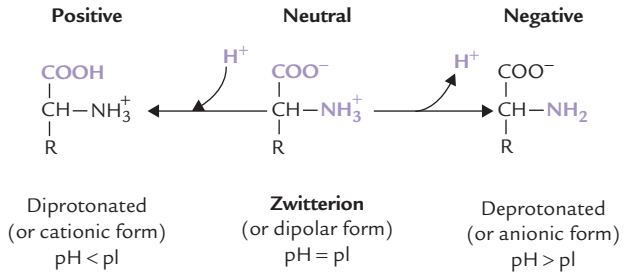


Fig. 4.6. Three charged states of amino acid.

As a result, overall charge on the amino acid molecule becomes negative.

$$pH > pI \rightarrow \text{Amino acid is negative}$$

Amino acids are zwitterionic structures, carrying both a positive and a negative charge. The pH value at which the number of positive charges equals the number of negative charges is isoelectric pH of the amino acid.

Table 4.3. Isoelectric pH (pI) values of some primary amino acids. At its pI an amino acid is neutral

Amino acid	pI
Aspartic acid	2.97
Glutamic acid	3.22
Serine	5.68
Alanine	6.02
Glycine	6.1
Histidine	7.6
Lysine	9.74

Titration Curve of a Monoamino-monocarboxylic Acid (Alanine)

The titration curve of alanine proceeds in the same way as that of a weak acid. There is, however, an important difference: the titration curve of this amino acid proceeds in two distinct stages. This is because an amino acid molecule, when fully protonated, is capable of donating two protons. Each of the two stages of the titration corresponds to loss of a proton (Fig. 4.7). In the first stage of titration, dissociation of the carboxyl group occurs while

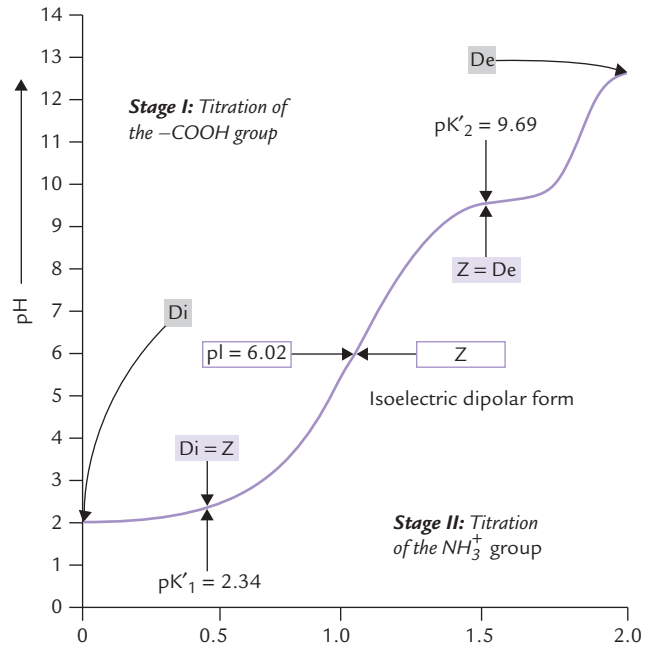


Fig. 4.7. The titration curve of 0.1 M alanine. The ionic species (Z = dipolar form, or zwitterion Di = diprotonated form, De = deprotonated form) predominating at various pH values are shown in boxes ($pK'_1 = pK'$ of the α-carboxyl group; $pK'_2 = pK'$ of the α-amino group; pI = Isoelectric pH).

the second stage involves dissociation of the amino group.

Stage I: Dissociation of the carboxyl group: At low pH both the carboxyl group as well as the amino group of the amino acid are protonated, and hence, the predominant ionic species is the diprotonated (positive) form. As the pH is raised by addition of NaOH, the COOH group donates its proton to neutralize the hydroxyl group of the alkali (i.e. sodium hydroxide). Thus, formation of the carboxyl (COO^-) ion occurs at the expense of COOH group, so that some diprotonated form is converted to the dipolar (zwitterion) form. At the end of this stage of titration, dissociation of carboxylate group is complete and the diprotonated form is almost completely converted to the dipolar (Z) form.

Midway through this stage, the following relations exist:

1. Concentrations of the Z form and the diprotonated form are equal.
2. The pH value of the solution equals the pK′ value of carboxyl group (pK′ is the negative log of the dissociation constant of the carboxyl group).

The solution has buffering power in this zone, as shown by flat shape of the titration curve. Normally, a system acts as buffer in the pH zone of pK′ ± 1. Since pK′ of the carboxyl group (pK'_1) of alanine is 2.34, it acts as a buffer in the pH zone of 2.34 ± 1 (i.e. from pH range of 1.34 to 3.34).

Stage II: Dissociation of the amino group: Dissociation of proton from the amino group of alanine occurs in this stage, resulting in the formation of the deprotonated (negative) form from the Z form (Fig. 4.11).

Midway through this stage, dissociation of amino group is half complete, and the following relations exist:

1. Concentration of the Z form equals that of the deprotonated form.
2. pH of the solution equals pK′ of the amino group (pK'_2).

The solution has buffering power at this stage as evidenced by a flat zone in the titration curve. Since pK′ of the amino group is 9.69, the second buffering zone of alanine lies in the pH zone of 9.69 ± 1 (i.e. from 8.69 to 10.69).

The following observations about the titration curve are noteworthy.

- *Alanine has two buffering zones*, the first in the pH range of 2.34 ± 1, and the second in the pH range of 9.69 ± 1. Since both lie beyond the physiological pH range, alanine lacks any buffering power in the body.
- At the end of the first stage of titration, alanine exists predominantly in the Z form. The pH value corresponding to this point is **isoelectric pH (pI)**. *It lies midway between the pK′ values of the two ionizable groups*.

$$pI = \frac{(pK' + pK'_2)}{2} = \frac{2.34 + 9.69}{2} = 6.02$$

where, pK'_1 and $pK'_2 = pK'$ values of α-carboxyl and the α-amino groups respectively, pI = isoelectric pH.

At this pH, both the amino group and the carboxyl groups of the amino acid are ionized, but the overall charge on the molecule in nil.

Titration Curve of Basic Amino Acids

These amino acids have a third ionizable amino, group in side chain. Since this group can reversibly gain or lose a proton, the basic amino acids have an additional third phase of titration. Accordingly, there are three regions of buffering; one region each corresponding to loss of a proton from the carboxyl group, the amino group, and the side-chain. The pK′ value of the imidazole group of histidine is 6.0, and therefore, it can serve as a buffer in pH range of 6.0 ± 1, which is very close to the physiological pH. Hence, histidine has the capacity to provide buffering action at the body pH.

Titration Curve of Acidic Amino Acids

These amino acids have a third dissociable group, i.e. the carboxyl group in side chain. Their titration curves, therefore, have an additional third stage corresponding to this carboxyl group.

F. Separation of Amino Acids

A number of techniques are available for the separation of a given amino acid from a mixture. The important ones discussed here are (paper) chromatography, electrophoresis, ion exchange chromatography, etc.

Chromatography: (Greek: *chroma* means colour; *graphin* means to write): It was discovered by a Polish scientist, **Tswett**, in 1901, who used this technique for separating solubilized plant pigments on solid adsorbents. In most modern chromatographic procedures, a mixture of substances to be fractionated is dissolved in a liquid (the **mobile phase**) and percolated on an inert support medium on which separation occurs (the **stationary phase**). The migrating solutes interact with the stationary phase and are differentially retarded; the retarding force depends on the properties of each solute. The solutes having higher affinity for the stationary phase are retarded more than those having relatively higher affinity for the mobile phase. The substances with different rates of migration are thereby separated.

The chromatographic procedures for separating amino acids (or proteins) are based on such interactions between the amino acids in the solution and the stationary phase. Strip of chromatography paper may be used as the inert support medium (i.e. **paper chromatography**), or alternately, a thin layer of silica spread on a glass plate (i.e. **thin layer chromatography** or **TLC**) is employed. A small quantity of the amino acid mixture is placed on a dry piece of chromatography paper (or silica-gel coated plate in case of TLC) and a solvent is allowed to permeate on it from one end (Fig. 4.8). The solvent is usually a mixture of organic solvents and water. As the solvent advances on the paper (or plate), the amino acids also travel with it. Different amino acids travel at different speeds, because each one has a different partition coefficient between the mobile phase (i.e. the solvent) and the stationary phase (i.e. the paper). In other words, the amino acid having higher affinity for the solvent moves faster than the one having lower affinity. As a result, the amino acids are separated from one another as the solvent ascends (in ascending chromatography) or descends (in descending chromatography) along the stationary phase.

The amino acids thus separated are located by spraying the paper with ninhydrin. Resolution may be improved by a second run at right angle to the first one. This is referred to as two-dimensional separation. The **two dimensional separation** may also be performed using chromatography and electrophoresis at right angle; the latter procedure has considerable diagnostic utility **(Case 17.1)**.

Electrophoresis: In electrophoresis, proteins dissolved in a buffer solution at a particular pH is placed in an electric field. Depending on the relationship of the buffer pH to its pI, a given protein moves towards the cathode (–), anode (+), or remains stationary (pH = pI).

Ion-exchange chromatography: In this technique preparative separation of amino acids by charge occurs on ion-exchange resins, which are packed in a column. Negatively charged resins bind cations and retard their movement, whereas positively charged resins retard movement of anions.

A detailed account of the latter two techniques is given later in this chapter.

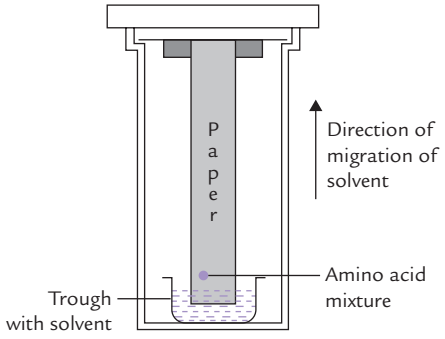


Fig. 4.8. Paper chromatography (ascending).

II. Peptides

A. Structure

*Short sequences of amino acids, linked covalently, are known as **peptides**.* The covalent linkage, termed peptide bond, is formed by reaction of the amino group of one amino acid with the carboxyl group of other (Fig. 4.9). The dipeptide so formed retains a carboxyl and an amino group, which are free to form peptide linkages with other amino acids. In this fashion, peptides of various lengths may be formed. On the basis of number of amino acids they contain, shorter peptides are termed di-, tri-, tetra-peptides, and so on. Peptides consist of a few

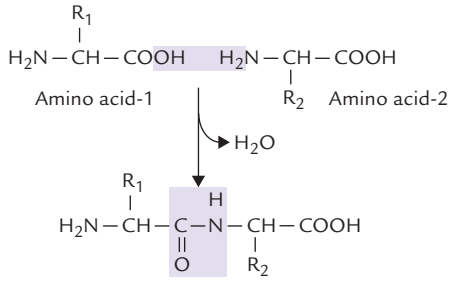


Fig. 4.9. Formation of a peptide bond.

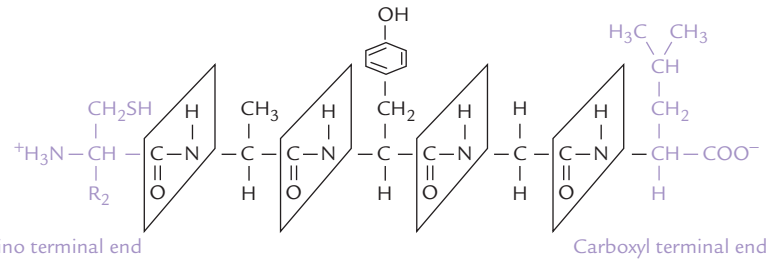


Fig. 4.10. Structure of a pentapeptide, Cys-Ala-Tyr-Gly-Leu. The peptide is shown in the ionized state in which it occurs at physiological pH.

(3 to approximately 50) amino acids, polypeptides contain up to 100 amino acids, and the polyamide chains larger than this are proteins.

The peptide has two ends: the amino terminal end, and the carboxyl terminal end. The amino acid present at the amino terminal end has a free amino group, and the amino acid at the C-terminal end possesses a free carboxyl group. The amino acid sequence of a peptide (**primary structure**) is always listed from the N-terminal; therefore, it is Cys-Ala-Tyr-Gly-Leu for the pentapeptide shown (Fig. 4.10). Note that the formation of peptide bond involves removal of water from the carboxyl group of one amino acid and the amino group of the other (Fig. 4.9). Therefore, the constituent amino acids of a peptide, having lost a portion of their carboxyl group and the amino group, can no longer be termed amino acids. Instead they are called the **amino acid residues**.

B. Reactions of Peptides

Acid-Base Behaviour

The acid-base behaviour of a peptide is a function of its N-terminal amino group, the C-terminal carboxyl group, and those R groups which can ionize. For example, a peptide having high content of dicarboxylic amino acids (glutamate and aspartate) has acid-like behaviour at the physiological pH. Conversely, a peptide with high content of lysine, arginine or histidine tends to have base-like behaviour at physiological pH.

Hydrolytic Cleavage

A peptide can be cleaved to its constituent amino acids by hydrolysis of the peptide linkages. Boiling with strong acids or bases brings about the hydrolysis. All the available peptide bonds can be cleaved by this treatment, irrespective of nature of the constituent amino acids.

Proteolytic Cleavage

Certain enzymes, namely trypsin and chymotrypsin, can also bring about hydrolysis of peptide bonds. Accordingly they are referred to as proteolytic, the word meaning *"protein dissolving"*. Their action is highly specific because a given enzyme attacks only certain specific regions of the peptide chain.

Labelling by FDNB

The amino terminal residue of a peptide can be labelled with 1-fluoro-2-4 dinitrobenzene (FDNB). This reaction is a useful step for determination of the amino acid sequence (i.e. primary structure) of the peptide.

C. Functions

Peptides are formed as intermediates during partial hydrolysis of the long polypeptides chains. In addition, some naturally occurring peptides are found in the living matter (Table 4.4).

Some of these peptides possess enormous biological activities; for example **thyrotropin-releasing hormone** (TRH) is a tripeptide released from the hypothalamus.

Pyroglutamate-Histidine-Proline: The TRH regulates the secretion of the thyroid stimulating hormone (TSH) from the anterior pituitary. The N-terminal glutamate residue is in the pyro form in which its γ-carboxyl group is covalently joined to its α-amino group via amide linkage. The C-terminal amino acid, proline, is amidated and thus also not free.

Another peptide of biological significance is **glutathione,** comprising glutamate, cysteine and glycine. The glutamate forms an amide bond with cysteine through its γ-carboxyl group, rather than its α-carboxyl group and so reffered to as *pseudopeptide linkage*.

γ-Glu-Cys-Gly: Glutathione acts as a co-enzyme in many reactions, e.g. for *transhydrogenases* and *peroxidases,* and participates in the second phase of the xenobiotic metabolism. When it is oxidized, many highly reactive and toxic compounds are reduced; one of the most important features of glutathione is to reduce hydrogen peroxide to water (Chapter 27).

Table 4.4. Some important biologically active peptides in humans

Peptides	Number of amino acid residues	Biological effect
Thyrotropin-releasing hormone	3	Stimulates anterior pituitary to release thyroid stimulating hormone
Glutathione	3	First line of defense against oxidative stress
Enkephalins	5	Binds with certain brain receptors to alleviate pain (i.e. opiate-like activity)
Oxytocin	9	Stimulates uterine contractions
Bradykinin	9	Inhibits inflammatory reactions in tissues
Vasopressin	9	Secreted from posterior pituitary it causes kidney to retain water from urine
Little gastrin	10	Secreted by mucosal cells in stomach; causes parietal cells of stomach to secrete the acid
Glucagon	10	Pancreatic hormone involved in glucose homeostasis (hyperglycaemic)
Substance P	10	Neurotransmitter
Gramicidin	10	An antibiotic, toxic to many microorganisms

The **enkephalins** are analgesia-inducing agents that are produced in the central nervous system (the word enkephalin means "in the head"). They bind with specific receptors in the brain cells, the receptors are the same to which the synthetic opiates like morphine, heroin and other addictive opiates bind. Therefore, the enkephalins are commonly referred to as **endogenous opiates**.

It is noteworthy, that it is the sequence of the amino acids that gives the peptides such intense biological effects while the constituent amino acids by themselves are devoid of any such activity.

Peptides formed by covalent linking of several amino acids may perform specific physiological roles.

D. Characteristics of Peptide Bonds

Some characteristics of peptide bonds are as follows.

Partial Double Bond Character

The peptide bond is conventionally described as a single bond (C–N), but it does not show the rotational freedom expected of a single bond. This rigidity is explained by the fact that the double bond between the C and the O is actually shared across the bond between the C and the N. Therefore, the peptide bond is actually resonance hybrid of two electron isomers:

Its "real" structure is halfway between these two forms. Thus, the peptide bond between the C and the N has partial characteristic of double bond and so does not rotate.

This is supported by spectroscopic measurements and X-ray diffraction studies. These studies show that **C to N bond length** (1.33 Å) of a peptide bond is approximately half-way between that of C–N single bond (~1.45 Å) and C = N double bond (~1.25 Å).

- Not only the bond length, but strength of peptide bonds also lies half-way: stronger than single bond and weaker than double bond. A consequence of the **partial double bond** character is that there is lack of rotational freedom, discussed above.

Trans-planar

The two atoms of peptide bond (C and N) have four substituents: a hydrogen, an oxygen, and two α-carbon atoms (see Fig. 4.10; an exception is the peptide bond formed by nitrogen of proline). All the substituents are fixed in the same plane and the O and the H are in the trans (opposite) positions.

Uncharged

Like other amide linkages, the peptide bond is uncharged.

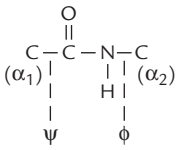
Hydrogen Bonding

The sharing of electrons between the C and the N results in unpairing of electrons between the C and the O. Thus, O has a slightly negative charge (δ^-). The electron of the H atom spends most of its time in the bond between the N and the H, which gives the H atom a slightly positive charge (δ^+). This means that these atoms, by virtue of

positive or negative charges, are able to participate in hydrogen bonding.

ϕ and ψ Angles

In addition to the peptide bond, the polypeptide backbone contains bonds between the α_1-carbon and the peptide bond carbon, and between the α-carbon and the peptide bond nitrogen. These two being single bonds show free rotation. Rotation around the nitrogen-and-α-carbon bond is measured as the *ϕ (phi) angle* and rotation around the peptide carbon and α-carbon bond as the *ψ (psi) angle*.



III. Proteins

Proteins are not only the most abundant biomolecules but are also centre of action in biological processes. They exhibit enormous diversity with regard to their structures. In general, structures of proteins consist of polypeptides, which are long, unbranched chains; each chain contains usually more than 100 amino acid residues. The chains containing fewer residues than that are simply called peptides.

A. Size

Proteins contain between 100 and 2000 amino acid residues. The mean molecular mass of an amino acid residue is about 110 dalton units (Da). Therefore, the molecular mass of most proteins is between 11,000 and 220,000 Da. The largest known polypeptide chain belongs to the 26,926 residue **titin,** a giant (2990 kD) protein that helps to arrange the muscle fibres. Some proteins consist of a single polypeptide; others are formed from two or more polypeptides held together by covalent bonds or by noncovalent interactions.

B. Functions

Range and scope of biological functions of proteins are very vast. Some important functions performed by proteins are as follows: (a) act as biological catalysts, called enzymes, (b) provide the structural framework of cells and tissues, (c) act as transport media in bloodstream for a variety of substances, such as lipids and oxygen, (d) act as hormones or regulatory proteins for controlling various biological processes, (e) perform mechanical work, such as in skeletal muscle contraction and the pumping of heart, (f) serve as essential nutrients, (g) act as antibodies in the bloodstream to provide natural defense against invading pathogens and play vital role in blood clotting mechanism, (h) regulate gene expression on chromosomes; and (i) play important role in food digestion respiration, vision, etc. These functions are summarized in Table 4.5.

Specialized functions are performed by certain proteins, which cannot be easily classified in one of the above categories. For example:

- **Antifreeze protein** is present in blood of some arctic and antarctic fishes. It lowers the freezing point of water, thereby saving aqueous plasma from freezing.
- **Reasilin** is the major component of the wing hinges of some insects. This protein has extraordinary elastic properties.

Biological functions of a given protein are same in different species. For example, haemoglobin serves as oxygen transporter in humans, mammals, birds, insects, etc. (refer to Box 4.2).

Studies have been conducted in homologous proteins obtained from different species, including mammals, fishes, reptiles, amphibians, birds, insects, plants and fungi. These studies intend to examine the biological significance of sequence homology. In one such study, **cytochrome C,** a mitochondrial protein of 100 amino acids (MW 125,000) and an essential component of electron transport chain, was studied in about 60 different species.

- At about 27 positions, identical amino acid residues were present (i.e. invariant residues). This suggests that these residues are important determinants of the biological activity of cytochrome C.

Table 4.5. Biological functions of body proteins

Protein	Example
Enzyme	*Amylase, ribonuclease,* trypsin
Structural proteins	Collagen, elastin, keratin, fibroin
Transport proteins	Haemoglobin, myoglobin, lipoproteins, serum albumin, membrane transporters
Hormonal proteins	Insulin, growth hormone, corticotropin
Regulatory proteins	Repressors and inducers
Contractile or motile proteins	Actin, myosin, tubulin
Nutrient and storage proteins	Gliadin (wheat), ovalbumin (egg), casein (milk)
Defense proteins	Antibodies
Blood proteins	Fibrinogen, thrombin
Toxin protein	Snake venoms
Vision protein	Rhodopsin

BOX 4.2

Homologous Proteins

The proteins that perform the same (or analogous) function in different species are called homologous proteins. They exhibit considerable structural similarities. For instance, haemoglobin serves as an oxygen carrier in different species, and the haemoglobin molecules isolated from these species have been found to have nearly the same chain length. Moreover, their primary structure (i.e. amino acid sequence) shows significant resemblance. A closer examination of the primary structure reveals that many of the (amino acid) positions are occupied by the same amino acid residues in nearly all the species studied. Such amino acids are called the **invariant residues.** These residues reflect similarities in the amino acid sequence of homologous proteins (called *sequence homology*). They indicate a common evolutionary origin of the homologous proteins. However, in some other positions there may be considerable variation in amino acids from one species to another. Such non-identical amino acid residues are called **variable residues.**

Table 4.6. Variation in positions of A_8, A_9 and B_{30} of insulin in different species

	A-chain (8, 9, 10)	B-chain position (30)
Human	Thr-Ser-Ile	Thr
Porcine	Thr Ser Ile	Ala
Bovine	Ala-Ser-Val	Ala
Dog	Thr-Ser-Ile	Ala

- The amino acid residues in other positions showed variation: the greater the phylogenic difference between the species, more the variation. Such information is helpful in getting insight into the evolutionary process.

Human insulin (a peptide of 51 amino acids) also exhibits sequence homology with the insulins obtained from other mammals (Table 4.6). For instance, human insulin and porcine insulin have identical amino acids in 50 positions, only at position B_{30} the amino acid is different. It is threonine in human insulin but alanine in porcine insulin. This permits use of porcine insulin for the treatment of diabetic patients. However, prolonged use of porcine insulin may result in side effects, which can be avoided by using modified insulins **(Case 4.1)**.

In bovine insulin, 8th and 10th amino acids are different in A chain also.

C. Classification

Proteins can be classified in different ways, based on the following criteria:

Based on Shape

Based on their three-dimensional shape (i.e. *conformation*), the proteins are divided into two classes:

Fibrous Proteins

The polypeptide chains of these proteins extend along a longitudinal axis without showing any sharp bends of folding (Fig. 4.11). Simplicity of structure of these proteins is responsible for their mechanical properties. They are essential components of several proteins, including collagen and elastin (for details see Chapter 5).

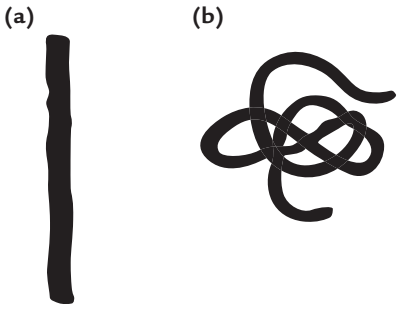


Fig. 4.11. Classification of proteins based on their shape. **(a)** Fibrous protein, **(b)** Globular protein.

Globular Proteins

Polypeptide chain of a globular protein is tightly folded and packed into a compact structure (Fig. 4.11). The hydrophilic groups of the constituent amino acids are exposed to the exterior, so that these proteins can diffuse through aqueous systems. Most enzymes, transport proteins, nutrient proteins, antibodies and hormones are globular.

Based on Structural Components

Two major classes of proteins are recognized depending on the nature of their structural components.

Simple Proteins

Simple proteins comprise solely of amino acids; no other chemical group is present in them. Pancreatic *ribonuclease* is an example.

Conjugated Proteins

Conjugated proteins contain a polypeptide group as well as a non-polypeptide group. In these proteins, the

non-polypeptide group is called the **prosthetic group,** and the polypeptide is called the **apoprotein**. A prosthetic group is often employed when the biological function of the protein requires a functional group that is not available in any of the 20 amino acids. For example, in oxygen transport proteins, haemoglobin and myoglobin, binding of oxygen occurs with the iron-porphyrin group. The prosthetic group is sugar in glycoproteins (Greek: *glycos* means sweet), lipid in lipoproteins, and a metal in metalloproteins (Table 4.7).

Note: The protein component and the prosthetic group of the conjugated proteins are covalently linked. However, lipoproteins are exceptions since the lipid and the protein components form molecular complexes, without being covalently linked.

D. Protein Structure

Every protein has a unique three-dimensional structure, which is referred to as its *native conformation.* Protein conformation is complex and is best analyzed by considering it in terms of the following organizational levels; primary, secondary, tertiary, and quaternary (Fig. 4.12a–d).

Primary Structure

Primary structure of proteins refers to the specific sequence of amino acids in a polypeptide chain. The amino acids are covalently linked to one another, e.g. Ala-Gly-Ser-Leu (Fig. 4.12a).

Thus, the polypeptide chains having different sequences of amino acids are said to differ with respect to their primary structure. It is noteworthy that it is the sequence of amino acids, and not the composition, that determines the primary structure. Thus, two polypeptide chains having identical set of amino acids, but arranged in different sequences, are said to have different primary structures.

Table 4.7. Conjugated proteins and their prosthetic groups

Class	Prosthetic group	Example
Lipoproteins	Lipids	Chylomicron
Glycoproteins	Carbohydrates	γ-Globulin of blood
Phosphoproteins	Phosphate groups	Casein of milk
Haemproteins	Haem (iron porphyrin)	Myoglobin
Flavoproteins	Flavin nucleotides	*Succinate dehydrogenase*
Metalloproteins	Iron	Ferritin
	Zinc	*Alcohol dehydrogenase*
	Calcium	Calmodulin

Such sequence variation is the most important element of protein diversity; for instance, a polypeptide with 100 amino acids can have 20^{100} different amino acid sequences.

The covalent backbone of protein, called its primary structure, is specified by the amino acid sequence. The most important element of proteins diversity is variability of the amino acid sequence.

Primary structure of insulin is shown in Figure. 4.13. The specific information that determines the final three dimensional form adapted by a protein is inherent in its primary structure which determines the function of a protein (**Case 4.2**). A slight change in the primary structure of a protein (e.g. modification of a single amino acid residue) results in loss of its biological activity, in sickle cell anaemia, change in primary structure results in deranged function of the affected protein (**Case 17.1**).

Secondary Structure

*The **secondary structure** refers to the local spatial arrangement of a polypeptides backbone atoms (without regard to the conformations of its side chains). It includes mainly the regional bends and local folding.* The **α-helix** and **β-pleated** sheets are some examples of the secondary structures that are commonly encountered in some proteins or portion of proteins.

α-Helix

It is a spiral structure consisting of a polypeptide chain that is coiled around a longitudinal axis in a helical fashion (Fig. 4.14a). The α-helical structure was first defined by Linus Pauling from his studies on fibrous proteins. However, short stretches of α-helix are found in the globular proteins as well. A vast variety of proteins contain the α-helical structure. They constitute almost the entire dry weight of hair, wool, feathers, and scales and form much of the outer layer of skin.

Characteristics: The polypeptide chain is coiled around the longitudinal axis to form the right handed helix (i.e. the coils turn in a clockwise fashion around the axis). Each turn of the α-helix contains 3.6 amino acid residues. The peptide bonds linking the successive amino acids are coiled inside and their R groups extend outward from the central longitudinal axis. This arrangement minimizes the stearic interference of the R groups with each other.

Forces stabilizing the α-helix: Two types of linkages stabilize the α-helix. These are hydrogen bonds and disulphide bonds.

- ***Hydrogen bonds:*** The hydrogen bonds are formed between the peptide bonds. Each peptide bond C = O

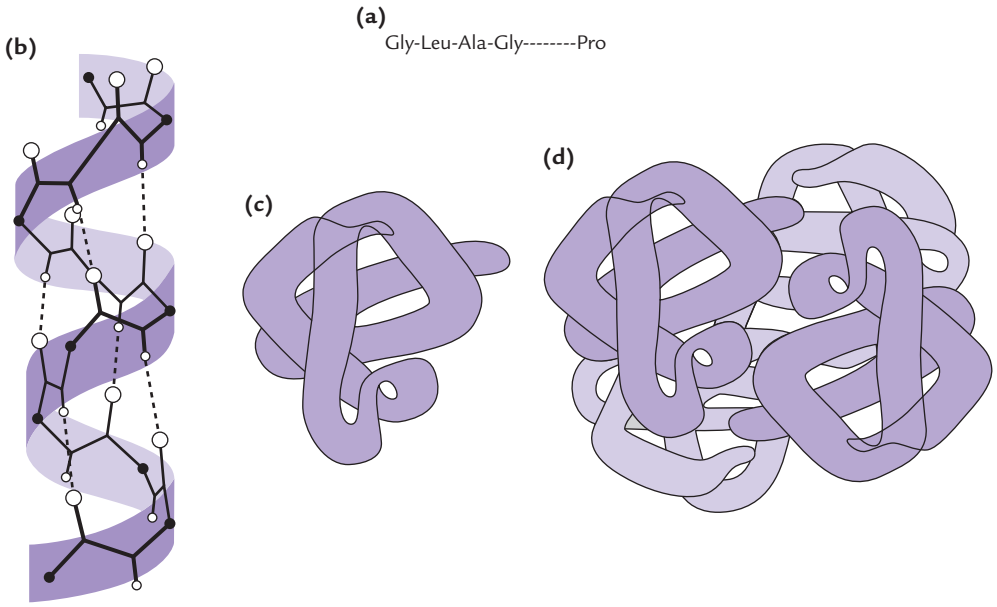


Fig. 4.12. Levels of protein structure. **(a)** Primary structure refers to the amino acid sequence in a polypeptide chain, **(b)** Secondary structure is the local spatial arrangement of the polypeptides backbone atoms, **(c)** Tertiary structure refers to overall folding pattern of the entire polypeptide chain, **(d)** Quaternary structure is formed by interactions in a non-covalent manner between different polypeptide subunits of a protein.

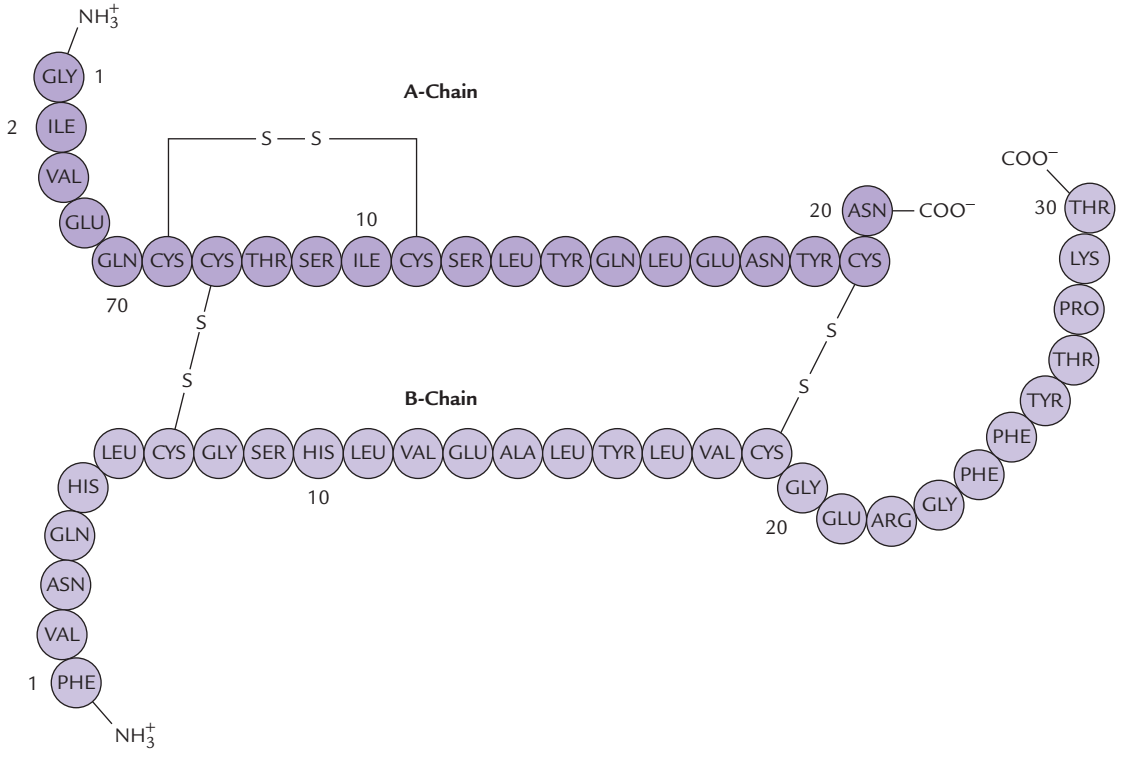


Fig. 4.13. The primary structure includes the location of covalent disulphide bonds between cysteine residues in addition to the linear sequence of amino acids, as shown in insulin.

is hydrogen bonded to the peptide bond N–H, four amino acid residues ahead of it. Each C = O and each N–H in the main chain is thus involved in hydrogen bonding (Fig. 4.14b).

- ***Disulphide bonds***: Because the polypeptide chain is rich in cysteine residues, disulphide cross-links may form between the cysteine residues of adjacent helical coils. This enhances strength of the structure (Fig. 4.14c).

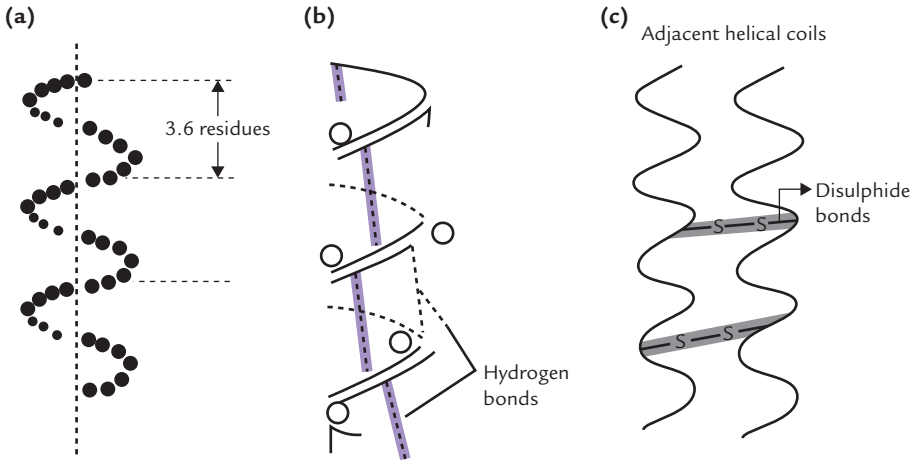


Fig. 4.14. The α-helical conformation. **(a)** A right handed helix with 3.6 residues per turn, **(b)** Intra-chain hydrogen bonding between the nth and (n + 4)th residues of α-helix, **(c)** Cysteine cross links between adjacent α-helical coils in α-keratin.

Amino acids not compatible with the α-helix: The following amino acids are not found in the α-helix because they disrupt the conformation.

Proline: The cyclic R group of proline is not geometrically compatible with the right-handed helical conformation. When incorporated in this structure, it causes the chain to bend sharply so that the regular helical conformation is disrupted. The **helix-breaking** property of proline has a significant effect on conformation of body proteins, including enzymes. Alteration of conformation may cause changes in biological activity of the affected protein **(Case 4.3).**

Amino acids with bulky R groups: Amino acids with bulky R groups, namely valine, leucine and isoleucine are too large to be accommodated in large numbers within the compact α-helix.

Amino acids with charged R groups: These amino acids (aspartate, glutamate, arginine and lysine) interfere with formation of α-helix because of electrostatic attraction or repulsion between the charged R groups.

β-Pleated Sheet

In 1951, the same year Pauling proposed α-helix, Pauling and Corney reported existence of a different type of secondary structure, the β sheet.

Characteristics: This structure is found in many fibrous proteins (e.g. silk fibroin) and some globular proteins. Features of β-pleated and α-helical conformation are given in Table 4.8. The β-sheet consists of two or more polypeptide chains in an extended conformation, i.e. each of the chains is almost fully extended into a zigzag (Fig. 4.15). This is in contrast to the α-helix, where the chain is coiled around a central axis. The structure is called pleated because it resembles a series of pleats.

Table 4.8. Comparative features of α-helix and β-pleated sheet structure

	α-Helix	β-Pleated
Hydrogen bonds	Intra-chain, between n and (n + 4) residues. Parallel to helix axis	Inter-chain perpendicular to chain axis
Residues	Small or uncharged residues, such as alanine, leucine, and phenylalanine (most common); proline never found	Alanine, glycine and serine are common
Covalent cross linking	Inter-chain disulphide	None
Chain direction and aggregation	Right-handed helical chains	Antiparallel chains

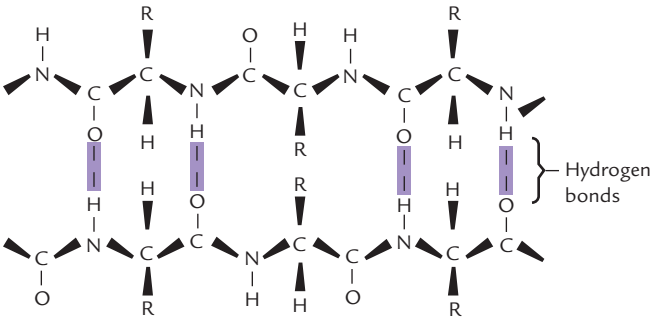


Fig. 4.15. Antiparallel, β-pleated sheet with inter-chain hydrogen linkages.

Stabilizing forces: Like the α-helix, the β-sheet uses the full hydrogen-binding capacity of the polypeptide backbone. In β-sheets, however, hydrogen bonds occur between neighbouring polypeptide chains (inter-chain

hydrogen bonds) rather than one chain, as in α-helix. The hydrogen bonds form between the peptide-bond C = O and N–H groups of polypeptides that lie side by side.

The interacting polypeptides are aligned either parallel or antiparallel. In the parallel arrangement, the N-terminals of the interacting polypeptide are together, whereas in antiparallel arrangement, the N-terminal and the C-terminal ends of the β-strands alternate. Sometimes the chain folds upon itself, as in the case of globular proteins. Intra-chain hydrogen bonds form between the adjacent folds in these proteins.

The secondary structure is a regular repetitive folding pattern of the polypeptide (α-helix and β-pleated sheets are the commonest secondary structure in proteins).

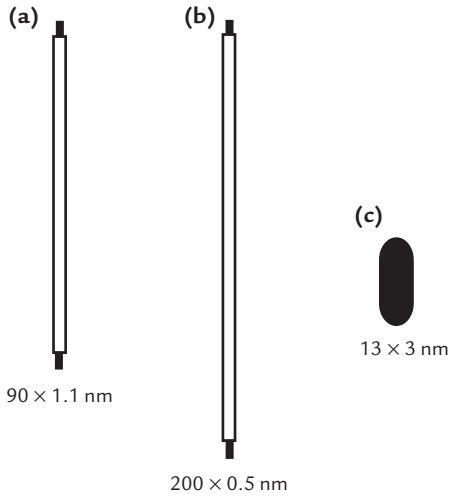


Fig. 4.16. Comparative dimensions of bovine serum albumin if it existed in **(a)** α-Helical form, **(b)** β-Pleated conformation, **(c)** Native globular conformation.

Tertiary Structure

The secondarily ordered polypeptide chain of the globular protein tends to fold onto itself (Fig. 4.12c) in aqueous solutions so that the protein molecule assumes a spherical or globular shape. *The term* ***tertiary structure*** *refers to the overall folding pattern of the polypeptide.* It not only describes folding of the secondary structural elements into a compact globular or spherical shape, but also depicts location of each atom in the protein. This includes **geometric relationship** between distant segments of the primary structure and the **positional relationship** of the side chains with one another. This is in contrast to the secondary structure, which refers to spatial relationship between side chains of the successively adjacent amino acids.

Globular conformation involves tight folding. This is evidenced by the fact that actual dimensions of a globular protein in its native conformation are much smaller than the approximate dimensions if it occurred as an α-helix or in β-pleated conformation (Fig. 4.16). Finally the hydrophobic side chains of the amino acids tend to aggregate towards the interior in globular proteins, whereas the hydrophilic groups strongly prefer to be on the exterior, exposed to water. Thus, globular proteins have a hydrophobic core but a hydrophilic surface, which enables them to interact with aqueous surroundings.

Folding of Proteins and Chaperones

Polypeptide chain of a given amino acid sequence (i.e. primary structure) always folds in the same characteristic fashion, suggesting that the information for protein folding is inherent in the primary structure itself. This is in line with the **Anfinsens law** stating that the secondary, tertiary and quaternary structures are determined by the primary structure. Indeed, protein folding starts, and often makes considerable headway, while translation is still in progress (Chapter 22).

In addition to the primary structure, certain other factors also determine the overall folding pattern. This is evident from the fact that in some cases, following denaturation of protein (i.e. unfolding) by temperature or pH alterations, the native conformation is not regained (by the unfolded protein) even when the original temperature or pH is restored.

It is now known that a *group of specialized proteins called* **"chaperones"** *are required for the native folding of many proteins.* The chaperones are also called **polypeptide chain binding** (PCB) proteins. Though they aid in folding during synthesis, they are not responsible for the stability of the final product. Chaperones bind to hydrophobic patches and stabilize intermediates during folding. They prevent aggregation of incompletely folded proteins by precluding adventitious contacts between exposed hydrophobic regions. Defective interaction of a polypeptide with the corresponding chaperone may cause disease (**Case 7.2**).

The energy requirements of chaperones can be substantial, and more than 120 ATP molecules can be expended during the folding of a single polypeptide chain. Many chaperones are heat shock or stress proteins, but they primarily serve the aforementioned function under physiological conditions.

Domains

Long polypeptides that contain more than 200 residues fold into two or more globular clusters known as domains. The domains are connected by more extended part of the polypeptide chain, which gives these polypeptides a bi- or multi-lobal appearance. Each subunit of

the enzyme *glyceraldehyde 3-phosphate dehydrogenase,* for example, has two domains connected by more extended part of the polypeptide chain.

The globular proteins have hydrophilic surface and hydrophobic core. Large proteins form domains. Folding of domains and their final arrangement in the polypeptide is the another way in which the tertiary structure is defined.

Many domains are structurally independent units, each having characteristics of a small globular protein. Generally, domains are formed from between 40 to 400 amino acid residues (mean size = 200 amino acid residues). Different domains may have different biological functions.

Forces Stabilizing the Tertiary Structure

Tertiary structure mostly represents a state of lowest energy and imparts greatest stability for the polypeptide in question. Five kinds of interactions, cooperate to stabilize it (Fig. 4.17). These are:

1. *Hydrophobic interactions:* These are the major cohesive forces in determining the native protein structure, and arise due to tendency of non-polar substances to minimize their contact with water. In water soluble globular proteins, the non-polar side chains aggregate in the interior hydrophobic core (whereas the hydrophilic side chains occupy surface), and this provides much of the driving force for protein folding.
2. *Hydrogen bonds:* These are formed by sharing of a hydrogen atom between one atom that has a hydrogen atom (donor) and another atom that has a lone pair of electrons (acceptor). The donors are –NH (peptide, imidazole), –OH (serine), and $-NH_2$ (lysine, arginine), and the accepting groups are COO^- (aspartate, glutamate), C = O (peptide) and S–S (disulphide). The hydrogen bonds make only a minor contribution to protein stability.
3. *Ionic interactions:* The association between two ionic side chains of opposite charge (e.g. + ve lysine or arginine; and −ve Glu or Asp) is known as an ion pair or **salt bridge.** Though the electrostatic attraction between the members of an ion pair is strong, these interactions contribute little to the stability of the tertiary structure. This is because of their relatively small number.
4. *Van der Waals interactions and London dispersion forces:* These are very short range interactions that occur between the tightly packed aliphatic side-chains in the interior of the protein. They are weak and develop only when atoms are packed very closely to each other. They serve to bring the atoms closer especially in the hydrophobic core in the interior, squeezing out gaps, if any, and thereby, contribute enormously towards stability of the native protein structure.
5. *Covalent cross linkages:* Cysteine residues, when present on the adjacent loops of the polypeptide(s), form covalent cross links (i.e. **disulphide bonds**). These linkages, however, do not establish the higher-order structure; they only stabilize the structure that has been formed by non-covalent interactions.

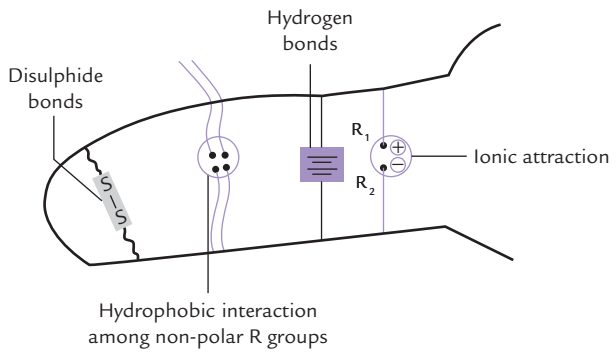


Fig. 4.17. Representation of interactions that stabilize the tertiary structure of globular proteins.

In addition to the peptide bonds between individual amino acid residues, the three-dimensional structure of a protein is maintained by a combination of non-covalent interactions (hydrophobic forces, hydrogen bonds, electrostatic forces, van der Waals forces) and covalent interactions.

Quaternary Structure

The proteins, discussed so far consisted of a single polypeptide chain. But several other proteins consist of two, three or more subunits (i.e. **oligomeric proteins**). The subunit chains may be structurally identical or different. Haemoglobin consists of four polypeptide chains of two different types—α and β. *Number, size and shape of component polypeptide chains and their spatial relationship with one another is called the* ***quaternary structure*** *of the protein* (Fig. 4.12d).

Non-covalent interactions, such as hydrophobic interactions, hydrogen bonds, electrostatic bonds, etc., hold the subunits together in some proteins, but others are stabilized by interchain disulphide bonds.

Different subunits may cooperate with one another (i.e. **subunit interaction**); for example, binding of oxygen to one of the subunits of the haemoglobin is followed by transfer of the ligand-binding information to other subunits, which in turn respond by increasing their affinity for oxygen (Chapter 17).

The formation of secondary, tertiary and quaternary structures of protein follows the ***thermodynamic principles***. In a polypeptide chain, the primary structure determines the higher levels of organization in such a way that the

final form represents the lowest energy stage. Therefore, the quaternary structure is the most stable of all the possible conformations. Simply stated, it represents the state of least energy.

The subunit composition and interactions of the proteins consisting of more that one polypeptide is referred to as quaternary structure.

E. Protein Denaturation

The term denaturation refers to disruption of the higher order structure of the protein. A number of agents are known, which act by breaking the physical bonds in a protein without splitting any peptide linkages. As a result, the well-ordered, neatly folded polypeptide is converted to a messy tangle work, called a **random coil,** and the process is protein denaturation. Note that only the non-covalent interactions that maintain the higher order structure are disrupted, but the covalent bonds (including peptide bonds) are left intact. Thus, **the primary structure is not altered during denaturation,** but denaturation may completely disrupt secondary, tertiary and quaternary structures (e.g. denatured oligomeric protein dissociates into subunits, each with a random coil conformation). Denaturation may be partial in which only certain aspects of the secondary and/or tertiary structures are broken.

Denaturation is always accompanied by a **loss of biologic functions,** e.g. enzymes are inactivated, antibodies fail to act with antigens and toxins lose their disease-producing effects. This is because biological functions of proteins depend on their higher order structures. Denaturation is **generally irreversible** boiled egg does not regain its original form when kept in cold. Under well-controlled laboratory conditions, renaturation is possible for some small proteins, e.g. *ribonuclease* is denatured by 8M urea, but when urea is removed by dialysis, the renaturation occurs and the enzyme activity is restored.

Changes in Physical Properties

A number of physical parameters of the protein change because of partial or complete denaturation:

1. *Solubility:* Water-soluble proteins become insoluble, the boiling of an egg is a familiar example of this phenomenon.
2. *Precipitation:* As the protein becomes less symmetric, it exerts resistance to various types of movements, and in many cases it precipitates.
3. *Optical rotation:* A more negative optical rotation results specific rotation of native proteins is between −20° and −40°, whereas that of denatured proteins is near −100°.
4. The sedimentation rate and diffusion rate fall and the intrinsic viscosity increases.

Agents Causing Denaturation

Proteins can be denatured by many treatments:

1. *Heat*: Most proteins are denatured at temperatures between 50°C and 80°C; at these temperatures the non-covalent interactions, which are individually very weak, are readily broken.
2. *Detergents and organic solvents*: They serve to disrupt hydrophobic interactions. Being non-polar, they insert themselves between the side chains of hydrophobic amino acids.
3. *Strong acids and bases*: They disrupt the intramolecular salt bonds by changing the charge pattern of the protein. In a strong acid, the protein loses its negative charges; and in a strong base, it loses its positive charges; both deprive the protein of the salt bonds.
4. *8M Urea or 6M guanidine hydrochloride*: Both are hydrophilic agents with high hydrogen binding potentials. They can denature proteins by disrupting the hydrogen bonds between the protein and the surrounding water molecules. Normally the water molecules assume an "ordered" position at an aqueous/non-polar interface, which is reduced by these agents, and consequently the hydrophobic interactions within the protein are weakened.
5. *Heavy metal ions*: They have high affinity for carboxylate groups, and for the sulphydryl groups in many proteins.
6. *Trichloroacetic acid (TCA)*: It is of special importance in laboratories because it can rapidly denature and precipitate proteins, which can then be removed by centrifugation.
7. *Physical agents*: X-Rays, ultraviolet rays, high pressure, vigorous shaking and surface tension also cause denaturation.

The non-covalent higher-order structure of proteins can be destroyed by heating, detergents, non-polar organic solvents, heavy metals and strong acids or bases—this is called denaturation and it leads to complete loss of the proteins biological activities.

F. Precipitation Reactions of Proteins

Solubility of proteins depends on pH and salt concentration. Precipitation (of proteins) occurs when their solubility decreases. Unlike denaturation, precipitation is reversible and does not cause permanent loss of the protein's activity.

1. *Salting-in and -out:* Salts at moderate concentration may cause increased protein solubility (**salting-in**) and at high concentration may decrease the solubility resulting in protein precipitation (**salting-out**).
   (i) Salting-in: It is shown in Figure 4.18b. The protein molecules have multiple salt bonds among them, which cause them to aggregate. As a result they are insoluble (Fig. 4.18a). In 5% NaCl, salt ions bind to surface charges of protein molecules, thereby preventing intermolecular salt bonds. This causes increased solubility.

   Thus higher solubility at a moderate concentration (1–10% salt) is because the salt ions shield the protein's multiple ionic charges, thereby weakening the attractive forces between different protein molecules.
   (ii) Salting-out: Excess salt, however, precipitates the proteins because most of the water molecules become tied up in forming the hydration shells around the salt ions. Effectively, so many salt ions are hydrated that there is significantly less bulk water available to dissolve other substances, including proteins. Ammonium chloride, one of the most soluble salts, is the most commonly used reagent for salting out the plasma proteins.
2. *Precipitation by organic solvents:* The water-miscible organic solvents like *ethanol* or *acetone*, because of their high affinity for water, tie up most of the available water molecules. Thus, like salts they cause precipitation of proteins, without denaturing them.
3. *Precipitation by altering pH:* The pH value also affects water solubility—a protein is least soluble at its isoelectric pH when equal numbers of positive and negative charges exist and so the protein molecules have maximal opportunity to form intermolecular salt bonds (Fig. 4.19). At pH values higher or lower than pI, the interactions are mainly repulsive. The pH may be adjusted to approximate the isoelectric point (pI) of the desired protein to cause its precipitation.
4. *Precipitation by heavy metals:* Heavy metals, such as lead, cadmium, and mercury bind to negative groups (carboxylate) or sulphydryl groups in many proteins, thus decreasing their solubility. Because of their affinity for functional groups, these metals tend to precipitate normal proteins of the gastro-intestinal tract, which explains their toxicity.

G. Separation and Purification of Proteins

To study and use a protein it is necessary to separate it from a biologic fluid and purify it. Purification is also necessary for proteins that are used for therapeutic purposes, such as insulin and clotting factor VIII used to treat haemophiliacs. Several techniques are available for this purpose which use fundamental properties of proteins, such as *solubility*, *molecular size*, *molecular charge* and s*elective binding of proteins to specific substances* (i.e. ligands). Generally, a combination of these techniques is employed to separate a given protein from other proteins and molecules. In clinical laboratories, protein separation is routinely carried out for diagnostic purpose. For example, separation of enzyme proteins (by electrophoresis) is useful for the diagnosis of various disorders (**Case 6.5**). Similar techniques are employed for therapeutic and research purposes as well.

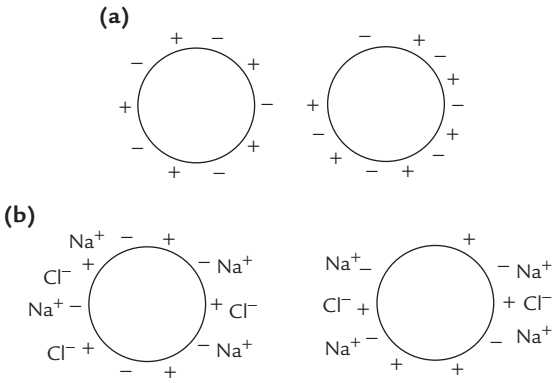


Fig. 4.18. Effects of salt on protein solubility. **(a)** Protein in distilled water. Salt bonds between protein molecules cause the molecule aggregate. The protein becomes insoluble, **(b)** *Salting-in*: Protein in 5% sodium chloride. Salt ions bind to the surface charges of the protein molecules, thereby preventing intermolecular salt bonds.

The aim of protein purification is to isolate one particular protein from all the others in the sample mixture. The purification techniques exploit solubility, molecular size, charge or/and specific binding affinity of the protein of interest.

Separation on the Basis of Protein Solubility

Protein solubility varies with salt concentration of the medium. Addition of salts to the medium results in decreased solubility of some proteins, an effect called *salting out*. The salt concentration at which a protein precipitates differs from one protein to another. For example, globulins precipitate at half saturation of ammonium sulphate (or 22% sodium sulphate) but full saturation of ammonium sulphate (or 28% sodium sulphate) is required for precipitating albumin.

Separation on the Basis of Molecular Size

Dialysis

Proteins can be separated from solute particles by a process termed dialysis which uses the fact that protein molecules

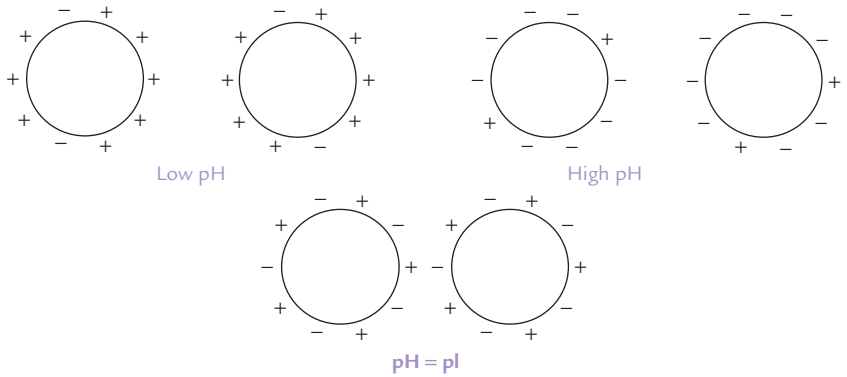


Fig. 4.19. The effect of pH on protein solubility: Formation of intermolecular salt bonds is favoured at the isoelectric point; at pH values greater or less than the pI, the electrostatic attractions between the molecules get weakened.

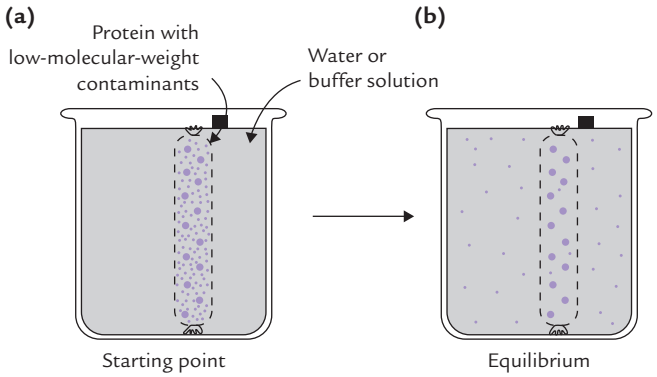


Fig. 4.20. **(a)** The dialysis bag containing the protein mixture and small solute particles is suspended in water, **(b)** Magnified view of a portion of the semi-permeable dialysis bag showing outward movement of solute particles.

are much larger (MW > 10,000) than solute particles. The mixture of protein and solute is put in a cellophane bag, which is made of a semi-permeable membrane. This membrane contains pores which allow the passage of smaller solute particles, but not of protein molecules (Fig. 4.20). When the bag is suspended in stirring buffer solution, the low-molecular weight solute particles gradually pass out.

The removal of low molecular weight waste products from blood of patients with renal failure is based on the same principle; the process called *haemodialysis*. The patient's blood is passed along a semi-permeable membrane through which the waste products (e.g. urea) are removed, but plasma proteins and blood cells are retained. The blood, however, is dialyzed against a solution with physiologic concentration of inorganic ions and nutrients, and not against any buffer.

Proteins can be separated from small molecules by dialysis through a semi-permeable membrane which has pores that allow solute particles to pass through but not proteins.

Gel Filtration (Column Chromatography)

Particles of different sizes can be separated from one another by gel filtration, also called, **size-exclusion** or **molecular-sieve chromatography.** The equipment consists of a column, which is packed with a beaded hydrophilic material such as Sephadex gel. Sephadex is a dextran derivative, obtained from bacterial cell walls. A gel bead contains numerous pores. The pore size is such that the smaller proteins can penetrate the porous gel beads, while the larger proteins bypass the beads altogether (Fig. 4.21). Having access to gels internal volume, the small proteins take a longer and tortuous route through the column. Therefore, they are retarded on the column with respect to the larger protein that cannot enter the porous gel beads. Thus, the latter traverse the column more rapidly and emerge earlier in the column effluent than the smaller ones. The proteins of intermediate size have partial access to gel's internal volume, and therefore, emerge in between, before the larger proteins but after the smaller proteins.

Fractions of a few milliliters (each fraction called an *elute*) are collected from the bottom of the column; the process is called *elution*. Each of these elutes would contain different size particle.

Ultracentrifuge

Proteins of different molecular sizes can be separated by *high-speed centrifugation* into different components. Rate at which a protein moves in a centrifugal field (i.e. sedimentation) depends on its size; larger proteins sediment faster than the smaller ones. This method can separate those proteins which have large differences in their molecular size.

Note: The data from an ultracentrifugal study are expressed in terms of *Svedberg units*. These are derived from the sedimentation constants which reflect the rate of sedimentation in the centrifugal field.

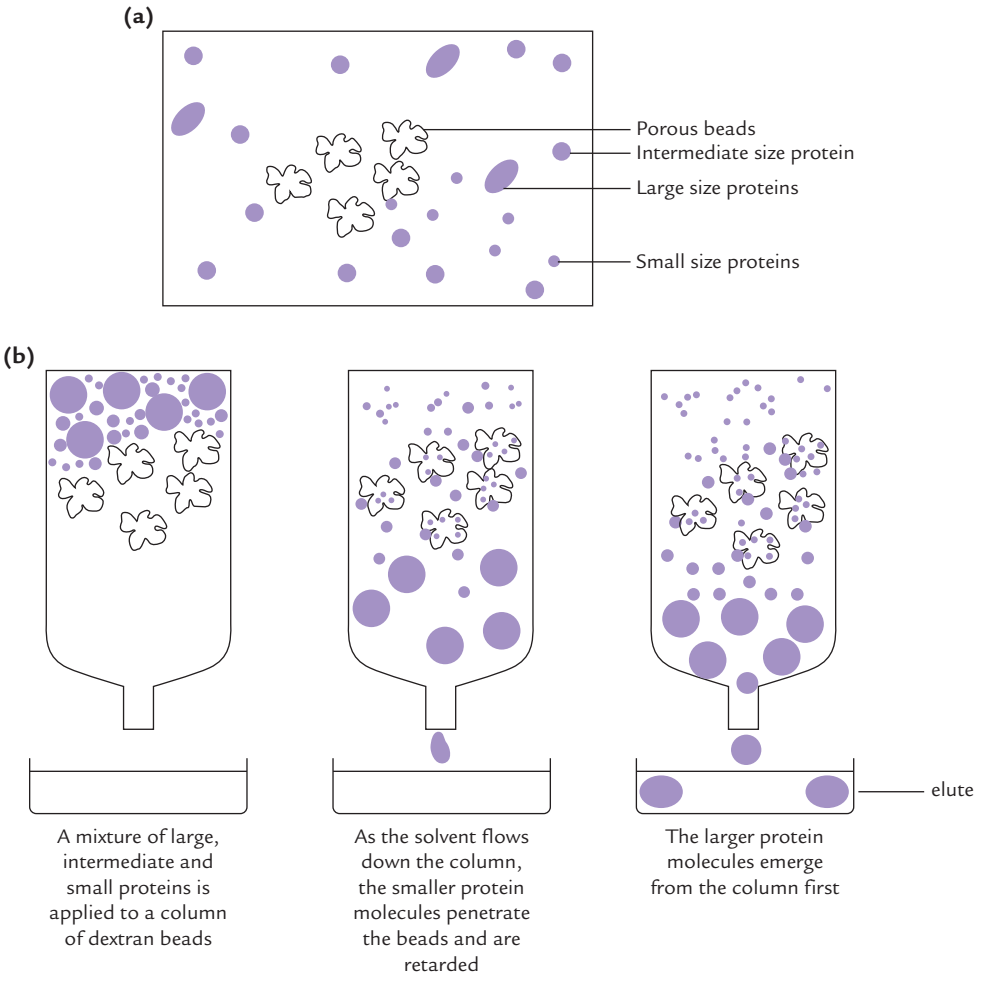


Fig. 4.21. Gel filtration (column chromatography). (**a**) Structure of a bead, (**b**) Process of separation.

Separation on the Basis of Molecular Charge

Differences in the nature and degree of electric charges carried by various proteins at a given pH forms the basis of their separation by several techniques, e.g. ion-exchange chromatography and electrophoresis. Difference in charges is due to difference in acid-base behaviour of these proteins; which in turn is due to different isoelectric pH values of the proteins (Table 4.3).

Ion-exchange Chromatography

Preparative separation of proteins can be achieved on the basis of charge-differences by this technique. The method takes advantage of differential affinity between the following:

- Charged ions or molecules in the sample mixture.
- Charged groups that have been immobilized on ion exchange resins.

The resins are packed in chromatography column (Fig. 4.22). They are cross-linked polymers (e.g. cellulose or agarose) containing charged groups as part of their structure. The charged groups most commonly attached are diethylaminoethyl (DEAE) group and carboxymethyl (CM) group:

DEAE: Resins–CH_2–CH_2–NH $(CH_2CH_3)_2^+$ **(Positive)**

CM: Resins–CH_2COO^- **(Negative)**

DEAE groups are protonated and have a positive charge associated with a negatively charged counterion, such as chloride; this system is called **DEAE-cellulose chromatography** (Fig. 4.23). If a protein solution at a specific pH is applied to this column:

- The negatively charged proteins replace the chloride ions and associate with the positively charged DEAE groups. This is **anion exchange.**
- The positively charged or uncharged proteins have relatively low affinity for the anion exchanger than the negatively charged proteins. They, therefore, flow faster and emerge from the column earlier (Fig. 4.22a).
- In this way, the negatively charged proteins get separated from the uncharged/or the +ve proteins (Fig. 4.22b). They (−ve proteins) can be subsequently eluted

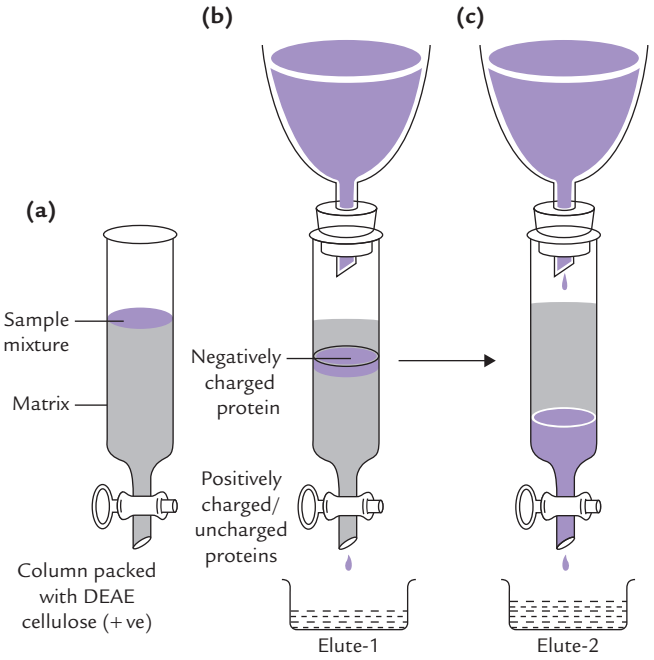


Fig. 4.22. Ion exchange chromatography. **(a)** Sample mixture consisting of positively charged, negatively charged and neutral proteins in a small volume of buffer applied at the top of the column, **(b)** The positively charged and neutral proteins pass through the column and are collected in elute-1. Flow of the negatively charged proteins is retarded, **(c)** The negatively charged proteins, earlier attached to the positive matrix, are made to pass through the column by applying a high-salt/low pH elution buffer.

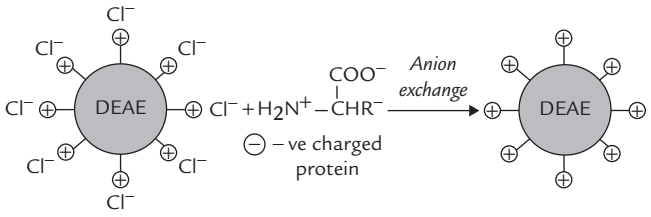


Fig. 4.23. DEAE cellulose chromatography (anion exchange). The protonated groups (+ve) on DEAE associated with chloride counterions (−ve). The negatively charged proteins COO⁻ / $H_3N^+-CHR^-$ exchange with chloride and get bound to resin.

(washed through the column) by changing elution conditions, such as applying a buffer that has a higher salt concentration, or at a pH that reduces the affinity with which the resin binds the protein (Fig. 4.22c).

CM groups with negative charge are associated with counterion Na^+. (Fig. 4.24). In this case, positively charged proteins exchange with Na^+ and get bound to the resin (i.e. **cation exchange**), whereas the negative ones are eluted faster.

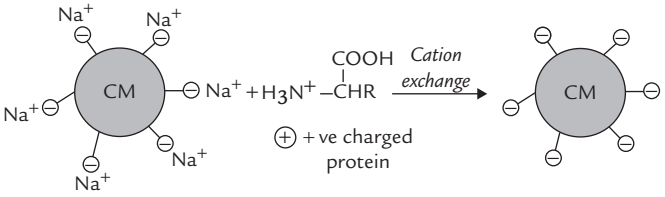


Fig. 4.24. Cation exchange: Positively charged proteins COO / $H_3N^+-CHR^+$ attach with carboxy methyl groups (negative) and get bound to resin.

High Performance Liquid Chromatography (HPLC)

HPLC resembles the conventional chromatographic techniques where the solution of proteins is passed through special resins that have attached side groups. These side groups can react with the protein sample ionically or hydrophobically. However, there is one important differences HPLC is carried out under *high pressure* **(5000–10,000 pounds per square inch)** in the chromatographic matrix. The matrix consists of specially fabricated 3–300 μm-diameter glass or plastic beads, which are coated with a uniform layer of chromatographic material. This greatly improves the speed, resolution, and reproducibility of the separation.

Electrophoresis

This process uses an **electric field to drive the movement of charged particles.** It is used for the separation of a wide variety of polyelectrolytes (carry positive and negative charges), including amino acids, polypeptides, proteins (and DNA).

At pH values greater than the isoelectric point (pI), the protein carries an excess of negative charge and moves to the anode; at pH values less than pI, it carries positive charge and moves towards the cathode; and at isoelectric point the net charge is zero and it does not move. If for example, an alkaline pH is used (greater than pI of most proteins) the serum proteins are negatively charged and move towards the anode, but are separated according to their net charge (Fig. 4.25). A porous support such as **paper, cellulose acetate** foil, **starch gel**, or other carrier material is commonly used to minimize diffusion and convection. Electrophoresis is the most common method of protein separation in the clinical laboratory, used for separation of plasma proteins and isoenzymes and for the detection of structurally abnormal proteins. **Case 5.1** illustrates diagnostic utility of electrophoresis in a bone disorder.

Polyacrylamide gel electrophoresis (PAGE): Electrophoresis can be performed in a cross-linked polyacrylamide gel with characteristic pore size (Fig. 4.26). Small molecules can move in the gel, but the larger ones "*get stuck*" in the gel and so their movement is retarded. The molecular

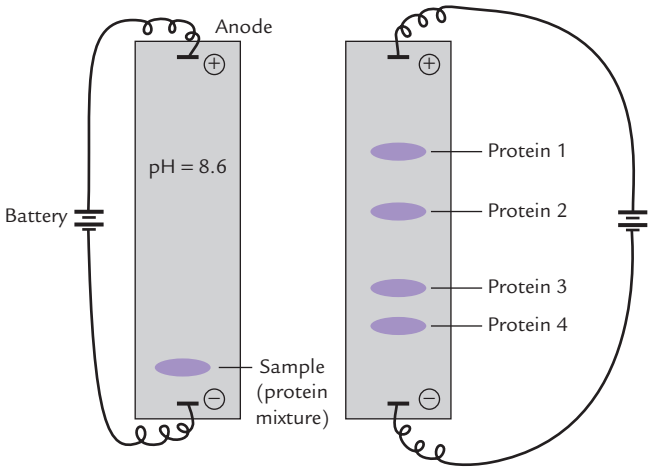


Fig. 4.25. Protein separation by electrophoresis on a wet cellulose acetate foil. At pH = 8.6, all proteins are negatively charged and move towards anode. They are separated into various bands according to their net charge.

separation is, therefore, **based on sieving effects as well as electrophoretic mobility**.

SDS-PAGE: Under pH conditions at which all proteins move to the same pole, they are better separated according their molecular mass rather than their charge. To accomplish this, the above method (PAGE) is used in the presence a strong detergent (e.g. **sodium dodecyl sulphate; SDS**). It causes denaturation by interfering with the hydrophobic interactions that normally stabilize proteins; the denatured proteins assume a rod-like shape. The large negative charges that the SDS imparts mask the protein's intrinsic charge. The net result is that SDS-treated proteins have similar shapes and charge-to-mass ratio. SDS-PAGE, therefore, separates proteins by **gel filtration effect**, i.e. according to molecular mass. The method is performed in presence of a reducing agent such as mercaptoethanol, which cleaves the disulphide bonds.

Electrophoresis separates molecules by charge and size; in SDS-PAGE however, SDS masks the intrinsic charges of proteins, and therefore, separation is primarily by size.

Isoelectric focusing: It is a variant of electrophoretic technique, that has extremely high resolution. It is carried out on a gel that contains a gradient of pH. The **pH gradient is created** across the supporting medium and electrophoresis of the protein mixture carried out through this gradient. A charged protein migrates through the gradient till it reaches a pH region of the gradient equal to its pI value. At this point it stops moving (net charge on the molecule being zero) and can be visualized. An extremely sensitive technique, isoelectric focusing can separate proteins that differ by as little as 0.0025 pI values.

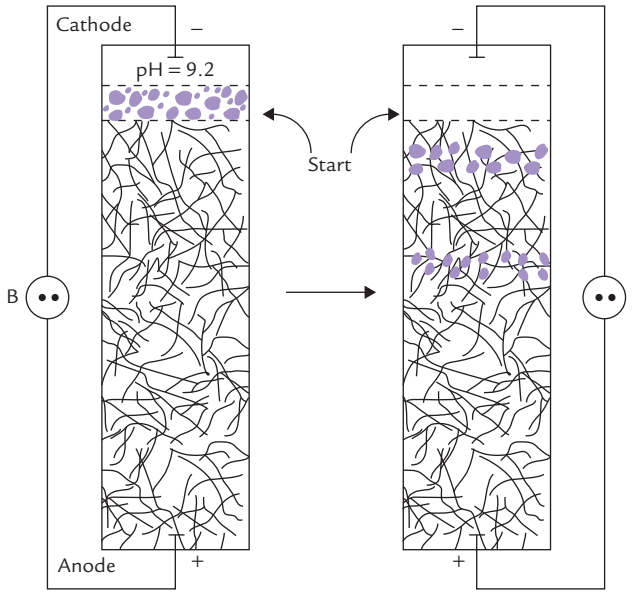


Fig. 4.26. Protein separation in a cross linked polyacrylamide gel. The pH is sufficiently high so that all proteins are negatively charged and move towards anode. At a suitable pH, this method separates more on the basis of molecular mass rather than charge.

Separation on the Basis of Affinity Binding

Affinity Chromatography

This is the only method of protein purification that exploits a property related to the function of the protein, namely the **ligand-binding ability**. This is a definite advantage over exploiting small differences in the physicochemical properties between proteins in the other chromatographic methods.

The column is packed with a chromatographic material that carries an immobilized ligand (Fig. 4.27).

The mixture of proteins is poured over the column. As it percolates down, the protein molecules with an appropriate ligand-binding site gets bound to the immobilized ligand, whereas the other proteins are washed through the column with the buffer. The bound protein can then be recovered in highly purified form by changing the elution conditions. (Generally, it is washed with solution having a high concentration of soluble-ligands.)

Precipitation by Antibodies

For separating a given protein from a protein mixture, antibodies are prepared against it. When these antibodies are added to the protein mixture, they form **antigen-antibodies complexes** with protein of interest. These complexes are large in size, and so can be separated from the rest of mixture by *centrifugation*. Since it is necessary to create complexes of sufficiently large size, the anti-globulin (anti IgG) is further added, so that the large **triple complex** (of antigen-antibody-anti IgG) is formed.

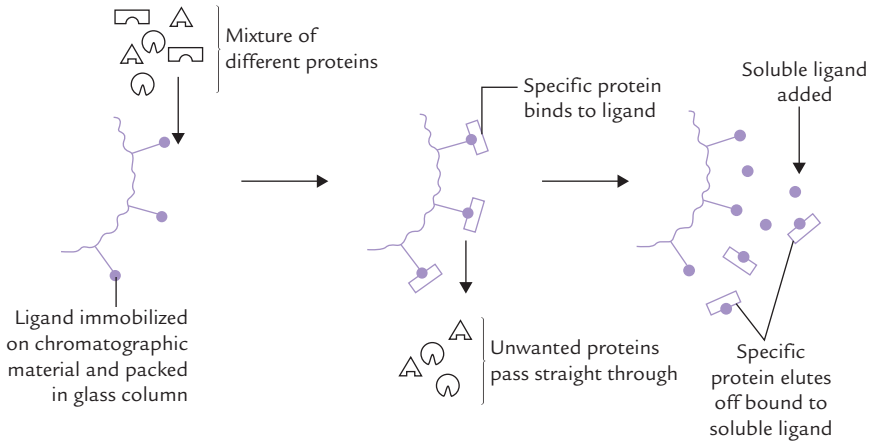


Fig. 4.27. Protein seperation by affinity chromatography.

Separation on the Basis of Hydrophobicity

In this technique, called **hydrophobic interaction chromatography,** separation of protein mixtures is accomplished on the basis of protein hydrophobicity. The proteins and the matrix material packed in the column interact hydrophobically. The matrix material is lightly substituted with acetyl or phenyl groups and the nonpolar groups on the surface of proteins interact with them by the hydrophobic interactions. Subsequently, the bound proteins are eluted by decreasing salt concentration, increasing concentration of detergent (which disrupts hydrophobic interactions), or by causing changes in pH.

Purification and characterization of proteins is essential for elucidating their structure and functions. By taking advantage of differences in their size, solubility, charge and binding capacity, proteins can be purified to homogeneity.

H. Determination of Amino Acid Sequence of Peptides and Proteins*

The primary structure of proteins can be determined by **chemical methods.**

- The preliminary steps (steps 1 and 2) provide clue to protein structure (e.g. size and amino acid composition of the protein), and
- The subsequent steps determine the exact amino acid sequence.

Step 1: The size of the protein

The molecular size of the protein is estimated usually by SDS-PAGE (sodium dodecyl sulphate-polyacrylamide gel electrophoresis), as discussed earlier. Since all proteins are covered with negative charges of SDS, they tend to be linear in shape and their movement in electric field solely depends on their size.

Step 2: Amino acid composition of the protein

It is determined by complete hydrolysis of the polypeptide, followed by the analysis of the liberated amino acids. The hydrolysis is accomplished by **acid hydrolysis** of the peptide linkages in 6N–HCl for 24–72 hours at 100°–110°C. This treatment, however, destroys tryptophan. **Alkaline hydrolysis** is, therefore, used to liberate this amino acid. The amino acids so liberated are then separated by chromatographic methods and determined quantitatively with suitable reagents. Modern amino acid analyzers can completely analyze a protein digest containing as little as 1 pmol of each amino acid in less than an hour.

Step 2 reveals the protein's amino acid composition, but not the amino acid sequence. It is, however, useful because the ratio of the amino acids can be compared with that from known proteins and this may give a clue to its structure.

Step 3: Determination of N and C terminal amino acids

The N-terminal amino acid is covalently bound to **dansyl chloride**. The tagged protein is then hydrolyzed to its constituent amino acids by acid hydrolysis. Because only the N-terminal amino acid is dansylated, it is identified by separating it out by column chromatography and comparing against dansylated standards.

The C-terminal amino acid is determined by treating intact polypeptide with **hydrazine**, which reacts hydrolytically at the carbonyl grouping (C–O) at each peptide bond (resulting in acyl derivatives of each residue). However, it cannot react with the C-terminal, which has carboxyl group (rather than a peptide linkage). The amino acids are then analyzed by column chromatography and

*This topic relates to higher level learning and it is meant only for postgraduates. Undergraduates or MBBS students may refer to it for higher learning.

compared with standards. The amino acid without hydrazine is the C-terminal amino acid.

Step 4: Sequence analysis

It is performed by **Edman degradation method** on automatic machines called, cyclic sequenators. This method sequentially removes one residue at a time from the amino end of a peptide. **Phenyliosthiocyanate** reacts with the N-terminal amino group of the peptide to form a phenylthiocarbamyl derivative in alkaline solution. It is released by gentle acid hydrolysis and identified by HPLC. The second amino acid now becomes the N-terminal amino acid and the cycle can be repeated by returning to alkaline conditions.

Only about 20–25 amino acids can be sequenced in this way. The large polypeptides, however, must be fragmented first. A number of **fragmenting agents** break specific peptide linkages, among which are digestive *endopeptidases*, trypsin and chymotrypsin.

- **Trypsin** cleaves only those peptide bonds in which carboxyl group is contributed by basic amino acids, lysine or arginine.
- **Chymotrypsin** cleaves only the bonds on the carboxyl side of the aromatic amino acid residues.
- **Cyanogen bromide** breaks those peptide bond in which methionine contributes the carboxyl groups.
- The short fragments so generated are more manageable in sequencing studies and the amino acid sequences in each fragment are then determined, as discussed above.

Step 5: Reconstructing the protein's sequence

After individual peptide fragments have been sequenced, their order in the original polypeptide must be elucidated. For this a second round of protein cleavage with a different *endopeptidase* is carried out. This generates an overlapping set of peptide fragments. These fragments are then separated and sequenced. The sequences of the overlapping peptides are then used, much as one would fit parts of a jigsaw puzzle, to obtain the whole protein structure.

Sequencing by Molecular Biology Techniques

This is an entirely different and relatively simple approach based on the fact that genes encode amino acid sequences of polypeptides; the **base sequence of the gene predicts the amino acid sequence.**

Step 1

The protein is partially purified and the first eight (or ten) amino acids at the N-terminal end are sequenced using an automatic sequenator.

Step 2

These eight amino acids are compared with an international protein/DNA sequence library available on the Internet, e.g. GenBank. If lucky, a match occurs in terms of molecular weight, which means that the protein (or a closely similar protein) has already been isolated and sequenced. Further sequencing may not be necessary in this case.

Step 3

If no match is found, then a cDNA molecule that codes for the N-terminal amino acid sequence is made. This is called cDNA primer.

Step 4

Using the cDNA as a probe, the RNA that codes for the protein is extracted from the cells. The two readily hybridize (attach together) because their sequences are complementary.

Step 5

The DNA complementary to the rest of the RNA sequence is then synthesized by polymerase chain reaction (Chapter 25). This results in the synthesis of large amounts of DNA that codes for the protein.

Step 6

The DNA is then sequenced using techniques described in Chapter 21 and, using the universal genetic code, and the protein sequence is derived from the DNA sequence.

This process sounds complicated but offers advantage that very little of the protein is needed and the protein does not have to be fully purified. Full purification is especially difficult in case of insoluble proteins (e.g. membrane proteins). Moreover, sequencing DNA is a relatively simple procedure.

I. Determination of Higher Order Structure of Proteins[†]

X-Ray Crystallography

This method is based on the scattering of the beam of x-rays passed through a protein crystal by high mass proton cores. The scattered x-rays form a diffraction pattern which is captured on film (or collected by an electronic device) in the form of an array of dots, from which bond lengths, folding and coiling and the overall three dimensional structure in proteins may be deduced. Because this pattern indicates the relative positions of atoms in the molecule, a careful study and analysis of it permits modelling of the 3-dimensional structure of the crystalline protein. Pauling and Corey were pioneers in interpreting such protein diffraction pattern to determine the structures of α-helix and β-pleated sheet.

[†] This topic relates to higher level learning and it is meant only for postgraduates. Undergraduates or MBBS students may refer to it for higher learning.

Ultraviolet Light Spectroscopy

Proteins absorb ultraviolet radiation with two absorption maxima: one at 190 nm is caused by peptide bonds, and the other at 280 nm is caused by the aromatic side chains of tyrosine and tryptophan (aromatic character of purine and pyrimidine bases also causes absorbance peak at 260 nm for nucleic acids).

Nuclear Magnetic Resonance (NMR)

Since mid 1980s advances in NMR spectroscopy have permitted the determination of the three-dimensional structure of proteins (and nucleic acids) in aqueous solution. The sample of protein is placed in a magnetic field so that the spin of protons is aligned. Radiofrequency pulses are applied, which cause excitation of protons. The excited protons then emit signals, whose frequency depends on the molecular environment of these protons. Interpretation of these signals helps to construct the three-dimensional protein structure. The technique is useful for the proteins that fail to crystallize. However, utility of this technique is limited to a size of less than 30 kD.

Computer Based Modelling

A number of proteins have been modelled till now, and based on them computer programs have been developed to predict the most likely tertiary structure from the primary amino acid sequence.

Deciphering the primary and three-dimensional structures of a protein leads to an understanding of structure-function relationship in proteins. Both chemical and physical methods are available for determining the amino acid sequence and higher-order structure for proteins.

IV. Chemical Synthesis of Proteins

Using chemical means, protein chemists have synthesized several small proteins and peptides in laboratory. The basic mechanism involves fixing of the carboxyl group of the last amino acid on the resin, followed by addition of other amino acids sequentially. Steps must be taken to assure that carboxyl and amino groups that are not participating in peptide bond formation must be blocked, and groups that are to participate in bond formation must be activated (for example, by PCl_5). Though synthesis of several polypeptides with more than 100 amino acids has been made possible, production of only small oligopeptides is cost effective. This limitation is deplorable because several naturally occurring proteins, such as hormones, and blood-clotting factor are useful therapeutic agents, and it is easier and cheaper to synthesize them rather than isolate and purify them from biological sources.

Exercises

Essay type questions

1. What do you mean by the term conformation as applied to a protein? Describe the various orders of structure of proteins.
2. Define secondary and tertiary structures. Mention the bonds that are responsible for maintaining these structures.
3. Give two examples each of fibrous and globular proteins. Describe the structural and functional relationship with respect to any one protein.
4. Describe the separation techniques that exploit the following molecular properties: charge, polarity, size and specificity.
5. Why are proteins called amphoteric molecules? Explain why histidine, but not alanine, can act as buffer at physiologic pH.

Write short notes on

1. Quaternary structure
2. α-Helix
3. β-Pleated sheets
4. Essential amino acids
5. Chaperones
6. Isoelectric pH
7. Gel chromatography
8. Edman degradation
9. Titration curve of histidine
10. Physiologically active peptides
11. Electrophoresis

CLINICAL CASES

CASE 4.1 A 38-year-old diabetic patient develops allergy to insulin

A 38-year-old man, having insulin dependent diabetes mellitus (IDDM) has been on insulin therapy for the last 23 years. A satisfactory control of blood glucose level had been achieved with porcine insulin. For about last 4–6 weeks, he had been feeling itching and local stinging at the site of injection. Examination showed swelling and a generalized urticaria. Analysis of random blood sample showed hyperglycaemia (blood glucose is 288 mg/dl), indicating poor control of diabetes. This was surprising for the patient as well as for the physician, because a fairly good glycaemic control was maintained with the recommended dose of porcine insulin till his last visit to the hospital. Presently, the patient was taking 40 units of the porcine insulin before breakfast every day, followed by 10 units in the evening.

Q.1. What is the rationale behind use of insulin from a different species (i.e. porcine insulin) in humans?
Q.2. What is the underlying cause of the above symptoms?
Q.3. Identify the cause of loss of glycaemic control in this patient.
Q.4. How is human insulin synthesized?

CASE 4.2 A chronic smoker with respiratory distress

α_1-Antitrypsin (ATN) is an inhibitor of certain enzymes. It contains a methionine residue at 358th position, which plays a vital role for the above stated function of ATN. In tobacco smokers, the primary structure of ATN is altered because of certain changes in this residue. As a result, its inhibitory action on the proteolytic enzymes is impaired, resulting in increased proteolytic activity. The ultimate outcome of this is an extensive alveolar damage; the condition is called **emphysema.**

Q.1. Name the proteolytic enzymes that are inhibited by ATN. Mention biochemical basis of the inhibition.
Q.2. State the cause of alveolar destruction in smokers.

CASE 4.3 A 10-year-old mentally retarded child

A 10-year-old child, who was incontinent, unable to feed himself, and spoke incoherently was referred for investigations. He presented a picture of severe mental retardation with an I.Q. of 65. Detailed biochemical investigations with the plasma and the urine samples, and with the cell free extract* prepared from the hepatocytes (obtained by liver biopsy), led to the conclusion that the child had an impaired ability to convert phenylalanine to tyrosine. This is due to a defect in the enzyme, phenylalanine hydroxylase.

The gene for this enzyme was identified and sequenced. Sequences of the complementary mRNA and the corresponding amino acid segment were determined, both for the patient as well as for a normal subject.

It may be observed that a point mutation caused a proline residue to be present at a certain position in polypeptide chain of the patient; the corresponding position was occupied by leucine in the normal subject.

	Normal subject	**Patient**
DNA	3′-CGAGTCTCA-5′	3′-CGAGGCTCA-5′
mRNA	5′-GCUCUGAGU-3′	5′-GCUCCGAGU-3′
Amino acids	Ala-Leu-Ser	Ala-Pro-Ser

Q.1. How might the mutation affect the enzymes?

***Cell free extract contains all the enzymes and coenzymes present intracellularly, and hence is a reliable indicator of the metabolic milieu within the cell.**

CHAPTER

PROTEINS WITH BIOLOGICAL ACTIVITY

5

Native conformation of a protein is adapted to serve specific biological functions. For example, β-conformation provides soft and flexible silk fibres, and the α-helical conformation provides tough, insoluble protective structures. In this chapter, some biologically active proteins are described with a special emphasis on structure-function relationships.

After going through this chapter, the student should be able to understand:

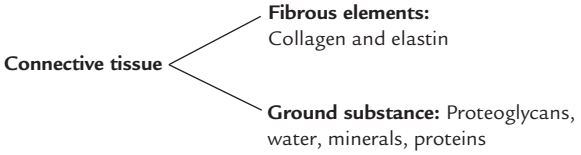
- Distribution, primary and secondary structures, and properties of fibrous proteins , e.g. collagen and elastin: synthesis and degradation of collagen, and molecular basis of various inherited or deficiency diseases that affect collagen structure and integrity, hybrid nature, structure and function; synthesis and degradation of glycoproteins and the role of dolichol.
- General structure and properties of immunoglobulin molecules; function and distinctive characteristics of different classes of antibodies; function and diagnostic significance of various plasma protein fractions; various disorders leading to over- or underproduction of plasma proteins and immunoglobulins.

I. Collagen

A. Overview

Multicellular organisms contain collagen, which is the major molecular component of connective tissue. Connective tissue is the principal structural and supportive element of the body—present in tendons, ligaments, cartilages, and organic matrix of bones. It also forms a structural layer under the skin, and envelops of blood vessels.

Connective tissue comprises of amorphous ***ground substance***, in which ***fibrous elements*** like collagen and elastin are present. The ground substance comprises mainly of proteoglycans, water, minerals and proteins.



Thus, major molecular components of connective tissue are: ***collagen, elastin*** and ***proteoglycans.*** The word collagen is derived from Greek language which means to produce glue. ***Collagen*** *is fibrous protein that possesses high tensile strength and cannot be stretched.* Collagen is most abundant of all human proteins, constituting about one-third of total protein mass of the body. It is present in all tissues and organs where it serves as the major structural component.

In addition to collagen, the other fibrous component of connective tissue is **elastin**. However, unlike collagen, elastin can be readily stretched.

Proteoglycans are hybrid molecules of proteins and polysaccharides. They form the ground substance in which various fibrous elements of connective tissue are embedded. The ground substance acts as a cushion or a lubricant.

All the three molecular components of connective tissue are synthesized in specialized cells. These are *fibroblasts* in connective tissue, *osteoblasts* in bones, and *chondroblasts* in embryonic cartilage.

The extracellular matrix consists of an amorphous ground substance of protein, proteoglycan, water and minerals into which are embedded the protein-fibres of collagen and elastin.

B. Basic Structure

Collagen is not a single protein but a large protein family with close to 20 members. Basic structural unit of all collagens is a trimer of polypeptides, called **tropocollagen**, that forms a characteristic **triple helix**. But the supermolecular structures of various collagen types differ, and they have different distributions and functions.

Tropocollagen is a rod-shaped molecule, about 300 nm long and 1.5 nm thick. *Several tropocollagen molecules are arranged head to tail in parallel bundles to form **fibrils** (Fig. 5.1a), commonly known as **microfibrils**.* The heads of tropocollagen molecules are staggered along the length of the fibrils meaning that adjacent molecules are displaced approximately one-fourth of their length (67 nm). This quarter-staggered array accounts for the characteristic 67 nm spacing of the cross-striations in most collagens. The fibrils further associate to form larger *fibres*, which have high tensile strength: a fibres of 1 mm diameter can withstand a load of 10–40 kg before it breaks. Hence, *the* **collagen infrastructure**—*a regular array of parallel, staggered tropocollagen molecules—provides a final product with a tremendous amount of strength.*

Triple Helix

The structure of collagens is "**triple-helical**". This is because *tropocollagen consists of three helical polypeptide chains*, each having approximately 1050 amino acids (Fig. 5.1b). These chains, known as the α-**chains** (nothing to do with the α-helix), are tightly wound around one another. Collagens also contain non-helical regions which are relatively small. The principal features of triple helix are:

1. Individual α-chains form a *tight, left-handed triple helix* with 3.6 amino acids per turn and a rise per amino acid of 0.15 nm. This structure is favoured by proline residues, which are present in large amounts in these chains (proline introduces a sharp bend in a polypeptide chain).
2. The three intertwined α-chains form a *superhelix with a right-handed sense*. Thus, the secondary structure of tropocollagen is helical on two levels:
   (a) each chain has left-handed sense and
   (b) the superhelix has right-handed sense.

The opposite direction of twists of the individual chains and the triple helix means that tropocollagen is very resistant to unwinding by tension.

3. In order to form a triple-helix a polypeptide *chain must contain glycine* as every third residue in the sequence. This is because side chain of only glycine is small enough to permit close contact between main chains.
4. The peptide bonds are internal, and therefore, collagens are very *resistant to digestion* by proteases. The triple helix is stabilized by **hydrogen bonds** between the peptide bonds of different chains.

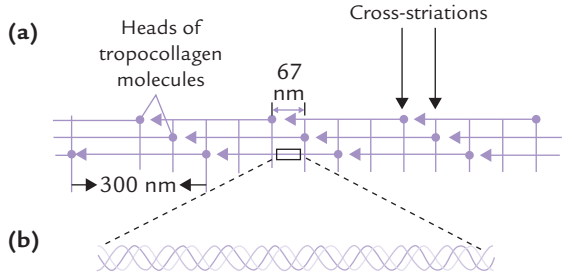


Fig. 5.1. Structure of collagen. **(a)** A fibril showing arrangement of tropocollagen molecules—in a regular, staggered parallel array, **(b)** Portion of tropocollagen molecule magnified to show the triple-helical structure of collagen—consisting of three helical polypeptide chains. Each chain is twisted into a left-handed helix and these chains wrap around each other to form a right handed triple helix.

Collagen, the most abundant vertebrate protein, is basically a trimer of three intertwined polypeptides, called tropocollagen—the so called triple helix.

Types of Collagens

The α-chains associate in different combinations in the trimer to form various types of collagen molecules, of which about **20 types** are known. The α-chains are identical in some collagen types and non-identical in others. As shown in Table 5.1, different types of α-chains are numbered according to the collagen types in which they occur. Thus, type I collagen is a heterotrimer, containing an α_1(I) chain that is present in two copies, and an α_2(I) chain in one copy, yielding the structural formula $[\alpha_1(I)]_2$ α_2(I). The type II and type III are homotrimers composed of three identical α_1(II) and α_1(III) chains respectively.

In most of these collagen types (notably types I, II, III, V, XI and XII), tropocollagen molecules associate in an ordered manner to form **fibrils**. The fibril formation depends on interactions between amino acid side chains in neighbouring molecules. Other types, such as type IV—a major constituent of basement membranes, do not form such fibrils, but are involved in the formation of an extensively branched mesh.

Table 5.1. Composition and distribution of different types of collagen

Type	Composition	Structural features	Distribution
I	$[\alpha_1\ (I)]_2\ \alpha_2\ (I)$	67 nm banded fibrils	Most abundant type,* in most connective tissues—skin, tendon, bone
II	$[\alpha_1\ (II)]_3$	67 nm banded fibrils	Cartilage, intervertebral disc, vitreous humor
III	$[\alpha_1\ (III)]_3$	67 nm banded fibrils	Fetal skin, cardiovascular system
IV	$[\alpha_1\ (IV)]_2\ \alpha_2\ (IV)$	Forms a branched network	Basement membranes
XIX	$[\alpha_1\ (XIX)]_3$	Non-fibrillar	On surface of collagen fibrils in rhabdomyosarcoma
X	$[\alpha_1\ (X)]_3$	Fibril formation	Calcifying cartilage

* Type I accounts for about 90% of all collagens in humans.

$-^{31}$Gly—Pro—Lys—Gly—Phe—Thr—Gly—Pro
—Hyp—Gly—Glu—Hyp—Gly—Glu—Hyp—Gly
—Ala—Ser—Gly—Pro—Met—Gly—Pro—Arg
—Gly—Pro—Hyp—Gly—Pro—Hyp^{60}

Fig. 5.2. Amino acid sequence of part of the α-chain of collagen.

Collagens consist of three polypeptide chains, the identity and distribution of which vary between different collagen types.

Amino Acid Composition and Sequence

Collagen has a unique primary structure, represented as **-Gly-X-Y**, which implies that **glycine** is present at every third position in the polypeptide chain. Thus, this amino acid constitutes about 33% of the total amino acid residues (Fig. 5.2). X and Y can be any amino acids, but certain ones predominate in these positions:

- **X** is often proline, 3-hydroxyproline, glutamate, histidine, leucine or phenylalanine, and
- **Y** is 4-hydroxyproline, threonine, lysine or arginine.

Collagens can thus be *identified by their amino acid sequence, and by high proportion of certain amino acids:* 33% **glycine** and 10% **proline**. Most remarkably, collagen contains **4-hydroxyproline** (10%), **3-hydroxyproline** (<0.5%), and **5-hydroxylysine** (1%), which do not occur in most other proteins. These hydroxylated amino acids are *not represented in the genetic code,* and therefore they have to be synthesized post-translationally from prolyl and lysyl residues in the polypeptides. Each of these amino acids plays a specific role in folding of the polypeptide chain, discussed as follows:

1. *Proline* residues introduce sharp bends in the α-chains (because R group of proline is part of a cyclic ring), which helps tight wrapping of these chains around one another. Such tight wrapping enhances strength of the triple helix.
2. *Glycine* residues play an indirect role in permitting extremely tight interwinding of the α-chains. The winding is so compact that there is little space available towards the interior, only the smallest amino acid, i.e. glycine, can be accommodated at that place. Importance of glycine in collagen is illustrated by the fact that substitution of this amino acid by a bulkier one results in marked reduction in the strength of collagen (**Case 5.1**).
3. *Hydroxyproline and Hydroxylysine* are modified amino acids in collagen, each containing an additional hydroxyl group (Fig. 5.3). Each hydroxyl group in turn can participate in the formation of an **additional hydrogen bond**. Formation of a large number of hydrogen bonds is thus possible, which gives enormous collective strength.

Formation of these special amino acids occurs from the corresponding primary amino acids, proline and lysine, by hydroxylation reactions. These reactions are catalyzed by the enzymes *proline hydroxylase* and *lysine hydroxylase*, respectively.

C. Biosynthesis of Collagen

The collagens are synthesized and secreted by specialized cells called **fibroblasts**. Like other secreted proteins, collagen polypeptides are synthesized by ribosomes on the rough endoplasmic reticulum (RER; Fig. 5.4). The polypeptide chain then passes through the RER and Golgi apparatus before being secreted. Along the way it is **post-translationally modified** by glycosylation, hydroxylation proteolytic cleavage, and cross-linking. Many of these modifications are unique to collagen.

Polypeptide Synthesis

Collagen genes are some 10 times longer than the mature mRNAs that they encode, and contain more than 50 short exons separated by introns of different sizes.

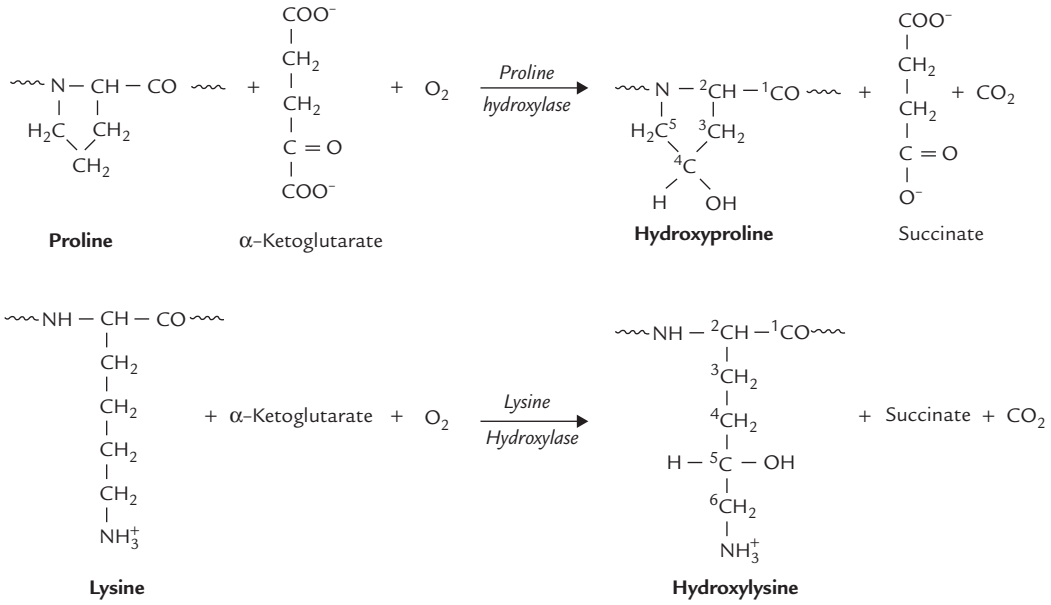


Fig. 5.3. Hydroxylated amino acids in collagen, synthesized post-translationally from proline and lysine residues in the polypeptide.

The nascent chain initially synthesized is called **pre-pro-α-chain**, or simply pre-pro-collagen. It contains a sequence of 15–30 hydrophobic amino acids at its N-terminal end, called **signal sequence**. This sequence directs entry of the pre-pro-α-chain into cisternae of endoplasmic reticulum, where it is converted to pro-α-chain by removal of this signal sequence.

Pre-pro-α-chain ⟶ Pro-α-chain + Signal sequence

Collagen is synthesized from a larger precursor, which is processed in the organelles of the secretory pathway and extracellularly.

Post-translational Modifications

Biosynthesis of collagen provides an excellent example of post-translational modification of a protein. As noted earlier, it involves a number of reactions: **hydroxylation** and **glycosylation** initially (followed by chain assembly, proteolysis and cross-linking, later on).

Hydroxylation

Selected proline and lysine residues undergo hydroxylation to form the corresponding hydroxylated amino acids: hydroxyproline and hydroxylysine. Action of *proline hydroxylase* and *lysine hydroxylase*, respectively catalyze these reactions. These enzymes have an Fe^{2+} ion at their active site and require **ascorbic acid** (vitamin C) for activity. The ascorbic acid acts as an anti-oxidant, keeping the Fe^{2+} ion in its reduced state. Both the enzymes are

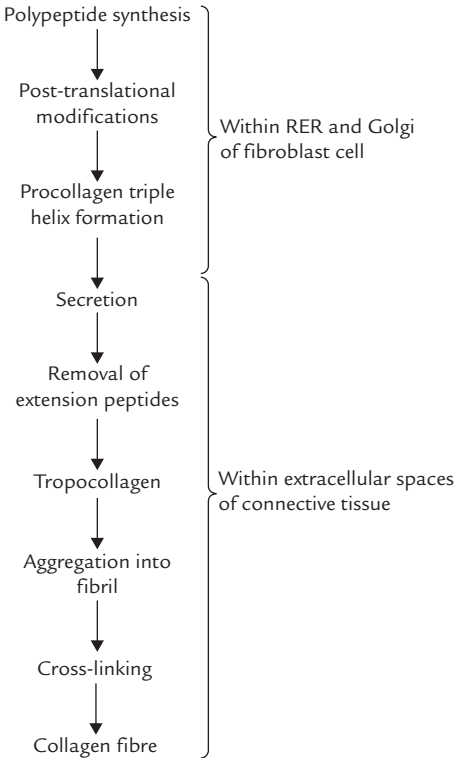


Fig. 5.4. Overview of the biosynthesis of collagen.

dioxygenases using molecular oxygen. α-Ketoglutarate, the TCA cycle intermediate, is a coreductant, which is oxidized to succinate during the reaction (Fig. 5.3). Moreover, both the enzymes hydroxylate only those proline and lysine residues that are incorporated in a polypeptide chain.

Because the hydroxylated amino acids are important for stabilizing the structure of collagen through hydrogen bond formation, in vitamin C deficiency decreased synthesis of the hydroxylated amino acids results in weakening of the collagen fibres. This leads to the fragile blood vessels, incomplete bone formation and poor wound healing—the characteristic of the disease **scurvy** (**Case 18.1**).

Glycosylation

Collagen is unique in having **covalently attached** monosaccharides (galactose) or disaccharides (glucose-galactose), linked to the hydroxyl of 5-hydroxylysine. These are attached in the endoplasmic reticulum by Mn^{2+} containing *UDP-glycosyltransferases*. Like hydroxylation, glycosylation occurs concomitantly with chain synthesis and is inhibited by triple helix formation. The carbohydrate content of collagens varies considerably from 0.5–1.0% in types I and III to nearly 15% in type IV, but precise significance of glycosylation is yet to be elucidated.

The collagen polypeptides are post-translationally modified on transport through RER and Golgi.

Procollagen Triple Helix Formation

Three pro-α-chains assemble together to form the triple helix. The process occurs while the chains are still in the endoplasmic reticulum. Note that when the collagen polypeptides are initially synthesized, they have additional amino acid residues (100–300) on both their N and C termini that are absent in the mature collagen fibre (Fig. 5.5). They are called **extension peptides**, and they help to correctly align the three polypeptides as they come together in the procollagen triple helix. This process may be aided by the formation of disulphide bonds between extension peptides on neighbouring polypeptide chains. The extension peptides also prevent the premature aggregation of the procollagen triple helices within the cell.

The procollagen is packaged into secretory vesicles within Golgi apparatus, and then secreted out of fibroblasts.

Procollagen to Tropocollagen Formation by Removal of Extension Peptides

On secretion out of the fibroblast the extension peptides are removed by the action of *extracellular peptidases*. *Procollagen peptidases,* removes them leaving behind the triple helical, mature collagen monomer, the tropocollagen.

Collagen Fibril Formation

The newly formed tropocollagen molecules aggregate to form collagen fibrils (which would cross-link extensively to form collagen-fibres). For the formation of a fibril, the tropocollagen molecules aggregate in a staggered head-to-tail arrangement. As shown in Figure 5.5, each tropocollagen associates itself in a parallel, overlapping staggered array with other tropocollagen molecules. A typical fibril consists of hundreds or even thousands of tropocollagen molecules in cross section.

The extension peptides at both the N- and C-terminal of polypeptide chain directs alignment of the three polypeptide chains. Their removal by *peptidases* forms tropocollagen, which aggregate in staggered array to form collagen fibril.

Cross-linking and Fibre Formation

The tropocollagen molecules (which have aggregated into fibrils) now become extensively cross-linked covalently to produce mature collagen fibres. The special tensile strength of collagen fibres is provided by these **covalent cross-links** both between and within tropocollagen molecules (though hydrogen bonds also have a stabilizing effect).

As there are few, cysteine residues in the final mature collagen, these cross-links are not disulphide bonds as commonly found in proteins, but rather are unique cross-links formed between lysine and its aldehyde derivative **allysine**. Allysine residues are formed from lysine by the action of *monooxygenase lysyl oxidase* (Fig. 5.6). It is a copper containing enzyme, also requiring *pyridoxal phosphate*, derived from vitamin B_6. The aldehyde group on allysine then reacts spontaneously with either the side-chain amino group of Lys or with other allysine residues on other polypeptide chains to form covalent interchain bonds.

Tropocollagen, formed after removal of extension peptide, aggregate into fibrils and are covalently cross-linked to form mature collagen fibres.

Decreased maturation due to impairment of formation of various linkages alters mechanical properties of skeletomuscular system (**Case 5.2**).

D. Degradation of Collagen

Collagen tissue undergoes constant but slow turnover, half-life being up to several months. The degradation occurs during tissue repair and during normal growth and development, and is initiated via the enzymes *collagenases*.

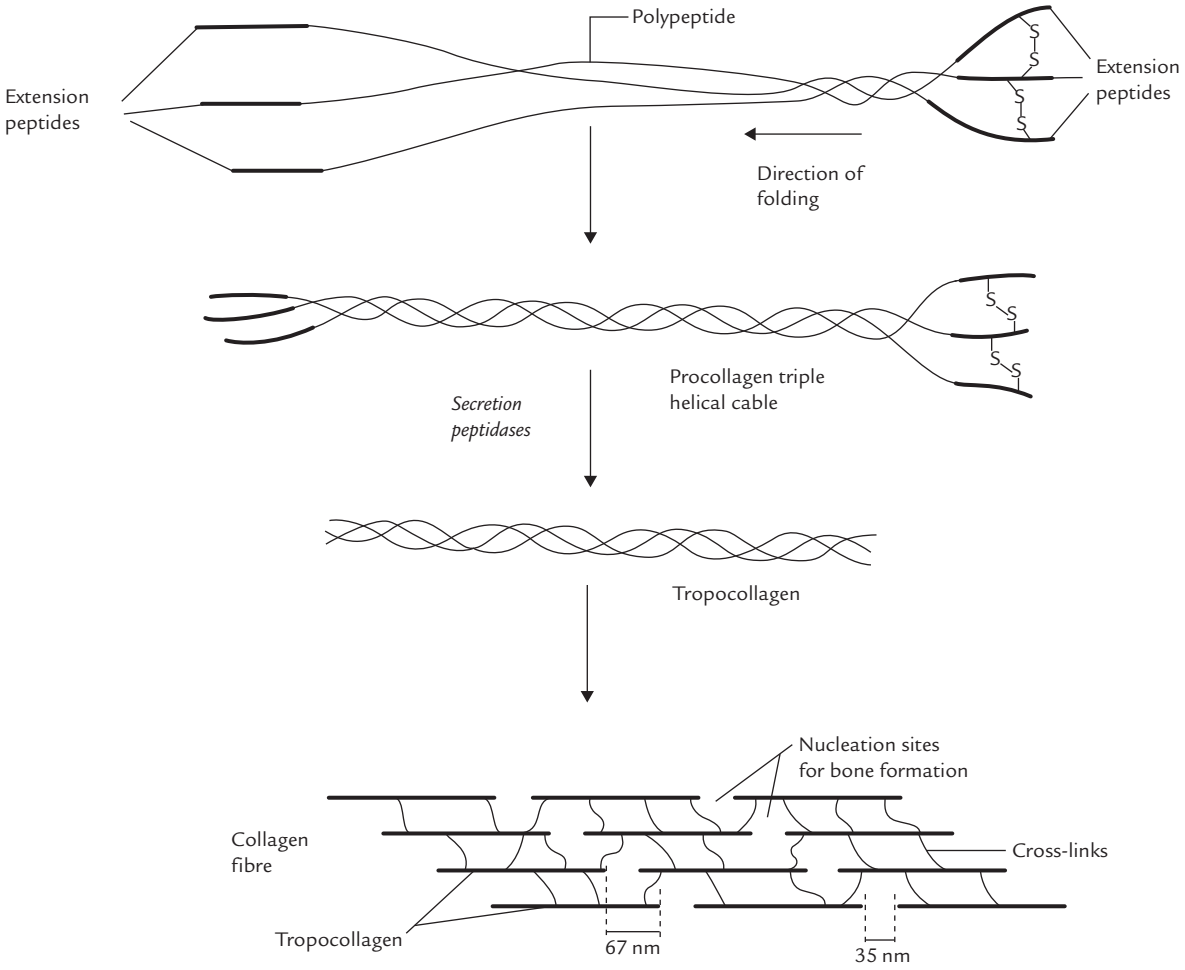


Fig. 5.5. Formation of procollagen and its processing by proteolysis to yield tropocollagen. The tropocollagen molecules aggregate to form collagen fibrils, which cross link extensively to form collagen fibres.

These are extracellular *zinc containing proteinases* that initiate degradation by splitting a single Gly-Ile bond in each of the three-tropocollagen strands. Two large fragments are produced, which then spontaneously denature and become susceptible to attack by several intracellular *proteases* (e.g. **cathepsins**). Further degradation is carried out by a multipronged attack by these enzymes.

Collagen (Triple-helix) —*Collagenases*→ Polypeptide fragments —*Cathepsins*→ Peptides

Action of *collagenases* is important because native collagen is resistant to attack by most *proteinases*, unless it is denatured. Ability to secrete large amount of *collagenase* is possessed by tumour cells, which helps their metastatic spread.

E. Genetic Defects of Collagen Biosynthesis and Metabolism

Inherited abnormalities in the procollagen chains or deficiencies of enzymes for their post-translational processing

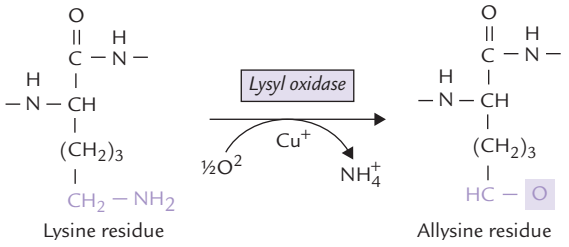


Fig. 5.6. The *lysyl oxidase* reaction.

may lead to certain genetic disorders. The underlying biochemical defects include abnormal collagen genes, abnormal post-translational modification of collagen, or deficiency of cofactors needed by the enzymes that carry out the post-translational modifications. Defective collagen formed in these diseases is unable to provide the required tensile strength.

Ehlers–Danlos Syndrome

It is heterogeneous group of genetic disorders characterized by stretchy skin and loose joints. At least 10 types are known, some of which may be traced to defective collagen

synthesis or to incomplete removal of propeptides (Table 5.2). Type IX, for example, arises due to X-linked recessive deficiency of ***lysyl oxidase***, resulting in impairment of cross-link formation. As a result, joint hypermobility and hyperextensible skin is seen. Autosomal recessive deficiency of ***lysyl hydroxylase*** occurs in type VI disorder. It manifests as hyperextensible skin, scoliosis and hyperextensible joints (**Case 5.2**).

The "India rubber man" who could bend and twist himself into incredible shapes and package himself into tiny boxes had Ehlers–Danlos syndrome probably.

Osteogenesis Imperfecta (OI)

It is a group of genetic disorders in which biosynthesis of type I collagen is defective. **Point mutations** in the genes for pro-α_1(I) or pro-α_2(I) have been identified as causing OI.

OI is characterized by fragile bones, hence also known as the **brittle bone syndrome**. Clinical features vary widely in severity: in mild forms occasional pathological fractures may be the only presentation, whereas in severe forms skeletal deformities and dwarfism, with even perinatal deaths are seen. Extra skeletal manifestations include a blue discolouration of sclera, hearing loss, and poor teeth development. The incidence of osteogenesis imperfecta is approximately 1 in 10,000, and the inheritance is autosomal dominant in most cases.

Epidermolysis Bullosa

It is a heritable disease in which glycosylation of collagen is impaired and the skin characteristically blisters. The type VII collagen is absent; this type of collagen forms anchoring fibrils at the dermis-epidermis junction, and they anchor the basement membrane to the underlying dermis.

Dermatosparaxis

There is a hereditary impairment in the removal of the terminal propeptide extensions. Because fibril formation can occur only when these propeptides have been removed, there is impairment in fibril formation and hence, reduced tensile strength of collagen.

Several inherited disorders of collagen structure and biosynthesis, including osteogenesis imperfecta and Ehlers–Danlos syndrome, are known.

Disorders of Collagen Degradation

Many collagen disorders are characterized by abnormal activities of collagen-degrading enzymes (**Case 5.3**). These enzymes are discussed earlier (p. xxx). Their activity is low in adult human tissues, but in some pathological states characterized by increased collagen degradation, such as Paget's disease, rheumatoid arthritis, and tumour metastasis, higher activity of these enzymes occur.

F. Effect of Aging and Disease on Collagen Metabolism

The gourmet knows that meat of young animal tastes better for being soft and tender, whereas that of old animals is tough and unpalatable. The same is true for human flesh. The reason for this age related change is that the *collagen of old animals and humans have more covalent cross links than those of the young*. Also the amount of collagen, relative to the proteins of parenchymal cells, increases with age.

Table 5.2. Ehlers–Danlos syndrome

Type	Inheritance*	Cause	Signs and symptoms
I	AD	Not known	Stretchy skin, loose joints, easy bruising with formation of characteristic cigarette-paper scars
II	AD	Not known	Like type I but milder
III	AD	Not known	Joint hypermobility, no scarring
IV	AD	Mutation in genes for type III collagen	Rupture of large blood vessels, the colon and the gravid uterus**
VI	AR	Deficiency of *lysyl hydroxylase*	Joint hypermobility, ocular fragility, extensible skin
VII	AD	Failure to remove the N-terminal propeptides in type-I collagen chain	Joint hypermobility, hip dislocation
IX	X-linked, recessive	Deficiency of *lysyl oxidase*	Joint hypermobility, hyperextensible skin

* AD, Autosomal dominant; AR, Autosomal recessive.
** These tissues are rich in type III collagen.

Collagen synthesis is also stimulated by tissue injury. During wound healing, the fibroblasts migrate to the site of injury and produce abundant collagen. Scar tissue consist mostly of collagen. The death of parenchymal tissue is followed by deposition of a collagen-rich extracellular matrix. For example, in **liver cirrhosis**, which is a common outcome of a variety of hepatic disorders, most of the parenchymal cells are replaced by fibrous (scar) connective tissue. Collagen metabolism is affected in a number of other diseases, e.g. **scurvy** (see **Case 18.3**) and **lathyrism**. In the latter condition cross-linking of collagen is impaired due to inhibition of the enzyme *lysyl oxidase* by a natural product (3-amino-propionitrile, an irreversible inhibitor). Copper deficiency also leads to enzyme inhibition resulting in failure of collagens to acquire the tensile strength they need to function.

II. Elastin

Elastin is a fibrous element which (unlike collagen) can be readily stretched. It is the major fibrous protein of yellow, elastic connective tissue of ligaments and large arteries.

The inextensibility of collagen suits it to forming extracellular matrices that have high tensile strength (Achilles tendon can withstand about 2000 kg cm^{-2}) or have ability to withstand compression (250 kg cm^{-2} for the intervertebral discs). In some tissues (skin, ligaments, and large arteries) there is requirement for elastic deformability. Such tissues have high content of elastic fibres (10% in skin to over 50% in large arteries and some ligaments). These fibres consist mainly of elastin.

Structure

The basic unit of elastin structure is **tropoelastin**, a polypeptide of 800–850 amino acid residues (MW 72000). Like tropocollagen, the tropoelastin has aberrant amino acid composition, with high proportions of *glycine* (31%), *alanine* (22%), and *hydrophobic amino acids* (40%). About 10–13% *proline* and 4-*hydroxyproline* are present, but there is no hydroxylysine. It has been proposed that tropoelastin units are present in the random coil conformation and are extensively cross-linked. This makes such a network kinetically free: free to stretch and recoil. Like in case of collagen, these cross-links are derived from allysine. Therefore, *lysyl oxidase* is required for the synthesis of elastin as well as that of collagen.

Elastin, the most important constituent of the rubber-band-like elastic fibres, guarantees the elasticity of such tissues, as ligaments, arteries, and lungs.

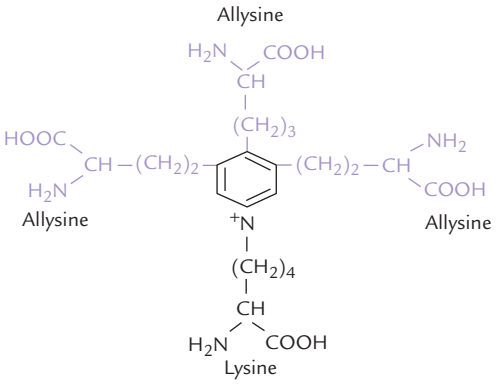


Fig. 5.7. The structure of desmosine showing 3 molecules of allysine and one molecule of lysine attached to pyridinium nucleus. Besides desmosine, elastin also contains some cross-links typical of collagen.

Cross-linking

The cross-linking among tropoelastin chains involves allysine and lysine residues. As shown in Figure. 5.7, three molecules of allysine and one molecule of lysine come together to form a new amino acids: **desmosine** and **isodesmosine**. Both contain a pyridinium nucleus, derived from the R groups of lysine and allysine residues. Thus, desmosine serves as the central connection that connects different polypeptides.

The non-specific inhibitors of collagen cross-linking, such as copper deficiency or lathyrism, inhibit cross-linking of elastin also. It is well known, for instance, that **copper deficiency** produces dissecting aneurysms in blood vessels of animals through incomplete elastin cross-linking. It appears, however, that elastin abnormalities observed in some inherited collagen diseases such as **Marfan's syndrome** (one of the glycoproteins in microfibrils, fibrillin, is defective due to gene mutation) may be secondary to aberrations in collagen structure. Patients with this dominantly inherited condition have unusually tall stature, with long, spidery fingers; the lens is displaced; and the media of the large arteries is abnormally weak. Many patients die suddenly in mid-life after rupture of dilated aorta.

III. Glycoproteins

A. Overview

Glycoproteins are conjugated proteins. Like all other conjugated proteins, the glycoproteins contain a non-protein portion (i.e. **prosthetic group**) that is covalently attached with the protein portion of the molecule. The non-protein portion in case of glycoproteins are short, branched oligosaccharide units that are covalently

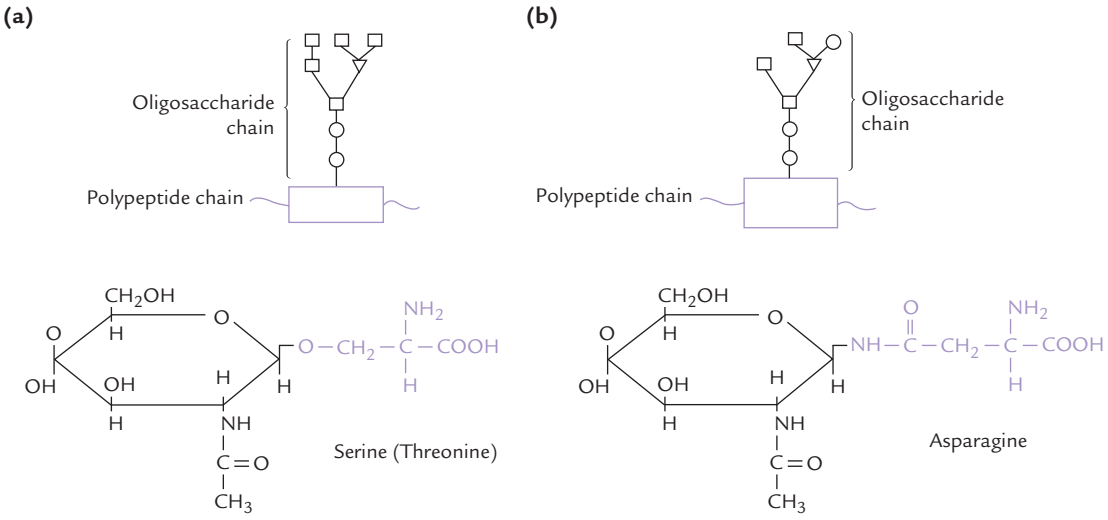


Fig. 5.8. Structure of glycoprotein. **(a)** O-Glycosidic linkage, **(b)** N-Glycosidic linkage.

attached to the polypeptide chain (i.e. **aglycone**) at several points (Fig. 5.8).

Glycoprotein = Polypeptide (aglycone) + Short oligosaccharides

Glycosaminoglycans also consist of same two components as glycoproteins. However, these are three remarkable differences:

(a) the length of the oligosaccharide chain is relatively short (less than 10 sugar residues) in glycoproteins, whereas it is longer in glycosaminoglycans
(b) the carbohydrate chains in glycoproteins are often branched
(c) The oligosaccharide chains of glycoproteins lack a serial repeating unit; this is in contrast to the carbohydrate chains in glycosaminoglycan that consist of repeating disaccharide units (amino sugar-acid sugar; Chapter 2).

B. Functions of Glycoproteins

Several cellular functions require participation of glycoproteins. Some of these are:

1. Glycoproteins are chief components of mucin of the gastrointestinal and urogenital tracts, where they act as ***protective lubricants***.
2. The cellular phenomena, including ***cell surface antigenicity*** (e.g. the blood group antigens, refer to Box 5.1 for details); and the cell surface recognition (by other cells, hormones, viruses) require glycoproteins.
3. Several *hormones* are glycoprotein in nature, including the follicle-stimulating hormone (FSH), luteinizing hormone (LH), chorionic gonadotropin, etc.
4. Several proteins that are secreted from cells into extracellular fluids are glycoproteins. Almost all *plasma proteins* are glycoproteins. Other proteins like *ceruloplasmin*, *plasminogen*, *prothrombin*, and *immunoglobulins* are also glycoproteins.
5. Glycoproteins are integral components of cell membranes, where they may play important role in the group behaviour of cells. They determine polarity of the biological membranes as well.
6. Several *structural* proteins, such as collagen, laminin, and fibronectin contain attached saccharide units.

C. Structure of Glycoproteins

The Carbohydrate Component

The oligosaccharide units of glycoproteins are **short and branched heteropolymers**. In general, the sugar moieties are in the D-form. The principal moieties present in these oligosaccharides are: D-mannose (common in N-linked oligosaccharides), D-galactose, L-fucose (α-6-deoxy hexose), N-acetyl glucosamine, N-acetyl galactosamine and N-acetyl neuraminic acid (NANA). The latter is usually found in terminal positions of oligosaccharides.

Oligosaccharides of glycoproteins show great **structural diversity**. They range in size from 2 sugar residues in the most simple (O-linked) oligosaccharides to more than 15 in some more complex (N-linked) oligosaccharides. The carbohydrate content of different glycoproteins accordingly varies from less than 5% to more than 50%. For instance: IgG contains as little as 4% carbohydrate by weight, whereas glycophorin, a red cell glycoprotein, contains as much as 60% carbohydrates.

The heterogeneity among glycoproteins is due to difference in the sugar component. For example, *pancreatic*

BOX 5.1

Blood Group Antigens

The surface of the human erythrocyte is covered with specific antigenic determinants, many of which are saccharides linked with polypeptides (i.e. glycoproteins). They determine the antigen specificity of the erythrocytes. There are about 160 blood group determinants, belonging to 21 independent human blood group systems. The most widely studied are the antigenic determinants of the ABO blood group system, and the closely related Lewis's system. In both these systems, the antigenic determinants are glycoproteins only. A minor alteration of the saccharide units of the glycoproteins can alter the antigenicity.

There are four main groups (A, B, O, and AB), distinguished on the basis of three blood group substances on erythrocytes. Individuals belonging to the group A have A antigen and those belonging to group B possess B antigen. In plasma, there are two naturally occurring antibodies (α and β) to these antigens. The group A persons have β-antibody and the group B persons have α-antibody. AB people have both A and B antigens but no antibody in their plasma. These people are universal recipients. The universal donors on the other hand are those that have no reactive antigens but have both α- and β-antibodies. The blood group of universal donors is denoted as O. They are said to carry H antigen, which represents the basic antigenic structure, as shown below:

Fucose—Gal—NAc—Protein

Gal NAc = N-acetyl galactose

The A and B antigens differ from the H antigen in having additional sugar units, as shown below:

A antigen : Fucose—Gal < Gal NAc—Protein
Gal NAc

B antigen : Fucose—Gal < Gal NAc—Protein
Gal

What is the genetic basis of different blood groups? Presence or absence of structural genes for the specific *glycosyl transferase* causes different blood groups. For instance, persons with blood group A have structural genes for *N-acetyl galactosyl transferase*. This enzyme transfers an additional sugar unit (GalNAc) to the basic antigenic structure. The people with group B have *galactosyl transferase* activity. Lack of both the structural genes leads to blood group O, whereas presence of both leads to blood group AB.

ribonuclease A and B have identical amino acid sequences, but their carbohydrate composition is significantly different. A given glycoprotein from different animal species usually has identical amino acid sequence in the polypeptide chains, but variable carbohydrate components, which accounts for their different properties. *Thus the carbohydrate component is responsible for the distinct identity of a glycoprotein.*

Role of Oligosaccharides: Oligosaccharides of glycoproteins are important for many of these biological functions of glycoproteins. They play a role in maintenance of their higher-order structure, water solubility, antigenicity, cell surface recognition and regulation of the metabolic fate of the protein.

Link between Carbohydrate and Protein

The saccharide unit of a glycoprotein are covalently linked to the polypeptide by specific glycosidic linkages, termed the ***glycopeptide bonds***. Two types of such linkages are present in glycoproteins: **O-glycosidic** or **N-glycosidic linkages** (Fig. 5.8). In general, the O-glycosidic linkage is formed by attachment of the sugar unit to a hydroxyl (–OH) group of the aglycone whereas the N-glycosidic linkage is formed by attachment of the sugar unit to an amino ($-NH_2$) group of the aglycone. In case of glycoproteins, these linkages are formed as here:

- **O-Glycosidic linkage** joins the sugar unit to the hydroxyl group at side chain of either *serine* or threonine.
- **N-Glycosidic linkage** joins the sugar unit to the amide group of an *asparagine* side-chain.

One or both types of linkages may be present in a given glycoprotein molecule.

Glycoproteins may contain N-linked oligosaccharides (attached to asparagine) or O-linked oligosaccharides (attached to serine or threonine) or both. Different molecules of a glycoprotein may contain different sequences and locations of oligosaccharides.

D. Synthesis of Glycoproteins

Synthesis of the N-linked and the O-linked glycosides differ from each other in certain important aspects, therefore, discussed separately.

Synthesis of O-linked Glycosides

The polypeptide chain of the glycoprotein is synthesized first; and later on the individual sugar units are sequentially attached to it, one at a time. The following steps are involved:

1. **Synthesis of the polypeptide chain**: It occurs in ribosomes that are associated with the membrane of endoplasmic reticulum (ER). The completed chain is transported into the lumen of the ER where the glycosylation would occur.
2. **Attachment of sugar units to polypeptide (Glycosylation):** It begins in ER. The polypeptide chain is then transported to Golgi apparatus, where the glycosylation continues and oligosaccharides are constructed by stepwise addition of monosaccharides. The enzymes catalyzing glycosylation are *glycosyl transferases* (*transglycosylases*), which are bound to the membranes of these subcellular organelles, i.e. the ER and the Golgi apparatus.

   Substrates for the *glycosyl transferases* are **nucleoside diphosphate sugars** (NDP derivatives). These sugar nucleotides represent high energy state, and hence are capable of donating sugar to the growing oligosaccharide chain; for instance, the UDP glucose can add a glucose moiety to the growing oligosaccharide chain (Fig. 5.9), UDP galactose can add galactose moiety, and so on.
3. The glycoproteins synthesized in this way are either released by the cell into extracellular fluid (ECF) or become part of the cell membrane.

Synthesis of N-linked Glycosides

Synthesis of N-linked glycosides is more complex and has additional processing steps. The characteristic feature of the synthetic process is that the polypeptide chain does not become glycosylated with individual sugar units; but rather a (mannose rich) core-oligosaccharide is constructed initially, attached to a lipid, **dolichol** (Fig. 5.10). Subsequently, the *core-oligosaccharide is transferred from the dolichol to the newly synthesized polypeptide chain*. The following steps are involved in the biosynthesis of the N-linked glycosides:

1. ***Synthesis of oligosaccharide unit***: It requires participation of **dolichol**, a polyisoprenol in the membrane of ER consisting of 15–18 isoprene units (Fig. 5.10a), which can be synthesized from acetyl CoA. Sugar units are attached to the dolichol pyrophosphate from their activated precursors in a stepwise manner. The enzyme *glycosyl transferase* catalyses the construction of branched oligosaccharides (Fig. 5.10b). The first sugars to be added by these membrane-bound enzymes are N-acetylglucosamine, mannose and glucose.

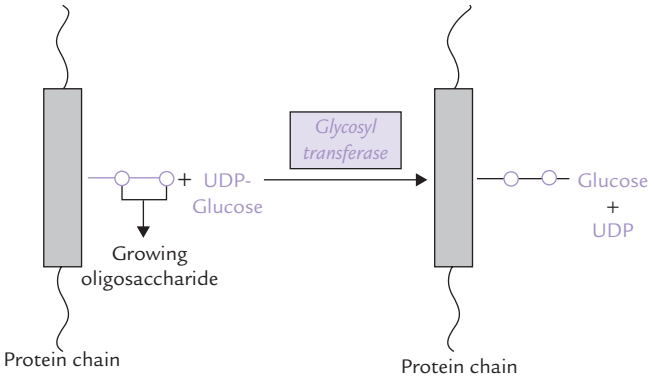


Fig. 5.9. Synthesis of O-linked glycoside by stepwise addition of individual sugar units (UDP = urine diphosphate).

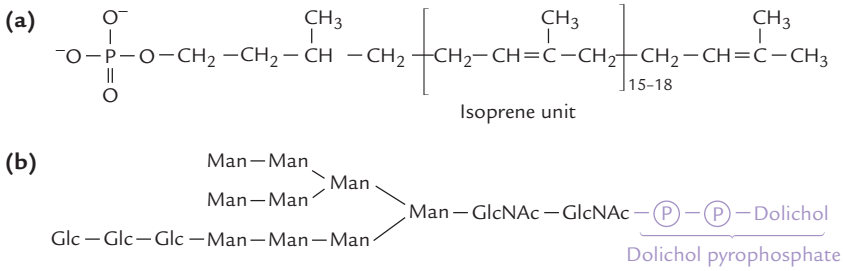


Fig. 5.10. Synthesis of N-linked oligosaccharides in glycoproteins. **(a)** Structure of dolichol phosphate. This lipid is present in the ER membrane, where it is used as a carrier of the core oligosaccharide, **(b)** Dolichol bound to the precursor oligosaccharide unit. This oligosaccharide is synthesized by the stepwise addition of the monosaccharides from activated precursors. Transfer of the precursor oligosaccharide to an asparagine side chain of the polypeptide. This transfer reaction is cotranslational.

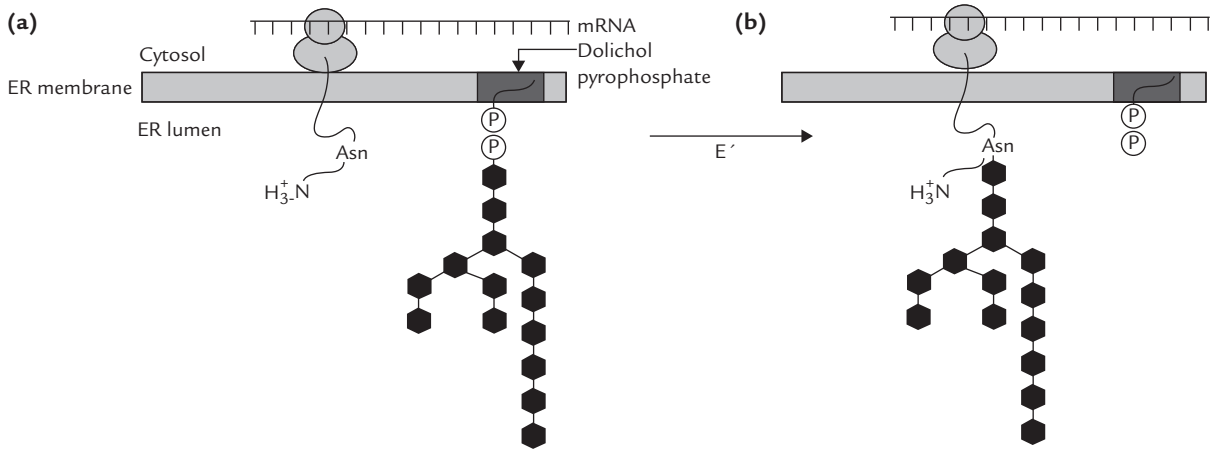


Fig. 5.11. Synthesis of N-linked glycoside. Polypeptide chain is synthesized in ribosomes associated with membrane of endoplasmic reticulum. There is a concomitant construction of oligosaccharide unit attached to dolichol phosphate. Transfer of the oligosaccharide unit to the amide group of asparagines occurs next; this transfer reaction is cotranslational.

2. ***Synthesis of polypeptide chain***: At the same time, a polypeptide chain is synthesized in ribosomes associated with membrane of endoplasmic reticulum (ER). It is then transported into the lumen of the ER (Fig. 5.11).
3. ***Transfer of oligosaccharides from dolichol to polypeptide***: The oligosaccharide unit is then transferred from the dolichol pyrophosphate to side group of an asparagine of the growing polypeptide chain. The enzyme catalyzing this transfer, *protein-oligosaccharyl transferase (E')*, is present in the lumen of ER.
4. ***Trimming and/or addition of monosaccharides***: After incorporation into the protein, the oligosaccharide is processed by removal of specific carbohydrate residues (mannose or glucose) as the glycoprotein moves through. The enzymes responsible for such processing are termed *glycosidases*. This process continues in the Golgi apparatus also. Finally, certain sugar units are added to complete the oligosaccharide chain synthesis.

Ultimately, the N-linked glycoproteins are either released extracellularly or become part of the cell membrane.

E. Degradation of Glycoproteins

Degradation of glycoproteins requires specific lysosomal enzymes called glycosidases. Usually these enzymes are *exoglycosidases*: they remove sugars in a series of reactions from the non-reducing end of the oligosaccharide chain. With each removal, the substrate is exposed for the action of the next enzyme. If any one of the *exoglycosidases* is absent, degradation of the glycoprotein by the next *exoglycosidase* cannot occur. Action of the *exoglycosidases* is supplemented by that of the *endoglycosidases* which attack glycosidic bonds that lie towards the interior. A concerted action by both *exo-* and *endoglycosidases* is highly effective in degrading the large glycoprotein molecules.

Genetic deficiency of one or more of the lysosomal hydrolytic enzymes results in accumulation of partially degraded glycoproteins and oligosaccharides in cells. These disorders are termed the **glycoprotein storage diseases**.

IV. Plasma Proteins

Plasma contains over 200 individual proteins that are structurally and functionally different from each other. Plasma concentration of these proteins is **6.3–8.8 g/dL**. They are broadly divided into three types: albumin, globulin, and fibrinogen; their plasma concentrations are:

Albumin = 3.6–5.4 g/dL
Globulins = 1.8–3.6 g/dL
Fibrinogen = 0.18–0.36 g/dL

The globulin fraction is subdivided into numerous components like α, β, γ, etc.

All these proteins except the γ-globulins are synthesized in the liver. It may be noted that **plasma** is the soluble fraction of blood, which contains all plasma proteins, including fibrinogen and other clotting factors. If, however, a blood-clot forms before centrifugation, the supernatant no longer contains fibrinogen, and some of the other clotting factors are consumed as well. This solution is called **serum**. For most diagnostic tests, either plasma or serum can be used. Most of the serum proteins as well as fibrinogen have been isolated and thoroughly characterized, forthwith discussed in this section.

Concentration of plasma proteins is between 6.3 g/dL and 8.8 g/dL. With important exception of γ-globulins (the immunoglobulins), most of the plasma proteins are synthesized in liver.

Separation of Serum Proteins

Individual proteins can be isolated by an array of techniques that depend on their size, shape, net charge, and solubility. Among these techniques the common ones are:

1. *Stepwise precipitation* of proteins with salts is known as **salting out**. Different proteins precipitate at different salt concentrations; for example, 0.8 M ammonium sulphate precipitates fibrinogen, whereas a concentration of 2.4 M is needed to precipitate serum albumin.
2. *Chromatography* on cation and anion exchangers.
3. *Fractionation* with organic solvents such as ethanol or ether.
4. *Gel filtration* on media of specific porosities.
5. *Electrophoresis*.

The basic principle of these techniques have been described in Chapter 3.

Electrophoresis: Among all the techniques enlisted above, electrophoresis is the most important for separation of serum proteins. Paper or agar gel electrophoresis with barbitone buffer at pH 8.6 yields an electrophoretic pattern shown in Figure 5.12. The fractions in the order of increasing mobility towards anode are albumin, α_1-, α_2-, β- and γ-globulin.

Albumin/Globulin (A/G) Ratio

The normal serum albumin concentration is 3.6–5.4 g/dL and that of globulins is 1.8–3.6 g/dL; therefore, the normal A/G ratio is **1.36** (mean), with a normal range of 1.2–1.6. It is lowered in various conditions where serum albumin concentration falls (hypoalbuminaemia) and/or when globulin concentration rises (hyperglobulinaemia). Various causes of decreased A/G ratio and its importance in clinical diagnosis is discussed later in this chapter.

A. Albumin

Albumin is the most abundant protein in plasma, which accounts for about 55–60% of the total plasma proteins. It is produced exclusively in liver; a healthy liver can synthesize 10–15 g of albumin per day. For this reason serum albumin levels are determined to assess a patient's liver function (decreased synthesis and thus fall of serum albumin occurs in ***hepatic disorders***).

Human albumin has a molecular weight of 69,000 and it consists of a single polypeptide chain of 585 amino acids with 17 disulphide bonds. Due to its low molecular weight it is one of the first proteins to appear in the urine after renal damage. Half-life of albumin is 20–25 days.

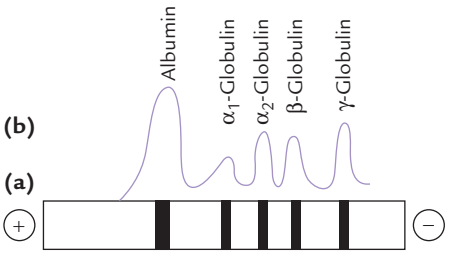


Fig. 5.12. Electrophoretic pattern of serum proteins. **(a)** After the electrophoretic run, the plate is stained with suitable dyes to visualize the protein bands, **(b)** Semi-quantification of the protein bands by scanning in densitometer:protein bands correspond with the peaks.

Functions

Albumin performs a variety of important functions described here.

Oncotic Pressure

Albumin plays an important role in osmotic regulation and fluid balance. As the capillary walls are relatively impermeable to plasma proteins, the latter exert an osmotic force across the capillary walls. This force helps to retain water in the intravascular compartment. The total colloidal osmotic pressure (oncotic pressure) is 25 mmHg, of which albumin accounts for about 80% (i.e. 20 mmHg). This is accounted by high concentration and low molecular weight of albumin. Hypoalbuminaemia, therefore, results in a significant reduction in osmotic pressure of blood, with consequent movement of water into interstitial spaces of the extremities and other parts of the organism, causing swelling. The condition is called **oedema**.

Transport

Albumin is involved in transport of several substances because of its predominantly polar nature. Nearly 40% of plasma calcium is bound-up with albumin. Bilirubin, steroid hormones, free fatty acids, and many non-polar drugs are also bound up with (and transported by) albumin.

Circulating Hormone Reservoirs

Albumin forms complexes with some of the circulating steroid hormones, thus preventing their filtration in the glomeruli. Such complexes act as circulating reservoirs of the hormones, which tissues can draw at the time of need.

Buffers

Albumin, along with other plasma proteins, is responsible for 15% of the total buffering capacity of blood. Among these proteins albumin has maximum buffering capacity.

Nutritive Function

Albumin degradation is an important source of essential amino acids during malnutrition, which are required for the maintenance of the organism's vital processes. Serum albumin concentration, is therefore, measured for assessing a person's protein nutritional state.

> Albumin, the most abundant of plasma proteins (60% of total) provides colloid osmotic pressure and is also a binding protein for several physiological substances and many drugs.

Diseases and Disorders

Concentration of serum albumin may increase (*hyperalbuminaemia*) or decrease (*hypoalbuminaemia*) in a number of conditions.

Hyperalbuminaemia

Dehydration, particularly if it develops rapidly, results in haemoconcentration. As a result, serum albumin concentration shows an apparent rise. The A : G ratio remains unaltered because serum globulin concentration also elevates proportionately. Vomiting and diarrhoea are major causes of dehydration.

In most cases, certain associated factors are present which oppose the effect of dehydration. For instance, vomiting and diarrhoea are associated with decreased protein absorption from the intestines. Consequently, serum albumin concentration may remain normal or even fall.

Hypoalbuminaemia

Certain conditions that result in hypoalbuminaemia are as follows:

Renal causes: The renal disorders in which loss of proteins in urine (i.e. proteinuria) occurs result in decreased serum albumin concentration. In **nephrotic syndrome**, a marked fall occurs because of a massive urinary loss of proteins (> 4 g/day) and the total serum protein concentration may fall below 4 g/dL.

Gastrointestinal causes: The gastrointestinal disorders in which the intestinal absorption of the digested molecules (including amino acids) is impaired result in hypoalbuminaemia. Some of the pathological conditions associated with **malabsorption** are enteritis, sprue, steatorrhoea, and malignant diseases involving intestine or pancreas.

Loss of proteins into the alimentary tract (known as **protein-losing enteropathy**) can occur in a number of diseases of the stomach and intestines. Occasionally, it may develop as a primary condition. Such a loss of proteins can be demonstrated by administration of intravenous Cr-51, followed by monitoring excretion of this compound in stools.

Insufficient dietary intake of proteins, for a prolonged period, may also lead to hypoalbuminaemia. This is because the serum albumin is degraded to provide amino acids for tissue protein synthesis.

Hepatic causes: As liver is the only organ for albumin synthesis, hepatic disorders cause decrease in serum albumin concentration (**Case 6.3**). In most cases, hypoalbuminaemia is accompanied by hypoproteinaemia as well. In some cases, however, there is an associated rise in the globulin fraction, so that total protein concentration remains normal or may even increase.

Other causes: Hypoalbuminaemia develops following severe haemorrhage, particularly if the blood loss is rapid. Initially there is loss of both plasma and proteins, but the plasma is restored earlier than the proteins. This results in decreased plasma albumin concentration. Similar effect is observed in shock, burns, crush-injuries, and in post-operative period. In burns, the fall in albumin concentration is greater because of extravasation of proteins.

Analbuminaemia

It is a rare genetic disorder where albumin is virtually absent. However, oedema does not generally develop since globulin levels are elevated in this condition, which tend to compensate for the absence of albumin.

Prealbumin (Transthyretin)

It is called so because it moves slightly ahead of albumin during electrophoresis. It participates in the transport of retinol from its major storage site in the liver to the extrahepatic tissues (Chapter 18): the retinol released from hepatic tissue is in tight binding to retinol binding protein (RBP). In blood circulation, the retinol-RBP forms a ternary complex with transthyretin.

$$\text{Retinol} + \text{RBP} + \text{Transthyretin} \rightarrow \text{Ternary complex}$$

The formation of this ternary complex is important because RBP (MW 21,000) is easily filtered at glomeruli and lost in urine. The ternary complex is large enough to escape this fate.

B. Globulins

Globulins comprise a number of proteins, which are separated into four distinct bands (α_1, α_2, β, and γ). Each of these bands is a mixture of proteins. Thus, the **γ-globulin** component consists of a variety of immunoglobulins

of the IgG type. Taken together, these immunoglobulins constitute the second most abundant component of serum proteins after albumin (Table 5.3). Other immunoglobulin types—IgA and IgM, have β-globulin mobility. The **β-globulin** group also contains haemopexin, transferrin, plasminogen, prothrombin and a lipoprotein called LDL (Chapter 12). Among **α_2-globulins** are—α_2-macroglobulin, haptoglobin and ceruloplasmin; while **α_1-globulins** include some hormone carrying proteins (e.g. thyroxine-binding globulin and the steroid-binding proteins, transcortin and orosomucoid) and α_1-antitrypsin (Table 5.3).

Biological significance of various globulins is summarized in Table 5.3; see also Box 5.2.

Important Globulins and their Functions

α_1-Antitrypsin

α_1-Antitrypsin is an α_1-migrating glycoprotein. It is a *serine protease* inhibitor, active against a number of proteolytic enzymes, particularly the *elastase* released by

Table 5.3. Plasma proteins and their functions in the human body

Proteins	Mean plasma concentration (g/100 ml)	Molecular weight (kD)	Function
Albumin	3.6–5.4	69	Osmotic regulation, transport, as hormone reservoir, buffering
Prealbumin	0.03	55	Transport of retinol and thyroxine
α_1-Globulins	0.3–0.5		
α_1-Antitrypsin	0.25	54	Inhibitor of *serine protease*
Orosomucoid	0.10	44	Binds with progesterone
Thyroxine-binding globulin	0.002	58	Major binding protein for T_3 and T_4
Transcortin	0.003	52	Binds glucocorticoids
α_2-Globulins	0.4–0.8		
α_2-Macroglobulin	0.20	725	*Non-specific protease* inhibitor
Ceruloplasmin	0.02	132	Copper transport, *Oxidase* activity
Haptoglobin	0.20	85	Binds haemoglobin
Prothrombin	0.02	63	Blood clotting
β-Globulins	0.6–1.1		
Haemopexin	<0.1	60	Transports haemoglobin
Transferrin	0.30	89	Transport of iron
Plasminogen	0.05	140	Retraction of fibrin clot
γ-Globulins	0.8–1.8	150–950	Very heterogenous
Fibrinogen	0.30	340	Precursor of fibrin clot

Immunoglobulins: IgA, IgG, IgM, IgD and IgE are mostly γ-globulins (some are β-globulins); their characteristics are outlined in Table 5.4.

BOX 5.2

Acute Phase Reactants

The acute phase reactants are a group of plasma proteins, concentration of which increases during inflammation. Some examples are α_1-antitrypsin, haptoglobin orosomucoid, ceruloplasmin, fibrinogen, C-3 complement protein and C-reactive protein. It is still not clear as to what triggers the production of these proteins. Probably, the elevation represents the body's attempt to bring inflammation under control. Measurement of one or more of these proteins may be of clinical value since their plasma concentration is elevated even before the signs and symptoms of the disease manifest.

The most sensitive of the acute phase reactants is C-reactive protein, α-, γ-globulin. It exhibits significantly elevated plasma levels within a few hours after an inflammatory stimulus. It is named so because of its ability to bind to a component of bacterial cell wall called C-polysaccharide. Its plasma concentration may rise up to 10-fold from normal concentration of 1 μg/dL. The elevated concentration is useful for predicting risk of heart disease.

neutrophils. Its role in the body is to protect lung tissue against proteolytic damage. It is implicated in two diseases: emphysema and liver disease.

Emphysema: It is a chronic disease characterized by abnormal lung distention. At least 5% of all cases of emphysema are due to impaired α_1-antitrypsin activity. Functional inactivation of α_1-antitrypsin occurs in smokers, and therefore chronic smokers are more prone to lung damage by *elastase*, and hence develop emphysema (**Case 4.2**). Smoking also increases neutrophil levels in the lungs (and hence release of *elastase*) and decreases α_1-antitrypsin levels in the bloodstream. α_1-Antitrypsin levels of emphysema patients are almost always below normal and this can also serve as a diagnostic tool for this disease.

Liver diseases: Some patients with α_1-antitrypsin deficiency develop neonatal hepatitis or infantile cirrhosis. In all likelihood this is caused not by lack of *protease* inhibition but rather by aggregation of mutant form of α_1-antitrypsin to form polymers. These non-secretable polymers accumulate in vesicles within the hepatocytes and cause damage by unknown mechanisms.

Deficiency of α_1-antitrypsin causes lung emphysema and liver cirrhosis (infantile).

Orosomucoid

Orosomucoid is an important constituent of α_1-globulins that binds progesterone and some drugs. Its concentration rises in trauma, malignancy, or inflammatory conditions.

α_2-Macroglobulin

α_2-Macroglobulin is an α_2-migrating glycoprotein of high molecular weight (725,000). It inhibits activities of *proteases*, such as trypsin, chymotrypsin, plasmin, thrombin and kallikrein.

Apart from α_2-macroglobulin, another α_2-migrating glycoprotein, α_2-antichymotrypsin, is a non-specific *protease* inhibitor, though less important than the former.

Ceruloplasmin

Ceruloplasmin is a copper containing α_2-globulin that binds approximately 60% of the copper in the serum and is itself loosely complexed to albumin or with histidine. A ceruloplasmin molecule binds with six atoms of copper, and also possesses *oxidase* activity, as discussed under copper metabolism (Chapter 19).

Haptoglobin

Haptoglobin is an α_2-migrating glycoprotein. It serves as a scavenger of the free intravascular haemoglobin. Haemolysis releases haemoglobin into plasma, from where it is at risk of being lost in urine. Haptoglobin prevents such loss by binding with haemoglobin to form a complex which is too large to be filtered in glomeruli. This complex is subsequently taken up by the reticulo-endothelial cells where the globin chain is degraded to release amino acids, and the haem is converted to bilirubin. The iron of haem is recycled for utilization. In this way, haptoglobin prevents wasteful loss of haemoglobin and its useful components (Fig. 5.13).

Normal plasma concentration of haptoglobin is 100–300 mg/dL. In haemolytic anaemia more of this protein is used for forming a complex with haemoglobin.

Half-life of this complex is much shorter (90 minutes) compared to that of haptoglobin (5 days), and so it is rapidly cleared from the plasma. As a result plasma concentration of haptoglobin falls.

Haemopexin

Haemopexin is a β-migrating glycoprotein. *It also serves as a scavenger protein for haemoglobin, but only after the haemoglobin-binding capacity of haptoglobin is saturated*. Excess haemoglobin is dissociated into the globin chains and haem; the haem is dissociated into αβ dimers, carried to liver taken up where it is by haemopexin, where iron atom is released for storage and reutilization. Thus, haemopexin and haptoglobin have analogous functions: both help in conserving the body iron (Fig. 5.13).

When binding capacities of both haptoglobin and haemopexin are exceeded, a complex of haem and albumin, called **methaemalbumin**, is formed. The latter continues to circulate until enough haemopexin becomes available.

Transferrin

Transferrin is a liver-derived β-globulin. It is the major iron transporter; essentially all iron in circulation binds with this glycoprotein.

Many plasma proteins bind small molecules: prealbumin binds thyroxine and is involved in retinol transport, haptoglobin binds haemoglobin, haemopexin binds haem, transcortin binds most of the circulating glucocorticoids, and transferrin is the major iron transporter.

Diseases and Disorders

Various pathological conditions are known to cause a generalized increase in serum globulin concentration (hyperglobulinaemia). Decreased serum globulin concentration (hypoglobulinaemia) however, never occurs.

Hyperglobulinaemia

It occurs in the following pathological conditions.

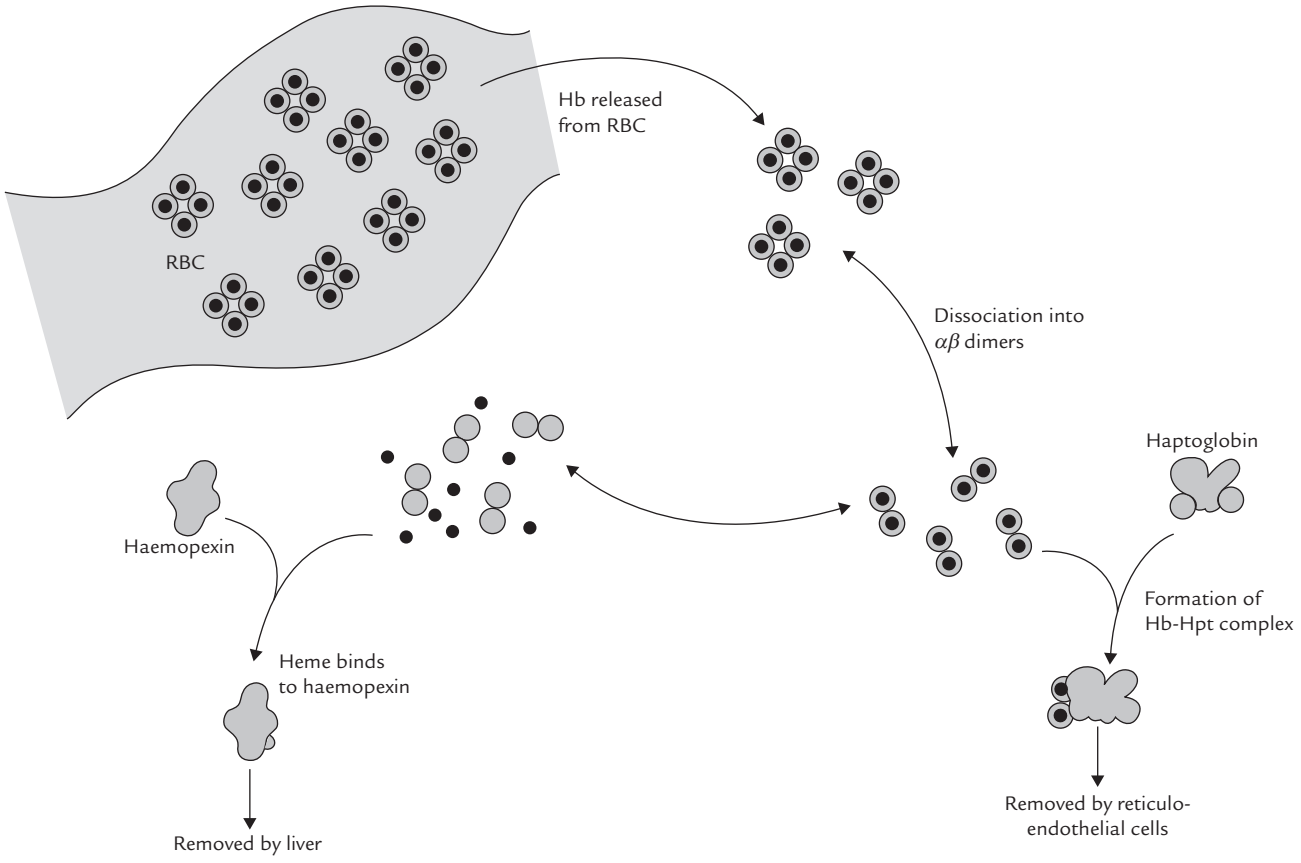


Fig. 5.13. Iron conservation by haemoglobin and haemopexin.

Advanced liver diseases: The serum globulin levels are typically elevated, often between 5–7 g/dL (normal = 1.8–3.6 g/dL). Since the albumin concentration is usually in the lower part of the normal range, or reduced (because of decreased synthesis), the A:G ratio is low.

Multiple myeloma: It is a group of diseases characterized by proliferation of a single plasma cell clone to cause extraordinary accumulation of immunoglobulin molecules in the patient's serum. Being derived from a single ancestral B cell, the proliferating plasma cells all produce the same (monoclonal) immunoglobin. Such immunoglobulins are termed paraproteins, and they produce a sharp peak in the γ-globulin fraction. The paraproteins may be of the IgG, IgA, IgM, IgD or IgE type. The malignant plasma cells grow in the bone marrow, where they cause bone pain, pathological fractures and radiological abnormalities. In many cases, a multiple myeloma patient excretes protein in the urine. This protein called **Bence-Jones protein,** is a dimer of κ or λ-chains that are formed in paraprotein.

Diagnosis: Demonstration of paraprotein upon serum electrophoresis is a reliable test for diagnosis of multiple myeloma. In addition, the **classical heat test** is performed. It involves the precipitation of Bence-Jones proteins when slightly acidified urine is heated to 40°−50°C and redissolve on further heating urine to boiling point. They reappear on cooling urine to about 70°C. A marked rise in serum globulin concentration is also observed in multiple myeloma, usually between 5 g/dL and 10 g/dL, but sometimes up to 14 g/dL (**Case 5.5**).

Chronic infections: These include rheumatoid arthritis, tuberculosis, kala-azar, lymphogranuloma venereum, Boeck' sarcoid and lupus erythematosus disseminatus. In most cases, the increase of globulins is only mild to moderate.

A/G Ratio and Electrophoretic Pattern in Some Diseases

As noted earlier, the mean A/G ratio of 1.36 decreases in various disease processes, for example in *nephrotic syndrome*, where albumin levels are low and/or globulin levels are high. Abnormal electrophoretic pattern, with decreased albumin and a prominent α_2-globulins are seen.

Some other important conditions where A/G ratio is reversed are *malnutrition* and *liver diseases* (albumin decreases), and *chronic infections* and *multiple myeloma* (globulin rises). Abnormal electrophoretic patterns are also seen; for example, in multiple myeloma a sharp distinct **M band** appears in γ-globulin region, in infections the α_1 and α_2 bands are increased.

C. Fibrinogen

Fibrinogen, the precursor of fibrin clot, appears to be formed in the liver. It is a soluble glycoprotein that consists of three pairs of identical polypeptide chains, called the **α-, β- and γ-chains** (MW 63,500, 56,000 and 47,000, respectively). Blood coagulation is initiated by the action of thrombin on fibrinogen, whereby two small peptides are removed from the fibrinogen. This results in formation of a fibrin monomer. Several monomers then aggregate via physical interactions to form a polymer that has properties of a soft gel. Further polymerization to form tough insoluble clots is accomplished by a specific enzyme called *factor XIII:* its action serves to join a number of linear fibrin polymers to form cross-linked fibrin lattices.

Diseases and Disorders

Decreased fibrinogen level in plasma is observed in severe liver diseases (e.g. acute hepatic necrosis). Low fibrinogen levels observed in antepartum haemorrhage, quickly return to normal after delivery.

Increased fibrinogen level is found in acute infections such as pneumonia, rheumatic fever, and tuberculosis. Since erythrocyte sedimentation rate (ESR) is affected by fibrinogen, it is markedly increased in these conditions. Hence, measurement of ESR is used for assessing the severity and course of these disorders. In contrast to most other acute infections, fibrinogen level is reduced in typhoid fever.

V. Immunoglobulins

The immune system is remarkable for its ability to generate a variety of specific proteins that include **antibodies** (**immunoglobulins**). The antibody reacts specifically with the antigen that triggered its production. **Antigens** are mostly proteins, but can be nucleic acids or polysaccharides. Thus, various bacterial and viral cell components may act as antigens. Antibodies recognize the antigen on the surface of the intruding organisms. The specific antigen-antibody interaction leads to activation of the complement system which destroys the foreign cell. Antibodies play important role in providing resistance against bacterial infections. Though it is mainly the γ-fraction of plasma proteins that comprises immunoglobulins, sometimes they separate with β- and α-globulin bands of electrophoretogram. Plasma concentration of immunoglobulins is about **1.0–2.5 g/dL**.

Immunoglobulins are synthesized by plasma cells. Plasma cells develop from B lymphocytes that have contacted foreign antigen. A single plasma cell can make only one type of antibody molecule. It is the accumulation of many plasma cells, each producing a distinct antibody, that gives rise to the multiple antibody specificities. Specific affinity of an antibody is not for the entire macromolecular antigen but for a particular site on it called **antigenic determinant** (or **epitope**).

This property is widely used in purification of proteins and in affinity chromatography (Chapter 4). The affinity of antibody for an epitope is due to the complementary structure of the two. If there is a cleft on the surface of the epitope, a corresponding elevation is envisioned on the antibody. If an antigen is injected, the body generates an antibody specific to each of the epitopes on the antigen.

Immunoglobulins or antibodies, a group of plasma proteins that are secreted by plasma cells, play an important role in body's defenses by specifically recognizing foreign molecules (antigens) and facilitating their selective elimination.

A. Structure

The basic structural unit of an immunoglobulin is a heterotetramer, made up of two light chains (**L**) and two heavy chains (**H**), joined through non-covalent interaction and disulphide bridges. Thus, the simplest description of an antibody molecule is H_2L_2 (Fig. 5.14). The light chains are simple proteins of 212 amino acids (MW 23,000 each) and the heavy chains are glycoproteins comprising approximately 440 amino acids (MW 53,000–75,000) each.

The light chains are of two types: κ or λ. A given immunoglobulin contains 2κ or 2λ chains, but never a mixture of the two. In humans, 60% light chains are of κ type.

There are five types of heavy chains: **γ**, **μ**, **α**, **ε** and **δ**. Depending on the heavy chain make up, the immunoglobulins are differentiated into **five classes**: **IgA** (α-chain), **IgD** (δ-chain), **IgE** (ε-chain), **IgG** (γ-chain), and **IgM** (μ-chain). All classes of immunoglobulins are present in every human being. (Within a class are found subclasses due to subtle differences in the amino acid sequences of heavy chains of the same class.)

Thus, structure of IgG can be represented as $\kappa_2\ \gamma_2$ or as $\lambda_2\ \gamma_2$; the IgA molecule as $\kappa_2\ \alpha_2$ or as $\lambda_2\ \alpha_2$ and so on. The heavy chains are hinged (flexible) in the vicinity of two inter-chain disulphide bridges. At these places, several prolyl residues are present, which confer a degree of flexibility on the molecule.

Human immunoglobulins exist as IgA, IgD, IgE, IgG and IgM classes, which contain α, δ, ε, γ and μ heavy chains respectively.

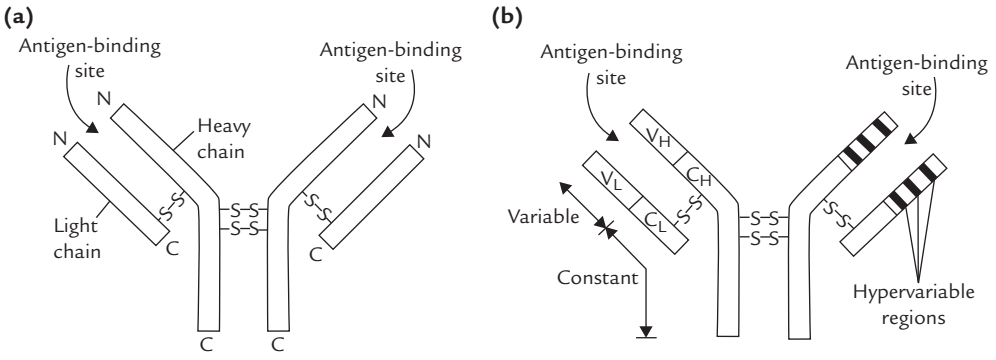


Fig. 5.14. Structure of an antibody molecule. **(a)** Each antibody (immunoglobulin) molecule has two identical light chains (L) and two identical heavy chains (H) consisting of 212 and 440 amino acids respectively, **(b)** Each chain has a variable region, where amino acid sequences shows considerable variations between individual immunoglobulin molecules and constant region where little variation occurs within one immunoglobulin class. The generic terms for these regions in the light chain are V_L and C_L and for heavy chain are V_H and C_H.

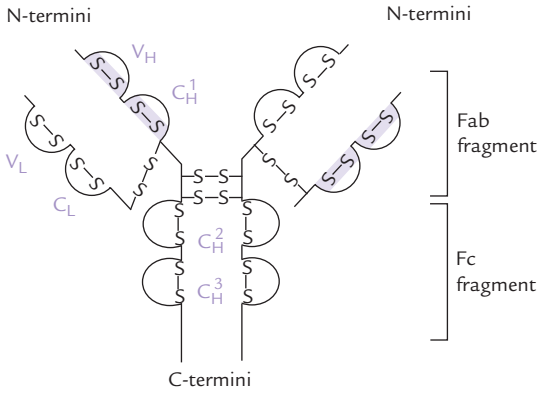


Fig. 5.15. Structure of an immunoglobulin molecule showing the disulphide-bridge limited domains. Domains on light chain are V_L and C_L; and the domains on heavy chain are: V_H, C_H^1, C_H^2, C_H^3.

Each of the four chains consists of variable region and a constant region, discussed later.

Both heavy and light chains have distinct globular regions called **domains** that have a highly organized tertiary structure. Each of these structural domains comprise about 110 aminoacyl residues. The light chain has two such domains, one in the variable region of the chain (V_L) and the other in the constant (C_L) region. The heavy chain, has four or five domains. The latter are termed V_H or C_H depending on whether they are present in the variable region or the constant region of heavy chain. Figure 5.15 shows domain arrangement of a typical immunoglobulin molecule.

Each domain is stabilized internally by a disulphide bond and generates a globular three-dimensional structure. Thus, antibody molecule can be visualized as being made up of distinct domains, connected by relatively non-ordered polypeptide segments. These domains have well defined biological functions.

- V_L and V_H are responsible for forming antigen-binding site.
- C_H^2 domain binds the complement.
- C_H^3 domain helps adherence to the cell surface.

All **immunoglobulins are glycoproteins**, and the degree of glycosylation in the region of the domain C_H^3 is variable.

An antibody consists of two light chains and two heavy chains joined by disulphide bonds. There are two antigen binding sites, each formed by a light chain and a heavy chain.

B. Variability

The amino acid sequences at the N-termini of both light and heavy chains show considerable variation between individual immunoglobulin molecules. Half of the light chain and quarter of the heavy chain constitute the **variable region**. Rest of the molecule, known as **constant region**, shows little variation within one immunoglobulin class (Fig. 5.14b). The light chain contains one variable (V_L) and one constant (C_L) region, whereas the heavy chain contains one variable and three or more constant regions. The variations in the amino acid sequences of the variable region permit the body to construct a large variety of three dimensional structures.

Moreover, within these V regions, certain segments are hypervariable. The 29–32, 48–52, and 93–96 aminoacyl residues in light chain and 30–36, 50–53 and 93–98 positions in heavy chains are termed **hypervariable regions**. They define the actual binding site for antigen. They are also termed the complementarity-determining regions (CDRs) as they form antigen binding site complementary to the topology of the antigen structure. In the

three-dimensional structure of the immunoglobulin molecule, the hypervariable parts of both the light and the heavy chains are looped together to form the **antigen-binding site**. Thus, they are referred to as the "hot spots" where amino acid sequence varies from one molecule to another. Such sequence variations are specific for each of the antibody molecules and are known as **idiotypic variations**.

Variability in the variable regions of both light and heavy chains is mainly localized to three hypervariable regions in each chain, which are looped together to form the antigen-binding site.

The remaining parts of variable region are far less variable and are called *framework regions*.

Proteolytic Cleavage of Immunoglobulin

Treatment with the proteolytic enzymes, **pepsin** and **papain**, cleaves the immunoglobulin (IgG) as shown in Figure 5.16. Papain digestion yields two univalent Fab fragments and an Fc fragment, whereas pepsin digestion yields a bivalent $F(ab')_2$ fragment. The cleavage occurs in the hinge region, where the antibody molecule can have limited degree of three dimensional movement. As discussed earlier, the hinge region lies in vicinity of two interchain disulphide bridges.

The **Fab (fraction antibody)** fragment binds with the corresponding antigen. It is constituted by whole of the light chain and N-terminal half of the heavy chain. **Fc (fraction crystallizable)** portion, constitutes the remaining part of the immunoglobulin molecule. It is concerned with activation of complement cascade and interaction with the cellular elements of the immune response. It triggers immune response that leads to the lysis of the intruding organisms.

Immunoglobulins from any class can also exist in two types: This is due to presence of either kappa (κ) or lambda (λ) light chain, giving rise to 10 different permutations of heavy and light chains. The κ chains are more common in humans than the λ chains.

While IgD, IgE, and IgG occur as a single Y-shaped antibody unit, IgM occurs as a pentamer and IgA can exist as a monomer, dimer, or trimer. Characteristics of various classes of immunoglobulins are summarized in Table 5.4.

There are five classes of heavy chains that define the class of immunoglobulin. Different classes of immunoglobulins have different properties.

Primary and Secondary Antibody Responses

Animals are capable of synthesizing specific antibodies virtually against all chemical groups. When the body system is

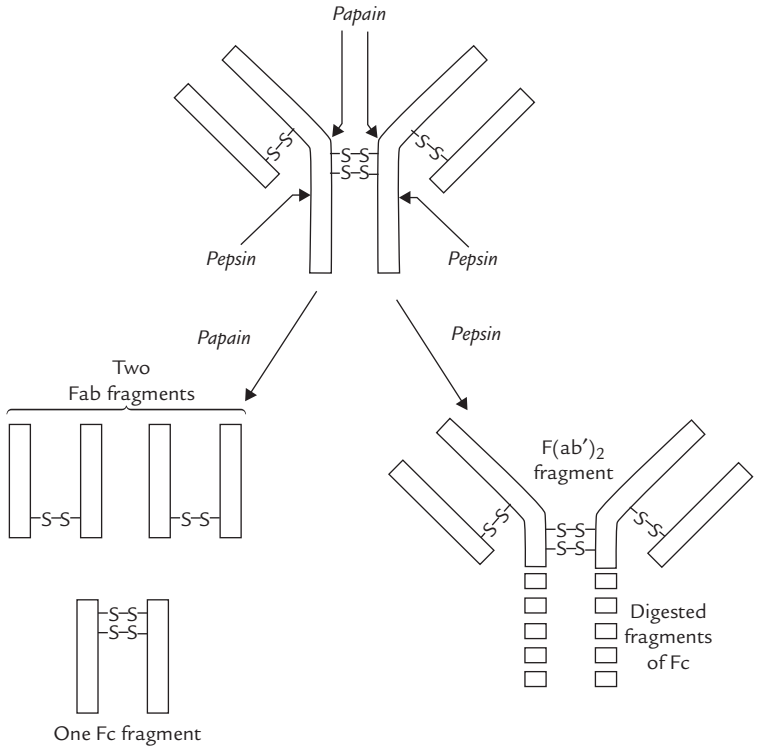


Fig. 5.16. Proteolytic change of IgG by pepsin and papain.

Table 5.4. Characteristics of the immunoglobulins

Class	Heavy isotype	Chain Mol wt (D)	Mean plasma concentration (g/dL)	Serum half-life (in vivo)–days	Mol wt (D)	Native form
IgG	γ	53,000	1.4	23	150,000	$\lambda_2\kappa_2$
IgM	μ	70,000	0.15	5	950,000	$(\mu_2\kappa_2)_5$
IgA	α	64,000	0.30	6	170,000	$\alpha_2\kappa_2$*
IgD	δ	58,000	0.003	3	160,000	$\delta_2\kappa_2$
IgE	ε	75,000	Trace	2.5	88,000	$\varepsilon\kappa_2$

* Can form polymer structures of the basic structural unit.

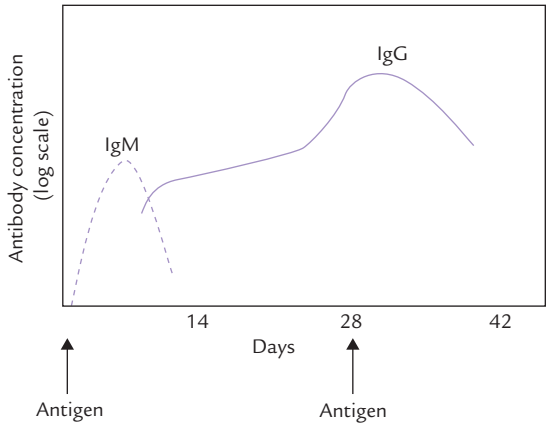


Fig. 5.17. Primary (---) and secondary (—) antibody response after the immune system is challenged by foreign antigen.

challenged by exposure to a foreign antigen, distinct primary response (IgM dominated) and secondary antibody responses are observed. IgM is the major antibody of primary response, whereas IgG dominates the secondary response. Figure 5.17 shows kinetics of the appearance of IgM and IgG following injection of foreign antigen.

The IgM are first to appear as they are synthesized at a faster rate than IgG. Thus, they serve as first line of defense until the particular type of IgG is produced. Level of serum IgM starts rising a few days after the (first) injection of antigen. Its level declines approximately 10 days after the injection. A booster dose of antigen at this time produces a further increase in the level of IgG in the serum.

C. Functions

The common function of all immunoglobulins is to serve as **the first line of defense** against foreign antigens. They recognize and bind with foreign antigens. In addition, Igs must interact with other cells and proteins for generating **effector functions** which are necessary for killing and eliminating foreign antigens. So functions of Igs can be grouped as below:

- Based on antigen recognition and binding.
- Based on effector response.

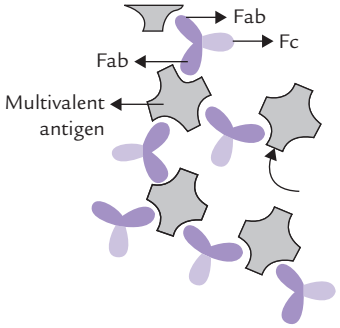


Fig. 5.18. Formation of large interconnected matrices by cross-linking of multivalent antigen and divalent antibody molecules.

Functions Based on Antigen Recognition and Binding

These are mediated by **Fab region** of the immunoglobulin, the region via which the immunoglobulin binds its antigen (ab stands for "antigen binding" F for "fragment"). Each Fab contains a binding site for an antigenic determinant. Being univalent (with a single binding site) a Fab alone cannot cross-link the antigen molecules that contain multiple determinants (i.e. multivalent antigens). Such cross-linking is possible with intact IgG molecule, which contains two identical antigen-binding sites. Each of these sites can bind to different antigen determinant on a multivalent antigen as shown in Figure 5.18. This opens the possibility of formation of large interlinked structures in which antibodies serve as bridges between the multivalent antigens. This may lead to precipitation of soluble antigens by cross-linking, or to agglutination by cross-linking of particulate antigens (e.g. viruses, bacteria, or parasites). Blocking of attachment of viruses or bacterial toxins (i.e. neutralization) may also occur in this manner.

Functions Based on Effector Response

The **Fc fragment**, called Fc because it is crystallized rapidly and does not bind with antigen. However, it mediates certain important protective responses called **effector functions** which enable the immunoglobulin to destroy the target cell. The following effector functions are known:

Complement activation: The complement factors, about 20 in number, get activated by Ig molecules and

cause lysis of the invading antigen. The Ig molecules that are capable of activating complement (IgG and IgM) have receptors for the complementary factors in their Fc region through which they bind these factors and activate the complement cascade. The complement activation enhances the immune response by

- forming a lethal complex of complement factors which kills bacterial cells
- increasing phagocytosis
- inducing inflammatory response.

Opsonization: It is the process in which Ig molecules promote uptake of antigen by phagocytic cells. The phagocytes have receptors for Fc portion of Ig, and therefore when an antigen binds an Ig molecule, it is easily recognized by phagocyte (Fig. 5.19).

Binding to high affinity Fc receptors also leads to activation of phagocytes which now display enhanced phagocytic activity.

The stimulation of phagocytosis by an Ig molecule bound to the surface of a foreign particle (antigen) is called opsonization; the Ig is called opsonin. Complement factors can also serve as opsonins.

***Antibody-dependent cell-mediated cytolysis*:** It is the process in which Ig molecules mediate destruction of virus infected cells, or tumour cells, by **killer cells** (Fig. 5.20). The killer cells involved in this process are called the *natural killer cells* (NK cells), or ADCC effector cells. They possess Fc receptors on their cell surface, which can bind with the Fc portion of the antibody.

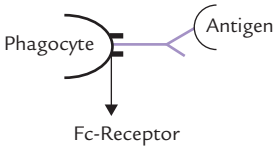


Fig. 5.19. Antibody acting as a bridge between the antigen and the phagocyte.

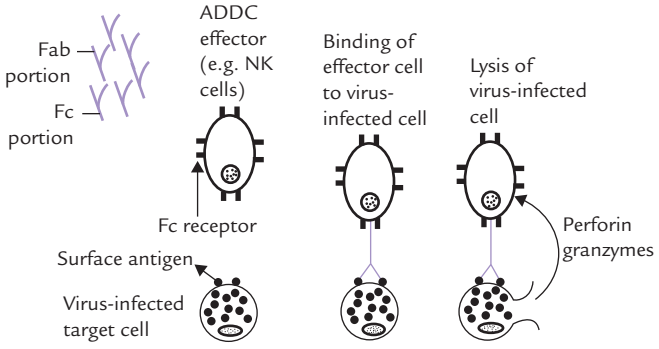


Fig. 5.20. Process of antibody-dependent cell-mediated cytolysis.

The virus-infected cells bind the antibodies through the Fab portion of the latter. The Fc portion of these antibodies bind with the Fc receptors of the NK cells. *Thus antibodies serve as a bridge between the infected target cell and the effector cell*. Lysis of the target cell is then brought about by the secretion of lytic enzymes (from the NK cells) such as the perforins and the powerful *proteases* called **granzymes**.

Classes of Immunoglobulins and their Effector Functions

Properties of various classes of immunoglobulins and their distinct effector functions are described here.

IgG

It is the major immunoglobulin that constitutes about 75–80% of the total immunoglobulins in circulation. The subclasses of IgG are: IgG1, IgG2, IgG3, IgG4. They differ in their degree of segmental flexibility and their capability to trigger complement fixation and other effector functions.

IgG1 makes up about 50% of the total IgG. It is the only type of immunoglobulin that can cross placenta and is localized in the vascular and extracellular compartments. IgG can enter the extravascular space and get involved in all the three effector functions (Table 5.5).

IgM

It is the antibody that is produced first when the immune system is challenged. Thus, IgM dominates the primary immune response. IgM, the largest immunoglobulin is a pentamer (Fig. 5.21) that is confined to vascular spaces. It cannot traverse blood vessels due to its large size. The five subunits of the IgM are joined by linkers called J peptides. The IgM serves as receptor on the B lymphocytes and helps in establishing humoral immunity, being the major antibody of primary immune responses. *Natural antibodies* are also IgM in nature. The natural antibodies are produced without any antigenic stimulation. For example, a person having blood group A will have anti-B group antibodies in his circulation.

IgA

IgA is the major immunoglobulin in seromucous secretions, e.g. saliva, tears, sweat, walls of intestine and bronchial mucous. It is abundant in colostrum, the initial secretion from the mother's breast after childbirth. It may occur as a single (monomer) or a double unit (dimer).

IgA provides surface immunity by binding with antigen on outer epithelial surfaces, and eliminating them. In this way it prevents entry of foreign antigen into the body.

Table 5.5. Functions of immunoglobulins

Class of Ig	Function
IgG	Crosses placenta to provide immunity to newborn.
	Acts as opsonin, i.e. binds to Fc-receptors on macrophages, monocytes and neutrophils and enhances the phagocytosis of the bound antigen.
	Potent activator of complement system.
	Is the only class of Ig that is active in mediating antibody-dependent cell-mediated cytolysis.
	Easily diffuses into extracellular spaces, so plays a major role in neutralizing toxins.
IgM	Most effective activator of classical complement pathway due to multiple binding sites.
	Acts as opsonin.
	Present on the B-cell surface and acts as a receptor for antigen binding.
	Especially potent against multivalent bacterial agents, which are agglutinated.
IgA	Present in breast milk, so provides immunological protection to infant's gut.
	Binds with antigen on epithelial cell surfaces and prevents their entry.
IgE	Responsible for immediate hypersensitivity reaction or allergic reaction.
IgD	Involved in B-cell activation: present on surface of B-cell and (along with IgM) acts as receptor for antigen binding.

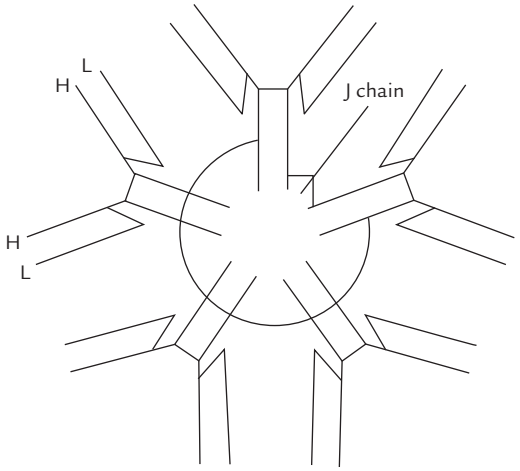


Fig. 5.21. IgM is present as a pentamer of molecular weight 90,000 in serum. IgM on the surface of B-cells, however, is present in monomeric form.

IgE

IgE is a Y-shaped monomer, plasma concentration of which is lowest among all the five immunoglobulin classes. It is a Y-shaped monomer, connected by the tail to the cell membrane of mast cells and basophils. Upon binding to an antigen, IgE stimulates degranulation of these cells and release of inflammatory substances (i.e. histamine). Its concentration increases in response to allergic reactions. Probably it has a role in immunity to helminths and is associated with immediate hypersensitivity.

IgD

IgD is present on the surface of B-lymphocytes but its role is not clear. Probably it is involved in the antigen recognition process.

In addition to the five major classes of immunoglobulins just described, subclasses within each class also exist. They differ from one another by slight variations in the sequence of amino acids in the heavy chain. The subclasses of IgG (IgG1, IgG2, IgG3 and IgG4) correspond to four different types of heavy chains (γ_1, γ_2, γ_3 and γ_4). IgM and IgA have two subclasses each.

D. Isotypes, Allotypes and Idiotypes

Isotypes are variants in molecules seen in all normal persons such as classes and subclasses of immunoglobulins (IgG, IgA, IgM, etc.). They depend on the heavy chain make up; for example, IgG possesses the γ heavy chain isotype, IgM has μ, IgA has α, and so on (Table 5.4).

Allotypes are variations within a subclass; for example, Gma^+ and Gma^- are the allotypes within the IgG class. In people with Gma^+, IgG1 molecule has an amino acid sequence, Asp-Glu-Leu Thr. The corresponding sequence in Gma^- people is Met-Glu-Glu-Thr. These markers show Mendelian recessive hereditary character.

Idiotypes: Unique sequence of amino acids is present in hypervariable regions of an antibody, and is individually specific for each antibody molecule. Such sequence variation is known as *idiotypic variation* and each individual determinant is called **idiotype** (or idiotypic determinant).

The idiotypic determinants are located in the variable part of an antibody (V_L and V_H). If three classes of antibodies (IgG, IgM and IgA) are raised against the same antigen, they will have identical variable domains, both in the light chain (V_L) and the heavy chain (V_H). These three antibodies will share an idiotype. Idiotypic variation is specific for each antibody molecule, as mentioned earlier.

E. Diseases Involving Antibodies

The diseases associated with antibodies are divided into two types. They are caused by:

- Underproduction of antibody (causing hypo- or agammaglobulinaemia).
- Overproduction of antibody (causing hypergammaglobulinaemia).

Hypogammaglobulinaemia

Various inherited disorders are known in which underproduction of immunoglobulins occurs. Most severe of these is **agammaglobulinaemia** a rare, X-chromosome associated disorder, in which there is total absence of immunoglobulins. The affected children are highly susceptible to the bacterial infections but react normally to the viral infections. In some other milder forms, termed **hypogammaglobulinaemia**, the underproduction may be restricted to a single class of immunoglobulins.

Hypogammaglobulinaemia may also be an acquired disease. It commonly occurs in hematological malignancies, such as chronic lymphatic leukemia, multiple myeloma and Hodgkin's disease. It can arise as a complication following the use of cytotoxic drugs or as a symptom of severe protein losing state such as the nephrotic syndrome.

Hypogammaglobulinaemia may arise in some physiological states also. At the time of birth, plasma concentration of all classes of immunoglobulins, except IgG, are low. The IgG molecules, present in relatively higher concentration, are of maternal origin that crossed placenta in the last trimester of pregnancy. The physiological hypogammaglobulinaemia in infants partially accounts for their increased susceptibility to infections.

Hypergammaglobulinaemia

Increased plasma immunoglobulin levels occur in various auto-immune disorders, such as rheumatoid arthritis, systemic lupus erythematosus (SLE), and in some liver diseases that have auto-immune basis. In auto-immune diseases, body rejects its own proteins which became antigenic. Probably, buried antigen sites in endogenous proteins (***epitopes***) get exposed; or certain exogenous triggers bind with "self-proteins" and make them antigenic. An immune response, involving production of antibodies against these proteins, is aroused. This leads to the destruction of self-proteins.

Markedly increased immunoglobulin levels occur in the patients suffering from myelomatosis (immunocytoma tumours). The condition is known as **multiple myeloma**, as discussed earlier. The immunoglobulins are produced from a single clone of cells of β-lymphocyte series, most frequently plasma cells. Due to common origin these immunoglobulins are identical and are known as **paraproteins**. They are mostly **monoclonal IgG**, but may also be IgA or IgD. They are seen as a single discreet M band on an electrophoretogram, usually in the gamma region. In the urine of these patients, massive excretion of *Bence-Jones proteins* occurs. The latter are identified as low molecular weight proteins (light chains). For clinical description of an immunoglobulin disorder (see **Case 5.4**).

The diseases involving antibodies are discussed further in Chapter 33.

The plasma of multiple myeloma (a malignant disease of the plasma cells) patients shows a distinct M-band, and their urine shows presence of Bence-Jones proteins. The latter are overproduced light chains of either the κ or the λ type.

Exercises

Essay type questions

1. Write briefly on plasma proteins and their functions. Discuss methods for their separation and explain why albumin deficiency causes oedema.
2. Draw electrophoretogram of plasma proteins in a normal subject and compare it with that of a multiple myeloma patient.
3. Describe the characteristics, structural features and functions of different immunoglobulins.
4. What are glycoproteins? Discuss their structural features and biomedical importance.
5. Explain the differences between N- and O-linked oligosaccharides in glycoproteins. How do the physical properties of glycoproteins relate to their biological roles.

Write short notes on

1. Electrophoresis of plasma proteins
2. Multiple myeloma
3. Opsonization
4. Bence-Jones proteins
5. Haemopexin
6. Elastin
7. Osteogenesis imperfecta
8. Electrophoresis in multiple myeloma
9. Disorders of collagen metabolism
10. Functions of albumin

CLINICAL CASES

CASE 5.1 A 3-year girl with recurrent fractures of long bones

A 3-year old girl was referred to the orthopedics emergency with multiple fractures. History revealed that ever since birth, she frequently sustained fractures of long bones following even minor injuries. On examination, musculoskeletal abnormalities were observed, the most prominent of them being loose and hypermobile joints. In addition, the child had hearing loss, blue sclera, and discolored bell-shaped teeth. X-ray revealed decreased bone density, suggestive of impaired mineralization. A collagen tissue disorder was suspected after it was ascertained that there was a positive family history. Fibroblasts were cultured and collagenous tissue was obtained from them. A detailed examination of the polypeptide chains of the tropocollagen showed a defect in the α_1 (I)-chain of the type I collagen. Analysis of the primary structure of the α_1 (I)-chain showed that α-cysteine residue was present at a certain position, instead of glycine of the normal chain.

Q.1. How does the above mutation result in brittle bones and skeletal abnormalities?

Q.2. How are the other symptoms of the patient accounted by the defect in the bones?

Q.3. Type I collagen was obtained from cultured fibroblasts. It was incubated with radioactive amino acids, and further investigated by performing polyacrylamide gel electrophoresis (PAGE). Some of the α-chains showed decreased mobility.

(a) Provide an explanation for the decreased motility.

(b) Why does mobility of only some of the α-chains decrease whereas that of the others is normal?

(c) Addition of mercaptoethanol results in increasing mobility of the α_1 (I)-chain that had retarded mobility earlier. Explain why.

CASE 5.2 A 4-year-old boy with fragile skin and loose joints

A 4-year-old child had loose, hypermobile joints, and soft, hyperflexible skin. His fingers could be extended to such an extent that they touched back of his hand. The skin was easily iron or scarred. In addition, he had scoliosis (bent spine). Sample of collagen tissue was obtained by biopsy from two sources: skin and cartilage. The skin collagen was deficient in hydroxylysyl residues, whereas the collagen obtained from cartilage had normal content of these residues. The patient did not have any features that were suggestive of vascular collagen involvement.

Q.1. Identify the biochemical lesion in this patient.

Q.2. How does the deficiency of hydroxylysyl residue result in reduced strength of collagen?

Q.3. Name the vitamin cofactor for the missing enzyme. Is it necessary to provide this vitamin in diet? Give reasons.

CASE 5.3 A 5-year-old child with ulcerative skin lesions

A 5-year-old child having chronic ulcerative dermatitis, was brought to the hospital OPD. Examination showed that he was mentally retarded. Urine chromatography revealed presence of dipeptides consisting of proline and glycine (i.e. glycylproline). The child had decreased ability to cleave this dipeptide as evidenced by the following experiment: The fibroblasts were isolated by biopsy, cultured, and a cell-free extract was prepared from them. Glycylproline was then added to the extract. Only a small fraction of the added dipeptide was cleaved to glycine and proline. In contrast, cell-free extract obtained from fibroblasts of a normal subject could cause complete degradation of this dipeptide.

Q.1. Identify the biochemical defect in this child.

Q.2. Delineate the cause of the ulcerative skin lesions observed.

CASE 5.4 A 62-year-old man with back pain, hypercalcaemia and impaired renal functions

A 62-year-old man presented with continuous aching pain in lower back and tiredness. He had been well until three months back when he noticed decrease in his stamina: he felt very weak in course of his everyday morning walk. Since then he had lost about 3 kg weight. He was a non-smoker, but had been suffering from recurrent episodes of respiratory tract infections for the last two years. Presently he felt short of breath even after slight exertion. On examination, there was nothing remarkable. He was normotensive, and his body systems were apparently normal. However, he showed pallor and had a number of tender spots on bones. A particularly tender spot was found on the spine.

Blood sample was obtained for biochemical analysis. The results are as below:

Investigations test	Patient's reports	Reference range
Haemoglobin	7.8 g/dl	12–16 g/dl
S. Calcium	12.6 mg/dl	8.5–10.5 mg/dl
S. Urea	70 mg/dl	15–45 mg/dl
S. Creatinine	2.1 mg/dl	0.6–1.4 mg/dl
S. Uric acid	10.4 mg/dl	2.7–7.0 mg/dl
S. Sodium	130 mmol/L	132–144 mmol/L
S. Protein	8.8 g/dl	6.2–8.2 g/dl
S. Albumin	2.7 g/dl	3.6–5.4 mg/dl
Alkaline phosphatase	60 U/L	40–100 U/L
ESR (in first hour)	>100	10.6

A blood film showed normochromic, normocytic anaemia; rouleaux were present on the blood film and there was increased background standing. Tentative diagnosis was made based on these findings. Plasma electrophoresis was carried out to confirm the tentative diagnosis: it showed a thick band in the γ-region, which was reported to be paraprotein of IgG class.

Q.1. What is the probable diagnosis? Explain the patient's signs and symptoms on the basis of the diagnosis.
Q.2. Suggest further tests to confirm the clinical diagnosis?
Q.3. Why are the parameters of renal function affected in this condition?
Q.4. What are Bence-Jones proteins? How are they detected by heating test?
Q.5. What are the prognostic indicators in this condition? Outline the treatment.

CHAPTER

ENZYMES

6

Body proteins perform a number of functions, of which the most remarkable and highly specialized function is their ability to act as biological catalysts. Catalytic proteins called enzymes, stimulate the rate of several reactions in the body without themselves getting consumed in the process. As *biological catalysts*, the enzymes, exhibit certain special characteristics which are rarely, if ever, exhibited by non-biological catalysts. These are:

1. The *catalytic power* of enzymes is *extraordinary*: they make occurrence of highly complex reactions possible under mild conditions of temperature and pH prevailing in the body.
2. Enzymes have a high degree of *specificity* for the substrates. Most of the enzymes use only one substrate, although some enzymes have broader specificity, with a capacity to act on more than one substrate.
3. Enzymes exhibit specificity for the reactions also. They accelerate specific chemical reactions without formation of any by-products.

Enzymes act as functional units of the cellular metabolism. They catalyze innumerable body reactions whereby the nutrient molecules are degraded, chemical energy is conserved and transformed, and complex biomolecules are synthesized from simple precursors. Virtually, all the body reactions are mediated by enzymes. In fact, cellular reactions would not occur rapidly enough to sustain life if enzymes were not present.

This chapter outlines the properties and characteristics common to most enzymes. After going through this chapter, the student should be able to understand:

- Concept of organic catalysts, the catalytic site and enzyme-substrate complex; role of cofactors and coenzymes in enzymatic catalysis.
- Nomenclature and classification of enzymes; mechanism of enzyme action; effect of various factors like temperature, pH, enzyme concentration, etc. on enzyme activity; qualification of enzyme activity.
- Michaelis-Menten theory of enzyme action.
- Various types of enzyme inhibitions and their recognition; enzyme regulation through modulation of activity or by induction/repression.
- Utility of enzyme and isoenzyme determination as a diagnostic tool.

I. Properties of Enzymes

A. Chemical Nature

Most known enzymes are proteins in nature. The typical enzyme is a **globular protein** that either is dissolved in cytoplasm or in a cellular organelle; or is bound to a membrane, or is located extracellularly in interstitial space or vascular space; or, as in case of the digestive enzymes, in an external compartment. The catalytic activity of an enzyme depends on integrity of the protein structure. This is evidenced by the fact that various factors which disrupt the native conformation of the

enzyme polypeptide chain(s) cause loss of the enzyme catalytic activity.

Enzyme, almost all of which are proteins, are highly selective biocatalysts that accelerate reactions by factors up to 10^{15}.

Some type of RNAs can also serve as enzymes. Such RNAs with catalytic activity are called the **ribozymes**. Most of them play specialized roles in gene expression (catalyze cleavage and synthesis of phosphodiester bonds), but are not involved in metabolic reactions.

B. Cofactors

Some enzymes are proteins, containing no chemical group other than amino acid residues (e.g. *pancreatic ribonuclease trypsin, chymotrypsin, elastase*). However, *most enzymes associate with non-protein chemical components, called* ***cofactors***, which are required for their catalytic activity. The complete active enzyme is called a **holoenzyme**, and it is made of the protein portion (**apoenzyme**) and the cofactor. The apoenzyme is inactive alone.

$$\text{Holoenzyme} \rightleftharpoons \text{Apoenzyme} + \text{Cofactor}$$

Cofactors may be:

- inorganic metal ions (Table 6.1) or
- organic molecules, called **coenzymes**

Many coenzymes are derived from the dietary, water-soluble vitamins. An example is flavin adenine dinucleotide (FAD) or flavin mononucleotide (FMN), which are derived from riboflavin (vitamin B_2). In general, a coenzyme actually participates in the overall reaction as another substrate, and mainly acts as a donor or acceptor of a particular chemical group. Specific coenzymes are concerned with transfer of specific groups, as shown in Table 6.2.

The coenzymes function as either co-substrates or prosthetic groups.

1. ***Co-substrates:*** The co-substrate associates with the active site of the enzyme only transiently for the purpose of the reaction. It is changed chemically during the reaction, and after completion of the reaction, the chemically modified co-substrate dissociates away and is free to participate in other enzymatic reaction cycle. For example, *pyridoxal phosphate* (a co-substrate in transamination reaction) brings about transfer of amino group from an amino acid to a keto acid by reacting with and serving as transient-carrier of the transferred amino group (Chapter 13).

Table 6.1. Metal ions as cofactors for specific enzymes (also called metalloenzymes)

Metal	Enzyme
Ca^{2+}	*Lipase*
Cu^{2+}/Cu^{+}	*Cytochrome oxidase, tyrosinase, lysyl/oxidase, superoxide dismutase*
Fe^{2+}/Fe^{3+}	*Cytochrome oxidase, xanthine oxidase peroxidase, catalase*
K^{+}	*Pyruvate kinase*
Mg^{2+}	*Hexokinase, phosphofructokinase, enolase, creatine kinase*
Mn^{2+}	*Arginase, glycosyl transferase, phosphoglucomutase*
Ni^{2+}	*Urease*
Zn^{2+}	*Carbonic anhydrase, alkaline phosphatase DNA polymerase, alcohol dehydrogenase*

Table 6.2. Coenzymes and the entity transferred

Coenzyme	Entity transferred
Biotin	Carbon dioxide
Coenzyme A	Acyl group
FAD	Hydrogen atoms
NAD^{+}	Hydride ion (H^{-})
Pyridoxal phosphate	Amino groups
Tetrahydrofolate	One carbon groups (other than CO_2)
Thiamine pyrophosphate	Hydroxy-ethyl
Coenzyme B_{12}	Alkyl groups or hydrogen atoms

FAD = flavin adenine dinucleotide, NAD^{+} = nicotinamide adenine dinucleotide.

Some enzymes require assistance of metal ion cofactors or organic coenzymes derived from vitamins. The coenzymes function as reversibly bound co-substrates, or as permanently associated prosthetic groups.

2. ***Prosthetic groups:*** The prosthetic group on the other hand is *bound permanently* to the active site of its enzyme, either covalently or non-covalently. Dissociation of the prosthetic group results in an irreversible loss of catalytic activity of the enzyme. For example, biotin, is the prosthetic group for a group of enzymes called *carboxylase.* It is an integral component of these enzymes, being attached by an amide linkage with the apoenzyme component (i.e. *apocarboxylase*). Dissociation of biotin from the *apocarboxylase* results in loss of activity of the *carboxylase* enzymes **(Case 6.1).**

The distinction between co-substrate and prosthetic group is, however, not absolute because a compound that

serves as a coenzyme for one enzyme can serve as a prosthetic group for another enzyme.

The name "coenzyme" is applied to two different types of cofactors that function either as co-substrates or as true prosthetic groups. Interestingly, behaviour of co-substrate sound "promiscuous" since it associates with the active site of the enzyme only transiently. Prosthetic group sounds monogamous: it is tied-up with enzyme permanently.

Metalloenzymes: The enzymes that require inorganic metal ions as their cofactors are termed metalloenzymes. Approximately two-third of the known enzymes are metalloenzymes. In most cases, the binding of metal ion is essentially required for the enzymatic activity (e.g. zinc for *carbonic anhydrase*), whereas in some the enzyme is active even without the metal ion, but its activity increases greatly when the metal ion is added (e.g. chloride and fluoride ions activate *salivary amylase*).

C. Catalytic Efficiency

Catalytic efficiency of enzymes is very high—most of them bring about several-fold (10^3 to 10^{15} times) increase in the reaction rate. Typically, each enzyme molecule is capable of transforming 100 to 1000 substrate molecules into the corresponding product molecules each second. *Carbonic anhydrase*, the enzyme catalyzing formation of carbonic acid from water and carbon dioxide, is highly active; a single enzyme molecule can transform 36,000,000 substrate molecules each second.

D. Enzyme Specificity: The Active Site

Enzymes possess most exquisite specificity of all the catalysts known. They are *highly selective for both the substrate and the reaction undergone by the substrate.* Typically, each individual reaction requires its own enzyme, and if an enzyme is lacking, only one particular reaction is generally blocked. Many enzymes show **absolute specificity** (e.g. *tyrosinase* acting on tyrosine only) for the substrate, and this includes *stereo-specificity* also. For example, *D-amino acid oxidase* binds only D-amino acids but not L-amino acids; *fumarase* hydrates fumarate to form malate, but the corresponding *cis* form of the former, maleic acid, is not acted upon. Other enzymes have **relatively broad specificity** and act on several different analogs of a specific substrate; for example, glucose, mannose, fructose, glucosamine and 2-deoxyglucose are phosphorylated by *hexokinase,* but at different rates. (*Glucokinase,* on the other hand, is specific only for glucose.)

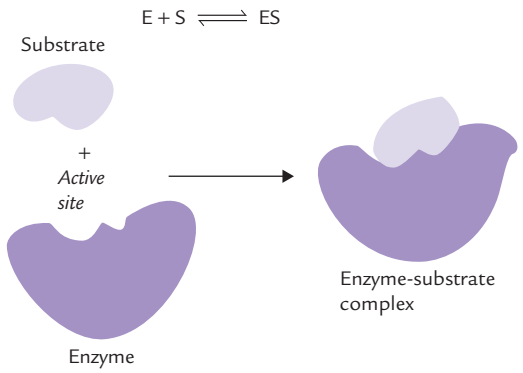


Fig. 6.1. Active site of an enzyme and enzyme-substrate complex (E = enzyme, ES = enzyme-substrate complex, S = substrate).

The enzyme specificity resides in a relatively small region of the enzyme, called the **active site.** It is generally a crevice, pocket or cleft that is three dimensional, having the correct molecular dimensions and appropriate topology to accommodate specific substrate (Fig. 6.1; see also Box 6.1). The active site is formed by collection of different amino acid residues (termed active-site residues) that may or may not be adjacent in the primary sequence. The interactions between these residues and the substrate occur via the same forces that stabilize protein structure: hydrophobic interactions, electrostatic interactions, hydrogen bonding and van der Waals interactions. The active site residues are directly responsible for the catalytic action. They not only bind the substrate to form enzyme-substrate (ES) complex, but also provide specific interactions that stabilize the formation of the transition state (highest energy arrangement of atoms) for the chemical reaction to proceed and to form the reaction product.

Enzymes exhibit reaction specificity and substrate specificity, which are determined by the geometric and electronic character of the active site where the substrate binds and undergoes transformation.

E. Location within the Cell

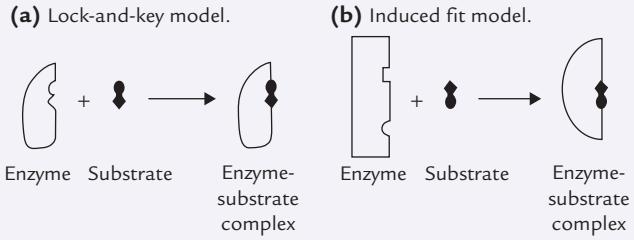
Most enzymes are compartmentalized in specific organelles within the cell; for example, enzymes that catalyze degradation of fatty acids (i.e. β-oxidation) are located in the mitochondrial matrix, whereas those involved in synthesis of fatty acids are present in the cytosol. Such arrangement helps to:

- organize thousands of enzymes of the cell in distinct pathways,
- provide favourable environment for cellular reactions,
- Isolate the substrate or product of a given reaction from other competing reactions.

BOX 6.1

Active Site Events: Lock-and-key and Induced Fit Modules

The lock-and-key model proposed by Fischer in 1894 suggests that the shapes of the catalytic site of the enzyme and its substrate are complementary to each other. The catalytic site forms a rigid pocket which permits exact fitting of the substrate into it in the manner of a key fitting into a lock. The substrate must be the right size and shape, have charges in the correct place, have the right hydrogen bond donors and acceptors and have just the right hydrophobic patches. Such rigidity is negated by Koshland's induced fit theory which proposes that binding of the substrate molecule may induce a conformational change in the enzyme molecule. This strains structure of the active site and also distorts the bound substrate in such way that the functional groups of the enzyme protein that participate in catalysis are brought to the substrate. Thus, the conformational changes help formation of a complex in which the substrate and catalytic groups on the enzyme are properly arranged to accelerate the reaction.



II. Mechanism of Enzyme Catalysis

The question, "how do enzymes work" can be answered from two different perspectives. The first one describes the energy changes that occur during the catalyzed and the uncatalyzed reactions. The second deals with changes in the enzyme active site which facilitates catalysis.

A. Free Energy Changes During Chemical Reactions

For any chemical reaction to take place, the free energy content of the products should be lower than that of the substrates. In these cases, the free energy change of the reaction (ΔG) has a negative sign and the reaction is said to be thermodynamically favourable (Chapter 8). However, not all reactions that are thermodynamically favourable occur spontaneously; the substrate has to be converted first to a higher energy form (called **transition state**) before it can form the reaction product. The transition state is structurally an intermediate between the substrate and the product, and represents the highest energy arrangement of atoms. Therefore, it is unstable; once formed, it decomposes almost immediately to form the reaction product (or sometimes the substrate again).

In the *reaction coordinate diagram* shown in Figure 6.2, the reaction is thermodynamically favourable but cannot occur spontaneously because the substrate has to be first converted to the transition state. Some energy must be put into

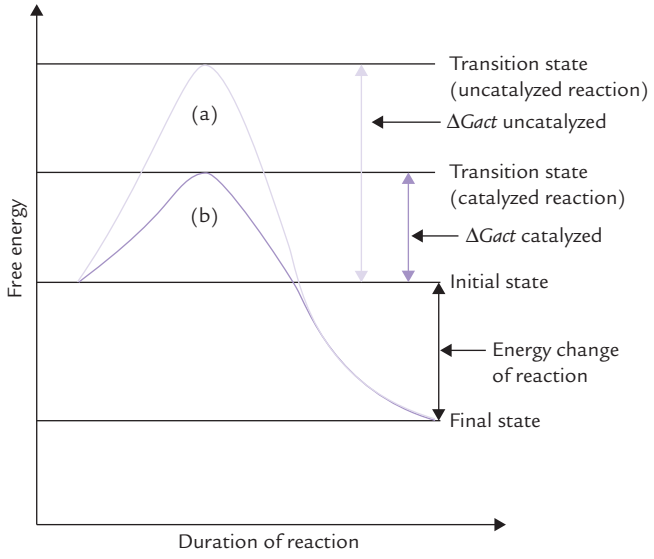


Fig. 6.2. Diagrammatic representation of the free energy changes that occur during a chemical reaction. **(a)** Uncatalyzed, **(b)** Catalyzed, Δ*Gact* = free energy of activation. Enzymes speed up a chemical reaction by lowering the magnitude of activation energy barrier without affecting equilibrium of the reaction.

the substrate for this conversion to the transition state. It is termed the ***free energy of activation (ΔGact)***, and it equals the free energy difference between the substrate and the transition state. The Δ*Gact* provides an energy barrier to the reaction—higher the barrier, the slower the reaction. The *enzymes speed up the chemical reaction by lowering the magnitude of the activation energy barrier.* This results in decreasing the Δ*Gact*. Lower Δ*Gact* means increased concentration of

the transition state molecules, so that the rate of catalyzed reaction increases. The role of enzyme can thus be compared to a tunnel that bypasses a mountain peak.

Note: Enzymes do not affect the free energy content of the reactants or the products. As a result, *enzyme catalysis does not affect equilibrium of reaction* and the energy change of the reaction remains the same in the catalyzed (Fig. 6.2b) and uncatalyzed reactions.

Enzymes can change only the reaction rate, not the equilibrium of the catalyzed reaction. They act by decreasing the activation free energy ΔG_{act}—the free energy required to reach the transition state (point of highest free energy in the reaction).

B. Changes in Enzyme Active Site

The active site is not a passive receptacle wherein the substrate molecule just fits in. A variety of complex changes occur within its vicinity, which facilitate the substrate transformation. Some of these changes are discussed.

Transition State Stabilization

The active site binds the substrate in a geometry that resembles the transition state of the substrate. This helps stabilization of the substrate molecule in its transition state, which being a reactive form, undergoes conversion into the product. This effect is sometimes referred to as **substrate-strain** because the enzyme effectively bends the substrate into the shape of the transition state.

Acid-base Catalysis

The aminoacyl residues of the active site may provide (or sometimes remove) certain chemical groups to the substrate, that enhance its conversion to the transition state. In most cases, the protons donated or accepted by some ionizable group on the enzyme help the formation of the transition state. In other cases, electron pair donors or acceptors, known respectively as *Lewis bases* and *Lewis acids*, participate in the reaction.

General acid-base catalysis is the most important reason for the pH-dependence of enzymatic reactions, as discussed later.

Covalent Catalysis

It is based on formation of transient covalent bonds (between amino acyl residues of the active site and the substrate molecule) during catalysis. Covalent catalysis, for example, is used by *serine proteases* which bind their substrate covalently at the serine side chain.

It is important to note at this stage that any favourable interaction—covalent or non-covalent—between the enzyme and the substrate decreases the free energy of activation and increases the reaction rate.

Proximity Effect

Proximity effect, also known as **entropy effect**, is based on approximation of the substrates in two-substrate reaction. In absence of the enzyme, a reaction takes place only when two substrate molecules collide in desired geometrical orientation, and with sufficient kinetic energy. This is certainly a rare event, with extremely low probability of occurrence. The enzymes, however, binds both the substrates to its active site in the correct geometric orientation, thereby increasing the frequency of productive collisions.

The catalytic mechanisms employed by enzymes are four: transient state stabilization, covalent catalysis, general acid-base catalysis, and proximity effect.

III. Nomenclature

Enzymes were assigned various trivial names as they were discovered, which are convenient to use. For example, *sucrase*, *lactase* and *urease* are the commonly used names for the enzymes that catalyze hydrolysis of sucrose, lactose and urea, respectively. These enzymes have been named by attaching the suffix-ase to the respective substrates. However, the nomenclature may lead to ambiguities; for example, *fumarase*, a citric acid cycle enzyme, does not catalyze hydrolysis of its substrate, fumarate. Rather, it catalyzes addition of a water molecule to fumarate.

To remove ambiguities associated with use of such trivial names and to establish a uniform system for enzyme nomenclature and classification, the International Union of Biochemistry and Molecular Biology **(IUBMB)** proposed a system in 1964 (modified in 1972 and 1978). In this system, enzymes are grouped into **six major classes**, based on the reaction that they catalyze. These are *oxidoreductase, transferase, hydrolase, lyase, isomerase* and *ligase* (Table 6.3).

Classification

1. ***Oxidoreductases*** are involved in the oxidation-reduction reactions. These include electron transfers, hydrogen transfers, and reactions in which oxygen participates. The most common *oxidoreductases* are *dehydrogenases;* others are *oxygenases, hydroxylases, peroxidases, oxidases* and *reductases*.

Table 6.3. Enzyme classification

Class	Subclass	Action
I *Oxidoreductases*	*Dehydrogenases, oxidases, peroxidases, catalase, oxygenases, hydroxylases, reductases*	Transfer of electrons ($A^- + B \rightarrow A + B^-$)
II *Transferases*	*Transaldolase, transketolase, kinases,* etc.	Transfer of functional groups ($A-B+C \rightarrow A+B-C$)
III *Hydrolases*	*Esterases, glycosidases, peptidases, phosphatases, thiolases, amidases,* etc.	Cleave a bond with water addition ($A-B+H_2O \rightarrow A-H+B-OH$)
IV *Lyases*	*Decarboxylases, aldolases, hydratases, dehydratases synthases*	Cleave without water, often forming a double bond ($A-B \rightarrow A=B+Z-W$)
V *Isomerases*	*Racemases, epimerases, mutases*	Interconvert isomeric structures ($A-B \rightarrow A-B$)
VI *Ligases*	*Synthetases, carboxylases*	ATP-dependent condensations ($A+B \rightarrow A-B$)

Dehydrogenases transfer hydrogen between a substrate and a coenzyme, most commonly NAD^+, $NADP^+$, FAD, or FMN:

$$AH_2 + \text{Coenzyme} \xrightarrow{\textit{Dehydrogenase}} A + \text{Coenzyme (reduced)}$$

Example:

$$\text{Malate} + NAD^+ \xrightleftharpoons{\textit{Malate dehydrogenase}} \text{Oxaloacetate} + NADH + H^+$$

Oxygenases use molecular oxygen as a substrate. *Monooxygenases* use one oxygen atom to oxidize their substrate and the other one to oxidize a reduced coenzyme; the *dioxygenases* use both oxygen atoms of the oxygen to oxidize their substrate. Most *hydroxylases are monooxygenases:*

$$A\text{–}H + O_2 + NADPH + H^+ \xrightarrow{\textit{Hydroxylase}} A\text{–}OH + NADP^+ + H_2O$$

Peroxidases use hydrogen peroxide or an organic peroxide as one of their substrates. *Catalase* is also, technically speaking, a *peroxidase*; it acts on a hydrogen peroxide molecule to form water and molecular oxygen:

$$H_2O_2 \xrightarrow{\textit{Catalase}} H_2O + \tfrac{1}{2}O_2$$

2. ***Transferases*** catalyze the transfer of functional group either from a donor molecule to an acceptor or within a single molecule. The ***kinases***, which transfer phosphate from ATP (or some other nucleoside triphosphate) to a second substrate, are typical examples:

$$A + ATP \xrightarrow{\textit{Kinase}} A\text{–}P + ADP$$

Example:

$$\text{Glucose} + ATP \rightarrow \text{Glucose 6-phosphate} + ADP$$

Name of this enzyme is *glucokinase*.

Other examples of *transferases* include *aminoacyl-, methyl-, glucosyl-,* and *phosphoryl-transferases, transaldolases,* and *transketolases*. A *transferases* that transfer a functional group within a single molecule is termed **mutase**.

$$\text{Glucose 6-p} \xleftrightarrow{\textit{Mutase}} \text{Glucose 1-p}$$

In all, there are eight subclasses among *transferases*.

3. ***Hydrolases*** make use of water to cleave a single molecule into two molecules. They are grouped according to their substrates, such as *esterases, glycosidases, lipases, peptidases, thiolases, phospholipases, amidases, ribonuclease* and *deoxyribonuclease*. The digestive enzymes and lysosomal enzymes are *hydrolases*. The cleavage specificities of those hydrolytic enzymes that act on polymeric substrates are indicated by the prefixes exo- (Greek, "outside") for they cleave only the bond next to the end; and endo- ("inside") which cleave a bond even at sites far removed from the ends.

4. ***Lyases*** split the substrate molecule by a non-hydrolytic process, mostly leaving a double bond; examples are the *dehydratases, aldolases* and *decarboxylases*. Many *lyase* reactions proceed in opposite directions, obliterating a double bond in one of the substrates, e.g. *fumarase*, which converts fumarate to malate.

$$\underset{\text{Fumarate}}{COOH\text{–}CH=CH\text{–}COOH} + H_2O \longrightarrow \underset{\text{Malate}}{COOH\text{–}CH_2\text{–}CH_2\text{–}COOH}$$

Some other *lyase* reactions create a new bond. The latter are termed *synthases* (synthetic enzymes, ATP-independent), for example, *citrate synthase*, which establishes a bond between acetyl CoA (2C) and oxaloacetate (4C) to form citrate (6C).

$$\text{Acetyl CoA} + \text{Oxaloacetate} \rightarrow \text{Citrate} + \text{CoASH}$$

5. ***Isomerases*** interconvert positional, geometric, or optical isomeric structures by intramolecular rearrangements. They include *racemases, epimerases, cis-* and *trans-isomerases* and *intramolecular transferases* (*mutases*).

BOX 6.2

Synthases and _Synthetases_ are Classified Differently

It is important to note that *synthases* and *synthetases* are both *biosynthetic* enzymes, but they differ from each other in an important aspect. Synthesis by *synthetase* is coupled with hydrolysis of a phosphoanhydride bond in ATP or some other energy rich compound; examples include *fatty acyl CoA synthetase, glutamine synthetase, amino acyl tRNA synthetase*, and *DNA ligase*. *Synthases* on the other hand carry out synthesis independent of ATP; for example, *citrate synthase, ALA synthase* and *glycogen synthase*. They are classified as *lyases*, whereas *synthatases* are placed in class VI (*ligases*).

For example, the enzyme *methylmalonyl CoA mutase* catalyzes conversion of *methylmalonyl CoA* to succinyl CoA.

$$\text{Methylmalonyl CoA} \xrightarrow[\text{CoA mutase}]{\text{Methylmalonyl}} \text{Succinyl CoA}$$

6. *Ligases* join two separate molecules by creating a new chemical bond at the expense of nucleoside triphosphate (e.g. ATP). They include enzymes that form C–O, C–S, C–N, and C–C bonds. Examples include *synthetases* and *carboxylases* (see Box 6.2).

$$\text{Fatty acid} + \text{ATP} + \text{CoASH} \xrightarrow[\text{synthetase}]{\text{Fatty acyl CoA}} \text{Fatty acyl CoA} + \text{AMP} + 2\text{Pi}$$

$$\text{Pyruvate} + \text{ATP} + CO_2 + H_2O \xrightarrow[\text{carboxylase}]{\text{Pyruvate}} \text{Oxaloacetate} + \text{ADP} + \text{Pi}$$

The classification of the enzymes is summarized in Table 6.3.

Enzymes are classified into six major classes on the basis of their reaction type: *oxido-reductases, transferases, hydrolases, lyases, isomerases* and *ligases*.

Each major class is divided into subclasses, each of which has numerous sub-subclasses. Every class, subclass, and sub-subclass is assigned a number.

Individual enzymes are assigned a *four-digit number*. The first three digits correspond to the class, subclass and sub-subclass, respectively and the fourth digit is the serial number of the enzyme in its sub-subclass. For instance, the enzyme *hexokinase* (HK) catalyzing the reaction shown below is assigned the classification number ***EC 2.7.1.1***

$$\text{ATP} + \text{D-Glucose} \xrightarrow{HK} \text{D-Glucose 6-phosphate} + \text{ADP}$$

EC stands for enzyme commission, and

(a) the first digit (2) stands for the class name (*transferase*),
(b) the second digit (7) stands for the subclass (*phosphotransferase*), indicating that a phosphate group is transferred,
(c) the third digit (1) is for the sub-subclass (*phosphotransferase* with hydroxyl group as an acceptor), and
(d) The fourth digit (1) denotes that the enzyme is ATP: D-glucose-6 *phosphotransferase*.

The IUBMB names can identify a given enzyme without any ambiguity, but are cumbersome. Therefore, the short names of the above enzyme, *hexokinase*, is still widely, used.

IV. Quantitative Assay of Enzymes

A. Catalytic Activity

The amount of an enzyme in a given solution or tissue extract can be assayed in terms of the catalytic effect it produces. In an enzyme-catalyzed reaction, the catalytic activity of the enzyme is reflected by the rate at which the substrate is transformed to the product. Greater the rate of transformation, more the enzyme activity, and vice versa. Rate of transformation may be estimated from the rate of appearance of product and/or the rate of disappearance of the substrate.

B. Measurement of Activity

The activity of the enzyme is measured in terms of the following:

Unit of enzyme activity: By international agreement, one unit enzyme activity is *defined as the amount causing transformation of 1.0 mmole of substrate per minute at 25°C*, under optimum conditions of measurement. It is usually expressed as mmole of substrate disappeared, or mmole of product formed per minute (mole = 10^{-6} μ moles).

Specific activity: It refers to the *number of enzyme units per milligram of protein*. It is a measure of enzyme purity: higher the enzyme purity, more is the specific activity.

Table 6.4. Turnover number of some enzymes (at 25°–37°C)

Enzyme	Turnover number
Carbonic anhydrase	36,000,000
β-Amylase	11,00,000
Phosphoglucomutase	1,240
Fumarase	800

Maximum value is attained when the enzyme is fully purified.

Turnover number of enzyme: This refers to the number of substrate molecules transformed per unit time by a single enzyme molecule (or by a single catalytic site), when the enzyme concentration alone is the rate-limiting factor. For example, *carbonic anhydrase* has a turnover number of 36,000,000. A high turnover number enables this enzymes to catalyze an otherwise slow and unfavourable reaction, that involves reversible hydration of dissolved carbon dioxide to form carbonic acid.

$$CO_2\,(d) + H_2O \xrightarrow{\textit{Carbonic anhydrase}} H_2CO_3$$

Turnover numbers of some of the enzymes are given in Table 6.4.

V. Enzyme Kinetics

Enzymes are the most efficient catalysts known: they increase the rate of a reaction without themselves undergoing a change, and without changing the equilibrium of the catalyzed reaction. Rates of enzymatic reactions are influenced by a variety of factors, such as concentration of substrates, products, inhibitors, temperature and pH levels. Study of these parameters is important because it provides a great deal of information on nature of a particular enzyme and the biochemical pathways involved. This is called **enzyme kinetics**.

Effects of various factors on enzyme activity are determined by the *in vitro* studies. These studies are conducted by isolating the enzyme from the cell, and then estimating its activity when these factors are varied.

A. Effect of Substrate Concentration

Rate of an enzyme-catalyzed reaction increases with increase in the substrate concentration (Fig. 6.3). As the concentration of the substrate molecules is gradually raised, more and more enzyme active sites are occupied by these molecules. Consequently, substrate transformation increases, reflecting enhanced enzyme activity.

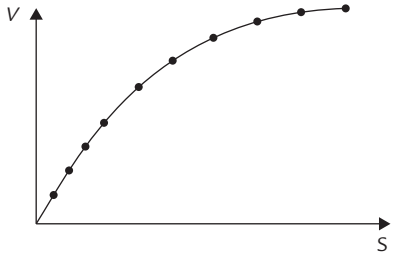


Fig. 6.3. The relationship between substrate concentration [S] and reaction velocity (V).

The rise is enzyme activity continues till all the active sites get occupied by the substrate molecules. This reflects a **saturation point;** further increase in substrate concentration does not increase the enzyme activity any further. The velocity of reaction at this stage is called the **maximum velocity (V_{max})**.

The rise in reaction velocity is not directly proportional to substrate concentration. Rather it shows a **hyperbolic** kinetics, indicating that there is discordance between the substrate concentration and the reaction rate. Two factors account for such discordance:

1. In ***initial stage*** of the reaction, most catalytic sites on enzyme are free and available to bind the substrate. Any substrate molecule that is added quickly finds a free active site to occupy. Thus, the reaction rate increases almost linearly with the substrate concentration. However, as the reaction progresses, more and more active sites get occupied and fewer remain free to bind with the added substrate. As a result, increase in the substrate concentration at this ***later stage*** has less effect on the reaction. Therefore, when the substrate concentration is plotted against the reaction velocity, the curve is initially steep but becomes more flat as the reaction progresses. This gives a **hyperbolic** shape to the curve.
2. The second factor accounting for discordance between the substrate concentration and the reaction velocity is the fact that the reaction occurs in two stages.

Stage I Binding of the substrate to the enzyme active site forms the enzyme-substrate complex (ES).

Stage II The ES complex dissociates to yield the final product (P). Formation of the ES complex is faster than the subsequent transformation to the product.

$$E + S \rightleftharpoons ES \text{ (Stage I, } \textbf{fast}) \quad (1)$$

$$ES \rightleftharpoons P + E \text{ (Stage II, } \textbf{slow}) \quad (2)$$

As a result, any newly added substrate does not undergo conversion to the product immediately, though it rapidly forms an enzymes-substrate complex. A part of it may even dissociate to form the substrate again, by reversal of the first reaction. Thus, the overall reaction velocity does not show a proportionate rise with the substrate concentration.

At lower substrate concentration, the reaction velocity almost shows a proportional rise with the substrate concentration. At higher substrate concentration, the reaction velocity does not increase as proportionately, but rather approaches a (fixed) maximum reaction velocity.

Note: In case of allosteric enzymes, a sigmoid curve is obtained when the reaction velocity is plotted against the substrate concentration.

B. Effect of Temperature

Increase in temperature increases the energy content of the reactant molecules, so that more of them are able to cross the activation energy barrier and form the reaction products. As a result, the reaction velocity rises. Increase in the reaction velocity continues till it reaches a peak (or maximum) value. The temperature at which the reaction velocity is maximum is called the **optimum temperature** (Fig. 6.4).

Further increase in temperature beyond the optimum temperature causes a fall in the reaction velocity. This is because at high temperature (typically somewhere between 42°C and 70°C) the enzymes get irreversibly denaturated. *This explains why an optimum curve is obtained under ordinary assay conditions, that declines steeply at high temperatures.*

The temperature dependence of enzymatic reactions explains the fact that metabolic rate increases during fever. Presumably, it is also responsible for the inability of the human body to tolerate body temperatures greater than 42°C. Moreover, protein denaturation is a time-dependent process, and an enzyme that survives the temperature of 45°C or more for some minutes may well denature gradually in the course of several hours.

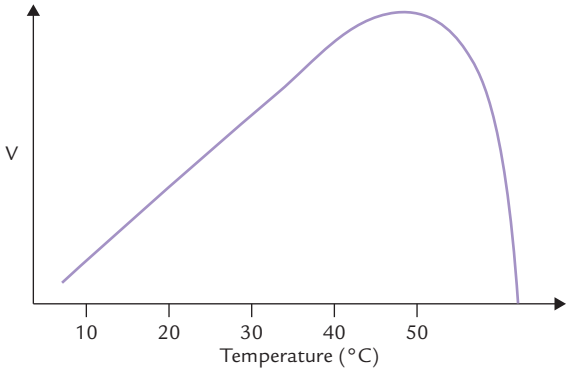


Fig. 6.4. Temperature dependence of enzyme. V = reaction velocity. Peak reaction velocity is reached at optimum temperature, and falls on increasing/or decreasing the temperature.

C. Effect of pH

The pH dependence of an enzyme protein is depicted in Fig. 6.5. The activity of a given enzyme is maximum at a specific pH which is the **optimum pH** of the enzyme.

Why does the pH exert such a significant influence on the catalytic properties of the enzyme? The reason that the catalytic process requires the enzyme active site and the corresponding substrate molecule to be in appropriate state of ionization. For example, if carboxyl group of some specific residue on the active site is in the anionic (−ve) form, the corresponding portion of the substrate should be in the cationic (+ve) form. Forces of attraction develop between the opposite charges, which permit optimization of the enzyme-substrate interaction. The pH value at which such a state exists corresponds to the optimum pH of the enzyme. The enzyme-substrate interaction being optimum at this pH, the activity of the enzyme reaches a peak (Fig. 6.5). *Extremes of pH decreases the enzyme activity by causing denaturation of the enzyme protein.*

Each enzyme has its characteristic optimum pH, corresponding to its maximum activity (Table 6.5). This has been depicted in Figure 6.5 also. The pH values of tissues and body fluids are tightly regulated. Deviations of more than 0.5 pH unit from the normal blood pH of 7.4 cause protein denaturation, and so rapidly fatal. The cytoplasm of most cells has a pH value between 6.5 and 7.0; the mitochondrial matrix has pH value between 7.5 and 8.0,

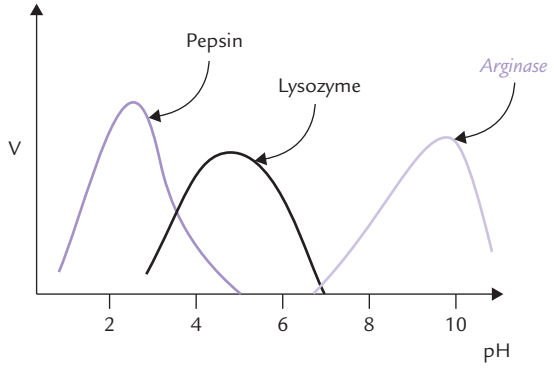


Fig. 6.5. pH dependence of some enzymes (V = reaction velocity. Each of the three enzymes has its characteristic optimum pH, corresponding to its maximum activity).

Table 6.5. Optimum pH values of some enzymes

Enzyme	Optimum pH
Pepsin	1.5
Lysozyme	4.8
Trypsin	7.7
Fumarase	7.8
Arginase	9.7

and the lysosomes are mildly acidic with pH values between 4.5 and 5.5. Lysosomal enzymes generally have pH optima in this (acidic) range.

D. Effect of Enzyme Concentration

Within reasonable limits, the rate of enzyme-catalyzed reaction increases linearly with enzyme concentration (Fig. 6.6). This is because the number of available substrate binding sites increase with the increase in enzyme concentration. For example, if the enzyme concentration is doubled, number of substrate binding sites is also doubled, and the rate of the reaction will increase twofold. This relationship holds good at all substrate concentrations.

The rate of enzymatic reaction increases with increasing temperature (typically doubling with every 10°C rise), pH, substrate concentration and concentration of enzyme.

VI. Michaelis–Menten Kinetic Theory of Enzyme Action

Michaelis–Menten model accounts for the kinetic properties of some enzymes. It helps to describe many enzymatic reactions under the following assumptions:

- The reaction has only one substrate
- The substrate concentration is much higher than that of the enzyme
- Only the initial reaction velocity (V_o) is measured.

Why is it necessary to consider only the initial reaction rate? At the beginning of the reaction the substrate alone, is present and the unoccupied active-sites of the enzyme molecule are free to bind them. Therefore, the substrate is rapidly converted to the product. As the product concentration builds up and the substrate concentration falls, the reaction keeps getting slower and eventually stops at equilibrium (Fig. 6.7). Since it is impossible to consider this "changing" reaction rate, only the initial-reaction-rate (V_o) is taken into account.

The relation between V_o and the substrate concentrate [S] is given by Michaelis–Menten equation:

$$V_o = \frac{V_{max}[S]}{K_m[S]}$$

$\mathbf{V_o}$ = initial reaction rate; $\mathbf{V_{max}}$ = maximum reaction rate; $\mathbf{K_m}$ = Michaelis constant; **[S]** = Substrate concentration.

The equation predicates a hyperbolic curve of V_o against [S].

$\mathbf{V_o}$ is measured at different substrate concentrations by incubating different amounts of the substrate with enzyme. The rate of disappearance of the substrate or the rate of appearance of the product *during the first few minutes of the reaction only is taken into account.*

$\mathbf{V_{max}}$ is the maximum velocity that can be reached by increasing the substrate concentration (Fig. 6.8). For any particular set of conditions it is a constant. It indicates a saturation state at which most enzyme active sites are occupied by the substrate molecules, and the reaction rate is no longer limited by substrate availability.

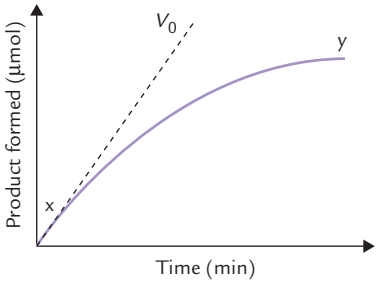


Fig. 6.7. The reaction velocity showing a progressive fall as the reaction proceeds; (x) represents the reaction rate at initial phase of the reaction, and (y) represents reaction rate at equilibrium.

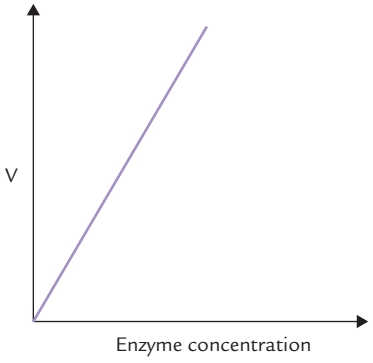


Fig. 6.6. Effect of enzyme concentration on enzyme activity (V = reaction rate).

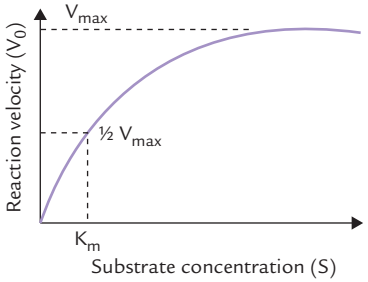


Fig. 6.8. The Michaelis–Menten model predicts a hyperbolic curve of the initial reaction velocity (V_o) against the substrate concentration. K_m is the substrate concentration at half the maximum velocity (V_{max}).

Table 6.6. K_m of certain enzymes

Enzyme	Substrate	K_m (mM)
Hexokinase	ATP	0.4
	Glucose	0.05
	Fructose	1.5
Glucokinase	Glucose	10
β-Galactosidase	D-Lactose	40

[S] is the substrate concentration at the initial phase of the reaction. This changes as the reaction proceeds, so the experimental data are taken, in the first few minutes of the reaction.

K_m is a special rate constant called the **Michaelis constant**. *It is defined as the substrate concentration at which the reaction rate is half-maximal* (Fig. 6.8). At this instance the enzyme is half-saturated with its substrate.

K_m has a characteristic value for a given enzyme-substrate pair. In case of the enzyme which acts on more than one substrate, the value of K_m is different for each of them. For example, K_m of the enzyme *hexokinase* for fructose is 1.5 mM, whereas for glucose it is 0.05 mM (Table 6.6).

K_m reflects binding affinity: *The Michaelis–Menten constant reflects the binding affinity of the enzyme for its substrate.* If affinity is more, less substrate is required to saturate the enzyme, so that V_{max} is reached at a relatively low substrate concentration. Consequently, the substrate concentration corresponding to the half V_{max} (i.e. K_m) is relatively low. For this reason, high enzyme-substrate affinity implies a low K_m value, and conversely, low-affinity implies high K_m. Higher K_m of glucose for *glucokinase* (10 mM) than that for *hexokinase* (0.05 mM) for example, indicates that *hexokinase* binds glucose with higher affinity than *glucokinase*.

The Michaelis–Menten equation describes the relationship between initial reaction velocity and substrate concentration under steady state conditions. The Michaelis constant K_m, defined as the substrate concentration at which the reaction velocity is half maximal, is related to the affinity of the enzyme for the substrate.

The following inferences can be drawn from the Michaelis–Menten kinetics.

Order of reaction: When the substrate concentration is relatively low (far below K_m), most of the enzyme active sites are free. An increase in substrate concentration at this stage results in rapid occupancy of the available sites. Therefore, the reaction rate is related almost linearly, with the substrate concentration, and the reaction shows **first order** kinetics (Fig. 6.8).

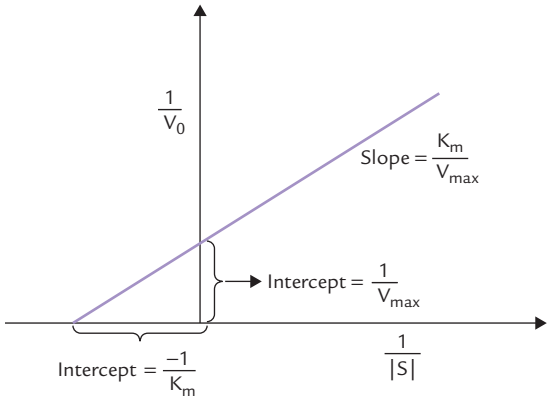


Fig. 6.9. $\frac{1}{V_0}$ against $\frac{1}{[S]}$. The straight line represents the Lineweaver-Burk transformation of the Michaelis–Menten equation.

With increasing substrate concentration, the reaction rate rises till a constant peak value (V_{max}) is reached. Nearly all enzyme active sites are occupied by the substrate at this stage and further increase in substrate concentration has no effect on the reaction rate. The rate of reaction at this stage is in **zero order** with respect to the substrate. It implies that the *reaction rate is independent of the substrate concentration* (Fig. 6.3).

Note: Derivation of the Michaelis–Menten equation is higher-level learning, hence omitted here.

Lineweaver–Burk plot: The slope of the Michaelis–Menten curve is gradual and upwards giving it a hyperbolic shape. From this shape, it is not possible to estimate the exact value of V_{max}. However, if reciprocal of these parameters: $\frac{1}{V_0}$ and $\frac{1}{[S]}$ are plotted against each other, a straight line is obtained (Fig. 6.9). From this line it is possible to determine the exact value of V_{max}. This plot, known as **double reciprocal plot** or the Lineweaver–Burk plot, can be used to determine the exact value of K_m. It is also used to determine the mechanism of action of enzyme inhibitors, discussed later.

The equation describing the double reciprocal plot is

$$\frac{1}{V_0} = \frac{K_m}{V_{max}[S]} + \frac{1}{V_{max}}$$

It is reciprocal of Michaelis–Menten equation. The intercept on the x-axis equals $\frac{-1}{K_m}$ and the intercept on the y-axis is equal to $\frac{1}{V_{max}}$, and the slope of the line is $\frac{K_m}{V_{max}}$.

V_{max} and K_m can be determined experimentally by measuring V_o at different substrate concentrations and then plotting $\frac{1}{V_0}$ against $\frac{1}{K_m}$. The intercept on X-axis is $\frac{-1}{K}$, that on Y-axis is $\frac{1}{V_{max}}$.

Bisubstrate reactions: Heretofore, we have considered simple, *single-substrate reactions*, which obey Michaelis–Menten kinetics. Yet, far more common (60% of total) are the enzymatic reactions requiring multiple substrates and yielding multiple products. Indeed those involving two substrates mostly yield two products

$$A + B \overset{\text{Enzyme}}{\rightleftharpoons} C + D$$

The bisubstrate reactions are classified as **sequential** (single displacement) and **ping-pong** (double displacement).

Sequential reactions: All substrates combine with the enzyme and only then the reaction occurs. In this type, the substrates encounter one another on the surface of the enzyme. Example: in *alcohol dehydrogenase* reaction, a hydride ion is transferred from ethanol to NAD^+.

$$\underset{\text{Ethanol}}{C_2H_5\text{–}OH} + NAD^+ \longrightarrow \underset{\text{Acetaldehyde}}{CH_3CHO} + NADH + H^+$$

A sequential reaction may proceed by an ordered or random mechanism.

Ping-pong reactions: In this type, all substrates do not combine with the enzyme together. The reaction starts before all substrates have been added, and often one or more products are even released prior to the addition of all substrates. Many enzymes, e.g. some *flavoenzymes* and *transaminases* react by this double displacement mechanism (Fig. 6.10).

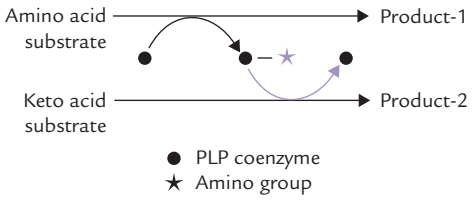


Fig. 6.10. Ping-pong reaction, also called double-displacement, in a transamination reaction. The first substrate, an amino acid, is converted to the corresponding keto acid by donating its amino group to PLP coenzyme before the second substrate (a keto acid) enters the reaction sequence.

In sequential reactions all substrates interact together at the enzyme active site. In double displacement they interact with the enzyme one by one.

VII. Inhibition of Enzyme Activity

Enzyme inhibitor is a substance that is capable of diminishing velocity of the enzyme-catalyzed reaction. Two types of inhibitors are recognized: reversible and irreversible. Inhibition by the **reversible inhibitor** can be reversed, so that the enzyme activity is recovered. Inhibition by the **irreversible inhibitor**, on the other hand, cannot be reversed.

Enzyme activities can be inhibited specifically in reversible (competitive, non-competitive and uncompetitive) and irreversible manner.

A. Reversible Inhibition

The reversible inhibitor (I) acts by forming a loose, dissociable complex (EI) with the corresponding enzyme

$$E + I \rightleftharpoons EI$$

Catalytic activity of the complex (EI) is lower than that of the enzyme alone. Therefore, the substrate transformation decreases after addition of the inhibitor. The reversible inhibition may be *competitive*, *non-competitive*, and *uncompetitive*.

Following formation of the EI complex, the reaction may still obey Michaelis–Menten kinetics but with altered K_m and V_{max} values that vary with the inhibitor concentration.

- If the inhibitor acts only on the apparent K_m, it is a competitive inhibitor;
- if the inhibitor affects only the apparent V_{max}, it is a non-competitive inhibitor; and
- if the inhibitor affects both the constants, it is an uncompetitive inhibitor.

Competitive Inhibition

Mechanism: The inhibitor exhibits is a close structural resemblance with the actual substrate. It impairs substrate-binding to the enzyme, because it occupies the active site of the enzyme making it unavailable for the substrate.

Occupancy of the enzyme active site by the inhibitor yields EI complex. Since the formation of EI depends upon the concentration of the inhibitor (just as the formation of ES is dependent on the substrate concentration), the *inhibition is strictly dependent upon the relative concentrations of the inhibitor and the substrate* at fixed concentration of the enzyme.

Example:

Some examples of competitive inhibitors are:

(a) Malonate ions inhibit the *succinate dehydrogenase* reaction of Kreb's cycle (Fig. 6.11).
(b) Acetazolamide causes competitive inhibition of *carbonic anhydrase*.

Inhibition of the *succinate dehydrogenase* reaction by malonate anions is a classical example of competitive inhibition (Fig. 6.11). Malonate is a structural analog of succinate, the actual substrate of the enzyme *succinate dehydrogenase*. It occupies the active site of the enzyme, thus preventing it from acting on succinate.

Some more of competitive inhibitors (of therapeutic significance) are given in the next section (see Table 6.8).

Effect on enzyme kinetics: The $\mathbf{K_m}$ shows an apparent **increase** in the presence of the inhibitor (Table 6.7). This reflects decreased affinity of the substrate with the enzyme active site. The decreased affinity is because the substrate must now compete with the inhibitor for the active site. This is depicted in a double reciprocal plot as a shift in the x-intercept ($-1/K_m$) and in the slope of the line (K_m/V_{max}) (Fig. 6.12). $\mathbf{V_{max}}$ (the measure of fastest possible rate) remains **unchanged** because the binding of inhibitor is reversible; and at infinite concentration of the substrate (obtained by extrapolation) the inhibitor can be completely displaced from the enzyme active site.

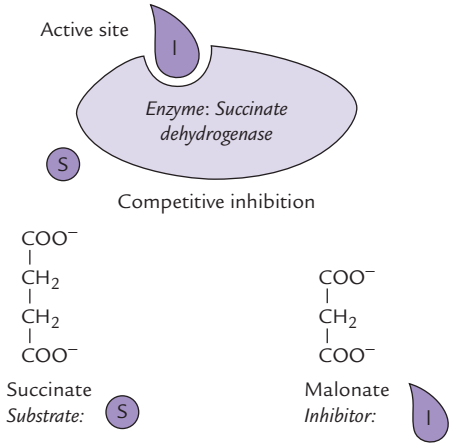


Fig. 6.11. Competitive inhibition of *succinate dehydrogenase* by malonate (I = malonate inhibitor, blocks active site of the enzyme).

Non-competitive Inhibition

Mechanism: In this type, the enzyme inhibitor is structurally unrelated to the structure of substrate and binds the enzyme at a site distinct from the enzyme active site. Occupancy of this site by the inhibitor alters the shape of the enzyme such that its catalytic activity is reduced or lost. The substrate is still able to bind the enzyme, but the enzyme cannot catalyze the reaction when inhibitor is bound. *Thus, the non-competitive inhibitor does not block active site of the enzyme, but behaves as though it were removing active enzyme from the solution*. Evidently, unlike competitive inhibition, increasing the substrate concentration does not reverse the inhibition.

The non-competitive inhibitor may bind the free enzyme as well as the enzyme substrate complex, forming *binary (EI)* and *ternary (ESI) complexes*, respectively. Both are catalytically inactive, and are therefore, dead-end complexes.

$$E + I \rightleftharpoons EI$$
$$ES + I \rightleftharpoons ESI$$

Some examples of non-competitive inhibitors are: (a) *Heavy-metal ions*, such as Hg, Ag, and Pb; they react reversibly with thiol groups (−SH) of enzymes. The thiol group participates in maintenance of the three-dimensional conformation of the enzyme proteins (b) *Chelating agents* such as ethylene-diaminetetra-acetic acid (EDTA). EDTA can chelate metals required for the activity of some enzymes.

Provided that the inhibitor has not reacted covalently with the enzyme, the non-competitive inhibition can often be reversed by exhaustive dialysis of the inhibited enzyme.

Effect on enzyme kinetics: The non-competitive inhibitor does not block the active site of the enzyme. Therefore, the $\mathbf{K_m}$ (indicating enzyme-substrate affinity) remains **unaffected**. Increasing the concentration of the substrate does not affect binding of the inhibitor to the enzyme. This is because the shape of the enzyme is different when the inhibitor is bound to it, and hence it does not matter how much substrate is present; the enzyme will always be less effective due to this shape change. This is reflected

Table 6.7. Comparison of two types of reversible inhibitors

	Competitive	Non-competitive
Site of action	Active site	Other than active site
Structure of inhibitor	Resembles substrate	Unrelated molecule
K_m	Increased	No change
V_{max}	No change	Decreased
Excess substrate	Reverses inhibition	No effect

in **decreased V_{max}**, and is seen graphically on the double reciprocal plot where both the slope and the y-intercept are changed (Fig. 6.12).

A competitive inhibitor blocks active site of the enzyme, whereas a non-competitive inhibitor acts by inducing a conformational change in 3-D shape of the enzyme, thereby lowering its catalytic activity.

Uncompetitive Inhibition

The inhibitor binds with only with the ES form of the enzyme to form the **enzyme-substrate-inhibitor (ESI) complex**. Binding may occur at a site distant from the active site, or at a site overlapping with the active site. In either case, the result is lowering of both: the catalytic efficiency of the enzyme and its apparent binding affinity for the substrate. These result in changes in K_m and V_{max} (both decreases). As a result, in the double reciprocal plot a line parallel to that of the uninhibited enzyme (slope remains unchanged), but with altered intercepts on both, the x and the y axis (Fig. 6.13).

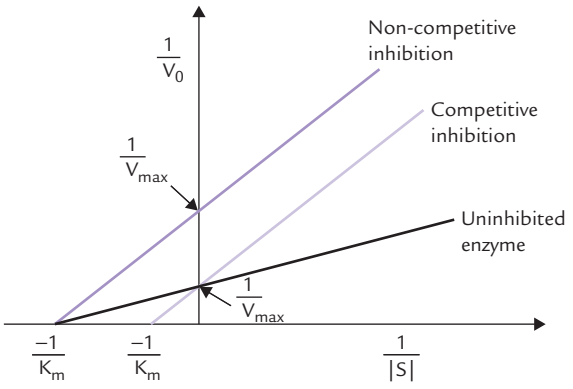


Fig. 6.12. Lineweaver-Burk plot showing the effect of competitive and non-competitive inhibitors. K_m increases in competitive and V_{max} decreases in non-competitive inhibition.

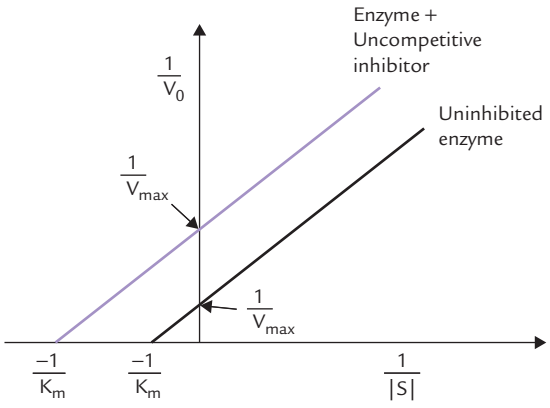


Fig. 6.13. The effect of uncompetitive inhibitor on enzymatic reaction.

The uncompetitive inhibition is uncommon in single-substrate reactions. But a noteworthy example in clinical enzymology is the inhibition of *intestinal alkaline phosphatase* by L-phenylalanine. It is more commonly observed in two-substrate reactions with a double-displacement reaction mechanism. However, the interpretation is complex in the latter case and will not be considered further.

Note: The types of inhibitions presented here are extreme cases. Actually, however, one may come across **mixed types of inhibitions**. For instance, the binding of an inhibitor to a site other than the active site may not only change the shape of the enzyme, but also affect the binding of the substrate at the active site, so that K_m as also V_{max} is altered. The double-reciprocal plot gives lines that intersect in the left hand quadrant of the graph.

Reversible inhibitors reduce an enzyme's activity by binding to the active site (competitive), to the enzyme-substrate complex (uncompetitive), or to both the enzyme and the ES complex (non-competitive). Various types of inhibitions can be distinguished according to the mode of interaction between the enzyme and the inhibitor, and the kinetic consequences of this interaction.

B. Irreversible Inhibition

The irreversible **inhibitor binds covalently with an enzyme** to form a stable complex. For all practical purposes, there is no dissociation of the enzyme and the inhibitor, so that the enzyme is permanently inactivated, or at best is slowly reactivated (requiring hours or days for reversal).

In some cases the inhibitor reacts at or near the active site of the enzyme with covalent modification of the active site. In other cases, the inhibitor has no structural resemblance with the substrate and may bind to a site other than the substrate-binding site. This type of inhibition resembles the non-competitive inhibition, except that covalent bonds are formed in this case.

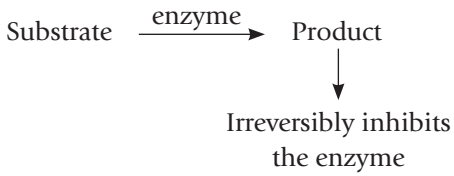
Irreversible inhibitor binds tightly, mostly covalently, to amino acids often at or near the active site of the enzyme, permanently inactivating the enzyme.

A classic example of the irreversible inhibition is that produced by *diisopropyl fluorophosphate* (DIPF), which irreversibly binds to the active serine residues of some hydrolytic enzymes. The reaction of aspirin with *cyclooxygenase* also represents an example of an irreversible enzyme inhibition. *Cyclooxygenase* catalyzes the first reaction in the biosynthesis of prostaglandin from arachidonate.

Aspirin acetylates the serine residue at the active site; the modification being highly stable leads to permanent loss of the enzyme activity. Thus, aspirin is used as an anti-inflammatory, antipyretic and anti-inflammatory drug. It is also used prophylactically to inhibit platelet aggregation and coronary thrombosis (Chapter 11).

A variety of poisonous substances, such as iodoacetate, heavy metal ions and oxidizing agents form covalent bonds, with functional groups of the enzyme, thereby causing irreversible inhibition.

Suicide inhibition refers to irreversible enzyme inhibition by a substrate analogue that, upon action of the enzyme, generates highly reactive species. This species forms a covalent adduct with an amino acid side chain at the active site, leading to irreversible inactivation of the enzyme. An example of this type of inhibitor is **allopurinol**, the suicide inhibitor of *xanthine oxidase* (**Case 6.2**). Also see **Case 20.1**.



Affinity labels are substrate analogues that possess a highly reactive group, not present on the natural substrate. This group reacts with amino acid residue(s) at the active site of an enzyme, thus blocking them irreversibly from the substrate.

C. Therapeutic Applications of Enzyme Inhibitors

Several synthetic drugs (and natural compounds) exert pharmacological effects by acting as enzyme inhibitors. This application is most easily appreciated with *antiviral*, *antibacterial*, and *antitumour drugs* that are administered under the condition of limited toxicity. Those drugs that resemble the substrate of an enzyme-catalyzed reaction are competitive inhibitors with respect to that substrate (and non-competitive with respect to other substrates).

Captopril, a commonly used drug in the treatment of hypertension acts by competitively inhibiting the *angiotensin converting enzyme* (ACE). This enzyme catalyzes the proteolytic cleavage of angiotensin I to angiotensin II; the latter elevates the arterial blood pressure. Captopril and other ACE inhibitors (e.g. lisinopril and enalapril) decrease formation of angiotensin II, thereby lowering blood pressure (Table 6.8). The ACE inhibitors contain a proline residue, which is an important substrate-binding determinant.

Lovastatin and **mevinolin** competitively inhibit activity of *HMG-CoA reductase,* the rate-limiting enzyme in cholesterol biosynthesis. Therefore, they are commonly used hypocholesterolaemic agents. Both the drugs resemble hydroxymethylglutaryl coenzyme A (HMG-CoA), the natural substrate for this enzyme.

A few more therapeutic applications of enzyme inhibitors are discussed in detail below:

Competitive Inhibition of Folate Biosynthesis

Competitive inhibition of a biosynthetic step in folate synthesis accounts for the antibacterial action of **sulphonamides**. Sulphonamides are structural analogues of p-aminobenzoic acid (PABA). Bacteria synthesize folate by combining PABA and pteroylglutamic acid.

$$\text{PABA} + \text{Pteroylglutamic acid} \rightarrow \text{Folate}$$

Being structural analogues of PABA, the sulphonamides inhibit synthesis of folate, which is required in several one-carbon transfer reactions. Because these reactions are essential for DNA synthesis, and for cell growth and division, inability to synthesize folate results in inhibition of bacterial cell growth and division (bacteriostasis).

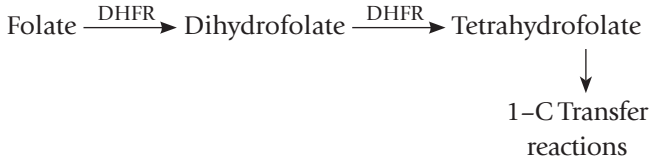
Table 6.8. Therapeutic applications of certain enzyme inhibitors

Therapeutic agent	Enzyme inhibited	Mode of inhibition	Clinical use
Captopril	Angiotensin-converting enzyme	Competitive	Hypertension
Lovastatin	*HMG-CoA reductase*	Competitive	Hypercholesterolaemia
Methotrexate	*Dihydrofolate reductase*	Competitive	Anticancer
Ethanol	*Alcohol dehydrogenase*	Competitive	Methanol poisoning
Digoxin	*Carbonic anhydrase*	Competitive	Diuretic
Acetazolamide	*Na^+–K^+ ATPase*	Competitive	Heart disease
5'-Fluorouracil	*Thymidylate synthetase*	Competitive	Anticancer
Isoxuridine	*Thymine kinase*	Competitive	Antiviral
Sorbinil	*Aldose reductase*	—	Diabetes
Carbidopa	*Dopa decarboxylase*	Irreversible	Parkinsonism
Cephalosporin	*Transpeptidase*	Irreversible	Antibacterial

Host cells (or other mammalian cells) are not affected because they use preformed folic acid.

Competitive Inhibition of *Dihydrofolate Reductase*

The enzyme *dihydrofolate reductase* (DHFR) is required for reduction of folate to its active form, tetrahydrofolate (THF), which then functions in several reactions including synthesis of the DNA (one-carbon transfer reactions).



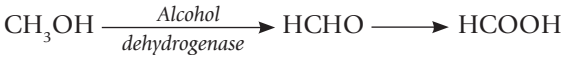
Methotrexate is a folate analogue; it therefore competes with the folate for an active site on DHFR, thus preventing its activation to THF. Since THF is required for the synthesis of DNA, a deficiency causes most harm to those cells which synthesize DNA rapidly, e.g. cancer cells (Chapter 20).

Competitive Substrate of Vitamin K, Dicoumarol

Competitive inhibition of vitamin K by the structurally similar dicoumarol accounts for its anticoagulant action.

Competitive Substrate in Treatment of Methanol Poisoning

Methanol, which is widely used industrially, is highly toxic. It is oxidized to formaldehyde and formic acid



Formaldehyde causes damage to retinal cells resulting in blindness, while formic acid produces severe acidosis that is potentially lethal. *Ethanol serves as an antidote to methanol poisoning*, being a natural substrate for the enzyme *alcohol dehydrogenase*. It is preferentially acted upon by this enzyme but the oxidation products are not as toxic. Thus, methanol oxidation and consequent generation of toxic products is averted.

Thus, competitive inhibition is the basis for the treatment of some intoxications, e.g. methyl alcohol and ethylene glycol. More examples of clinically useful enzyme inhibitors are listed in Table 6.8.

VIII. Regulation of Enzyme Activity

Activities of some enzymes, called the regulatory enzymes, are influenced by a variety of effectors. Regulation of activities of these enzymes is essential for (a) coordinating numerous metabolic process, and (b) altering the rate of a pathway according to the overall metabolic needs of the body. For example, during vigorous exercise, the energy requirement of the exercising muscle drastically increases. To meet this demand, the energy-yielding catabolic pathways are speeded up. There is a concomitant suppression of the energy-requiring anabolic pathways, which is consistent with the overall metabolic needs of the exercising tissue.

There are three major mechanisms for regulating the enzyme activity: *allosteric modulation, covalent modification and induction-repression of enzyme synthesis*.

A. Allosteric Modulation

Some enzymes, called allosteric enzymes, exist in alternate higher order structures. The equilibrium between these alternative structural forms can be affected by binding of regulatory ligands, termed **allosteric modulators**. The binding of a ligand can either enhance the activity of the enzyme (allosteric stimulation) or inhibit it (allosteric inhibition).

Properties of allosteric enzymes: The allosteric enzymes possess a site distinct and physically separate from the substrate-binding site; it is known as the **allosteric site**. The term allosteric is derived from the Greek word, *allo* which means, other and *stereos* means space or site. Allosteric modulators bind with the allosteric site by reversible non-covalent interactions (Fig. 6.14). This binding alters conformation of the enzyme protein, which in turn leads to alteration of the enzyme activity. Thus, occupancy of the allosteric site influences, the enzyme activity, thereby providing important means of its regulation. The process is relatively rapid, and hence, the first response of cells to any change(s) in intracellular metabolic state.

The *allosteric site is specific for the modulator*, just as the active site is for the substrate. Some modulators, called the positive modulators, increase the enzyme activity in this manner, whereas others, called the negative modulators, decrease the enzyme activity. The allosteric site at which the positive modulator binds is referred to as an *activator site*; the negative modulator binds at an *inhibitory site*.

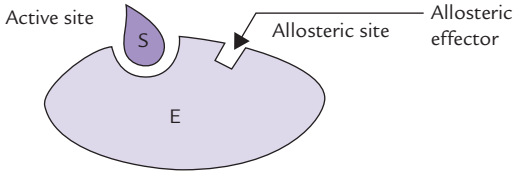


Fig. 6.14. An allosteric enzyme with active (catalytic) site and allosteric site. E = enzyme, S = substrate.

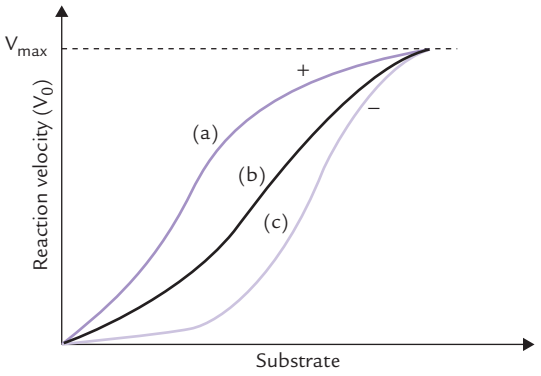


Fig. 6.15. The sigmoid curve showing the activity of an allosteric enzyme. **(a)** After addition of an allosteric activator, **(b)** Without any allosteric modulator, **(c)** After addition of an allosteric inhibitor.

Most allosteric enzymes are oligomeric, consisting of several subunits. Identical subunits are designated as **protomers**, and each protomer has its own active site, and may consist of one or more polypeptide chains. (Two notable exceptions are *ribonucleoside diphosphate reductase* and *pyruvate-UDP-N-acetylglucosamine transferase*, which are monomeric allosteric enzymes.)

Deviation from Michaelis–Menten kinetics: The allosteric enzymes do not show simple Michaelis-Menten kinetics and yield **sigmoid shape** of the V_0/[S] graph, rather than the hyperbolic curve of the non-regulated enzymes (Fig. 6.15). The sigmoid curve resembles the oxygen binding curve of haemoglobin. The cooperative interaction between different subunits of haemoglobin is responsible for the sigmoid shape. This concept of cooperativity applies to allosteric enzymes as well. Indeed, the allosteric site and the active site exhibit cooperativity: binding of a modulator to the former influences catalytic properties of the active site.

Homotropic and heterotropic effects: Effects of allosteric modulators on enzyme activity may be termed homotropic or heterotropic. When the substrate itself is an allosteric modulator, this is referred to as **homotropic effect**; whereas if it is different, it is called **heterotropic effect** (Fig. 6.16).

1. Homotropic effect is observed when the reaction of one substrate molecule with an enzyme affects the reaction of a second substrate molecule with (a different active site on) the same enzyme. Homotropic interactions are almost always positive.
2. Heterotropic effects may be positive or negative and are mediated by cooperativity between subunits of the enzyme. Usually the end product of a metabolic pathway modulates activity of a key enzyme (**feedback regulation**).

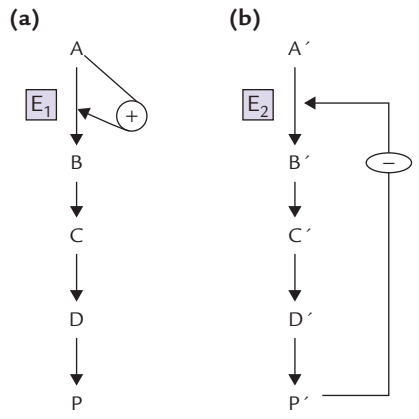


Fig. 6.16. Diagrammatic representation of **(a)** homotropic, and **(b)** heterotropic effect. Homotrophic effects are almost always positive, the heterotrophic effects are positive or negative (E_1, E_2 = enzymes).

In metabolic pathways, the end-product often feedback inhibits an enzyme earlier in the pathway to prevent build-up of intermediates and the unnecessary use of metabolites and energy.

A classical example of such effects is observed in the allosteric enzyme *aspartate transcarbamoylase* (ATCase) from *Escherichia coli*, which catalyzes the committed step in pyrimidine biosynthesis, i.e. formation of N-carbamoylaspartate from aspartate and carbamoylphosphate (Fig. 6.17).

CTP, an end product of the pathway heterotropically inhibits ATCase. This provides an example of feedback inhibition of the pathway by the end product. In contrast, ATP is an intermediate earlier in the pathway and it homotropically activates the enzyme activity (Chapter 20).

K- and V-class enzymes: Allosteric enzymes are divided into two classes: the K-class and the V-class, based on the effect of the allosteric modulator on K_m and V_{max}. In the **K-class** the modulator alters the K_m but not V_{max}.

In the **V-class** the modulator alters V_{max} but not K_m. There are a few enzymes which affect both K_m and V_{max}.

Allosteric effectors convert enzymes into more active or less active forms. Allosteric enzymes do not obey Michaelis–Menten kinetics: a plot of V_0 against |S| gives a sigmoid-shaped curve.

B. Covalent Modification

There are two general types of covalent modifications of enzymes that regulate their activities. These are the

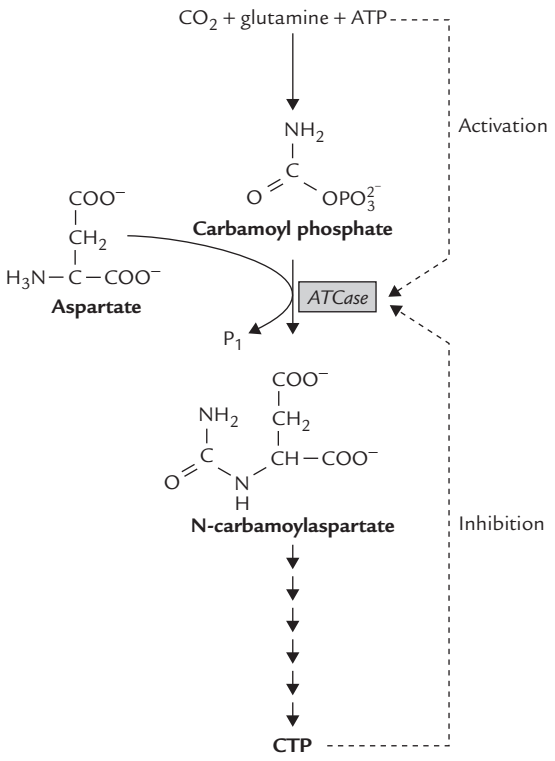


Fig. 6.17. Formation of N-carbamoylaspartate by *aspartate transcarbamoylase* (ATCase) is the committed step in pyrimidine biosynthesis and a key control point. CTP is allosteric inhibitor and ATP is allosteric activator.

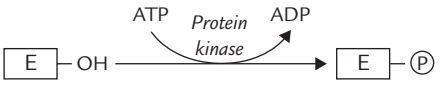
(a) **reversible interconversion** of active and inactive states of an enzyme by covalent attachment of specific group(s), and (b) **irreversible activation** of inactive enzyme precursor by proteolytic cleavage.

Reversible Covalent Regulation

Activities of several enzymes are regulated by covalent attachment of specific groups, such as phosphate, calcium or nucleotide. This results in altered charged-state of the protein molecule, which in turn causes change in shape of the protein and hence its activity.

Phosphorylation-dephosphorylation

R groups of some amino acids present at the active site of enzymes carry hydroxyl group (e.g. threonine, serine and tyrosine), which can be phosphorylated by *protein kinase*. This enzyme causes attachment of a phosphate group obtained from ATP to hydroxyl group(s) of these amino acid(s).



This puts a large negative charge in this region of the protein, which causes it to change its shape and its activity. Activities of the regulatory enzymes of glycogen metabolism, cholesterogenesis and *lipase* of adipose tissue are regulated in this manner.

Importantly, the changes to these proteins are **reversible**, and hence their activities can be switched on and off.

Calcium Binding

Several calcium binding proteins can attach calcium ions, which puts a positive charge on them, thereby modifying their activities. Examples of some enzyme proteins regulated in this way are *pyruvate carboxylase, glycogen synthase, phosphorylase kinase,* etc.

Adenylation

Some proteins bind cyclic nucleotides (cyclic nucleotide binding proteins), which alters their activity. This mechanism is mostly seen in bacteria.

Irreversible Activation by Proteolytic Cleavage

Several proteins are synthesized in inactive forms. These are called ***zymogens***, e.g. protein digesting enzymes and blood clotting proteins. They are activated when a small length of the protein is cleaved off from one end through action of specific *proteases*. This causes an irreversible rearrangement of the tertiary structure to yield the active form of the protein.

Trypsinogen → Trypsin + peptide fragment
Pepsinogen → Pepsin + peptide fragment

Enzyme activity may be regulated by allosteric effectors, covalent modulation, activation of proenzymes, synthesis or degradation of enzymes, etc.

C. Induction-repression of Enzyme Synthesis

The allosteric modulation and the covalent modifications, discussed so far, change the activity of the regulatory enzymes. In addition, there exist another mechanism that operates at gene-level to regulate the enzyme synthesis (or breakdown). This may cause increased synthesis (**induction**) or decreased synthesis (**repression**) of the enzyme protein and hence its intracellular concentration also changes accordingly.

The sex hormones oestrogen and testosterone are two very important regulators of genes. Insulin induces some key glycolytic enzymes and represses some enzymes of gluconeogenesis (Chapter 15). Levels of metabolites may also regulate gene expression.

Synthesis of the enzyme protein can be changed at genetic level. Induction implies enhanced synthesis and repression means decreased synthesis. Both alter intracellular concentration of enzyme protein.

IX. Isoenzymes

These are closely related enzyme proteins that catalyze the same reaction but differ in their molecular structure, kinetic parameters, electro-phoretic mobility, or other properties, e.g. different sensitivity to inhibitors or response to allosteric modulators. They may act on the same substrate, but with different K_m and V_{max} values.

Mechanisms for the Production of Isoenzymes

1. Most common mechanism for formation of isoenzymes is by way of different combinations of polypeptide subunits to form an active polymeric enzyme. These subunits arise from different genetic loci; for example,

- *Creatine kinase* occurs as a dimer with two types of subunits: M (muscle) and B (brain), which are products of loci in chromosomes 14 and 19, respectively. Three combinations of these subunits are possible—**MM**, **MB**, and **BB**.
- *Lactate dehydrogenase* occurs as a tetramer with two types of polypeptide subunits (H and M) that are arranged differently to yield five isoenzymic forms, described later (see Table 6.11).

2. Some isoenzymes, called **allelic isoenzymes**, arise from the *same genetic locus*, and yet they exhibit structural variations among different individuals. They are referred to as different alleles or alternate forms of the same enzyme. For example, about 400 different allelic isoenzymes of *glucose 6-phosphate dehydrogenase* have been identified; only one form is present in a given individual, but all different forms are seen in the whole population.
3. Post-translational modification of the same enzyme polypeptide may occur in different ways to generate different isoenzyme, called **iso-forms**. Several isoforms of the enzyme *alkaline phosphatase* are known, differing in the number of sialic acid residues, and therefore, the number of the charged group.

Separation of Isoenzymes

1. ***Electrophoresis***: Different subunit compositions or primary structures yield different ionic charges to the polypeptide chains—this permits their separation by electrophoresis. Several isoenzyme studies have relied upon different electrophoretic migration patterns.
2. ***Use of inhibitors***: High temperature or some inhibitor may inhibit an isoenzyme, which forms the basis for its identification. For example, the α-2 isoenzyme of *alkaline phosphatase* (ALP) of placental origin is heat stable even at 65°C, but the α-2 isoenzyme produced by hepatic cells loses its activity at this temperature in 30 minutes. The γ-ALP of intestinal origin is inhibited by phenylalanine.

Isoenzymes Useful in Diagnosis

CK and *LDH* have tissue-specific isoenzymes that are determined on a routine basis in most clinical laboratories. Their estimation makes clinical diagnosis more discriminating as discussed later in this chapter.

Isoenzymes are multiple forms of an enzyme, catalyzing the same reaction but differing in kinetic properties. LDH and CK isoenzymes are very important in the diagnosis of myocardial infarction.

X. Enzymes in Clinical Diagnosis

A. Overview

Enzyme assays are performed in blood samples and other body fluids to further the diagnosis of a wide range of clinical conditions. Cellular enzymes are usually confined to the cells in which they are synthesized and where they function. During normal cellular turnover they are released into plasma, from where they are subsequently removed and eliminated from the body. In plasma they do not perform a physiological function and their plasma levels in healthy state are low to absent. Under pathological conditions, the serum levels of cellular enzymes can increase, because disease process can increase the cell-membrane permeability and release of intracellular enzymes into serum. Metabolic stress without necrosis can also lead to elevated enzyme levels, apparently by a transient increase of the membrane permeability.

Each cell type has its characteristic complement of enzymes (Table 6.9), and so identification of the pattern or profile of the elevated enzymes in plasma can point towards the damage of particular tissues. Following a block in coronary blood vessel, for example, oxygen deprivation damages the heart muscles and activities of the myocardial enzymes, e.g. *lactate dehydrogenase* and *creatine kinase* in blood rise markedly. Neoplastic diseases also result in

Table 6.9. Tissue localization of some important enzymes

Enzyme	Main localization
Aspartate transaminase	Heart, liver, skeletal muscle, kidney, brain
Alanine transaminase	Liver, skeletal muscle, kidney, brain
Alkaline phosphatase	Bone (osteoblasts), intestinal mucosa, liver, placenta, saliva
Acid phosphatase	Prostate, erythrocytes
Lactate dehydrogenase	Heart, liver, skeletal muscle, kidney, erythrocytes, pancreas, lung
γ-Glutamyl transpeptidase	Liver
Creatine kinase	Skeletal muscle, heart, brain
α-Amylase	Pancreas, saliva
Aldolase	Skeletal muscle, heart
Arginase	Liver
Acetylcholinesterase	Brain, nervous tissue, erythrocytes
5′-Nucleotidase	Hepatobiliary tract, pancreas
Glucose 6-phosphatase	Liver

elevated enzyme levels when the tumour invades and destroys tissues. The tumour cells themselves represent an increased tissue source for the enzyme. Moreover, the areas of tumour necrosis can develop when the tumour outgrows its blood supply.

Measurement of the amount of enzyme that has leaked provides an estimate of how much tissue damage has occurred. Cytosolic enzymes appear in plasma before mitochondrial enzymes, which are released only in severe damages, so that the total enzyme activity is markedly elevated. Thus, measurement of plasma enzyme activity indicates the severity of the disease and the prognosis for the patient.

An enzyme gives best diagnostic clue if it is specific for a particular tissue or even for a disease. Few enzymes meet such requirements: *alcohol dehydrogenase,* which is unique to liver; and *acid phosphatase* with its only established clinical use in diagnosis of metastatic prostatic carcinoma. Other than these two examples, few enzymes are disease or tissue specific, which limits their diagnostic utility. Fortunately, many of these enzymes occur as different **isoenzymes** in different tissues and so estimation of isoenzymes makes the diagnostic enzymology more discriminatory.

Assays of enzymes in blood are useful for diagnosis of some diseases: *amylase* and *lipase* in pancreatitis, alanine transaminase in hepatitis, *aspartate transaminase*, *lactate dehydrogenase* and *creatine phosphokinase* in myocardial infarction, etc. These diseases involve death of the affected tissue, with release of the cellular contents into the blood.

B. Clinical Applications of Diagnostic Enzymes

The disease process may cause changes in the cell membrane permeability and release the intracellular enzymes into the plasma. The circulating enzymes can be quantified by measuring their catalytic activity or kinetic properties. Thus, clinical enzymology is a useful diagnostic tool, and several enzymes are estimated routinely in most clinical laboratories (Table 6.10). Some of the most important ones are:

Transaminases

These are important enzymes of intermediary metabolism that catalyze reversible transfer of an amino group from an α-amino acid to a α-keto acid (Fig. 6.10). *Aspartate transaminase* (formerly known as *glutamate oxaloacetate transaminase,* or **GOT**) and *alanine transaminase* (formerly known as *glutamate pyruvate transaminase*, **GPT**) are two important transaminases, providing important diagnostic clues in hepatic, cardiac and skeleto-muscular disorders.

Aspartate Transaminase (AST)

Distribution: *Aspartate transaminase* has a wide distribution, being present in heart, liver, kidneys, skeletal muscles and erythrocytes.

Diagnostic utility: Heart and liver are richest in this enzyme, and the plasma AST activity is elevated in diseases involving these organs. In myocardial infarction (MI), gross elevation of the AST activities are seen; the elevation follows *creatine kinase*. In viral and toxic hepatitis, the enzymes activity rise up to 25-folds. Moderate elevation (up to fivefold) occurs in cirrhosis and obstructive liver diseases. Other conditions characterized by increased plasma AST include acute pancreatitis, degenerative diseases of skeletal muscles, severe haemolytic anaemia and infectious mononucleosis (with liver involvement).

Alanine Transaminase (ALT)

Distribution: *Alanine transaminase* is also widely distributed among tissues, but by far the highest concentration is found in liver. Next to liver, highest ALT concentration is found in skeletal muscles.

Diagnostic utility: ALT is an early and a specific indicator of hepatocellular damage (Fig. 6.18). Rise in its activity occurs in skeleto-muscular disorders, and trauma also (Table 6.10).

Specificity of ALT combined with sensitivity of AST, has made the two *transaminases* the enzymes of choice in studying liver diseases (**Case 6.3**).

Table 6.10. Commonly assayed enzymes for specific diagnosis

Enzyme	Diseases
Aspartate transaminase (AST)	Rises in myocardial infarction after CK and returns to normal in 4–5 days. Early indicator of hepatocellular damage
Alanine transaminase (ALT)	Marked elevation in acute hepatitis (viral or toxic) and in other parenchymal liver disease
Alkaline phosphatase (ALP)	Marked elevation in obstructive liver disease and in bone diseases with increased osteoblastic activity, e.g. rickets
Acid phosphatase (ACP)	Marker for carcinoma prostate. Rises in metastatic bone diseases, especially from primary from prostate
Lactate dehydrogenase (LDH)	Rises in myocardial infarctions after CK and AST; LDH_1 becomes more than LDH_2 (called flipped pattern)
γ-Glutamyl transpeptidase	Liver disease, especially alcoholism
Creatine kinase (CK)	Marked increase in muscle disease (CK-MM) and in myocardial infarction (CK-MB is the first enzyme to rise)
Amylase	About 1000-fold rise in acute pancreatitis
Lipase	Highly elevated in acute pancreatitis
Prostate specific antigen	Marker for carcinoma prostate

LDH_1 and LDH_2 = *lactate dehydrogenase* isoenzymes 1 and 2.

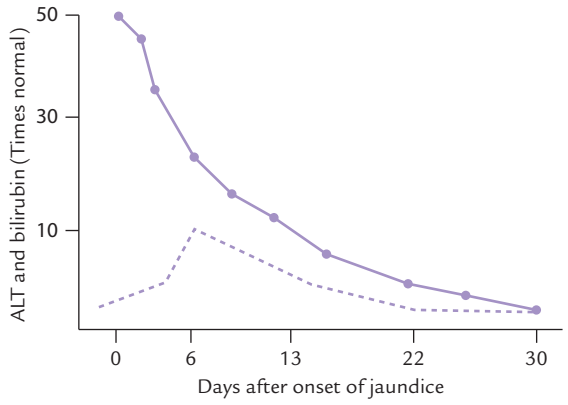


Fig. 6.18. *Alanine transaminase* is an early indicator of hepatocellular damage.

Phosphatases

Alkaline Phosphatases (ALP)

Alkaline phosphatase catalyzes hydrolytic removal of phosphate group from organic phosphate esters at an alkaline pH (optimum pH is approximately 9.0). Their substrate specificity is low, and they can act on a large number of organic phosphate esters.

Distribution: Highest concentration of ALP is found in liver, bone, intestine and placenta. Several isoenzymes, secreted by the above tissues are present in the plasma. The serum ALP activity, measured by routine methods, includes activities of all these isoenzymes. To distinguish the isoenzymes, serum electrophoresis is the preferred technique.

Diagnostic utility: Due to multiple origins of the enzyme, ALP activity is raised in a number of diseases. However, raised activity is especially useful in diagnosis of bone and liver pathology, since these are the major sources of the enzyme (Table 6.10). ALP is most predictably raised in diseases of the liver, particularly when there is an obstructive pathology (i.e. cholestasis). In hepatocellular damage, the degree of elevation is much less. Hence, ALP is a sensitive indicator of cholestatic lesions; its diagnostic utility in hepatocellular damage is much less (**Case 6.3**).

In bone ALP probably plays a role in calcium phosphate deposition. This enzyme is often elevated in several bone disorders where osteoblastic activity increases, e.g. Paget disease, rickets, osteomalacia, hyperparathyroidism.

Malignant disorders, primary or metastatic, may raise the enzyme activity because of two factors:

- Some malignant cells are capable of producing ALP.
- Secondary infiltration of bone or liver by malignant cells cause damage to these organs.

Isoenzymes of alkaline phosphatase: A number of different isoenzymes of *alkaline phosphatase* are known, named according to their electrophoretic mobility: α-1 form moves in the α-1 position of the electrophoretogram and α-2 moves in the α-2 position. The *various isoenzymes originate from different tissues*.

1. **α-1 ALP** is produced in the epithelial cells of biliary canaliculi. Normally, it forms about 10% of the total ALP activity, and shows elevation in obstructive jaundice.
2. **α-2 ALP** is synthesized by hepatic cells, constituting about 25% of the total ALP activity. Its activity rises

in hepatitis. Placenta also secretes the α-2 ALP form that differs from the hepatic α-2 form in two aspects: (a) it is more heat stable, and (b) it is inhibited by phenylalanine.

3. **Pre-β ALP** originates from bone cells, forming about 50% of the total ALP activity. Elevation in its activity is seen in bone diseases.
4. **γ-ALP** originates from intestinal cells. Its activity is normally 10% of the total ALP activity, and it increases in ulcerative colitis. This form is inhibited by phenylalanine.
5. A distinct type of ALP originates from leucocytes. It is increased in lymphomas and decreased in chronic myeloid leukaemia.

Regan isoenzyme: It is an isoenzyme of ALP (named after the patient in whom it was first detected) that is present in plasma of about 15% of the patients with carcinoma of lung, liver or gut. It is also seen in chronic smokers. Structurally, the Regan isoenzyme closely resembles the placental form.

Acid Phosphatase

Acid phosphatase also catalyzes the hydrolysis of phosphate from a variety of phosphate esters. It is so named because it exhibits an acid pH optimum, i.e. pH 5–6. Elevation of *acid phosphatase* occurs with metastatic carcinoma of prostate. Normal prostate does not release significant *acid phosphatase* into the circulation, whereas metastatic prostate tissue does so.

Dehydrogenases

Lactate Dehydrogenase

It is an enzyme of anaerobic glycolysis that catalyzes the reversible interconversion of lactate and pyruvate.

Distribution and diagnostic utility: It is widely distributed in all human tissues; therefore, its plasma level is elevated in a wide variety of diseases, including liver diseases and myocardial infarction, or haemolytic anaemia when erythrocytes are degraded more rapidly than normally. Fortunately, different tissues contain different isoenzymes, which enhance its diagnostic utility.

Isoenzymes of LDH: The isoenzymes are formed from two different subunits, designated H (heart) and muscle (M), which have essentially the same molecular weight (34,000) but different charge patterns. They combine in different combinations to yield five isoenzymes (Table 6.11), which can be separated by electrophoresis.

- The isoenzyme 1, a tetramer of four H subunits moves fastest at pH 8.6.
- The isoenzyme 5, consisting of four M subunits, moves slowest at this pH.

Table 6.11. Tissue distribution of lactate *dehydrogenase* isoenzymes

Isoenzymes	Subunit composition	Tissue	Percentage in serum
LDH_1	HHHH	Myocardium, RBC	30
LDH_2	HHHM	Myocardium, RBC	35
LDH_3	HHMM	Brain, kidney	20
LDH_4	HMMM	Skeletal muscle, liver	10
LDH_5	MMMM	Skeletal muscle, liver	5

Apart from their electrophoretic mobilities, the LDH isoenzymes can be distinguished from one another on the basis of other properties, such as:

- Sensitivity to heat and inhibitors.
- Tissue distribution.

The isoenzyme pattern of a given tissue depends on the relative amounts of H and M subunits produced by its cells, which differs markedly among different tissues. Thus, liver and skeletal muscles produce mostly M subunits; myocardium and bone marrow produce mostly H subunits; and most other tissues (lung, brain, kidney, pancreas) produce both. Thus LDH_1 (HHHH) is the predominant isoenzyme in cardiac muscle and erythrocytes. Therefore, in myocardial infarction and haemolytic crises of sickle cell anaemia, the LDH_1 is raised. LDH_4 and LDH_5 are predominantly elevated in muscle trauma (**Case 6.4**). In liver disorders, such as infectious hepatitis, the hepatic isoenzymes, LDH_4 and LDH_5 rise in blood.

Rise in LDH_5 occurs in heart failure also. Note that *heart failure is not the same as myocardial infarction*; whereas MI refers to heart cell death, heart failure refers to the inability of the heart to adequately maintain cardiovascular output. An infarction can occur without heart failure and vice versa. In right heart failure, blood accumulates in the venous side of the circulation and causes liver engorgement. As a result, liver cells release LDH_5.

Note: Isoenzyme determination, as also a simultaneous estimation of various enzymes, are of invaluable help for the clinician to reach a conclusive diagnosis. For example, an individual with chest pain may be suffering from a myocardial infarction, pulmonary infarction, or (rarely) both. Either disorder can occur with or without changes in electrocardiogram (ECG). In myocardial infarction, the isoenzymes 1 and 2 are elevated (1 or 2 days) after an episode of chest pain; whereas, elevations of isoenzymes 3, 4, and 5 occur after pulmonary infarction. Also AST, CK and *β-hydroxybutyrate dehydrogenase*

activities are more or less normal in pulmonary infarction (only LDH is elevated), whereas they are all raised in myocardial infarction.

β-Hydroxybutyrate Dehydrogenase

β-Hydroxybutyrate dehydrogenase catalyzes the oxidation of β-hydroxybutyrate by NAD^+ to yield acetoacetate and NADH. The enzyme has a wide distribution, being present in most human cells. Its activity is elevated following a myocardial infarction, and the elevation follows *creatine kinase* and AST.

Transferases

γ-Glutamyl Transpeptidase (GGT)

It catalyzes reaction between glutathione and an amino acid to form a γ-glutamyl amino acid and cysteinylglycine. It is a sensitive indicator of liver diseases, especially of alcoholism where it shows pronounced elevation when there are no other serum enzyme abnormalities.

Creatine Kinase (CK)

It is an important enzyme in energy metabolism that catalyzes transfer of the phosphate group of creatine phosphate to ADP to form ATP, thereby serving as an immediate source of ATP in contracting muscles. It is a reasonably specific muscle enzyme, found in heart and skeletal muscle. Besides muscle tissue, only the brain contains appreciable amounts of CK. Its activity rises in Duchenne muscular dystrophy and several other muscular afflictions. The most important diagnostic clue is given in myocardial infarction as it is the **first** *enzyme to be elevated following an episode of MI*; the enzyme is elevated in about half of the affected people 6 hours after the initial insult. Whereas, isoenzyme measurements are more useful in the diagnosis of MI, the main use of total CK measurement is in the diagnosis of muscle diseases. Injuries, intramuscular injections, vigorous exercise, polymyositis, and muscular dystrophies increase the serum CK level (as well as increase LDH, AST and myoglobin.

Isoenzymes of CK: The isoenzymes are tissue specific and are very useful for the specific diagnosis of myocardial infarction. Two slightly different gene products contained in the muscle (M) and the brain (B) correspond to CK. The active CK is a dimer and three isoenzyme forms are therefore possible: **CK_1 (BB), CK_2 (MB)** and **CK_3 (MM)**. The numbering is based on electrophoretic mobility; CK_1 has the greatest anodic mobility at pH 8.6. CK_1 occurs in brain, CK_2 occurs in heart and CK_3 in skeletal muscles (Table 6.12).

Rise in activity of CK_2 is the cornerstone for the diagnosis of acute myocardial infarction because of its abundance in the heart and absence from other cells (**Case 6.5**).

Table 6.12. Characteristics of CK-isoenzymes

Isoenzymes	Sub-unit composition	Tissue	Percentage in serum
CK_1	BB	Brain	1
CK_2	MB	Heart	5
CK_3	MM	Skeletal muscle	80

Besides **LDH** and **AST**, **CK_2** is the most revealing enzyme in the diagnosis of acute MI. For instance, in differentiating MI from pulmonary infarction (both may have similar clinical picture, but without any perceptible changes in ECG), measurement of CK_2 activity is important: it never rises in pulmonary infarction (lungs lack this enzyme).

The brain isoenzyme is rarely elevated in the blood, even after cerebrovascular accidents, although it is increased in cerebrospinal fluid of patients with various CNS diseases.

The electrophoresis method for determining the enzyme is too slow and insensitive to be of much value in cardiac emergencies. Therefore, **ELISA** based on monoclonal antibodies to CK_2 is used. This method is quick and sensitive enough to detect this isoenzyme in serum within an hour or so of an episode of MI.

Diagnostic utility of some enzymes in MI is highlighted in Box 6.3.

Amylase and Lipase

These are the digestive enzymes secreted by exocrine pancreas, and their serum levels are most notably increased in acute pancreatitis. Estimation of their serum levels provides important diagnostic clues in patients who present with severe abdominal pain of sudden onset (**Case 26.1**).

In such emergencies, acute pancreatitis has to be differentiated from a variety of other causes of acute abdomen, such as cholelithiasis, perforated peptic ulcer and peritonitis. Though most pronounced elevations occur in acute pancreatitis, modest elevations of these enzymes may occur in various extrapancreatic disorders; *amylase* levels are elevated in inflammation of salivary glands such as that produced by mumps, and *lipase* levels are elevated in intestinal infarction, peritonitis or perforation.

Plasma Cholinesterase (Pseudocholinesterase)

Distribution: This enzyme, unlike other diagnostic enzymes is normally secreted by hepatic cells, and hence always present in serum, though its physiological role is uncertain. Unlike *acetylcholinesterase* of cholinergic synapses, it has a broader specificity and participates in metabolism of some drugs, including cocaine and succinylcholine (Box 6.4).

BOX 6.3

Diagnostic Enzymes in Myocardial Infarction

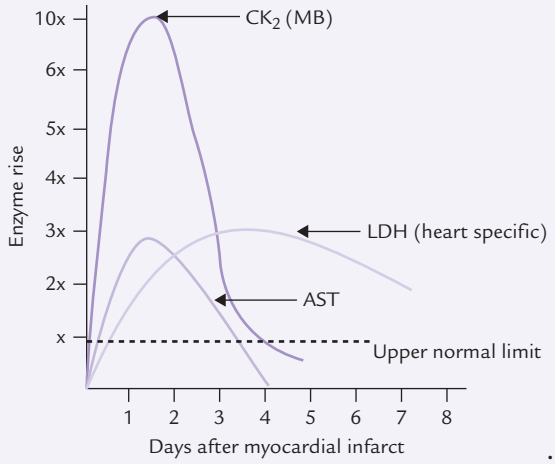
Measurement of serum activities of several enzymes which are normally present in myocardium in high concentration, have diagnostic utility in MI. These enzymes are: *lactate dehydrogenase* (LDH), *creatine kinase* (CK), and *aspartate transaminase* (AST). Serum activities of these enzymes are raised in myocardial infarction.

In myocardial infarction, sudden and complete occlusion of coronary artery occurs, which cuts off blood supply to a portion of myocardium. Oxygen and glucose supply to the myocardium is thereby blocked, resulting in tissue necrosis. The intracellular enzymes leak out from the damaged cells into the blood circulation. These enzymes provide chemical evidence of damage to the myocardial tissue.

In MI, a simultaneous release of the CK, AST and LDH occurs from the damaged myocardium (see the Figure in this box). But they are removed from blood circulation at different rates. Serial determinations of enzyme activities, therefore, show transient elevations, which persist for different periods. For instance, elevation of CK activity lasts for 2–3 days, and for about a week or more in the case of LDH.

CK (the MB isoenzyme) is the first enzyme to respond (within 3–6 hours), peak activity is reached in about 12–24 hours and normal values are attained by the third day.

AST activity starts rising 6–8 hours after the attack. It peaks at about 24–36 hours and returns to normal by 4–6 days. **LDH** activity starts rising about the same time, or a little later than AST. Peak activity reaches in 48–72 hours and returns to normal by 7–10 days.



Enzyme	Abnormal activity detectable (hours)	Time for peak activity rise (hours)	Time for return to normal (days)
CK_2 (MB-isoenzyme)	3–6	12–24	2–3
AST	6–8	24–36	4–6
LDH (heart specific)	8–10	48–72	7–10

BOX 6.4

Scoline Apnoea in Patients with Low *Plasma Cholinesterase*

Scoline apnoea is a fatal condition that may develop after an overdose of succinylycholine, a muscle relaxant. However, if the plasma levels of *cholinesterase* are low or absent (this is a genetically transmitted condition), the patient may develop apnoea even after a standard dose of succinylcholine. *Why?* The reason is that succinylcholine is structural analogue of acetylcholine and so competitively occupies the post-synaptic receptors of actelycholine. Succinylcholine is inactivated by *plasma cholinesterase* in about 2–4 minutes in a normal subject.

In patient with low *plasma cholinesterase*, metabolism of succinylcholine slows down, so much so that it may take several hours to metabolize succinylcholine. The muscle relaxing effect is prolonged and the patient is at high risk of developing fatal apnoea. It is therefore desirable that prior to administration of succinylcholine, the *plasma cholinesterase* activity is estimated and the normal levels ascertained.

Diagnostic utility: Most important use of the *plasma cholinesterase* is in the diagnosis of organophosphate poisoning. Organophosphates, which include several commonly used insecticides, are toxic for both insects and humans because they inhibit *acetylcholinesterase* (at cholinergic synapse, irreversibly). The organophosphates also inhibit the *pseudocholinesterase*, therefore, the enzyme activity is decreased in organophosphates poisoning.

Decreased *pseudocholinesterase* levels in serum are found in viral hepatitis, cirrhosis, primary and metastatic cancers of liver and malnutrition.

Note: In contrast to the non-plasma specific enzymes, discussed so far, some other enzymes have a functional role in plasma. They are called ***functional plasma enzymes***. Examples include the enzymes associated with processing of low density lipoproteins and chylomicrons (i.e. *lipoprotein lipase*), blood coagulation (e.g. *thrombin*), and fibrin dissolution (e.g. *plasmin*). Plasma concentration of such enzymes is relatively higher.

XI. Enzymes as Therapeutic Agents

Enzymes find widespread use in treatment of a number of diseases:

- *Streptokinase* (from Streptococcus) or *Urokinase* (from urine) can cause lysis of intravascular clots and therefore used in myocardial infarction.
- *Hyaluronidase* has degradative action against hyaluronic acid, a mucopolysaccharide in extracellular tissue. Therefore, it is used for increasing tissue permeability, and in treatment of traumatic or post-operative oedema.
- *Asparginase* is used as an anticancer drug; most commonly in some types of leukaemia.
- *Pepsin* and *trypsin* are useful in patients with chronic indigestion and in pancreatic insufficiency.
- *Alpha-1-antitrypsin* (AAT) is used in treatment of a type of emphysema that is caused by deficiency of AAT (Chapter 5).
- *Collagenase* degrades collagenous tissue, and so used in management of severe burns and in debridement of dermal ulcers.
- *Lysozyme* (endogenous antibiotic) is found in human tears. It has antibacterial properties being active against cellulose of bacteria, and is used in the infection of eye.
- *Penicillinase* is a bacterial enzyme which degrades penicillin. Therefore, it finds use in management of patients who are allergic to penicillin.
- *Papain* is anti-inflammatory
- *Streptodornase* is a DNAase, applied locally.

Exercises

Essay type questions

1. What are enzymes. What properties distinguish them from inorganic catalysts.
2. What are cofactors and why are they needed for some enzymatic reactions.
3. What factors influence enzyme activity?
4. Write double-reciprocal equation and describe the features of a Lineweaver-Burk plot. Discuss the effects of different types of inhibitions on this plot.
5. Discuss some mechanisms for regulating enzyme activity.
6. Describe the importance of serum enzymes in the diagnosis of diseases.

Write short notes on

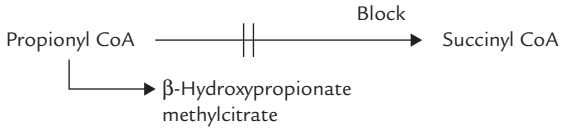
1. Allosteric enzymes
2. Active site
3. K_m value
4. Isoenzymes
5. CK isoenzymes
6. Michaelis constant
7. Competitive inhibition
8. *Lactate dehydrogenase* isoenzymes

CLINICAL CASES

CASE 6.1 A 3-year-old girl with fast and shallow breathing

A 3-year-old girl was brought to the Pediatrics emergency in a disorientated state, having fast and shallow breathing. Blood analysis indicated an acidotic state (reflected by reduction of plasma bicarbonate concentration and decreased blood pH; and marginally lowered pCO_2). To elucidate the underlying cause of the metabolic acidosis, relevant biochemical tests were performed. Lactic acid and ketone bodies were measured in blood and urine samples with possible diagnosis of lactic acidosis or ketoacidosis in mind. (These two are the most common causes of metabolic acidosis.) However, the test results ruled out these possibilities. Urine was analyzed by chromatography. The following abnormal constituents were detected; β-hydroxypropionate and methylcitrate. Since these two compounds are derived from propionyl CoA as alternate metabolites, it was speculated that normal metabolism of this compound (propionyl CoA) was affected.

To investigate this possibility the following experiment was carried out.



The patient's fibroblasts were cultured and a cell-free extract* was prepared from them. Radioactive-labeled propionyl CoA was added to the extract and its conversion to succinyl CoA was measured. It was found to be much less than expected in a normal subject. Similarly, conversions of radioactive pyruvate to oxaloacetate, and of 3-methylcrotonyl CoA to 3-methylglutaconyl CoA were decreased.

Q.1. What is the most likely biochemical defect in this child?
Q.2. Intracellular biotin concentration was measured and found normal. What is your diagnosis in view of this observation?

*Contains all the enzymes and cofactors normally present within a cell.

CASE 6.2 Enzyme inhibitors for treatment of gouty arthritis

A 30-year-old man has recurrent arthritic attacks. On examination signs of inflammation were present in several joints. Serum uric acid was grossly elevated, as was the urinary excretion of this compound. Further investigation revealed deposition of urate crystals in the inflamed joints and tissues. Treatment was started with allopurinol at a daily dose of 50 mg twice a day. Over a period of 2–3 weeks, the dose was gradually increased to 400 mg per day. Ibuprofen was added to control pain and inflammation.

Q.1. What is the rationale behind treating this patient with allopurinol?
Q.2. State the mechanism of action of allopurinol.
Q.3. State the biochemical basis of using ibuprofen in this patient.
Q.4. Allopurinol is recommended for the leukaemia patients along with the anticancer drugs. What is the logic behind using this combination?

CASE 6.3 Enlarged liver with yellow sclera and dark urine

A 15-year-old boy complains of nausea, vomiting and abdominal discomfort for the past one month. In general, he had been feeling "yellow" for the past few weeks. On examination, yellow discoloration of skin and sclera were detected. Liver was palpable 3 cm below the costal margin.

The urine and serum samples, obtained for biochemical analysis, were deep yellow in colour, and the stool sample was clay coloured.

Investigations test	Patient's reports	Reference range
Serum bilirubin	18 mg/dl	0.1–1.0 mg/dL
Aspartate transaminase (AST)	240 U/L	10–35 U/L
Alanine transaminase (ALT)	278 U/L	10–40 U/L
Alkaline phosphatase (ALP)	186 U/L	40–100 U/L
Total protein	6.2 g/dl	6.2–8.2 g/dL
Serum albumin	3.2 g/dl	3.6–5.4 g/dL
Serum globulin	3.0 mg/dl	1.8–3.6 g/dL
Urine bilirubin	+ ve	

Q.1. Interpret the results of the enzyme measurements and identify the underlying defect in this patient?
Q.2. Account for a low A : G ratio.
Q.3. In general, what causes elevation of ALP activity.
Q.4. The serum *alanine transaminase* is an early indicator of hepatocellular damage. Give reason.
Q.5. Mention the significance of relative values of AST and ALT activities in assessing the severity of hepatocellular damage.
Q.6. Why were the stools "clay colored" and why was the urine brown?

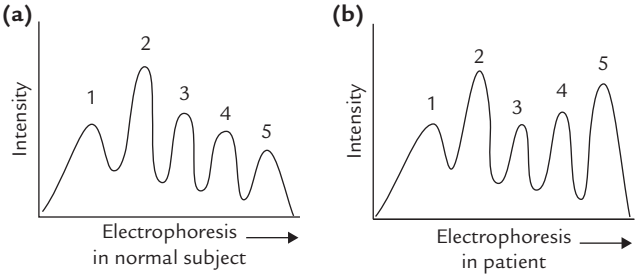
CASE 6.4 A 28-year-old man having chest pain

A 28-year-old salesman felt chest pain while exercising in a local health club. Initially moderate in intensity, the pain gradually became increasingly severe, and soon he felt severe constriction around the chest which caused difficulty in breathing. The patient felt apprehensive because his father had died of heart attack two months earlier having suffered a similar agonizing experience.

The patient was rushed to the nearest nursing home, where he was admitted as a suspected case of myocardial infarction. ECG and X-ray of the chest were, however, normal. Blood sample was analyzed in the biochemistry laboratory with the following results:

Investigations test	Patient's results	Reference range
Lactate dehydrogenase (LDH)	1080 U/L	100–300 U/L (in males)
Creatine phosphokinase (CPK)	6010 U/L	30–120 U/L (in males)
Aspartate transaminase (AST)	130 U/L	0–35 U/L
Alanine transaminase (ALT)	48 U/L	0–48 U/L

CK-MB activity was normal. Electrophoretic separation of the LDH isoenzymes was carried out. The results are shown in the following graph.



The bands were obtained by scanning with a photoelectric densitometer. In this patient, sample for electrophoresis was obtained 8 hours after the chest paint.

Q.1. Would you recommend that the patient be shifted to coronary care unit on the basis of above test results?
Q.2. Interpret the results of electrophoresis.
Q.3. Which further tests would be of relevance in this patient?

CASE 6.5 A tight constricting band around chest in a hypertensive patient

A 48-year-old man experienced a sense of tightness in middle of the chest early in the morning (5.00 AM). It was like a tight 'constricting band', which lasted for a few minutes. It was accompanied by pain in the left arm, wrist and hand. He felt breathless and had vomiting during this period.

On examination, the patient showed pallor, profuse sweating and appeared out of breath. His heart rate was increased (102 per minute), but the pulse volume appeared low. Blood pressure was low, although he was known to be hypertensive for the last five years. His family physician made a diagnosis of myocardial infarction (MI) and had given him intramuscular diamorphine. He referred the patient to the hospital where the latter was hospitalized immediately. In casualty, the patient complained of chest pain and shortness of breath; however, the ECG results were not typical of myocardial infarction though they indicated ischaemic changes. Blood sample was taken soon (7.30 AM) for biochemical analysis.

Q.1. The results of enzyme activities, returned by the laboratory after a few hours are as below: CK 230 U/L (normal 30–120 U/L), AST 40 U/L (normal 10–35 U/L); LDH 218 U/L (normal 100–300 U/L). In view of these results, do you believe that the patient suffered an attack of myocardial infarction?
Q.2. Discuss the significance of isoenzyme estimation in this case.
Q.3. What other biochemical tests would be of relevance in this case?

CHAPTER

7

CHARACTERISTICS OF CELLS AND MEMBRANE TRANSPORT

Cell is the basic functional unit of all forms of life. It is the live unit containing a number of living components that are sequestered from the external environment by a membranous barrier. The cell membrane is selectively permeable since it permits passage of only selected substances across it, being impermeable to most others.

In prokaryotes, such as bacteria, algae and rickettsia, the cell itself is the organism. In some eukaryotes, like yeast and fungi, aggregation of the cells occurs to form tissues. In the higher forms of life, such as plants and animals, the tissues are further organized to form the whole organ systems. The crucial difference between the eukaryotic and the prokaryotic cell is the presence in the former of internal membrane structures, particularly the nucleus. Such structures are not seen in prokaryotic cells.

Most eukaryotic cells are 10–30 times longer in their linear dimension than the prokaryotic cells. For example, the *Escherichia coli* cell is about 1 mm long, whereas the hepatocyte is about 30 mm in diameter. Thus, the relative surface to volume ratio is much larger in this microorganism compared to that in the hepatocyte but the size is much smaller.

For all cell types, it is advantageous to have membrane surfaces on which various membrane-related phenomena, such as enzyme-catalyzed reactions and cellular transport takes place. Bacteria make use of the cell membrane for this purpose, whereas eukaryotic cells have acquired internal membranes for the same. The surface area of these internal membranes account for 96–98% of the total membrane area associated with the hepatocyte. However, exceptions do occur in eukaryotes. For instance, the plasma membrane of the intestinal epithelial cell has a large surface area, a very useful adaptation which enables it to absorb nutrients (Chapter 26).

This chapter outlines the structure and function of a typical eukaryotic cell. It also describes characteristics of biological membranes and membrane transport. After going through this chapter, the student should be able to understand:

- Chemical composition and architecture of a eukaryotic cell; functions of various subcellular components.
- Biological membranes: lipid bilayer structure, fluid mosaic model, the role of phospholipids, proteins (extrinsic and intrinsic), and other biomolecules in the bilayer structure; membrane asymmetry and fluidity, specialized membrane structures.
- Mechanisms for transport of substances across the permeability barrier, i.e. passive and carrier mediated transport, exocytosis and endocytosis, various types of channels and ionophores.

I. Cell Structure

Different types of cells vary enormously in their size, shape and specialized functions. Despite these outward differences, various kinds of cells are remarkably similar in their basic structural features. All cells are rendered self-contained by the surrounding *plasma membrane*, the basic molecular architecture of which consists of two layers of lipid molecules. A number of specialized proteins are embedded in the plasma membrane. Enclosed by the membrane there is the cytoplasm in which a vast array of chemical reactions of metabolism occurs. Also present in the cytoplasm is a *nucleus* or *nuclear body* in which genetic material is replicated and stored (as deoxyribonucleic acid). A number of organelles are also present, each of which

performs a specialized task; e.g. *mitochondria* are power plants of eukaryotic cells, *lysosomes* are pockets containing lethal enzymes, *peroxisomes* are peroxide-destroying vesicles, *Golgi bodies* are secretory vesicles, *ribosomes* synthesize proteins and endoplasmic reticulum form a maze of membrane channels through the cytoplasm.

A. Composition: Chemical Viewpoint

An average eukaryotic cell contains 70% or more water. About 10% of the weight of the cell is contributed by inorganic matter, and the rest is accounted by organic material.

Inorganic Constituents

Inorganic constituents of the cell comprise mineral elements. Although more than 100 mineral elements are found in nature, only 22 are present in cells. Most abundant of these are *carbon, hydrogen, nitrogen* and *oxygen*. Together, these four elements account for about 98% of the total mineral elements present in living systems. The relative abundance of these elements in the human body and earth's crust is shown in Table 7.1.

These four elements make up thousands of intracellular compounds. Most are highly reduced and energy rich. Other cellular inorganic constituents are cations like Na^+, K^+, Ca^{2+}, Fe^{2+}, Cu^{2+}, Mg^{2+} and Zn^{2+}; and anions like $C1^-$, SO_4^{2-} and HPO_4^{2-} About 1.0% of the weight of the cell is accounted for by inorganic matter.

Organic Constituents

The organic molecules in the cell, about 500 in number, are of diverse types. Some of them having relatively low molecular weight (MW 50–350 D) account for about 4.0% of the wet weight of the cell, e.g. monosaccharides, amino acids, fatty acids, nitrogenous bases, etc. The high molecular weight organic compounds (MW 2–2000 kD) are mainly proteins; the latter roughly accounts for 10–20% of the wet weight of the cell. Lipids, complex carbohydrates, DNA and RNA contribute approximately 3.0%, 3.0%, 1.0%, and 5.0% of the cellular wet weight, respectively. However, per cent composition of the above-mentioned cellular components varies depending on the cell type; a given cell may be rich in one type of bio-molecule but lack in others. For example, adipocytes contain about 98% triacylglycerols, and RBCs contain about 40% proteins.

B. Subcellular Organelles of Eukaryotic Cell

Ultrastructure of a cell showing various subcellular (internal) organelles is illustrated in Figure 7.1. Each organelle is specialized for performing a specific task (as outlined later in Table 7.3). Relative proportions of different organelles vary from one cell type to another. Moreover, a given organelle may carry out a certain function in a given cell type only. For example, release of glucose from glucose 6-phosphate can occur in hepatocytes (because of the presence of the enzyme, *glucose 6-phosphatase*), but not in cells of other tissues.

Table 7.1. Relative abundance of carbon and nitrogen in the living systems and earth's crust

	In human body (%)	In earth's crust (%)
Carbon	10.5	0.08
Hydrogen	60.3	Negligible
Nitrogen	2.42	0.0001
Oxygen	25.5	62.5

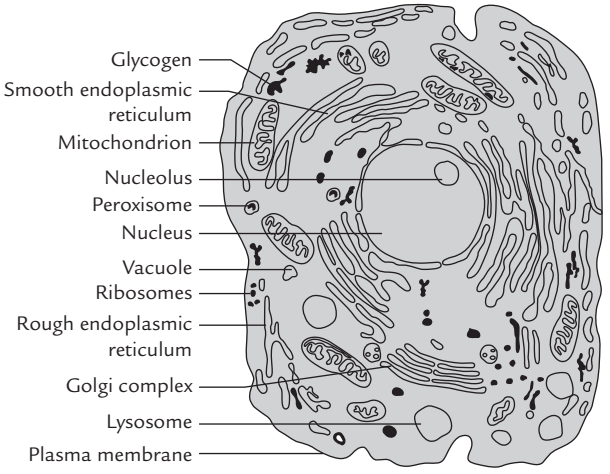


Fig. 7.1. Diagrammatic representation of a eukaryotic cell.

Cells in eukaryotes have a membrane bound nucleus, and a number of other membrane-bound subcellular organelles, each of which carries out a specific function.

Separation of Subcellular Organelles

The technique for separating subcellular organelles (subcellular fractionation) involves breaking open the cell by homogenization and separating various organelles from one another by centrifugation.

Separation of organelles for biochemical studies was first studied by Schneider and Pallade in late 1940s, and Albert Claude was awarded Nobel Prize in 1974 for fractionating subcellular organelles. The latter showed that different organelles, present in homogenate of a tissue, such as rat liver, could be separated by a procedure known as **differential velocity centrifugation**.

The following steps are involved in subcellular fractionation:

- The tissue is suspended in 0.25 M sucrose solution and the cells are disrupted by means of the shearing forces generated in a potter homogenizer.
- The homogenate thus prepared is then subjected to high-speed centrifugation. This process brings about separation of various organelles because different organelles sediment at different speeds in a centrifugal field, depending on their density and size. More the density and the size, more readily does the sedimentation take place. To some extent, shape of the subcellular organelle also determines its sedimentation.
- Centrifugal forces of sufficient magnitude and duration are used to produce separation of the organelles. Nucleus being the heaviest particle, sediments most readily. Centrifugation at 10^4 g/min results in pelleting of nucleus.
- Centrifugal force of 4×10^5 g/min results in sedimentation of mitochondria, lysosomes and peroxisomes (Fig. 7.2).

Separation of other organelles is then achieved in the same way, i.e. by increasing the centrifugal force.

At lower centrifugal forces, nuclei, mitochondria lysosomes and peroxisomes pellet to bottom of the centrifuge tube, whereas higher forces are required to pellet endoplasmic reticulum, Golgi apparatus and membranes.

Some enzymes, called the **marker enzymes**, are preferentially located in particular organelles, as shown in Table 7.2. They are used to identify purity of an organelle preparation: a given organelle is identified (after centrifugation) by measuring the activity of its marker enzyme in various subcellular fractions.

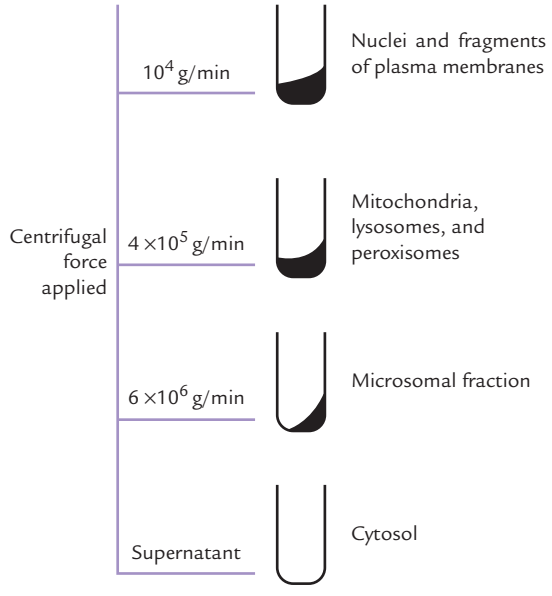


Fig. 7.2. Separation of subcellular organelles by differential centrifugation.

Why do Eukaryotic Cells Subcellular Organelles?

The subcellular organelles provide several benefits to the eukaryotic cell. They create compartments within the cell. Each compartment is specialized for a specific role. Major advantages of having compartments are as below:

1. *Enhances efficiency of reactions:* In the metabolic pathway, the reactions are sequential, which means that product of a reaction serves as substrate for the next reaction. Such reactions work much more efficiently if they are held in close proximity within a compartment.
2. *Storage:* Specific substance can be stored within a distinct intracellular compartment, from where it can be released when the need arises. For example, calcium ions are stored in sarcoplasmic reticulum and released to promote muscle contraction.
3. *Reciprocal regulation of opposing pathways:* The pathways that carry out opposing groups of reactions (such as fatty acid synthesis and degradation) do not occur simultaneously. A mechanism for their reciprocal regulation exists by which when one of the pathways is operative, the opposite pathway is inhibited. A smooth working of this arrangement is possible only when the opposite pathways are located in different compartments. For instance, the pathway for fatty acid synthesis is located in cytosol, whereas the pathway for its degradation is located in the mitochondrial matrix.
4. *The pH within an organelle can be maintained at a different level* than the rest of the cell (e.g. lysosomes). Lysosomes are membrane-bound digestive vesicles of varying sizes and shapes (average diameter = 0.5 μm) that contain lytic enzymes (called *hydrolase*) at a relatively low pH (4.5–5.5).

   The *hydrolases* cause intracellular digestion of macromolecules. Each lysosome contains over 30 *hydrolases*, which can digest all types of biomolecules (e.g. proteins, nucleic acids, carbohydrates, and lipids).

Table 7.2. Marker enzymes. A given marker enzyme is present within only one compartment of the cell

Subcellular organelle	Marker enzyme
Mitochondria	Inner membrane: ATP *synthase*
Lysosome	Cathepsin
Golgi complex	*Galactosyl transferase–mannosidase I (cis); Fucosyl transferase (trans)*
Microsomes	*Glucose 6-phosphatase*
Cytoplasm	*Lactate dehydrogenase*

Therefore, lysosomes are referred to as the potential *lethal bags* of the cell. Since the lysosomal membrane separates these enzymes from the cytosol, the cytosolic contents cannot be digested by them under normal circumstances. Even when the *acid hydrolases* leak out of the organelle because of membrane damage, they can cause only minimal damage to the cytosolic contents. This is because their optimum pH is low, and so they are readily inactivated at the relatively high pH prevailing in the cytosol.

Cytoskeleton

Electron microscopic studies have shown that a *complex network of protein filaments is present in the cytoplasm. The filaments form a flexible, framework within the cell, called* ***cytoskeleton***. Pivotal role is played by the cytoskeleton in a variety of cellular functions, such as intracellular transport, maintenance of shape of the cell, motility and cell division.

Three types of filaments are usually found in the mammalian cell: (a) **microfilaments** of 5 nm diameter, (b) **microtubules** of 25 nm diameter, mainly present in the long cells of the nervous system, and (c) myosin, which is mainly present in the muscle cells.

Eukaryotic cells have internal scaffold, the cytoskeleton, that control shape and movement of the cell.

Cytoplasm

The cellular matrix in which the subcellular organelles are embedded is known as cytoplasm. It accounts for approximately 50% of the cell volume. Cytoplasm was previously believed to be an inert jelly. However, it is now known that cytoplasm contains a variety of enzymes that participate in a number of metabolic reactions. The other components of the cytoplasm are glycogen granules and fat droplets.

Role played by various organelles is given in Table 7.3. For details the student is advised to refer to a textbook of cell biology.

Cytosol is soluble part of cytoplasm, wherein numerous metabolic reactions take place.

Table 7.3. Functions of subcellular organelles

Organelle	Function
Nucleus	Contains chromatin composed of DNA and proteins, RNA synthetic apparatus, and a fibrous matrix
Nucleolus	A nuclear subcompartment where most of the cell's rRNA is synthesized
Endoplasmic reticulum (smooth)	Site of biosynthesis of several biomolecules, detoxifies certain hydrophobic compounds
Ribosome	Protein synthesis
Golgi apparatus	Functions in the synthesis, processing and sorting out of secreted proteins, lysosomal proteins and certain membranes
Mitochondrion	Principal site of oxidative pathways (e.g. TCA, β-oxidation), ATP production, urea cycle (partly) and haem synthesis (partly)
Lysosome	Acidic organelle, contains a battery of lytic enzymes that degrade material internalized by the cell, and worn-out cellular membranes and organelles
Peroxisome	Oxidation of long chain fatty acids, D-amino acids and α-hydroxy fatty acids
Cytoplasm	Site of several metabolic pathways

II. Biological Membranes

Biological membranes are large flexible sheets that are universal elements of cell structure. Each cell is surrounded by a plasma membrane, which forms a boundary between the cell and its environment. It plays important role in:

(a) cell-cell recognition and communication,
(b) maintenance of the shape of cell,
(c) cellular movement and
(d) controlling movements of molecules between the inside and the outside of the cell.

Eukaryotic, but not prokaryotic cells, possess intracellular membranes as well, which encompass various cell organelles, thus segregating one part of the cell from the other and also enabling each organelle to carry out its characteristic cellular function.

Membranes are very active metabolically and have significant influence on cell metabolism. Most membranes are around 50 Å (5 nm) thick. In general, the membranes of organelles are thinner than the plasma membrane. All membranes have similar structural organization, which may be pictured as a *mosaic of globular proteins embedded* in a fluid-like phospholipid bilayer.

A number of diverse functions are performed by biological membranes, as outlined below:

Membrane lipids—Diffusion barrier, controlling movement of specific molecules, maintain shape of cell.

Membrane proteins—Enzymes, carrier activities, signal transduction, link between cytoskeleton and extracellular matrix.

Carbohydrates—Cell-cell recognition, adhesion and receptor action.

Biological membranes are diffusion barriers and sites of specific regulated transport around cells and intracellular compartments. They are also involved in cell recognition and intercellular communication, and have many associated enzymes.

A. Chemical Composition of Membranes

Lipids and **proteins** are the two major components of all biological membranes, as mentioned earlier. Relative proportions of the two vary greatly amongst different types of membranes. For example, lipids account for about 20% of the total weight of the rat liver cell membrane to over 54% in myelin sheath. In most membranes, 50–65% of the total membrane mass is accounted by proteins. Membranes of organelles have a higher percentage of proteins because of their greater participation in enzymatic and carrier (i.e. transport) activities. Inner mitochondrial membrane (IMM) contains highest proportion of proteins (75%).

Other membrane components, present in smaller quantities (5–8%) are **glycolipids** and **glycoproteins**. Carbohydrates do not exist in free form in membranes. Small quantities of free and esterified **cholesterol** are also present in most membranes.

The membrane lipids: phospholipids, glycolipids and cholesterol, are amphipathic molecules with hydrophilic heads on the outside and hydrophobic tails oriented inside.

B. The Lipid Bilayer

The bilayer structure of membranes is due to amphipathic nature of the major membrane lipids, i.e. the phospholipids. A phospholipid molecule is oriented in such a way that its polar head group is exposed on the external surface of the membrane, and the fatty acyl chain is oriented to the inside of the membrane. This forms a sheet-like phospholipid bilayer, which is two molecules thick (Fig. 7.3). The fatty acyl chains of the phospholipids in each layer, or leaflet, form a hydrophobic core that is 2–3 nm thick in most membranes. The diffusion barrier mentioned above is because of this hydrophobic core, which is highly impermeable to polar molecules and ions.

Understanding of the membrane structure has evolved gradually over a period of several years and still new information is being added. As a result, the proposed model for membrane structure continues to undergo modifications and refinements. It was in 1925 that *Gorter and Grendel* first proposed the **lipid bilayer structure** for the biological membranes. In 1935, *Davson and Danielle* suggested another model in which phospholipids were the major constituents that formed the matrix or continuous part of the membrane. Few globular proteins were also present, but only towards the polar exterior. A major question with these earlier models was as to how the membrane proteins interacted with the lipid bilayer. No satisfactory explanation was in sight, which casted doubt on general validity of these models.

In 1972, *Jonathan Singer and Garth Nicolson* proposed the **fluid mosaic model** (Fig. 7.4), in which some proteins called *intrinsic proteins* are actually immersed in the lipid bilayer, while others are loosely attached to the surface of the membrane, i.e. *extrinsic proteins*. The important features of the **fluid-mosaic model**, the most accepted model for the overall structure of biological membranes, and supported by a wide variety of experimental observations, are excerpted as below:

- The bilayer organization of lipids in membranes can be viewed as two dimensional solutions of oriented lipids; and there are proteins embedded in it. The proteins may be intrinsic (tightly associated with hydrophobic core) or extrinsic (loosely bound to membrane), as discussed.
- The integral membrane proteins can be considered as '*icebergs*' floating in two dimensional lipid '*sea*'. The bilayer organization of the lipids acts both as solvent for the amphipathic integral membrane proteins and as a **permeability barrier**.
- Membrane proteins are free to diffuse laterally in the plane of the bilayer unless restricted in some way, but would not be able to flip from one side of the bilayer to the other.
- Some lipids may interact with certain membrane proteins. These interactions are essential for the normal functioning of the protein.

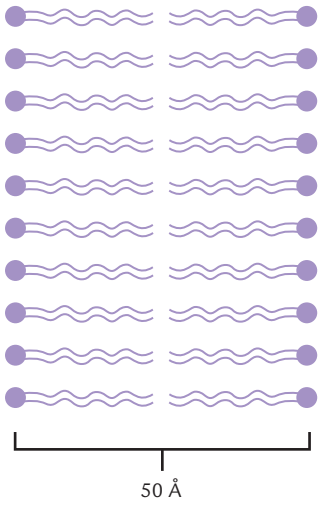


Fig. 7.3. The bimolecular leaf arrangement of phospholipids in biological membranes.

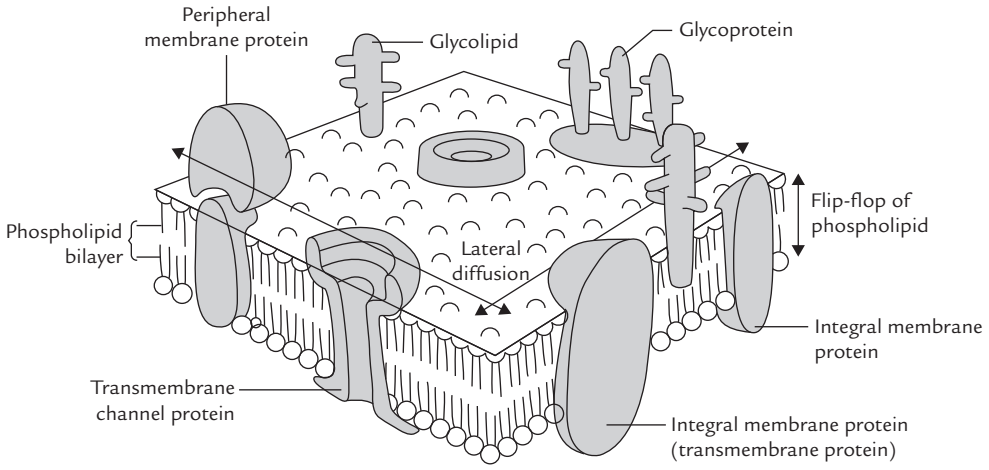


Fig. 7.4. Membrane structure: Fluid mosaic model (Singer and Nicolson).

Biological membranes are sheet like-structures that are only two molecules thick (60–100 Å). Basis of all cellular membranes is lipid bilayer onto which the globular proteins are embedded to form the fluid mosaic model.

C. Membrane Components

Membrane Lipids

Phospholipids, sphingolipids, and cholesterol constitute the major membrane lipids.

Phospholipids: These are the most predominant molecular components of all membranes. The principal membrane phospholipid is **phosphatidylcholine (lecithin)**, which accounts for 40–50% of the total phospholipid content. Other phospholipids, present in relatively small quantities, are *phosphatidylethanolamine* (*cephalin*), *phosphatidylserine, cardiolipins* and *inositol phosphoglycerides*.

Sphingolipids: These derivatives of sphingosine, an amino alcohol, with a long hydrocarbon chain, are the second important class of lipids found in biomembranes. The *sphingomyelins*, which contain a phosphocholine head group, are major components. Other sphingolipids are glycolipids in which a single sugar residue or branched oligosaccharide is attached to the sphingosine backbone (Chapter 3). Glycolipids constitute 2–10% of the total lipids in plasma membranes; they are most abundant in nervous tissue.

Cholesterol: The third important class of membrane lipids is cholesterol and its derivatives. Though cholesterol is especially abundant in the plasma membrane of mammalian cells, it is absent from most prokaryotic cells. It appears almost entirely hydrophobic in composition, but it is actually *amphipathic* because its hydroxyl group can interact with water. It is oriented in such a way in the membrane that its hydrophilic hydroxyl group faces the exterior and the cyclopentanophenanthrene ring fits into the hydrophobic lipid phase of the membrane.

In membranes three major classes of lipids are phospholipids, sphingolipids and cholesterol. The (glycero-) phospholipids have glycerol backbone and the sphingolipids are based on sphingosine, an amino-alcohol. Cholesterol is most rigid lipid in animal cells; in plants β-sitosterol is present instead.

The proportion of lipid component of different membranes vary. Neutral lipids and sphingolipids occur in high concentration in plasma membranes. The inner mitochondrial membrane (IMM) is especially rich in cardiolipins and phosphatidylethanolamine. High concentration of sphingolipids is present in myelin sheaths of axons of neural tissue, whereas the intracellular membranes primarily contain phospholipids with relatively smaller amount of sphingolipids or cholesterol. The cholesterol content of a given membrane may vary with the nutritional status of an individual.

Membrane Proteins

The membrane proteins are involved in a number of specific functions. While the continuity of biological membranes, and their ability to act as permeability barrier for water-soluble molecules are due to the properties of the lipid bilayer, other membrane functions can be explained by properties of the membrane proteins. The most important of these are:

1. *Enzymatic activities:* A number of biochemical reactions are catalyzed by the membrane proteins which

possess enzymatic activity; e.g. Na^+-K^+ ATPase that maintains a steep gradient of Na^+ and K^+ across the plasma membrane.

2. *Carrier activities:* Several water-soluble intermediates, nutrients and waste products must be carried across the lipid bilayer. Simple diffusion of these substances is not possible due to selectively permeable nature of the membrane. Their movement across the membranous barrier is made possibly by certain transport proteins (i.e. *translocases*).
3. *Signal transduction:* The message inherent in a hormone or neurotransmitter or other such extracellular substances must cross the membranous barrier so that it may influence intracellular events. This is made possible by signal transduction by specific proteins (Chapter 29).
4. *Interactions with the extracellular matrix and the cytoskeleton:* These interactions are essential for the integrity of the cell and of the tissue as a whole. Proteins such as fibronectin, ankyrin and spectrin participate in these interactions.
5. *Some membrane proteins regulate permeability for inorganic ions:* This forms the basis for the excitability properties of some biological membranes.

Proteins are in charge of specialized functions, acting as enzymes, as structural links between the cytoskeleton and the extracellular matrix, as components of signalling pathways, and as carriers.

Peripheral and Integral Proteins

As mentioned earlier, the membrane proteins are divided in two classes: peripheral (or extrinsic) and integral (or intrinsic). The classification is based on the location and the ease with which they can be removed from the membrane.

The **peripheral proteins** can be released from the membrane by relatively mild treatment such as by salt solutions of different ionic strengths, or through alteration of the surrounding pH. These proteins usually possess enzymatic activity. They may be bound to either face of the membrane by electrostatic interactions and hydrogen bonds (Fig. 7.4). Removal of peripheral proteins causes little or no change in the integrity of the membrane.

Each membrane contains a unique complement of proteins, which may be intrinsic (embedded in the membrane) or extrinsic (bound to it). Carbohydrates in form of oligosaccharide chains are attached to lipid or protein.

The **integral proteins** are embedded deeply in the bilayer, and are attached by hydrophobic bonds or van der Waals forces. Mostly they span the whole bilayer and when they do so they are called ***transmembrane proteins*** (Fig. 7.4). They serve as important transport proteins.

It requires use of detergents or organic solvents to separate the integral proteins from the membrane. The integral proteins are usually tightly bound with lipids. Removal of the lipid component during the separation causes denaturation of the protein, and consequent loss of its activity.

Note: Some membranes contain a specific type of hydrophobic proteins called **proteolipids**, which consist of covalently linked lipid and protein components. These proteins are particularly abundant in **myelin**, where they represent about 50% of the membrane proteins, e.g. lipophilin.

Other Membrane Components

Carbohydrates are relatively minor components, forming 5–8% of the total membrane mass. Most carbohydrates occur as oligosaccharide units that form part of membrane **glycolipids** and **glycoproteins**. Glucose, galactose, mannose, fucose, N-acetylglucosamine, and sialic acid are some of the major carbohydrates present in these hybrid molecules. It is noteworthy that carbohydrates never exist in free form in the membranes.

The carbohydrates are generally located towards the exterior of plasma membrane. They play an important role in cell-cell recognition, adhesion and receptor action. However, they may be located in the inner leaflet of the membrane, e.g. in endoplasmic reticulum where they face the membrane enclosed channel.

D. Membrane Asymmetry

Membranes have asymmetric structure. Proteins are mainly responsible for it because they are inserted in an asymmetric fashion (Fig. 7.4). Further contributors to asymmetry are the oligosaccharide units that always project towards the exterior. Lipid components are also distributed in an asymmetric fashion, for example, in the erythrocyte membrane, the outer leaflet of the bilayer contains mostly phosphatidylcholine and sphingolipids, whereas the inner one contains phosphatidylethanolamine and phosphatidylserine.

The fluid mosaic model depicts asymmetric membrane structure with different lipid compositions in two leaflets. Further contributors to membrane asymmetry are oligosaccharides that always project towards exterior and membrane proteins having only one orientation.

E. Membrane Fluidity

The fact that some membrane components can move in the lipid bilayer indicates fluid nature of biological membranes. A number of factors influence the fluidity of membranes, which in turn influences its physiological function. Temperature and composition of the membrane are the two major factors that influence the membrane fluidity. At a relatively low temperature the fluidity is less, and therefore, limited mobility of membrane components is possible. As the temperature is raised, a phase transition occurs that results in tremendous increase in fluidity. The **phase transition temperature (melting temperature; T_m)** provides a measure of the temperature at which the transition occurs. The T_m value itself depends on the membrane composition.

The nature and relative abundance of membrane constituents have significant effect on the membrane fluidity. For example:

- The short chain fatty acids increase the fluidity while the long chain fatty acids tend to decrease it.
- Unsaturated fatty acids with their *cis*-double bonds do not pack closely; this results in more fluid structures in which the T_m is lowered. More the number of double bonds, greater is the fluidity of the membrane (and hence lower is the value of T_m).

Cholesterol content of the membrane also alters the fluidity of the membrane. With its rigid ring system, it decreases the fluidity of most membranes. It can, however, also insert itself between the fatty acid chains and prevent their crystallization. In this respect, cholesterol acts like an impurity that decreases the melting point of a chemical.

F. Specialized Membrane Structures

Fibronectin (fibro = fibre, nectin = connect): This is a protein (600 Å long) that permits binding of the cell with components of the extracellular matrix. It is made up of two polypeptide chains (250 kD each), which are linked by disulphide bridges at their carboxyl terminals. It contains distinct domains, each one of which is capable of binding specific molecules, as shown in Figure 7.5. Malignant cells are deficient in fibronectin which is responsible, at least partly, for their invasive character.

Several important functions are performed by fibronectin. Fibroblasts and other such cells, involved in repair mechanisms, adhere to a clot by attaching themselves through fibronectin.

Tight junctions: These are present between the two cells that lie in close approximation. They form narrow hydrophilic channel through which calcium and other small molecules pass from one cell to another. As shown in Figure 7.6, only three (and not four) layers of plasma membrane are present at the tight junctions.

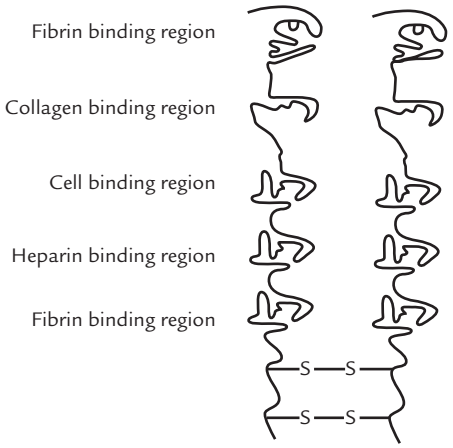


Fig. 7.5. Binding regions of fibronectin.

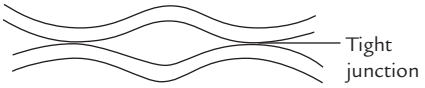


Fig. 7.6. Tight junctions between cells.

Desmosomes: These provide attachment to cells on the basal tissues. They are mostly seen in epithelial cells.

G. Red Blood Cell (Erythrocyte) Membrane

The erythrocyte membrane is ideal for studying membrane structure because it is easily isolated without any contamination with other subcellular membranes. Study of this membrane has given insight into the role of various membrane components, particularly the membrane proteins. Three major proteins—glycophorin, anion-exchanger and ankyrin are present, in addition to several other minor proteins.

Glycophorin: It is an integral glycoprotein (MW 30 kD), with several oligosaccharide units attached to the amino-terminal, which faces extracellularly (Fig. 7.7). These oligosaccharides determine the antigenic specificity of the membrane. Function of glycophorin is still not fully understood because the individuals lacking it to do not suffer from any specific abnormality.

Anion exchanger: It is an integral glycoprotein, accounting for one-third of the total membrane proteins (MW 106 kD). It facilitates extrusion of bicarbonate ions from erythrocyte in exchange for chloride ions. This transport mechanism plays an important role in acid-base homeostasis.

Ankyrin: It is a peripheral protein located towards cytoplasmic side of the erythrocyte (Fig. 7.7). It is cross linked to another peripheral protein called **spectrin**, which is filamentous in nature and about 1000 Å long. These two interconnected proteins, together with **actin**, form a mesh-work that underlies the erythrocyte membrane. This mesh-work is responsible for the resilience and stability of the erythrocyte membrane.

H. Micelles and Liposomes

When a suspension of pure phospholipids or a mixture of phospholipids is mechanically dispersed in aqueous solution, the phospholipids aggregate into one of the two spherical forms: **micelles** or **liposomes**. In both the structures, the hydrophobic effect causes the fatty acyl chains to aggregate and exclude water molecules from the core. The type of structure formed depends on several factors, including the length of the fatty acyl chains, their degree of saturation, and temperature.

Micelles

These are small globular structures with hydrophilic surface and hydrophobic core (Fig. 7.8a). They can be formed by all polar lipids, including ordinary detergents. However, micelles are rarely formed from the natural phosphoglycerides, whose fatty acyl chains are too bulky to fit into the interior of a micelle.

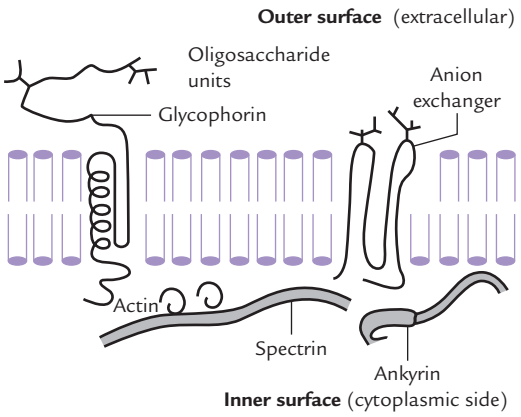


Fig. 7.7. Structure of erythrocyte membrane.

Liposomes

A lipid bilayer may close in on itself, forming a spherical vesicle separating the external environment from an internal compartment. Such vesicles are termed liposomes (Fig. 7.8b). They have diameter in the range of 50–1000 nm. Wall of a liposome consists of two layers of lipids and encloses an inner aqueous compartment in which small ions or molecules can be trapped. Liposomes can be artificially prepared by agitating a polar lipid such as lecithin suspended in water, with high frequency sound waves.

Biomedical importance of liposomes

1. Liposomes are used to study membrane properties and their role as permeability barrier. They are also used to introduce a number of impermeant substances into the cells.
2. Liposomes have important therapeutic applications because they can carry various drugs and enzymes.

Liposomes are prepared with specific drugs or enzymes encapsulated inside. They carry these substances to their target tissues. The biodegradable nature of the liposomes permits release of the encapsulated substance within the vicinity of the target tissue. Many antibiotics, antiviral, antifungal and anti-inflammatory agents have been found to be highly effective when delivered in this fashion. It may be possible to prepare liposomes with a high degree of tissue specificity.

Liposomes are aqueous compartments enclosed by a lipid bilayer, that can be used to study membrane properties or to deliver drugs to specific target tissues.

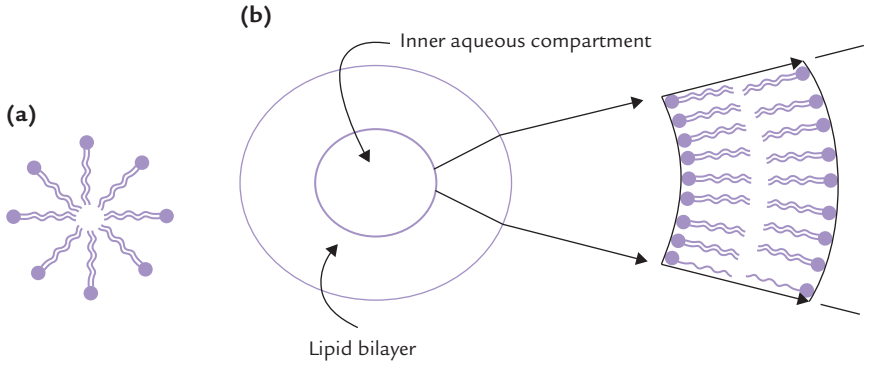
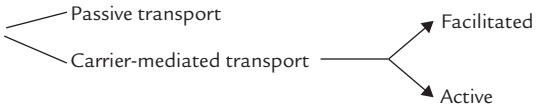


Fig. 7.8. Diagrammatic representation of **(a)** micelle, and **(b)** liposome.

III. Transport Across Cell Membrane

Biological membranes mediate interaction of the cell with environment. Several useful substances present extracellularly enter the cells through the membranous barrier. Likewise, the intracellular waste products leave the cell by moving across the membrane. During the transport across the membrane, these substances encounter varying degrees of resistance—some can freely diffuse through the membrane (**passive diffusion**), whereas movement of others is restricted and requires a carrier, usually integral membrane proteins (**carrier-mediated transport**).

Some types of carrier-mediated transport processes occur along the concentration gradient (**facilitated transport**); others are driven by hydrolysis of ATP (**active transport**) against a concentration gradient.



Transport of molecules through membrane occurs either by passive diffusion, or via mediation of integral membrane proteins: the mediated transport can be along the concentration gradient (facilitated), or against the concentration gradient (active).

A. Passive Diffusion

It is a spontaneous, unaided movement of the solute particles along their concentration gradient. The particles move from the higher concentration to the lower concentration, till they equilibrate. At equilibrium point, the entropy reaches a maximum possible level under the prevailing circumstances, and therefore, no further movement of the solute particle occurs.

Permeability coefficient *is an expression that indicates relative ease of diffusion of a given substance through the lipid bilayer.* Figure 7.9 represents the permeability coefficient values of some substances in a descending order. The membrane is: (i) highly permeable to gases (e.g. CO_2, NO, and O_2), small hydrophobic molecules and small uncharged polar molecules (e.g. steroids, ethanol, and urea); (ii) moderately permeable to water (permeability coefficient 10^{-2} cm/sec); and (iii) essentially impermeable to ions and to large polar molecules (sugars, amino acids, etc.), which have extremely low permeability coefficients, and require presence of specific transport mechanisms to move across the membrane.

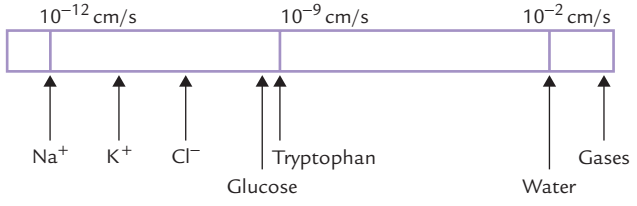


Fig. 7.9. Permeability coefficients of some polar substance in descending order (right to left). Some molecules (gases, water, urea) can pass directly through the bilayer unaided, whereas other molecules (glucose, amino acids, ions, etc.) require presence of transport proteins.

B. Carrier-mediated Transport

It occurs through mediation of integral membrane proteins, which have been variously known as **permeases, porters** or **translocases.** These proteins are highly specific. For example, the erythrocyte glucose transporter has high affinity for D-glucose, but low affinity for the related sugars, D-mannose and D-galactose (10–20 times lower).

The carrier-mediated transport may be facilitated or active; both are specific for the solute transport and both can be inhibited by structural analogues. For example, 1,5-anhydroglucocitol can inhibit transport of glucose across the membrane. When a solute molecule is transported singly, the system is known as **uniport**. In some cases, transport of a molecule requires an obligatory simultaneous cotransport of another molecule, either in the same direction or in the opposite direction. When both substances move in the same direction, the process is known as **symport**. Conversely, when they move in opposite directions, process is called the **antiport** (Fig. 7.10).

Facilitated Diffusion

In ***facilitated diffusion*** *movement of particles occurs along their concentration gradient.* It requires mediation of specific integral membrane proteins for "facilitating" the movement of the solutes (glucose, other sugars, amino acids). The transport proteins serve to hasten the process so that equilibrium is attained far more rapidly than it occurs in passive diffusion.

Mechanism of action of these proteins is not clearly understood. One possible explanation is that a transport protein oscillates between two conformations. In one conformation, the solute-binding site is exposed towards the extracellular space (Fig. 7.11a). When the solute binds to this site, a conformation change is induced in the transport-protein, which permits the solute to be transferred to the cytosolic side of the membrane (Fig. 7.11b and c). Release of the solute on this side occurs next because its affinity for the transport protein in the changed conformation is low (Fig. 7.11d). The release triggers the return of the protein to its original conformation. This whole process is reversible, meaning that the solute can be

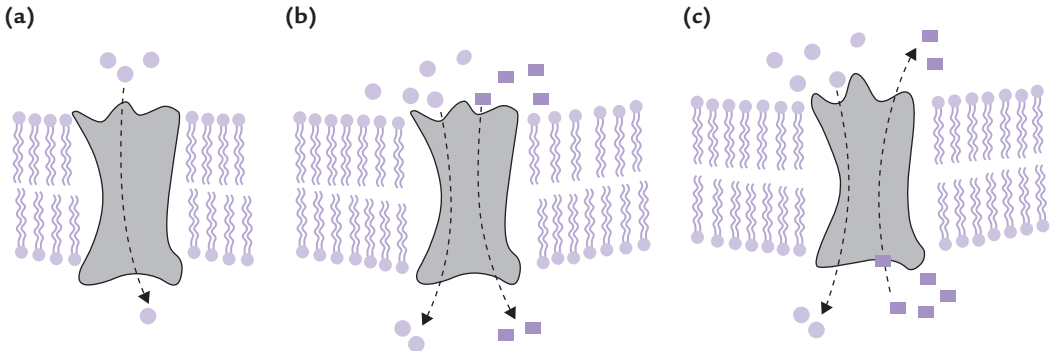


Fig. 7.10. Schematic representation of **(a)** uniport, **(b)** symport, and **(c)** antiport.

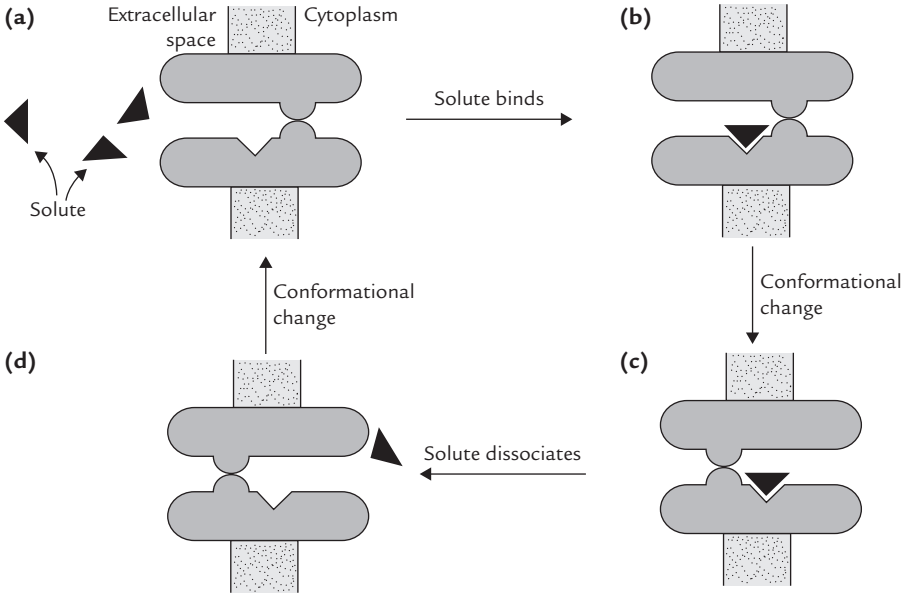


Fig. 7.11. Schematic representation of faciliated transport. There is no external energy source, so the net transport is down the concentration gradient.

extruded from the cell by reversal of the above events. The direction of movement depends on the relative concentration of solute on two sides of the membrane.

Kinetics of facilitated transport obeys the Michaelis-Menten rate law. A plot of the initial rate of transport (V_0) versus the concentration gradient, is **hyperbolic** (Fig. 7.12). As the concentration gradient increases, the rate of transport also rises. Rate of transport depends on the per cent saturation of the carrier protein. It rises to maximum (V_{max}) when all the carrier protein molecules are fully saturated with the solute. Thus, systems subjected to facilitated diffusion are saturable and the K_m value for the transport protein corresponds to the concentration of solute at which the carrier protein is half-saturated.

The transport protein involved in facilitated transport is specific for a particular molecule, is saturable, displays binding kinetics and is influenced by inhibitors, pH and temperature.

Certain well characterized transport systems are discussed below.

Glucose Transporters

Glucose is transported across the cell membrane by facilitated diffusion system. The glucose transport family comprises five members, named GLUT-1 to GLUT-5. They shuttle between two conformational states, one in which the substrate-binding site faces outward and the other in which the binding site faces inward. All transmembrane proteins are similar in size, having about 500 amino acid residues and 12 transmembrane helices. Their primary structures are remarkably similar, and they display a tissue-specific pattern of expression; for example, GLUT-1 is abundant in erythrocytes and GLUT-4 in muscles.

- GLUT-1 molecule in erythrocytes has a K_m of 15–20 mmol/L; it is mostly active under fasted state.
- GLUT-2 in pancreatic cells has higher K_m (10 mmol/L), which, in response to raised blood glucose

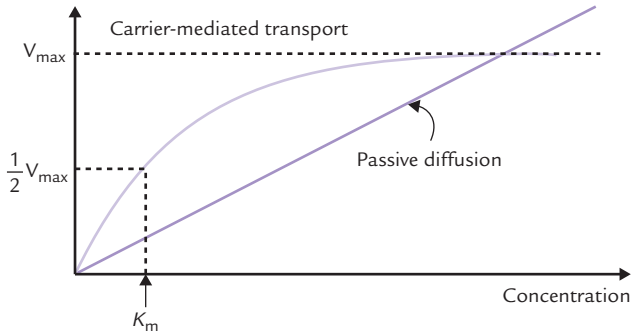


Fig. 7.12. Kinetics of facilitated diffusion by a transport protein. The protein is saturable and displays Michaelis–Menten-type binding kinetics in a similar manner to enzymes. On the other hand, in passive diffusion. The rate of transport is directly proportional to concentration gradient.

concentration, mediates an increase in the cellular uptake of glucose, leading to insulin secretion.

- GLUT-4, present in insulin sensitive tissues like muscles and adipose tissue, is translocated from intracellular vesicles to the plasma membrane by insulin. This facilitates glucose uptake during meals.
- GLUT-3 and -5 are found in brain and testis; and in intestinal epithelial cells, respectively. Their properties have not yet been elucidated in detail.

Chloride Transporters

There are two major types:

1. *Cl^-/HCO_3^- antiport*: This transport system is located on the erythrocyte membrane. It moves the bicarbonate ions, generated intracellularly by action of *carbonic anhydrase*, out of the erythrocytes with a concomitant movement of chloride ions into the cell. In this way, electroneutrality is maintained (Chapter 1).
2. *Cystic fibrosis transmembrane conductance regulator (CFTR):* This carrier protein is located on cells of exocrine glands. It forms a cAMP regulated Cl^- channel, through which the movement of chloride ions across the epithelial cells occurs. Defective action or faulty localization of CFTR in cell membrane leads to diseases (**Case 7.1**).

Others: Specific transport proteins for a number of other substances, including amino acids, have been identified. These transporters facilitate movement of various amino acids across renal tubular and intestinal mucosal cell membranes. Defective transport may lead to a number of clinical disorder, as exemplified in **Case 13.4**.

Active Transport

In ***active transport,*** *the solute molecules are moved against a concentration gradient through expenditure of energy*. Therefore, active transport is coupled to the energy state of the cell. The active transport is classified as primary and secondary.

The **primary active transport systems** use energy obtained by hydrolysis of ATP to drive transport of the molecules (Na^+ K^+ Ca^+ or H^+) across the membrane. Some examples are: the Na^+-K^+ ATPase, the proton-pump in stomach, etc.

In **secondary active transport systems** the movement of the substrate molecule across the membrane is coupled to the movement of an ion (Na^+, H^+) down its concentration gradient. Thus, a pre-existing ionic gradient of Na^+ or H^+ is used to drive the "*uphill*" movement of a substrate. Example: Coupled movement of glucose and Na^+ across the intestinal cells.

The energy source for active transport may be ATP or a pre-existing ion gradient. When ATP-driven, it is primary active transport; and when ion-gradient-driven, it is secondary active transport.

Primary Active Transport

In this section, three transport systems involved in the primary active transport are discussed. These are: Na^+-K^+ dependent ATPase for antiport of sodium and potassium ions, Ca^{2+}-dependent ATPase for moving calcium ions, and proton pump in the stomach that antiports protons and potassium.

Na^+-K^+ dependent ATPase: The *Na^+-K^+* dependent ATPase was first detected in the crab nerves by Jens Skou in 1957. It was named so because it requires both Na^+ and K^+ (apart from Mg^{2+}), and uses ATP energy for its action. The Na^+-K^+-dependent ATPase is an antiporter which actively pumps K^+ into the cell from extracellular fluid (ECF), with a concomitant extrusion of Na^+ from the cell. It is located in the plasma membrane of all cells, with highest activity in nervous tissue and muscle. It maintains the steep gradients of sodium and potassium across the plasma membrane, as shown in Figure 7.13. It

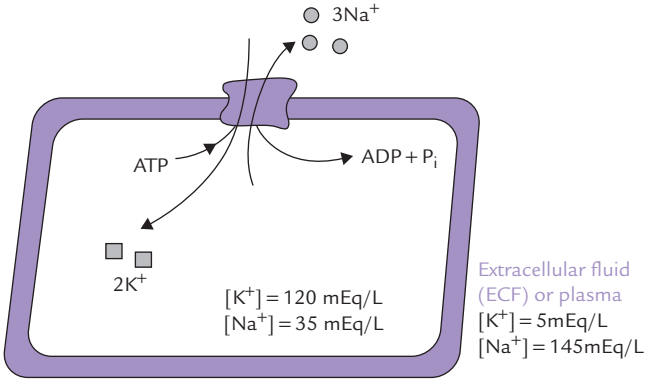


Fig. 7.13. A tentative model of Na^+–K^+ ATPase, also showing concentration gradients of sodium and potassium across the biomembrane.

is a major consumer of metabolic energy, hydrolyzing about 100 molecules of ATP per second under optimal conditions. *About a third of total energy requirement of the cell is used to drive the Na^+–K^+ ATPase;* in electrically excitable cells, it accounts for about 70% of the cell's energy. Thus, a significant proportion of the basal metabolic rate (20–25%) is allotted to sodium-potassium pumping.

Na^+–K^+ pump maintains high K^+ and low Na^+ inside the cells, using a significant proportion of the cellular energy.

Subunit structure: The Na^+–K^+ ATPase is an oligomeric integral membrane protein composed of two types of subunits; the stoichiometry is $\alpha_2\beta_2$ (Fig. 7.14).

The **α-subunit** (MW = 112,000) brings about hydrolysis of ATP. In addition, each α-subunit has binding sites for Na^+ on the cytosolic side and for K^+ towards the extracellular side.

The **β-subunit** is smaller (MW = 55,000). It is a glycoprotein, whose function is unknown.

Sequence of ionic pumping activities: Three sodium ions are ejected for every two potassium ions pumped into the cell, and one ATP molecule is hydrolyzed in the process. The sequence of ionic pumping activities (that proceed in **four steps**) is represented in Figure 7.15.

Step I: The sodium ions bind with α-subunits towards cytosolic side of the latter. Binding of sodium induces phosphorylation of this subunit (termed Na^+-*induced phosphorylation*); the phosphate group is obtained from ATP. This group gets attached to a specific aspartyl residue of the α-subunit.

Step II: The phosphorylation in turn induces a change in the conformation of the pump protein (from E_1 to E_2 form), whereby its orientation changes in such a way that its ion-binding site is everted (Fig. 7.15). The ion-binding site now faces the exterior, and consequently the sodium ions are moved out into the extracellular fluid (ECF).

Step III: The phosphorylated protein has low affinity for sodium, so that these ions are released into ECF. Binding affinity of the phosphorylated protein is high for K^+ and so it binds these ions. Two K^+ ions attach with specific binding sites present on the extracellular side of the α-subunit.

Step IV: The K^+ ions trigger release of the phosphate ions (termed K^+-*induced dephosphorylation*). The dephosphorylation favours the E_1 conformation so that the pump protein reverts to the original E_1 conformation. With change of conformation ($E_2 \rightarrow E_1$), the ion-binding site is inverted so that potassium ions move into the cytoplasm. Intracellularly they are released on account of low affinity of the E_1 conformation for K^+.

The oppositely directed movements of Na^+ and K^+ are on account of (a) **oscillation** of the pump between an inside open conformation with a high sodium affinity and an outside open conformation that has a high affinity for potassium, and (b) **stabilization** of the inside open conformation by dephosphorylation and of outside open conformation by phosphorylation.

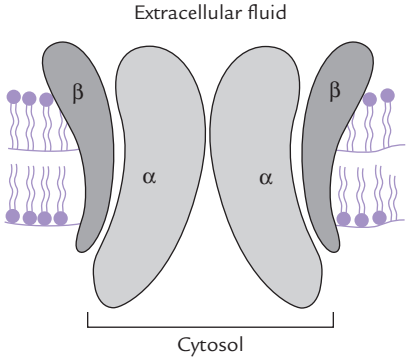


Fig. 7.14. Structure and orientation of Na^+–K^+ ATPase.

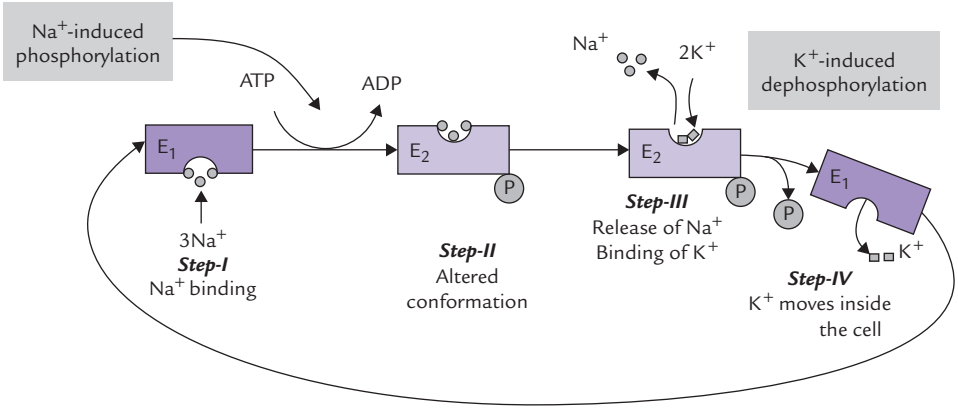


Fig. 7.15. Sequence of ionic pumping activities of the Na^+—K^+ ATPase. E_1 and E_2 refer to protein conformations in dephosphorylated and phosphorylated forms, respectively. E_1 is inside open conformation with high affinity for Na^+; and E_2 is outside open conformation with high affinity for K^+.

Inhibitors of Na^+-K^+ ATPase

(a) *Ouabain*, a steroid extracted from plants, inhibits activity of Na^+–K^+ ATPase. It binds to the extracellular face of the α-subunit and prevents the K^+-induced dephosphorylation process. This disrupts the functions of the Na^+–K^+ ATPase resulting in inhibition of the ion translocation.

(b) *Digitalis* is another steroid of plant origin (obtained from *Digitalis purpurea*), which is therapeutically used for its ability to increase contractile activity of the cardiac muscle (i.e. cardiotonic effect). This action is due to the inhibitory effect of digitals on the Na^+–K^+ ATPase. Inhibition of the pump activity results in a raised intracellular concentration of Na^+ due to decreased extrusion of this ion. Consequently, the Na^+ gradient across the cell membrane falls. The diminished gradient in turn causes building up of the intracellular calcium concentration, which enhances the force of contraction.

How does the diminished sodium gradient lead to rise in cytosolic calcium concentration? The major mechanism for removal of calcium out of the myocardial cell is by a Ca^{2+}/Na^+ antiporter in the plasma membrane. This antiporter depends on sodium gradient for effective calcium transport. With the rise in the cytosolic sodium concentration, the sodium gradient diminishes, and more calcium remains in the cytoplasm.

Digitalis is very toxic, and overdosage results in fatal arrhythmias because a normal sodium gradient is essential for the normal excitability of the myocardial cells.

Ca^{2+}–dependent ATPase: The *Ca^{2+}*–dependent ATPase is present in most membranes but is most abundant in the sarcoplasmic reticulum of skeletal muscles, where it accounts for about 80% of the integral membrane proteins. *It actively moves calcium from the cytosol into the sarcoplasmic reticulum against the concentration gradient*. The process requires expenditure of ATP energy.

Sarcoplasmic reticulum is a repository of calcium ions in the resting muscle. A nerve impulse causes release of these ions into the cytosol. This results in about 100-fold increase in the cytosolic calcium concentration. The cytosolic calcium in turn induces muscle contraction. Subsequent muscle relaxation depends on removal of calcium from the cytosol. It is at this stage that the Ca^{2+}-dependent ATPase comes into play. It removes calcium from the cytosol by pumping it into the sarcoplasmic reticulum.

The Ca^{2+}-dependent ATPase resembles the Na^+–K^+-dependent ATPase in three aspects.

(i) Both can exist in two conformational states, which depends on the state of phosphorylation. Conformation of the phosphorylated form is different from that of the dephosphorylated form.

(ii) Both contain a specific aspartyl group to which attachment of the phosphoryl group occurs.

(iii) The phosphorylation–dephosphorylation cycle is responsible for the oscillation of these transport proteins between two conformational states.

Proton pump in the stomach: is expressed in gastric parietal cells in response to stimuli such as histamine and gastrin. The pump antiports two cytoplasmic protons and two extracellular potassium ions, coupled with hydrolysis of a molecule of ATP. The protons are moved against a steep concentration gradient (cytoplasmic concentration, 10^{-6} M vs luminal concentration, 0.15 M). With the movement of H^+, Cl^- is secreted through a chloride channel, producing hydrochloric acid (HCl) in the lumen.

Secondary Active Transport

It is an **ion-driven cotransport** mechanism, in which movement of a substrate against its concentration gradient is coupled to movement of another ion (Na^+ or H^+) along its concentration gradient. An example is glucose-Na^+ symport. As shown in Figure 7.16, Na^+ is moving along its concentration gradient, and glucose against the gradient. Though not coupled to ATP hydrolysis, this method of transport is classified as **active transport** because there is energy expenditure: the energy is derived from electrochemical gradient of Na^+ (Table 7.4). It is primarily used for absorption of glucose as also amino acids in the intestinal mucosa. It is also employed in the proximal renal tubules for the reabsorption of glucose and amino acids from the primary filtrate. Indeed, kidney and intestine often use the same sodium cotransporter, and therefore, many transport defects are expressed in both the organs.

Knowledge of the coupled movement of Na^+ and glucose forms the basis of glucose rehydration therapy in

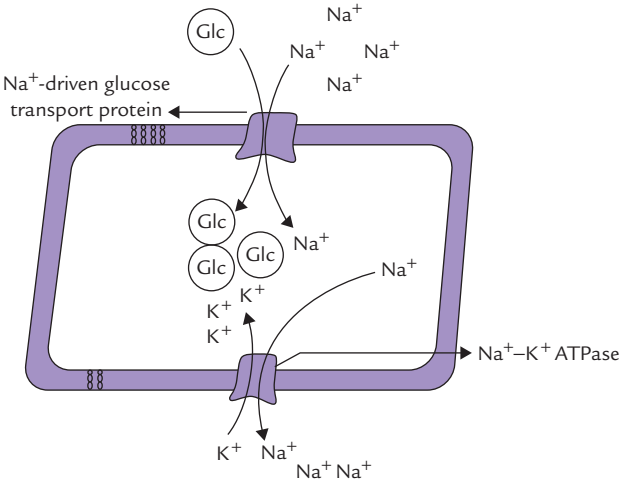


Fig. 7.16. Secondary active transport: coupled movements of Na^+ and glucose. The Na^+ is subsequently moved out of the all by action of a Na^+/K^+ ATPase on the basolateral membrane.

Table 7.4. Comparative features of various transport mechanisms

Property	Passive diffusion	Carrier-mediated transport		
		Facilitated diffusion	Primary active transport	Secondary active transport
Carrier required	–	+	+	+
Transport against gradient	–	–	+	+
ATP hydrolysis	–	–	+	–
Energy required	–	–	+	+

cholera. Oral administration of glucose prevents loss of Na^+ in the diarrhoea fluid to a considerable extent. Glucose transport is powered by Na^+ concentration gradient in renal tubules and choroid plexus as well.

A transport protein couples movement of Na^+ down its concentration gradient to the movement of glucose against its concentration gradient.

Models for Action of Transporters

The **transport proteins** are a special class of integral membrane proteins that bind ligands on one side of the membrane with high affinity and specificity, and release them on the other side. Details about the transport processes at the molecular level are not clear. Following are the conceptual models that have been proposed to explain some of these processes.

Mobile carrier system: Binding of the ligand physically alters the orientation of the transport protein, possibly by rotation through plane of the membrane. As a result, the ligand moves to the other side of the membrane (Fig. 7.17a). This is also referred to as the **ping-pong model**.

Shuttle system: The transport protein, after binding with the ligand, shuttles back and forth across the lipid bilayer (Fig. 7.17b).

Transmembrane pore system: Some transport proteins (referred to as the **pore ionophores**) contain a pore through which specific substances are transported (Fig. 7.17c). The system is firmly embedded in the membrane because the hydrophobic side chains of the constituent amino acids intercalate with the membrane matrix. Only certain selected substances can move through the pore. This model is also visualized as being guarded because it opens only when such substances arrive.

Evidence for existence of these models comes from studies with the following antibiotics: *valinomycin*, a diffusion (shuttle) type ion transporter, and *gramicidin*, a pore type ion transporter. These antibiotics are also known as the **ionophores** since they permit greater mobility of ions across the membranous permeability barrier. They are important experimental tools for studying the transport

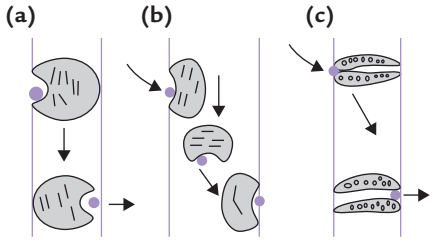


Fig. 7.17. Models for action of transporters. **(a) Mobile carrier** system, **(b) Shuttle** system, **(c) Transmembrane pore** system.

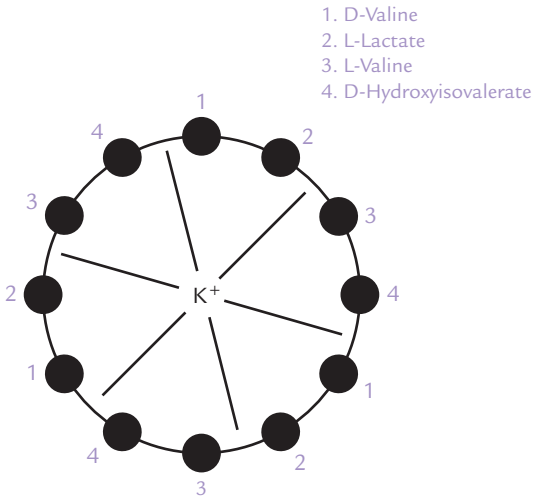


Fig. 7.18. Structure of valinomycin, holding K^+ in the centre. It is a shuttle type ion transporter.

phenomena, in addition to being important therapeutic agents.

Valinomycin is, cyclic trimer, comprising a three tetrapeptide units (Fig. 7.18). It is obtained from *Streptomyces*. It permits transport of K^+ across cell membranes. It has a polar interior that interacts with K^+, and a non-polar exterior which enables it to interact with the membrane. K^+ is held in the centre of the cyclic structure through electrostatic interactions with carbonyl oxygen of the peptide bonds. The non-polar exterior is due to the hydrophobic side chains of the non-polar amino acids that constitute this antibiotic. These side chains easily intercalate with the hydrophobic membrane matrix, which enables the valinomycin molecule to shuttle back

and forth across the membrane. During the process, the K^+, held in the centre of the antibiotic, is carried across the permeability barrier.

Gramicidin A is an antibiotic isolated from *Bacillus brevis*. It forms a tail-to-tail dimer, which makes a pore across the membrane (Fig. 7.19). The two helical monomers span the membrane. Length of the helix is about 3 nm, which is consistent with the hydrophobic region of a typical phospholipid bilayer. The helical pore, having a diameter of 0.4 nm, can transport several cations, such as H^+, Na^+, K^+, etc. Ion transport by gramicidin is 1000 times faster than that by valinomycin.

Specific channels can rapidly transport ions across the hydrophobic barrier of membranes.

Exocytosis and Endocytosis

Exocytosis

Exocytosis refers to the process by which outward movement of certain intracellular substances occurs (Fig. 7.20a). Intracellular enzymes, hormones, lipoproteins and a variety of other endogenously synthesized molecules move out of the cell through exocytosis. A classical example is extrusion of the lipoprotein, VLDL, from hepatocytes. The nascent VLDL particles, synthesized in the rough endoplasmic reticulum, undergo glycosylation in the Golgi apparatus. The glycosylated particles are then packaged into membrane-bound vesicles. The latter move towards the cell membrane, the vesicular contents (i.e. the VLDL particles) are released extracellularly.

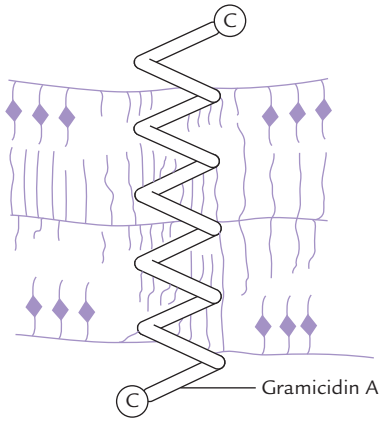


Fig. 7.19. Structure of gramicidin A: a pore type transporter.

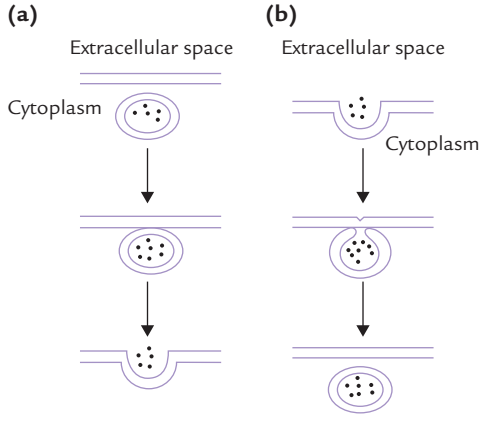


Fig. 7.20. Cellular transport of substances. **(a)** Exocytosis, **(b)** Endocytosis.

Endocytosis

Endocytosis *refers to the process of cellular uptake of certain extracellular substances* (Fig. 7.20b). Uptake of variety of extracellular macromolecules, that are important for the cell, occurs through endocytosis. The process begins with invagination of a segment of the plasma membrane; the invaginated portion encloses the extracellular macromolecule to be internalized. The invaginated segment then closes upon itself to form a vesicle. The latter pinches off next, and enters the cell. In many instances, the internalized vesicle fuses with lysosome and its contents are digested by the *lysosomal hydrolases*.

Receptor-mediated endocytosis: In case of certain extracellular macromolecules, the internalization is preceded by interaction with certain cell surface receptors. The process is called receptor-mediated endocytosis and is responsible for uptake of a variety of essential metabolites, including low density lipoprotein (LDL), vitamin B_{12} bound to the carrier protein (i.e. transcobalamin II), and iron bound to transferrin. Several growth factors (e.g. nerve growth factor), hormones (e.g. insulin), toxins and some viruses also enter the cell through this process.

Action of LDL receptors is presented in Figure 7.21. The receptors are transmembrane glycoproteins. They are located in some specific indentations of the cell membrane called the **coated pits.** Cytoplasmic aspect of these pits have high concentration of a protein, called *clathrin* which forms a lattice work. Interaction of ligands with the receptors is followed by invagination of coated pits and subsequent sealing off to form coated vesicles. Clathrin helps in the process of sealing off. In cytosol, the clathrin is shed and several coated vesicles fuse together to form large structures called **endosomes**.

Certain important changes take place in the endosome which permit recycling of the receptor. A proton pump builds up a high proton concentration within the endosome which drops the pH to about 5. At this pH, dissociation of the ligand and the receptor occurs. The separated receptor is then enclosed in a new vesicle and recycled back to the plasma membrane. The remainder of the endosome fuses with lysosome and its contents are degraded by the lysosomal enzymes.

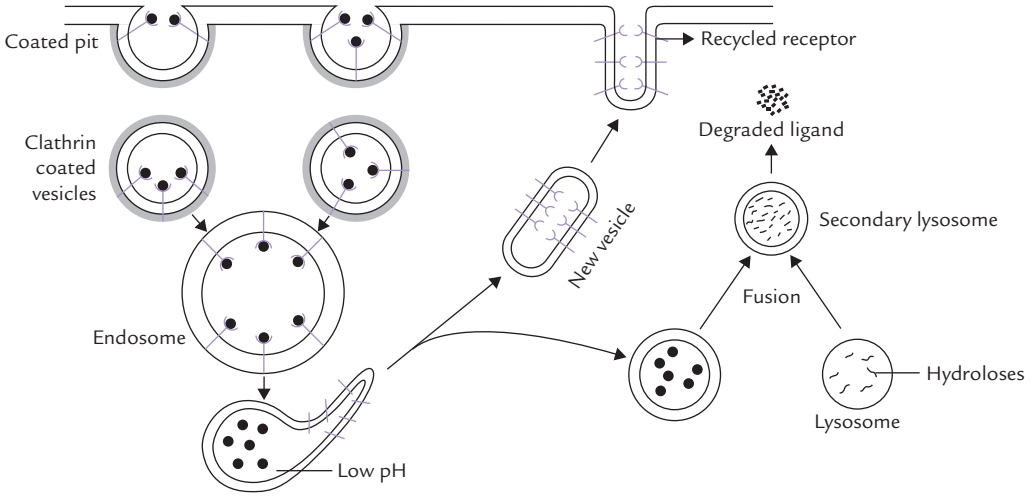


Fig. 7.21. Receptor-mediated endocytosis: action of LDL receptors (● = ligand, ⟙ = receptor).

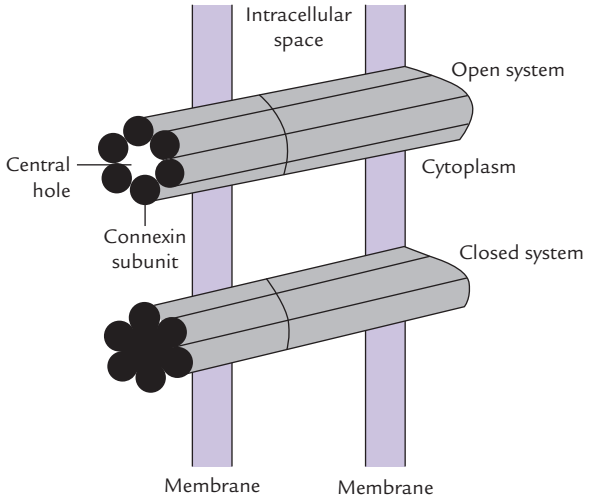


Fig. 7.22. Gap junction connecting two opposed cells.

Receptor-mediated endocytosis is the selective uptake of extracellular macromolecules (e.g. cholesterol) through their binding to specific cell-surface receptors. The receptor-macromolecule complex then accumulates in clathrin coated pits and is endocytosed via clathrin coated vesicles.

Transport Through Gap Junctions

Cell-to-cell transport is made possible through passageways connecting the interiors of the contagious cell (Fig. 7.22). These passageways span the intervening space, or gap between the apposed cells and hence, are termed the **gap junctions**, or **cell-to-cell channels**. All polar molecules with molecular weight of 1 kD or less can be transported through gap junctions. They move along a concentration gradient by passive diffusion.

Structure: A transmembrane protein, known as **connexin** (MW 32 kD) is the major structural element of gap junction. Total of 12 connexin subunits form a single gap junction. Six of these subunits are arranged in a hexagonal array in such a manner that they enclose a central hole about 20 Å wide (Fig. 7.22).

These six subunits form a half-channel called **connexon** (i.e. hemichannel). Two such hemichannels join ends in the intercellular space to form a gap junction, which encloses a functional channel between the communicating cells. These cell-to-cell channels differ from other membrane channels in two important aspects.

- They traverse two membranes rather than one. Therefore, they form connection from cytosol to cytosol; rather than cytosol to extracellular space, or cytosol to lumen of an organelle.
- They have a functional nature; a high concentration of calcium or low pH can close a channel. The subunits of connexin slide and rotate in such a way that closure of the central hole occurs. Switching between the open and closed state is smooth because free energies of these two states do not differ markedly. As a result, passage of molecules occurs rapidly as per the cellular requirements.

Functions: Primary function of gap junctions is to serve as passage-ways for inorganic ions and a variety of metabolites, such as sugars, amino acids and nucleotides. In contrast, proteins, nucleic acids and polysaccharides are too large to traverse these channels. Evidently, gap junctions are of prime importance for **intercellular communication**.

Other functions of gap junctions are:

1. They are essential for nourishment of the cells that lie distant from blood vessels, such as the cells of lens and bone.
2. They ensure synchronous and rapid response of cells of excitable tissues, such as the heart muscle. This is because, in response to a stimulus, the ions rapidly

flow from one cell to another, thus ensuring a coupled and quick response.
3. Current evidence suggests that the communicating channels are important for development and tissue differentiation.

Exercises

Essay type questions

1. What are liposomes? Mention their uses in biology and medicine.
2. Describe the fluid mosaic model.
3. Distinguish passive, facilitated, active- and secondary active-transport systems.

Write short notes on

1. Difference between integral and peripheral membrane proteins
2. Inophores
3. Na^+–K^+-Pump
4. Secondary active transport
5. Gap junctions
6. Lipid bilayer

CLINICAL CASE

CASE 7.1 A 3-year-old boy with salt disposition on skin

A 3-year-old child was brought to the hospital OPD with complaints of cough, difficulty in breathing, and indigestion. He had recurrent episodes of infections of respiratory tract for which he was treated with antibiotics by his father, a general practitioner. The latter had noticed that during hot weather, acute salt deposition occurred on child's skin. Moreover, a fond kiss left a salty taste in mouth, which he affectionately referred to as the 'salty kiss'. A first cousin of the child had somewhat similar signs and symptoms, though of a milder degree.

On examination, the child appeared weak and malnourished. Analysis of the sample obtained by rectal biopsy showed thick mucus, that was blocking various tubular structures of the glands. The epithelial cells (obtained by biopsy) were cultured in appropriate medium, and transport of chloride ions across these cells was measured. The results were compared with the chloride transport in epithelial cells of a normal subject. The transport in the patient was less than 5% of that in case of the normal subject. ^{14}C–PABA test, a tubeless pancreatic function test, was performed. It showed decreased urinary elimination of the labeled compound which suggested impaired exocrine pancreatic activity. Chloride content of the sweat was markedly enhanced (88 mEq/L).

Q.1. Identify the biochemical defect in this child. Which clinical disorder does the above defect lead to?
Q.2. Provide a biochemical explanation for the signs and symptoms of the child.
Q.3. The CFTR protein was isolated from the airway. What changes are likely to be present in this protein?
Q.4. The CFTR protein of the patient was found missing in phenylalanine at the 358th position. However, the protein was found to be functionally active because when it was reconstituted with phospholipids to form lipo-some, the chloride transport across the latter was normal. Provide an explanation for the patients disease stated in spite of having a functionally normal protein.

CHAPTER

8

OVERVIEW OF METABOLISM AND CELL BIOENERGETICS

Metabolism is defined as the sum total of all the chemical reactions that are taking place in the body. Metabolism is derived from a Greek word, *metabellein,* which means "to change". It includes the process by which cells use food material to obtain energy, store excess calories for future use and build up various substances. Metabolism also includes degradation and excretion of unnecessary compounds. In short, metabolism is sum total of all those processes that turn food into flesh.

Bioenergetics is the field of biochemistry that deals with transformation and utilization of energy in biological systems. It concerns only with the initial and the final energy states of reaction components, and predicts the energetic feasibility of chemical reaction. However, it provides no information about the mechanism or the rate of reaction. These aspects are measured in kinetics.

In this chapter, general aspects about metabolism and cellular bioenergetics are discussed. After going through this chapter, the student should be able to understand:

- Fundamental design of metabolic network; interdependence of metabolic reactions.
- Regulation of metabolic pathways and various techniques used to study the details of metabolism.
- Principles of bioenergetics: laws of thermodynamics; free energy, entropy and enthalpy; equilibrium constant.
- Role of ATP as energy carrier; concept of high-energy and super high-energy compounds and substrate level phosphorylation; other energy rich nucleoside triphosphates.

I. Overview of Metabolism

A. General Considerations

Metabolism serves two important purposes:

1. To release energy from the ingested food material through catabolic degradation, and to convert this energy into a form that can be used for cellular work.
2. To transform small organic compounds into macromolecules. This aspect of metabolism also includes transformation of one group of organic compounds into another.

During catabolic degradation, the energy inherent in the organic molecules (particularly carbohydrates and lipids) is released. It is then trapped and stored as **adenosine triphosphate** (**ATP**). The stored energy can be released from ATP when needed and used to perform cellular work (Fig. 8.1). The major cellular works are:

- **Transport** of organic molecules and inorganic ions across the cell membrane.
- **Mechanical work**, such as muscle contraction.
- **Electrical work** (e.g. nerve conduction).
- Ensure fidelity of **information transfer**.

The second major purpose of metabolism is to synthesize a vast array of macromolecules, which include carbohydrates, proteins, lipids and nucleic acids. It is amazing that so many diverse biomolecules are intracellularly synthesized from a limited number of organic compounds. Evidently, thousands of reactions are

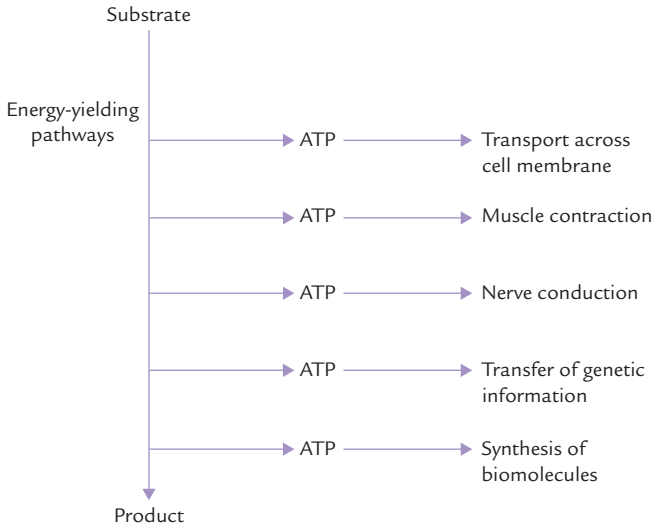


Fig. 8.1. The major cellular works.

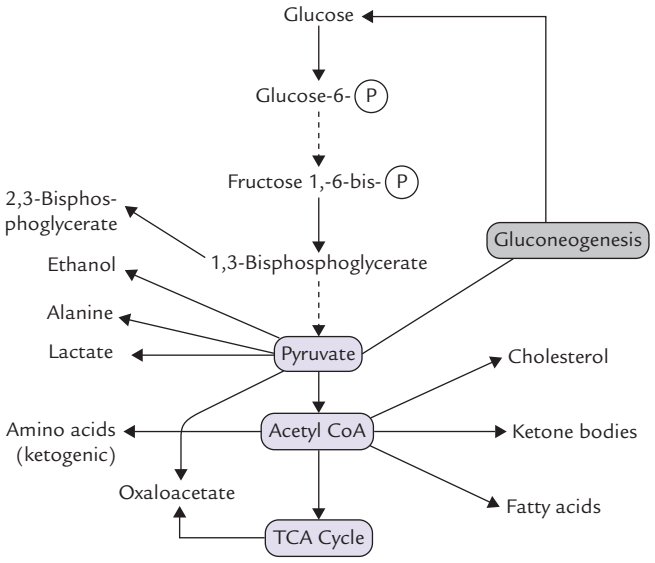


Fig. 8.2. Metabolic pathways connected through branch point compounds, pyruvate and acetyl CoA.

involved in the processes which split, join and rearrange the atoms of organic compounds, thus resulting in generation of complex biomolecules. Each of these reactions are catalyzed by a specific enzyme.

Metabolism comprises a highly integrated network of chemical reactions, which can be subdivided into **catabolism** and **anabolism.** Catabolic reactions are used to extract energy from fuels, and anabolism comprises reactions that use this energy for biosynthesis.

A bird's eye view of the scheme of metabolism is presented in this chapter with a focus on fundamental principles underlying metabolism in the cell. The details of reactions of individual metabolic pathways are given in subsequent chapters.

B. Metabolic Reactions are Interdependent and Interconnected

Metabolic reactions do not occur in isolation, or in a random or haphazard manner. Rather they are organized into multi-step sequences called metabolic pathways. Each reaction forms just one step in a metabolic pathway and is part of a larger scheme that involves several other interrelated reactions.

In metabolic pathways, the product of one reaction serves as a substrate for the next one; the product of second reaction is substrate for the third reaction, and so on.

$$A \rightarrow B \rightarrow C \rightarrow D$$

Such series of consecutive reactions allows the cell to carry out highly complex molecular conversions intracellularly. It is noteworthy that these conversions occur in mild conditions of temperature and pH that prevail within the cell.

An important example that illustrates this design is the sequence of reactions that converts glucose to pyruvate (i.e. glycolytic sequence). As soon as the glucose enters the cell, a phosphate group from ATP is added to it to form glucose 6-phosphate. Glucose 6-phosphate becomes the substrate for the next reaction, in which an *isomerase* converts it into fructose 6-phosphate. The latter then serves as substrate for another enzyme-catalyzed reaction, and the sequence continues through six more reactions until glucose is converted to pyruvate. The reactions of this metabolic pathway are summarized in Figure 8.2.

The energy inherent in the substrate glucose is released in small packets in a stepwise fashion and is effectively captured. If the glucose to pyruvate conversion occurred in a single step, the energy inherent in the glucose molecule could not have been trapped as ATP so effectively. A stepwise transformation ensures efficient and effective trapping of the energy. Further, some of the intermediates of this pathway are channeled into other pathways; for example, glucose 6-phosphate can enter glycogenesis or the pentose phosphate pathway, and 1,3-bisphosphoglycerate can form 2,3-bisphosphoglycerate.

Metabolic pathway appears like an **intricate and integrated web of chemical reactions** wherein the individual threads are interconnected at several points. In subsequent chapters, individual metabolic pathways would be discussed separately with an aim to simplify and categorize. This might give an erroneous impression that each pathway is self-contained and isolated.

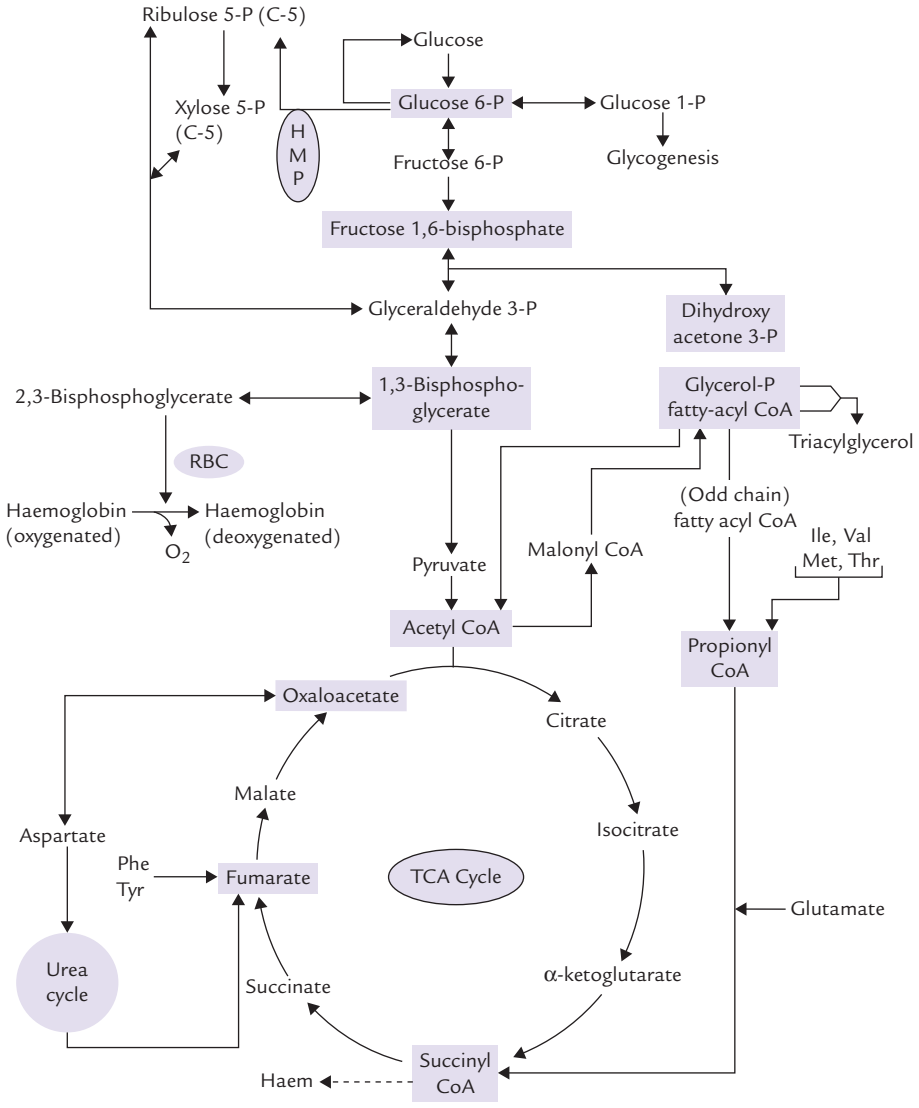


Fig. 8.3. Some important interconnections between metabolic pathways.

However, in reality, all the pathways are interdependent. In fact, if any reaction sequence is followed through the metabolic network sufficiently, we can find it to be connected to every other. For example, the end product of glycolytic sequence is pyruvate which serves as the starting point for *gluconeogenesis*. It can also be converted to lactate (i.e. anaerobic glycolysis), oxaloacetate (by carboxylation), or acetyl CoA by *pyruvate dehydrogenase* complex (Fig. 8.2). Acetyl CoA can serve as a precursor for fatty acids, cholesterol, ketone bodies, etc.

Similarly, various other metabolic intermediates, such as glucose 6-phosphate, dihydroxyacetone phosphate, succinyl CoA, etc. also serve as common links between different pathways. Such metabolites are referred to as the **branch point compounds**.

Figure 8.3 shows important interconnections through branch point compounds (other than pyruvate and acetyl CoA), e.g. oxaloacetate, succinyl CoA and glucose 6-phosphate.

Metabolic pathways do not occur in isolation: they are interdependent and interconnected through branch point compounds. Through these compounds different pathways intersect and form a network of chemical reactions, called metabolism.

C. Regulation of Metabolism

How are metabolic pathways regulated and why is it necessary to regulate them? As noted earlier, all cells are capable of carrying out the central metabolic pathways (such as glycolysis, critic acid cycle, gluconeogenesis, β-oxidation, lipogenesis, etc.). However, it is not desirable to run them all at the same time. For example, when the cell is adequately supplied with energy, the energy-yielding pathways like β-oxidation and TCA must be impeded; and at the same time, the available energy should be

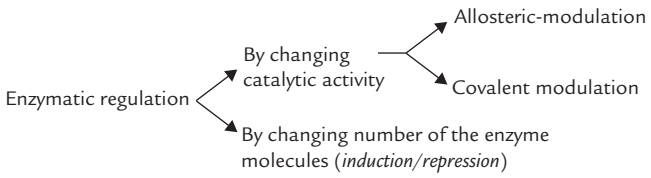
used to drive forward the energy requiring pathways, such as lipogenesis and gluconeogenesis.

Moreover, regulation of the metabolic pathways ensures *economical use of nutrients*. Imagine a cell present in a glucose rich medium. Glucose serves as its major carbon source in this environment. If this cell is shifted to a medium containing alanine as well, it would be a wasteful expenditure if the cell continued to produce alanine from glucose. An efficient cell would turn off its own alanine producing machinery and instead obtain alanine from the medium.

Control of metabolic reactions is accomplished through regulation of enzyme activity. Within a cell, the enzyme activity can be regulated in two ways.

1. By increasing or decreasing the *catalytic activity* of the enzyme molecule. This is accomplished through (i) **allosteric modulation**, or by (ii) **covalent modulation**.

   Both the mechanisms induce change in conformation of the enzyme protein, thereby altering its activity.
2. By increasing or decreasing the *total number of enzyme molecules*. This is accomplished by inducing changes in the enzyme synthesis (by alteration in either transcription or translation), or the enzyme degradation. Hormones play an important role in regulating the enzymes by this mechanism, termed ***induction-repression.*** (For details on regulation of enzyme activity, see Chapter 6.)



The most economical way to regulate a pathway is to change the activity of the first enzyme of the pathway. Inhibition of the first enzyme stops the production of the end product without wasting any energy of nutrients in the production of unnecessary early intermediates. Thus, *the enzymes catalyzing the production of early substrates in a pathway are prime target of metabolic regulation.*

Catalytic activities of enzymes of metabolism are regulated by allosteric interactions (as in feedback inhibition) and by covalent modulation. Critical enzymes of metabolism are controlled by regulation of the rate of protein synthesis and degradation.

D. Study of Metabolic Pathways

A number of methods are employed for detailed study of a metabolic pathway so as to elucidate its sequence and its relation with other metabolic pathways. One of the earliest methods, though elementary in nature, involved determination of nature and amount of various compounds present intracellularly. This information was then used to discern the metabolic pathway. However, such analysis yielded a long list of chemicals present intracellularly, but did not provide much information about the metabolic interrelations.

Studies with **isotopic tracers** and **metabolic blocks** have been invaluable in this regard. Using these aids it has become possible to know the exact sequence of metabolic pathways. In addition, examination of **cell-free extract** has revealed finer details of metabolic processes and their regulation.

Isotopic Tracers

Metabolic fate of a compound can be traced by putting a label (**tag**) on it. Follow up of the tag then serves as a useful indicator of the metabolic route taken by the labelled compound. This strategy is analogous to putting a flag atop a car which helps to identify it in busy traffic.

Earlier, certain dyes were used for the purpose of labelling (i.e. *chemical tagging*). However, it was found that these dyes might interfere with the normal metabolism of the compound. The *chemical tagging* is, therefore, of a limited value. Thereafter, search for tags that would not alter metabolism continued. Isotopes were found to be ideal tags.

Isotopes are alternative forms of the naturally occurring elements. Chemical properties of an isotope are identical to that of the more abundant form of the element. Its atomic number is also the same, but atomic weight is different from that of the natural element. For example, atomic number of all isotopes of carbon is 6, but molecular weight may be 12, 13 or 14 (Table 8.1). The chemical properties of the isotopes are identical. Therefore, cells cannot distinguish between isotopes of the same element. Cells cannot distinguish glucose that

Table 8.1 Isotopes used for study of metabolic pathways

Element	Abundant isotope	Rare isotope	Distinguishing physical property of rare isotope
Nitrogen	^{14}N	^{15}N	Heavy, non-radioactive
Carbon	^{12}C	^{13}C ^{14}C	Heavy, non-radioactive Radioactive
Hydrogen	^{1}H	^{3}H	Radioactive
		^{2}H	Heavy, non-radioactive
Phosphorus	^{31}p	^{32}p	Radioactive
Iodine	^{127}I	^{131}I	Radioactive
Sulphur	^{32}S	^{35}S	Radioactive

contains carbon with an atomic weight of 14 (^{14}C-glucose) from the natural one having an atomic weight of 12 (^{12}C-glucose). However, fate of ^{12}C-glucose can be followed distinctly from the ^{14}C-glucose, lending support to the statement, "isotopes are the compounds that cells cannot see but a scientist can see."

Isotopes have unusual physical properties which help their identification. For example, heavy isotopes of nitrogen (**^{15}N**) when incorporated in DNA, gives the latter a higher density. The DNA molecule produced from heavy nitrogen settles at a lower position than its ^{14}N counterpart on performing centrifugation. This strategy was employed by Stahl and co-workers to elucidate the semi-conservative nature of DNA replication (Chapter 21). Some isotopes of carbon are radioactive in nature and emit ionizing radiations. The latter can be detected on radioactivity counter. As a result, the labelled compounds can be easily identified in the complex metabolic maze. For instance, where ^{14}C-acetyl CoA is fed to experimental rabbits, the label appears in tissue fatty acids and tissue cholesterol. Therefore, it can be concluded that acetyl CoA serves as a precursor for fatty acids and cholesterol.

Metabolic Blocks

Metabolic pathways can be blocked by using inhibitors of specific enzymes or by use of the mutant cells which lack one of the other enzyme of the pathway. As a result of the block, the substrate of the blocked reaction tends to accumulate. Sometimes the earlier intermediates of the pathways (i.e. the ones lying a few steps prior to the block) also accumulate. Characterization of the accumulated compounds helps in understanding of the pathway.

$$V \rightarrow W \rightarrow X \rightarrow Y \rightarrow Z$$

If the first step is blocked, the first intermediate of the pathway (V) accumulates. Blocking the second step results in accumulation of first two intermediates (V & W). This information can be used for elucidating the pathway.

Metabolic blocks provide useful information about the sequence of events, i.e. the order in which the pathway intermediates are produced. Suppose, in the above pathway a metabolic block leads to production of V and W, but not of X, Y and Z. This may lead to the conclusion that V and W are produced earlier than X, Y and Z. If a different block allows the cell to produce V, W, X and Y, but not Z, one may logically conclude that Z is the final product of the pathway. In the same manner different metabolic blocks are employed and the results are summed up. This permits placement of intermediates in a defined order.

Cell-free Extracts

The most elaborate metabolic analysis is conducted in cell-free systems. These systems contain all the enzymes and coenzymes that are present intracellularly and which are necessary for the reaction to proceed. In this way, the reactions taking place within the cell are made to occur in vitro.

The information gained by isotopes and metabolic block, when combined with information from studies with the cell-free extract, makes it possible to reconstruct the pathway. For example, to elucidate the first step of a given pathway, the initial substrate is mixed with the cell-free extract. Characterization of the product formed helps us to be sure of correct elucidation of the first step. For example, when phenylalanine (labelled) is mixed with cell-free extract obtained from hepatocytes, the lable appears in the tyrosine molecule after some time. It may, therefore, be concluded that the first reaction of phenylalanine metabolism is:

$$\text{Phenylalanine} \rightarrow \text{Tyrosine}$$

If, however, the labelled tyrosine is not produced in this system, the obvious conclusion is that there is a metabolic block, i.e. phenylalanine to tyrosine conversion is blocked. Mostly likely, the enzyme catalyzing this reaction (*phenylalanine hydroxylase*) is deficient, or the coenzyme for the reaction is not available **(Case 4.3)**. Study of various reactions in this manner help in correct identification and placement of the metabolic intermediates.

Studies with cell-free extract also permit detailed study of a given metabolic pathway, separately from the others. Further, effect of various extraneous conditions, such as pH, temperature and substrate concentration can be studied by changing these factors. In this manner not only the sequence of the pathway but also its regulation can be studied.

Steps of Cell-free Extracts Preparation from the Isolated Cells

In a number of clinical cases in this book, diagnostic utility of studies with cell-free extracts is expressed.

Step I: Rupture of the cell membrane, i.e. cell lysis, as described in Chapter 7. The intracellular contents are thereby obtained.

Step II: Separation of the intracellular contents from the solids, such as cell membrane and cellular debris by ultracentrifugation. The supernatant represents the cell-free extract.

A number of methods are employed to study metabolism, such as application of isotopes and use of metabolic blocks and cell-free extracts.

II. Bioenergetics: Principles

Energy is defined as capacity to do work; it is required for performing various activities. The living cells need energy for performing various activities, discussed on page 146.

Billions of years of evolution has taught the living cells to use energy very economically and efficiently.

Energy transformations, both by the living systems as well as by the artificial devices, take place according to the laws of thermodynamics. Thermodynamics is a branch of physical science that deals with energy changes.

A. Laws of Thermodynamics

The First Law

The first law of thermodynamics, also referred to as the *law of conservation of energy, states that energy can neither be created nor destroyed*. In course of any physical or chemical reaction, one form of energy may change to some other form, but the total amount of energy in the universe always remains constant.

For example, electric energy changes to heat energy in room heater, heat energy changes to mechanical energy in rail engine and mechanical energy changes to electric energy in hydroelectric plants. These examples show that various forms of energy are interconvertible. However, none of these transformations brings about any net generation or loss of energy.

The Second Law

All physical and chemical reactions tend to proceed in such a direction that useful energy of the reacting system is irreversibly converted to a randomized and useless form, known as ***entropy***. The reactions proceed in this direction till entropy reaches maximum possible under the prevailing circumstances. At this point, called the equilibrium point, no further progress of the spontaneous reaction is possible.

Thus in all reacting systems, disorder or randomness is favoured at the cost of orderliness. To be more explicit, fall in useful energy content of the system occurs with a concomitant rise in the randomized energy of the universe. The reacting system implies collection of matter undergoing a reaction, and the universe includes both, the system and its surroundings. This literally includes the whole of earth, or even the outer space.

Bioenergetics is the field of biochemistry that deals with the study of energy changes in metabolic reactions. All living organisms conform to the laws of thermodynamics.

B. Free Energy, Entropy and Enthalpy

Useful energy is the form of energy that is capable of performing work. It is broadly classified in two major types:

1. **Heat energy**, which is capable of performing work through change of temperature.
2. **Free energy** which is capable of performing work at a constant temperature.

Since human body maintains a constant temperature (i.e. isothermic), it cannot utilize the heat energy. **Free energy** is the useful form of energy in humans because of its ability to function at constant temperature. It performs various functions (mentioned on page 146) and, in the process, gets converted to the randomized form, i.e. **entropy**. These relations can be expressed in the equation form as below:

$$\Delta G = \Delta H - T\Delta S$$

- **ΔG** is change in the free energy content of the reacting system. It equals the difference in energy level between the substrates and products.
- **ΔS** is change in entropy of the universe.
- **T** is absolute temperature measured in Kelvin.
- **ΔH** implies change in heat content of the system, which is termed as **enthaply** (the word means "warming within").

In course of any ongoing reaction, entropy of the universe (system + surroundings) always increases. *The value of ΔS, therefore, has a positive sign*. Since rise in entropy is accompanied by a corresponding fall in free energy of the system, *the value of ΔG is negative*. The enthalpy change may be positive or negative, depending on whether the system absorbs or releases heat. It is positive when the heat absorption occurs and is negative when heat release occurs. However, in the isothermic state prevailing in the body, its value is zero.

Finally, it must be emphasized that entropy or disorder is not an entirely useless activity. Since increase in entropy is an irreversible process, it gives direction to all biological activities.

C. Standard Free Energy Change

The free energy change, ΔG is the most important thermodynamic function because it is the driving force of the reactions. Like ΔH, it is measured in kilocalories per mole (kcal/mol) or kilojoules per mole (KJ/mol). A negative sign of ΔG indicates an **exergonic reaction**, the one that proceeds spontaneously and forms product from subratate. A positive sign of ΔG indicates an **endergonic reaction** which can proceed only in the backward direction. *At equilibrium, ΔG equals zero.* ΔG is not a property of the reaction as such but is affected by relative reactant concentrations. To obtain a convenient energetic expression that predicts the equilibrium of the reaction, we have to define the **standard free energy change** ($\Delta G^{0'}$). The standard conditions are as below:

- Temperature of 25°C (298°K).
- Pressure of 1.0 atm (760 mmHg).

- pH 7.0.
- The reactants and products are all present in the concentration of 1.0 M.

The standard free energy change of a given reaction is an ***immutable constant*** and it equals the difference in energy level between the substrates and products, under standard conditions, and is shown in Table 8.2. It predicts equilibrium of a reaction.

The free energy change (ΔG) is the driving force for chemical reactions. In spontaneous exergonic reactions there is release of free energy (ΔG is negative), and in endergonic reactions there is requirement for free energy (positive ΔG).

The standard free energy change ($G^{0'}$) must be differentiated from the *actual free energy change* (ΔG) of a reaction:

1. $\Delta G^{0'}$ is measured under standard conditions, stated above and has a constant value for a given reaction. In contrast, the actual free energy change is measured at the conditions prevailing during the reaction. It is a function of the conditions of concentration, pH and temperature actually prevailing during the reaction, which are not necessarily the standard conditions.
2. Further, the value of the actual free energy change keeps changing as the conditions change. It has negative value in the beginning of the reaction, which becomes less and less negative as the reaction progresses towards equilibrium. At equilibrium its value falls to zero.

$\Delta G^{0'}$ has a definite relation with ΔG and with equilibrium constant of the reaction, as illustrated in Boxes 8.1 and 8.2, respectively.

$\Delta G^{0'}$ is an immutable constant; it has a characteristic unchanging value for a given reaction. On the other hand, value of ΔG in an ongoing reaction keeps changing with the progress of the reaction.

Table 8.2. Standard free energy of hydrolysis of some common metabolic intermediates

Metabolite	$\Delta G^{0'}$
Phosphoenol pyruvate	−14.8
Carbamoyl phosphate	−12.3
1,3-Bisphosphoglycerate	−11.8
Acid phosphate	−11.2
Creatine phosphate	−10.3
Arginine phosphate	−7.6
ATP to ADP + P_i	−7.3
ATP to AMP + PP_i	−7.7
ATP to AMP + P_i + P_i	−14.6
Glucose 1-phosphate	−5.0
Glucose 6-phosphate	−3.3
Glycerol 1-phosphate	−2.2

Standard Free Energy Changes of Sequential Reactions

Total standard free energy change of a reaction sequence equals the sum of the standard free energy changes of the individual reactions of the sequence.

Reaction 1

Glucose + ATP → Glucose 6-phosphate + ADP
($\Delta G^{0'} = -4.0$ kcal/mole)

Reaction 2

Glucose 6-phosphate + H_2O → Glucose + Pi
($\Delta G^{0'} = -3.3$ kcal/mole)

The sum of these two reactions is

$$ATP + H_2O \rightarrow ADP + P_i$$

It has a standard free energy change of $(-4.0) + (-3.3) = -7.3$ kcal/mole.

Note: Value of the standard free energy of ATP hydrolysis can be thus obtained in an indirect manner only.

BOX 8.1

Relationship Between the Standard Free Energy Change (ΔG^0) and the Actual Free Energy Change (ΔG)

The actual free energy change (DG) and the standard free energy change ($\Delta G^{0'}$) of a chemical reaction ($\rightleftharpoons$) are related to each other as shown in the equation below:

$$\Delta G = \Delta G^{0'} + 2.303\ RT \log \frac{[B]}{[A]}$$

where T is absolute temperature measured in kelvin, R is the gas constant, and [A] [B] are the actual concentrations of the reaction substrate and product respectively.

D. Role of ATP in Cell Bioenergetics

ATP (adenosine triphosphate) serves as the mediator of biological energy transfers. It acts as an energy carrier in biological systems. It links the energy-yielding (i.e. exergonic) and the energy-requiring (i.e. endergonic) processes.

Composition: ATP is a nucleotide consisting of the following three components (Fig. 8.4):

1. *Purine base* which is adenine.
2. *Ribose sugar* which is a 5-carbon sugar (i.e. a pentose). The first carbon of ribose is linked to N-9 of adenine through N-glycosidic linkage.
3. *Phosphate groups*, three in number, are designated α-, β- and γ-phosphates. The α-phosphate is linked to C-5 of the ribose.

Functions: ATP serves as a link between the energy-yielding (exergonic) pathways and the energy-requiring (endergonic) processes, as mentioned earlier.

Normally, the exergonic processes are thermodynamically favoured and therefore, occur spontaneously. The endergonic processes, on the other hand, are thermodynamically unfavoured, and hence, require input of energy to be driven forward. These two types of processes are coupled, so that the energy of exergonic reaction can be used to drive progression of the endergonic reaction.

In most biological transformations, ATP is generated during exergonic reactions, using the free energy released (Fig. 8.5). Thus each molecule of ATP represents stored free energy.

Endergonic reactions being thermodynamically unfavourable can be driven by coupling them to thermodynamically favourable exergonic ones (which in many cases is the hydrolysis of ATP).

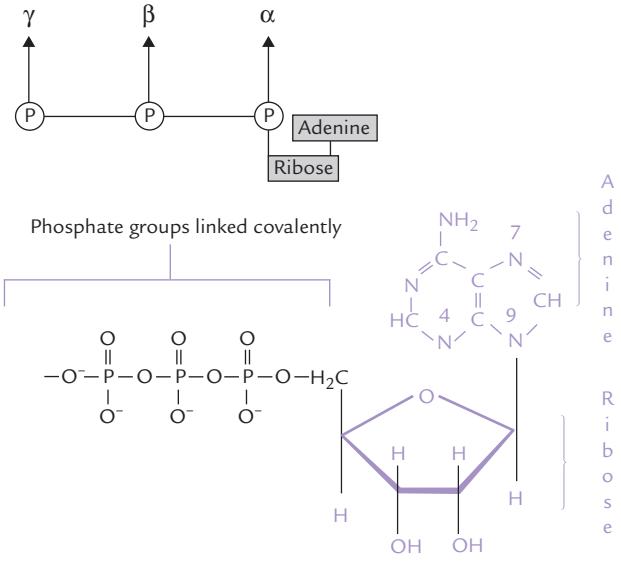


Fig. 8.4. Structure of ATP.

BOX 8.2

The Standard Free Energy Change is Related to Equilibrium Constant

In a reversible chemical reaction ($A \rightleftharpoons B$) equilibrium describes a state when rate of the forward reaction equals that of the reverse reaction. In this state no net chemical transformation occurs, so that equilibrium concentrations of both A and B are invariable. Their ratio is termed as equilibrium constant (K'_{eq} of the reaction)

$$K'_{eq} = \frac{[B]}{[A]}$$

where [A] and [B] refer to the equilibrium concentration of the reactant and the product, respectively. At equilibrium the relation between $\Delta G^{0'}$ and ΔG is as follows:

$$\Delta G = \Delta G^{0'} + 2.303 \log RT \frac{[B]}{[A]}$$

Since $\Delta G = 0$ **at equilibrium,** it can also be rewritten as:

$$\Delta = G^{0'} + 2.303\ RT \log K'_{eq}$$

When $K'_{eq} = 1$, value of $\log K'_{eq}$ is zero. Hence, $\Delta G^{0'} = 0$, meaning that product contains the same amount of free energy as the reactant, and therefore the reaction stands at equilibrium. Likewise, if the $K_{eq} > 1.0$, its $G^{0'}$ is negative and if $K_{eq} < 1.0$, its $G^{0'}$ is positive and the reaction respectively proceeds in the forward and the reverse direction. Table 8.3 summarizes these points.

The standard free energy change of a chemical reaction is simply a different mathematical way of expressing its equilibrium constant.

Some of the energy requiring processes include biosynthesis, active transport, transfer of genetic information and movement (Fig. 8.1). Thus, ATP serves as a primary and universal carrier of free energy. It is also referred to as the *currency of free energy in the body.*

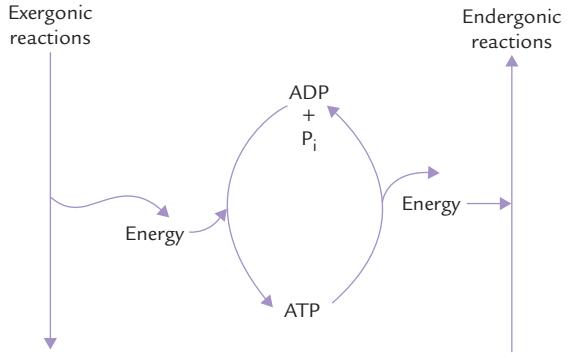


Fig. 8.5. ATP serves as common link between the exergonic and the endergonic reactions.

Table 8.3. Relationship between Keq, $\Delta G^{0'}$ and the direction of chemical reactions under standard conditions

K_{eq}	$\Delta G^{0'}$	Direction of reaction (starting with 1M components
>1.0	Negative	Forward
1.0	Zero	Stands at equilibrium
<1.0	Positive	Reverse

Exergonic reactions are used for the synthesis of ATP and the chemical bond energy of ATP drives the energy dependent processes. Thus, ATP is used as the energetic currency of the cell.

Standard free energy ($DG^{0'}$): The standard free energy of hydrolysis of ATP is **−7.3 kcal/mole**. It implies that 7.3 kcal of free energy is released, under standard conditions, from 1.0 mole of ATP by the following reactions.

$$ATP \rightarrow ADP + P_i$$
$$(\Delta G^{0'} = -7.3 \text{ kcal/mole})$$

Conversely, reformation of ATP occurs by attachment of a terminal phosphate group to ADP. The process requires input of the same amount of free energy. Reasons for high value of the standard free energy of hydrolysis of ATP are given in Box 8.3.

The energy obtained from various catabolic pathways is used for the ATP generation: 7.3 kcal of energy is required for the synthesis of 1.0 M of ATP (from 1.0 M each of ADP and P_i). Therefore, each mole of ATP represents 7.3 kcal of stored energy. During the energy-requiring processes, hydrolysis of ATP occurs, which liberates the required energy. In this way, generation of ATP is followed by its hydrolysis, which is again followed by regeneration, and so on. Thus the γ-phosphate group of ATP

BOX 8.3

Reasons for Relatively High Standard Free Energy of Hydrolysis of ATP

Standard free energy of hydrolysis of ATP is higher than that of most of the phosphorylated compounds such as glucose 6-phosphate, glucose 1-phosphate or glycerol 3-phosphate. Certain structural features of ATP, described below account for higher $\Delta G^{0'}$ of ATP.

1. Degree of ionization. At pH 7.4, ATP is almost completely ionized as shown below.

$$ATP^{4-} + H_2O \rightarrow ADP^{3-} + HPO_4^{2-} + H^+$$

Under standard conditions, ATP^{4-}, ADP^{3-}, H^+, HPO_4^{2-}, would be present at 1.0 M concentration. However, at pH 7.4, concentration of H^+ is only about 10^{-7} M. This is very small compared to standard concentration of (1.0 M) ATP, and therefore, the reaction tends to be pulled far to the right. In contrast, no such effect is seen in case of glucose 6-phosphate.

2. The second major reason for the higher $\Delta G^{0'}$ of ATP is that the four closely spaced negative charges present on ATP molecule tend to repel one another strongly. This builds up considerable electrical stress. When terminal phosphate group is removed, the stress is relieved. Moreover, the reaction products, ADP^{3-} and HPO_4^{2-} are negatively charged and therefore, have little tendency to approach one another and recombine to form ATP.
3. Third reason for large $\Delta G^{0'}$ is the fact that both the reaction products, ADP and P_i are resonance hybrids. Resonance hybrids are special stable forms because the electrons sink to much lower energy levels in them. As a result, free energy content of ADP^{3-} and HPO_4^{2-} is much lower than that of ATP.

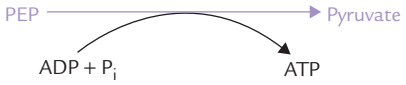
Bond between the terminal phosphate groups (i.e. γ and β) of ATP, which yields large energy upon hydrolysis, is a high-energy phosphate bond. Likewise, the bond between α- and β-phosphate is also a high-energy phosphate bond.

continually undergoes removal and replacement in a cyclic fashion: the process is termed the **ATP cycle**. It ensures formation of ATP with energy release and transfer of this energy to processes which require it.

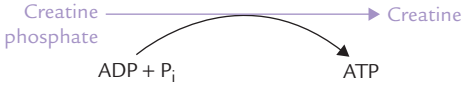
E. Low-Energy, High-Energy and Super High-Energy Compounds

ATP is high-energy compound placed higher than several compounds on the thermodynamic scale (Table 8.2). Such compounds having $\Delta G^{0'}$ less than that of ATP are termed the low-energy compounds. On the other side of the thermodynamic scale lie certain compounds, $\Delta G^{0'}$ of which is higher than that of ATP. Such compounds are referred to as the **super high-energy compounds**; phosphoenol pyruvate (PEP) and creatine phosphate having $\Delta G^{0'}$ of -14.8 kcal/mole and -10.3 kcal/mole respectively are some examples.

Phosphoenol pyruvate: Phosphoenol pyruvate (PEP) is an intermediate in glycolysis. Its conversion to the next glycolytic intermediate, pyruvate, under standard conditions, yields 14.8 kcal/mole energy. This is more than sufficient for the formation of ATP, which needs 7.3 kcal/mole. Thus, ATP, formation (i.e. phosphorylation) is coupled with transformation of a substrate (i.e. PEP in this case). This process is called **substrate level phosphorylation**.



Creatine phosphate: It is stored in the skeletal muscles, where it helps in quick in generation of ATP. During muscle exercise, ATP is rapidly hydrolyzed to meet the energy needs of the exercising muscle. Consequently, the ATP levels within the muscular tissue tend to get depleted. Creatine phosphate replenishes the ATP under these circumstances. Conversion of this compound to creatine and phosphate releases energy, which is used for ATP generation.



This is also an example of the substrate level phosphorylation. It permits the skeletal muscles to perform the intermittent strenuous work by maintaining the intracellular ATP levels normal. Since ATP lies midway in the thermodynamic scale, i.e. between the super high energy compounds and the low energy compounds (Table 8.2), it is a suitable compound to serve as a currency of free energy in the body.

Phosphoenol pyruvate, creatine phosphate and other such high energy compounds play a crucial role in the energy transfer of biochemical reactions.

F. Cleavage of ATP to AMP and Pyrophosphate

In course of most endergonic reactions, ATP is converted to ADP by removal of the terminal γ-phosphate group (**orthophosphate cleavage**). However, in certain other endergonic reactions, the β-and the γ-phosphate groups are simultaneously removed in one piece to form AMP and pyrophosphate (**pyrophosphate cleavage**).

$$ATP \rightarrow AMP + PP_i \quad (1)$$
$$\Delta G^{0'} = -7.7 \text{ kcal/mole}$$

This followed by cleavage of the pyrophosphate, which yields two phosphate groups.

$$PP_i + H_2O \rightarrow P_i + P_i \quad (2)$$
$$\Delta G^{0'} = -6.9 \text{ kcal/mole}$$

Sum of these two reactions and its $G^{0'}$ is

$$ATP + H_2O \rightarrow AMP + P_i + P_i \quad (3)$$
$$G^{0'} = (-7.7) + (-6.9)$$ ($G^{0'}$ of sequential reactions bring additive.)
$$= -14.6 \text{ kcal/mole}$$

This is double than the usual 7.3 kcal/mole yield by orthophosphate cleavage of ATP. Release of a larger amount of energy in this manner can drive certain complex biosynthetic reactions forward. For example, 7.3 kcal/mole, obtained by orthophosphate cleavage of ATP would be insufficient to drive the activation of fatty acid to completion, which requires an input of about 7.9 kcal/mole.

Fatty acid + CoASH → Fatty acyl CoA
($\Delta G^{0'} = +7.9$ kcal/mole).

Release of 14.6 kcal/mole of energy by the pyrophosphate cleavage of ATP can drive this reaction to completion.

Amount of energy provided by pyrophosphate cleavage of ATP is evidently far in excess of that required for the fatty acid activation. The surplus (i.e. $14.6 - 7.9 = 6.7$ kcal/mole) dissipates as entropy. However, *dissipation is not a wasteful expenditure, since it helps to drive the reaction forward and further ensures its irreversibility.*

Energy released by pyrophosphate cleavage of ATP is double (-14.6 kcal/mole) than that released by orthophosphate cleavage, which can drive highly energy dependent processes.

In contrast to ADP, which returns to the ATP cycle by directly accepting a phosphate group, AMP requires participation of an additional enzyme, *adenylate kinase*, to form ATP. This enzyme catalyzes reversible phosphorylation of AMP to yield two ADPs:

$$ATP + AMP \rightleftharpoons 2\ ADP$$

The ADPs so formed directly accept phosphate group(s) to form ATP. Acting in the reverse direction, the *adenylate kinase* has another important role: it helps to maintain a constant intracellular ATP level by catalyzing the transfer of the β-phosphate of one ADP to another ADP molecule. Thus it supplements the ATP generation by *creatine kinase* in the contracting muscle.

G. Other Energy Rich Nucleoside Triphosphates

There exist several other nucleoside 5′-triphosphates which are analogous in structure to ATP, and which are also energy-rich like ATP. Some examples are guanosine triphosphate (GTP), cytidine triphosphate (CTP), and uridine triphosphate (UTP). They are present in all cells but in a much lower concentration than ATP. Although ATP is the major carrier of the phosphate group, the other types of nucleoside triphosphates are specialized to serve in certain biosynthetic pathways. For example, *CTP is used in phospholipid biosynthesis and UTP in glycogen synthesis*. Synthesis of the nucleoside triphosphates is catalyzed by the enzymes called *nucleoside diphosphokinases*: ATP mostly serves as the donor of the terminal phosphate groups.

$$ATP + UDP \rightleftharpoons ADP + UTP$$
$$ATP + GDP \rightleftharpoons ADP + GTP$$
$$ATP + CDP \rightleftharpoons ADP + CTP$$

Low concentration of corresponding 2′-deoxyribonucleoside 5′-triphosphates is also present in cells. These are 2′-deoxyadenosine 5′-triphosphate (dATP) and 2′-deoxyguanosine 5′-triphosphate (dGTP), 2′-deoxycytidine 5′-triphosphate (dCTP), and 2′-deoxythymidine 5′-triphosphate (dTTP).

Besides ATP, all cells contain GTP and UTP, which are used as energy source in some enzymatic reactions.

Exercises

Write short notes on

1. High energy phosphates
2. Laws of thermodynamics
3. Standard free energy change
4. Currency of free energy
5. Entropy
6. Substrate level phosphorylation
7. Bioenergetics' significance of ATP
8. Isotopic tracers

METABOLISM OF CARBOHYDRATES I: MAINLINE METABOLIC PATHWAYS

CHAPTER

9

Carbohydrates are important for both the generation of metabolic energy and biosynthetic purposes. They are the structural basis of cellulose, bones, cartilage, lubricants such as mucus, cell recognition, DNA, RNA, and some membrane lipids (sphingosine). They are source of carbon skeleton of certain amino acids and are basis of some intracellular messenger systems. As an energy source they are broken down, and when there is a surplus, they are stored as starch and glycogen.

Glucose occupies a key position in carbohydrate metabolism, being a building block of all major dietary carbohydrates and the principal transported carbohydrate in humans. Fructose and galactose, obtained from dietary sucrose and lactose respectively, can be converted to glucose in liver. Therefore, the largest portion of dietary carbohydrates reaches the consuming tissues in the form of glucose. Evidently, an understanding of glucose metabolism forms the core of study of carbohydrate metabolism.

A detailed account of metabolic pathways is given in this chapter. After going through this chapter, the student should be able to understand:

- Transport of glucose into cells: facilitated transport and secondary active transport across cell membrane.
- Pathway of glycolysis: reaction sequence, generation of ATP and regulation.
- Feeder pathways: reaction sequences for monosaccharides, disaccharides and polysaccharides, and the associated metabolic diseases.
- Tricarboxylic acid cycle: reactions, energy yield, synthetic function and regulation.
- Pathway of gluconeogenesis; bypass reactions and reversible steps; energetics and regulation.
- Glycogen metabolism reactions of glycogen synthesis (glycogenesis) and degradation (glycogenolysis), their reciprocal regulation through cAMP activated cascade; role of hormones in the regulation; and biochemical basis and clinical presentations of glycogen storage diseases.

I. Transport of Glucose Into Cells

Glucose cannot passively move across the cell membrane to enter the cells. The polar nature of glucose hinders its movement across the predominantly non-polar lipid bilayer. Two kinds of transport mechanisms are present in the cell membrane, which permit glucose to cross the membranous barrier and enter the cell. These are (a) the **facilitated transport**, and (b) the **secondary active transport**.

Both are *carrier-mediated* processes, requiring transport protein(s).

Facilitated transport occurs along a concentration gradient and is mediated via a family of at least five transport proteins located in the cell membrane. These proteins are designated as GLUT-1, GLUT-2, GLUT-3 GLUT-4 and GLUT-5. Glucose moves from higher extracellular glucose concentration to a lower concentration inside the cell. A given transport protein exists in two conformational states, which alternate during the transport process (Fig. 7.11).

The glucose transport proteins are tissue specific. GLUT-1 is present in RBCs, whereas GLUT-4 is abundant in skeletal muscles and adipose tissue. Insulin increases

number and activity of GLUT-4, thereby promoting entry of glucose in these tissues (Chapter 7).

Note: Insulin is not required for glucose uptake by some tissues, such as liver, brain and red blood cells.

Secondary active transport is the mechanism of sodium-glucose cotransport across the membranous barrier, against a concentration gradient of glucose (i.e. from lower glucose concentration to a relatively higher concentration). The "*uphill movement*" of glucose is powered by the movement of sodium along its concentration gradient (Chapter 7). This process is operative during the movement of glucose from lumen of intestine into the intestinal mucosal cells (Fig. 7.16). Another type of such Na^+/glucose co-transporter (called SLUT-1) is known to operate in renal tubules.

Abnormalities in some of the transport proteins may lead to diseases; for example, renal glycosuria results in case of a defect in the renal transport mechanism for glucose.

Specific integral membrane proteins facilitate movement of glucose along concentration gradient, without expenditure of any energy in **facilitated transport;** whereas in **secondary active transport** a specific transport protein moves glucose against concentration gradient through expenditure of energy; the energy comes by co-transport of sodium ions along its concentration gradient.

Glucose 6-phosphate: a branch-point compound: After entering the cell glucose is rapidly converted to glucose 6-phosphate. The latter serves as a common link between various metabolic pathways, therefore called a branch point compound. Figure 9.1 shows some important interconnections. Glucose 6-phosphate can enter a number of pathways, e.g. glycolysis, uronic acid pathway, HMP shunt, glycogenesis, and can again form glucose by hydrolytic removal of the phosphate group. It can be generated by glycogenolysis and gluconeogenesis also. Further, glucose 6-phosphate cannot simply diffuse out of the cell by crossing the cell membrane and hence is committed to intracelluar metabolism. Detailed description of various pathways mentioned above has been given in Chapters 9 and 10.

Glucose enters the call by a carrier-mediated transport mechanism (facilitated transport or secondary active transport), where it is rapidly phosphorylated to glucose 6-phosphate—a compound that can enter several metabolic pathways.

II. Glycolysis

Of all the metabolic transformations that glucose 6-phosphate can undergo, the sequence of reactions leading to the formation of pyruvate is quantitatively the most

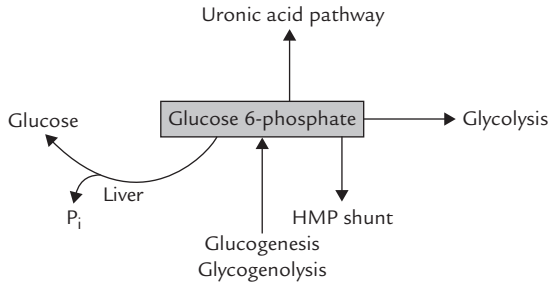


Fig. 9.1. The role of glucose 6-phosphate in carbohydrate metabolism.

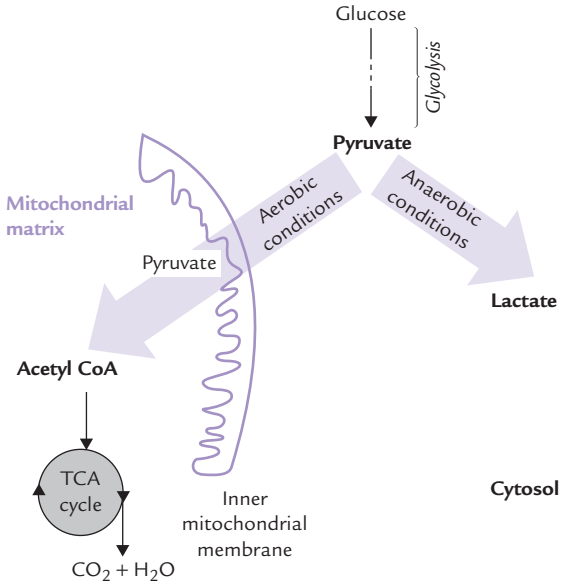


Fig. 9.2. Pyruvate metabolism under aerobic and anaerobic conditions. Under anaerobic conditions, pyruvate is reduced to lactate, whereas in aerobic conditions pyruvate is oxidized to acetyl CoA, which is further oxidized in TCA cycle.

important. This pathway is known as ***glycolysis or the Embden-Meyerhof pathway***. *It is an anaerobic process that involves the breaking down of a molecule of glucose into two molecules of pyruvate or lactate.*

$$\underset{(6\text{-c})}{\text{Glucose}} \dashrightarrow (2) \times \underset{(3\text{-c})}{\text{Pyruvate}}/\underset{(3\text{-c})}{\text{Lactate}}$$

Glycolysis is common to most organisms, and in humans it **occurs virtually in all tissues**. Although the glycolytic sequence usually begins with glucose, other sugars may also enter it via appropriate intermediates. Glycolysis serves a dual role:

- to generate metabolic energy
- to provide intermediates for several other metabolic pathways.

Glycolysis occurs in 10 sequential reactions, that occur in the **cytosol**. The fate of pyruvate, the end product of these reactions, is different in anaerobic and aerobic conditions (Fig. 9.2).

Under **aerobic conditions:** Pyruvate enters the mitochondrial matrix and is oxidized to acetyl CoA, which is the

major metabolic fuel for the citric acid cycle. Further oxidation of acetyl CoA in the citric acid cycle releases free energy which is used to generate ATP molecules. This process occurs in all tissues except those which lack mitochondria (e.g. erythrocytes, leucocytes) and the exercising muscles.

Under **anaerobic conditions:** Pyruvate is reduced in cytosol by NADH to form lactate. This process, called *anaerobic glycolysis* is discussed later. (Note that the term anaerobic literally means without air, but in practice it means without oxygen).

Glycolysis is a cytoplasmic reaction sequence of 10 enzyme-catalyzed reactions which produces two molecules of the three-carbon compound, pyruvate (or lactate), from the six-carbon substrate, glucose.

A. Reactions

Glycolysis can be divided into two main stages. The stage I, which comprises the first five reactions, is an endergonic (i.e. energy-requiring) stage in which glucose molecules are activated through the expenditure of ATP molecules. Thus, this stage involves energy investment. Stage II leads to energy generation and hence referred to as the *"payoff stage"*.

Reactions of glycolysis are shown in Figure 9.3.

Stage I: The Investment Phase

Reaction 1: Phosphorylation of Glucose

Glucose is phosphorylated at the sixth carbon to form glucose 6-phosphate. This places a large negative charge on the glucose molecule so that it cannot simply diffuse out of the cell. The phosphate group is obtained from ATP which is converted to ADP in the process, which means that energy is lost in this step. However, this is better than losing glucose out of the cell.

$$\text{Glucose} + \text{ATP} \rightarrow \text{Glucose 6-phosphate} + \text{ADP}$$

The enzyme involved in this reaction is *kinase*, (an enzyme that catalyzes a phosphorylation and uses ATP as its source of phosphate is called *kinase*). There are two possible enzymes for this reaction, depending on the tissue. One is *glucokinase* (found in liver) and the other is *hexokinase* (muscle and fat), which will catalyze phosphorylation of most hexoses, including glucose. Kinetic properties of the two enzymes are different (Table 9.1).

Hexokinase has low K_m, i.e. high affinity for glucose, low V_{max}, and is subject to feedback inhibition by the reaction product, glucose 6-phosphate. The *hexokinase* reaction is not specific for glycolysis; it only commits glucose to intracellular *metabolism* because glucose 6-phosphate is not transported across the plasma membrane.

Glucokinase, on the other hand, is specific for glucose. It has high V_{max} and, unlike *hexokinase,* is not inhibited by the reaction product, glucose 6-phosphate. Moreover, it has high K_m (20 mmol/L), i.e. low affinity for glucose and therefore, comes into play only when intracellular glucose concentration rises steeply, such as following a carbohydrate rich diet. These properties make the *glucokinase* suitable for rapidly phosphorylating the glucose that is presented to liver following a meal. Within the hepatocytes, the glucose molecules are rapidly phosphorylated by *glucokinase* (due to high V_{max} and high K_m), without any hindrance to its action (due to lack of feedback inhibition). The above transformation is further enhanced because of induction of *glucokinase* synthesis after a meal, especially following a carbohydrate rich diet. The enzyme induction is probably brought about by insulin, that is released at such time. Large quantities of glucose 6-phosphate are thereby produced, which are channeled into the pathway of glycogen synthesis (i.e. glycogenesis).

Glucokinase can rapidly phosphorylate large amount of glucose after meals, which then enters the pathway of glycogen synthesis.

Reaction 2: Isomerization of Glucose 6-Phosphate

Glucose 6-phosphate, an aldose sugar, is now prepared for subsequent cleavage into two 3-carbon molecules. To split anything is easier if it is bilaterally symmetrical. Glucose 6-phosphate is, however, not very symmetrical, and so converted to fructose 6-phosphate, a little more symmetrical molecule. It is a freely reversible reaction catalyzed by *phosphohexose isomerase*.

Reaction 3: Phosphorylation of Fructose 6-Phosphate

Fructose 6-phosphate, formed in the previous step, is made more symmetrical by esterifying a phosphate to C-1. This yields fructose 1,6-bisphosphate, a molecule that is almost symmetrical. The phosphate group comes from ATP and the enzyme is *phosphofructokinase* (PFK_1).

Note: A diphosphate differs from bisphosphate (and a triphosphate from trisphosphate) in that, for a diphosphate (e.g. ADP) the phosphates are joined to each other, whereas in a bisphosphate (e.g. fructose 1, 6-bisphosphate), the phosphate groups are attached at different places on the sugar.

This reaction is irreversible and is the *"rate-limiting step"* for glycolysis. It is also the *committed step*, meaning that once fructose 1, 6-bisphosphate is formed it must go for the glycolytic pathway only. Activity of PFK_1 is controlled by a variety of positive and negative regulatory modulators, which make this step the most important regulatory step of the pathway. Regulation of glycolysis is discussed later in this chapter.

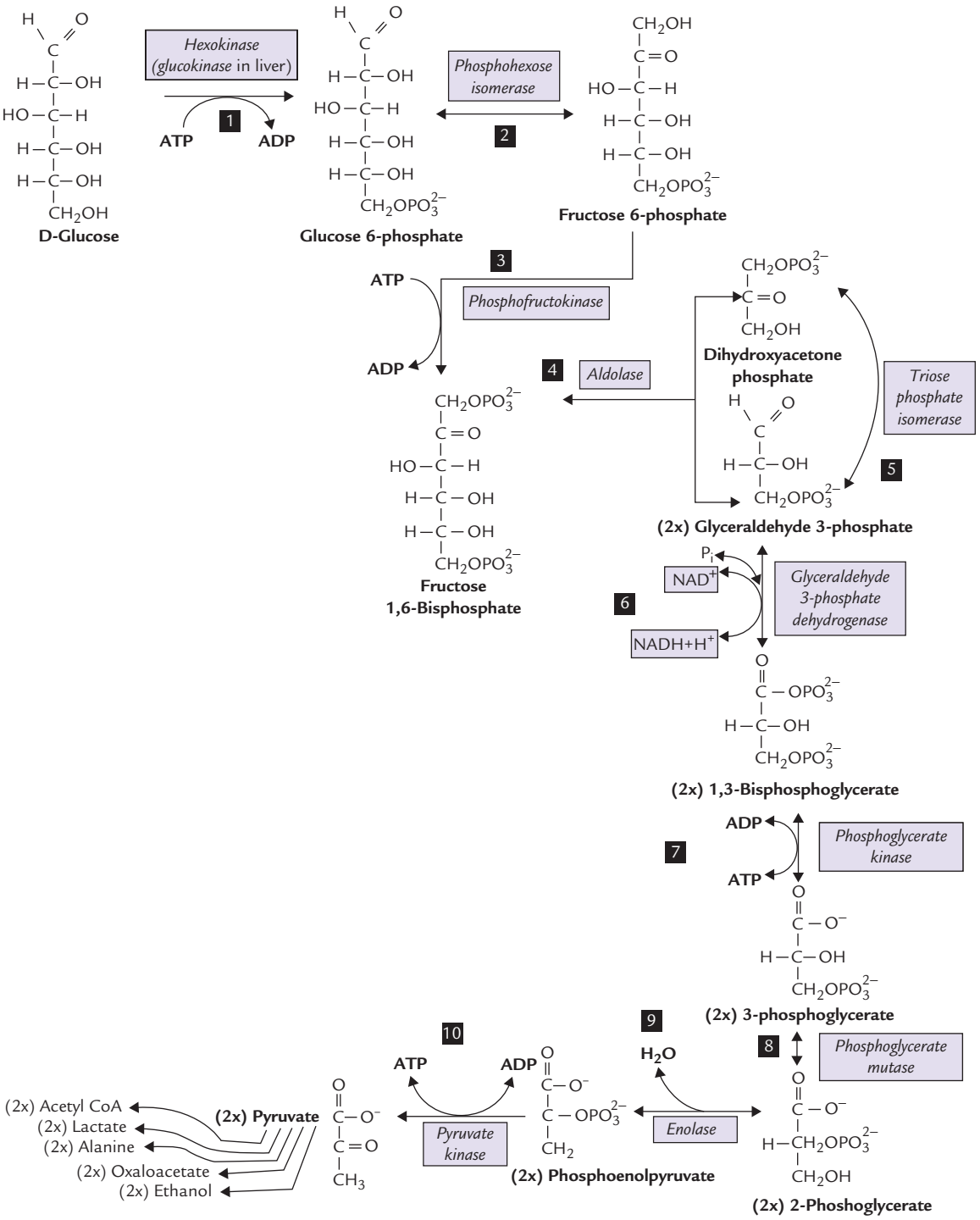


Fig. 9.3. The pathway of glycolysis and fate of pyruvate. **Stage I:** Reactions 1–5—energy-consuming reactions and **Stage II:** 6–10—energy-yielding reactions. **Irreversible steps:** 1st, 3rd and 10th).

Table 9.1. Comparative properties of *hexokinase* and *glucokinase*

	Hexokinase	Glucokinase
Specificity	Broad (hexoses)	Narrow (glucose only)
Feedback inhibition	Yes	No
Affinity for glucose	High	Low
K_m value	10^{-2} mmol/L	20 mmol/L
V_{max} value	Low	High

Reaction 4: Cleavage of Fructose 1,6-Bisphosphate

Fructose 1,6-bisphosphate, a 6-carbon compound, is cleaved by the enzyme *aldolase* into two trioses: glyceraldehyde 3-phosphate and dihydroxyacetone phosphate. This enzyme, which splits an aldose, is not specific for fructose 1,6-bisphosphate, hence its general name.

Reaction 5: Isomerization of Dihydroxyacetone Phosphate

Of the two trioses, only glyceraldehyde 3-phosphate can be further metabolized through the glycolytic sequence.

Dihydroxyacetone phosphate, therefore, must be converted to glyceraldehyde 3-phosphate for further metabolism. This aldose-ketose isomerization is catalyzed by the enzyme, *triose phosphate isomerase*.

Fructose 1,6-bis Ⓟ ⟶ Dihydroxyacetone-Ⓟ + Glyceraldehyde 3-Ⓟ

Sum: Fructose 1,6-bis-Ⓟ ⟶ (2x) Glyceraldehyde 3-Ⓟ

It is necessary to understand that two glyceraldehyde 3-phosphate molecules are generated from the original glucose molecule and enter this part of the pathway, and therefore, *all of the following steps occur twice, once for each of the two glyceraldehyde 3-phosphate molecules*.

In the first stage of glycolysis, glucose is phosphorylated, isomerized, phosphorylated again and cleaved to yield two triose molecules. These reactions consume 2ATPs per glucose. All reactions beyond this point occur twice for each glucose molecule.

Stage II: Energy Pay-off Phase

The next few steps are remarkable for they bring about oxidation of each of the three carbons of glyceraldehydes. There is addition of oxygen at C-1, loss of hydrogen at C-2, and loss of phosphate at C-3. In these steps a large amount of energy is liberated, enough to replace the energy lost so far (2ATPs) plus some more.

Reaction 6: Dehydrogenation of Glyceraldehyde 3-Phosphate

Dehydrogenation of glyceraldehyde 3-phosphate is catalyzed by the enzyme *glyceraldehyde 3-phosphate dehydrogenase*, resulting in the formation of 1,3-bisphosphoglycerate.

P_i, NAD^+ ⟶ $NADH, H^+$

Glyceraldehyde 3-phosphate ⟶ 1,3-Bisphosphoglycerate

There are two interesting aspects of this reaction: firstly, the pair of hydrogen atoms removed is replaced by a phosphate group. This hydrogen is transferred to NAD^+ to form NADH, which is a valuable product (its oxidation in the respiratory chain is a source of ATP). Second, the phosphate group added to the glyceraldehyde 3-phosphate molecules is inorganic phosphate. This means that energy is not lost, which would have been the case if it had come from ATP.

Reaction 7: ATP Production from 1,3-Bisphosphoglycerate

The 1,3-bisphosphoglycerate (1,3-BPG) contains a carboxylic anhydride with phosphate at C-1. This type of anhydride, called acyl-phosphate, has a very high standard free energy of hydrolysis; the $\Delta G^{0\prime}$ of 1,3-bisphosphoglycerate is −11.8 kcal/mole (Chapter 8). The cell uses this energy by transferring the phosphate to ADP to form ATP (i.e. **substrate level phosphorylation**).

This reaction is catalyzed by the enzyme *phosphoglycerate kinase*. Since two molecules of 1,3-bisphosphoglycerate are produced from a single molecule of glucose, two ATPs are actually generated in this step from a single glucose molecule.

The 1,3-BPG is converted to an alternate metabolite, 2,3-bisphosphoglycerate (2,3-BPG) in erythrocytes which enhances unloading of oxygen (Box 9.1).

Reaction 8: Inter-Molecular Shift of Phosphate Group

The phosphate group is transferred from the C-3 of the 3-phosphoglycerate to the C-2 by the enzyme *phosphoglycerate mutase*, resulting in the formation of 2-phosphoglycerate.

Note: The term *mutase* is used to designate the enzyme catalyzing transpositioning of functional group, within the same molecule.

Reaction 9: Dehydration of 2-Phosphoglycerate

Removal of a water molecule from 2-phosphoglycerate is catalyzed by the enzyme *enolase*, resulting in formation of phosphoenolpyruvate (PEP). In this reaction, redistribution of energy occurs within the same molecule, so that the phosphate ester (PEP) becomes unusually energy rich ($\Delta G^{0\prime} = -14.8$ kcal/mole), which is sufficient for generation of ATP (i.e. *substrate level phosphorylation*).

Inhibition of *enolase* catalyzing this step has important role in clinical biochemistry (Box 9.2).

Reaction 10: ATP Production from Phosphoenol Pyruvate

The last step of the glycolytic sequence involves transfer of high energy phosphate group from PEP to ADP to yield a molecule of ATP (Fig. 9.4).

This is the **third irreversible step of glycolysis (reaction 1 and 3 are the other two)**. *Pyruvate kinase*, the enzyme catalyzing this step, is a regulatory enzyme.

Deficiency of *pyruvate kinase* causes decreased production of ATP from glycolysis. Red blood cells are predominantly affected in this disorder. They are unable to generate sufficient ATP for their sodium pumps. Consequently, the ionic gradient across the erythrocyte membrane cannot be maintained, resulting in lysis of the cell. This results in *haemolysis*, and hence haemolytic anaemia.

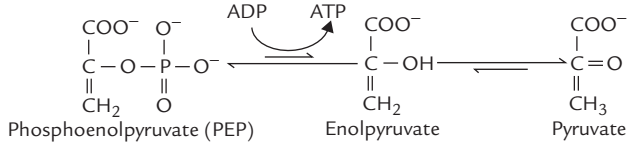


Fig. 9.4. The *pyruvate kinase* reaction. Pyruvate shows keto-enol tautomerism, the keto form being energetically far more stable than the enol form.

BOX 9.1

Role of 2,3-Bisphosphoglycerate in Erythrocytes

In erythrocytes, 1,3-BPG is converted to 2,3-bisphosphoglycerate (2,3-BPG) by catalytic action of the enzyme *2,3-bisphosphoglycerate mutase*. The 2,3-BPG concentration sometimes reaches 5 mmol/L, comparable with the molar concentration of haemoglobin in the RBC, and much higher than that of ATP (1–2 mmol/L) or inorganic phosphate (1 mmol/L). 2,3-BPG is a negative allosteric effector of the oxygen affinity of haemoglobin: it decreases the affinity to promote the release of oxygen (from oxyhaemoglobin) in peripheral tissues. Its concentration increases in the RBCs, during adaptation to higher altitudes and in anaemia, thus promoting the release of oxygen to tissues when the oxygen transport capacity of blood is decreased. Fetal haemoglobin is less sensitive than adult haemoglobin to the effects of this compound, so that 2,3-BPG in maternal erythrocytes is one factor that promotes efficient transfer of oxygen across the placenta from HbA to HbF (Chapter 17).

Finally, 2,3-BPG is hydrolyzed to 3-phosphoglycerate by *bisphosphoglycerate phosphatase*. It is postulated that it is a bifunctional enzyme catalyzing both the synthesis of 2,3-BPG (from 1,3-BPG) and its subsequent hydrolysis.

More detailed discussion on 2,3 BPG-is given in Chapter 17.

Note: This pathway of synthesis and breakdown of 2,3-BPG in erythrocytes causes bypassing of the *glycerophosphate kinase* reaction (step 6 of glycolysis) and is referred to as **Rapaport–Leubering cycle** (or simply BPG cycle). In this shunt pathway, no ATP is generated. Normally, about 15–25% of the glucose converted to lactate in erythrocytes is routed via this cycle. This is useful not only in the oxygen unloading by oxyhaemoglobin, but also in dissipation of energy, and therefore, advantageous to the cell when the energy requirement is minimal.

BOX 9.2

Inhibition of Glycolysis

The glycolytic enzyme *enolase* is inhibited by fluoride, which thereby inhibits the glycolytic sequence. Sodium fluoride is mixed with potassium oxalate, an anticoagulant, in the blood sample collected for glucose estimation. Because it inhibits *enolase*, in vitro glycolysis is prevented. In absence of sodium fluoride, glucose will be used.

Some other inhibitors of glycolysis are as below:

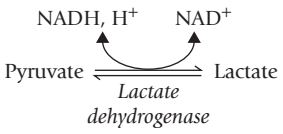
1. Iodoacetate, which inhibits the enzyme *glyceraldehyde 3-phosphate dehydrogenase*.
2. Arsenite, which inhibits *phosphoglycerate kinase*.

In the second stage of glycolysis, a series of changes convert glyceraldehyde 3-phosphate to pyruvate. This stage produces 4 ATPs per glucose molecule for a net yield of 2 ATPs per glucose.

B. Fate of Pyruvate

There are several pathways into which pyruvate can enter (Fig. 9.3). The pathway chosen in a given tissue depends on its state of oxygenation and the prevailing metabolic conditions, as described below:

1. *Oxidative decarboxylation to acetyl CoA:* In tissues that are adequately perfused with oxygen (i.e. under aerobic conditions), pyruvate undergoes oxidative decarboxylation to form acetyl CoA, which is further oxidized via citric acid cycle.
2. *Reduction to lactate:* Lactate is produced by glycolysis under anaerobic conditions and the process is called **anaerobic glycolysis**. The pyruvate to lactate reduction is brought about by the cytosolic enzyme *lactate dehydrogenase* and requires NADH.



The NADH is produced in the *glyceraldehyde 3-phosphate dehydrogenase* reaction (Reaction 6) of glycolysis (Fig. 9.5). The pyruvate to lactate conversion regenerates NAD^+ for the above reaction so that glycolysis can proceed normally. It is noteworthy, that without regeneration of NAD^+ by *lactate dehydrogenase* reaction, the Reaction 6 would stop and the glycolytic pathway would also soon be halted for lack of NAD^+. Thus anaerobic glycolysis

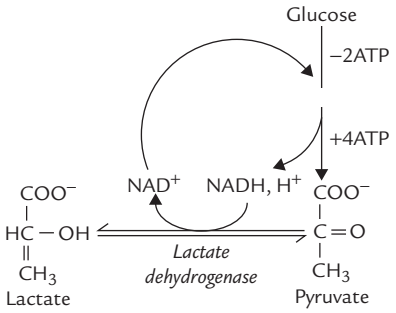


Fig. 9.5. Outline of reactions of anaerobic glycolysis. It regenerates NAD^+ (for the *glyceraldehyde 3-phosphate dehydrogenase* reaction) so that glycolytic sequence can proceed smoothly with generation of 2ATP molecules.

provides elegantly simple solution to ensure that glycolysis proceeds uninterruptedly.

More about anaerobic glycolysis is discussed at end of this section.

Under aerobic conditions, pyruvate enters the mitochondrion, where it is oxidatively decarboxylated to acetyl CoA, and under anaerobic conditions, pyruvate is reduced to regenerate NAD^+ for glycolysis.

3. *Carboxylation to oxaloacetate:* Pyruvate is converted to oxaloacetate by biotin-dependent carboxylation catalyzed by the enzyme *pyruvate carboxylase*. The oxaloacetate can form glucose via gluconeogenesis.
4. *Conversion to malate*: *Malic enzyme* converts pyruvate to malate, which can also enter gluconeogenic sequence.
5. *Conversion to alanine*: Pyruvate may form the amino acid, alanine by transamination.
6. *Reduction to ethanol*: Conversion of pyruvate to ethanol occurs in yeasts and some bacteria, including those of the intestinal flora.

Pyruvate can be funneled into various metabolic pathways. It can also be generated from a number of compounds, as shown in Figure 9.6. Evidently, *pyruvate is said to be at the metabolic cross-roads*.

Anaerobic Glycolysis

Anaerobic glycolysis **occurs in tissues that lack mitochondria**, such as *erythrocytes, lens* and *cornea of the eye, renal medulla*, and *leukocytes*. In oxygen deficient tissues such as **severely exercising muscles**, this pathway is chosen. Oxygen deficiency occurs because the blood is squeezed out of the contracting muscles. Moreover, production of pyruvate in exercising muscle is very fast due to rapid rate of glycolysis, but its oxidation is much slower due to oxygen deficiency. The surplus pyruvate is diverted for lactate formation (Fig. 9.5).

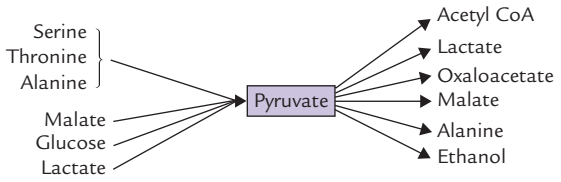


Fig. 9.6. Pyruvate can be generated from several compounds, and can be channeled into various pathways.

Finally, during hypoxia resulting from vascular obstruction or other causes, for example, in acute **myocardial infarction** – anaerobic glycolysis occurs to help the cells survive the brief episodes of oxygen depletion.

Significance

Anaerobic glycolysis provides less energy: net gain of ATPs is only two molecules (Fig. 9.5). However, it is an important process because of two reasons:

1. NAD^+ is generated during their process, which is required for step 6 of glycolysis (Fig. 9.5). As discussed in the previous section, NAD^+ regeneration must occur for, glycolysis to continue.
2. Carbohydrates are the only metabolic substrates that can produce ATP under anaerobic conditions.

C. Generation of ATP by Glycolysis

Energy yield in the anaerobic and the aerobic states are different, and therefore, discussed separately.

Aerobic State

In course of conversion of one molecule of glucose through glycolytic sequence to two molecules of pyruvate, there occurs generation of four molecules of ATP and, utilization of two molecules of ATP.

Initially, two ATPs are utilized to phosphorylate glucose to glucose 6-phosphate and fructose 6-phosphate to fructose 1,6-bisphosphate (see reactions 1 and 3; Fig. 9.3). Subsequently, cleavage of fructose 1,6-bisphosphate, followed by isomerization of dihydroxyacetone phosphate, produces two molecules of glyceraldehyde 3-phosphate. Therefore, from each molecule of glucose, two molecules of triose phosphate are produced. Each molecule of triose phosphate then generates two molecules of ATP: one ATP is produced by *phosphoglycerate kinase* (Reaction 7) and the other by *pyruvate kinase* (Reaction 10). Therefore, two molecules of triose-phosphate generates four ATPs. *With two molecules of ATP invested and four generated, net gain of ATPs is two molecules* (Table 9.2).

In addition to these, generation of more ATPs occurs by oxidation of the NADH, (generated earlier in *glyceraldehyde 3-P dehydrogenase* reaction). Each of these NADH

Table 9.2. Energy yield during the conversion of one molecule of glucose to two molecules of pyruvate in aerobic glycolysis*

Reaction	Enzyme	Product
1st	*Hexokinase*	−1 ATP
3rd	*Phosphofructokinase*	−1 ATP
6th	*Glyceraldehyde 3-phosphate dehydrogenase*	+ 2 NADH
7th	*Phosphoglycerate kinase*	+ 2 ATP
10th	*Pyruvate kinase*	+ 2 ATP
		2 ATP + 2 NADH

* Note that all reactions beyond the *aldolase* reaction occur twice for each glucose molecule.

molecules yields three ATPs by oxidative phosphorylation (Chapter 14). Thus, a total of six ATPs are produced from this source. When two ATPs generated earlier are added, the *net production of ATPs in aerobic glycolysis becomes* **eight**.

Anaerobic State

Generation of two molecules of ATP and two molecules of NADH occurs initially in this type also. The latter are oxidized to NAD^+ during reduction of pyruvate. Thus, there is no net production of NADH (Fig. 9.5) in anaerobic glycolysis and, therefore, net generation of only **two ATP** occurs.

Lactate, the end product of anaerobic glycolysis, still contains a large amount of inherent free energy, but is not degraded further. Thus anaerobic glycolysis releases only a small fraction of the energy of the glucose molecules. Still it is a valuable source of energy in tissues lacking mitochondria. Moreover, in case of oxygen depletion, the affected tissues depend more on anaerobic glycolysis for getting the required energy. Production of lactate is thereby increased. However, excessive production of lactate may have serious consequences **(Case 9.1).**

Cori Cycle or Lactic Acid Cycle

In an **actively exercising muscle,** less than 10% of pyruvate is utilized by the citric acid cycle, and the rest is reduced to lactate (i.e. anaerobic glycolysis). Accumulation of lactate and consequent fall in pH is potentially hazardous and should be prevented. Mechanisms exist in body which prevent such excessive accumulation of lactate.

The lactate first diffuses out into the blood circulation, the plasma membrane of muscle cells being freely permeable to lactate. It is carried to liver, and is oxidized to pyruvate in hepatocytes by the *lactate dehydrogenase* reaction. Pyruvate, so produced, enters the gluconeogenic sequence and converted to glucose, which is then transported to skeletal muscles (Fig. 9.7).

Thus, a cyclic process is set up between liver and muscle, which ensures efficient reutilization of lactate by the body. It is referred to as the Cori cycle, named after Carl Cori and Gerty Cori who were awarded Nobel prize in 1947 for this discovery.

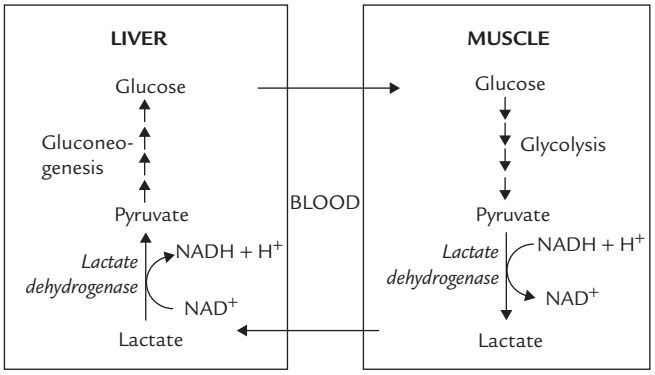


Fig. 9.7. Cori cycle. The exercising muscles generate lactate via anaerobic glycolysis, which is transported to liver for resynthesis of glucose.

Vigorously exercising muscles generate lactate via anaerobic glycolysis. During recovery, some of this lactate is transported to liver and used to form glucose via gluconeogenesis. The overall pathway (glucose → lactate → glucose) constitutes the Cori cycle.

D. Regulation

Glycolysis is central to metabolism and is integrated with a number of other metabolic pathways. Therefore, the regulatory enzymes of glycolysis respond to appropriate signals from several other pathways. This makes the regulation of glycolysis a complex event. There are three glycolytic reactions catalyzed by *hexokinase* (or *glucokinase*), *phosphofructokinase* (PFK_1) and *pyruvate kinase*, which are metabolically irreversible and subject to regulation. *The main control step is that catalyzed by PFK_1 but hexokinase and pyruvate kinase are additional control sites.*

Regulation of *Hexokinase/Glucokinase*

Hexokinase activity is subject to feedback inhibition by glucose 6-phosphate. This prevents intracellular accumulation of glucose 6-phosphate.

Glucokinase, that specifically effects phosphorylation of glucose in liver, is an inducible enzyme. Availability of substrate glucose induces the enzyme synthesis, probably through insulin. It is free from feedback inhibition by glucose 6-phosphate.

Regulation of *Phosphofructokinase*

Most important control point for glycolysis is through regulation of *phosphofructokinase* (PFK_1), a complex allosteric enzyme. Its activity is influenced by **cellular energy charge** (high energy charge inhibits PFK_1 and low energy

BOX 9.3

Rate of Glycolysis is Regulated by Cellular Energy Charge

An overview of regulation of glycolysis shows that its rate is adjusted according to the energy content of the cell. In general, when the cellular energy content is low, the catabolic pathways are accelerated so as to provide the required energy; and conversely, when the cell has sufficient energy, inhibition of these pathways occurs. The anabolic pathways are regulated in a reciprocal manner, i.e. stimulated by high energy content and inhibited by low energy content of the cell. Evidently, cellular energy content is an important determinant of the rate of various metabolic pathways, including glycolysis. It is reflected by the following equation:

$$\frac{\text{ATP}}{\text{ATP} + \text{AMP}}$$

It is commonly referred to as the energy charge of the cell. Low energy charge signals that the cell needs energy, and accordingly the catabolic pathways, including glycolysis, are stimulated. When the energy charge is high, inhibition of glycolysis occurs.

charge stimulates it; Box 9.3), and by a number of positive and negative **allosteric modulators** (Table 9.3):

1. ***ATP:*** ATP is an important inhibitory modulator of the enzyme activity.
   - When the intracellular ATP level rises (indicating high cellular energy charge), the enzyme activity is inhibited. This is because of binding of ATP to the allosteric site of the PFK_1, which results in diminished affinity of the enzyme for fructose 6-phosphate.
   - Conversely, low intracellular ATP level results in removal of ATP from the allosteric site, which results in enhanced activity of the enzyme.
2. ***Citrate:*** The inhibitory effect of ATP on PFK_1 is enhanced by citrate, an intermediate of TCA cycle. Regulation of glycolysis by an intermediate of citric acid cycle ensures that rates of these two pathways, keep pace with each other. Whenever TCA cycle is impeded, accumulation of citrate results, which in turn slows down glycolysis as well.
3. ***AMP:*** Effect of AMP on PFK_1 is stimulatory since it opposes the inhibitory effect of ATP on this enzyme. Reciprocal influences of ATP and AMP have additive effect on PFK_1. This is because when intracellular AMP level is high, ATP level is correspondingly low and both (low ATP and high AMP) stimulate PFK_1, thereby speeding up the rate of glycolysis.
4. ***Fructose 2,6-bisphosphate (F-2,6-BP):*** This compound is the *most important* allosteric modulator (**activator**) of PFK_1 and, therefore, of glycolysis in liver. It is synthesized from fructose 6-phosphate by *phosphofructokinase 2* (PFK_2), a different enzyme from PFK_1, and is hydrolyzed back to fructose 6-phosphate by *fructose 2,6-bisphosphatase 2* (FBPase2). Interestingly, both PFK_2 and FBPase2 are activities catalyzed by the same polypeptide; hence this is a **bifunctional enzyme** (Fig. 9.8).

Table 9.3. Allosteric modulators of *phosphofructokinase* (PFK_1)

Activators	Inhibitors
AMP	ATP
Fructose 2,6-bisphosphate	Citrate
ADP	Ca^{2+}
K^+	Mg^{2+}
Phosphate	Low pH

Covalent Modulation of PFK_2

Activity of PFK_2 is regulated by phosphorylation-dephosphorylation mechanism.

- The *dephosphorylated* enzyme acts as *kinase*
- The *phosphorylated* enzyme acts as *phosphatase.*

The following hdormones regulate interconversion of the phosphorylate/dephosphorylated forms:

Glucagon: It acts through the second messenger cAMP (and cAMP activated protein *kinase A*) and phosphorylates the enzyme. This results in activation of the *phosphatase* activity of PFK_2, causing hydrolysis of fructose 2,6-bisphosphate. This decreases activity of PFK_1 and, therefore, the rate of glycolysis falls. Thus, the overall effect is:

Glucagon → cAMP induced phosphorylation → stimulation of *phosphatase* activity of PFK_2 → ↓ concentration of fructose 2,6-bisphosphate → ↓ activity of PFK_1, and decreased rate of glycolysis.

Insulin: opposes glucagon by lowering the cellular cAMP concentration that promotes dephosphorylation of the enzyme and hence stimulates its *kinase* activity.

Thus, by adjusting the cellular concentration of fructose 2,6-bisphosphate, insulin and glucagon can influence the

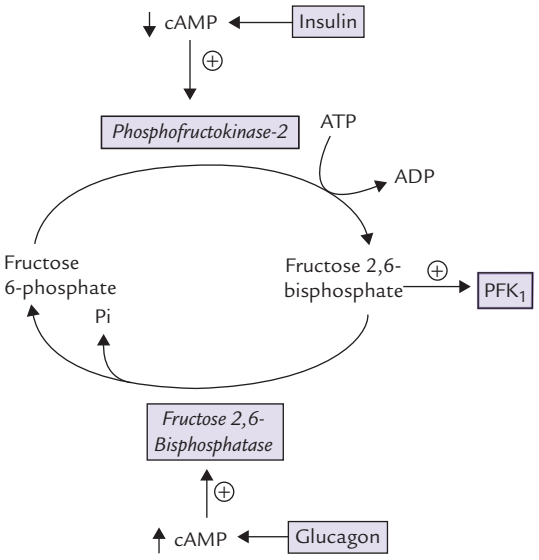


Fig. 9.8. Regulation of *phosphofructokinase-2* (PFK_2), an enzyme having *kinase* activity and *phosphatase* activity. The *phosphatase* activity is stimulated by glucagon (via cAMP induced phosphorylation), and the *kinase* activity by insulin.

glycolytic activity (and gluconeogenic activity, discussed later) within minutes.

Fructose 6-phosphate is a potent modulator of PFK_2-activity. It not only stimulates the synthesis of F-2,6-BP, but also inhibits its hydrolysis. F-2,6-BP in turn strongly activates PFK, and hence stimulates glycolysis. **The overall effect is that when fructose 6-phosphate levels are high, such as after meals, the PFK, (and hence) glycolysis is stimulated**.

Because of various aforementioned regulatory influences the activity of PFK_1 can increase up to **700-fold** from the basal resting level.

The glycolytic reactions catalyzed by *hexokinase, phosphofructokinase* and *pyruvate kinase* are metabolically irreversible and subject to regulation. *Phosphofructokinase* is most important, of which fructose 2,6-bisphosphate is most important regulator.

Regulation of *Pyruvate Kinase*

Pyruvate kinase is next in importance to PFK_1 in the regulation of glycolysis. It has three isozyme forms, all catalyzing the final reaction (Reaction 10) of the glycolytic sequence. Its activity is regulated allosterically and by covalent modulation.

Allosteric regulation: *Pyruvate kinase* is inhibited allosterically by ATP and alanine and activated by fructose 1,6-bisphosphate. The concentration of fructose 1,6-bisphosphate is increased whenever PFK_1 activity increases, and so its effect on *pyruvate kinase* is an example of **feed-forward stimulation.**

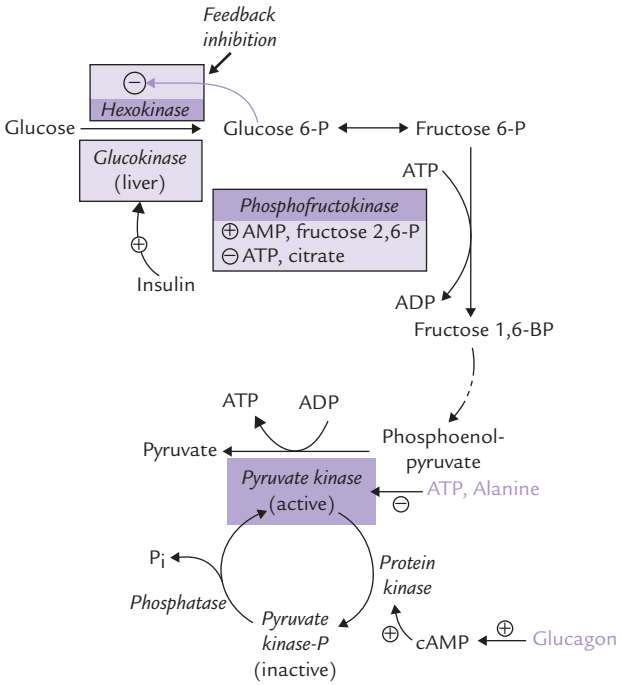


Fig. 9.9. Regulation of glycolysis.

Covalent modulation: *Pyruvate kinase* is inhibited by the cAMP-induced phosphorylation of a serine side chain in the enzyme protein (Fig. 9.9). Thus *glucagon*, which acts via cAMP-dependent *protein kinase, inhibits* the enzyme activity (thereby inhibiting hepatic glycolysis); and conversely insulin stimulates it.

Pyruvate kinase is activated by fructose 1,6-bisphosphate but allosterically inhibited by ATP and alanine. Like PFK, it is also regulated hormonally by glucagon.

E. Diseases Associated with Glycolysis

Excessive accumulation of lactic acid occurs in a number of conditions, described in Chapter 1, to cause **lactic acidosis**. It causes severe metabolic acidosis, and is a potentially lethal condition (**Case 9.1**).

Deficiencies of glycolytic enzymes, most commonly of *pyruvate kinase* and *hexokinase*, have been reported. In these rare conditions, the predominant clinical manifestation is *haemolytic anaemia* because energy depleted erythrocytes are easily destroyed. A rare inherited disease causing *pyruvate dehydrogenase* (*PDH*) *depletion* has been reported (incidence 1 in 250,000 births). Metabolism of pyruvate being blocked, *lactic acidosis* frequently develops. Blockage of the PDH reaction can also occur in **beriberi** (thiamine pyrophosphate being a cofactor in PDH) and among **alcoholics** (**Case 18.1**).

III. Feeder Pathways

Glycolysis is not exclusively for the catabolism of glucose; many other carbohydrates enter the glycolytic sequence in course of their metabolism. *The metabolic pathways by which various monosaccharides, disaccharides and polysaccharides enter the glycolytic sequence are called* ***feeder pathways.*** Separate feeder pathways exist for various monosaccharides, disaccharides and polysaccharides.

A. Feeder Pathways for Monosaccharides

Metabolism of Fructose

D-Fructose is present in free form in many fruits, and honey. The major dietary source is sucrose (cane sugar), a disaccharide consisting of glucose and fructose. In the body, **entry of fructose into the cells is not dependent on insulin**. This is in contrast to glucose, which requires insulin for this purpose.

The Pathway

Some of the fructose in the cells is phosphorylated to the glycolytic intermediate fructose 6-phosphate by ***hexokinase*** in muscle and adipose tissue. Affinity of *hexokinase* for fructose, however, is extremely low (Km approximately 20 times higher than that for glucose), so this pathway is normally insignificant, and is important only when the fructose concentration is very high. Most of the fructose is phosphorylated by the enzyme ***fructokinase*** in the liver, kidney, and intestine. The liver accounts for nearly half of the total fructose metabolism. *Fructokinase* catalyzes phosphorylation of fructose at C-1 to form fructose 1-phosphate (Fig. 9.10). The latter is cleaved into dihydroxyacetone phosphate (DHAP) and glyceraldehyde by the enzyme ***aldolase B***. Defective action of this enzyme leads to a disorder called **hereditary fructose intolerance (Case 9.2)**.

Aldolase B is an isoenzyme of the *aldolase* (see Reaction 4-glycolysis), and it can cleave both fructose 1,6-bisphosphate and fructose 1-phosphate. Glyceraldehyde is phosphorylated by the enzyme ***triokinase*** to glyceraldehyde 3-phosphate, which along with DHAP, is metabolized further by glycolysis or gluconeogenesis.

Fructose is absorbed from the intestine less rapidly than glucose, but once in the blood it is metabolized nearly twice as rapidly as glucose. It bypasses the *phosphofructokinase* step in liver, and therefore, its metabolism is simpler than that of glucose (Box 9.4).

Disorders of Fructose Metabolism

Essential fructosuria: It is a rare medical problem associated with fructose metabolism. The enzyme *fructokinase* is deficient in this disorder and therefore, fructose cannot be metabolized as rapidly as in normal subjects. Plasma fructose levels rise, and fructose may appear in urine. The condition is asymptomatic, sometimes detected incidentally during urinalysis when

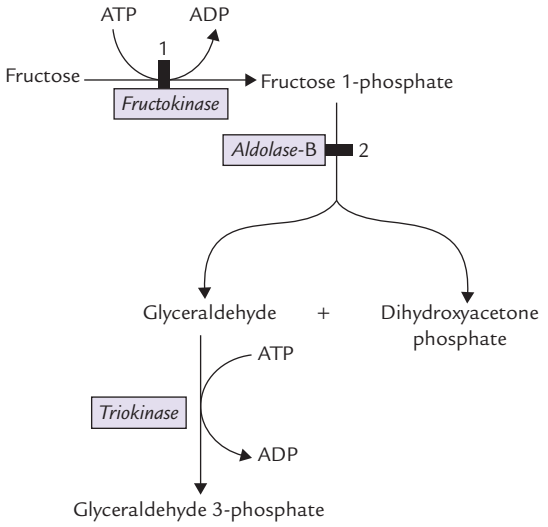


Fig. 9.10. Conversion of fructose to intermediates of glycolysis in liver. (▬ = metabolic blocks: 1. in essential *fructokinase*, 2. in hereditary fructose intolerance).

BOX 9.4

Is Fructose Useful in Parenteral Nutrition

Metabolism of fructose bypasses the *phosphofructokinase* step in liver, and therefore simpler than metabolism of glucose. Because of this, fructose was considered to be more useful than glucose in patients requiring parenteral nutrition. However, use of fructose in such patients has severe limitation. Fructose is rapidly phosphorylated to fructose 1-phosphate in hepatocytes. This reaction is so fast that a much larger amount of fructose 1-phosphate is generated than can be metabolized. This ties up the available inorganic phosphates, thereby depleting the liver cell of P_i and ATP. Consequences of such depletion are highly hazardous. Therefore, use of fructose is not recommended in parenteral nutrition.

"reducing sugar" (fructose) is present. Restriction of dietary fructose is effective to treat this.

Hereditary fructose intolerance: These are disorders caused by the deficiency of one of the fructose metabolizing enzymes. The enzymes involved are:

1. *Aldolase B:* Deficiency of *aldolase B* results in intracellular accumulation of fructose 1-phosphate. There is vomiting, jaundice, and hypoglycaemia caused by intracellular accumulation of fructose 1-phosphate. This compound allosterically inhibits liver *phosphorylase* to block glycogenolysis, thus leading to hypoglycaemia. Besides tying up phosphate, and thereby impairing ATP synthesis, fructose 1-phosphate inhibits *aldolase* and *phosphohexose isomerase*; these changes result in **liver** damage, and jaundice is commonly seen. More about hereditary fructose intolerance is given in **Case 9.2.**
2. *Fructose 1,6-bisphosphatase:* Deficiency of this gluconeogenic enzyme causes fructose intolerance similar to *aldolase-B* deficiency, but these patients also have fasting hypoglycaemia. They can form glucose from stored glycogen, but gluconeogenesis is blocked. Glycogen degradation can maintain a reasonably normal blood glucose level for many hours, but the defect in gluconeogenesis results in dangerous hypoglycaemia when the period of fasting exceeds 14–18 hours.

Inborn errors of fructose metabolism cause hypoglycaemia and liver damage.

Excess Fructose is Hazardous for Body

Based on the reasoning that insulin-dependent PFK_1 reaction is bypassed by fructose, diabetic diets were formulated in the past with fructose as the predominant carbohydrate. It was soon found, however, that excess fructose is toxic because it can cause damaged liver and cause increased lactate formation, hypertriglyceridaemia and hyperuricaemia, not only in individuals with diabetes but in normal subjects as well.

Liver damage: The activity of *fructokinase* far exceeds that of *aldolase B*, so fructose 1-phosphate tends to accumulate (concentration in liver can reach up to 10 μmol/g). This compound allosterically affects several enzymes of carbohydrate metabolism and also ties up substantial portion of the phosphate in the cell. This can impair oxidative phosphorylation, and lower the synthesis of ATP from ADP, with consequent damage to liver cells.

Hyperlactataemia: This can be traced to a rapid metabolism of fructose to pyruvate, which is in equilibrium with lactate.

Hypertriglyceridaemia: Fructose is rapidly metabolized to yield acetyl CoA, which is channeled into lipogenesis (fatty acids and triglycerides synthesis) in the liver.

Hyperuricaemia: This is due to excessive accumulation of ADP and AMP (due to lack of Pi) followed by their degradation to uric acid.

Metabolism of Galactose

D-Galactose is present in milk as one of the monomeric constituents of lactose; the other one is glucose. Galactose is liberated in the intestine by action of the brush border enzyme, ***lactase***.

The Pathway

Entry of galactose into peripheral cells is independent of insulin. Galactose is phosphorylated at C-1 by the enzyme *galactokinase* to form galactose 1-phosphate. Galactose 1-phosphate is converted to glucose 1-phosphate by the action of *galactose 1-phosphate uridyl transferase*. In this reaction, uridine diphosphate (UDP) acts as a carrier of hexose molecule (Fig. 9.11). The net reaction of the pathway is the ATP-dependent conversion of galactose to glucose 1-phosphate. The latter enters glycolytic sequence via glucose 6-phosphate.

UDP-galactose is an important intermediate formed in the second reaction (from galactose 1-phosphate and UDP-glucose). It can be epimerized to UDP-glucose, as shown in the Figure 9.11. The reversibility of epimerization makes it possible to obtain UDP-galactose from UDP-glucose, so that *galactose is not an essential nutrient in diet*.

UDP-Galactose has synthetic roles, being involved in the following reactions:

1. It is an active donor of galactose residue during synthesis of lactose.

   UDP-Galactose + Glucose → Lactose + UDP

2. Glycolipids, glycosaminoglycans, glycoproteins and cerebrosides are some other important compounds, which depend on UDP-galactose as the donor of galactose.

Galactosaemia

Inborn errors due to deficiency of one of the enzymes of galactose metabolism cause ***galactosaemia.*** The deficient enzyme may be *galactokinase*, *galactose 1-phosphate uridyl transferase*, or rarely *epimerase*. Deficiency of the *uridyl transferase*, known as classical galactosaemia, is best known.

Classical galactosaemia: This is an autosomal recessive disease with an estimated population incidence of 1 in 40,000. Deficiency of the *uridyltransferase* leads to accumulation of galactose and galactose 1-phosphate after ingestion of milk or food containing galactose . There is elevated concentration of galactose in blood (*galactosaemia*) and

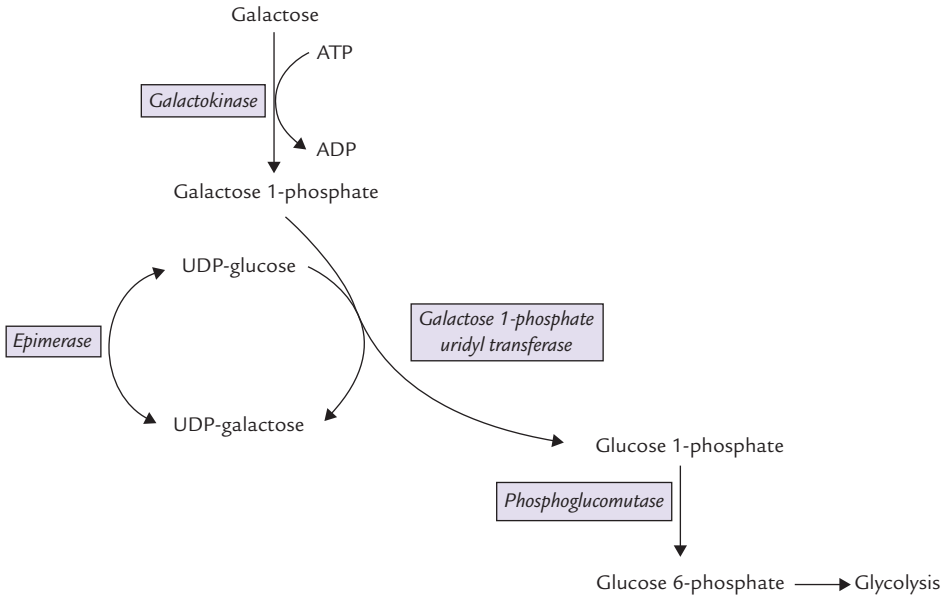


Fig. 9.11. Galactose entering glycolysis via the galactose-glucose interconversion pathway, a four-step reaction sequence.

urinary elimination (*galactosuria*). Other salient features are as follows:

1. Excess galactose is reduced to *galactitol* (dulcitol) by *aldose reductase*, the same enzyme that forms sorbitol from glucose in the polyol pathway. Galactitol, like sorbitol, accumulates in lens and causes cataract.
2. Accumulation of galactose 1-phosphate and concomitant depletion of inorganic phosphate causes *liver dysfunction*. It is evident within weeks after birth: feeding difficulties and vomiting result in poor weight gain, and jaundice is present that often is misdiagnosed as physiological jaundice of newborn. Excess galactose 1-phosphate inhibits *glycogen phosphorylase* and *phosphoglucomutase*; both impede glycogenolysis and glycogen accumulates in liver.
3. The central nervous system is affected as well, as evidenced by an *abnormal lethargy* or *irritability*. *Hypoglycaemia* is commonly observed because galactosaemia provides persistent stimulus for insulin secretion from the pancreas. Galactose in erythrocytes inhibits *glucose 6-phosphate dehydrogenase* and hence HMP-shunt, which reduces the supply of NADPH. This results in *haemolysis*, as discussed in Chapter 10.
4. In severe, untreated cases of galactosaemia, mental deficiency, liver cirrhosis, cataract, aminoaciduria and albuminuria develop.

Classical galactosaemia, caused by deficiency of enzyme *galactose 1-phosphate uridyl transferase*, results in accumulation of galactose and galactose 1-phosphate. Diversion of galactose in the galactitol is implicated in pathogenesis of cataract.

***Diagnosis*:** Presence of reducing material (galactose) in urine with a negative *glucose oxidase* suggests diagnosis. In addition, *galactose 1-phosphate uridyl transferase* activity can be measured in RBCs: the patients completely lack the enzyme, and the unaffected heterozygotes have a reduced enzyme activity.

Treatment: The patient is placed on a milk-free diet. Synthetic diets like soya milk, which are free of galactose, are recommended. The galactosaemic child, even when sustained on galactose-free diet, thrives well because galactose needed for the synthesis of biomolecules, can be obtained from UDP-glucose by epimerization reaction.

Galactokinase deficiency: Generally, *galactokinase*-deficient patients do not suffer from liver or renal complications, but develop *cataract* at a very early age (within 10 months after birth).

Metabolism of Mannose

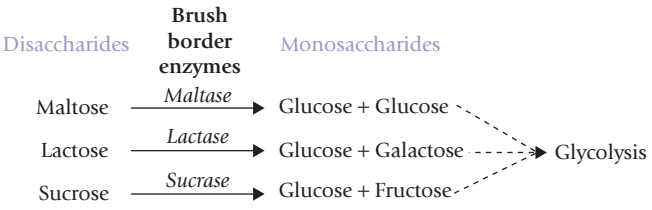
D-Mannose is a constituent of various polysaccharides and glycoproteins. It is phosphorylated at C-6 by *hexokinase* to form mannose 6-phosphate. The latter is isomerized by the enzyme *phosphomannose isomerase* to fructose 6-phosphate, a glycolytic intermediate.

B. Feeder Pathways for Disaccharides and Polysaccharides

The disaccharides are cleaved into the constituent monosaccharides by disaccharidases such as ***lactase, maltase,*** and ***sucrase.*** Lactose is cleaved by *lactase* into glucose and

galactose, sucrose by *sucrase* into glucose and fructose, and *maltase* cleaves maltose into two glucose molecules.

The *disaccharidases* are present in outer surface of the epithelial cell lining of the small intestine. The monosaccharides, generated by their action are carried to liver by portal blood. In liver, the monosaccharides are further metabolized through glycolytic sequence, as described earlier.



Similarly, the polysaccharides are first hydrolyzed into the constituent monosaccharides, which are then funneled into the central glycolytic sequence.

Monosaccharides, e.g. fructose, galactose, and mannose, as also the common disaccharides and polysaccharides, are enzymatically channeled into the pathway of glucose metabolism.

Lactose intolerance is a commonly encountered clinical condition, due to deficiency of the *intestinal lactase*. Lactose cannot be digested and is oxidized by bacteria in gut, producing gas, bloating and watery diarrhoea. Deficiency disorders of some other *disaccharidases* are also known to exist.

Metabolism of Amino Sugars

The amino sugars are required for the synthesis of glycolipids, glycoproteins, and proteoglycans. They are synthesized from fructose 6-phosphate. The amino group is derived from amide group of glutamine, the reaction being catalyzed by ***amido transferase.***

Glucosamine 6-phosphate forms its acetylated derivative (N-acetyl glucosamine 6-phosphate) by condensing with acetyl CoA. A ***mutase enzyme*** then converts it to N-acetyl glucosamine 1-phosphate, which reacts with UTP to produce the nucleotide activated form, UDP-N-acetyl glucosamine. The latter is epimerized to form UDP-N-acetyl galactosamine (Fig. 9.12).

Both the nucleotide activated forms (UDP-N-acetyl glucosamine and UDP-N-acetyl galactosamine) are then used for synthesis of complex molecules, e.g. glycoproteins, proteoglycans and glycolipids. UDP-N-acetyl glucosamine can enter an alternate pathway to form NANA.

IV. Tricarboxylic Acid Cycle

Tricarboxylic acid (TCA) cycle is also called **Krebs cycle** or the **citric acid cycle.** It is a cyclic pathway occurring in **mitochondria,** also referred to as the "central catabolic pathway". It occupies a central place in catabolism because catabolic pathways involving various biomolecules of diverse origin and chemical nature ultimately terminate in TCA cycle. The end product of these pathways is one of the intermediates of TCA cycle. For example, the glycogenic amino acids yield pyruvate, oxaloacetate, or α-ketoglutarate; the ketogenic amino acids yield acetyl CoA (which is the main precursor of TCA cycle), and the end product of degradation of various fatty acids and carbohydrates is also acetyl CoA (Fig. 9.13). Thus, TCA cycle is spoken of as the meeting point where catabolic pathways converge.

TCA cycle is the final common pathway of the oxidation of the acetyl CoA, formed from carbohydrates, fatty acids and amino acids.

TCA cycle serves **two main functions** in metabolism:

(a) to provide an important pool of metabolic intermediates for synthesis of many useful compounds, and

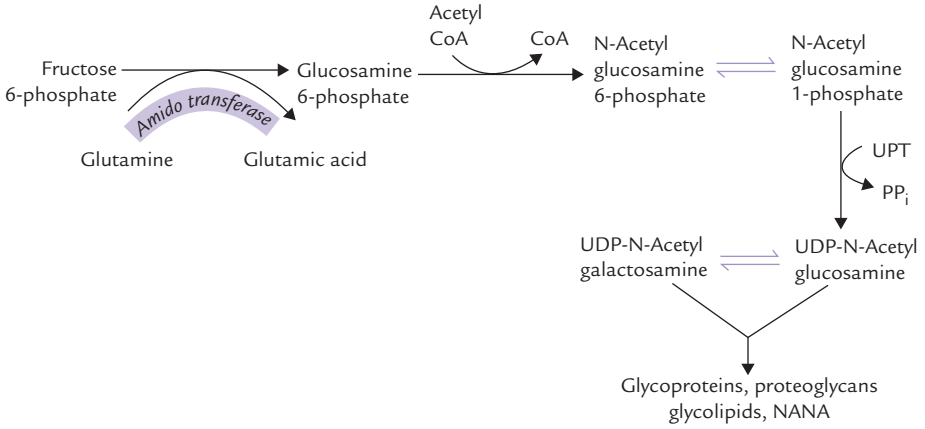


Fig. 9.12. Metabolism of amino sugars (UDP-N-Acetyl glucosamine and UDP-N-Acetyl galactosamine).

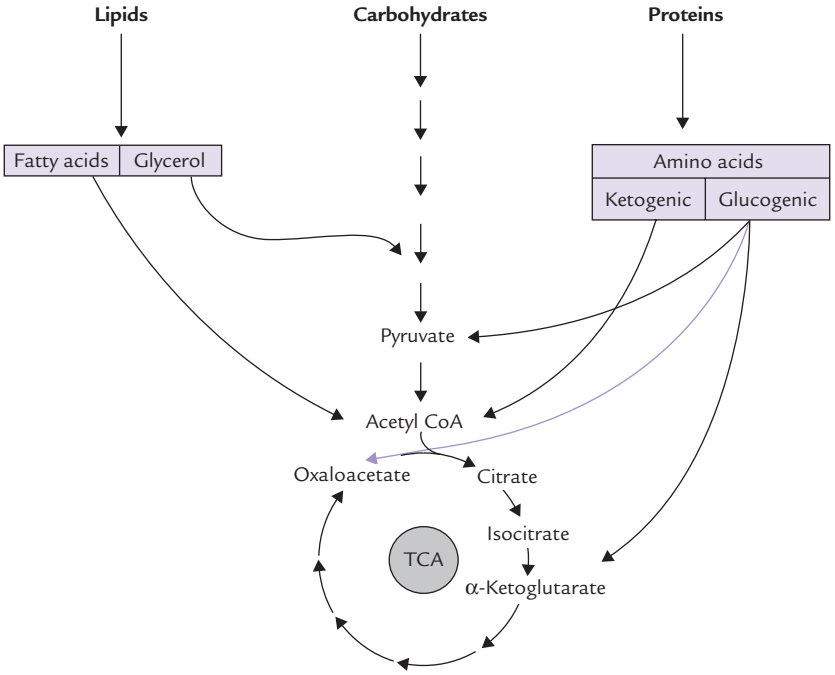


Fig. 9.13. TCA cycle as the meeting point of various catabolic pathways.

(b) to break down some compounds for generation of energy. The energy obtained is captured by reducing NAD^+ and FAD to NADH and $FADH_2$, respectively. To obtain full usage of the energy generated in this cycle, NADH and $FADH_2$, are processed by another pathway (oxidative phosphorylation) where their energy is converted to ATP.

A. *Pyruvate Dehydrogenase* Complex: A Bridge Between Glycolysis and TCA Cycle

Acetyl CoA, one of the two major reactants in the first reaction of TCA cycle (the other being oxaloacetate), is mainly generated from pyruvate. This conversion involves a series of complex reactions, catalyzed by a multienzyme complex, called ***pyruvate dehydrogenase* complex** *(PDH)*. Strictly speaking, this reaction sequence is not a part of the TCA cycle. It is, however, discussed here along with TCA cycle, because it serves as a bridge between glycolysis and TCA cycle.

The *pyruvate dehydrogenase* complex is located in the *mitochondrial matrix*. It is a multimolecular aggregate (MW 9×10^6) consisting of three enzymes and five coenzymes.

Enzymes: *Pyruvate decarboxylase*, *dihydrolipoyl transacetylase* and *dihydrolipoyl dehydrogenase* are the three enzymes present in the complex. There are about 60 molecules of *dihydrolipoyltransacetylase* and about 20–30 molecules each of the other two enzymes in each complex.

Coenzymes: Lipoic acid, FAD, NAD^+, coenzyme A and thiamine pyrophosphate (TPP) are the coenzymes in this complex. They serve as transient carriers of functional groups or participate in the oxidation–reduction reactions.

Aggregation of various enzymes in a complex greatly increase the efficiency of the transformation. This is because the intermediate products are not released to the medium but are channeled to the succeeding enzymes of the pathway. Thus a metabolic intermediate is available for the next reaction immediately after its production. Outline of the reaction sequence and role of various enzymes and coenzymes of this complex are summarized in Figure 9.14.

The pyruvate *dehydrogenase* multienzyme complex, which contains three enzymes and five coenzymes, generates acetyl CoA from the glycolytic product, pyruvate.

B. TCA Cycle as a Cyclic Pathway

TCA cycle consists of eight sequential reactions. It is begins with condensation of a four carbon oxaloacetate (OAA) molecule with an acetyl CoA molecule (2-carbon) to form a six carbon citrate molecule (Reaction 1; Fig. 9.15). In subsequent reactions of TCA cycle, two carbon atoms are lost from the citrate in the form of CO_2 (Reactions 3 and 4, Fig. 9.14). A series of modifications occur, in the remaining four carbon atoms, in successive steps, to ultimately form oxaloacetate (OAA). Thus, the last intermediate of one cycle (i.e. OAA) is ready for use

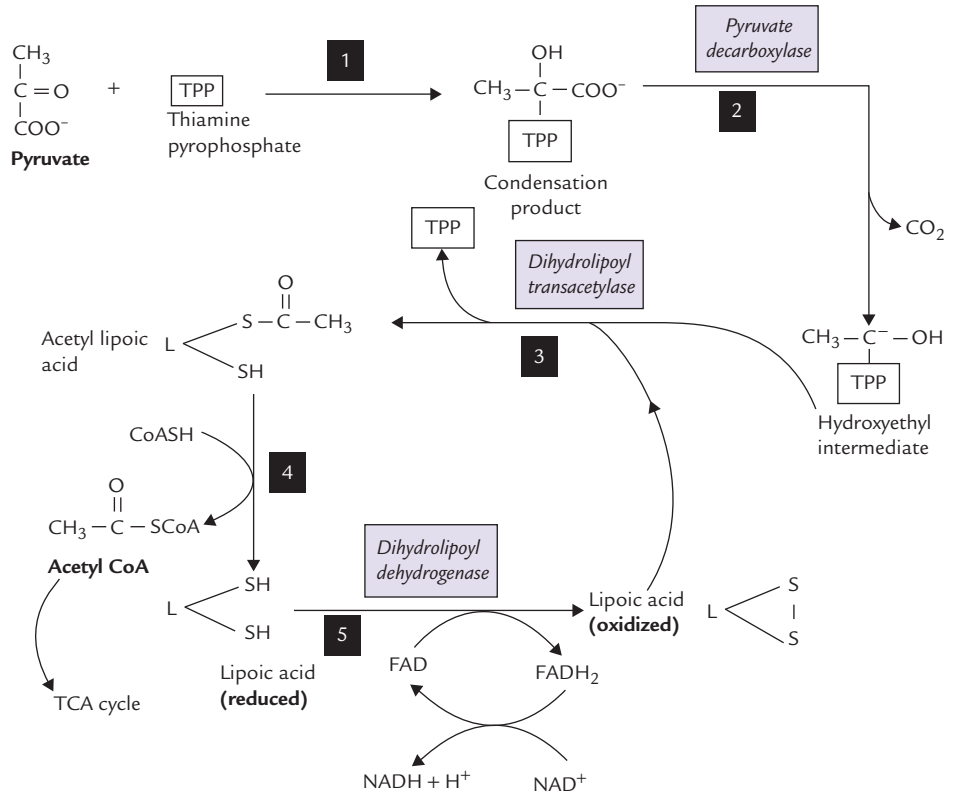


Fig. 9.14. Mechanism of conversion of pyruvate to acetyl CoA by *pyruvate dehydrogenase* complex. ***Step 1***: Pyruvate and thiamine pyrophosphate (TPP) combine to form a condensation product. ***Step 2***: *Pyruvate decarboxylase* catalyzes release of carbon dioxide from the condensation product to form a hydroxyethyl intermediate. The latter is attached to the reactive carbon of the TPP. ***Step 3***: Transfer of acetyl group from the hydroxyethyl intermediate to the lipoic acid (oxidized) occurs to form acetyl lipoic acid. The reaction is catalyzed by *dihydrolipoyl transacetylase*. ***Step 4***: Transfer of an acetyl group from the acetyl lipoic acid to coenzyme A (CoASH) occurs next, converting the latter to acetyl CoA. The acetyl CoA then enters the TCA cycle. The other product of this reaction is reduced lipoic acid. ***Step 5***: This step regenerates the oxidized lipoic acid (from the reduced lipoic acid), which then participates in the next cycle of reactions. This conversion is catalyzed by the *dihydrolipoyl dehydrogenase* component of the enzyme complex, which catalyzes transfer of the reducing equivalents first to FAD and then to NAD^+. **Overall:** Pyruvate + NAD^+ + CoASH → Acetyl CoA + CO_2 + NADH + H^+.

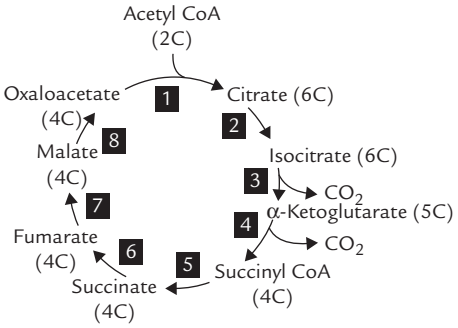


Fig. 9.15. TCA cycle: Oxaloacetate (OAA), the last intermediate is also a reactant in the first step. The eight reactions of the pathway are numbered 1 to 8.

as a substrate in the next cycle. In this way, *there is no net generation of OAA, or of any of the cycle intermediates.*

The reactions of TCA cycle take place in the **mitochondrial matrix**. This is in close proximity to the reactions of oxidative phosphorylation, which occur in the inner mitochondrial membrane (Chapter 14). This arrangement ensures that the free energy liberated during reactions of the TCA cycle is promptly trapped in the form of ATP.

C. Reactions of TCA Cycle

The reactions brought about by different enzymes of the TCA cycle are depicted in Figure 9.16. All the enzymes are present in **mitochondrial matrix** except *succinate dehydrogenase* which is located in the inner mitochondrial membrane.

Reaction 1: Synthesis of Citrate from Acetyl CoA and Oxaloacetate

Citrate is produced by condensation of acetyl CoA with oxaloacetate in an irreverssible reaction catalyzed by the enzyme, *citrate synthase*. The reaction equilibrium lies far towards the right (i.e. towards the formation of citrate).

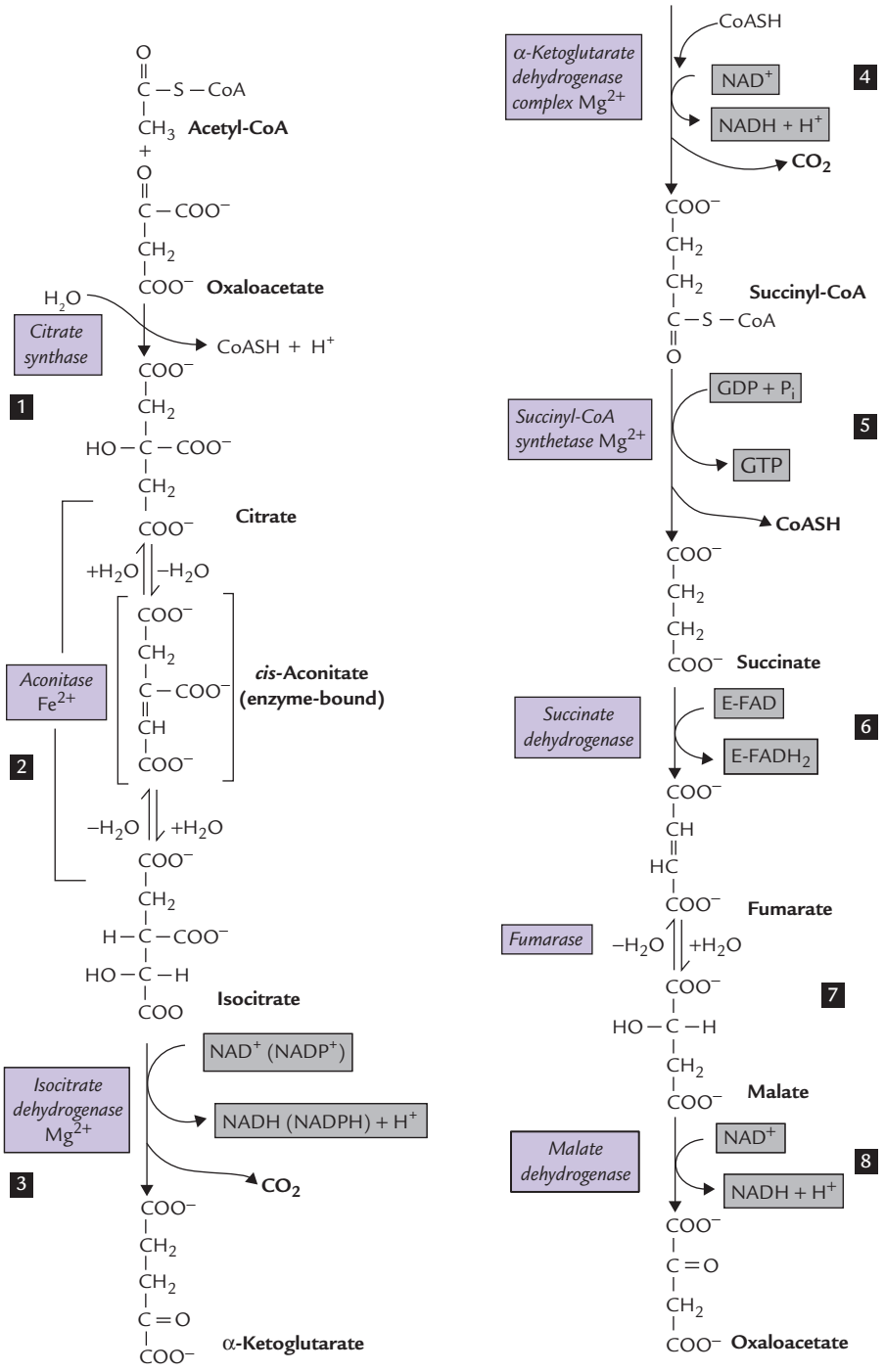


Fig. 9.16. Reactions of tricarboxylic acid (Krebs) cycle.

TCA cycle derives its other name (citric acid cycle) from this first intermediate, citrate. However, citrate may leave the citric acid cycle to participate in other metabolic pathways as well (Fig. 9.17). Excessive citrate crosses the inner mitochondrial membrane through specific tricarboxylate carriers and reaches the cytosol, where it provides acetyl CoA:

$$\text{Citrate} \xrightarrow{\textit{Citrate lyase}} \text{Acetyl CoA} + \text{OAA}$$

In cytosol, acetyl CoA serves as a precursor for fatty acid synthesis (lipogenesis). Moreover, citrate stimulates *acetyl CoA carboxylase*, the rate-limiting enzyme of fatty acid synthesis. Thus citrate enhances fatty acid synthesis by:

- Providing the substrate (i.e. acetyl CoA).
- Stimulating the key lipogenic enzyme (i.e. *acetyl CoA carboxylase*).

In the cytosol, citrate has another important regulatory role to play. *It adjusts the rate of glycolysis to that of TCA cycle.* It does so by inhibiting *phosphofructokinase* (PFK_1) the rate-limiting enzyme of glycolysis (Table 9.3). Inhibition of this enzyme decreases the rate of glycolysis,

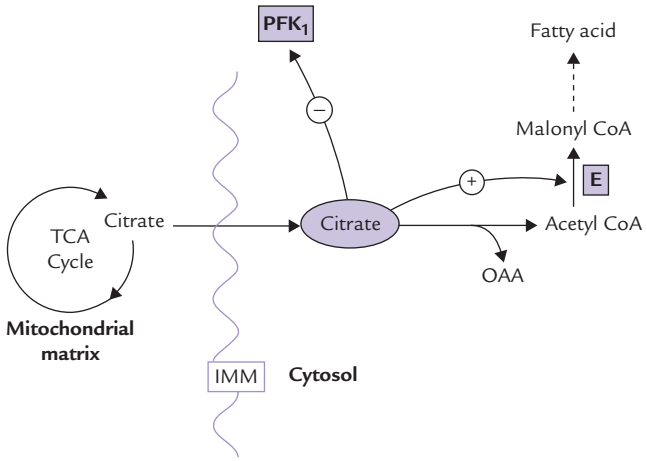


Fig. 9.17. Citrate, a TCA intermediate, plays an important regulatory role in lipogenesis and glycolysis by modulation of activities of the enzymes, *phosphofructokinase* (PFK_1) and *acetyl CoA carboxylase* (E) (IMM = inner mitochondrial membrane).

and therefore, the production of acetyl CoA falls. This results in decreased rate of TCA cycle since acetyl CoA is needed in the first reaction of TCA cycle. In this way, rates of glycolysis and TCA keep pace with each other.

Thus, when citrate concentration is high, implying that the cell is adequately supplied with fuel molecules and, therefore, further production of energy is not required, the energy-yielding catabolic pathways (e.g. glycolysis and TCA cycle) are inhibited. The biosynthetic pathway (i.e. fatty acid synthesis) is favoured at the same time. This illustrates a fundamental principle of biochemistry, i.e. the catabolic and the anabolic pathways are regulated reciprocally. If one pathway is favoured, the other is inhibited.

Reaction 2: Isomerization of Citrate

Isomerization of citrate by the enzyme *aconitase* yields isocitrate. During this reaction, a transient enzyme-bound intermediate, *cis*-aconitate is formed.

Reaction 3: Oxidative Decarboxylation of Isocitrate

The enzyme *isocitrate dehydrogenase* (IDH) catalyzes removal of two hydrogen atoms from isocitrate (i.e. oxidation) with a concomitant release of a CO_2 molecule (i.e. decarboxylation). α-Ketoglutarate is the reaction product.

NAD^+ serves as a coenzyme in this step, and is converted to NADH by accepting a pair of hydrogen atoms. $NADP^+$ can also serve as a coenzyme in this reaction. *Isocitrate dehydrogenase* is one of the few enzymes that are capable of using both NAD^+ and $NADP^+$ as coenzymes.

Reaction 4: Oxidative Decarboxylation of α-Ketoglutarate

Like the previous step, this one also involves oxidation and decarboxylation. The reaction is catalyzed by the enzyme *α-ketoglutarate dehydrogenase*, to produce succinyl CoA.

The mechanism of this conversion is similar to that of the conversion of pyruvate to acetyl CoA (Fig. 9.14). The same set of coenzymes, i.e. thiamine pyrophosphate, lipoic acid, FAD, NAD^+, and CoASH, is used. Each of these performs a function analogous to that performed in the *pyruvate dehydrogenase* complex.

The reaction releases the second CO_2 molecule of the cycle and produces the second NADH. The equilibrium of the reaction lies towards the right, i.e. towards succinyl CoA formation.

Reaction 5: Cleavage of Succinyl CoA

The enzyme *succinyl CoA synthetase* (also called *succinate thiokinase*) cleaves the high-energy thioester bond in succinyl CoA ($\Delta G^{0'}$ is –8.0 kcal/mole) to release large amount of free energy, which is used to produce a GTP molecule.

This reaction provides an example of **substrate level phosphorylation,** since the production of a high-energy phosphate (e.g. GTP) is coupled with enzymatic transformation of a substrate molecule. GTP can produce ATP by action of the enzyme *nucleoside diphosphokinase*.

$$GTP + ADP \rightarrow GDP + ATP$$

Note: Succinyl CoA may also serve as a substrate for haem synthesis.

Reaction 6: Oxidation of Succinate

A pair of reducing equivalents is removed from succinate by the enzyme *succinate dehydrogenase* to form fumarate. FAD serves as a coenzyme in this reaction. Unlike the other enzymes of the TCA cycle, which are located in the mitochondrial matrix, *succinate dehydrogenase is anchored in the inner mitochondrial membrane*. It catalyzes removal of a hydrogen pair from succinate and transfers it to FAD which becomes $FADH_2$. This in turn transfers its electrons to ubiquinone, which becomes ubiquinol, and is then transferred to complex III for oxidation (Chapter 14).

Reaction 7: Hydration of Fumarate

Addition of a water molecule to fumarate forms malate. It is a reversible reaction, catalyzed by the enzyme *fumarase*.

Note: In spite of being reversible, this reaction always proceeds unidirectionally, i.e. towards formation of malate. This is because of the thermodynamic pull in TCA cycle (discussed later).

Reaction 8: Oxidation of Malate

Removal of a pair of reducing equivalents from malate by the enzyme *malate dehydrogenase* produces oxaloacetate. The reaction generates the third NADH molecule of the cycle. Malate to oxaloacetate conversion is the last reaction of the cycle. Oxaloacetate, generated in this step,

BOX 9.5

Inhibitors of TCA Cycle

TCA cycle is inhibited by a number of chemicals, which include the following:

- *Fluoroacetate:* It inhibits the enzyme *aconitase* (non-competitively).
- *Arsenite:* It causes non-competitive inhibition of *α-ketoglutarate dehydrogenase.*
- *Malonate*: It causes competitive inhibition of *succinate dehydrogenase.*

condenses with another molecule of acetyl CoA to initiate another cycle.

The following equation summarizes all the reactions of the TCA cycle.

$$\text{Acetyl CoA} + 3NAD^+ + FAD + GDP + P_i + 2H_2O \rightarrow 2CO_2 + 3NADH + 2H^+ + FADH_2 + GTP + \text{CoA-SH}$$

Several chemicals can inhibit reactions of the TCA cycle, as outlined in Box 9.5.

A four-carbon compound oxaloacetate reacts with acetyl CoA to form six-carbon citrate, which is then converted back to oxaloacetate in the remaining reactions of the TCA cycle.

The following points about TCA cycle are noteworthy:

1. Two carbon atoms enter the cycle as acetyl CoA (and condense with oxaloacetate) and two carbons leave in the form of two molecules of CO_2. Thus, TCA basically involves oxidation of acetyl CoA to carbon dioxide and, as such there is *no net consumption or regeneration of oxaloacetate or any of the other cycle intermediates*.
2. Some of the reactions of TCA cycle are reversible and yet they always proceed unidirectionally. This is because equilibrium of some of the reactions of TCA cycle (e.g. Reactions 1, 3 and 4, Fig. 9.12) lies far towards the right. These reactions are irreversible and have strong tendency to proceed unidirectionally, towards the right. This generates a strong "*thermodynamic pull*" so that rest of the (reversible) reactions are also pulled in that direction. As mentioned earlier, fumarate to malate conversion which is freely reversible otherwise, always proceeds in direction of malate formation because of the thermodynamic pull.

D. Energy Yield from TCA Cycle

There are four dehydrogenation reactions in TCA cycle (Reactions 3, 4, 6 and 8), which generate three NADH and one $FADH_2$ molecules. These reduced coenzymes donate electrons to the electron transport chain and generate ATPs by oxidative phosphorylation. Oxidation of each

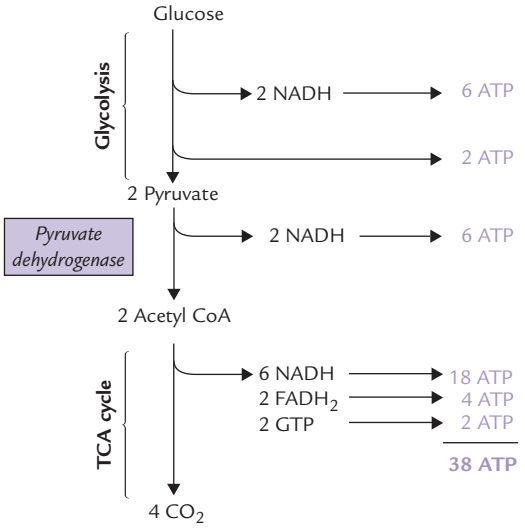


Fig. 9.18. Energetics of glucose oxidation. 38 ATPs are generated by complete oxidation of glucose via glycolysis, PDH reaction, and TCA cycle.

NADH in this manner generates three ATPs, whereas two ATPs are produced from one $FADH_2$ molecule. In addition, one GTP is produced during enzymatic transformation of succinyl CoA (Reaction 5), which generates an ATP through action of *nucleoside diphosphokinase*. Hence, each cycle of TCA cycle, produces **12 ATP** molecules.

3 NADH (Reaction 3,4 and 8)	→	9 ATP
1 $FADH_2$ (Reaction 6)	→	2 ATP
1 GTP (Reaction 5)	→	1 ATP
TOTAL (Each cycle)	=	12 ATP

However, two acetyl CoA molecules are generated from each glucose, therefore, this cycle occurs twice, generating 6NADH, $2FADH_2$ and 2GTPs (Fig. 9.18) Hence , total 24 ATPs are generated. Thus, complete oxidation of glucose via glycolysis, *pyruvate dehydrogenase*, the Krebs cycle and the oxidative phosphorylation pathway yields **38 ATPs**. Since two ATPs were initially used during stage I reactions of glycolysis, the **net yield is 36 ATPs**. Compared with anaerobic pathway, the oxidative pathway thus yields 18 times more energy in the form of ATP. However, this is still far less than the energy obtained by burning of a molecule of glucose in calorimeter (2780 kJ), which is sufficient to generate about

91 ATPs (one ATP has energy bond equivalent to 30.5 kJ). Evidently, *only about 40% of the energy locked in chemical bonds of glucose molecule is captured for ATP synthesis in our body.*

Pasteur Effect

It has been observed that under anaerobic condition a tissue or microorganism utilizes much more glucose than it does under aerobic conditions. This is the ***Pasteur effect***, and it reflects inhibition of glycolysis by oxygen. It is also observed that the levels of glycolytic intermediates from fructose 1,6-bisphosphate onwards decrease while the earlier intermediates increase. Evidently, the Pasteur effect is because of inhibition of the enzyme *phosphofructokinase*. The inhibitory effect is caused by citrate and ATP, the compounds produced in presence of oxygen.

Crabtree Effect

When oxygen supply is kept constant and glucose concentration is increased, the oxygen consumption by cells falls. This is called *Crabtree effect*, which is basically *opposite of Pasteur effect*. It is seen in the cells that have a high rate of aerobic glycolysis. In such cells the glycolytic sequence consumes much of the available P_i and NAD^+, which limits their availability for oxidative phosphorylation. As a result, rate of oxidative phosphorylation decreases, and oxygen consumption also shows a corresponding fall.

E. Synthetic Functions of the TCA Cycle

TCA is not only the final oxidative pathway for various biomolecules, its intermediates are also substrates for various biosynthetic pathways (Fig. 9.19). Thus, TCA is both catabolic and anabolic in nature, and hence regarded as **amphibolic.**

Some important synthetic pathways connected with TCA are as follows:

Synthesis of Glucose

Intermediates of TCA cycle can serve as substrates for **gluconeogenesis**, the pathway that produces glucose from non-carbohydrate sources. As glucose is synthesized in this manner some intermediates of TCA cycle are removed from the cycle. These intermediates may be subsequently replenished by anaplerotic reactions.

Synthesis of Fatty Acids

Citrate is used for fatty acid synthesis (Fig. 9.17).

Synthesis of Amino Acids

Oxaloacetate and α-ketoglutarate can form aspartate and glutamate, respectively by transamination reactions,

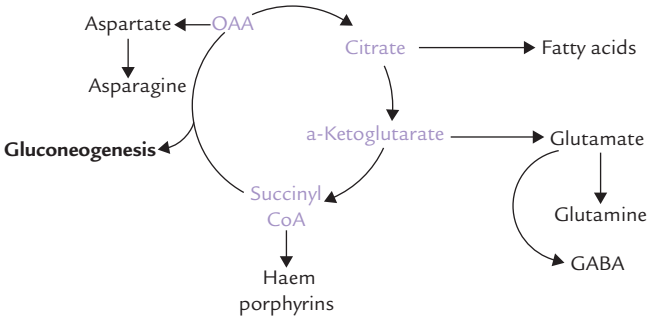


Fig. 9.19. Synthetic functions of TCA cycle.

which in turn are required for the synthesis of other non-essential amino acids (Chapter 13), as also of purines and pyrimidines.

Others

Succinyl CoA can be diverted from TCA cycle for biosynthesis of porphyrins and haem (for further information refer to Box 9.6). The α-ketoglutarate can form glutamate, from which γ-aminobutyric acid (GABA), an inhibitory neuro-transmitter is produced.

Some TCA intermediates are substrates for gluconeogenesis, amino acid synthesis, and fatty acid synthesis. Anaplerotic reactions such as the *pyruvate carboxylase* reactions replenish TCA cycle intermediates.

F. Regulation of TCA Cycle

A number of modulators regulate activities of various enzymes of TCA cycle. Two types of regulatory mechanisms—allosteric modulation and covalent modulation of the enzyme activities operate. Regulation of the *pyruvate dehydrogenase* complex, the bridge between glycolysis and TCA cycle, and the regulation of the cyclic pathway are discussed below.

Regulation of the *Pyruvate Dehydrogenase Complex*

Activity of the *pyruvate dehydrogenase complex* (PDH) is switched-on or switched-off based on cellular energy needs. When the cellular energy charge is high and surplus fuel molecules are present intracellularly, inhibition of activity of the enzyme complex occurs. Two mechanisms of regulation have been recognized.

Allosteric regulation: The enzyme complex is subject to allosteric inhibition by the reaction products: acetyl CoA and NADH. It is stimulated by AMP.

Elevated intracellular concentrations of both acetyl CoA and NADH indicate surplus intracellular fuel molecules

BOX 9.6

Anaplerotic ('Filling-up') Reactions

The TCA cycle intermediates may be "*siphoned off*" for various biosynthetic processes, which may deplete the important pool of these metabolites. For example, removal of succinyl CoA for haem synthesis could gradually deplete intramitochondrial concentration of succinyl CoA and hence, that of the other cycle intermediates. If the cycle intermediates were not replenished, TCA cycle would cease to function. Anaplerotic reactions provide the TCA cycle with metabolic intermediates, maintaining the activity of the cycle and ensuring that the cycle is never stalled because of lack of these metabolites. *Pyruvate carboxylase* (PC) reaction is the classical example of the anaplerotic reaction. PC converts pyruvate to oxaloacetate, which is required for the initiation of the cycle. Some other important anaplerotic reactions are as below:

- Conversion of pyruvate to malate by the cytoplasmic *malic-enzyme*. Malate can then enter the mitochondrion as a substrate for the TCA cycle.
- Pyruvate may react with aspartate or glutamate in *transaminase* reactions, producing the TCA cycle intermediates oxaloacetate and α-ketoglutarate, respectively.
- Several glycogenic amino acids may serve as source of TCA cycle intermediates (Chapter 13).

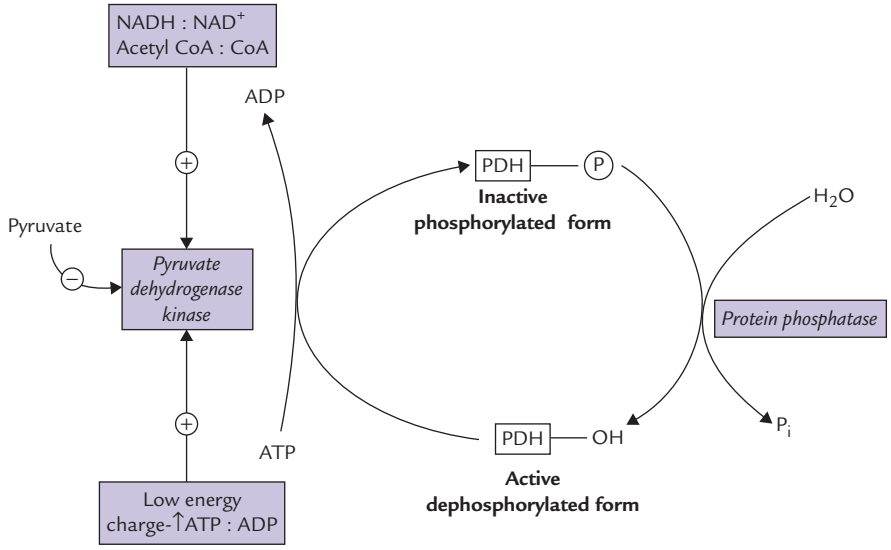
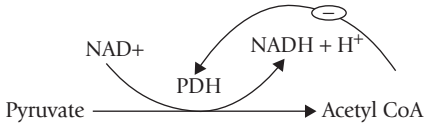


Fig. 9.20. Regulation of *pyruvate dehydrogenase* complex (PDH).

and excessive cellular energy charge. Inhibition of the PDH, under these circumstances, stops further production of energy and fuel molecules.



Covalent modulation: The PDH exists in two forms:

- **Phosphorylated form**, which is inactive.
- **Dephosphorylated form**, which is active.

Thus, covalent attachment of a phosphate group renders the PDH complex inactive, whereas removal of this group results in its activation. The following enzymes are involved in interconversion of the phosphorylated-dephosphorylated forms.

1. *Pyruvate dehydrogenase kinase:* This enzyme phosphorylates the enzyme complex, thereby inactivating it. Increased activity of the *kinase* is, therefore, associated with decreased activity of PDH. Conversely, inhibition of the *kinase* activates the enzyme complex (Fig. 9.20).

Pyruvate dehydrogenase kinase is allosterically activated by high energy charge, high [NADH] : [NAD^+] ratio, and high [acetyl CoA] : [CoA] ratio and is inhibited by pyruvate. Conversely the enzyme activity is inhibited by low energy charge, reflected by elevated ratio of ADP : ATP, which in turn results in activation of the PDH. The elevated ADP : ATP ratio signals increased demand for energy production, which is met through activation of the PDH.

2. *Protein phosphatase:* This enzyme removes a phosphate group from the phosphorylated form of the PDH complex to produce the active dephosphorylated form (Fig. 9.20). Thus, it activates the PDH.

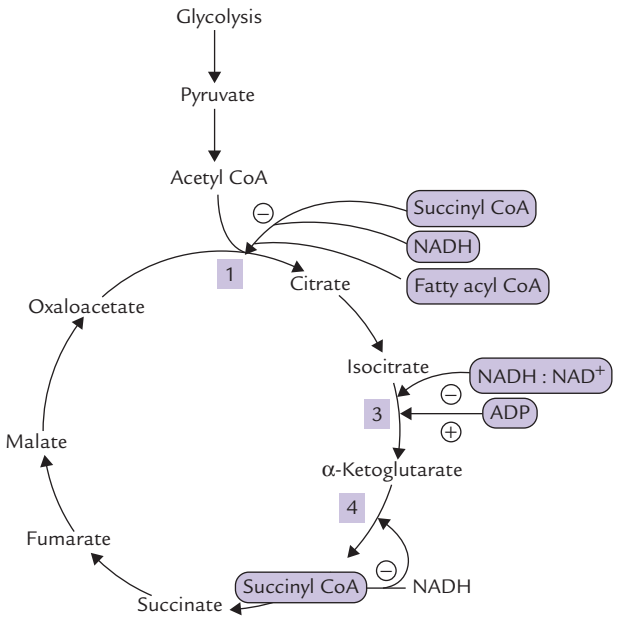


Fig. 9.21. Regulation of TCA cycle (⊖ = inhibits reaction, ⊕ = stimulates reaction, numbers 1, 3 and 4 are steps at which regulation occurs).

Entry of acetyl CoA into the TCA cycle is regulated at the *pyruvate dehydrogenase* step by product inhibition (by NADH and acetyl CoA) and by covalent modification.

Regulation of the Cyclic Pathway (Fig. 9.21)

It occurs at the following steps:

1. Synthesis of citrate (Reaction 1).
2. Oxidative decarboxylation of isocitrate (Reaction 3).
3. Oxidative decarboxylation of α-ketoglutarate (Reaction 4).

Regulation of Citrate Synthesis

This is the most important regulatory step of the pathway. As in case of most metabolic pathways, where an initial step is the rate setting step for the pathway as a whole, the first step of TCA cycle is also most important in regulation. Rate of this reaction is determined by the following factors:

1. *Concentration of the substrate molecules* (i.e. oxaloacetate and acetyl CoA): Concentration of oxaloacetate is the most critical factor in determining the reaction rate. Concentration of acetyl CoA is also an important determinant for pushing the reaction forward.
2. *Allosteric modulation of the enzyme:* Activity of *citrate synthase*, the enzyme catalyzing this step, is mainly regulated by succinyl CoA, which inhibits the enzyme activity by decreasing its affinity for acetyl CoA. Fatty acyl CoA and NADH also act as negative allosteric modulators of the enzyme. Raised levels of these metabolites indicate that the cell has adequate supply of fuel and energy. Under these circumstances, TCA cycle, which is an energy-yielding pathway, is inhibited so that there is no further production of energy.

Regulation of Oxidative Decarboxylation of Isocitrate

The enzyme catalyzing this step, *isocitrate dehydrogenase*, is activated by ADP and inhibited by high energy charge and high [NADH] : [NAD^+] ratio. Elevated levels of mitochondrial ADP indicate low energy state of the cell and signals a need for generation of more high-energy phosphate molecules (i.e. ATP). Stimulation by ADP accelerates the cycle, generating the required ATP.

Regulation of Oxidative Decarboxylation of α-ketoglutarate

The enzyme catalyzing this step, *α-ketoglutarate dehydrogenase*, is allosterically inhibited by its own products (succinyl CoA and NADH) and by high energy charge.

It is evident that low energy charge stimulates TCA so as to generate more energy; conversely, high energy charge inhibits it. The other energy-yielding pathways discussed so far, i.e. glycolysis and oxidative decarboxylation of pyruvate are also regulated similarly.

The TCA cycle is regulated at the steps catalyzed by *citrate synthase, isocitrate dehydrogenase* and *α-ketoglutarate dehydrogenase*. Regulation is accomplished mainly by substrate availability, allosteric modulation, and feedback inhibition.

Under normal conditions, the rates of glycolysis and TCA cycle are integrated with each other so that only that much glucose is metabolized as required to supply the initial substrate, i.e. the acetyl CoA for the TCA cycle. Rates of these two pathways are matched because of inhibition of *phosphofructokinase* (PFK_1), the key regulatory enzyme of glycolysis, by a TCA cycle intermediate, citrate. ATP and NADH, produced during TCA cycle also inhibit glycolysis by acting as negative modulators of this enzyme.

G. Glyoxylate Cycle

Synthesis of carbohydrate from fat is not possible in the mammalian cell due to irreversibility of the *pyruvate dehydrogenase* reaction and lack of alternate pathways. However, plant cells and some microorganisms can accomplish this conversion because of a cyclic pathway, termed glyoxylate cycle (Fig. 9.22). It shares several reactions with

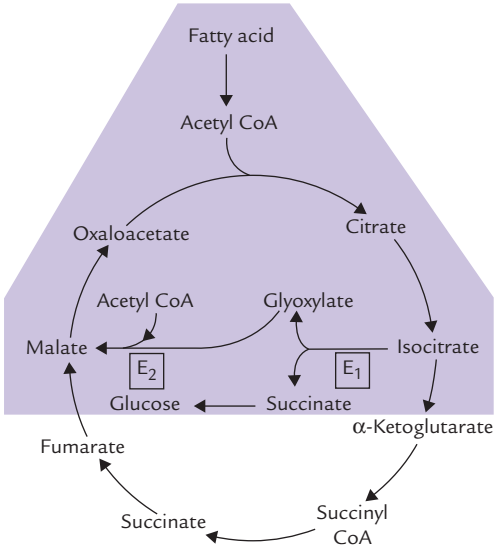


Fig. 9.22. Reactions of glyoxylate cycle (shaded area) (E_1 = *isocitrate lyase*, E_2 = *malate synthase*).

TCA cycle and is considered an **anabolic variant** of the latter. It is useful in germinating seeds where the stored fat is converted to glucose for meeting the energy needs of the cell. It operates in **glyoxysomes,** the specialized cellular organelles where fatty acid oxidation also occurs.

Reactions

The initial few reactions up to isocitrate formation, are common with TCA cycle, but the isocitrate bypasses the TCA cycle and is cleaved by the enzyme *isocitrate lyase* to succinate and glyoxylate. The latter compound forms malate (by condensing with another molecule of acetyl CoA), which is an intermediate of the TCA cycle. Net result of these reactions is conversion of two 2-carbon fragments of acetyl CoA to malate, which is a substrate for gluconeogenesis, and can also enter TCA cycle.

In addition to glyoxylate, the other product of cleavage of isocitrate is succinate, which is converted to glucose via reactions of gluconeogenesis, thereby ensuring net synthesis of carbohydrate from fat.

The glyoxylate pathway, which operates only in plants, is a variation of TCA cycle. It permits net synthesis of glucose from acetyl CoA.

V. Gluconeogenesis

Synthesis of glucose from compounds other than carbohydrates is called ***gluconeogenesis*** (i.e. formation of new sugar). A variety of biomolecules such as lactate, pyruvate, glycerol (derived from triacylglycerols) and α-keto acids (derived from amino acid catabolism) are substrates for gluconeogenesis. The pathway is important because some organs, like the **brain** and the *erythrocytes, depend exclusively on glucose* for their energy needs. The brain is said to be a voracious eater of glucose, consuming about 120 grams per day (out of about 160 grams needed by the entire body) and requires a blood glucose concentration of 70–100 mg/dL. Though degradation of stored glycogen can also provide glucose, gluconeogenesis is the only source of glucose during prolonged fasting and starvation.

Gluconeogenesis occurs **mainly in the liver,** and to a **lesser extent in renal cortex.** On weight basis, the gluconeogenic capacity of the renal cortex equals that of the liver, but because of the size difference between these tissues, the gluconeogenic activity in the kidney amounts to only 10% of that in the liver. However, in certain abnormal metabolic states, such as prolonged starvation, kidneys become the major glucose-producing organ.

The gluconeogenic pathway is located mainly in **cytosol**, although some precursors are produced in the mitochondria.

A. Gluconeogenesis: Not a Reversal of Glycolysis

Though most of the reactions of the gluconeogenic sequence (7 out of 10) are reversal of those of glycolysis the following three reactions are unique to gluconeogenesis, and are catalyzed by a different set of enzymes. They are called the bypass steps:

First bypass Conversion of pyruvate to phosphoenolpyruvate (PEP).

Second bypass Conversion of fructose 1, 6-bisphosphate to fructose 6-phosphate.

Third bypass Conversion of glucose 6-phosphate to glucose.

The corresponding steps in glycolysis are formation of pyruvate from phosphoenol pyruvate (for first bypass); fructose 1,6-bisphosphate from fructose 6-phosphate (for second bypass) and glucose 6-phosphate from glucose (for third bypass). Recall that only these three reactions in glycolysis are irreversible. Therefore, an alternate set of reactions, unique to gluconeogenesis, is needed to circumvent these three irreversible glycolytic reactions (Fig. 9.23). Apart from these three steps, rest of the seven steps of the glycolytic sequence are reversible and are operative in the gluconeogenesis as well.

Three reactions are unique to gluconeogenesis, which correspond to the three irreversible steps of glycolysis. Rest of reactions are reversible and common to both pathways.

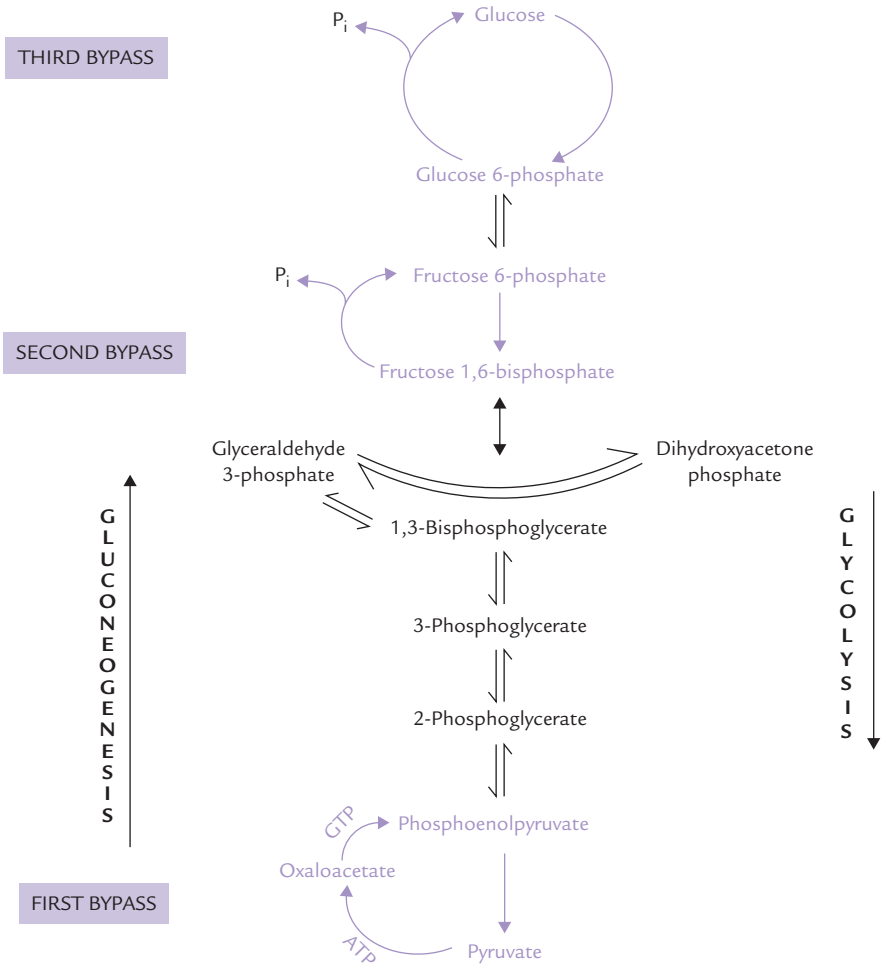


Fig. 9.23. The opposing pathways of glycolysis and gluconeogenesis. Steps made in colour are unique to gluconeogenesis.

B. The Bypass Reactions of Gluconeogenesis

First Bypass: Conversion of Pyruvate to Phosphoenolpyruvate (PEP)

During conversion of phosphoenolpyruvate to pyruvate in glycolytic sequence, release of a large amount of free energy ($\Delta G^{0'}$ –14.8 kcal/mole) occurs under standard conditions. For the reverse reaction (i.e. conversion of pyruvate to PEP during gluconeogenesis), input of an equivalent amount of free energy is required. This is a thermodynamically unfavourable situation, which is circumvented by making this conversion occur in two distinct reactions:

- Conversion of pyruvate to oxaloacetate: 1 ATP hydrolyzed
- Conversion of oxaloacetate to PEP: 1 GTP hydrolyzed

Sum: Pyruvate to PEP: 1 ATP + 1 GTP are hydrolyzed.

The conversion involves the following **four steps**, which occurs partly in cytosol and partly in mitochondrial matrix (Fig. 9.24):

1. *Transport of pyruvate into mitochondrial matrix:* Pyruvate moves across the inner mitochondrial membrane (IMM) to enter the mitochondrial matrix through mediation of specific transport proteins, called the monocarboxylate carriers monocarboxylate carrier **monocarboxylate carriers (MC).**
2. *Pyruvate to oxaloacetate conversion:* Pyruvate, a 3-carbon molecule is converted to 4-carbon oxaloacetate by carboxylation reaction catalyzed by *pyruvate carboxylase* (E_1), a biotin-dependent mitochondrial enzyme.

$$\text{Pyruvate} + CO_2 + H_2O + \text{ATP} \longrightarrow \text{Oxaloacetate} + \text{ADP} + P_i$$

3. *Transport of oxaloacetate into cytosol:* Rest of the enzymes of gluconeogenesis are located in the cytosol. Therefore, the mitochondrial oxaloacetate, produced in the step 2, must be transported out of the mitochondria into cytosol. However, oxaloacetate cannot directly cross the IMM because the latter is impermeable to most molecules except for the ones for which specific carrier proteins exist (such as monocarboxylate carrier for pyruvate, dicarboxylate carrier for malate, and tricarboxylate carrier for citrate). Therefore, the oxaloacetate

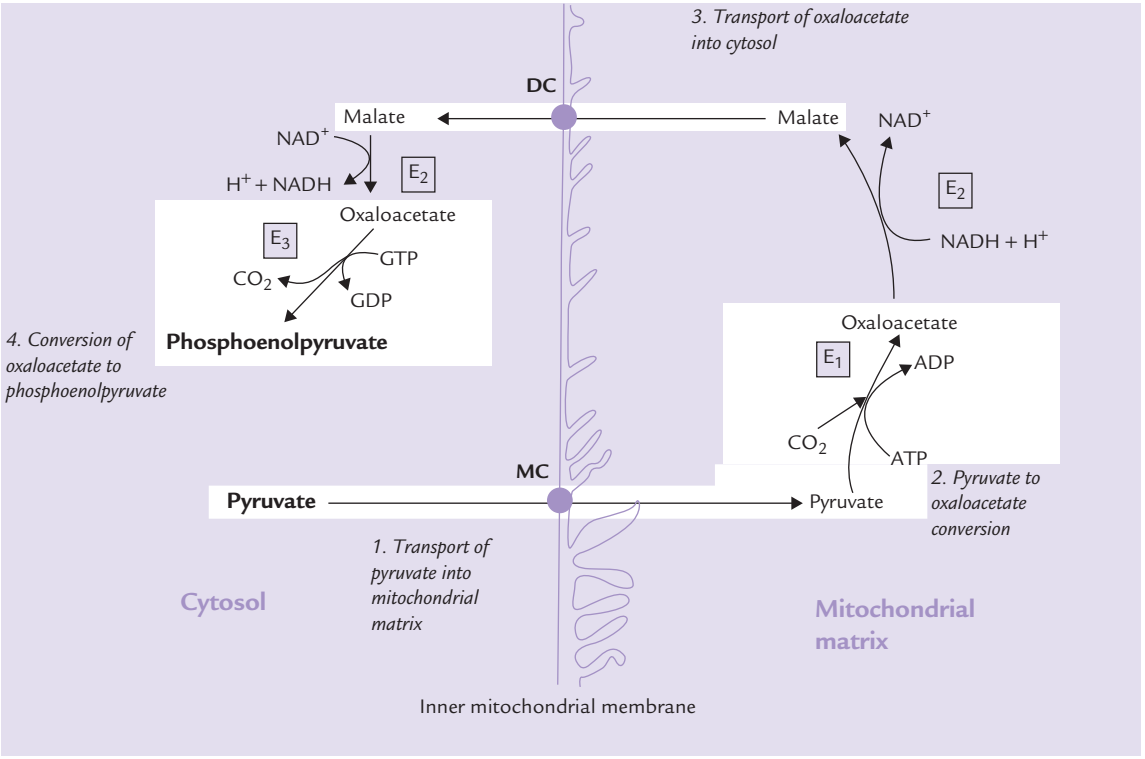


Fig. 9.24. Synthesis of phosphoenolpyruvate from pyruvate. The conversion occurs partly in cytosol, partly in mitochondria and in four steps (MC = monocarboxylate carrier, DC = dicarboxylate carrier, E_1 = *pyruvate carboxylase*, E_2 = *malate dehydrogenase*, E_3 = *phosphoenolpyruvate carboxy kinase*).

is converted to malate, a molecule which is capable of crossing the IMM. The enzyme *malate dehydrogenase* (E_2) catalyzes this reaction.

$$\text{Oxaloacetate} + \text{NADH} + \text{H}^+ \rightarrow \text{Malate} + \text{NAD}^+$$

Malate is carried out of the mitochondria into cytosol by the **dicarboxylate carrier (DC)**. In cytosol, oxaloacetate is recovered from malate by reversal of the above reaction.

4. *Conversion of oxaloacetate to phosphoenolpyruvate:* In cytosol, the oxaloacetate undergoes decarboxylation and phosphorylation to yield phosphoenolpyruvate (PEP).

$$\text{Oxaloacetate} + \text{GTP} \rightarrow \text{PEP} + \text{CO}_2 + \text{GDP}$$

This reaction is catalyzed by the enzyme *phosphoenolpyruvate carboxykinase (PEPCK)*, and is driven forward by hydrolysis of GTP.

Thus, a concerted action of several enzymes and carrier proteins permits conversion of PEP to pyruvate; a conversion which is thermodynamically unfavourable otherwise. Energy of two high-energy phosphate bonds (equivalent to 14.6 kcal/mole), one each from ATP and GTP, are required to drive the conversion across the energy "**road-block**". The overall equation for this set of reactions is as below:

$$\text{Pyruvate} + \text{ATP} + \text{GTP} \longrightarrow \text{PEP} + \text{ADP} + \text{GDP} + \text{P}_i$$
$$\Delta G^{0\prime} = +0.2 \text{ kcal/mole}$$

A series of reversible reactions follow next, which are shared by gluconeogenesis and glycolysis. These reactions result in conversion of PEP to fructose 1,6-bisphosphate (Fig. 9.23).

Second Bypass: Dephosphorylation of Fructose 1,6-bisphosphate

Hydrolysis of fructose 1,6-bisphosphate by *fructose 1,6-bisphosphatase* is the second reaction that is unique to gluconeogenesis. It bypasses the irreversible *phosphofructokinase* reaction. It also provides an energetically favourable pathway for the formation of fructose 6-phosphate ($\Delta G^{0\prime}$ = –3.9 kcal/mole). This reaction is an important control point for gluconeogenesis, discussed later in this chapter.

Third Bypass: Hydrolysis of Glucose 6-phosphate

Glucose 6-phosphatase catalyzes hydrolytic cleavage of the phosphate ester to liberate free glucose. Unlike the other gluconeogenic enzymes, which are cytoplasmic (except *pyruvate carboxylase*), this enzyme resides on the luminal surface of the endoplasmic reticulum (ER) membrane. Like the fructose *1,6-bisphosphatase* reaction, this step is also irreversible, and so provides an energetically favourable pathway for the formation of free glucose ($\Delta G^{0\prime} = -2.9$ kcal/mole).

Gluconeogenesis, the only source of glucose during long-term fasting, is based on a reversal of glycolysis with specific enzymes bypassing the three irreversible reactions of glycolysis.

C. Reversible Steps of Gluconeogenesis

In addition to the three bypass steps, there are seven reversible reactions common to both these pathways (Fig. 9.23). During gluconeogenesis, their equilibrium is pushed in favour of glucose synthesis.

A summary of all reactions of gluconeogenesis, including the reversible ones is as below:

$$2\ \text{Pyruvate} + 4\text{ATP} + 2\text{GTP} + 2\text{NADH} + 2\text{H}^+ + 4\text{H}_2\text{O} \rightarrow \text{Glucose} + 2\text{NAD}^+ + 4\text{ADP} + 2\text{GDP} + 6\text{P}_i + 6\text{H}^+$$

$$\Delta G^{0\prime} = -9\ \text{kcal/mole}$$

D. Gluconeogenesis is an Expensive Process

Several reactions of gluconeogenesis are energetically unfavourable. It requires input of considerable free energy to drive these "uphill" reactions: Thus six phosphoanhydride bonds are required for the synthesis of **one glucose** molecule from **two pyruvate** molecules (Table 9.4).

Pyruvate carboxylase reactions consume two ATPs, *PEP-carboxykinase consume* two GTPs, and *phosphoglycerate kinase* consume two ATPs in the reversal of the substrate level phosphorylation (see reaction 7; Fig. 9.3).

In addition, two NADH molecules are used up in the *glyceraldehyde 3-phosphate dehydrogenase* reaction, for each molecule of glucose synthesized. Since each NADH can generate 3ATPs by oxidative phosphorylation, this is equivalent to the input of another **six ATPs per glucose synthesized**.

In spite of being expensive, gluconeogenesis is fairly efficient—the liver can make a kilogram of glucose per day by gluconeogenesis.

E. Substrates for Gluconeogenesis

Pyruvate and TCA cycle intermediates are important substrates. In addition, a variety of other molecules can also yield glucose via the gluconeogenic sequence, as discussed here.

Lactate

Lactate produced in the skeletal muscles is the major substrate for gluconeogenesis (Cori's cycle). It is converted to pyruvate (in liver) by the *lactate dehydrogenase* reaction, which then enters gluconeogenic sequence.

Amino Acids

Glycogenic amino acids (all except lysine and leucine) are catabolized to pyruvate or some intermediate of citric acid cycle, which can generate glucose (Chapter 13). During fasting and starvation, the glycogenic amino acids become most important precursors for gluconeogenesis.

Glycerol

Glycerol is obtained during degradation of adipose triacylglycerols. It enters the gluconeogenic sequence in the liver. First, it reacts with ATP to form glycerol 3-phosphate, which is oxidized to dihydroxyacetone phosphate. The latter being a glycolytic intermediate can be converted to glucose.

Propionate

Catabolism of some amino acids (methionine, isoleucine) and odd chain fatty acids yields propionyl CoA, which enters a reaction sequence to ultimately yield succinyl CoA (Chapter 11). Being a TCA cycle intermediate, succinyl CoA is convertible to glucose.

***Fats cannot generate glucose*:** It is important to note that *fatty acids are NOT substrates for gluconeogenesis,* so it is generally said that glucose cannot be synthesized from fat. This is because fatty acids are oxidized to acetyl CoA via β-oxidation (Chapter 11). Acetyl CoA cannot be converted to pyruvate, because the *pyruvate dehydrogenase* reaction is irreversible and there are no alternative reactions to channel acetyl CoA into gluconeogenesis. However, *odd chain fatty acids are exceptions because their terminal three carbons may serve as gluconeogenic substrate.*

The substances that are degraded to acetyl CoA, including ketone bodies, ethanol and even chain fatty acids, are also not substrates of gluconeogenesis.

Table 9.4 Energy expenditure for synthesis of one glucose molecule from two molecules of pyruvate

Enzyme	Reaction	Energy consumption
Pyruvate carboxylase	2 × Pyruvate ⟶ 2 × Oxaloacetate	2 ATPs
PEP-carboxy-kinase	2 × Oxaloacetate ⟶ 2 × PEP	2 GTPs
Phosphoglycerate kinase	2 × 3 Phosphoglycerate ⟶ 2 × 1,3 Bisphoglycerate	2 ATP
Sum: 6 High energy phosphate bonds		

F. Regulation

Certain bypass steps of gluconeogenesis serve as important control points.

Regulation of Pyruvate to Oxaloacetate Conversion

Pyruvate carboxylase, the enzyme catalyzing this bypass step is a regulatory enzyme; it is allosterically inhibited by ADP and stimulated by acetyl CoA. *So potent is the effect of acetyl CoA that the enzyme is virtually inactive in its absence*. Stimulation by acetyl CoA ensures that sufficient oxaloacetate is generated (by *pyruvate carboxylase*) for condensing with it (acetyl CoA) to form citrate (Fig. 9.25a).

Regulation of Fructose 1,6-bisphosphate to Fructose 6-phosphate Conversion

ATP is a positive modulator, while AMP affects *fructose 1,6-bisphosphatase* negatively (Fig. 9.25b). Therefore, rate of gluconeogenesis is determined by cellular energy charge, reflected by the ratio of ATP : (ADP + AMP)

1. When the energy charge is low, as indicated by increased intracellular AMP concentration, rate of gluconeogenesis is impeded through inhibition of *fructose 1,6-bisphosphatase* by AMP.
2. Conversely, when the energy charge is high, rate of gluconeogenesis is enhanced due to stimulation of this enzyme by ATP. Since the corresponding glycolytic enzyme, PFK_1, is stimulated by AMP and ADP, and inhibited by ATP, these two pathways are regulated in a reciprocal manner. For example, when energy charge of the cell is low, glycolysis is favoured but gluconeogenesis is inhibited.

Hormonal Regulation

The above mechanisms bring about short-term regulation of gluconeogenesis. The day-to-day regulation is effected by the action of hormones on the amount of the key enzymes. **Insulin** represses the gluconeogenic enzymes (and induces most of the glycolytic enzymes); and these effects are counter-balanced by **glucagon** and other insulin antagonists.

Regulation by Glucagon

Glucagon, a peptide hormone produced by the α-cells in the endocrine pancreas stimulates gluconeogenesis by following mechanisms:

1. Glucagon raises the intracellular concentration of the second messenger cyclic AMP (cAMP), which causes conversion of the active form of *pyruvate kinase* to the inactive form (Fig. 9.9). This reduces the PEP to pyruvate conversion and the former is diverted into the gluconeogenic sequence.
2. Glucagon **reduces** the intracellular concentration of fructose **2,6-bisphosphate** (Fig. 9.8). This compound is a negative allosteric modulator of *fructose 1,6-bisphosphatase* and, as explained earlier, a positive modulator of *phosphofructokinase* (PFK_1); both these factors ultimately favour gluconeogenesis.

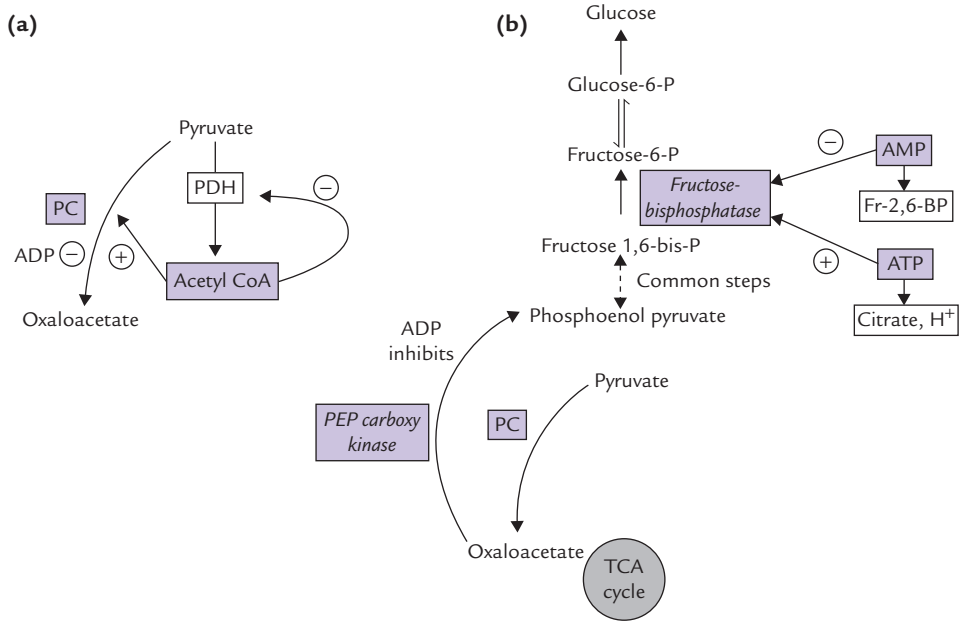


Fig. 9.25. Regulation of gluconeogenesis. **(a)** Acetyl CoA as the positive allosteric modulator of *pyruvate carboxylase* (PC) and inhibitor of *pyruvate dehydrogenase* (PDH) complex, **(b)** Role of various other allosteric modulators (⊖ = inhibits reaction ⊕ = accelerates reaction, Fr-2,6-BP = fructose 2,6-bisphosphate).

3. Glucagon increases fatty acid mobilization and their oxidation to yield acetyl CoA, NADH and ATP; these compounds inhibit the PDH complex and acetyl CoA stimulates *pyruvate carboxylase* (Fig. 9.25a). Thus, the available pyruvate is not converted to acetyl CoA but to oxaloacetate for gluconeogenesis.
4. Glucagon causes **induction** of the key *enzymes of gluconeogenesis*, i.e. *phosphoenolpyruvate carboxykinase* and *glucose 6-phosphatase* (and possibly *pyruvate carboxylase* also).

Regulation by Insulin

Insulin causes repression of severed gluconeogenic enzymes, a detailed account of which is given in Chapter 15.

Reciprocal Regulation of Gluconeogenesis and Glycolysis

It is evident from the above discussion that glucagon not only stimulates gluconeogenesis but concomitantly inhibits glycolysis. Allosteric modulations also ensure that the opposing enzymes of these two processes are not active at the same time, which in turn prevents futile cycles (refer to Box 9.8 for details):

1. As shown in Figure 9.26, compounds reflecting high energy state of cell, e.g. ATP, citrate and acetyl CoA, allosterically stimulate the key gluconeogenic enzymes but inhibit the corresponding enzymes of glycolysis.
2. Fructose 2,6-bisphosphate also has opposite effects on *phosphofructokinase* (stimulatory) and *fructose 1,6-bisphosphatase* (inhibitory). *Levels of fructose 2,6-bisphosphate are high in fed state* (and low in starvation) so that glycolysis is accelerated and glyconeogenesis is inhibited in fed state. During starvation, glyconeogenesis predominates because levels of fructose 2,6-bisphosphate are low (refer to Box 9.7 for details).

GN Ratio or DN Ratio

It refers to the *ratio between the dextrose (glucose) to nitrogen excreted in urine*. In the experimental animals treated with alloxan (causes destruction of pancreatic β-cells) or phlorizin (causes renal tubular cell necrosis) glucose is

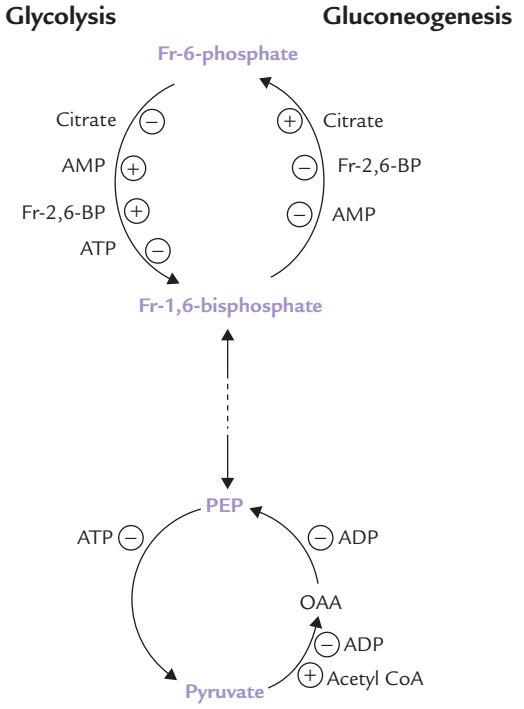


Fig. 9.26. Reciprocal regulation of glycolysis and gluconeogenesis.

BOX 9.7

Gluconeogenesis in Starvation and Well-fed State

Intramitochondrial concentration of acetyl CoA rises in starvation due to suppression of TCA cycle and stimulation of β-oxidation (Chapter 15). Acetyl CoA causes stimulation of the *pyruvate carboxylase*, which accelerates the pyruvate to oxaloacetate conversion. Since acetyl CoA inhibits activity of *pyruvate dehydrogenase*, pyruvate to acetyl CoA conversion is concomitantly inhibited. Thus, most of the available pyruvate is directed to oxaloacetate formation (Fig. 9.25). Oxaloacetate, so formed, is mostly channeled into gluconeogenic sequence for the generation of glucose (OAA cannot enter the TCA cycle since this pathway is suppressed in starvation). Generation of glucose in this manner ensures that normal blood glucose levels are maintained in starvation.

Conversely, when intracellular level of acetyl CoA falls, *pyruvate carboxylase* becomes inactive. Such situation arises in the well-fed state, when insulin release is enhanced, which funnels pyruvate into the other available route, i.e. oxidation by *pyruvate dehydrogenase* pathway to acetyl (CoA).

In short, during starvation pyruvate is mostly used for the production of OAA, a gluconeogenic substrate, whereas in well-fed state it is converted to acetyl CoA, which enters TCA cycle and lipogenic sequence.

lost through urine and body attempts to generate glucose through gluconeogenesis. The body proteins are degraded to provide amino acid substrates for gluconeogenesis, and their catabolic end-product, urea, is excreted in urine. The ratio of glucose to urea nitrogen in such animals is found to be **3.65**, meaning that one gram of nitrogen (from proteins) forms 3.65 grams of glucose.

GN ratio is increased when catabolism is enhanced, such as in starvation, hyperthyroidism and cancer.

Because 3.65 grams glucose is synthesized from one gram of nitrogen of protein, and proteins contain 16% nitrogen, it is calculated that 58% of protein is glycogenic.

VI. Glycogen Metabolism

Glycogen is stored mainly in liver and skeletal muscles as an energy reserve, just as starch in plants. The tissue concentration of glycogen in liver (6–8%) is higher than that in muscle (1–2%), but because of the relative masses of muscle and liver, the majority of glycogen in the body is stored in muscles. Though total hepatic glycogen stores (50–100 g) are significantly less than the muscle stores (200–300 g), they are involved in a vital activity, that is to serve as our first line of defense against declining blood glucose level, particularly between meals. Muscle glycogen is essential for muscle energy metabolism during bursts of physical activity even though muscle relies primarily on fats as a source for energy.

Glycogen is the storage form of carbohydrates in liver and muscles. Hepatic glycogen generates glucose (due to presence of *glucose 6-phosphatase*), whereas muscle glycogen serves as a source of metabolic fuel for use in muscle.

Glycogen is an extensively branched homopolysaccharide, consisting of chains of α(1→4) linked glucosyl residues with branches that are joined by α(1→6) linkages (Chapter 2). The latter are spaced about every 4–6 residues along the α-1,4 chain. The gross structure of glycogen is dendritic in appearance, expanding from a core sequence bound to a tyrosine residue in the protein, **glycogenin**.

A. Glycogen Synthesis (Glycogenesis)

Glycogen is synthesized from glucose. Like most other anabolic processes, glycogenesis occurs in the cytosolic fraction of the cell and requires input of energy. The energy for glycogenesis comes from high energy nucleotides, ATP and UTP. These nucleotides provide energy for the conversion of the precursor glucose molecule to its energized form, UDP-glucose. The *UDP-glucose (UDPG) serves as activated donor of glucose residues* that are added to the growing glycogen molecule. ***Glycogen synthase*** is the major enzyme of this complex process, which occurs in following four stages:

Synthesis of UDP-glucose

Activation of glucose to form UDP-glucose (UDPG) occurs in three sequential reactions, outlined in Figure 9.27.

Reaction 1

Glucose is phosphorylated at C-6 by *glucokinase* in liver (*hexokinase* in muscle) to form glucose 6-phosphate.

Glucokinase has high K_m (i.e. low affinity for glucose) and high V_{max} which permits it to rapidly phosphorylate large quantities of glucose after meals, when the ambient glucose concentration is high.

Reaction 2

The phosphate group of glucose 6-phosphate is then shifted from the sixth carbon to the first carbon of the molecule by *phosphoglucomutase* to form glucose 1-phosphate.

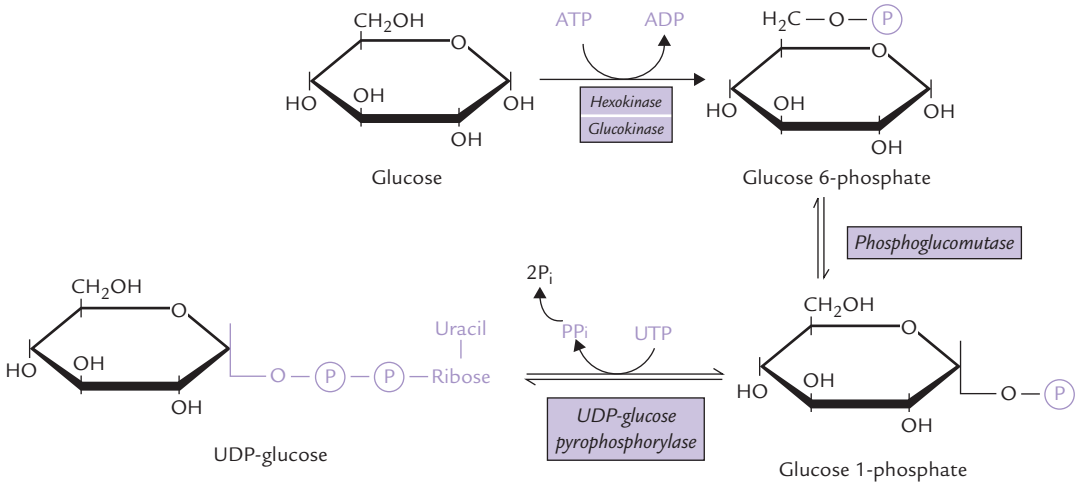


Fig. 9.27. Synthesis of UDP-glucose, the activated form of glucose for glycogen synthesis.

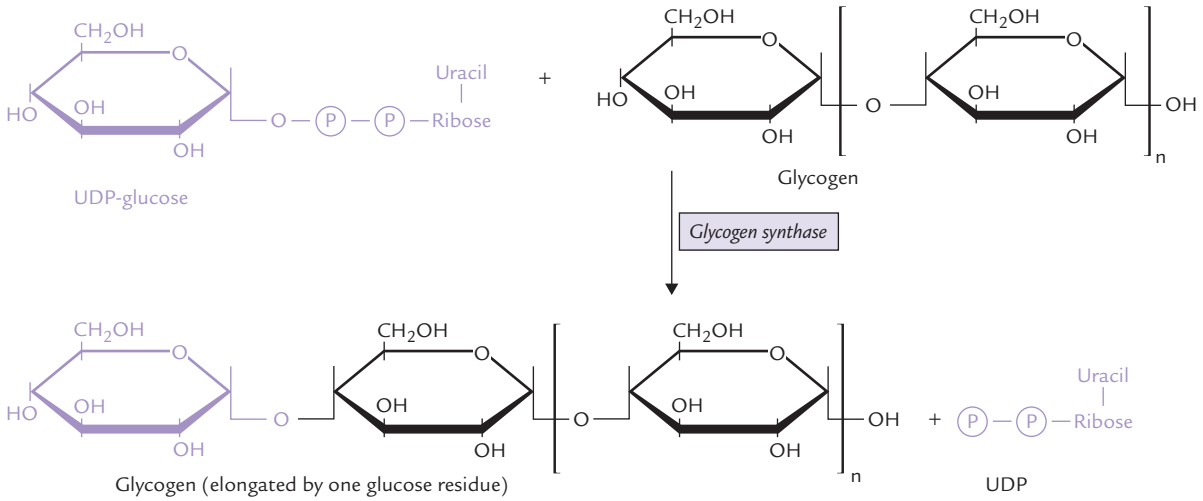


Fig. 9.28. The *glycogen synthase* reaction; UDPG serving as an activated donor of a glucose residue.

Reaction 3

Activation of glucose 1-phosphate to the sugar nucleotide, uridine diphosphate glucose (UDPG) occurs by the enzyme *UDPG pyrophosphorylase*.

Glucose 1-phosphate + UTP → UDPG + PPi

Pyrophophosphate produced in this reaction is hydrolyzed to inorganic phosphates by *pyrophosphatase*; this ensures irreversibility of the reaction.

Synthesis of Primer to Initiate Glycogen Synthesis

UDPG donates glucose residues to an existing α(1→4) glucosyl chain called **primer,** which will accept the incoming glucose residues.

UDPG + Glycogen primer → UDP + Glycogen
(n residues) (n + 1 residues)

Normally a fragment of glycogen serves as a primer. Such fragments are obtained from glycogen molecules that had been partially degraded in liver during fasting or in muscle during exercise. However, when the glycogen stores are depleted, a specific protein, known as **glycogenin**, provides the site at which the primer is built. Glycogenine contains a specific tyrosyl residue, which accepts the first glucose donated by UDPG.

–Tyrosyl–OH + UDPG ⟶ –Tyrosyl–O–glucose + UDP

The above reaction is catalyzed by glycogenin (**autocatalysis**) or by the enzyme *glycogen synthase*.

More of glucosyl residues coming from UDP glucose, are attached sequentially in the same manner. Thus a short α(1→4) glucosyl chain (i.e. the primer) is built which is attached to the protein.

Elongation of Chains by *Glycogen Synthase*

The primer is elongated by sequential addition of glucose residues. The enzyme *glycogen synthase* catalyzes this reaction. It transfers the glucose residues from UDPG to the non-reducing end of glycogen in α(1→4) linkage (Fig. 9.28).

UDPG + Glycogen → UDP + Glycogen (elongated by one glucose residues)

This reaction has been elaborately shown in Figure 9.28.

UDP produced in the above reaction is reconverted to UTP by the enzyme *nucleoside diphosphate kinase*.

UDP + ATP → UTP + ADP

Formation of Branches in Glycogen

The reactions discussed so far result in formation of linear, unbranched chains that are linked by α(1→4) linkage. However, glycogen is a highly branched structure; the branch points are created by the action of the "**branching enzyme**". This enzyme comes into action after glycogen synthase has added at least ten glucosyl units (Fig. 9.29).

The branching enzyme removes a block of five to eight glucose residues from the non-reducing end of the chain and transfers it to a more internal location on the same or other chain. The transferred block is attached via α(1→6) linkage to a glucose residue at the new location. The branching enzyme is also referred to as a ***glucosyl 4 : 6 transferase*** because it forms a new α(1→6) bond instead of an α(1→4) bond.

The transfer of a block results in creation of a new non-reducing end. More of glucose residues can be added to this end. In this way, branching increases the number of non-reducing ends to which new glucose residues can be added, thereby greatly accelerating the rate at which glycogen synthesis occurs.

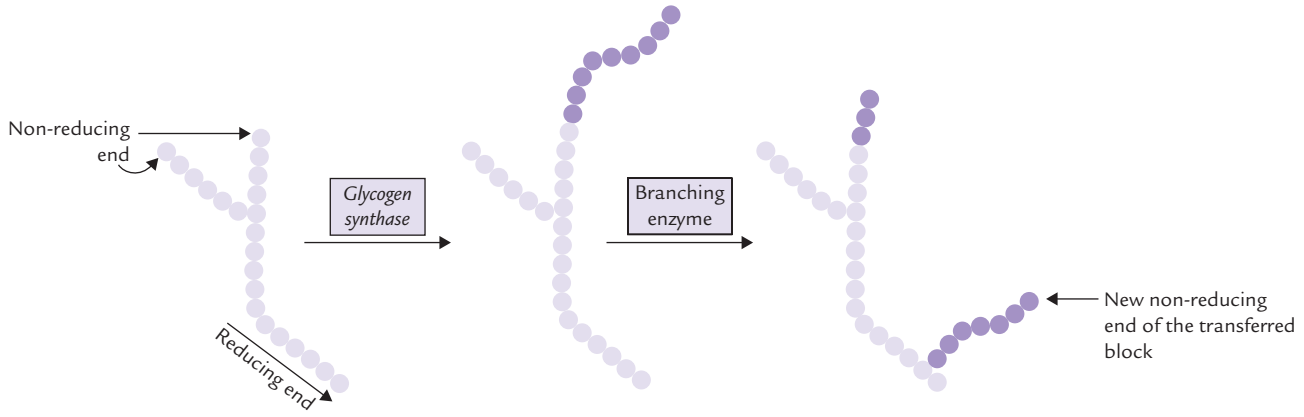


Fig. 9.29. The action of branching enzyme. It forms branches by transferring a block of glycosyl residues, typically seven in length, from the end of an unbranched chain to a more interior location (on the same or other chain). The transferred block is attached via α1–6 linkage. → indicates non-reducing ends [each circle represents a glucose residue].

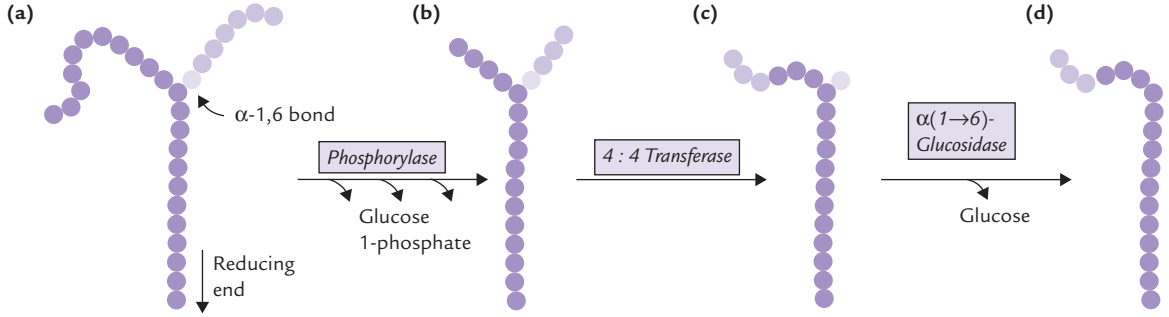


Fig. 9.30. Glycogen degradation. Designation of glucose residues as in Figure 9.28.

B. Glycogen Degradation (Glycogenolysis)

The principal enzyme of glycogenolysis is ***glycogen phosphorylase***. It acts in association with a debranching enzyme (Fig. 9.30).

Action of *Glycogen Phosphorylase*

Glycogen phosphorylase removes glucose residues, one at a time, from the non-reducing ends of glycogen (Fig. 9.30a). It utilizes inorganic phosphate (P_i) to cleave the α(1→4) bonds. This results in release of the terminal glucosyl residue as glucose 1-phosphate. The reaction is termed as **phosphorolysis**, that is, breakage of a covalent bond by addition of a phosphate group.

$$\underset{(n \text{ residues})}{\text{Glycogen} + P_i} \rightleftharpoons \underset{(n-1 \text{ residues})}{\text{Glycogen} + \text{Glucose 1-phosphate}}$$

Pyridoxal phosphate is an essential cofactor in this reaction; it is covalently bound to the enzyme protein.

Phosphorylase is specific for α(1→4) linkages only, it cannot cleave α(1→6) linkages. Further, this enzyme cannot approach the branching glucose residues efficiently. Thus, as shown in Figure 9.30b, *phosphorylase* cleaves the external glucose residues until the branches are about four residues long. Then the action of *glycogen phosphorylase* stops and debranching enzyme comes into play.

Removal of Branches

The key enzyme for removing branch points of glycogen is the debranching enzyme, which possesses dual activity, namely *glucosyl 4 : 4 transferase* activity and *α(1→4) glucosidase* activity.

4 : 4 transferase activity: Three of the four external glucosyl residues that remain are removed as a trisaccharide and transferred to the non-reducing end of a nearby chain (Fig. 9.30c). This action involves cleaving of an α(1→4) bond at one site and formation of a new α(1→4) bond elsewhere.

α(1→6) glucosidase activity: The single glucosyl residue that remains at the branch point is removed by the α(1→6) *glucosidase*, to liberate *free glucose* (Fig. 9.30d).

Removal of branch point in this manner exposes another set of α(1→4) linkages, till the next branch point. When another branch point is reached, it is again removed by the debranching enzyme, and the cycle thus continues.

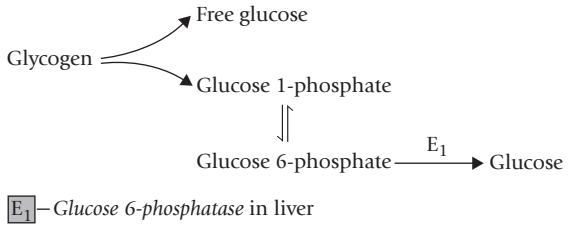
Thus, the phosphorolysis/debranching processes, acting alternately, can cleave a large glycogen molecule having thousands of glycosyl residues.

Fate of Glucosyl Units Released from Glycogen

The glucosyl units are released from glycogen in two forms: glucose 1-phosphate and free glucose. *About 90% of the glucose is released as glucose 1-phosphate, and only the α-1,6 branching residue, is released as free glucose*. Glucose 1-phosphate is then converted to glucose 6-phosphate by the enzyme *phosphoglucomutase.*

Further, fate of glucose 6-phosphate varies depending on the tissue. **Liver** possesses the enzyme *glucose 6-phosphatase*, which forms free glucose from glucose 6-phosphate. This glucose is released into blood for use by needy tissues.

Muscles *lack glucose 6-phosphatase*, hence cannot contribute to the blood glucose. Rather muscle glycogen is used for the generation of metabolic energy in the exercising muscles themselves.



Role of liver glycogen and muscle glycogen are different: Liver releases free glucose in blood circulation, and thereby plays an important role in glucose homeostasis. However, hepatic glycogen stores are barely sufficient for maintenance of blood glucose concentration during a 12-hour fast.

Muscle glycogen stores are used for liberating fuel molecules (i.e. glucose 6-phosphate), which in turn are used in the exercising muscles for generating metabolic energy.

C. Regulation

Synthesis and breakdown of glycogen are important intracellular activities, having implications beyond the cell. Balance between the two processes is important for:

(a) sustaining adequate glycogen stores, and
(b) for maintaining the normal blood glucose levels.

Normal blood glucose levels are especially important for the tissues that use **glucose as their primary substrate**, such as, **brain erythrocytes**, **renal medulla**, **lens** and **cornea of the eye**. Therefore, tight regulation of these two processes (i.e. glycogenesis and glycogenolysis) is essential.

Regulation of Activity of *Glycogen Phosphorylase*

The regulatory enzyme for glycogenolysis is *glycogen phosphorylase*. Activity of this enzyme is regulated by (a) covalent modulation through phosphorylation-dephosphorylation, (b) allosteric regulation modulation, and (c) calcium ions.

Covalent Modulation

Studies concerning regulation of *glycogen phosphorylase* have been mostly conducted in skeletal muscles. The muscle enzyme is a dimeric protein consisting of two identical subunits. Each subunit contains an essential serine residue, where phosphate group can covalently attach. Thus, the *glycogen phosphorylase* exists in two forms:

- Phosphorylated form called *phosphorylase a* which is catalytically active.
- Dephosphorylated form called *phosphorylase b* which is much less active.

These two forms of the enzyme are interconvertible. Conversion of inactive *phosphorylase b* to active *phosphorylase a* is a catalyzed by the enzyme ***phosphorylase kinase***, which phosphorylates, and thereby activates *glycogen phosphorylase*. Another enzyme, ***protein phosphatase-1***, dephosphorylates, and thereby inactivates *glycogen phosphorylase a* (Fig. 9.31).

Through action of these enzymes, ratio of the active and the inactive *phosphorylase* can be varied, which ultimately controls the rate of glycogenolysis. The same mechanism is operative in liver as well.

Allosteric Regulation

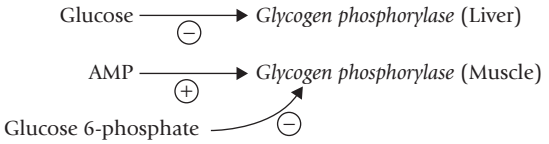
Superimposed upon the covalent modulation are the actions of allosteric effectors:

1. *AMP: Glycogen phosphorylase* in muscles and other extrahepatic tissues is stimulated powerfully by AMP (Fig. 9.31). This allosteric effect ensures that glycogen is degraded rapidly in severely contracting muscles, to provide substrate for an anaerobic glycolysis. Concentration of AMP quickly builds up in contracting muscles due to its rapid production by pyrophosphate cleavage of ATP

   $$ATP \rightarrow AMP + PPi$$

   Note: Unlike other forms of stored energy, only glycogen can be used for ATP synthesis under anaerobic conditions.

2. *Glucose*: *Glycogen phosphorylase* in liver is inhibited by glucose. Because the intracellular glucose concentration in the liver approximates the blood glucose level, glycogen degradation in liver is regulated directly by the blood glucose level.
3. *Glucose 6-phosphate*: It inhibits muscle glycogenolysis.
4. *ATP*: Both liver and muscle glycogenolysis are inhibited by ATP.



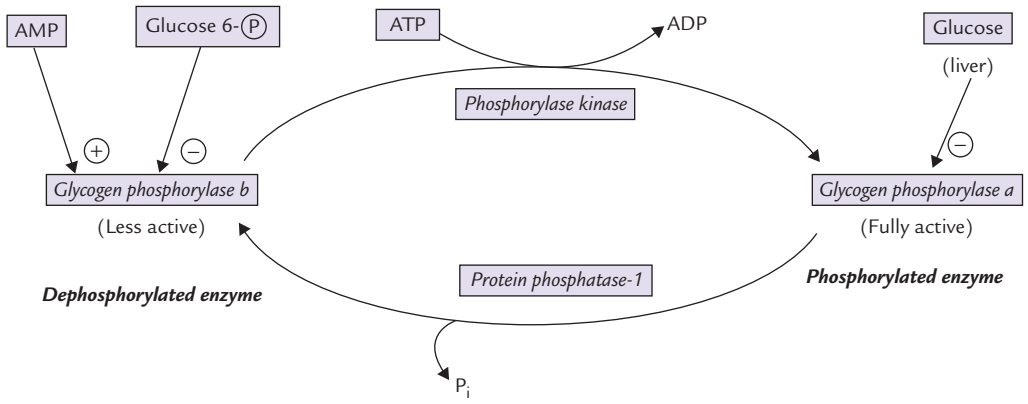


Fig. 9.31. Regulation of *glycogen phosphorylase* by covalent modification and allosteric effectors.

Table 9.5 Difference between glycogenolysis in liver and muscles

	Muscle	Liver
1.	Little glucose formed, since *glucose 6-phosphatase* is absent	Free glucose forms by *glucose 6-phosphatase*
2.	Stimulated by adrenaline, glucagon has no effect	Glucagon is major stimulator of glycogenolysis
3.	Stimulated by AMP but inhibited by glucose 6-phosphate	No effect of AMP, but hepatic *glycogen phosphorylase* is inhibited by glucose

To sum up, the various mechanisms described ensure that *glycogenolysis is stimulated when glucose concentration and energy levels are low, and inhibited when these are high.*

Regulation by Calcium Ions

During muscle contraction, calcium is released from sarcoplasmic reticulum and activates *phosphorylase kinase* (via a calmodulin-calcium modulating protein; Chapter 29). This enzyme causes phosphorylation of *glycogen phosphorylase* and hence enhances glycogenolysis (without involving cAMP).

Hepatic and muscle glycogenolysis serve different roles and respond to different regulatory signals. Major differences between the two are given in Table 9.5.

Regulation of Activity of *Glycogen Synthase*

The regulatory enzyme of glycogenesis is *glycogen synthase*. Regulation of its activity is effected by the following mechanisms: (a) allosteric regulation, and (b) covalent modulation.

Covalent Modulation

Activity of *glycogen synthase* is also regulated by phosphorylation-dephosphorylation, as in the case of *glycogen phosphorylase*. However, there is a major difference: *glycogen synthase* is more active in the dephosphorylated form and less active in the phosphorylated form (Fig. 9.32). Thus, it is activated by *protein kinase* and rendered less active by *protein phosphatase-1*.

It is important to note that *simultaneous phosphorylation of both glycogen-phosphorylase and synthase will switch the cell from glycogen synthesis to glycogen degradation*. This prevents futile cycles (Box 9.8).

Allosteric Regulation

Glucose 6-phosphate is a potent activator of glycogenesis. It does so by stimulating the *protein phosphatase-1*, so that *glycogen synthase* is converted to the active (dephosphorylated) form. It also causes direct allosteric stimulation of the *glycogen synthase* activity. In **muscle tissue,** glucose 6-phosphate not only stimulates glycogen synthase, but also inhibits activity of *glycogen phosphorylase*.

Availability of glucose 6-phosphate is high in fed state, which ensures activation of glycogenesis and hence glycogen deposition.

Hormonal Regulation of Glycogen Metabolism

Glucagon and epinephrine, acting through cAMP, are primarily involved in regulation of the glycogen metabolism. ***Glucagon* acts on liver cells and *epinephrine* acts on** both **liver and muscle cells**. These hormones stimulate glycogenolysis and inhibit glycogenesis. ***Insulin***, on the other hand, promotes glycogenesis and inhibits glycogenolysis.

Role of Glucagon and Epinephrine

Both these hormones activate the membrane enzyme *adenylate cyclase*, which catalyzes formation of cAMP from ATP (Fig. 9.33). The cAMP binds to *protein kinase A* (**PKA**, also called cAMP-dependent *protein kinase*), a tetrameric protein. The cAMP binding activates this enzyme by letting its catalytic subunits free (Chapter 29). Activation of PKA has the following consequences: (i) *Inhibition of glycogenesis*, and (ii) *Stimulation of glycogenolysis*.

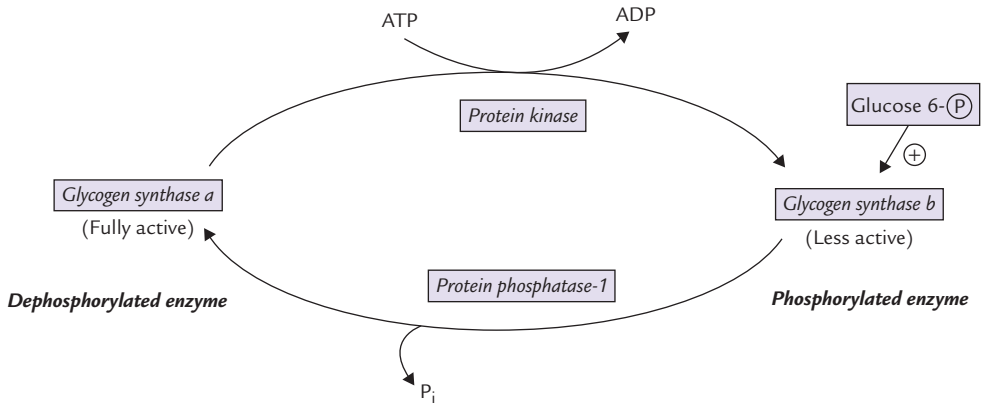


Fig. 9.32. Regulation of *glycogen synthase* by covalent modification and allosteric effectors.

BOX 9.8

Futile Cycles

The synthetic and the degradative pathways of metabolism are fine-tuned by allosteric effectors and hormone-induced enzyme phosphorylation in order to meet the cellular demands for metabolites and energy, and at the same time minimize their wastage. Normally, opposing pathways (e.g. glycolysis and gluconeogenesis; glycogen synthesis and its degradation) are reciprocally regulated: activation of one is accompanied by inactivation of the other, and vice versa. This is necessary because simultaneous activity of these opposing pathways would lead to a wasteful metabolic exercise, called **futile cycle**, that achieves little but ATP hydrolysis.

Futile cycles are minimized through reciprocal regulation of the opposing enzymes. For example, the key enzymes of glycogen synthesis (*glycogen synthase*) and glycogen degradation (*glycogen phosphorylase*) can both be phosphorylated on specific serine side chains by *protein kinases* (and dephosphorylated by *protein phosphatase*). *Glycogen synthase* is more active in the dephosphorylated form and *glycogen phosphorylase* is more active in the phosphorylated form. Therefore, the simultaneous phosphorylation of both enzymes will switch the cell from glycogen synthesis to glycogen degradation. Conversely, simultaneous dephosphorylation will switch the cell from glycogen degradation to glycogen synthesis, simultaneous activation of both (i.e. futile cycling) being prevented in either case.

- *Inhibition of glycogenesis*: The active PKA cause phosphorylation of *glycogen synthase.* Since the phosphorylated *glycogen synthase* is less active, glycogenesis is inhibited.
- *Stimulation of glycogenolysis*: The active PKA stimulates glycogenolysis by a dual mechanism, as discussed below:
  (a) It catalyzes phosphorylation of *phosphorylase kinase,* which exists as an inactive dephosphorylated form and an active phosphorylated form. Thus, *phosphorylase kinase* is now activated, and in turn phosphorylates the *glycogen phosphorylase b,* converting it to *phosphorylase a,* that now carries out a rapid dehydration of glycogen (Fig. 9.33).
  (b) Active PKA phosphorylates a protein called inhibitor-1, which thereby gets stimulated. This protein inhibits *protein phosphatase-1,* the enzyme that normally causes inactivation of *glycogen phosphorylase a.* The overall effect is that the *glycogen phosphorylase* a dose not get inactivated, and is *"locked"* in activated (phosphorylated) form providing a persistent stimulus for glycogenolysis (Fig. 9.34).

Note: Inhibition of *protein phosphatase-1* locks *glycogen synthase* in its inactive (phosphorylated) form, hence suppressing glycogenesis.

Cascade with amplification: The cAMP activated process is a cascade in which the initial hormonal signal is amplified several folds. Single hormone molecule of glucagon or adrenaline activates the enzyme *adenylate cyclase* for long enough to produce hundreds of cAMPs which activate *protein kinase* molecules. Although four cAMP molecules are required to activate a *protein kinase* (Chapter 29), each active PKR molecule can convert hundreds or thousands of molecules of *phosphorylase kinase* to active form. Similarly, each active molecule of *phosphorylase kinase,* can convert thousands of molecules of *glycogen phosphorylase b* to *phosphorylase a,* and so on. *The net result*

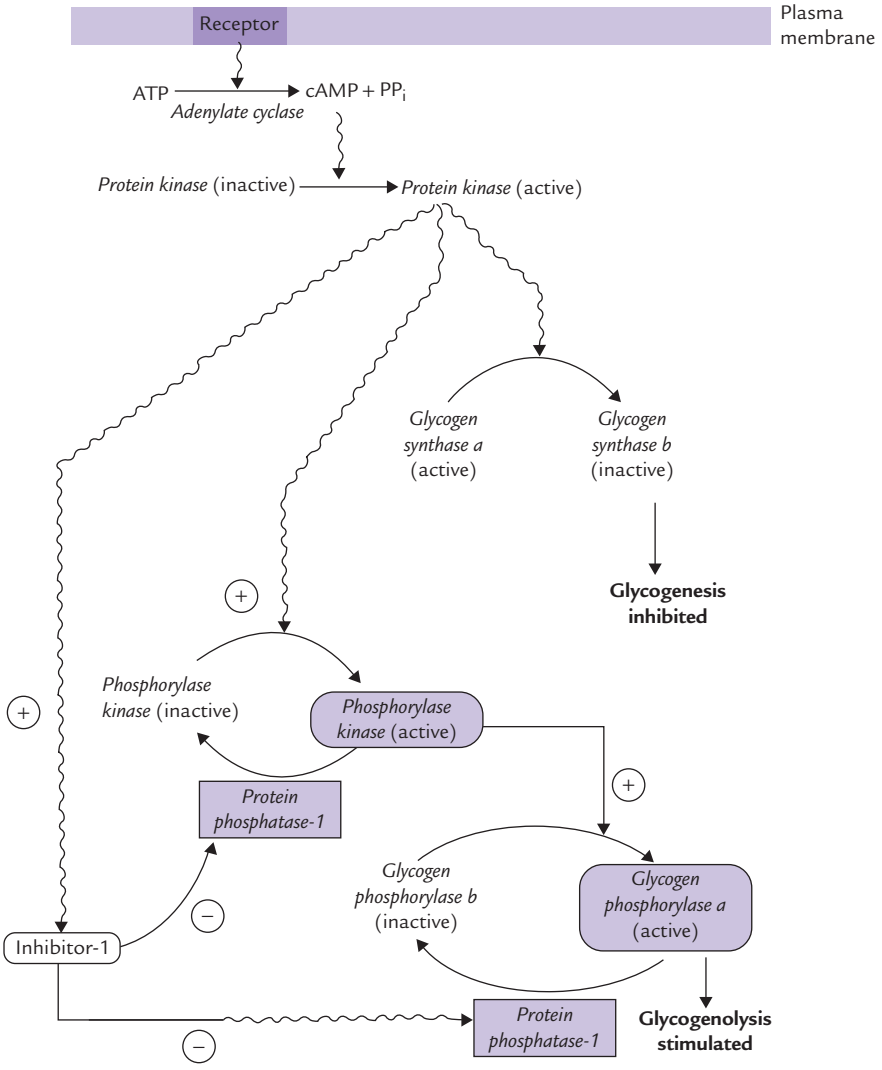


Fig. 9.33. Hormonal regulation of glycogen synthesis and degradation. The cAMP generated by hormones activates *protein kinase A*, which cause phosphorylation of *glycogen synthase* (inactivation), *phosphorylase kinase* (activation) and the inhibitor-1 (activation). The last inhibits *protein phosphatase*. Activated *phosphorylase kinase* causes phosphorylation of *glycogen phosphorylase b*, activating it to *glycogen phosphorylase* (⊖ = inhibitor, ⊕ = stimulator).

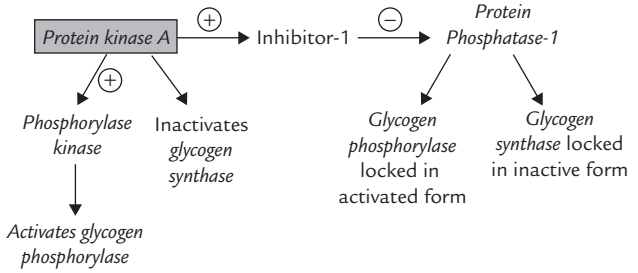


Fig. 9.34. Dual control of glycogenesis and glycogenolysis by *protein kinase A*.

is that a single hormone molecule can generate thousands of molecules of glucose 1-phosphate by glycogenolysis.

Role of Insulin

Insulin is the key hormone that **stimulates glycogen synthesis**. It is released in the fed state and **activates** an intracellular ***phosphatase.*** The latter causes dephosphorylation of *glycogen synthase.* Dephosphorylation causes this enzyme to be activated, resulting in increased glycogen synthesis.

Moreover, insulin stimulates the uptake of glucose by muscle, providing substrate molecules for glycogen synthesis. Dephosphorylation of *glycogen phosphorylase* causes a concomitant **fall** in the rate of **glycogenolysis**. Finally, insulin promotes activity of *phosphodiesterase* (it degrades cAMP) in liver, which decreases cAMP levels. This enhances glycogenesis and suppresses glycogenolysis.

Glycogen is synthesized in pathway of glycogenesis (major enzyme, *glycogen synthase*) and degraded by phosphorolytic cleavage (major enzyme, *glycogen phosphorylase*) in pathway of glycogenolysis. The two pathways are reciprocally regulated by hormones and metabolites.

D. Glycogen Storage Diseases

Any deficiency of a glycogen degrading enzyme causes abnormal accumulation of glycogen in the affected tissue. The resulting diseases, summarized in Table 9.6, are collectively known as the glycogen storage diseases. These diseases are quite rare (overall incidence of 1 in 40 000) and are mostly inherited as autosomal recessive trait. Normally a defect in an enzyme of glycogenolysis (or glycogenesis) leads to abnormal accumulation of glycogen. However, defective action of a glycolytic enzyme, *phosphofructokinase,* may also result in abnormal glycogen accumulation.

In some forms, the main abnormality lies in liver, whereas in others muscle tissue is primarily affected. Accordingly, glycogen storage diseases can be classified in two major types: the **hepatic** forms and the **myopathic** forms. Certain forms (e.g. Type II) cannot be put into either of these categories. Some rare forms, e.g. IX, X and XI have also been identified which are due to defects in enzymes that activate or deactivate liver *phosphorylase*.

Hepatic Forms

Type I, *III*, *IV*, *VI* and *VII* are classified as the hepatic forms of glycogen storage disease. These are characterized by excessive deposition of glycogen in liver, and hence liver enlargement (hepatomegaly) is the prominent feature. The biochemical manifestations are hypoglycaemia, lactic acidosis, hyperuricaemia and hyperlipidaemia, which are most prominent in the type I form (von Gierke's disease), but may occur in varying severities in the other forms. Enzymatic defects in various forms are shown in Table 9.6.

von Gierke's disease (Type I): It is a classical hepatic glycogen storage disease, caused by deficiency of *glucose 6-phosphatase*. It is a rare disorder with an incidence of 1 per 200,000 persons. It is characterized by severe hepatomegaly and dangerous hypoglycaemia, that may develop within hours of last meal (not surprising, because in between meals *glucose 6-phosphatase* is required for formation of glucose by both glycogenolysis and gluconeogenesis). *The patient may be kept alive only by feeding carbohydrates at regular intervals, day and night.* The hypoglycaemia is unresponsive to glucagon or adrenaline because of the metabolic block in formation of free glucose from liver glycogen stores. Other manifestations of von Gierke's disease are as follows:

1. ***Lactic acidosis:*** It may develop because the glucose 6-phosphate cannot be degraded to glucose, and is therefore, diverted into the glycolytic sequence to form pyruvate and then lactate. Lactate level in blood increases and pH is lowered (acidosis).

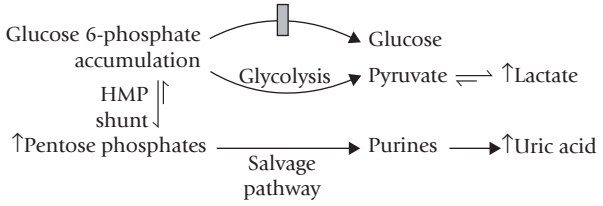
Table 9.6. Glycogen storage diseases

Type		Enzyme deficiency	Organ (s) affected
I	**von Gierke's disease**	*Glucose 6-phosphatase*	Liver, kidney
II	**Pompe's disease**	$\alpha(1\rightarrow4)$ *Glucosidase* (acid maltase)	All organs
III	**Cori's disease**	Debranching enzyme	Muscle, liver
IV	**Andersen's disease**	Branching enzyme	Liver, myocardium
V	**McArdle's disease**	*Phosphorylase*	Muscle
VI	**Hers' disease**	*Phosphorylase*	Liver
VII	**Tarui's disease**	*Phosphofructokinase*	Muscle, RBCs
VIII		*Phosphorylase kinase**	Liver

* There is also on X-linked form of *phosphorylase kinase* deficiency. This is sole exception as all other glycogen storage diseases are inherited as autosomal recessive trait.

2. ***Hyperuricaemia:*** It may develop because accumulation of glucose 6-phosphate enhances the reactions of HMP shunt (2nd stage). This stage involves interconversion of hexose phosphates (e.g. glucose 6-phosphate) and pentose phosphates (e.g. ribose 5-phosphate). Therefore, accumulation of glucose 6-phosphate leads to increased formation of ribose 5-phosphate. This leads to increased uric acid production since ribose phosphate enhances formation of purine nucleotides by "salvage pathway" (Chapter 20), which are then degraded to uric acid.

   Moreover, lactic acidosis inhibits renal excretion of uric acid, which aggravates hyperuricaemia.



3. *Hyperlipidaemia:* It is the result of decreased glucose availability, leading to mobilization of the stored fats.

The above features may occur, in varying degrees, in the other hepatic forms of glycogen storage diseases (Table 9.6).

Cori's disease (Type III): Branched chain glycogen accumulates because of absence of the debranching enzyme. Clinical course runs like the type I, but is much milder.

Anderson's disease (Type IV): A rare disease in which glycogen having few branches accumulates because of absence of the branching enzyme. Death from liver cirrhosis usually occurs before the age of two years.

Her's disease (Type VI): Liver glycogen stores cannot be degraded due to *phosphorylase* deficiency, so liver enlarges and mild hypoglycaemia and ketosis are often seen. The disease runs a mild course.

Type VIII: Mild hepatomegaly and hypoglycaemia characterize this condition.

Myopathic Forms

These include types V (McArdle's disease) and VII (Tauri's disease). Muscle glycogen stores are very high but are not available for use. Because glycogen is the primary substrate in the exercising muscles, the latter are most severely affected. Clinical course of both type V and type VII runs similarly: the affected patients suffer from muscle cramps and pain upon exertion and are easily fatigued. Apart from this they lead a normal life. This shows that the utilization of muscle glycogen is not essential for life. Interestingly, after muscular activity, these patients do not show the expected increase in blood lactate. This demonstrates that the most important source of lactic acid during muscular activity is not the free glucose from blood but stored muscle glycogen.

Pompe's disease (Type II), which involves all organs, needs a special mention because of its peculiar features. In this disease, glycogen accumulation occurs in lysosomes since ability to hydrolyze it is impaired. This is because of impairment (or absence) of the lysosomal enzyme *α(1→4) glucosidase*. Consequently the lysosomes get distended with glycogen, especially in the myocardium, skeletal muscle, liver and motor nuclei of spinal cord.

The infantile form of this disease is very severe, and the child dies in first few months of life due to cardiac enlargement, that leads to cardiac failure. The juvenile form of this disease is also fatal, but the adult form is relatively milder, being marked by slowly developing myopathy.

Glycogen accumulates in several rare enzyme deficiencies, collectively termed glycogen storage diseases. These diseases (except one) are inherited as autosomal recessive traits.

Exercises

Essay type questions

1. Enumerate the rate regulating enzymes of glycolysis and glyconeogenesis. How are these enzymes regulated in a reciprocal manner?
2. Describe glycolysis and its regulation. Explain energetics of aerobic and anaerobic glycolysis.
3. List distinctive features of *glucokinase* and *hexokinase.* Explain why the high Km of *glycokinase* is important for the role of the liver in buffering blood glucose.
4. Describe the conditions under which the Cori's cycle and the glucose-alanine cycle operate.
5. Describe the reactions of the *pyruvate dehydrogenase* multienzyme complex. How is entry of acetyl CoA into the TCA cycle regulated at this complex?
6. Mention the reactions of glycogenesis and glycogenolysis in flow diagram. Describe how these pathways are regulated. Describe the distinctive features of glycogen synthesis and degradation in liver and skeletal muscle.
7. With at least one example each related to glucose metabolism, discuss how enzyme activity is regulated by (a) allosteric mechanism, and (b) covalent modification.
8. Highlight the role played by *protein kinases* in glycogen metabolism. Explain the differences in glycogenolysis between liver and muscle.

Write short notes on

1. Cori's cycle
2. Glycogen storage diseases
3. Anaerobic glycolysis
4. Pasteur effect
5. McArdle's disease
6. GLUTs
7. Fluoride as inhibitor of glycolysis
8. Essential fructosuria
9. BPG shunt
10. Congenital galactosaemia
11. Renal glycosuria
12. Distinctive features of *glucokinase* and *hexokinase*
13. Metabolism of amino sugars
14. Glyoxylate pathway

CLINICAL CASES

CASE 9.1 A 48-year-old man in a comatose state

A 48-year-old moderately obese man was brought to the medical emergency in an unconscious state. He was breathing deeply and rapidly. The breath did not smell of acetone or alcohol. Signs of mild dehydration, such as dry tongue, soft eyeballs, weak and rapid pulse and low blood pressure were present.

Examination of the past medical record showed that the patient was diagnosed as having type 2 diabetes mellitus (NIDDM) six years back. Initial treatment with sulphonylureas proved effective in controlling blood glucose levels, and the patient apparently remained in good health all these years. However, about four months back, elevated blood glucose levels were detected on two separate occasions (210 mg/dL and 236 mg/dL). Increasing the dose of sulphonylureas did not help. Biguanide therapy was initiated, after the renal and the hepatic functions were found normal. Adequate control of diabetes was achieved with metformin in three daily doses of 1.0 g each.

For the last two days, the patient had hectic sessions of physical activity when he went to a hill station for recreation. He had rich meals during this period, often accompanied by drinking till late night. While driving back home early in the morning, the family members noticed that he was disoriented and his speech was incoherent. Soon afterwards, he lost consciousness and was rushed to the nearest hospital.

Blood and urine samples were obtained for analysis.

Investigations test	Patient's reports	Reference range
Venous plasma glucose (random)	85 mg/dL	< 140 mg/dL
pH	7.32	7.35–7.45
Plasma bicarbonate	20 mmol/L	21–28 mmol/L
pCO_2	32 mmHg	36–45 mmHg
Lactic acid	9.8 mmol/L	0.44–1.4 mmol/L

Urine examination: Normal: sugar, proteins or ketone bodies absent.

Q.1. What is the biochemical abnormality in this patient?
Q.2. Diagnose the clinical condition that might have led to the above biochemical abnormality?
Q.3. What is the role of the following in causing the patient's problems: (a) severe muscular exercise, (b) acute alcohol intoxication.

CASE 9.2 An unconscious 5-year old child

A 5-year-old child was brought to the medical OPD in a comatose state. He had felt headache and dizziness only a few hours before. His father noticed profuse sweating and some abnormal behavior at that time and rushed him to the hospital. A similar episode had occurred two months earlier after the child had a glass of sugarcane juice. However, he had not lost consciousness on that occasion; his mother had made him drink a glass of heavily sugared milk thinking that the child was feeling weak. This seemed to have alleviated the symptoms.

Presently, physical examination showed tachycardia and rapid breathing. Liver was markedly enlarged, being palpable 4 cm below costal margin. Blood and urine samples were obtained for biochemical analysis. Blood glucose level was very low (52 mg/dl). No other abnormality was found in the test result. The child was given intravenous dextrose to treat hypoglycaemia. He responded well and the hypoglycaemic symptoms promptly disappeared. Two days later, liver biopsy was performed and the sample was analyzed. It revealed large deposits of fructose 1-phosphate within the hepatocytes.

Q.1. What is the probable biochemical defect in this child? Suggest a biochemical test to evaluate the above diagnosis.
Q.2. Provide a biochemical explanation for the child's signs and symptoms.
Q.3. Would you expect liver function to be normal or sluggish in this child? Give reason.
Q.4. Suggest treatment for this condition.

METABOLISM OF CARBOHYDRATES II: SECONDARY PATHWAYS AND REGULATION OF BLOOD GLUCOSE LEVEL

CHAPTER

10

The mainline metabolic pathways of glucose utilization generate ATP energy and fuel molecules. In addition to such pathways, there are other minor (or secondary) pathways taken by glucose, which are specialized for synthesis of glycolipids, glycoproteins and other special products needed by the cells. Two such pathways are **pentose phosphate pathway**, which leads to formation of ribose 5-phosphate and NADPH; and **uronic acid pathway**, which generates D-glucuronate and ascorbic acid. Rate of flux of metabolic intermediates through the secondary pathways is much less compared to that through the mainline pathways.

Significance of certain other minor metabolic pathways that metabolize sugars other than glucose, such as fructose, galactose, mannose, or the disaccharides and the polysaccharides is also outlined this chapter. Maintenance of blood glucose level within the normal range has been discussed.

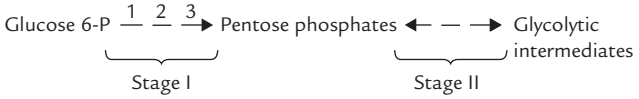
After going through this chapter, the student should be able to understand:

- Reactions of pentose phosphate pathway and uronic acid pathway, and their significance.
- Minor pathways of metabolism of carbohydrates other than glucose.
- Metabolic significance and clinical consequences of enhanced rate of sorbitol pathway in diabetes mellitus.
- The interplay of various factors that are responsible for maintenance of normal blood glucose levels.

I. Pentose Phosphate Pathway

The pentose phosphate pathway also called **hexose monophosphate (HMP) shunt** or **phosphogluconate pathway** is required for provision of five-carbon sugars and NADPH. Activity of this cytosolic pathway is minimal in muscle and brain, where almost all of the glucose is degraded by glycolysis. However, it is *highly active in liver, adipose tissue, adrenal cortex and lactating (but not the non-lactating) mammary gland.* These tissues are involved in synthesis of cholesterol and fatty acids, the processes dependent on NADPH. Because NADPH is important for the antioxidant defenses of the body, the pentose phosphate pathway is well developed in cells that are exposed to a high oxygen partial pressure, such as erythrocytes and cornea of the eye. It accounts for about 60% of the total oxygen consumption in cornea.

The pathway consists of three irreversible oxidative reactions (**Stage I**), followed by a series of reversible sugar-phosphate interconversions (i.e. non-oxidative reactions **Stage II**). An overview is presented below.



The **Stage I**, irreversible oxidative reactions result in formation pentose phosphates, while a series of reversible non-oxidative interconversions take place in the **Stage II**, which convert these pentose phosphates into glycolytic intermediates.

A. Specialized Products Generated by Pentose Phosphate Pathway

1. *Reduced nicotinamide adenine dinucleotide phosphate (NADPH)* serves as a biochemical reductant in a number of reactions. Such reductive reactions form important steps of several biosynthetic pathways, such as **steroidogenesis** and **lipogenesis**. Therefore, rate of the pentose phosphate pathway is high in tissues that synthesize steroids and fatty acids.

   NADPH generation through HMP shunt is essential for maintaining integrity of the erythrocyte membranes because glutathione reduction is brought about by NADPH (**Case 10.1**). These functions are discussed in detail later in this chapter.
2. *Ribose 5-phosphate* is an important constituent of the purine and pyrimidine nucleotides, that perform a variety of important functions. It is an essential structural component of various coenzymes such as coenzyme-A, FAD and NAD^+. Ribose 5-phosphate can be reconverted to hexose phosphates through reversible sugar-phosphate interconversions by the non-oxidative reactions of the pentose phosphate pathway. This provides an important mechanism for the metabolism of five carbon sugars.

B. Reactions of Pentose Phosphate Pathway

Stage I: Oxidative Phase

The oxidative reactions of Stage I bring about conversion of glucose 6-phosphate to ribulose 5-phosphate (Fig. 10.1). Two NADPH molecules are also generated during this reaction sequence comprising of three irreversible oxidative reactions. The reactions are shown in Figure 10.2.

Dehydrogenation of glucose 6-phosphate

Glucose 6-phosphate is irreversibly converted to 6-phosphogluconolactone by the enzyme *glucose 6-phosphate dehydrogenase*. $NADP^+$ as a co-enzyme is specific for this reaction. $NADP^+$ acts as an acceptor of the reducing equivalents (i.e. a pair of hydride) that are removed from the glucose 6-phosphate molecule and gets converted to NADPH.

This step is the ***regulatory step*** of the pathway. NADPH acts as a negative modulator, causing competitive inhibition of *glucose 6-phosphate dehydrogenase* (G_6PD).

Hydrolysis of 6-phosphopgluconolactone

6-Phosphogluconolactone is converted to 6-phosphogluconate by addition of a water molecule. This reaction is catalyzed by the enzyme *gluconolactonase* and is irreversible, like the previous step.

Formation of ribulose 5-phosphate

Oxidative decarboxylation of 6-phosphogluconate, catalyzed by *6-phosphogluconate dehydrogenase*, then yields the ketose sugar, ribulose 5-phosphate, plus a molecule of NADPH.

The non-equilibrium nature of the Stage I reactions enables the cell to maintain a cytoplasmic ratio of **NADPH : $NADP^+$ = 100**. Interestingly, the ratio of NADH : NAD^+ in the cytoplasm is nearly the inverse, less than 0.01. For this reason the *cells employ NADPH rather than NADH whenever a strong reducing agent is required*.

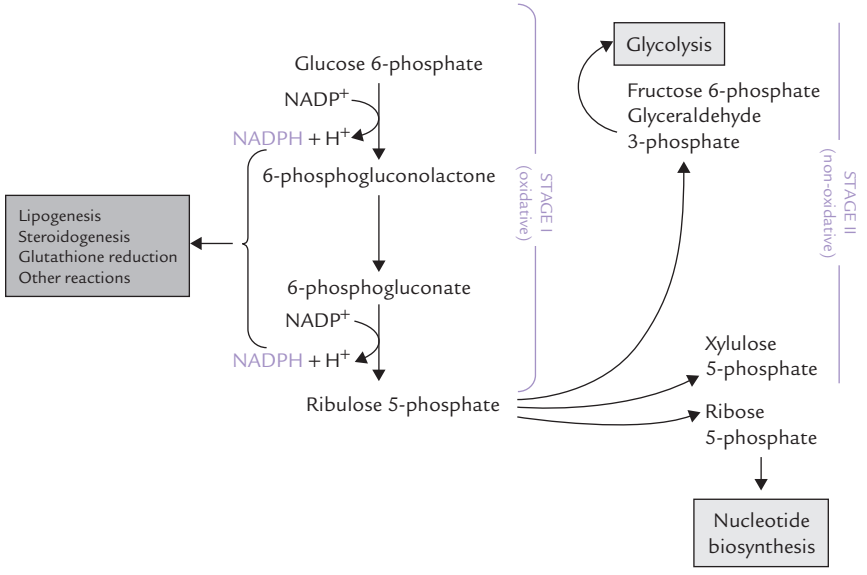


Fig. 10.1. Overview of the pentose phosphate pathway.

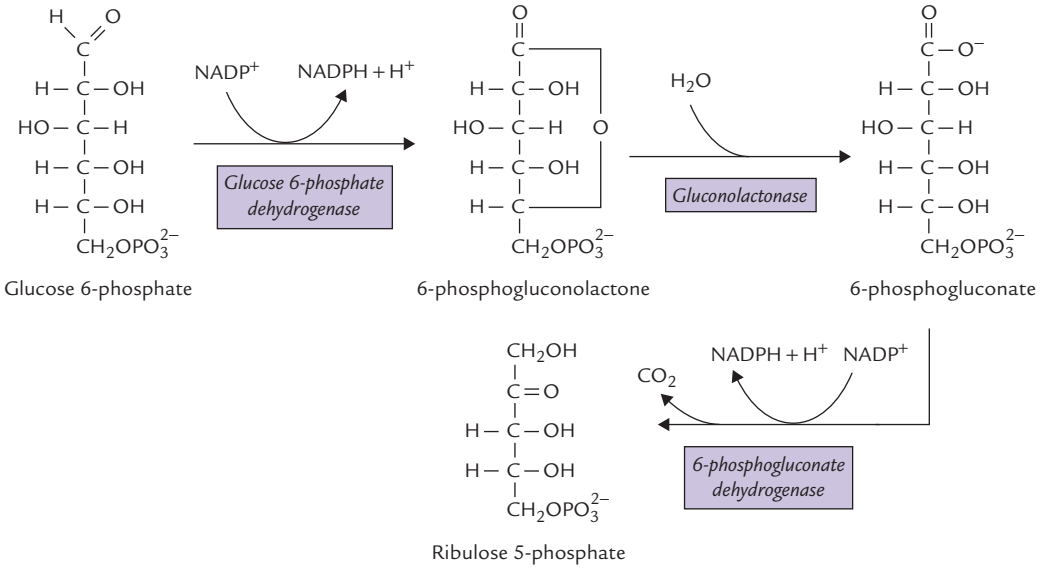
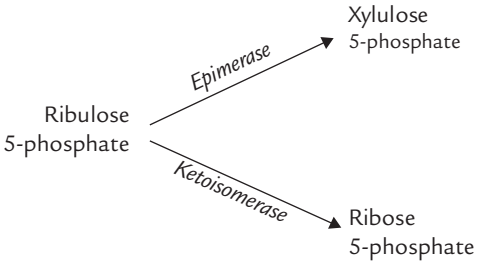


Fig. 10.2. The oxidative reactions of pentose phosphate pathway.

The ribulose 5-phosphate, the end product of the Stage I, is convertible to other pentose phosphates: xylulose 5-phosphate and ribose 5-phosphate. The enzymes for these reactions are:

1. *Epimerase*, which converts ribulose 5-phosphate to xylulose 5-phosphate.
2. *Ketoisomerase*, which catalyzes conversion of ribulose 5-phosphate to ribose 5-phosphate.



Stage II: Non-oxidative Interconversion Phase

The pentose phosphates generated in Stage I, ribose 5-phosphate, xylulose 5-phosphate and ribulose 5-phosphate, are converted to glycolytic intermediates in the second stage of HMP (Fig. 10.3). This stage comprises a series of non-oxidative reversible interconversions, as noted earlier. The enzymes of this stage are *transketolase* and **transaldose**, which catalyze the following three reactions:

Reaction 1

This reaction is catalyzed by *transketolase*, which transfers a two carbon unit from a ketose sugar to an aldose sugar. When the two-carbon unit is transferred from a xylulose 5-phosphate (a 5-C ketose) to ribose 5-phosphate (a 5-C aldose), formation of a seven-carbon sugar (sedoheptulose 7-phosphate) and a three carbon-sugar (glyceraldehyde 3-phosphate) results.

$$C_5 + C_5 \rightleftharpoons C_7 + C_3$$

Thiamine pyrophosphate, a coenzyme for *transketolase*, serves as transient carrier of the two carbon units in this reaction.

Reaction 2

This is catalyzed by ***transaldolase***, which transfers a three-carbon unit from the sedoheptulose 7-phosphate to glyceraldehyde 3-phosphate. This results in the formation of a erythrose 4-phosphate and fructose 6-phosphate.

$$C_7 + C_3 \rightleftharpoons C_4 + C_6$$

Reaction 3

This is again catalyzed by ***transketolase***. The tetrose phosphate (erythrose 4-phosphate) generated in the previous step, accepts a two-carbon unit from a new pentose phosphate (xylulose 5-phosphate) molecule to form a fructose 6-phosphate and glyceraldehyde 3-phosphate.

$$C_4 + C_5 \rightleftharpoons C_6 + C_3$$

Net result of these reactions is formation of two hexose phosphates (fructose 6-phosphate) and a triose phosphate (glyceraldehyde 3-phosphate), which directly enter the glycolytic sequence.

Thus, the non-oxidative interconversions (of the Stage II) permit conversion of the pentoses (generated in Stage I) into intermediates of glycolysis. This establishes a link between the pentoses and the other metabolizable

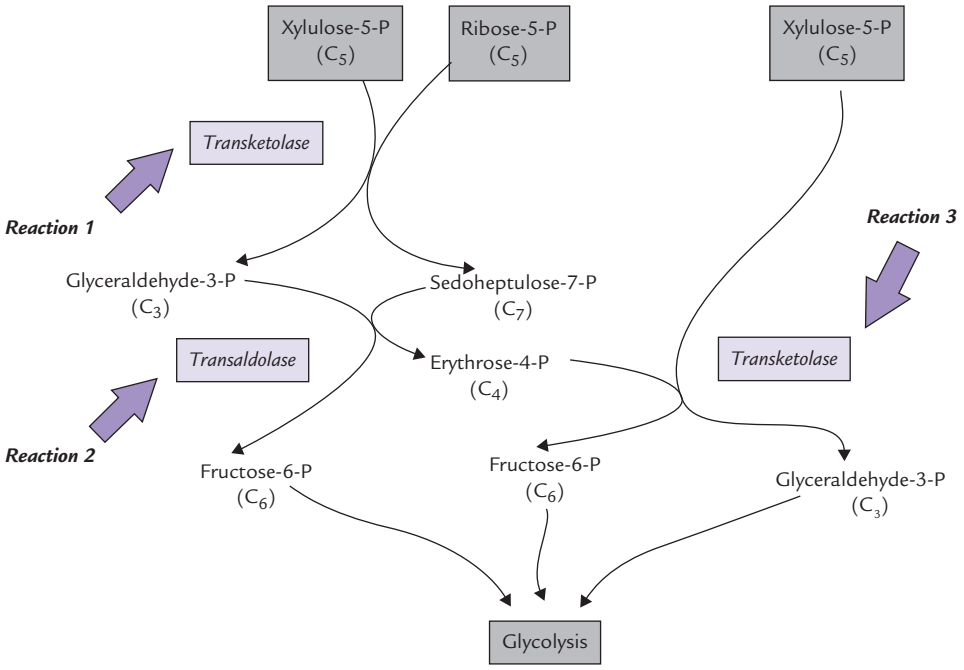


Fig. 10.3. The non-oxidative reactions of pentose phosphate pathway. The *transketolase* transfers a 2-carbon unit, and the *transaldolase* causes transfer of a 3-carbon unit.

$$C_5 + C_5 \rightleftharpoons C_7 + C_3 \quad \textbf{(Reaction 1; } \textit{Transketolase}\textbf{)}$$
$$C_7 + C_3 \rightleftharpoons C_6 + C_4 \quad \textbf{(Reaction 2; } \textit{Transketolase}\textbf{)}$$
$$C_5 + C_4 \rightleftharpoons C_6 + C_3 \quad \textbf{(Reaction 3; } \textit{Transketolase}\textbf{)}$$
$$\textbf{Sum: } C_5 + C_5 + C_5 \rightleftharpoons C_6 + C_6 + C_3$$

sugars, thereby integrating the HMP shunt pathway with glycolysis.

A *summary of reactions* of the pentose phosphate pathway, as shown in Figures 10.2 and 10.3 can be written as:

$$3\ \text{Glucose 6-phosphate} + 6NADP^+ \longrightarrow$$
$$2\ \text{Fructose 6-phosphate} + \text{Glyceraldehyde 3-phosphate} + 6NADPH + 6H^+ + 3CO_2$$

The 1st phase of HMP shunt is the oxidative pathway in which glucose 6-phosphate is oxidized and decarboxylated to produce two NADPH, a carbon dioxide and a ribulose 5-phosphate. The interconversion phase consists of reversible rearrangements of the phosphorylated pentoses. It links ribulose 5-phosphte to the glycolytic pathway.

C. Regulation

The regulatory enzymes of the pentose phosphate pathway are *glucose 6-phosphate dehydrogenase* and *6-phosphogluconate dehydrogenase*. Synthesis of both the enzymes is induced by **insulin**. Thus, in fed state, their intracellular concentrations rise, leading to enhanced oxidation of glucose through this pathway, and conversely in starvation and diabetes mellitus, the pathway is suppressed.

Activity of *glucose 6-phosphate dehydrogenase is competitively inhibited by NADPH*, as mentioned earlier. It is the ratio of NADPH : $NADP^+$ that determines the overall rate of the pathway. Under normal circumstances, the ratio of $NADPH/NADP^+$ is sufficiently high (up to 100), which keeps the activity of G_6PD inhibited. However, when demand for NADPH increases, its utilization also increases resulting in the fall of the intracellular NADPH concentration. This removes the inhibition on the enzyme activity. Consequently, this reaction is speeded up, resulting in enhanced production of the required NADPH.

D. Functions

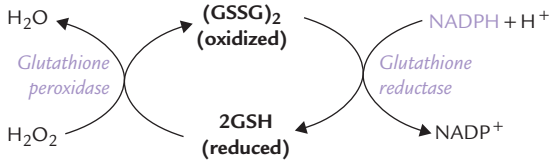
The pentose phosphate pathway is a multifunctional pathway, unique in generation of two important products: **pentoses** and **NADPH**.

- **Pentoses:** The most important pentose is ribose 5-phosphate, which is required for a number of biosynthetic and other functions, as mentioned earlier in this chapter.
- **NADPH:** It has the same redox potential as NADH, but the functions of the two coenzymes are different. Whereas NAD^+ collects hydrogen from catabolic substrates for transfer to respiratory chain, NADPH is

required for reductive biosynthesis and several other functions:

Reductive biosynthesis: The synthesis of fatty acids and cholesterol from acetyl CoA require reductive power of NADPH.

Antioxidant role: NADPH is necessary for the maintenance of a reducing environment inside the cell. This is necessary for avoiding damage to unsaturated fatty acids, proteins and DNA by peroxides or other oxidizing agents, to which the cell is continuously exposed. NADPH plays a key role in antioxidant defenses by converting the oxidized glutathione into the reduced glutathione, which is protective:



As diagrammed above, the reduced glutathione (GSH) detoxifies hydrogen peroxide; the enzyme *glutathione peroxidase* catalyzes this reaction. However, GSH is oxidized during the reaction, and must be regenerated to continue detoxification. NADPH participates in the reaction responsible for the regeneration of reduced glutathione from the oxidized one; *glutathione reductase* catalyzes this reaction.

NADPH maintains the glutathione in reduced state by converting the dimeric, oxidized glutathione $(GSSG)_2$ into the protective reduced form (GSH). This recycling of GSSG to GSH is crucial for maintaining antioxidant defense.

Phagocytosis (literally "cell eating"): It is the engulfment of solid particles into the cell, such as granulocytes, monocytes and macrophages. These cells participate in our immune system by eating up and destroying infectious microorganisms (Chapter 33). NADPH is required for production of superoxide anion radicals by macrophages, which have bactericidal activity (**Case 27.1**).

Metabolism of xenobiotics: Lipophilic xenobiotics are metabolized to water-soluble products by hydroxylation in the microsomal cytochrome P_{450} system; the process requires NADPH (Chapter 15).

Special function in erythrocytes: NADPH is especially important to prevent oxidative damage to the RBC membrane and intracellular proteins in erythrocytes because the cell cannot replace these by new synthesis during its 120-day lifespan. NADPH maintains the antioxidant defenses by converting the oxidized, dimeric glutathione into the protective reduced form. This restores the level of reduced glutathione, which is oxidized while detoxifying peroxides and other oxidizing agents. Thus, *requirement of NADPH is relatively higher in erythrocytes*. The extra demand is met by a slight modification in the HMP-shunt, as follows.

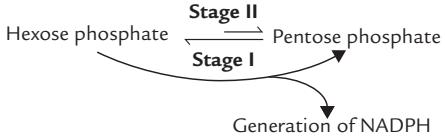
HMP-shunt can be Modified

Based on demand of the tissue, the HMP shunt is modified as below:

(a) Extra demand for NADPH: In erythrocytes requirement for NADPH is relatively more, which is met as below:

- Conversion of the pentoses to hexoses through the reversible Stage II reactions is favoured.
- The hexose phosphates so formed are then funneled into another cycle of oxidative reactions to produce more of NADPH.

This establishes a cyclic pathway, with each of these cycles generating 2 NADPH molecules.



(b) Extra demand for pentoses: Demand for pentoses is greater than that for NADPH in some tissues. The non-oxidative reactions can provide the ribose 5-phosphate molecules directly from fructose 6-phosphate via the Stage II reactions in such tissues.

Hexose phosphate $\rightleftharpoons$ Pentose phosphate

Depending on cellular needs, ribulose 5-phosphate is converted (via ribose 5-phosphate and xylulose 5-phosphate) to hexose phosphates, which can re-enter into another cycle of oxidative reactions to produce NADPH. Conversely, hexose to pentose conversion is favoured in tissues having more demand for pentose.

Glucose 6-phosphate Dehydrogenase Deficiency

Some persons have a genetic defect in this enzyme, typically yielding an unstable enzyme, that has shorter half-life in the RBC, or an enzyme that is unusually sensitive to inhibition by NADPH. In either case, insufficient flux of the HMP-shunt and so decreased production of NADPH results, and the *cell's ability to recycle GSSG to GSH is impaired*. Drug-induced oxidative stress leads to lysis of RBCs (haemolysis); *haemolytic anaemia* is the obvious consequence (**Case 10.1**).

Glucose 6-phosphate dehydrogenase deficiency causes drug induced haemolytic anaemia because of lack of NADPH, which maintains glutathione in its reduced state.

Wernicke-Korsakoff Syndrome

Thiamine deficiency in **alcoholics** does not cause beriberi but Wernicke-Korsakoff syndrome. Reduced activity of the thiamine-dependent *transketolase* enzyme occurs in this condition resulting is impairment of HMP shunt.

This condition is diagnosed by clinical features that include mental derangements, delirium and motor incoordination, which may progress to chronic stage (called *Korsakoff psychosis*) characterized by amnestic syndrome. Early diagnosis and immediate treatment is essential, because brain damage in this condition is irreversible.

II. Uronic Acid Pathway

Uronic acid pathway is a source of *glucuronic acid*. It also produces ascorbic acid in some animals. Glucuronic acid is a useful product required in the synthesis of mucopolysaccharides, detoxification of some drugs and for conjugation of bilirubin and steroid hormones. The unutilized glucuronic acid (glucuronate) is converted to xylulose 5-phosphate, which is metabolized via the pentose phosphate pathway.

A. Reactions of the Pathway

Glucose is first converted to UDP-glucose by a series of reactions, as discussed earlier in glycogenesis. The UDP glucose further proceeds as below:

1. The glucose portion of the UDP-glucose is enzymatically oxidized, resulting in formation of UDP-glucuronate. The reaction is catalyzed by the enzyme *UDP-glucose dehydrogenase*. UDP-glucuronate is the metabolically active compound, as discussed later.
2. From the UDP-glucuronate release of D-glucuronate occurs (Fig. 10.4).
3. D-Glucuronate is reduced by the NADPH-dependent enzyme *glucuronate reductase* to form L-gulonate. Glucuronate contains an acid carboxylate group at C-6, whereas gulonate contains this group at C-1.
4. L-Gulonate then loses a water molecule to form L-gulonolactone. The enzyme is *aldonolactonase*.
5. Removal of a pair of hydrogen atoms from the L-gulonolactone by the enzyme *gulonolactone oxidase* yields L-ascorbate or vitamin C.

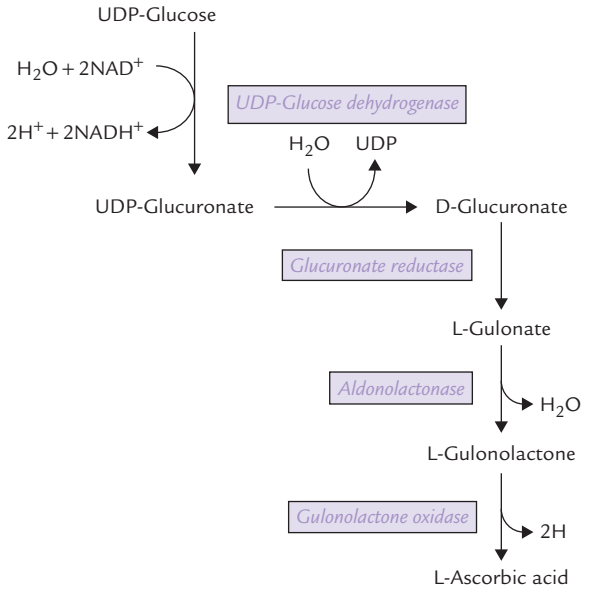


Fig. 10.4. Reactions of uronic acid pathway.

This pathway is used by all plants and animals for the **synthesis of vitamin C**. However, some primates, such as humans, monkeys, guinea pigs, as well as some birds and fishes lost the capability to do so during evolution. This is because of genetic deficiency of the enzyme *gulonolactone oxidase*, which causes conversion of L-gulonolactone to L-ascorbic acid. It appears that the capacity to synthesize the ascorbic acid was lost by these species because of a mutation, which was not lethal. These species require vitamin C in diet.

A single enzyme defect in the uronic acid pathway (deficiency of *gulonolactone oxidase*) causes inability of primates to synthesize ascorbic acid.

B. UDP-Glucuronate is Metabolically Active Compound

UDP-Glucuronate can readily denote the glucuronate residue for the following functions:

1. *To detoxify foreign compounds or drugs.* During detoxification, the glucuronate residues are covalently attached to these substances. Since glucuronate residues are strongly polar, their attachment imparts polar character to these compounds, thus making possible their renal excretion. Bilirubin and steroid hormones are also rendered more polar for excretion in this manner.

Uronic acid pathway, a minor pathway of carbohydrate metabolism, is a source of glucuronic acid for conjugation of several endogenous and exogenous substances before excretion as glucuronides in urine and bile.

2. *To synthesize the acid mucopolysaccharides,* such as hyaluronic acid and heparin, which contain glucuronic acid as an essential component (Chapter 2).

C. Alternate Route for Oxidation of L-Gulonate

- The L-gulonate (produced from unutilized D-glucuronate) is converted to a pentose sugar, L-xylulose, by oxidative decarboxylation.
- L-xylulose is then converted to its D-isomer, D-xylulose (via an intermediate xylitol).
- D-xylulose is then phosphorylated at the expense of an ATP. The xylulose 5-phosphate so formed is an intermediate of the pentose phosphate pathway (Fig. 10.5).

Thus, *the uronic acid pathway is connected with the pentose phosphate pathway.*

Essential Pentosuria

In the rare hereditary disease termed "*essential pentosuria*", the enzyme that catalyzes reduction of L-xylulose to xylitol is deficient. As a result, excretion of significant amounts of L-xylulose in the urine occurs.

Effect of Drugs

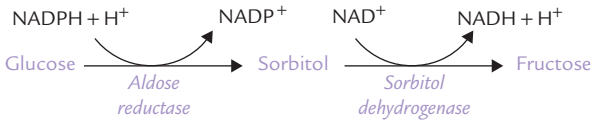
Certain drugs increase the flux of intermediates through the uronic acid pathway and thereby enhance generation of the products of the pathway. In rats, synthesis of ascorbic acid is enhanced by these drugs. Barbitol, aminoantipyrine and chlorobutanol are some examples. These drugs increase excretion of L-xylulose in the pentosuric patients.

III. Metabolism of Other Sugars

The metabolic pathways of other monosaccharides, disaccharides and polysaccharides, referred to as the feeder pathway, have been described in the previous chapter. Though these pathways appear quantitatively insignificant, they play an important supplementary role.

IV. Sorbitol Pathway

Sorbitol pathway also known as **polyol pathway**, causes conversion of glucose to fructose via sorbitol. *The pathway is greatly enhanced in uncontrolled diabetes mellitus, and is implicated in progression of chronic diabetic complications.* It comprises of two sequential reactions, as shown below.



The first step involves conversion of glucose to sorbitol by the enzyme ***aldose reductase***. This enzyme is present in a significant concentration in those cell types that do not require action of insulin for the transport of glucose across the cell membrane (e.g. insulin unresponsive cells). Some examples are epithelial cells of lens, Schwann cells of peripheral nerves, seminal vesicle cells and papillae of kidneys. The second enzyme of the sorbitol pathway, i.e. ***sorbitol dehydrogenase***, is present in liver and seminal vesicle cells. It can oxidize sorbitol to fructose. Fructose is preferred over glucose for supplying energy in sperm cells because of the presence of this enzyme.

Sorbitol Pathway in Diabetes Mellitus

Normally, flux of metabolic intermediates through the sorbitol pathway is not significant because *aldose reductase* has a high K_m (i.e. low affinity) for glucose. Therefore, glucose is taken up by this pathway in small amount, the rest being rapidly phosphorylated to glucose 6-phosphate. However, when blood glucose concentration is markedly elevated, such as in **uncontrolled diabetes mellitus**, large quantities of glucose enters the insulin-unresponsive cells. Large concentration of glucose substrate overcomes the effect of low affinity (of *aldose reductase* for glucose) and pushes the equilibrium of the first reaction towards sorbitol formation. Consequently, **large quantities of sorbitol is produced in prolonged hyperglycaemia**.

Sorbitol cannot be further metabolized in lens and nerve cells (because *sorbitol dehydrogenase* has either low activity or is absent in these tissues) and cannot diffuse

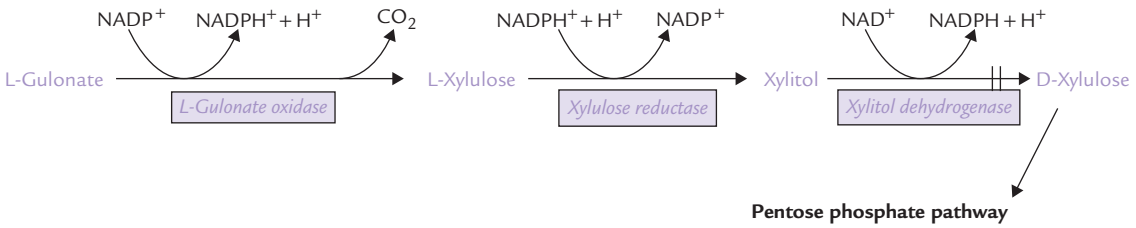


Fig. 10.5. Oxidation of L-gluonate (||= block in essential pentosuria).

out of the cell easily. Therefore, *concentration of sorbitol rapidly builds up, exerting an osmotic effect, which is responsible for many of the physiologic and pathologic alterations associated with diabetes*. These alterations lead to development of **cataract**, **peripheral neuropathy**, **retinopathy**, and **vascular problems** (see Chapter 15).

Glucose can be converted to fructose via sorbitol pathway. The pathway is enhanced in prolonged hyperglycaemia (diabetes mellitus) so that sorbitol accumulates leading to chronic diabetic complication (cataract, nephropathy, peripheral neuropathy, etc.).

Sorbitol Intolerance

Sorbitol, used as a sugar-free sweetener, is converted to fructose in liver by *sorbitol dehydrogenase*. Abdominal discomfort occurs in some individuals with consumption of sorbitol, probably due to impaired metabolism.

V. Regulation of Blood Glucose Level

Normally the blood glucose level is kept within the range of **60–100 mg/dL** in **fasting state**. The fasting state here implies that the there has been no food intake for the last *12–16 hours*. After a meal, glucose is absorbed from intestines and blood glucose level starts rising, reaching a maximum in about one hour. *In a normal individual, the maximum value remains less than 1.5 times the fasting value.*

Thus, the blood glucose level does not rise above **140 mg/dL**, or fall below about 60 mg/dL. This limit is not crossed in a normal subject even though he may take food at certain fixed times and fast between the meals. Several mechanisms of blood glucose regulation remain constantly in operation for this purpose. It is hazardous for the body if above-stated limits are crossed. For example:

- A fall below about 60 mg/dL will deprive the nervous tissue of adequate energy, because under normal circumstances, nerve tissue is heavily dependent on glucose for its energy needs (**Case 10.2**).
- Conversely, a rise above 180 mg/dL (called **renal threshold**) will result in urinary elimination of glucose since the amount of glucose filtered by the glomeruli will exceed the amount which can be reabsorbed by the tubules.

Concentration of blood glucose is determined by a balance between the amount of glucose that pours into the blood circulation and the amount that leaves it. Various sources of glucose generation and utilization are shown in Figure 10.6.

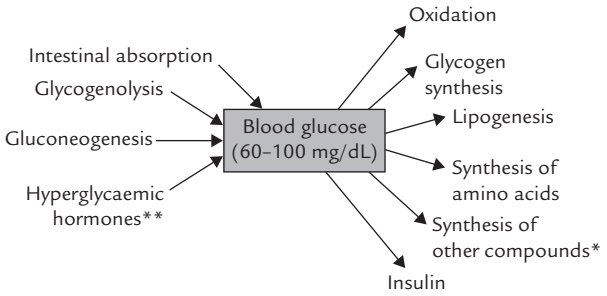


Fig. 10.6. Sources of glucose generation and utilization. (*These include lactose, ribose, mucopolysaccharides, glycolipids, glycoproteins. They are synthesized from glucose or its metabolites. **Hyperglycaemic hormones are glucagon, growth hormone, thyroxine, epinephrine, glucocorticoids, etc.).

A regulated interplay of these factors ensures blood glucose regulation. **Liver** and **endocrine glands** play a vital role in this.

A. Role of Liver

Liver is capable of utilizing large amounts of glucose that is delivered to it at high concentration (in portal blood) during and following meals. This is because:

High permeability of hepatocytes to glucose: Liver cells are highly permeable to glucose because high-capacity, low-affinity ($K_m > 10$ mmol/L) glucose transporter, **GLUT-2**, are located on surface of hepatocytes.

Rapid utilization of the internalized glucose: Liver is also rich in ***glucokinase***, an enzyme that is specific for glucose and converts it into glucose 6-phosphate. As discussed earlier, *glucokinase is* inducible by continued consumption of high-carbohydrate diet, not subject to product inhibition. These properties permit *glucokinase* to efficiently phosphorylate the intracellular glucose and force it into all of the major pathways of glucose metabolism.

Most of the glucose is channeled into glycogen, providing a carbohydrate reserve for maintenance of blood glucose during the post-absorptive state. The glucose 6-phosphate in liver, beyond that needed to replenish glycogen reserves, is funneled into glycolysis, both for energy production and for conversion into triacylglycerols, which are exported for storage in adipose tissue.

B. Role of Endocrines

Several hormones participate in the regulation of blood glucose levels. All except insulin tend to raise blood glucose level, hence termed hyperglycaemic or **diabetogenic**

hormones (Fig. 10.6). Insulin is the only hormone that brings down the blood glucose level.

Role of Glucagon

It is a peptide hormone secreted by the α-cells in endocrine pancreas that is released in response to a decreased blood glucose level. It causes rapid stimulation of glucose-producing pathways (including hepatic glycogenolysis and gluconeogenesis), and depresses glycogen synthesis and glycolysis, thereby contributing to the hyperglycaemic effect during fasting.

Role of Glucocorticoids

Glucocorticoids cause stimulation of gluconeogenesis during **chronic stress** by causing induction of the key enzymes, especially *PEP carboxykinase*. Other effects are: induction of *fructose 1,6 bisphosphatase, glucose 6-phosphatase* and *amino transferases;* stimulation of release of amino acids and lactate from muscles, which act as precursors for gluconeogenesis. However, overall effect of glucocorticoids is relatively slow.

Role of Catecholamines

The catecholamines – epinephrine and norepinephrine, are released during **acute stress** (physical exercise and cold exposure) and psychological emergencies from adrenal medulla. They induce slight elevation of cAMP level to increase glycogenolysis in liver and muscles. Liver contributes to increased level of blood glucose, whereas muscles generate lactate for gluconeogenesis. Other effects are:

1. Increased gluconeogenesis (mainly through stimulation of *fructose 1,6-bisphosphatase*).
2. Increased release of glycerol by lipolysis, which is substrate for gluconeogenesis.
3. Decreased activates of key glycolytic enzymes, e.g. *PFK, glucokinase* and *pyruvate kinase*.

Role of Growth Hormone

It inhibits glycolysis (inhibits PFK) and mobilizes fatty acids from adipose tissue stores.

Role of Insulin

It is a peptide from pancreatic β-cells, released in response to elevation of blood glucose level. Its plasma concentration is 10 times higher after a carbohydrate rich meal than during fasting. It reduces the cAMP level in the liver, probably by activating a cAMP degrading *phosphodiesterase*, which means that it opposes the effects of glucagon and catecholamines. It stimulates several glucose-utilizing pathways, and inhibits the glucose-generating pathways. Details are given in Chapter 15 (see Table 15.2).

Regulation of blood glucose levels requires several hormones and a concerted interplay between various organs.

An interplay of various factors maintains blood glucose in different metabolic states: for example, in **fasting state**, glucose does not enter blood circulation through intestinal absorption, but some amount leaves it to enter tissues for utilization. The blood glucose level, therefore, tends to fall. An excessive fall is prevented by the following metabolic changes in tissues:

1. Instead of glucose, tissues start using alternate fuels, such as fatty acids and ketone bodies. Glucose is spared for utilization by erythrocytes and brain (Chapter 15).
2. Glycogen synthesis in the liver and muscle stops, whereas glycogen breakdown is stimulated. These adaptations are brought about by increased secretion of glucogen and catecholamines, and decreased secretion of insulin. The glucose produced by way of hepatic glycogenolysis is poured into the blood circulation.
3. Formation of glucose in liver occurs by way of stimulation of gluconeogenesis also. For this purpose, the protein degradation in muscles is enhanced and the amino acids so released are transported to liver where they serve as substrates for gluconeogenesis. Glucocorticoids play a key role in this process.
4. In the adipose tissue, lipolysis is stimulated, resulting in excessive generation of fatty acids.

Thus, adipose tissue becomes a source of alternate fuel molecules (e.g. free fatty acids).

These factors ensure that the blood glucose does not fall steeply during fasting and starvation. Likewise, in the **fed state**, metabolic changes occur to prevent excessive rise of blood glucose concentration. Failure of the regulatory mechanisms occurs in some pathological conditions, such as diabetes mellitus, hyperpituitarism, hyperthyroidism, Cushing's syndrome, severe liver diseases, etc. Further details about these are given in Chapter 30.

Exercises

Essay type questions

1. Describe the pentose phosphate pathway of glucose metabolism, giving names of enzymes and coenzymes involved. What are the functions of this pathway?
2. Explain the relationship between the integrity of erythrocyte membrane to the operation of HMP shunt pathway. How is the demand for the extra NADPH in red cells met?

3. What is glutathione and what role does it play in the human body? Why does it fail to prevent the haemolytic anaemia in deficiency of *glucose 6-phosphate dehydrogenase*?
4. Describe the role of various factors involved in maintenance of normal blood glucose level.

Write short notes on

1. Sorbitol pathway
2. Glutathione
3. Uronic acid pathway
4. Sorbital intolerance
5. Wernicke-Korsakoff syndrome

CLINICAL CASES

CASE 10.1 A 23-year-old man with dark urine and yellow sclera following antimalarial treatment

A 23-year-old male developed fever about two weeks back. He had bouts of shivering, temperature of 40.4°C and was delirious. The family physician suspected malaria and started treatment with primaquine after identification of the parasites in a blood smear. The fever subsided the next day, but the patient continued to feel weak and listless. By the next day these symptoms aggravated and he felt fatigue, dizziness, breathlessness on slightest exertion, headache, insomnia and paresthesia of the fingers and toes. Three days later, the patient noticed dark (black) coloured urine. On examination, he showed pallor of the skin and the mucus membrane. Oedema feet, jaundice, tachycardia (heart rate 110/min), and systolic murmurs were the other prominent examination findings. His sclera was yellow and the spleen was marginally enlarged.

Investigations test	Patient's reports	Reference range
Haemoglobin	10.2 g%	11–14 g%
Reticulocyte count	6.3%	Up to 2%
Serum bilirubin (total)	8.3 mg/dl	0.1–1 mg/dl
Urine bile pigments	Absent	

The red cells, on microscopic examination, were found to contain small inclusion bodies (Heinz bodies).

Since the above tests were suggestive of acute haemolytic crisis, further tests were done, including estimation of the *RBC* enzymes. Activity of *glucose 6-phosphate dehydrogenase* was found deficient, i.e. less than 10% of the normal.

Q.1. State the biochemical basis of development of haemolytic disease in this patient.
Q.2. What is glutathione and what is its role in body? How does glutathione fail to perform its function following administration of primaquine in this patient?
Q.3. Does the disorder affect other tissues of the body as well? Explain.
Q.4. Comment on the following observations:
(i) The dark coloured urine
(ii) Presence of Heinz bodies in red cells.
Q.5. Does the disorder in this patient offer any advantage as well? If yes, explain how.

CASE 10.2 A newborn having tremors and convulsions

A premature infant of 2.3 kg body weight, born of diabetic mother, was shifted to nursery after the attending paediatrician noticed muscle twitching, tremors and profuse sweating. Soon after, the infant had convulsions and lapsed into comatose state. Blood and urine samples were obtained for biochemical analysis.

Investigations

Venous plasma glucose	38 mg/dl
Urine examination	Normal; no sugar or ketone bodies detected

Q.1. What is the probable diagnosis?
Q.2. State the cause of loss of consciousness in hypoglycaemia?
Q.3. Could the diabetes in mother be responsible for the infant's condition? Give reason for your answer.
Q.4. Explain why infants are more susceptible to hypoglycaemia than adults.

LIPID METABOLISM I: METABOLISM OF FATTY ACIDS AND COMPOUND LIPIDS

CHAPTER

11

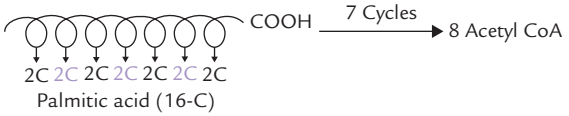
Fatty acids are major sources of energy in humans. They are used as fuel by most tissues except red and white blood cells, nervous tissue, retina and adrenal medulla. They are stored as triacylglycerols (TAGs) in adipose tissue. When the need for energy arises, the fatty acids are mobilized from the TAG stores and released in circulation. They are transported to various parts of the body after being complexed with albumin; each molecule of albumin can bind 6 to 8 fatty acid molecules. Subsequently they are taken up by peripheral tissues where they are oxidized to release energy. The principal oxidative pathway is β-oxidation, though the other pathways (such as α-oxidation and ω-oxidation) play a supplementary role. The de novo synthesis of fatty acids (i.e. lipogenesis) occurs from acetyl CoA with the help of *fatty acid synthase* complex.

In this chapter, various aspects of metabolism of fatty acids and the compound lipids (e.g. triacylglycerol, phospholipids, and sphingolipids), particularly in adipose tissue and liver, have been described. After going through this chapter the student should be able to understand:

- The oxidative pathways (e.g. β-, α-, ω-, or peroxisomal oxidation) for degradation of fatty acids and catabolism of unsaturated and odd chain fatty acids.
- Mechanism of fatty acids synthesis; *fatty acid synthase* system; and *desaturase* and chain elongation systems.
- Metabolic processes occurring within adipocytes; mechanism of mobilization of depot fat from adipose tissue and the factors influencing it, particularly the role of hormones.
- Pathways of biosynthesis and catabolism of compound lipids such as phosphoglycerides, triacylglycerols and sphingolipids.

I. β-Oxidation

β-Oxidation is the principal pathway for catabolism of fatty acid. The scheme of β-oxidation, whereby fatty acids are degraded by successive loss of two-carbon units, was first elucidated by a German scientist *Fray Knoop* in 1904. It consists of repeated cycles of a series of reactions. With each cycle, a two-carbon unit (i.e. acetyl CoA molecule) is removed from the carboxyl terminal of the fatty acid. Thus complete oxidation of the 16-carbon fatty acid (e.g. palmitic acid) requires seven such cycles and generates eight molecules of acetyl CoA.



The pathway occurs in **three stages**:

- Activation of fatty acid in the cytosol
- Transport of activated fatty acid into mitochondria
- Standard β-oxidation process in mitochondrial matrix.

Oxidation of fatty acids occurs in the mitochondrial matrix of most cells by the pathway of β-oxidation. There occurs sequential removal of two-carbon acetyl CoA units from the end of the acyl chain.

A. Activation of Fatty Acid

As the priming step for catabolism, the fatty acids are activated through formation of a thioester linkage between a fatty acid molecule and coenzyme A. The product is an acyl coenzyme A, and therefore the name of the enzyme that catalyzes this reaction is *acyl CoA synthetase*, also called ***thiokinase***. This reaction requires a great deal of energy and in the process ATP is converted to AMP

$$\text{Fatty acid} + \text{ATP} + \text{CoASH} \rightarrow \text{Fatty acyl CoA} + \text{AMP} + \text{PPi}$$
$$(\Delta G^{0\prime} = -0.20 \text{ kcal/mole})$$

This was the first reaction in biochemistry found to yield pyrophosphate by cleavage of ATP. The reaction favours the formation of fatty acyl CoA, since the pyrophosphate formed is hydrolyzed by the enzyme, *pyrophosphatase*: PPi 2→Pi. A large drop of free energy (−6.9 kcal/mol) accompanies the *pyrophosphate cleavage*, which ensures irreversibility of the overall reaction. Thus activation of fatty acid to fatty acyl CoA with an energy-rich thioester bond requires expenditure of two high-energy phosphate bonds.

The fatty acid is activated by forming a thioester link with coenzyme A before entering the mitochondria. The reaction uses a molecule of ATP, and is irreversible due to subsequent hydrolysis of PPi to two molecules of Pi.

The activation reaction actually takes place in **three steps:**

1. The carboxyl group is first activated to an enzyme-bound, high-energy, *acyl-adenylate intermediate* (fatty acyl-AMP).
2. The acyl group then reacts with coenzyme-A to give fatty acyl CoA.
3. The pyrophosphate is now hydrolyzed to ensures irreversibility of the overall reaction.

$$\text{Fatty acid} + \text{ATP} \longrightarrow \text{Fatty acyl-AMP} + \text{PPi}$$
$$\text{Fatty acyl-AMP} + \text{CoASH} \longrightarrow \text{Fatty acyl CoA} + \text{AMP}$$
$$\text{PPi} \longrightarrow \text{Pi} + \text{Pi}$$

Sum:

$$\text{Fatty acid} + \text{ATP} + \text{CoASH} \longrightarrow \text{Fatty acyl CoA} + \text{AMP} + 2\text{Pi}$$

Types of Thiokinase Enzymes

At least four different *thiokinase* enzymes, one each for short chain, medium chain, long chain fatty acids, and one for arachidonate, have been identified (Table 11.1). The chemistry and bioenergetics of each of the enzyme-catalyzed reactions for the biosynthesis of variable length acyl CoA molecules are the same.

Table 11.1. Substrate specificity and locations of *thiokinases*

Fatty acid substrate chain length	Location
Short: 2–3 carbon atoms	Mitochondrion
Medium: 4–12 carbon atoms	Mitochondrion
Long: 12 or more carbon atoms	Endoplasmic reticulum
Arachidonate specific (20:4 ω–6)	Endoplasmic reticulum

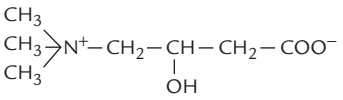


Fig. 11.1. Structure of carnitine.

B. Transport of Activated Fatty Acid into Mitochondria

Formation of fatty acyl CoA takes place in the cytosol. Further degradation occurs in the mitochondrial matrix. This creates a problem because the inner mitochondrial membrane (IMM) is impermeable to fatty acyl CoA. To overcome this problem, participation of a carrier, **carnitine** (Fig. 11.1) is required. It is *beta-hydroxyl-gamma-trimethyl ammonium butyrate*, synthesized from lysine and methionine in liver and kidney.

The transport process, referred to as **carnitine shuttle**, comprises three steps:

Step 1: The acyl group of fatty acyl CoA is transferred to carnitine, resulting in formation of fatty acyl carnitine. The reaction is catalyzed by the *rate-limiting* enzyme of the pathway, *acyl CoA carnitine transferase* (E_1), which is located on outer surface of the inner mitochondrial membrane.

$$\text{Fatty acyl CoA} + \text{Carnitine} \xrightarrow{E_1} \text{E Fatty acyl carnitine} + \text{CoASH}$$

Step 2: Fatty acyl carnitine is translocated across the inner mitochondrial membrane into the mitochondrial matrix. Entry of the acyl-carnitine is linked to the exit of carnitine, both mediated by a *translocase*, called ***carnitine/acylcarnitine translocase*** **(E_2)** (Fig. 11.2).

Step 3: Once inside the mitochondrial matrix, the fatty acyl carnitine is reconverted to fatty acyl CoA by the enzyme ***acyl CoA carnitine transferase II*** **(E_3)**. The enzyme is located on the inner surface of IMM.

$$\text{Fatty acyl carnitine} + \text{CoASH} \xrightarrow{E3} \text{Fatty acyl CoA} + \text{Carnitine}$$

The inner mitochondrial membrane is not permeable to long chain fatty acids and so these are transported into the mitochondria as carnitine derivatives by *carnitine/acylcarnitine translocase*.

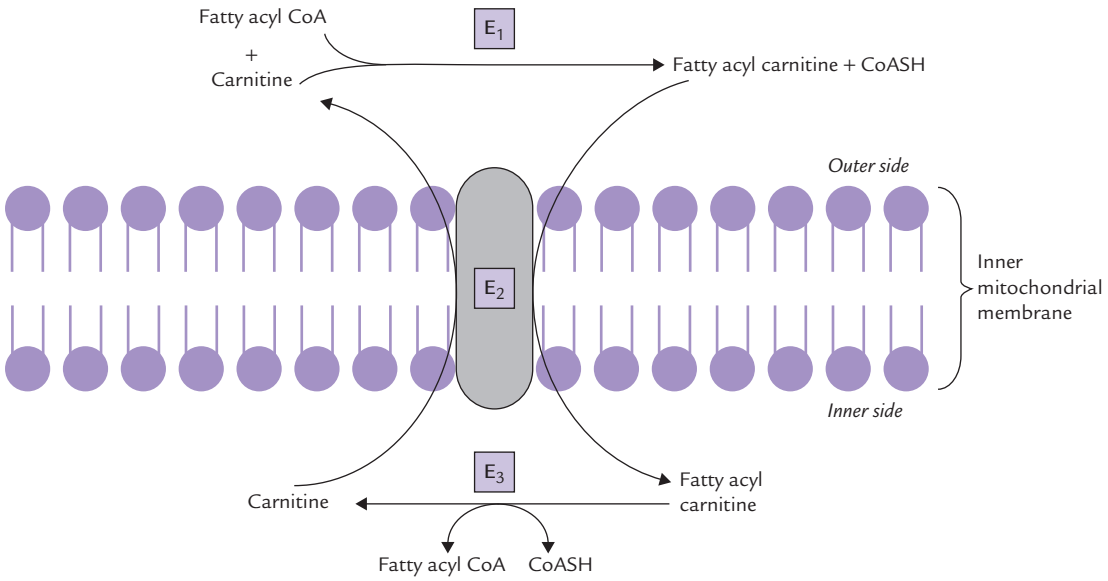


Fig. 11.2. Transport of fatty acids (16 ± 4 carbons) from cytosol into the mitochondrion (E_1 = *acyl CoA-carnitine transferase I*, E_2 = *translocase*, E_3 = *acyl CoA carnitine transferase II*).

Since carnitine is an essential component of the transport of fatty acyl CoA, its deficiency leads to decreased entry of fatty acids into mitochondrial matrix. This results in impaired fatty acid utilization, specially in muscles (**Case 11.1**).

The carnitine shuttle is primarily for the transport of the long chain fatty acids (16 ± 4 carbons).

- *Short and medium-chain fatty acids* can cross the mitochondrial membrane by **passive diffusion**, and are activated to their CoA derivative within the mitochondrion.
- *Very long-chain fatty acids* are shortened to long-chain fatty acids in peroxisomes and then transported by carnitine shuttle into the mitochondrion.

C. Standard β-Oxidation Process of Activated Fatty Acid

The principal fate of fatty acyl CoA in the mitochondrion is β-oxidation.

Reactions of β-Oxidation

The standard β-oxidation process comprises a series of identical cycles. Each cycle shortens the fatty acid by two carbons and produces an acetyl residue in the form of acetyl CoA. Repeated cycles of β-oxidation generate several acetyl CoA molecules which enter TCA cycle and are further degraded to carbon dioxide and water. The reaction sequence is *called β-oxidation because it is the β-carbon (C-3) that is oxidized*.

In each cycle of β-oxidation, fatty acyl CoA is acted upon in a sequential manner by four enzymes—**FAD-linked *dehydrogenase, hydratase, NAD⁺-linked dehydrogenase,*** and ***thiolase***. These reactions (Fig. 11.3) proceed as follows:

1. ***First dehydrogenation:*** A pair of hydrogen atoms is removed across the α- and β-carbons of the fatty acyl CoA by action of *fatty acyl CoA dehydrogenase* resulting in the formation of *trans*-Δ^2-enoyl CoA. Four distinct *dehydrogenases* are known to exist, each one is specific for a given range of fatty acid chain length. All four are flavoproteins and contain a tightly bound molecule of flavin adenine dinucleotide (FAD).
2. ***Hydration*:** Addition of a water molecule converts *trans*-Δ^2-enoyl CoA to L-3-OH acyl CoA. The reaction is catalyzed by *enoyl CoA hydratase*. In fact there are two *hydratases*: one preferring short chain fatty acids and the second long chain. Both are highly specific and act only on the *trans* isomer (i.e. *trans*-Δ^2-enoyl CoA).
3. ***Second dehydrogenation:*** The L-3-OH-acyl CoA then loses a pair of hydrogen atoms to NAD^+. The reaction is catalyzed by the enzyme *L-3-OH-acyl CoA dehydrogenase*. Two hydrogen atoms are removed from C-3 so that the hydroxyl group at that position is converted to a keto group. The reaction product is accordingly called 3-ketoacyl CoA.

   Thus, by a sequential action of three enzymes, a keto group is introduced at C-3 position of the fatty acyl CoA molecule (compare structure of fatty acyl CoA with that of 3-ketoacyl CoA; Figure 11.3).
4. ***Thiolysis:*** Finally, a (thiolytic) cleavage occurs at the β-carbon atom of the 3-ketoacyl CoA, liberating an acetyl CoA molecule. The enzyme catalyzing this step is *thiolase*.

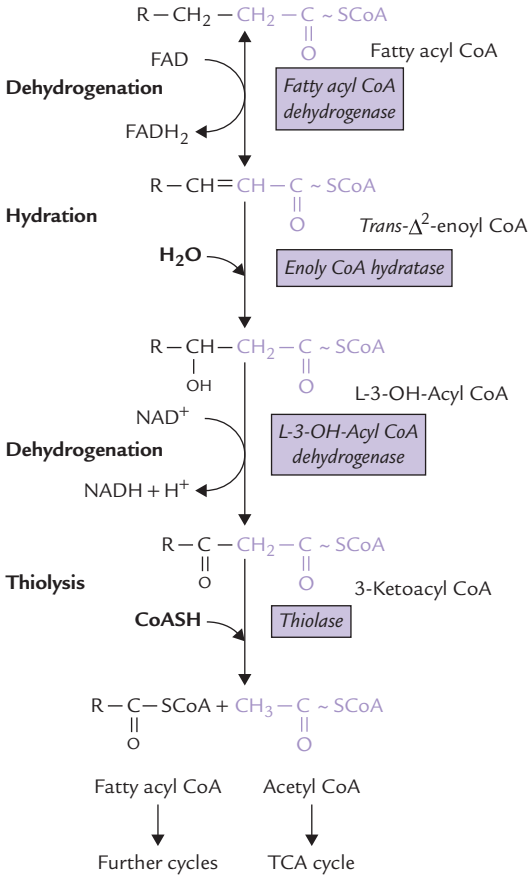


Fig. 11.3. Reactions of β-oxidation: dehydrogenation, hydration, dehydrogenation and thiolysis. These four reactions form one 'round' of degradation and their overall effect is to remove a 2-C (acetyl CoA) unit.

The active fatty acids (acyl CoA) are transported into mitochondrion by a carrier, carnitine, where they are oxidized at the β-carbon (β-oxidation) by oxidation (removal of hydrogen) of the acyl group, followed by an hydration, and a second oxidation (removal of hydrogen) at the β-carbon, followed by thiolysis (cleavage).

Repeating Sequence of Reactions

In the reaction sequence discussed above, one NADH and one $FADH_2$ are produced. At the end of the one cycle, the fatty acyl CoA is shortened by two carbons. The shortened fatty acyl CoA then undergoes a second cycle to liberate another acetyl CoA. These cycles are continued till whole of the acyl chain is degraded to acetyl CoA molecules. In case of palmitic acid there are seven such cycles, and the overall reaction can be represented by the following equation:

$$\text{Palmitoyl CoA} + 7\text{FAD} + 7\text{NAD}^+ + 7\text{H}_2\text{O} + 7\text{CoASH} \longrightarrow 8\ \text{Acetyl CoA} + 7\text{FADH}_2 + 7\text{NADH} + 7\text{H}^+$$

The NADH and $FADH_2$ feed directly into oxidative phosphorylation, while acetyl CoA feeds into TCA cycle, where further NADH and $FADH_2$ are produced.

D. Energy Production from β-Oxidation

For each round of degradation, one $FADH_2$, one NADH and one acetyl CoA molecule are produced.

Energy is generated by oxidation of these products of β-oxidation: 2ATPs are produced by oxidation of $FADH_2$; 3ATPs by NADH and 12 ATPs by degradation of acetyl CoA (through TCA cycle). Therefore, when a 16-C palmitic acid is fully oxidized a total of **131 ATPs** ($8 \times 12 + 7 \times 2 + 7 \times 3$) are generated:

8 Acetyl CoA	→	96 ATP
7 $FADH_2$	→	14 ATP
7 NADH	→	21 ATP
		131 ATP

2 ATPs (i.e. two phosphodiester bonds) are used during initial activation of the fatty acid. Hence, net gain in terms of ATP molecules is 131–2, i.e. **129 ATPs**.

Each cycle of β-oxidation produces an acetyl CoA and an acyl CoA with two carbons less, as also one $FADH_2$ and NADH each. The cycles are repeated until complete degradation occurs. Complete oxidation of one mole of palmitate generates 129 ATPs.

Regulation of β-oxidation

The regulation occurs at the following two levels:

- Supply of fatty acids
- Mitochondrial uptake of fatty acids

Supply of Fatty Acids

The β-oxidation is primarily regulated by the supply of fatty acids. Within rather wide limits, the *use of fatty acids by tissues is proportional to the plasma-free fatty acid level,* and therefore fatty acid oxidation is for the most part regulated at the level of adipose tissue metabolism. During fasting and uncontrolled diabetes, for example, a large amount of free fatty acids are released from adipose tissue (Chapter 12) and β-oxidized to acetyl CoA.

Mitochondrial Uptake of Fatty Acids

An additional regulatory step is the uptake of fatty acids into the mitochondrion. The enzyme, *acyl CoA carnitine transferase I*, which regulates entry of fatty acids into mitochondrial matrix, is *allosterically inhibited by* **malonyl-CoA.** This is important because malonyl-CoA is a substrate

for fatty acid synthesis and its cytoplasmic concentration is high whenever fatty acid synthesis is stimulated. It would be futile to have fatty acid synthesis occurring at the same time as fatty acid degradation, and hence, a substrate for fatty acid synthesis inhibits fatty acid degradation (in this case by preventing the acyl group being transferred into the mitochondrial matrix).

β-Oxidation is regulated by the supply of fatty acids, and by the enzyme which regulates entry of these fatty acids into mitochondrial matrix.

Note: Non-esterified fatty acids in plasma are referred to as *free fatty acids*, but it should be noted that they are neither free (99% bound to albumin) nor acids (at pH 7.4 present as anions). Plasma concentration ranges from 0.2 to 0.6 mM depending on nutritional status, but may rise as high as 2.0 mM in severe stress.

E. Defects of β-oxidation

The tissues such as muscle and liver, which derive most of their energy from fatty acids, at least during fasting are compromised by defects of β-oxidation. Among these common defects are discussed below:

Deficiency of Enzymes of β-Oxidation

Several defects of enzymes within the **β-oxidation** are known, the most common of which is *medium-chain acyl-CoA dehydrogenase deficiency*. The affected enzyme catalyzes the first dehydrogenation reaction of the β-oxidation of medium-chain (C_4 to C_{12}) fatty acids. This defect presents with episodic *non-ketotic hypoglycaemia* initiated by fasting during the first 2 years of life. Many of such patients initially diagnosed with "Rey's syndrome" or "sudden infant death syndrome" later turned out to suffer from medium chain *acyl-CoA dehydrogenase* deficiency. The hypoglycaemia is often lethal and at autopsy the patients who die of this disease show a typical fatty infiltration of the liver. These patients should be advised to avoid excessive fasting and recommended a low fat diet.

Deficiency of Translocase

A few cases of *translocase* deficiency have been reported, where the transport of long chain fatty acids across the mitochondrial membrane is impaired. *Muscle cramps*, precipitated by fasting, exercise and high fat diet, is the predominant clinical feature.

Carnitine Deficiency

Defects in carnitine biosynthesis, defective transport into the cells, or excessive renal excretion may result in deficiency of this carrier protein, which most commonly affects liver and muscles. When liver is affected, *hypoketotic hypoglycaemia* during periods of extended fasting is the most obvious consequence. *Muscle weakness* and *muscle cramps* on exertion are the most typical complaints when muscle suffers in carnitine deficiency **(Case 11.1)**. Histologically, many patients show an unusual abundance of fat droplets in muscle tissue and liver, and some develop fatty degeneration of the liver. This is to be expected, because excess fatty acyl CoA that cannot be transported into the mitochondrion is diverted into triglyceride synthesis.

Jamaican Vomiting Sickness

Unripe ackee fruits contain hypoglycin, an unusual toxic amino acid that inhibits the enzyme *acyl CoA dehydrogenase*. Severe hypoglycaemia, vomiting, convulsions and coma are prominent features.

II. Other Oxidative Pathways

Mitochondrial β-oxidation can smoothly oxidize unbranched saturated fatty acids with an even number of carbons and a chain length up to 18 or 20 carbons. Fatty acids that do not fit this description require additional enzymatic reactions:

- Methylated fatty acids require **α-oxidation** initially, before they can be beta-oxidized.
- Very long chain fatty acids (VLCFA) are handled initially by **peroxisomal oxidation** pathway.
- Medium chain fatty acids are degraded by **ω-oxidation**.
- Unsaturated and odd chain fatty acids also require additional processing before they are beta-oxidized.

Oxidation of some fatty acids (e.g. unsaturated, branched, odd chain) requires specialized reactions.

A. α-Oxidation of Fatty Acids

α-Oxidation is a minor pathway occurring in the endoplasmic reticulum and mitochondria. It removes one carbon unit at a time from the carboxyl end. Unlike β-oxidation, the function of this pathway is not the production of any high-energy phosphate bonds, but rather the degradation of methylated fatty acids.

A methylated fatty acid cannot be β-oxidized without prior modification. Its degradation is initiated by oxidation of the carbon-2 (the α-carbon) and then release of carbon-1 as CO_2, thus shortening the fatty acid by one

carbon at a time. The shortened fatty acid then undergoes a round of β-oxidation yielding propionyl CoA rather than acetyl CoA (Fig. 11.4).

Refsum's Disease

It is a rare recessively inherited defect of α-oxidation. The disease is also called **phytanic acid storage disease** because the patients accumulate excessive quantities of phytanic acid (Fig. 11.4) in their body lipids. This unusual fatty acid is derived from the alcohol phytol, an isoprenoid, that occurs as a constituent of chlorophyll and in milk.

The β-carbon of phytanic acid is methylated, which blocks β-oxidation (Fig. 11.4). The phytanic acid levels in blood increase, reaching up to 100 mg/dl (normal value is around 0.3 mg/dl). Phytanic acid also accumulates in liver, sometimes to such an extent that nearly 50% of the total fatty acids in the liver is phytanic acid. *Clinical features* include abnormalities of the nervous system like cerebellar ataxia, motor weakness and distal sensory loss. Such abnormal neurological features probably result due to incorporation of the branched chain fatty acids in the membrane lipids. This increases fluidity of the neuronal membranes, which affects nerve conduction. The patients respond to dietary restriction of green vegetables and of ruminant milk and meat, which contain large amounts of phytanic acid.

α-Oxidation oxidizes carbon 2 (the α-carbon) and then releases carbon-1 as CO_2, thereby shortening the fatty acid by one-carbon at a time. It is used for oxidation of methylated fatty acids, such as phytanic acid. The latter cannot be β-oxidized without prior modification because the β-carbon is methylated.

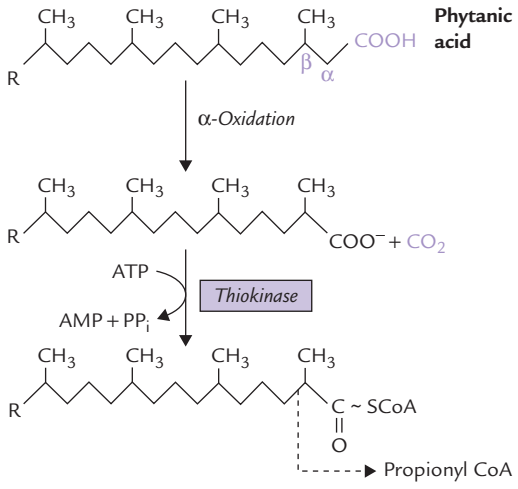


Fig. 11.4. Degradation of a methylated acid (phytanic acid) is initiated by α-oxidation, followed by a round of β-oxidation, which yields propionyl CoA rather than acetyl CoA.

B. Fatty Acid Oxidation in Peroxisomes

The peroxisomal β-oxidation handles **very long chain fatty acids** (VLCFA), of chain length **>20 carbons**, which are poor substrates for mitochondrial β-oxidation. It uses essentially the same reaction sequence as the mitochondrial system, except for the very first reaction, which is catalyzed by an H_2O_2-producing *flavoprotein oxidase* (it is *FAD-dehydrogenase* in mitochondria).

The *flavoprotein oxidase* acts on the CoA derivatives of VLCFA, after the latter crosses the peroxisomal membrane (without involvement of carnitine). It catalyzes removal of a pair of hydrogen atoms from the hydrocarbon chain of the fatty acid molecule and transfers it directly to oxygen to form hydrogen peroxide. Removal of hydrogen atoms introduces a double bond in the hydrocarbon chain to form *trans*-Δ^2-enoyl CoA. The latter is an intermediate of β-oxidation. It undergoes similar sequence of reactions as in mitochondrial β-oxidation to liberate an acetyl CoA molecule, leaving behind a fatty acyl CoA molecule shortened by two carbons. In this manner, action of the *flavoprotein oxidase introduces a shortcut, whereby the substrate fatty acid directly enters the standard β-oxidation cycle.*

The above cycle is repeated several times, each time an acetyl CoA is removed, the fatty acid chain is shortened by two carbon atoms. This cycle proceeds only to the stage of **octanoyl-CoA** (8-carbon); the latter moves to the cytosol and then into the mitochondrial matrix with the help of specific membrane transport proteins, where it undergoes further degradation by the standard β-oxidation process. Thus it may be said that *main purpose of peroxisomal β-oxidation is to shorten very long chain fatty acids, thereby boosting* β-oxidation.

VLCFA (>20-C) ⟶ Peroxisomes ⟶ Sequential release of 2-C (acetyl CoA) units till octanoyl CoA (8-C) is left ⟶ Enters mitochondria ⟶ β-oxidation

High fat diet and the hypolipidaemic drugs like clofibrate cause a marked increase in the number of peroxisomes in a cell in order to increase efficiency of fatty acid utilization.

Zellweger syndrome: It is an *inborn error of VLCFA oxidation* because of absence of peroxisomes. Substrate accumulation in tissues, particularly brain, liver and kidneys, and excretion in urine (shown by chromatography) and non-ketotic hypoglycaemia are the predominant features of this defect.

Peroxisomal β-oxidation degrades VLCFA by sequentially removing acetyl CoA till octanoyl CoA (8-C) is left. All enzymes are same as in mitochondrial system except the first: an H_2O_2 producing *flavoprotein oxidase*, which directly generates an intermediate of standard β-oxidation process.

C. Omega (ω)-Oxidation of Fatty Acids

Omega oxidation is *a microsomal system* for *oxidation of the carbon atom most remote from the carboxyl group in a fatty acid to produce a dicarboxylic acid*. The basic reaction, catalyzed by a *monooxygenase* that requires NADPH, O_2, and cytochrome P-450, occurs in smooth endoplasmic reticulum. First, methyl group at the ω-carbon is converted to CH_2OH and subsequently oxidized with the help of NAD^+ to a COOH group to produce a dicarboxylic acid. The latter can be activated at either end, followed by β-oxidation (Fig. 11.5).

Medium chain fatty acids mobilized from the adipocyte triacylglycerol stores are oxidized in liver by ω-oxidation. Moreover, ω-oxidation becomes important when β-oxidation is defective. Dicarboxylic acids (C-6 and C-8 acids) are excreted in urine under these circumstances and the condition is termed as **dicarboxylic aciduria**.

Alpha oxidation makes possible oxidation of methylated fatty acids; ω-oxidation is for medium chain fatty acids and for oxidizing very long chain fatty acids a distinctive pathway in peroxisomes exists.

D. Oxidation of Odd Chain Fatty Acids

Even chain fatty acids, by far the most abundant fatty acids in human organism, are β-oxidized to acetyl CoA. Odd chain fatty acids are also β-oxidized normally but the last step produces a 3-carbon **propionyl CoA** instead of acetyl CoA. If C-17 fatty acid, for example, were catabolized, 7 acetyl CoA molecules and one propionyl CoA molecule would result after seven cycles of the β-oxidation process (Fig. 11.6).

Propionyl CoA is catabolized to succinyl CoA, a Krebs cycle intermediate via methylmalonyl-CoA (Fig. 11.7). In this series of reactions, bicarbonate is incorporated into propionate by a biotin-dependent enzyme *propionyl CoA carboxylase* to yield D-methylmalonyl CoA. The latter undergoes intramolecular rearrangements in presence of a *racemase* to give L-methylmalonyl CoA, which is then converted to succinyl-CoA in the presence of a *vitamin B_{12}-containing mutase*. The last step involves migration of the bulky COSCoA moiety to the methyl carbon.

Odd chain fatty acids can therefore give net biosynthesis of glucose through propionyl CoA. Thus, *three carbon*

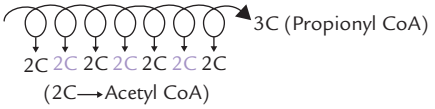


Fig. 11.6. Odd chain fatty acid produces several acetyl CoA and a single propionyl CoA (3-C) molecule in the final round of fatty acid degradation.

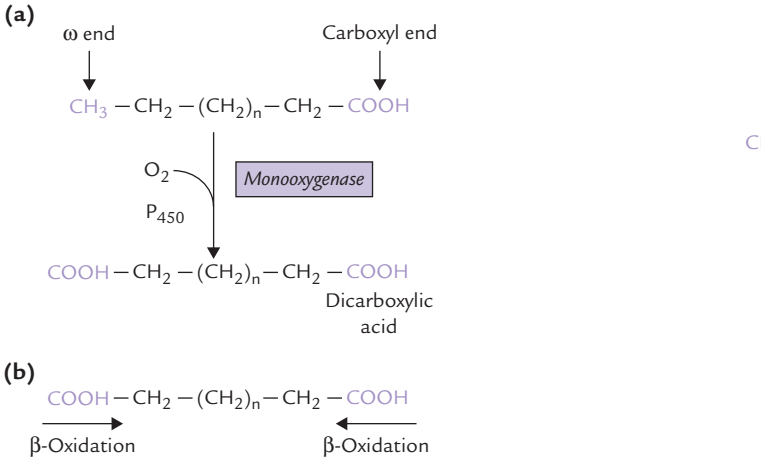


Fig. 11.5. The ω-oxidation of medium chain fatty acid. **(a)** The last carbon (ω) of a fatty acid is oxidized to a carboxyl group to produce a dicarboxylic acid. **(b)** The dicarboxylic acid is activated at either end, followed by β-oxidation.

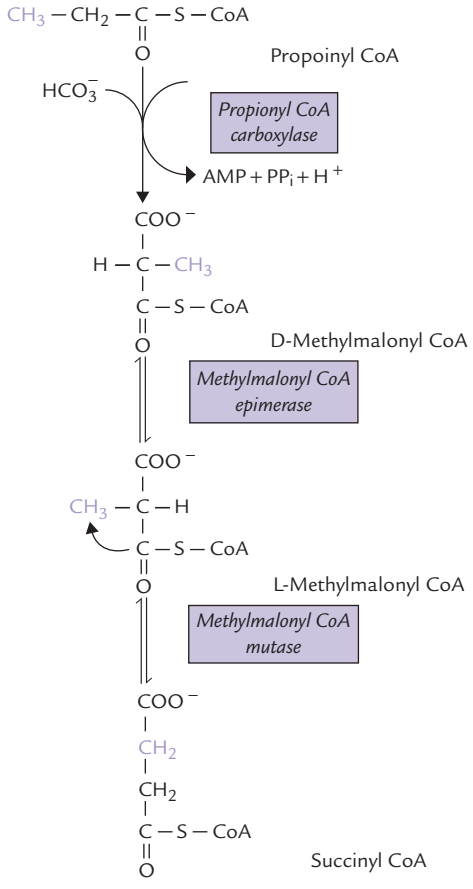


Fig. 11.7. Pathway for metabolism of propionyl CoA: converted to succinyl CoA via methylmalonyl CoA.

units from odd chain fatty acids (comprising ω-1 to ω-3 carbons) are glucogenic.

Inborn Errors of Propionate Metabolism

Propionic Acidaemia

It results when *propionyl-CoA carboxylase* is deficient because of a genetic lesion. It is characterized by developmental abnormalities and ketoacidosis.

Methylmalonic Aciduria

This results when the B_{12}-dependent *methylmalonyl CoA mutase* reaction is blocked. Methylmalonic acid accumulates in the blood and is excreted in urine. Methylmalonic aciduria can be caused by (a) inherited **defect in the *mutase*** enzyme, or (b) inherited **defect in cobalamin** (B_{12}) metabolism. In the latter condition, there is inability to convert dietary B_{12} to the appropriate coenzyme form, i.e. adenosylcobalamin. (Lack of vitamin B_{12} from dietary sources may also lead to this condition.)

The inherited forms of methylmalonic aciduria are severe maladies (overall frequency 1 in 10,000), with life threatening acidosis in affected infants and children. The patients suffering from inherited defect in cobalamin metabolism respond to injected adenosylcobalamin (**Case 13.3**), whereas those with inherited defect in the *mutase* enzyme do not. In either case accumulation of methylmalonic acid causes severe metabolic acidosis, CNS damage and growth retardation. It is desirable to restrict intake of odd chain fatty acids. Restriction of dietary valine, isoleucine, threonine and methionine is also required because propionyl CoA is derived from the metabolism of these amino acids also (Chapter 13).

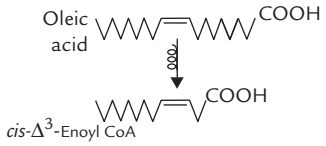
Note: The disorders of metabolism of fatty acids, such as methyl malonic aciduria, propionic acidaemia, medium chain *acyl CoA dehydrogenase* deficiency and *flavoprotein oxidase* deficiency are collectively referred to as **organic acidurias**. The accumulation of organic acids in body tissues and their excretion in urine characterize them all. Their collective incidence is about one in 3000 live births.

E. Oxidation of Unsaturated Fatty Acids

Oxidation of unsaturated fatty acids initially proceeds by the β-oxidation process in the same way as described for saturated fatty acids. It begins at the carboxyl terminal, and successive removal of two carbon units occurs with each cycle of β-oxidation. But *when a double bond is reached, the cycle cannot proceed any further* because the intermediate formed at this stage is not a natural substrate for the enzymes of β-oxidation. Participation of additional enzymes is required at this stage. These enzymes convert the above mentioned intermediate to a natural intermediate of β-oxidation.

Monounsaturated Fatty Acids

The monounsaturated fatty acids, for example, oleic acid (*cis*-Δ^9-octadecanoic acid), has 18 carbon atoms with a *cis* double bond in position 9. After three β-oxidation cycles, the chain is shortened by six carbons, and now the oxidative mechanism encounters *cis*-Δ^3-enoyl CoA.



This compound is not a natural substrate for β-oxidation because there is a double bond between C-3 and C-4 in *cis* configuration. For β-oxidation to proceed, we need a double bond in *trans*-Δ^2-configuration in enoyl CoA (Fig. 11.3). This problem is circumvented by the action of an ***isomerase*** which converts the *cis*-Δ^3 bond into trans-Δ^2 bond. As a result *cis*-Δ^3-enoyl CoA is converted to *trans*-Δ^2-enoyl CoA (Fig. 11.8). The latter being a normal substrate for β-oxidation, the reaction cycle can now proceed further.

Polyunsaturated Fatty Acids (PUFA)

The enzymatic difficulty in linolenic acid (*cis cis* 9, 12 octadecanoic acid) arises due to presence of two *cis* double bonds at position 9 and 12. The initial reaction sequence take place as in case of monosaturated fatty acid (Reactions I–III; Fig. 11.9). The oxidative process moves past the *cis*-Δ^9 double bond through action of the same *isomerase*, which convert the *cis*-Δ^3 bond into *trans*-Δ^2 bond, leaving behind Δ^4-*cis*-enoyl CoA. This compound, needs additional processing before it can be completely degraded by β-oxidation, as outlined below:

(a) Δ^4-*cis* enoyl CoA is acted upon by an *acyl CoA dehydrogenase* which introduces a double bond at the C-2

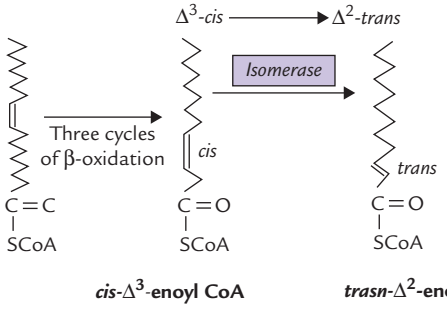


Fig. 11.8. Conversion of *cis*-Δ^3-enoyl CoA to *trans*-Δ^2-enoyl CoA. The enzyme is Δ^3-*cis* → Δ^2-*trans isomerase*.

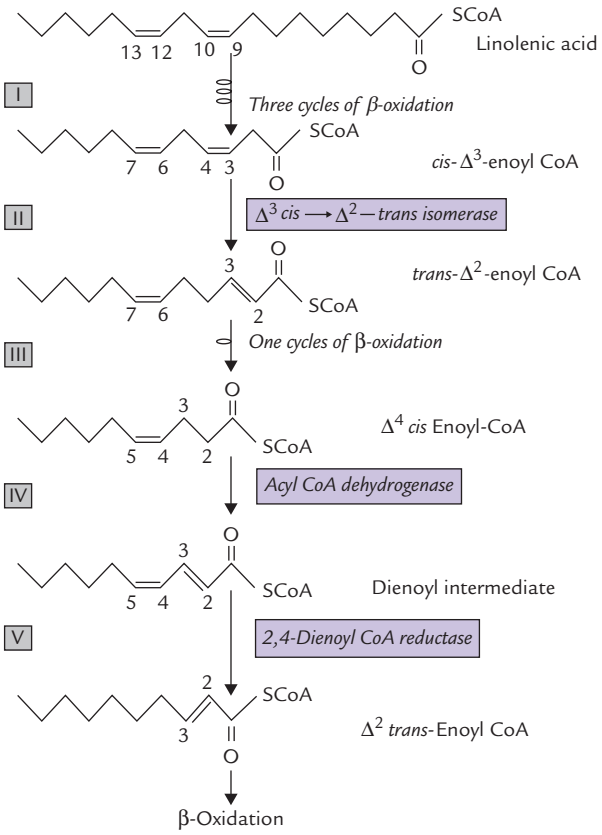


Fig. 11.9. Oxidation of linolenic acid.

(Reaction IV). This produces a 2,4-dienoyl intermediate (Δ^2-*trans*-Δ^4-*cis*-dienoyl CoA).

(b) The 2,4-dienoyl intermediate is a poor substrate for *enoyl CoA hydratase*, the next enzyme of the standard β-oxidation process. This difficulty is solved by an NADPH-dependent enzyme ***dienoyl CoA reductase*** (Reaction V), which reduces the double bond at C-4 to produce Δ^2-trans-enoyl CoA.

(c) The Δ^2-*trans* enoyl CoA is a normal substrate for the *enoyl CoA hydratase*, and hence for β-oxidation.

Unsaturated fatty acids require additional processing before they can be degraded by β-oxidation. Monounsaturated fatty acids (e.g. oleic acid) requires an isomerase, and polyunsaturated fatty acids (e.g. linoleic acid) requires additional actions of an isomerase, a dehydrogenase and a reductase.

III. Ketone Body Production and Metabolism

Ketone bodies is a term that refers to the three biosynthetically related compounds **acetoacetate, β-hydroxybutyrate** and **acetone**. They are produced in the liver from acetyl

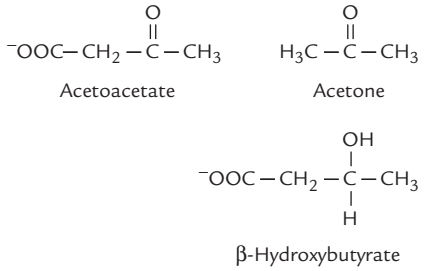


Fig. 11.10. Structures of ketone bodies. The ketone group (C = O) is present in acetoacetate and acetone, but not in β-hydroxybutyrate.

CoA substrate and are carried to extrahepatic tissues where acetoacetate and β-hydroxybutyrate are metabolized to produce energy. Being energy rich they serve as *concentrated packets of metabolic energy.* Acetone is however, an exception, since it cannot be metabolized and is readily exhaled. As diagrammed Figure 11.10 acetoacetate and acetone are true ketones, while β-hydroxybutyrate (or 3-hydroxybutyrate) does not posses a keto group.

Because normally ketone bodies are produced in liver in such amount that can be rapidly extracted from circulation by various organs, their *plasma level remains relatively low: 0.3 mM (1 mg/dl) and only traces are excreted in urine (daily excretion less than 1 mg)*. However, they are synthesized in inordinate amounts in some conditions (e.g. starvation and diabetes mellitus), which exceeds the ability of extrahepatic tissues to utilize them. Consequently, they get accumulated in blood leading to ketonaemia and excretion in urine (ketonuria). *Ketonaemia and ketonuria together constitute a condition called* **ketosis**.

In starvation and diabetes mellitus, the liver converts excess fatty acids to ketone bodies: acetoacetate, β-hydroxybutyrate and acetone, which are released into the blood and oxidized in extrahepatic tissues.

A. Ketone Body Synthesis

Ketone bodies are synthesized exclusively in **liver mitochondria** from acetyl CoA or acetoacetyl CoA; the process is called **ketogenesis** (Fig. 11.11). Acetyl CoA is obtained by oxidation of pyruvate, fatty acids or the ketogenic amino acids (lysine, leucine, tyrosine, phenylalanine, etc.). The following **steps** are involved in ketogenesis:

1. Acetoacetyl CoA is produced by condensation of two molecules of acetyl CoA. The condensation of acetyl CoA molecules is catalyzed by *thiolase*, the enzyme involved in final step of β-oxidation (where it cleaves

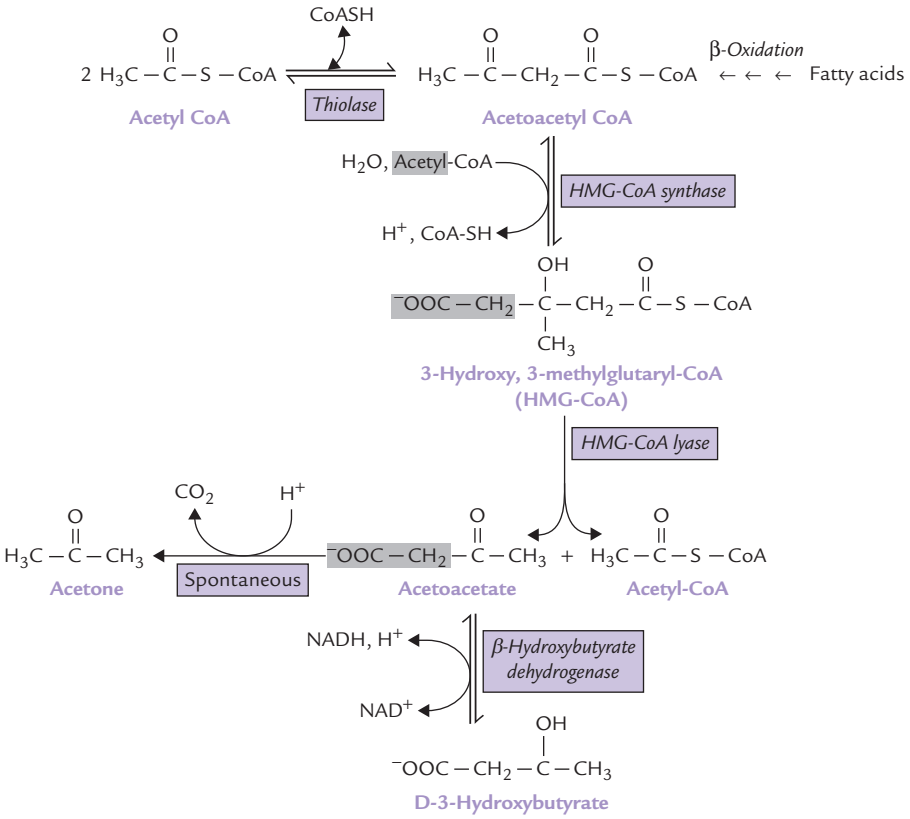


Fig. 11.11. Synthesis of ketone bodies in liver mitochondria. Only acetoacetate and β-hydroxybutyrate are present in plasma; the small amount of acetone normally formed by the non-enzymatic decarboxylation of acetoacetate is exhaled through lungs.

acetoacetyl CoA). When, however, the concentration of acetyl CoA is high, the reaction occurs in the reverse direction, producing acetoacetyl CoA (Fig. 11.11).

$$\text{Acetyl CoA} + \text{Acetyl CoA} \rightleftharpoons \text{Acetoacetyl CoA} + \text{CoASH}$$

2. Acetoacetyl CoA combines with another acetyl CoA molecule to produce HMG-CoA (3-hydroxy-3-methylglutaryl CoA). The enzyme catalyzing this step, *HMG CoA synthase,* is ***rate-limiting*** for the pathway; it regulates the synthesis of ketone bodies.

$$\text{Acetoacetyl CoA} + \text{Acetyl CoA} \overset{H_2O}{\rightleftharpoons} \text{HMG CoA} + \text{CoASH}$$

3. HMG CoA undergoes an aldose-*lyase* reaction, catalyzed by *HMG CoA lyase*. The cleaved products are **acetoacetate** and acetyl CoA.
4. Acetoacetate is reduced by NADH to form β-**hydroxybutyrate** by the enzyme β-*hydroxybutyrate dehydrogenase*, and a small fraction undergoes spontaneous decarboxylation to form **acetone.**

Note: Oxidation of some ketogenic amino acids directly yields acetoacetate, a ketone body (Chapter 13). Acetoacetyl CoA, an important intermediate in ketogenesis can be directly formed by β-oxidation of fatty acids.

Ketone bodies, synthesized in liver, are rich in metabolic energy. Acetoacetate and β-hydroxybutyrate are used in preference to glucose as energy source by certain tissues, as heart, muscle and renal contex. Brain also switches to using predominatly acetoacetate in starvation and diabetes.

B. Ketone Body Oxidation

The two main ketone bodies, acetoacetate and β-hydroxybutyrate, are water soluble. They are readily transported from the liver to various tissues, which use them as important sources of energy, e.g. skeletal muscles, cardiac muscle, renal cortex, etc. Within mitochondria of these tissues, the ketone bodies are metabolized in following steps:

- First, oxidation of β-hydroxybutyrate produces acetoacetate, catalyzed by β-*hydroxybutyrate dehydrogenase* by reversal of the synthetic reaction (Fig. 11.12).
- Acetoacetate is activated to acetoacetyl CoA via the mitochondrial *thiophorase.* Donor of the coenzyme group in this reaction is succinyl CoA.

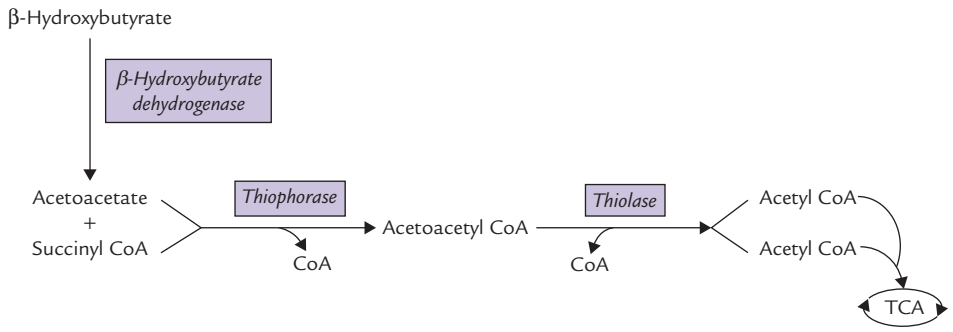


Fig 11.12. Oxidation of ketone bodies: acetoacetate and β-hydroxybutyrate. Acetone cannot be thus oxidized and most is exhaled through lungs.

- Acetoacetyl CoA is cleaved by *thiolase* to yield two acetyl CoA molecules, which are oxidized in TCA cycle.

Acetone, however cannot be metabolized and most is exhaled through lungs.

Erythrocytes and liver cannot oxidize ketone bodies. Because oxidation of ketone bodies occurs in mitochondria, tissues that lack mitochondria (e.g. erythrocytes) cannot metabolize ketone bodies. Liver also cannot utilize ketone bodies for it lacks an important enzyme (*thiophorase*).

C. Excessive Production of Ketone Bodies

Normally, rates of ketone body synthesis and oxidation are balanced so that their serum concentration is maintained around 1 mg/dl and urinary excretion is negligible (1 mg/24 hours). When, however, rate of ketogenesis exceeds their utilization, their concentration in blood rises, and the condition is known as ***ketonaemia***. Excess ketone bodies are eliminated in urine, this is known as ***ketonuria***. Ketonaemia and ketonuria, as mentioned earlier, present an overall picture of **ketosis**, a risky and potentially lethal condition.

Any substrate, including glucose, that is degraded to acetyl CoA in the liver can be converted to ketone bodies, *But ketogenesis is usually associated with excessive fatty acid oxidation*, which occurs most notably in ***starvation*** *and* ***diabetes mellitus***: 3-day-fast produces ketone bodies amounting to plasma level of 2–3 mM and in uncontrolled diabetes this may rise to 25 mM.

Starvation

It is a low insulin : high glucagon state, which causes flow of fatty acids from the adipose tissue and to the liver, where they become a major energy source (Chapter 12).

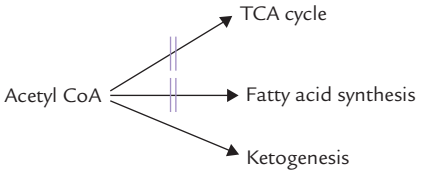


Fig. 11.13. Excessive amounts of acetyl CoA generated in starvation and diabetes mellitus are channeled into ketone body production; TCA cycle and fatty acid synthesis being retarded in these conditions.

Excessive generation of acetyl CoA results, which cannot be fully handled by TCA cycle because of the following reasons:

Oxaloacetate depletion: In the liver, an enhanced β-oxidation results in depletion of oxaloacetate because the excess of NADH produced converts it to malate. This causes dearth of oxaloacetate, thereby limiting activity of the TCA cycle.

Exhanced gluconeogenesis: Oxaloacetate is directed towards glucose synthesis for meeting the essential requirements for tissues like brain, and this further exacerbates the dearth of oxaloacetate and retards TCA cycle.

TCA being virtually blocked, the acetyl CoA must find another outlet. One of the possible outlets, the fatty acid biosynthesis, is largely shut down under such circumstances (reasons given later), and acetyl CoA is instead channeled into ketone body production. Consequently, large amounts of acetyl CoA (produced by β-oxidation in liver) are diverted towards overproduction of ketone bodies (Fig. 11.13).

Production of acetyl CoA increases in starvation and TCA is hampered. The acetyl CoA molecules in excess of that required for entry into TCA is diverted for ketone-body production.

Diabetes Mellitus

Severe and uncontrolled diabetes mellitus is the most common cause of ketosis. Persons with type 1 diabetes have little or no plasma insulin and the *insulin : glucagon ratio falls very low* in decompensated patients, resulting in overproduction of ketone bodies by the aforementioned series of events.

Acetoacetate and β-hydroxybutyrate decrease the pH of blood and cause *metabolic acidosis*; the condition called **diabetic ketoacidosis** is life threatening. In type 1 diabetes it develops very quickly, even after missing a single insulin dose, whereas in type 2 diabetes, ketoacidosis is relatively rare but may be precipitated by a major stress, such as myocardial infarction. Production of acetone being significantly enhanced, its smell in the patient's breath is a valuable aid to diagnosis of diabetic ketoacidosis. The breathing is characteristically deep and fast (i.e. *Kaussamal's breathing*).

Other Causes of Ketosis

In addition to starvation and severe diabetes, excessive quantities of ketone bodies are present in the circulation of patient with alcoholism, carbohydrate deficiency, toxaemia of pregnancy, prolonged labour and overdose of salicylates and methanol.

Dangers of Ketosis

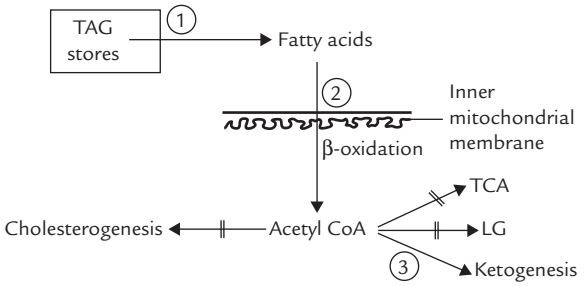
Ketosis causes **metabolic acidosis** and **loss of sodium** and **potassium ions** from the body, which are potentially lethal conditions. Large amounts of ketone bodies are excreted in urine, and because of their anionic character they also carry away Na^+. The problem is aggravated because uncontrolled diabetes is accompanied by an osmotic diuresis. Moreover, insulin increases potassium uptake by cells, therefore lack of insulin in diabetes leads to release of potassium, which is then uncontrollably lost in urine.

Loss of circulating Na^+ may aggravate acidosis, because the Na^+ depletion often results in decrease in circulatory bicarbonate ion concentration. One may view this problem in terms of ketone bodies displacing bicarbonate ions even if Na^+ concentration remains constant **(Case 15.3).**

D. Regulation of Ketogenesis

Synthesis of ketone bodies is regulated at the following levels:

1. *Availability of fatty acids*: Fatty acids are precursors of ketone bodies, and the source of free fatty acids is depot fat adipose tissue. Therefore, factors influencing mobilization of fatty acids from adipose tissue also influence the rate of ketogenesis. These are described later in this chapter.
2. *Entry of fatty acid into mitochondrial matrix*: *Carnitine acyl CoA transferase-I* regulates entry of fatty acids into mitochondrial matrix for oxidation, so its activity regulates ketogenesis as well. Regulation of the enzyme activity by malonyl CoA has been discussed earlier.
3. *Channelization of acetyl CoA into ketogenesis*: Liver has a very high capacity for fatty acid oxidation, and the acetyl CoA formed by β-oxidation has a choice between progressing into TCA cycle, fatty acid biosynthesis, or ketogenesis or even cholesterol synthesis. However, TCA cycle is impeded and acetyl CoA is not used for biosynthetic activity during fasting. This results in channelization of acetyl CoA into ketogenesis.



4. *Effect of hormones*: Insulin inhibits and glucagons stimulate ketogenesis, as discussed earlier.

IV. De Novo Synthesis of Fatty Acids

Fatty acids are obtained from both diet and ***de novo*** (new) synthesis. Because most fatty acids have multiples of two carbon atoms, they are synthesized from successive addition of two carbon units, the donor of which is acetyl CoA. **NADPH** is required for the reductive reactions of the pathway which occurs in the cytosol of the cell. The elongation of the fatty acid chain stops upon formation of palmitate (16-C). Further elongation (and desaturation) are carried out by other enzyme systems.

In humans, liver and lactating mammary glands are the main organs for the endogenous synthesis of fatty acids, although adipose tissue, brain and kidneys are involved to a lesser extent. Liver convert excess dietary carbohydrates into fatty acids and triacylglycerols.

The pathway of synthesis of fatty acids involves condensation of two carbon units, in the form of acetyl CoA, to form long hydrocarbon chain a series of reaction. The reactions are carried out on *fatty acid synthase* complex using NADPH as reductant.

The biosynthetic pathway occurs in 3 stages:

(a) Transport of 2-carbon units (acetyl CoA) to cytosol.
(b) Conversion of acetyl CoA to malonyl CoA.
(c) Reactions of *fatty acid synthase* complex.

A. Transport of Two Carbon Units to Cytosol

Fatty acid synthesis occurs in the ***cytosolic compartment*** of the cell. However, acetyl CoA, the precursor, is generated in the mitochondrial matrix as end product of various catabolic processes. It must reach cytosol by crossing the mitochondrial membranes. But the inner mitochondrial membrane (IMM) is impermeable to most molecules and ions, including acetyl CoA. Transport of acetyl CoA across the inner mitochondrial membrane requires that it is first converted to citrate which is capable of moving across the IMM.

In the cytosol the citrate is cleaved to regenerate acetyl CoA (Fig. 11.14).

$$\text{Citrate} + \text{CoASH} + \text{ATP} \xrightarrow{\textit{ATP-Citrate lyase}} \text{Oxaloacetate} + \text{Acetyl CoA} + \text{ADP} + \text{Pi}$$

Thus, by the coordinated action of two enzymes–*citrate-synthase* and *-lyase*, the acetyl CoA is effectively transported from mitochondrial matrix to the cytosol. While it is used for fatty acid synthesis, *oxaloacetate is shuttled back into the mitochondrion either as malate or as pyruvate*. It is converted to malate via *malate dehydrogenase* that uses NADH as the proton donor. Malate is then converted to pyruvate by the $NADP^+$-dependent *malic enzyme*. The reduced NADPH so formed becomes available for lipogenesis. Pyruvate is then translocated across the mitochondrial membrane to be cycled again.

$$\text{Oxaloacetate} + \text{Acetyl CoA} \xrightarrow{\textit{Citrate synthase}} \text{Citrate}$$

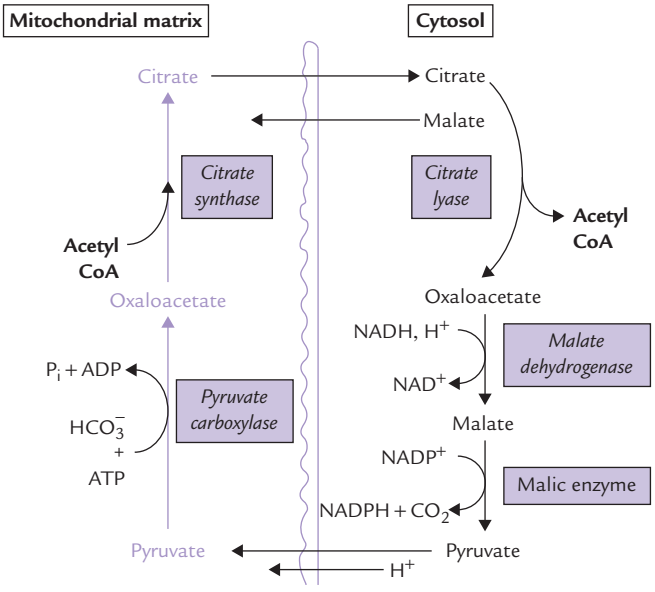


Fig. 11.14. Transport of acetyl CoA across inner mitochondrial membrane (IMM). It combines with oxaloacetate to form citrate, which readily crosses the membrane. In the cytosol citrate is cleaved to regenerate oxaloacetate.

The malic enzyme reaction can theoretically supply half of the NADPH required for fatty acid synthesis. The remaining NADPH has to be obtained from the pentose phosphate pathway.

B. Conversion of Acetyl CoA to Malonyl CoA

The key reactions of fatty acid synthesis are the carbon-to-carbon condensations. They require input of considerable energy and are, therefore, thermodynamically unfavourable. To overcome this energy barrier, acetyl CoA needs to be converted to activated form, which will serve as the donor of carbon units to the growing fatty acid chain. **Malonyl CoA**, a 3-carbon compound, is one such activated form. It is produced by carboxylation of acetyl CoA. The reaction is catalyzed by the enzyme *acetyl CoA carboxylase*, and requires free energy which is provided by ATP hydrolysis.

$$\text{Acetyl CoA} + HCO_3^- \xrightarrow{\text{ATP}\ \ \text{ADP+Pi}} \text{Malonyl CoA}$$

This is a typical ATP-dependent carboxylation reaction, with enzyme-bound biotin serving as a carrier of the carboxyl group (Chapter 18). Since the malonyl CoA is not used in other metabolic pathways, this reaction is the **committed step** of fatty acid synthesis. It also serves as a control point, since activity of *acetyl CoA carboxylase* is allosterically modulated by **citrate** (positive modulator) and **palmitoyl CoA** (negative modulator).

The first committed step in fatty acid synthesis is the carboxylation of acetyl CoA to form malonyl CoA which is catalyzed by the biotin containing enzyme, *acetyl CoA carboxylase*.

C. Reactions Catalyzed by *Fatty Acid Synthase*

Synthesis of fatty acid from acetyl CoA precursor occurs on a cytoplasmic polyprotein called *fatty acid synthase* (MW 400,000). *A poly-protein is a single protein with more than one enzymatic activities*. *Fatty acid synthase* is a dimer formed from two identical chains of such polyprotein (Fig. 11.15). Each one of these polyproteins exhibits multifunctionality, possessing seven enzyme activities (designated E_1 to E_7). Electron microscopic studies with this enzyme complex have shown that contiguous stretches of the polypeptide chain are folded to form seven autonomous domains, each with a different enzyme activity.

(For additional information on *fatty acyl synthase*, refer to Box. 11.1). These domains are arranged in such a way that they catalyze the successive steps in the fatty acid synthesis cycle. The biosynthetic intermediates do not diffuse away from the polyprotein but remain attached to terminal sulphydryl (SH) group of an acyl carrier protein (ACP), which passes them from one enzyme active site to the next.

The growing fatty acid chain is bound covalently to two types of sulphydryl groups (SH) during its synthesis. These groups, two in each polyprotein chain, serve as carriers of acyl groups:

(a) One of these, called the ***central thiol***, is contributed by ACP. The reactive unit of ACP is 4-phosphopantetheine (Pan) (Fig. 11.16) that is covalently bound to a swinging arm, 2.0 nm long, carrying the acyl intermediates from one enzyme active site to another. In other words, it acts as mobile carrier for shuttling the growing chain to successive catalytic domains.

(b) The second important thiol group, called the ***peripheral thiol***, is contributed by a cysteine residue (Cys) of the enzyme *3-ketoacyl synthase* (E_3).

The two types of thiol groups lie in close proximity in the dimer, suggesting that they are present on separate (subunits) that interact in a head-to-tail manner.

Each subunit (or monomer) contains all the seven enzyme activities, but the actual **functional unit** consists of one-half of a monomer interacting with the complementary half of the other monomer. Thus, two fatty acid chains can be synthesized simultaneously.

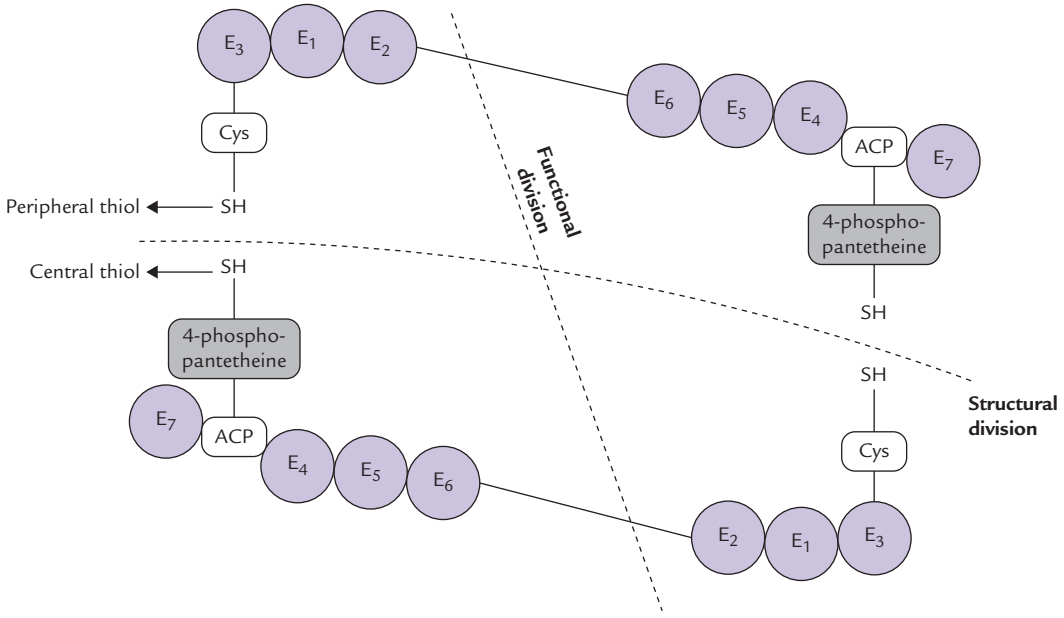


Fig. 11.15. The enzyme activities of fatty acid synthesis in higher organisms are present in a single, multifunctional polypeptide chain (as a diamer) called *fatty acid synthase*. The two subunits of this dirner interact in head-to-tail manner. Each subunit contains seven distinct enzyme activities (E_1 to E_7) and an acyl carrier protein (ACP). Cys = cysteine.

BOX 11.1

Evolutionary Origin of Fatty Acyl CoA as a Multifunctional Protein

No other enzyme in eukaryotes exhibits so remarkable multifunctionality as *fatty acid synthase*. Such an aggregation of different enzyme activities into a single multienzyme functional unit is, however, not seen in prokaryotes. Studies with cell-free extracts of *E. coli*, have shown that the reactions of the biosynthetic pathway are catalyzed by independent enzymes. Individual enzymes with these activities occur in chloroplasts also (in plants, fatty acid synthesis does not occur in the cytosol).

Such independent enzymes are not found in yeast, where the fatty acid synthase is a cytosolic, 2500 kD multifunctional enzyme. Its subunit composition is $\alpha_6\beta_6$. In eukaryotes it is much smaller (534 kD), consisting of only two polypeptide chains, though it contains all the enzyme activities needed for synthesizing fatty acids. A comparative analysis in different organisms suggests that evolution of *fatty acid synthase* as a multifunctional protein probably occurred by the joining of previously independent enzymes.

Fig. 11.16. The 4-phosphopantetheine group of the acyl carrier protein (ACP). It contains pantothenic acid and cystamine (shown in colour) with a terminal SH group, and thus resembles coenzyme A.

Fatty acid synthesis occurs in cytosol through involvement of a polyprotein (*fatty acid synthase*) that has seven enzymatic sites (in which successive reactions take place). The intermediates are linked to the terminal sulphydryl group of the phosphopantetheine reactive unit in ACP.

The reactions catalyzed by the seven domains of *fatty acid synthase* are described below. The first two reactions (called **priming reactions**) ensure that both the sulphydryl groups are 'loaded' with correct acyl groups, and the remaining reactions called **elongation steps** are involved in actual building of the fatty acid chain. The sequence of reactions is shown in Figure 11.17.

Priming Reactions

The two priming reactions are catalyzed by enzymes *acetyl transferase* and *malonyl transferase*.

1. ***Acetyl transferase:*** This enzyme catalyzes transfer of an acetyl group from acetyl CoA to the cysteine – SH group, i.e. the peripheral thiol group furnished by E_3.
2. ***Malonyl transferase:*** This enzyme catalyzes transfer of a malonyl group from malonyl CoA to the central thiol group furnished by ACP.

At this state, both central and peripheral-thiol groups are charged by malonyl and an acetyl group respectively (Fig. 11.16). Since the two acyl groups are close to each other on the enzyme complex, the stage is now set for their condensation and further elongation steps.

Elongation Steps

Condensation, reduction, dehydration and reduction are the four elongation steps.

1. ***Condensation by 3-Ketoacyl ACP synthase:*** The condensation reaction between the previously attached acetyl group and the malonyl group forms 3-ketoacyl ACP. This frees the peripheral thiol group of E_3 that had been occupied by acetyl CoA. A CO_2 molecule is liberated from the malonyl group during condensation, and this (decarboxylation) provides the "pull" for the reaction.
2. ***Reduction by 3-Ketoacyl ACP reductase:*** The keto group is reduced by this enzyme to form D-3-OH-acyl ACP.

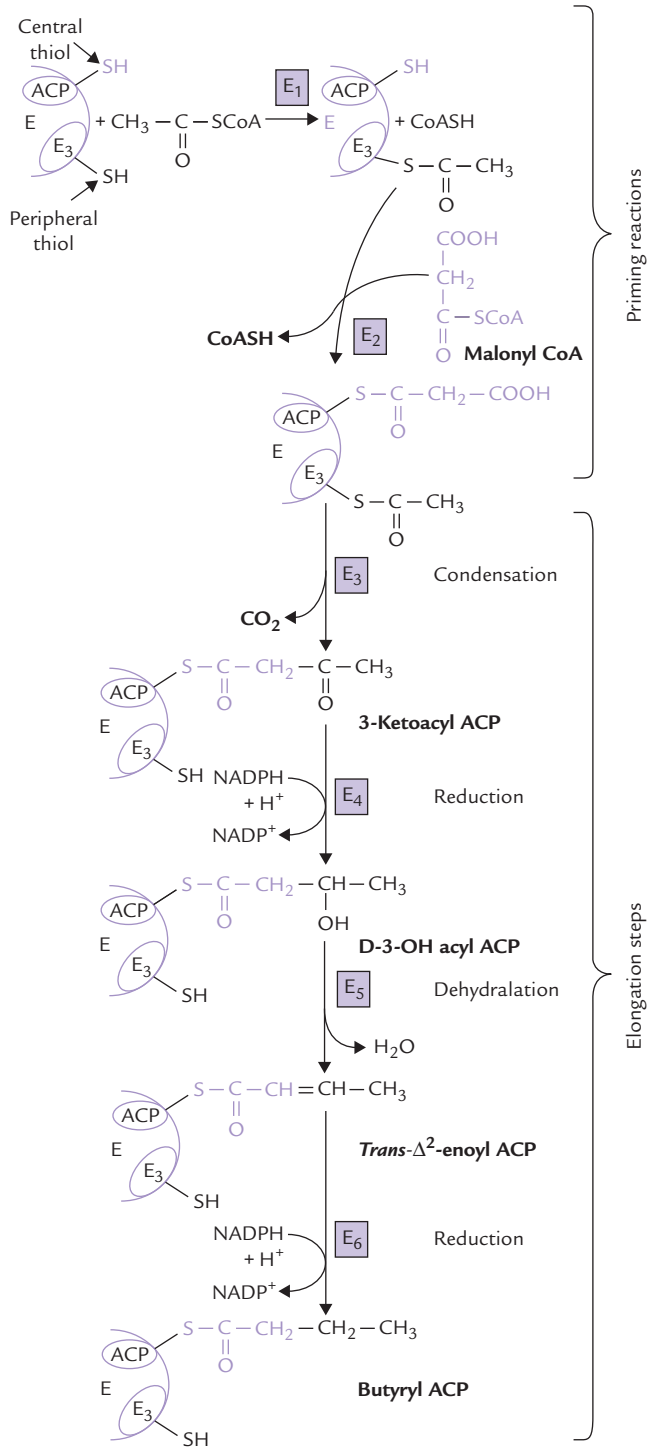


Fig. 11.17. Reactions in a single cycle of fatty acid synthesis (E_1 = *acetyl transferase*, E_2 = *malonyl transferase*, E_3 = *3-ketoacyl ACP synthase*, E_4 = *ketoacyl ACP reductase*, E_5 = *D-3-OH-acyl ACP dehydratase*, E_6 = *trans-Δ^2-eonyl-ACP reductase*).

The reducing power for this reaction is provided by NADPH.

3. ***Dehydration by D-3-OH-acyl ACP dehydratase:*** A water molecule is removed from the D-3-OH acyl ACP to form *trans*-Δ^2-enoyl ACP.
4. ***Reduction by Trans-Δ^2-enoyl-ACP reductase:*** The *trans*-Δ^2-enoly-ACP, formed in the previous reaction, is reduced by this enzyme to form butyryl ACP. NADPH serves as a coenzyme in this reaction also.

This final reaction of elongation produces the four carbon butyryl-ACP; thus the acyl group has grown longer by two carbon atoms.

All the above enzymatic steps are carried out in different enzyme domains of fatty acid synthase with the flexible arm of the phosphopantetheine group moving the acyl group from one active site to the next. When it gets to the last part of the sequence, the new acyl group (here butyrate) is passed to the peripheral thiol-SH group of cysteine residue (that was previously occupied by acetyl CoA) on the other monomer. The central thiol group of ACP is now reloaded with malonyl group and the elongation cycle is repeated in the same way, but the acyl group is now longer by two carbon atoms.

Further cycles: The elongation steps described above, are repeated with malonyl ACP adding 2-carbon units in each cycle. It is noteworthy that pairs of carbons from the malonyl group are added at the carboxyl (C-1) end and not the methyl (ω) end of the growing chain. This continues until the acyl group is 16-carbon long (palmitoyl-ACP). This is not a substrate for the condensing enzyme (E_3), but for *thiolase* (*deacylase*; E_7), which catalyzes the hydrolysis of palmitoyl-ACP. As a result, release of palmitic acid from the *fatty acid synthase* occurs and HS-ACP is regenerated. This completes synthesis of palmitic acid.

Cycle summary: Only two carbons in palmitic acid come directly from acetyl CoA. The remaining 14 are obtained from malonyl CoA, which in turn is produced from acetyl CoA. The overall stochiometry for the synthesis of palmitate therefore is

$$8\text{Acetyl CoA} + 14\text{NADPH} + 7\text{ATP} + 6\text{H}^+ \longrightarrow \text{Palmitic acid} + 14\text{NADP}^+ + 7\text{ADP} + 7\text{P}_i + 7\text{H}^+ + 8\text{CoASH} + 6\text{H2O}$$

The coenzyme in lipogenesis is NADPH (not NAD^+), which differs from NAD^+ in being phosphorylated at C-2 of ribose. In general, *$NADP^+$* is associated with synthetic pathways and NAD^+ is associated with degradative pathways.

Palmitic acid is produced in *liver* and *adipose tissue;* but in *lactating mammary gland, medium chain fatty acids*, 10-carbon capric acid, and 12-carbon lauric acid, are produced. Human milk contains both; cow's milk by contrast contains odd numbered fatty acids.

Table 11.2. Comparative features of β-oxidation and fatty acid synthesis

Location	β-oxidation	Fatty acid synthesis
	Mitochondrial matrix	Cytosol
Carbon units	Acetyl CoA	Malonyl CoA
Polyprotein participation	No	Yes
Coenzyme(s)	NAD^+, FAD	NADPH
Acyl group carrier	CoA	ACP
CO_2 participation	No	Yes
Stereoisomeric form of 3-OH-acyl group	L	D

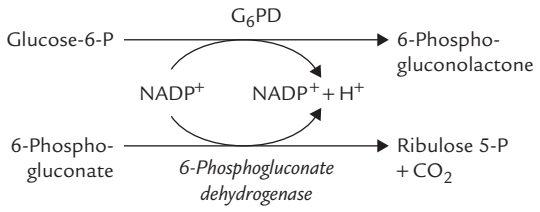
Comparison of Fatty Acid Synthesis and Oxidation

Fatty acid synthesis and β-oxidation are distinct pathways, and not a simple reversal of each other. Comparative features of the two are outlined in Table 11.2.

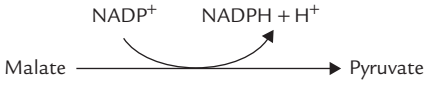
Sources of NADPH

NADPH, which plays a key role in fatty acid synthesis, is generated from various sources.

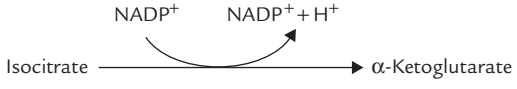
1. *Reactions of HMP shunt:* Two of the reactions of this pathway yield an NADPH each. These reactions are catalyzed by the enzymes *glucose 6-phosphate dehydrogenase* (G_6PD) and *6-phosphogluconate dehydrogenase*, respectively.



2. *Malate to pyruvate conversion* also yields a NADPH molecule. The reaction is catalyzed by the *malic enzyme.*



3. *Isocitrate to α-ketoglutarate conversion* by the cytosolic *isocitrate dehydrogenase* (IDH) is a minor source of NADPH. In fact, this enzyme can use both NAD^+ and $NADP^+$ as coenzymes.



D. Regulation of Fatty Acid Biosynthesis

Availability of Citrate in Cytoplasm

The citrate formed in mitochondria has a choice between progressing further in the TCA cycle or diffusing into cytoplasm, where it can supply the acetyl CoA precursor for synthesis of fatty acids. As discussed in Chapter 9, in high-energy states, when NADH:NAD^+ ratio is high and ADP low, the pathway for progression of citrate into the TCA cycle is blocked and hence its concentration builds up in the mitochondrial matrix. This excess citrate is then passed into the cytoplasm, where it supplies the acetyl CoA precursor. Citrate also allosterically stimulates the key enzyme of lipogenesis, *acetyl CoA carboxylase*.

The key control point of fatty acid synthesis is *acetyl CoA carboxylase*, which is regulated by citrate—a positive allosteric modulator.

Regulation of Enzyme Activities

Acetyl CoA carboxylase: Fatty acid synthesis is regulated at the *acetyl CoA carboxylase* step by allosteric mechanism, covalent modulation and induction-repression mechanism of enzyme synthesis. This provides a good example of ***control at committed step*** of a pathway.

(a) **Allosteric modulation:** *Acetyl CoA carboxylase* can exist in an inactive monomeric form (MW 400 kDa) and an active polymeric form (MW 6000–8000 kDa). The most important allosteric regulators are *citrate*, which stabilizes the active polymeric form, and *palmitoyl-CoA*, which dissociates the complex into inactive monomers. Thus, *citrate has stimulatory effect and palmitoyl CoA has inhibitory effect*. This is entirely logical: under conditions of high citrate concentration, energy storage is desirable, but decreased synthesis of fatty acids is appropriate when the product (palmitoyl CoA) accumulates.

(b) **The covalent mechanism** involves hormone-dependent *protein kinase A/phosphatase*: phosphorylation inhibits the enzyme activity and dephosphorylation activates the enzyme activity. These effects are under influence of hormones—insulin promotes dephosphorylation and thereby stimulates the *acetyl CoA carboxylase* activity. Glucagon or epinephrine, on the other hand, inactivates the enzyme by cAMP-dependent phosphorylation. Thus, insulin promotes fatty acid synthesis, while epinephrine and glucagons inhibit it (Fig. 11.18).

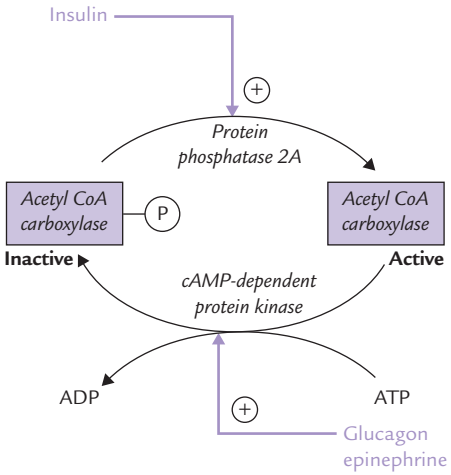


Fig. 11.18. Hormonal regulation of *acetyl CoA carboxylase*, the key lipogenic enzyme, by covalent mechanism.

(c) **Induction-repression** enzyme synthesis effected by diet: synthesis of *acetyl CoA carboxylase* is up regulated with high carbohydrate/low fat intake (insulin:glucagon ratio is high) and down regulated by high-fat low-carbohydrate diet and starvation (low insulin : glucagon ratio).

Finally, activity of *acetyl CoA carboxylase* is subject to control by energy charge of cell. Short-term regulation is effected by allosteric and covalent modulation, and a long term mechanism relates to induction or repression of enzyme activity.

The key control point of fatty acid synthesis is *acetyl CoA carboxylase*, which is inactivated by phosphorylation (by a cAMP dependent *protein kinase*), and reactivated by dephosphorylation (by a *protein phosphatase 2A*). It is allosterically stimulated by citrate and inhibited by palmitoyl CoA.

Fatty acid synthase: In common with *acetyl CoA carboxylase system*, *fatty acid synthase* activity also is similarly affected by regulatory mechanisms, such as, allosteric effect, and induction-repression of the enzyme. For instance, high carbohydrate low fat diet increases synthesis of *fatty acid synthase* and conversely low carbohydrate/high fat diet or fasting decreases the synthesis.

These regulatory mechanisms ensure that the *excess dietary carbohydrates are converted to fatty acids in liver*. The latter are esterified to triacylglycerols, which are released as constituents of very low density lipoprotein. This carbohydrate-to-fat pathway is important only on a high carbohydrate diet, as the regulated enzymes are stimulated by insulin. This explains why *people with carbohydrate-based diets are prone to develop obesity, even when they restrict fat intake*.

The syntheses of fatty acids takes place when carbohydrates and energy are plentiful and when fatty acids are scarce.

V. Chain Elongation and *Desaturase* Systems

Fatty acids of longer chain (above 16-carbon) length cannot be synthesized by *fatty acid synthase* system. But these fatty acids are important for the body: they are required in triacylglycerols and membrane lipids, which have fatty acids of chain lengths of up to 24 or 26 carbons, (Carbon-18 fatty acids being the commonest). Also only approximately 50% of the fatty acids of our body are saturated; another 40% are monounsaturated and approximately 10% are polyunsaturated. These long chain and unsaturated fatty acids are synthesized respectively by chain elongation system and *desaturase* systems.

A. The Chain Elongation System

Chain elongation beyond C-16 palmitate is possible in endoplasmic reticulum (the microsomes) and in the mitochondrion.

The **microsomal system** elongates palmitate by the addition of 2-carbon fragments derived from malonyl CoA. The system of reactions during the chain elongation is similar to that involved in fatty acid synthesis, except that here the fatty acid is attached to CoA, instead of the pantetheine residue of acyl-carrier protein. **Mitochondrial elongation system** is less active and uses acetyl CoA as a donor of carbons. It is NADH-dependent and the substrates for chain elongation are short- and medium-chain fatty acids containing fewer than 16 carbon atoms. Elongation of fatty acids is greatly reduced during fasting and starvation.

B. The *Desaturase* System

Monosaturated fatty acids can be obtained by introduction of double bond in the corresponding saturated fatty acids by a membrane bound *desaturase system* in the ER of the liver and other organs. The bond is most commonly introduced at position Δ^9 (between C-9 and C-10) of palmitic or stearic acid, producing palmitoleic acid or oleic acid, respectively. The system for desaturation requires molecular oxygen, NADH, and cytochrome B_5. The *desaturases* are *monooxygenases*: molecular oxygen is the oxidant and the process of desaturation results in the oxidation of both the fatty acid and the NADH.

$$\text{Saturated fatty acyl CoA} + \text{NADH} + \text{H}^+ + \text{O}_2 \longrightarrow \text{Monounsaturated fatty acyl CoA} + \text{NAD}^+ + 2\text{H}_2\text{O}$$

Polyunsaturated fatty acids may be formed by introduction of additional double bonds between the first double bond and the carboxyl end of the fatty acid, but not beyond Δ^9. Hence, linoleic acid (having two double bonds at the positions Δ^9 and Δ^{12}) cannot be synthesized from the monounsaturated oleic acid, but the fatty acid having two double bonds at Δ^6 and Δ^9 (called γ-linoleic acid) can be synthesized.

Most fatty acids can be synthesized from palmitate by *desaturase* and *elongase* enzymes, which introduce double bonds and lengthen the existing acyl chains, respectively.

Some polyunsaturated fatty acids are nutritionally ***essential***. Because of the inability of the human *desaturase* system to introduce double bonds beyond Δ^9, ***linoleic acid*** (C-18, $\Delta^{9,12}$) and **α-linolenic acid** (C-18, $\Delta^{9,12,15}$)—the parent compounds of the ω^6 and the ω^3 classes of polyunsaturated fatty acids—cannot be synthesized in the human body. They must be present in the diet or synthesized from other fatty acids in the diet, hence called **essential fatty acids**. Arachidonic acid (ω^6 20-C, $\Delta^{5,8,11,14}$) can be formed if the dietary supply of linoleic acid is adequate. The reaction sequence requires a combined action of the *desaturase* and chain elongation systems (Fig. 11.19).

1. First linoleic acid is converted to γ-linolenic acid (C-18; $\Delta^{6,9,12}$) by introduction of a double bond at C-6.

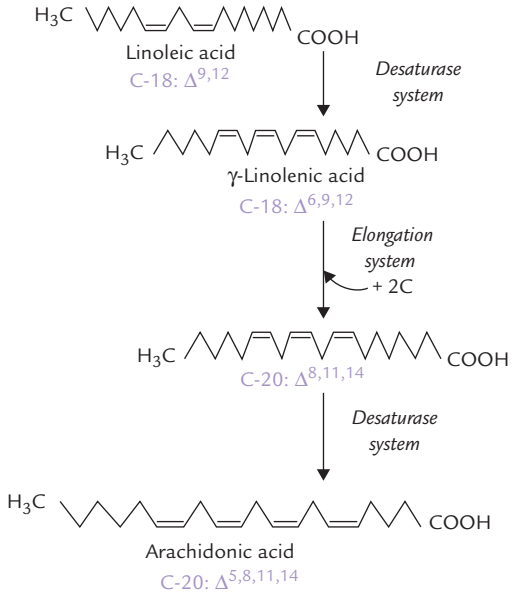


Fig. 11.19. Arachidonic acid can be synthesized from linoleic acid in the human body by the elongation and the *desaturase* systems.

2. Conversion of the γ-linolenic acid to arachidonic acid (C-20; $\Delta^{5,8,11,14}$) occurs by combined action of the elongation and *desaturase* systems, as shown here.

Thus, arachidonic acid, generally regarded as an essential fatty acid, can be synthesized provided linoleic acid is available. Therefore, strictly speaking *arachidonic acid is not an essential fatty acid*.

VI. Metabolism of Triacylglycerol

A. Synthesis of Triacylglycerols

Triacylglycerols are synthesized mainly in liver, adipose tissue and intestinal cells. There are two pathways: the *general pathway*, which is operational in liver, adipose and other organs where fatty acid biosynthesis occurs; and the *intestinal pathway*, which is responsible for resynthesis of triacylglycerols following digestion and absorption of dietary fats.

General Pathway

Glycerol-3-phosphate serves as the primary starting material and phosphatidic acid is the key intermediate of the pathway which proceeds in the following steps:

- Synthesis of phosphatidic acid from glycerol-3-phosphate and fatty acyl CoA.
- Synthesis of triacylglycerol from phosphaitidic acid.

1. ***Synthesis of phosphatidic acid:*** Glycerol-3-phosphate (glycerol-3-P) is produced in most tissues, including adipose, by reduction of the glycolytic intermediate, dihydroxyacetone phosphate (DHAP). In liver and kidneys, glycerol-3-P can also be formed directly via phosphorylation of glycerol by a specific *kinase* (*glycerol kinase*).

   It is then acylated by transfer of two long chain fatty acids from fatty acyl CoA to the hydroxyl groups at C-1 and C-2 producing phosphatidic acid (Fig. 11.20). The first fatty acid, usually a saturated fatty acid, is linked to C-1 forming monoacyl-glycerol-P, also called lysophosphatidic acid; the prefix lyso indicates that one of the hydroxyl groups is not acylated. Then a second fatty acid, usually an unsaturated fatty acid, establishes an ester bond to C-2 to form phosphatidic acid. Two different enzymes catalyze these two acylation steps—*glycerol phosphate acyltransferase* (E) and *monoacyl glycerol acyltransferase (E′):*

$$\text{Fatty acyl CoA} + \text{Glycerol-3-P} \xrightarrow{E} \text{Monoacylglycerol-P} + \text{CoASH}$$

$$\text{Fatty acyl CoA} + \text{Monoacylglycerol-P} \xrightarrow{E} \text{Diacylglycerol-P} + \text{CoASH}$$

(Phosphatidic acid)

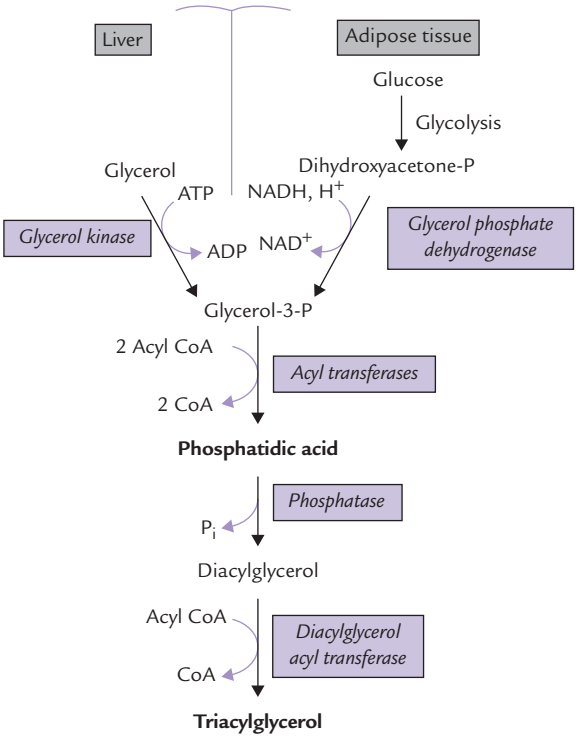


Fig. 11.20. Synthesis of TAG in liver and adipose tissue. The source of glycerol-3-P is different in the two tissues because there is no *glycerol kinase* in adipose tissue.

2. ***Synthesis of triacylglycerol from phosphatidic acid:*** Phosphatidic acid is first converted to diacyglycerol (DAG) by a specific *cytosolic phosphatase*, which remove a phosphate group. Transfer of an acyl group from fatty acyl CoA to DAG by the enzyme *1,2-diacylglycerol acyl transferase* (E″) finally generates triacylglycerol.

$$\text{Fatty acyl CoA} + \text{DAG} \xrightarrow{E'} \text{Triacylglycerol} + \text{CoASH}$$

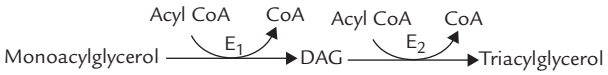
Note that the ester bonds of triacylglycerol are low energy in nature. *Fatty acyl CoA, the acyl group donor, on the other hand, has a high energy thioester.* Therefore, the acyl transfer reactions proceed with a decrease in the number of high energy bonds, and are therefore exergonic. This general principle provides rationale for understanding nearly all acyl transfer reactions in lipid metabolism.

Phosphatidic acid is synthesized from glycerol phosphate (which is mostly derived from glycolytic intermediate, dihydroxy-acetone phosphate) and fatty acyl CoAs. Acyl CoA molecules are added to glycerol phosphate to form first lysophosphatidic acid and then phosphatidic acid.

Intestinal Pathway

2-Monoacylglycerol, rather than glycerol-3-P is a precursor in intestinal epithelium. The pathway is simpler and

more straightforward. An intestinal *monoacylglycerol acyl transferase* (E_1) catalyzes the formation of DAG and *1,2-diacylglycerol acyl transferase* (E_2) catalyzes acylation of the remaining free carbon of DAG to yield a triacylglycerol.



B. Hydrolysis of Triacylglycerols (Lipolysis)

Lipases are the enzymes involved in degradation of triacylglycerols, catalyzing sequential removal of the fatty acids from glycerol backbone. Fatty acids are beta-oxidized to produce energy. The glycerol is converted to dihydroxyacetone phosphate which enters glycolysis.

Various types of *lipases* are known, most important of which are *hormone-sensitive lipase*, *lipoprotein lipase* and *hepatic lipase*. Their actions are described below:

1. *Hormone sensitive lipase* (HSL) acts on triacylglycerol stores in adipose tissue, catalyzing the first cleavage to remove the fatty acid esterified to C-1 of glycerol. *Lipases* specific for monoacylglycerol and diacylglycerol then cleave the fatty acids linked to C-2 and C-3.

$$\text{Triacylglycerol} \xrightarrow{\text{HSL}} \text{DAG}$$

2. *Lipoprotein lipase* (LPL) causes hydrolysis of triacylglycerols present in some lipoproteins (very low density lipoproteins and chylomicrons) and is thus required for utilization of circulating triacylglycerols. This enzyme is synthesized in adipose tissue, skeletal muscle, the myocardium, the lactating mammary gland, spleen, lung, kidney and aorta (bound noncovalently to heparan-sulphate proteoglycans of the endothelial plasma membrane). As the chylomicrons or very low density lipoproteins (very low density lipoproteins and chylomiarons) pass through the capillaries of these organs, they bind to *LPL*, and their triacylglycerol is hydrolyzed to free fatty acids and 2-monoacylglycerol. Most of these breakdown products diffuse directly into the tissue cells.
3. *Hepatic lipase* hydrolyzes triacylglycerols present in other lipoprotein fractions, e.g. intermediate- and high-density lipoproteins (described in Chapter 12).

Lipases release fatty acids from glycerol backbone. *Hormone sensitive lipase* degrades TAG stores in adipose tissue, whereas circulating TAGs are mostly degraded by *lipoprotein lipose*.

VII. Adipose Tissue Metabolism

Metabolic processes in the adipose tissue have consequences that extend beyond the tissue itself. For instance, the plasma level of free fatty acids is determined by these processes. The circulating free fatty acids in turn have profound effect upon the metabolism of other tissues, particularly of liver and muscles.

Source and determinant of the circulating free fatty acids are the triacylglycerol (TAG) molecules that are stored in the adipose tissue. TAGs account for approximately 90% weight of the adipose tissue. Also termed **depot fats**, the stored triacylglycerols form major fuel reserve of the body. These stores exist in a dynamic steady state, being continually degraded and resynthesized (Fig. 11.21); life span of a stored triacylglycerol molecule is only 2–3 days. The degradation of TAG, known as lipolysis, is catalyzed by enzymes called *lipases*. Lipolysis results in production of a glycerol and three fatty acid molecules from each triacylglycerol molecule. The fatty acid molecules thus produced are re-used for synthesis of triacylglycerol.

A. TAG Synthesis

Synthesis of TAG requires glycerol phosphate and three fatty acid molecules, activated to their CoA-thioesters.

- The fatty acids are mostly derived by the action of *lipoprotein lipase* on chylomicron TAGs, after a mixed meal.
- The glycerol phosphate is derived from adipose tissue by the *glycerol phosphate dehydrogenase* reaction. In this reaction, dihydroxyacetone phosphate, a glycolytic intermediate, is reduced in an NADH-dependent step (Fig. 11.22). The alternative reaction for the generation of glycerol phosphate (the phosphorylation of glycerol by the enzyme *glycerol kinase*) is important in

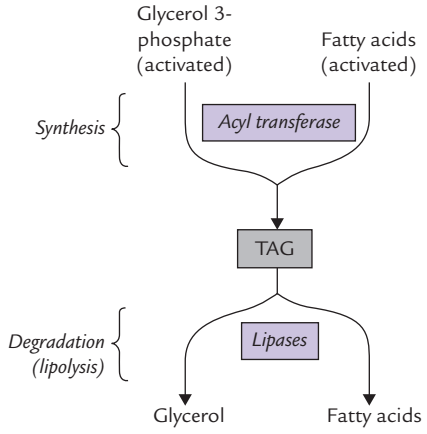


Fig 11.21. Diagrammatic representation of the reactions in an adipocyte.

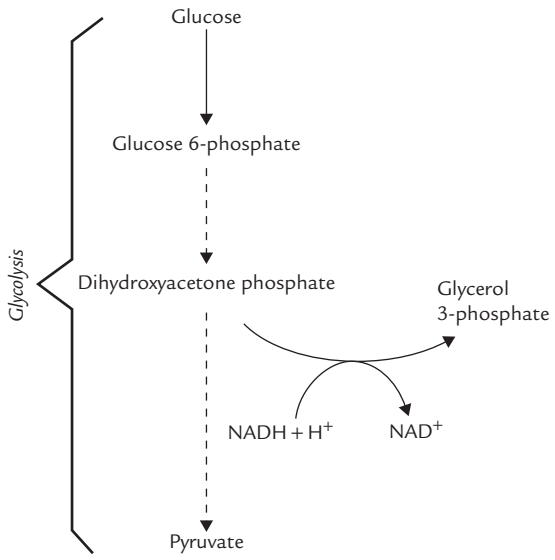


Fig. 11.22. Formation of glycerol 3-phosphate from glucose in adipose tissue.

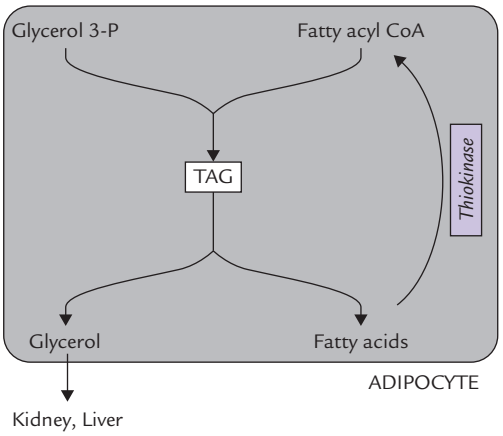


Fig. 11.23. Fatty acids are recycled for triacylglycerol (TAG) synthesis in adipose tissue but glycerol is metabolized in other tissues.

liver, but not in adipose tissue, because of extremely low concentration of this enzyme in adipocytes.

The acyl groups of the activated fatty acids are then sequentially transferred to glycerol 3-phosphate by *acyl transferases* to form triacylglycerol, as described earlier (Fig. 11.20).

- The first acyl group is linked to C-1 of glycerol phosphate by the enzyme *glycerol phosphate acyl transferase* ($\mathbf{E_1}$),
- The second acyl group is transfersed to C-2 of glycerol phosphate by *monoacylglycerol acyl transferase* ($\mathbf{E_2}$),
- The third acyl group is transfersed to C-3 by *diacylglycerol acyl transferase* ($\mathbf{E_3}$).

$$\text{Glycerol P} + 3\ \text{Fatty acyl CoA} \xrightarrow{E_1 - E_2 - E_3} \text{Triacylglycerol} + 3\ \text{CoASH} + \text{Pi}$$

B. TAG Hydrolysis (Lipolysis)

This process requires the *sequential removal of three fatty acids from glycerol*. The first cleavage which is **rate-limiting** is catalyzed by the enzyme *hormone-sensitive lipase* (HSL). Subsequent cleavages are brought about by other *lipases* which are specific for monoacylglycerol and diacylglycerols. A different fate awaits each of the two products of lipolysis: glycerol and fatty acids.

- **Glycerol** cannot be utilized within the adipocytes because low concentration of *glycerol kinase* in adipose tissue prevents its conversion to glycerol 3-phosphate in any substantial amount. Therefore, intracellular level of glycerol builds up. Subsequently, it diffuses into the plasma from where it is taken up by tissues, mainly liver and kidney. In these tissues, activity of *glycerol kinase* being high, the glycerol is phosphorylated to glycerol 3-phosphate which then forms dihydroxyacetone phosphate, a glycolytic intermediate. Thus, glycerol is ultimately used via the glycolytic sequence in other tissues (Fig. 11.23).
- **Fatty acids**, on the other hand, are used within the adipose tissue for TAG synthesis. Under normal circumstances, rates of TAG synthesis and hydrolysis are nearly balanced and most of the fatty acids obtained by lipolysis are activated to Co-thioesters which are recycled for TAG formation. However, imbalance between the two processes disturbs the dynamic steady state.

Adipose tissue is specialized for storage of triglycerides—a highly concentrated source of metabolic energy and also a source and determinant of the circulating free fatty acids. The latter are the most important energy source during fasting.

C. The Dynamic Steady State

TAG production and hydrolysis are not forward and reverse phases of the same reaction. They are entirely different processes using different reactants and catalyzed by different enzymes. Under normal circumstances, a dynamic steady state exists where the rates of the two processes are balanced. Hormonal, nutritional and metabolic, factors regulate the metabolism of adipose tissue by influencing these two processes. Net balance of the two determines the amount of TAG stores in the adipose tissue, as also the levels of circulating free fatty acids in the plasma.

D. Hormonal Regulation of Adipose Tissue Metabolism

The metabolic processes operative within adipocytes are regulated by a number of hormones which control activities of both the key enzymes of adipose tissue metabolism, viz. the ***hormone-sensitive lipase*** and the ***glycerophosphate acyl transferase***.

The hormonal influences on the *lipase* are summarized in Figure 11.24. The enzyme activity is regulated by the **cAMP-dependent cascade**; activation occurs when the enzyme is phosphorylated by the cAMP-dependent *protein kinase-A* (PKA). It is the same enzyme that operates in regulation of glycogen metabolism.

c-AMP $\xrightarrow{(+)}$ PKA $\xrightarrow{(+)}$ *HSL*

Thus, the hormones which increase intracellular level of cAMP activate the *hormone sensitive lipase.* Such hormones are lipolytic.

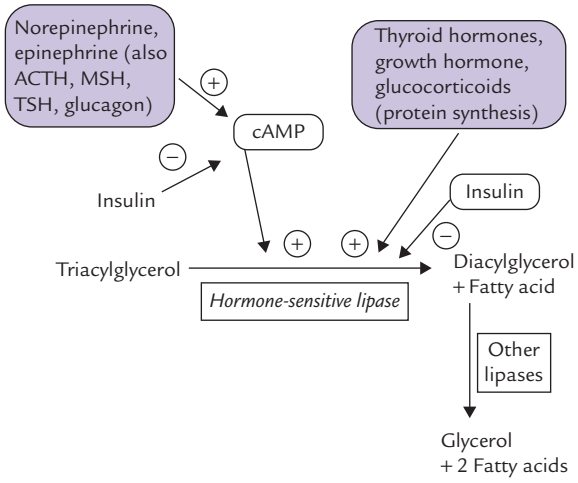


Fig. 11.24. Hormonal regulation of lipolysis in adipose tissue. Some lipolytic hormones activate HSL, others enhance synthesis of the enzyme-protein; and insulin antagonizes effects all lipolytic hormones.

The most important lipolytic agents are ***catecholamines*** which act through β-adrenergic receptors, cAMP and *protein kinase A (PKA)*. *Norepinephrine*, released from sympathetic nerve terminals in adipose tissue, is a more potent lipolytic agent than epinephrine. The catecholamines ensure that the TAGs are hydrolyzed during cold exposure, stress and physical exercise through stimulation of the *lipase* activity.

Catecholamines elevate intracellular level of cAMP, which allosterically activate *PKA*. The *PKA* in turn phosphorylates *hormone sensitive lipase*, activating it.

Glucocorticoids, growth hormone and the thyroid hormones enhance lipolysis by inducing synthesis of lipolytic proteins. Thus, glucocorticoids induce the *de novo* synthesis of the *hormone-sensitive lipase*, and thereby accelerate the lipolytic response to catecholamines. This effect shows regional differences in action of glucocorticoids (For additional information on lipolytic action of corticoids refer to Box 11.2).

Lipolytic hormones facilitate lipolysis either by direct stimulation of cAMP formation or by inducing synthesis of lipolytic protein (HSL).

Role of Insulin

In contrast to the lipolytic hormones-action of **insulin is antilipolytic**, i.e. it antagonizes these hormones by causing inhibition of the *hormone-sensitive lipase*. The mechanism of action is poorly understood. Either insulin decreases the level of cAMP in the cell (by activating a *phosphatase*) or induces dephosphorylation of the HSL. Adipose tissue is highly sensitive to insulin: a moderate rise in insulin level (during feeding) is able to halt lipolysis.

BOX 11.2

Regional Differences in Actions of Glucocorticoids Manifest Clinically

The lipolytic action of glucocorticoids is not uniform in different regions of the body: some areas are highly responsive whereas others are resistant to their action. For example, glucocorticoids effectively induce lipolysis in extremities but the triacylglycerol stores in trunk respond poorly to their lipolytic action. This regional difference shows clinically in patients with **Cushing's syndrome** (excess glucocorticoids). The striking features of this condition are truncal obesity and thin extremities. Rapid hydrolysis of the TAG stores in extremities by excessive glucocorticoids reduces adipose tissue mass, resulting in thin and wasted appearance of extremities. The mobilized fats are redistributed and get deposited in trunk, leading to truncal obesity, commonly referred to as the "buffalo hump".

Insulin enhances TAG synthesis: In addition to inhibiting lipolysis, insulin can also enhance TAG synthesis in adipocytes by several mechanisms:

1. *By increasing production of glycerol 3-phosphate substrate:* Insulin increases the rate of glycolysis, which enhances production of dihydroxy-acetone phosphate. The latter is then reduced to glycerol 3-phosphate.

   In adipose tissue, the most important control site for glycolysis is tissue up-take of glucose (unlike in other tissues where it is the PFK reaction). Insulin enhances the glucose uptake by increasing the number of functional glucose carriers in plasma membrane.
2. *By increasing synthesis of fatty acid substrate:* Insulin enhances lipogenesis by bringing about the following metabolic alternations. It increases production of acetyl CoA from glucose (by augmenting glycolytic sequence and *pyruvate dehydrogenase* complex activity), and acetyl CoA is the substrate for lipogenesis. Further, it stimulates activity of *acetyl CoA carboxylase*, the key enzyme of lipogenesis. Insulin favors generation of NADPH, the coenzyme required in lipogenesis, by promoting the pentose phosphate pathway.
3. *By increasing lipoprotein lipase activity:* Stimulation of activity of this enzyme results in increased production of fatty acids from the TAG fraction of lipoprotein (Chapter 12). This further increases the supply of fatty acid substrate.
4. *By increasing esterification of fatty acid and glycerol substrates:* Insulin induces the enzyme *glycerophosphate-acyl transferase* which adds the first fatty acid to glycerol phosphate in the bio-synthetic pathway of TAG.

Insulin ensures that triacylglycerols are synthesized after a meal, and degraded during fasting. Under its stimulatory influence, the adipose cells use the fatty acids (generated by *lipoprotein lipase*) for synthesis of storage triglycerides.

Thus, *insulin has a dual role—inhibition of lipolysis and stimulation of triacylglycerol production*. This helps to maintain a balance between these two pathways, under normal circumstances. In **diabetes mellitus**, loss of this balance occurs. Consequently, lipolysis is accelerated, but triacylglycerol synthesis slows. The net result is that the rate of lipolysis now exceeds that of triacylglycerol synthesis. Quantity of fatty acids produced is far in excess than that can be reutilized for re-esterification. The surplus flows out into the circulation to elevate the plasma levels of free fatty acids. This explains *a marked elevation of plasma free fatty acid levels in diabetes mellitus*.

E. Effect of Nutritional States on Adipose Tissue Metabolism

The metabolic events in adipocytes show dramatic changes in different nutritional states. TAG synthesis depends on the supply of glycerol phosphate, which is synthesized from glucose via dihyroxyacetone phosphate (Fig. 11.22). *The insulin-dependent uptake of glucose into the cell is the most important rate-limiting step in this pathway*. Thus, in the fed state, increased availability of glucose and elevated insulin level in serum enhance TAG synthesis. Conversely, during starvation synthesis of TAG is impeded.

Utilization of chylomicron for providing fatty acids is also altered in different nutritional states. The utilization depends on the enzyme *lipoprotein lipase* which is induced by insulin in adipose tissue (but not in other tissues like cardiac and skeletal muscles). Therefore, the lipoprotein triacylglycerols are routed preferentially through adipose tissue in the well-fed state but not during fasting.

Insulin ensures that TAGs are synthesized after a meal and degraded during fasting. Fatty acids are major fuel for most tissues during fasting. Glycerol, which is released from adipose tissue together with fatty acids is a substrate for gluconeogenesis. This has a role in maintenance of normal glucose levels in blood during fasting and starvation, which is essential for survival.

Adipose tissue is the main storehouse of triacylglycerol in the body, mobilization of which is regulated by a variety of intricate mechanisms, including hormonal regulation.

VIII. Metabolism of Complex Lipids

Complex lipids comprise a diverse group of saponifiable compounds found in biological membranes. They may be **polar** (phospholipids, sphingolipids) or **non-polar** (cholesterol esters, triacylglycerols). In this section metabolism of two major classes of the polar lipids, phospholipids and sphingolipids, are discussed; metabolism of non-polar lipids are discussed elsewhere.

A. Metabolism of Phospholipids

The term phospholipids generally refers to two types of compounds: (a) **glycerophospholipids**, that contain a glycerol backbone, and (b) **sphingomyelins**, that contain an amino alcohol sphingosine instead of glycerol (the latter are discussed along with sphingolipids).

B. Glycerophospholipid Metabolism

The commonest glycerophospholipids are phosphatidylcholine, phosphatidylethanolamine, and phosphatidylserine. A number of more complicated structures, such as phosphatidylglycerol or cardiolipin, found in inner mitochondrial membrane; and phosphatidylinositol, which has a role in signal transduction and in anchoring proteins in the plasma membranes, is also classed as glycerophospholipid (Chapter 3).

Biosynthesis of Glycerophospholipids

Glycerophospholipids are synthesized from phosphatidic acid and diacylglycerol, the intermediates in production of triacylglycerol (Fig. 11.20). The same set of enzymes is used till diacylglycerol. These enzymes are located in the endoplasmic reticulum.

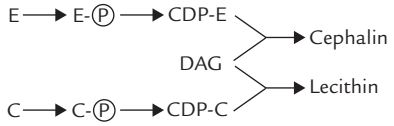
Synthesis of Lecithin and Cephalin

It involves transfer of choline/ethanolamine to the diacylglycerol, from an **activated donor**, e.g. CDP-choline or CDP-ethanolamine. How is the activated donor synthesized? Synthesis of CDP-choline occurs in a series of reactions given below (Fig. 11.25):

- Choline is first converted to phosphocholine, and then activated to CDP-choline by a *pyrophosphorylase* reaction.
- The pyrophosphate bond is cleaved to drive forward the reaction.

The CDP-choline then transfers its phosphocholine group to diacylglycerol (DAG) to form **phosphatidylcholine (*lecithin*)**. The last reaction is analogous to the transfer of glucose from UDP-glucose during glycogenesis.

Phosphatidylethanolamine (*cephalin*) is generated by a similar pathway using CTP and phosphoethanolamine to form CDP-ethanolamine; the latter transfers its phosphoethanolamine group to DAG to form phosphatidylethanolamine (Fig. 11.25).



E = ethanolamine; **C** = choline; **DAG** = diacylglycerol

In biosynthesis of lecithin or cephalin, the head group (i.e. choline or ethanolamine) is activated by CTP first, and then transferred to DAG.

Synthesis of Phosphatidylserine

Phosphatidylethanolamine can react with free serine to form phosphatidylserine and the free base, ethanolamine (Fig. 11.25). Likewise, phosphatidylcholine can form phosphatidylserine by such an **exchange reaction.**

(Liver has another route to phosphatidylethanolamine, involving decarboxylation of phosphatidylserine by a specific *mitochondrial decarboxylase*.)

Synthesis of Cardiolipins or Phosphatidylinositol

In contrast to the pathways of lecithin and cephalin synthesis, where the head group (choline or ethanolamine)

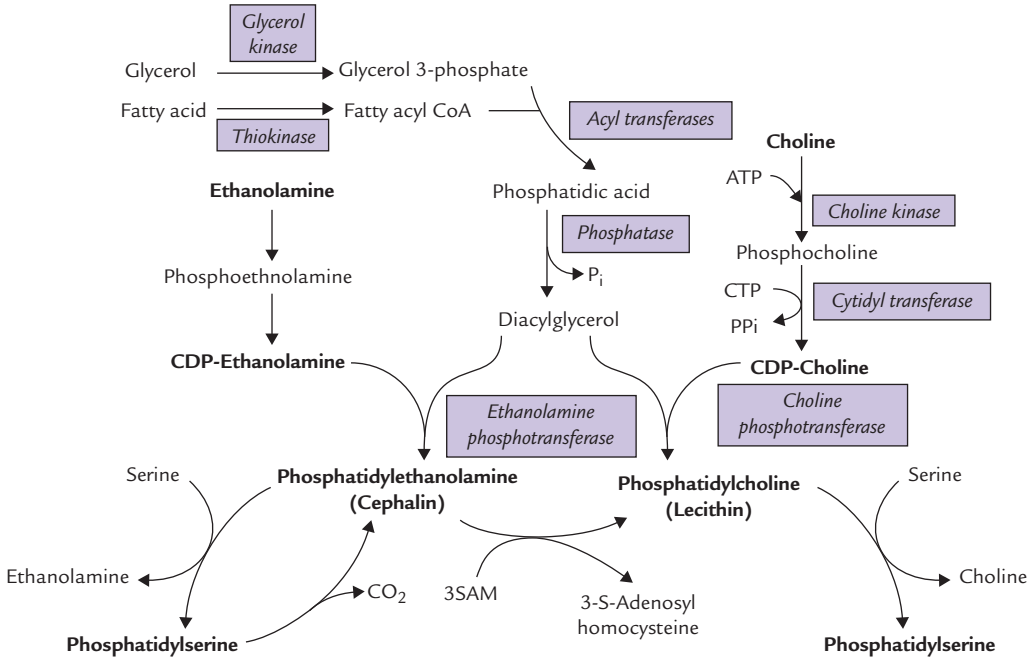


Fig. 11.25. Synthesis of phospholipids (SAM = S-adenosyl methionine, CTP = cytidine triphosphate, CDP = cytidine diphosphate).

was activated by cytidine nucleotide, an alternate pathway exists for synthesis of other phospholipids (e.g. cardiolipin and phosphatidylinositol). In this pathway, phosphatidic acid is activated, rather than head group, yielding CDP-diacylglycerol (CDP-DAG).

The CDP-DAG is activated donor. It can readily transfer the phosphatidic acid group to free glycerol or inositol to form phosphatidylglycerol or phosphatidylinositol, respectively. A second phosphatidic acid may also be likewise added to phosphatidylglycerol to form diphosphatidylglycerol or cardiolipin (Fig. 11.27).

Others

In a secondary pathway, phosphatidylcholine is formed by transfer of methyl groups to phosphatidylethanolamine with the methyl donor, *S-adenosylmethionine* (SAM). The methylation pathway involves the sequential transfer of three activated methyl groups from SAM (Fig. 11.26).

De novo and salvage pathways: The above description makes it clear that there are two ways to synthesize a glycerophosphate. First, the head group is synthesized anew (this is *de novo* pathway). Second, the head group (choline or ethanolamine) is not synthesized anew, rather preformed choline and ethanolamine are reutilized or **salvaged**.

As depicted in Figure 11.25, when (i) phosphatidylethanolamine is derived from phosphatidylserine by decarboxylation (Fig. 11.25), or (ii) when phosphatidylcholine is derived from phosphatidylethanolamine by methylation pathway (Fig. 11.26), new ethanolamine and choline moieties, respectively are formed. Therefore, there pathways are the **de novo pathways**. However, ethanolamine and choline *per se* are the starting substances in some reaction sequences depicted in the diagram on page 228, 11.25 and are called **salvage pathways.**

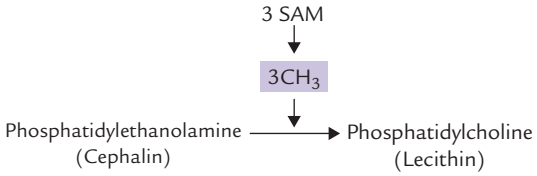


Fig. 11.26. Formation of lecithin by transfer of 3 methyl groups to cephalin, with the methyl donor S-adenosyl methionine (SAM).

Most cells are able to synthesize phosphoglycerides, sphingolipids and triacylglycerols. Phosphatidic acid, a key intermediate, is synthesized from glycerol phosphate and CoA activated fatty acids.

Degradation of Glycerophospholipids

The glycerophospholipids are in a continuous state of turnover in most membranes; the turnover especially increases in response to oxidative damage during inflammation, or in response to hormonal stimuli. As shown in Figure 11.28, a number of *phospholipases* are involved in degradation of phosphoglycerides. These enzymes are present in essentially all tissues and are classified according to their cleavage specificity: each one acts on a specific bond in the phospholipid structure.

Phospholipase A$_1$ and -***A***$_2$ remove the first and the second acyl group, respectively to produce the corresponding lysophospholipids. PLA_2 acts on phosphatidylinositol to liberate arachidonic acid for the synthesis of prostaglandins. It is particularly active during the inflammatory response. Snake venom is especially a rich source of PLA_2.

Phospholipase B (not shown) is a *phospholipase* that removes the second acyl group from lysophospholipid after the action of PLA_1 or PLA_2.

Phospholipase C (PLC) cleaves off the bond between the glycerol and phosphate to yield a phosphoryl base and a diacylglycerol molecule. It is especially active during signal transduction. Toxins isolated from *Clostridium* and other bacilli are rich sources of *PLC*.

Phospholipase D (PLD) liberates the polar base to form phosphatidic acid. It is found in plants, such as cabbage and cotton seeds.

The products obtained by degradation of glycerophospholipids enter the metabolic pool and are funneled into various important pathways.

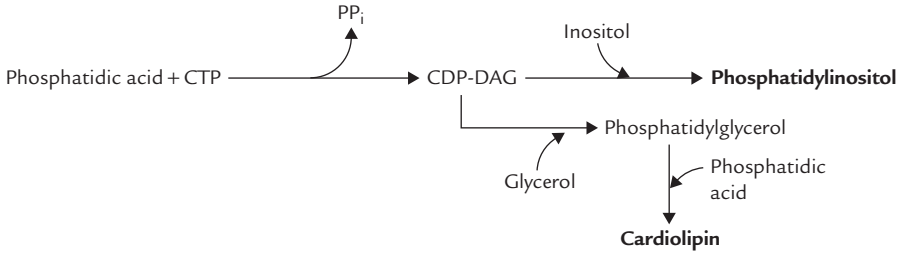


Fig. 11.27. Alternate pathways for synthesis of cardiolipin and phosphatidylinositol (CDP-DAG = cytidine diphosphate diacylglycerol, CTP = cytidine triphosphate).

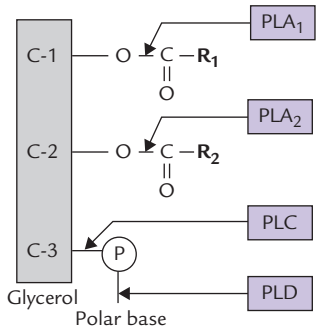


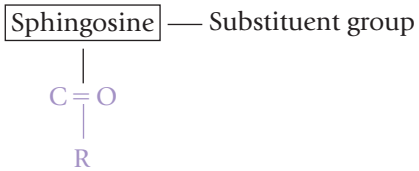
Fig. 11.28. Enzymatic degradation of phospholipid (PLA_1 = *phospholipase* A_1, PLA_2 = *phospholipase* A_2, PLC = *phospholipase* C, PLD = *phospholipase D*, (P) = phosphate group, R_1 and R_2 = acyl groups).

Finally, *phospholipases* are used not only for the final degradation of the lipid molecule, but also for the **remodelling** of phosphoglycerides. The fatty acids in positions 1 and 2 for example, can be exchanged by hydrolysis and reacylation to form a new molecule. The cleavage specificities of *phospholipases* are largely responsible for the presence of specific fatty acids in positions 1 and 2 of phosphoglycerides : C-1 is usually occupied by a saturated fatty acid, and the fatty acid in position 2 is usually unsaturated.

The phospholipids are remodelled and eventually degraded by various *phopholipases*.

C. Sphingolipid Metabolism

Sphingolipids are a complex group of amphipathic polar lipids that are built on a core structure of the long chain (18-C) amino alcohol, **sphingosine**. They are divided into three general classes: sphingomyelins, cerebrosides and gangliosides, which differ with respect to the substituents attached to the C-1 hydroxyl group of sphingosine (Chapter 3).



Ceramide = Sphingosine + Fatty acid (R.COOH)

Sphingomyelins contain phosphoryl base, cerebrosides contain a monosaccharide, and gangliosides contain an oligosaccharide unit.

Biosynthesis of Sphingolipids

The pathway consists of two stages: firstly, synthesis of sphingosine/ceramide, and secondly, conversion of ceramide to any of the above classes of sphingolipids.

Synthesis of sphingosine and ceramide: Sphingosine is synthesized in the ER of most cells. The carbon atoms that contribute to sphingosine (C-18) synthesis come from palmitoyl CoA (C-16) and serine (which contributes two carbons). The condensation reaction requires reductive power of NADPH and is followed by a series of reactions to ultimately yield sphingosine.

$$\text{Palmitoyl CoA} + \text{Serine} + \text{NADPH} + H^+ \dashrightarrow \text{Sphingosine} + CO_2 + \text{CoASH} + NADP^+$$

Ceramide, the next important intermediate, is formed by N-acylation of sphingosine:

$$\text{Sphingosine} + \text{Fatty acyl CoA} \rightarrow \text{Ceramide} + \text{CoASH}$$

Formation of three classes of sphingolipids: Subsequent conversion of ceramide to sphingolipids occurs in ER and Golgi apparatus. It involves transfer of a substituent group to ceramide, usually from an activated donor, e.g. a nucleotide sugar (UDP glucose) or CDP choline (Fig. 11.29).

(a) ***Conversion of ceramide to sphingomyelin:*** Transfer of phosphocholine to the terminal hydroxyl group of ceramide produces sphingomyelin. It is the only sphingolipid that contains phosphorus. The formation of sphingomyelin from ceramide is unusual in that a high-energy precursor is not involved. The most important phosphocholine donor in this exchange reaction is phosphatidylcholine. The reactants and products are low-energy compounds. Alternatively, CDP-choline may act as the phosphocholine donor.

(b) ***Conversion of ceramide to cerebrosides:*** The donor of carbohydrate(s) of cerebrosides are energy-rich uridine diphosphate sugars. The monosaccharide residues are introduced from these activated precursors; for example, formation of galactocerebrosides (a cerebroside) involves a reaction of ceramide with UDP-galactose, and formation of glucocerebroside involves UDP-glucose.

(c) ***Conversion of ceramide to gangliosides:*** The oligosaccharide chains of these complex glycosphingolipids are constructed by the stepwise addition of monosaccharyl units, e.g. activated hexoses, hexosamine, and N-acetyl neuraminic acid (NANA). UDP serves as carrier of the sugar units in these reactions also. Each one of these exergonic reactions is catalyzed by an individual *glycosyltransferase*. Various gangliosides, e.g. GM_1, GM_2 and GM_3 are formed at different stages.

Sulphatides are the sulphate esters, formed by the combination of a sulphate with an alcohol, group present in glycosphingolipid. The donor of sulphate in biosynthetic reactions is 3′-phosphoadenosine-5′-phosphosulphate or **PAPS**. The standard free energy of hydrolysis of PAPS (to form 3′-phosphoadenosine monophosphate plus sulphate) is about –75 kJ/mol, which drives the biosynthetic reaction.

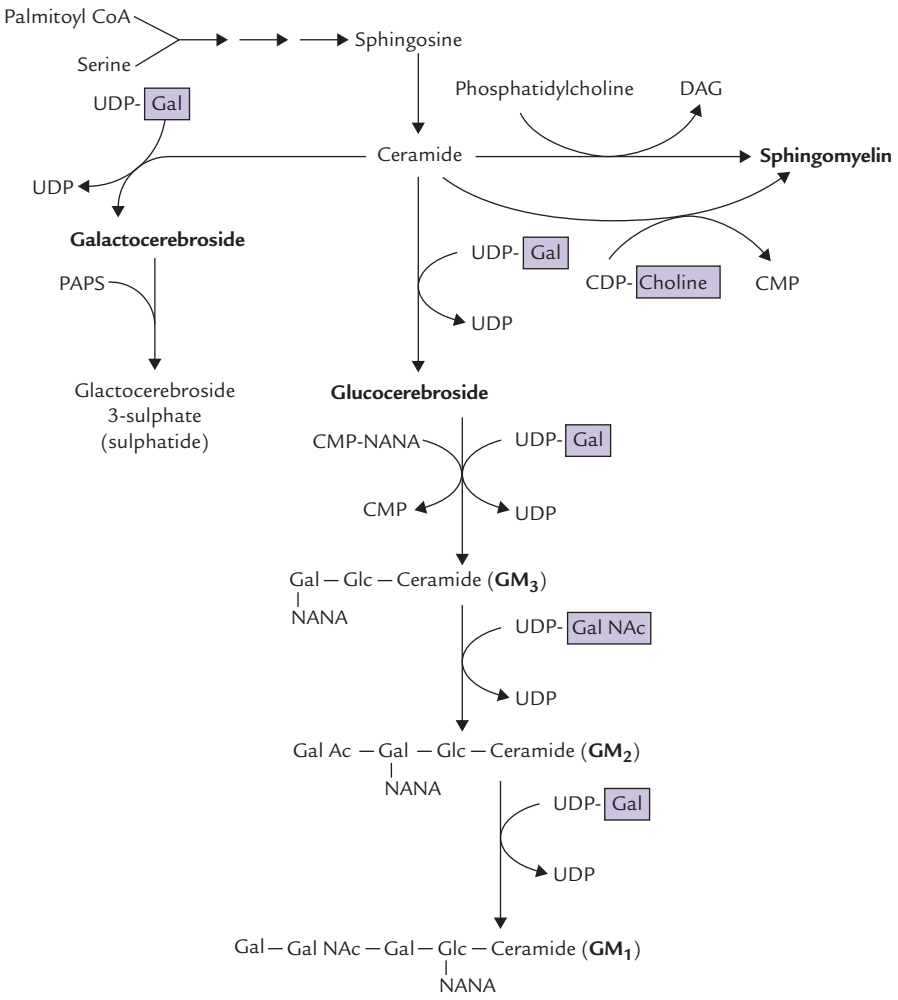


Fig. 11.29. Biosynthesis of sphingolipids from ceramide and nucleotide-activated monosaccharide or polar head group (e.g. choline) (PAPS = 3-phosphoadenosine-5-phosphosulphate, NANA = N-acetyl neuraminic acid or sialic acid, GM_1, GM_2, GM_3 = gangliosides. Gal NAc = N-acetyl galactosamine).

Globosides: When carbohydrate portion of the cerebroside contains two or more sugars plus an N-acetyl galactosamine group, it is referred to as globoside. The globosides are synthesized as gangliosides; UDP serving as carrier of the sugar units.

Glycosphingolipids are synthesized from ceramide by the addition of substituent glycosyl units usually donated by nucleotide sugars.

Degradation of Sphingolipids

Sphingolipids are degraded by lysosomal enzymes. In case of sphingomyelin, there are two enzymes: *sphingomyelinase*, which cleaves off the phosphocholine head group to form *ceramide*; and *ceramidase*, which further degrades the ceramide (Fig. 11.30).

In case of glycosphingolipids (cerebrosides and gangliosides), these degrading enzymes are specific *lysosomal exoglycosidases*, which remove terminal sugars from the oligosaccharide. They have narrow specificities for the monosaccharide and the type of bond they cleave. Figure 11.30 shows degradation of some important sphingolipids along with the associated inborn errors.

D. Sphingolipidoses

Partial or complete deficiencies of these sphingolipid-degrading enzymes result in lysosomal accumulation of the missing enzyme substrate, which is insoluble, non-degradable, and non-excretable. This either causes lysosomes to lyse and release their hydrolytic enzymes into a cell, or prevents the lysosome from functioning normally. In either case, the resulting disease is called a lipid storage disease **lipid storage disease** or **sphingolipidosis**. The enzyme deficiency is expressed in all tissues but clinical effect is most obvious in nervous tissue because of its high sphingolipid content and turnover. Enlarged liver and spleen (hepatosplenomegaly) is also common because

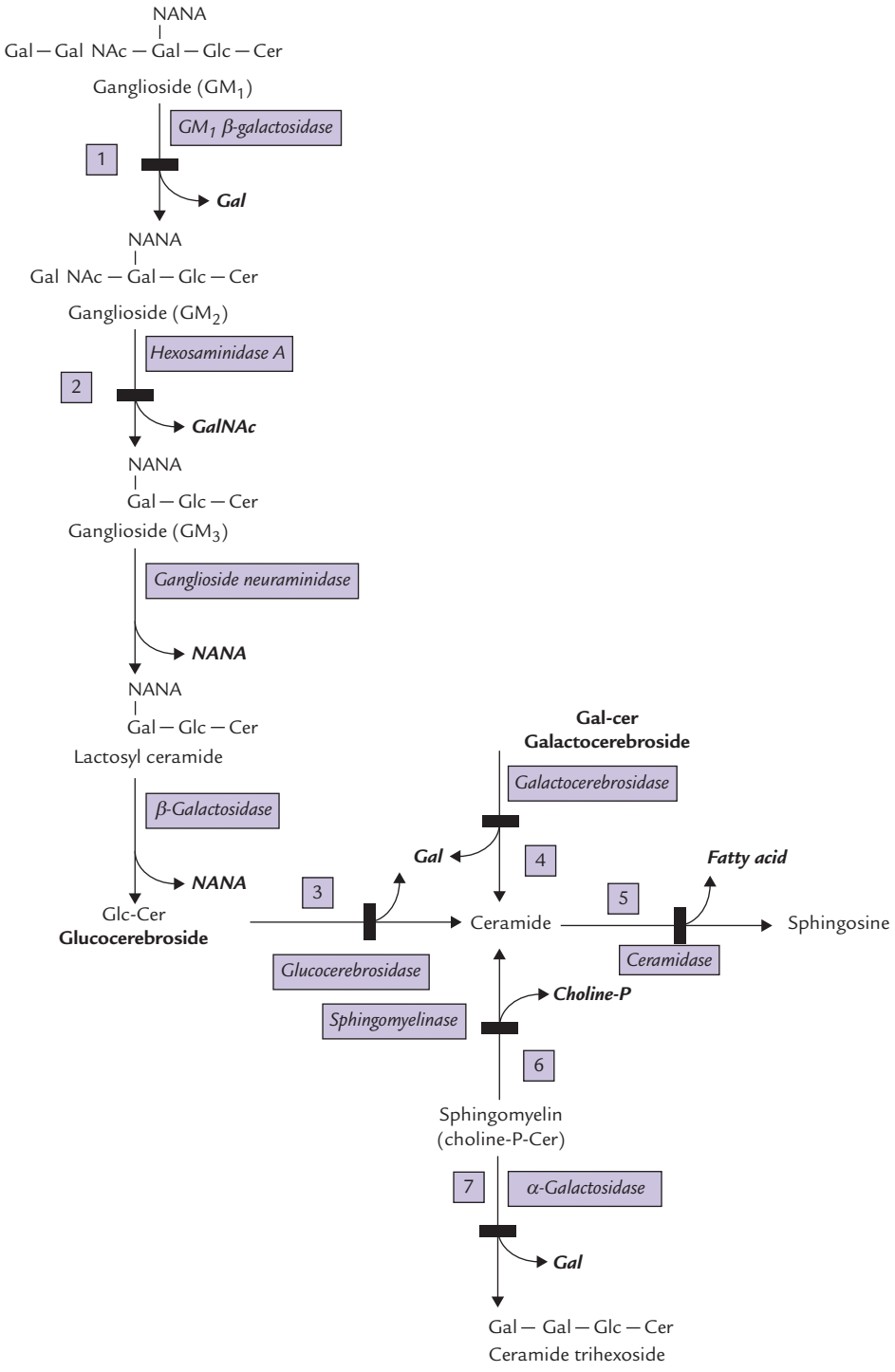


Fig. 11.30. Lysosomal degradation of sphingolipids, and the associated storage diseases ((1) generalized gangliosidosis, (2) Tay–Sachs disease, (3) Gaucher's disease, (4) Krabbe's disease, (5) Farber's disease (6) Niemann–Pick disease, (7) Fabry's disease).

Table 11.3. Some less commonly described sphingolipidoses

Disease (No)	Enzyme deficiency	Incidence	Accumulated lipid	Clinical causes
Farber's (5)	*Ceramidase*	Rare	Ceramide	Mental retardation, dermatitis skeletal abnormalities, early death
Niemann-Pick (6)	*Sphingomyelinase*	Rare	Sphingomyelin	Hepatosplenomegaly, mental retardation, early death
Fabry's (7)	*α-Galactosidase A*	1 in 40,000	Ceramide trihexoside	Skin rash, renal failure

* The numbers refer to the numbered reactions of Figure 11.30.

phagocytic cells in these organs remove erythrocytes from the circulation, and non-degradable lipids from the RBC membrane accumulate in these tissues.

Deficiencies of individual sphingolipid-degrading lysosomal enzymes result in lipid storage diseases.

Depending on the enzyme affected, several types of sphingolipidoses have been recognized (Fig. 11.30). For example, in Niemann-Pick disease, *sphingomyelinase* is defective, leading to sphingomyelin accumulation (Fig. 22.27; number 6).

General Features

Certain features are common to all types of sphingolipidoses. These are:

1. ***Inheritance:*** Sphingolipidoses are autosomal recessive diseases. The only exception is Fabry's disease, which is X-linked.
2. ***Severity:*** With exception of Fabry's disease and Gaucher's disease (adult form), all other types of sphingolipidoses are relentlessly progressive, causing death in early life.

   Milder variants of lipid storage disease also occur, in which affected homozygotes have greatly decreased, but not completely absent enzyme activity. The severity and prognosis of disease depends largely on the residual enzyme activity.
3. ***Diagnosis:*** Sphingolipidoses rests on determination of enzyme activities in cultured leukocytes, cultured skin fibroblasts; or, for prenatal diagnosis, cultured amniotic cells are used for determination of the enzyme activities.

   Heterozygotes can be identified because their enzyme activity is approximately one-half of normal; the homozygotes on the other hand have near-zero enzyme activity.

Major Types of Storage Diseases

Some major types of sphingolipidoses are depicted in Figure 11.30; numbered reactions refer to storage diseases in the Figure.

1. *Generalized gangliosidosis* is because of deficiency of ***β-galactosidase***, so that there is impaired degradation of GM_1 gangliosides and it accumulates in tissues. Mental retardation, skeletal deformities and hepatomegaly are the prominent features.
2. *Tay-Sachs disease* is a rare disease (except in Ashkenazi Jews) because of complete deficiency of ***hexosaminidase A***. It leads to accumulation of the substrate, ganglioside GM_2. The affected child appears normal at birth but within the first year of life develops signs of mental and neurological deterioration, and also hepatosplenomegaly. The lipid laden cells, the ganglion cells in retina (fovea centralis) cause visual impairment. Death occurs before the child reaches the age of 3 years.
3. *Gaucher's disease* is a rare disease caused by deficiency of the enzyme ***glucocerebrosidase***, so that tissue levels of glucocerebrosides elevate. The clinical course of the infantile form is similar to Tay-Sachs disease; and the adult onset form is without mental retardation.
4. *Krabbe's disease* results because of deficiency of ***β-galactosidase***, therefore there is accumulation of galactocerebrosides. There is near complete absence of myelin in the nervous tissue. The clinical features include mental retardation, blindness, deafness and hypertonia.

Some other disorders associated with abnormal sphingolipid metabolism are also shown in Table 11.3.

Exercises

Essay type questions

1. Describe β-oxidation of a 16-carbon saturated fatty acid and its energetics. Explain the role of carnitine in transport of the fatty acid into mitochondrion.
2. Describe synthesis and oxidation of ketone bodies and discuss regulation of ketogenesis.
3. Describe the shuttle systems for transport of fatty acids into the mitochondria and acetyl CoA into the cytosol.
4. Describe the functions and metabolism of phospholipids.
5. Describe in detail the pathway of de novo synthesis of fatty acids and its regulation.
6. Describe the major mechanisms of regulating fatty acid metabolism in humans.
7. How are triacylglycerols, glycerophospholipids and sphingolipids synthesized?

Write short notes on

1. Carnitine
2. Ketone bodies
3. Gaucher's disease
4. *Acetyl CoA carboxylase*
5. Acyl carrier protein
6. *Lipoprotein lipase*
7. Refsum's disease
8. Omega oxidation
9. Sphingolipidoses
10. Ketosis

CLINICAL CASE

CASE 11.1 Muscle wasting in a 2-year-old child

A 2-year-old child was brought to the hospital OPD with complaints of difficulty in performing movements. Two months back his mother had first noticed that he had problem in climbing the staircase. He could not easily get up from chair and found running difficult. There had been a progressive deterioration since then, and presently he had problem even in walking. Further enquiry brought out that there had been delay in development of motor milestones. A generalized weakness was observed on physical examination. Muscles of hands and feet showed some wasting. Blood sample was obtained for biochemical analysis. It showed the following abnormal results:

1. *Serum creatine phosphokinase* (CK) activity was elevated about six-fold.
2. Serum carnitine level was slightly elevated. X-Ray chest showed evidence of cardiomyopathy. Muscle biopsy showed massive lipid accumulation.

Q.1. What is the most probable diagnosis?
Q.2. Discuss the mechanism by which muscle wasting occurs in this child.
Q.3. What treatment would you recommend for this patient?
Q.4. Episodes of hypoglycaemic coma are quite common in this disorder. Why?

LIPID METABOLISM II: LIPOPROTEINS, CHOLESTEROL AND PROSTAGLANDINS

CHAPTER

12

Compared to glucose and amino acids, which are freely transported in the aqueous plasma, transport of water insoluble lipids involves the participation of a range of complex molecules. Unesterified ("free") fatty acids are transported in non-covalent binding to serum albumin, but triacylglycerols and cholesterol esters are too hydrophobic for such transport. They are, therefore, packaged together with proteins (apoproteins) and amphipathic lipids by hepatocytes and enterocytes into particles called lipoproteins. Lipoproteins are important for they provide means for fat transport between different organs and tissues. In this chapter, a detailed account of normal functions of lipoproteins and their relationship to abnormal states, particularly atherosclerosis, is given. The chapter also deals with derived lipids such as cholesterol and eicosanoids. Cholesterol is a major component of cell membranes and serves as a precursor of bile salts and steroid hormones. Eicosanoids, which include prostaglandins, thromboxanes and leukotrienes act as regulators of cellular functions.

After going through this chapter, the student should be able to understand:

- Composition, separation, properties, metabolism and significance of different types of lipoproteins; and function of each type of lipoprotein and apolipoprotein.
- Pathogenesis of fatty liver.
- The de novo pathway of cholesterol biosynthesis and its regulation, with special emphasis on role of hormones.
- Classification, origin and structure of prostaglandins, thromboxanes, and leukotrienes; and important physiological effects and therapeutic uses of these compounds.

I. Lipoproteins

A. Overview

Lipoproteins are molecular complexes consisting of lipids and proteins that provide means for lipid transport between different organs and tissues. The lipids, by virtue of their immiscibility with aqueous solutions, depend on protein carriers for transport within bloodstream and extracellular fluids. A lipoprotein particle is a spherical structure having a hydrophobic core wrapped in an amphipathic coating.

- The **core** contains cholesterol esters and triacylglycerols, which, because of their non-polar nature, always avoid contact with water (Fig. 12.1).
- The amphipathic **coat** contains phospholipids, free cholesterol, and proteins. It interacts with water in plasma, thereby promoting solubility of lipoproteins.

It is noteworthy that the interactions between lipids and proteins in the lipoprotein particle are of a purely non-covalent nature. Protein part of lipoprotein is called **apoprotein** or **apolipoprotein**. The apolipoproteins on surface of lipoproteins help to solubilize the lipids and target the lipoproteins to the correct tissues.

The most characteristic structural feature of apolipoproteins is the *amphipathic helix*, an α-helix in which one side is formed by hydrophobic- and the other by hydrophilic-amino acid side chains. The latter faces the aqueous exterior (being found at the water-lipid interface), so that lipoproteins are soluble in water.

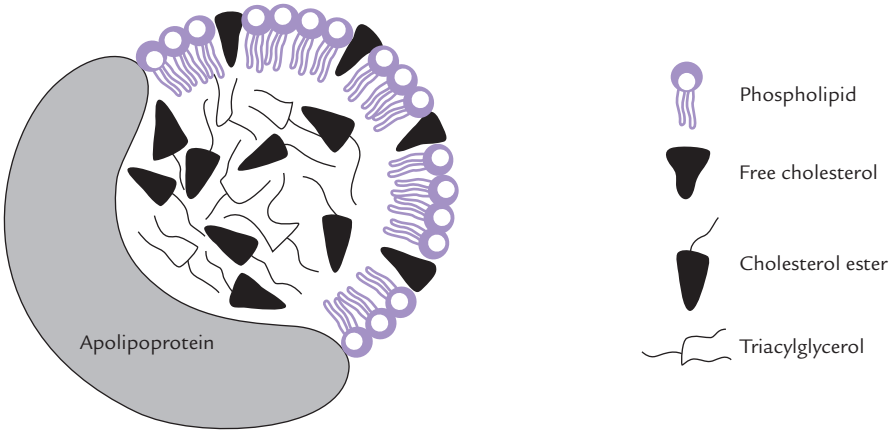


Fig. 12.1. Structure of a lipoprotein: a globular particle consisting of a hydrophobic core of triacylglycerol and cholesterol ester, surrounded by an amphipathic coat of proteins, phospholipids and cholesterol.

Table 12.1. Plasma lipid levels in adults (360–600 mg/dL) of which approximately 1/3rd is phospholipids and less than 1/3rd is triacylglycerols

Lipids*	Concentration (mg/dL)
Triacylglycerol (TAG)	40–140
Cholesterol (C)	150–200
Phospholipids (PL)	145–250
Free fatty acids (FFA)	4–20
Total	360–600

* TAG, C and PL are carried by lipoproteins but FFAs are transported after forming complexes with serum albumin.

Most plasma lipids are components of lipoproteins, the non-covalent aggregates of lipids and proteins, that serve as transport vehicles for triacylglycerols, phospholipids and cholesterol around the body.

The amount and types of lipids can change rapidly in the circulation according to dietary habits because most lipids are exchanged readily between different lipoprotein particles. The normal range for the lipid levels are shown in Table 12.1.

B. Nomenclature

Several different classes of lipoproteins exist, each having characteristic lipid and protein compositions. Their densities are inversely related to their lipid content: the higher the lipid content, the lower the density (Table 12.2). Apart from the largest species, the chylomicron, lipoproteins are named according to their density (because they are most commonly isolated by ultracentrifugation). The various classes are: (a) chylomicrons (**CM**), (b) very low density lipoproteins (**VLDL**), (c) low density lipoproteins (**LDL**), and (d) high density lipoproteins (**HDL**).

C. Functions

The lipoprotein particles have three important functions in human body. First, they transport the dietary fat from the intestinal mucosa, where it is absorbed, to liver by **exogenous lipid transport** (this role is performed by chylomicrons). The second role is to transfer triacylglycerols and cholesterol from liver to other tissues by **endogenous lipid transport** (VLDL and LDL carry out this function). The third function is to transport cholesterol from extrahepatic tissues to the liver by **reverse cholesterol transport** (HDL performs this role).

D. Constituents

Lipids

Typical analysis of the lipid composition of plasma lipoproteins are shown in Table 12.2. The following observations are noteworthy:

1. All lipoproteins contain various types of lipids: **triacylglycerols**, **phospholipids**, and **cholesterol**-*free* and *esterified*, although the quantities and relative proportions of these vary widely.
2. The bulk of plasma cholesterol is transported associated with LDL, and the bulk of plasma triacylglycerol is transported associated with chylomicrons or VLDL. Therefore, *elevated cholesterol level most often reflects an increase in the LDL fraction*, whereas elevation of the plasma triacylglycerol level is usually caused by increased VLDL or chylomicrons.
3. Density is inversely proportional to lipid content and directly proportional to the protein content.

Proteins

Five main classes of apolipoproteins (A–E) have been identified in lipoproteins. They are mainly **synthesized**

Table 12.2. Composition of plasma lipoproteins

	Chylomicron	VLDL	LDL	HDL
1. Particle mass (kD)	400,000	10,000–80,000	2300	175–360
2. Density (g\cm^3)	< 0.95	0.95–1.006	1.018–1.063	1.063–1.21
3. Diameter (Å)	1000–10,000	300–700	150–250	75–100
4. Apolipoproteins	A, B, C, E	B, C, E	B	A, C, D, E
5. Components (% dry weight)				
Apolipoproteins	1.5–2.5	5–10	20–25	40–55
Triacylglycerols	84–89	50–65	7–10	3–5
Cholesterol	4–8	15–23	40–50	15–18
Phospholipids	7–9	15–20	15–20	20–35

in liver, although smaller quantities are produced in almost all organs. The **group-A** are found in highest amounts in HDL, while the **group-B** are associated with VLDL, LDL and chylomicrons. The latter are tightly integrated into the phospholipid bilayer, whereas members of **C** and **E group** are less tightly associated and exchange among lipoproteins with HDL acting as the distributor.

Functions of Apolipoproteins

The apolipoproteins participate in lipoprotein metabolism by performing following functions:

1. **Transport:** The primary role is to impart a hydrophilic character to the lipoprotein particles so that they can be transported in the aqueous plasma.
2. **Enzyme activation:** They activate various enzymes of lipoprotein metabolism. For example, apoC-II activates the enzyme *lipoprotein lipase*. The activated enzyme then catalyzes degradation of the triacylglycerol component of VLDL and chylomicrons.
3. **Ligand:** The apoproteins recognize the cell surface receptors, thus serving as ligands. For instance, the apoB-100 serves as a ligand for the LDL receptor.
4. They facilitate the transfer of lipids between different lipoprotein classes or between lipoproteins and cells. **ApoD**, for example, acts to facilitate the transfer of esters of cholesterol between lipoproteins.

Further details about roles of apolipoproteins are given in Table 12.3.

In some cases there is more than one gene for a particular class of apolipoprotein and these are specified by post-scripts (e.g. apoA-I, apoC-II). In addition, different alleles at a particular locus have been recognized within a given population, i.e. there is **polymorphism**. Presence of a particular allele is sometimes found to be associated with increased susceptibility to certain diseases. For example, there are four alleles of apoE designated as ε_1 to ε_4 and the individuals expressing the ε_4 allele are at greatly increased risk of developing Alzheimer's disease and senile dementia.

E. Separation of Lipoproteins

Plasma lipoproteins can be separated by either their electrophoretic mobility or their density. The techniques most commonly used are electrophoresis on agarose gel, and ultracentrifugation.

Electrophoresis

This method separates lipoproteins on the basis of their charge and mass to yield four major bands. These bands are shown in Figure 12.2, which also shows the alternative names used to specify lipoproteins. In mildly alkaline pH (8.3), with the fasting serum or plasma, two most prominent lipoprotein bands are found in the α and the β fractions. They are designated accordingly as α-lipoprotein and β-lipoprotein, which consist respectively of HDL and LDL. A less prominent band the pre-β-lipoprotein (= VLDL) moves slightly ahead of the β-lipoprotein. After a fatty meal, a fourth band, which does not show any mobility in electrophoresis, consists of chylomicrons.

Ultracentrifugation

This method allows separation according to density and also yields four major fractions. Since lipids are less dense (density approximately 0.9 g/cm^3) than proteins, the lipoproteins with high lipid content have a low density (from 0.9 g/cm^3 to 1.006 g/cm^3) and so float on centrifugation. With increasing protein content, the densities of lipoprotein particles increase to well above 1.0 g/cm^3, and so they sediment easily.

Currently, the terms used to designate density of the lipoproteins are also used in naming them. In the order of increasing density, these are very low density lipoproteins, VLDL; low density lipoproteins, LDL and high density

Table 12.3. Classes of apolipoproteins

Type	Found in	Site of synthesis	Plasma concentration (mg/dL)	Mean particle mass (kD)	Function and comments
A.					
A-I	HDL, Chylomicron	Liver, intestine	70–160	29.01	Major protein of HDL, activates LCAT
A-II	HDL, Chylomicron	Liver, intestine	20–55	17.4	Primarily in HDL, enhances *hepatic lipase* activity
B.					
B-100	LDL, VLDL	Liver	60–120	513.0	Structural protein of VLDL and IDL; the only apoprotein of LDL. Mediates tissue uptake of LDL particles. One of the longest proteins in humans
B-48	Chylomicron	Intestine	Variable	241.0	Structural protein of chylomicrons, derived from apoB-100 gene by RNA editing in intestinal epithelium; lacks the LDL receptor-binding domain of apoB-100
C.					
C-I	Chylomicron, VLDL, HDL	Liver	3–11	6.6	Readily transferred between different classes C-II activates *lipoprotein lipase*
C-II	-do-	Liver	1–6	8.9	
C-III	-do-	Liver	3–23	8.7	
D.	HDL	Not known	8–12	19.0	Associated with HDL, cholesterol-ester transfer
E.	Chylomicron VLDL, HDL	Liver	2–6	34.0	Mediates uptake of chylomicron remnants and IDL by liver. Rich in arginine residues

HDL = high density liporotein, IDL = intermediate density lipoproteins , LCAT = *lecithin cholesterol acyl transferase*, LDL = low density lipoprotein, VLDL = very low density lipoproteins.

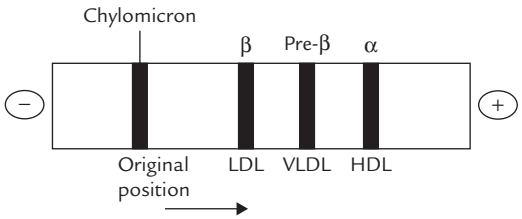


Fig. 12.2. Separation of the lipoprotein fractions by electrophoresis (Buffer pH = 8.3). (α, β, pre-β and chylomicrons = names of fractions of lipoproteins, HDL = high density lipoprotein, LDL = low density lipoprotein, VLDL = very low density lipoprotein).

lipoproteins, HDL (density of chylomicrons are comparable to that of VLDL).

Electrophoresis separates lipoproteins on basis of their charge and mass, while ultracentrifugation separates on the basis of density.

F. Lipoprotein Metabolism

Chylomicrons

The dietary triacylglycerols are transported by chylomicrons, the largest of lipoproteins. They are synthesized in the intestine and transport dietary triacylglycerols to skeletal muscles and adipose tissue, and dietary cholesterol to the liver. As the chylomicrons pass through the capillaries of various tissues, the triacylglycerols are hydrolyzed by the action of *lipoprotein lipase* (LPL), an enzyme located on the endothelial cells. The fatty acids and monoacylglycerols released by the hydrolysis diffuse directly into the tissue either for metabolism or for storage.

Synthesis

Following absorption in intestine, the dietary lipids are incorporated in chylomicrons. Since chylomicrons carry lipids (mainly triacylglycerols) of dietary origin, they are synthesized and appear in circulation only after a meal rich in fats.

The relative content of triacylglycerols, cholesterol, phospholipids and fat-soluble vitamins present in chylomicrons reflect the lipid composition of the preceding meal. These lipid components are assembled in the SER and the Golgi apparatus of the mucosal cells. Then the apolipoproteins (B-48 and A) synthesized in RER are also incorporated. The particles so formed, called **nascent chylomicrons**, are exocytosed into the laceteals of the intestinal villi. From these lymph vessels (i.e. laceteals) the nascent chylomicrons reach the blood circulation via the thoracic duct.

Metabolism

Initially, as the chylomicrons are synthesized by intestinal cells, they contain only apoB-48 and apoA. But upon entering the circulation, the nascent particles acquire apoC and apoE from plasma HDL to form **mature chylomicrons** (Fig. 12.3). The acquisition of these apolipoproteins makes the mature chylomicrons functionally competent. In particular, the **apo C-II** in mature particles activates the enzyme *lipoprotein lipase* (LPL).

In the peripheral tissues, such as muscles and adipose tissue, the activated LPL causes hydrolysis of about 80–90% of the chylomicron triacylglycerols. This is accompanied by the transfer of most of the A- and C-apolipoproteins to HDL. These changes convert the chylomicron into a smaller particle, known as a **chylomicron remnant**. The fatty acids released from the hydrolyzed triacylglycerols enter muscle and adipose tissue cells, and the glycerol part enters the liver, where it is used for synthesis of TAG.

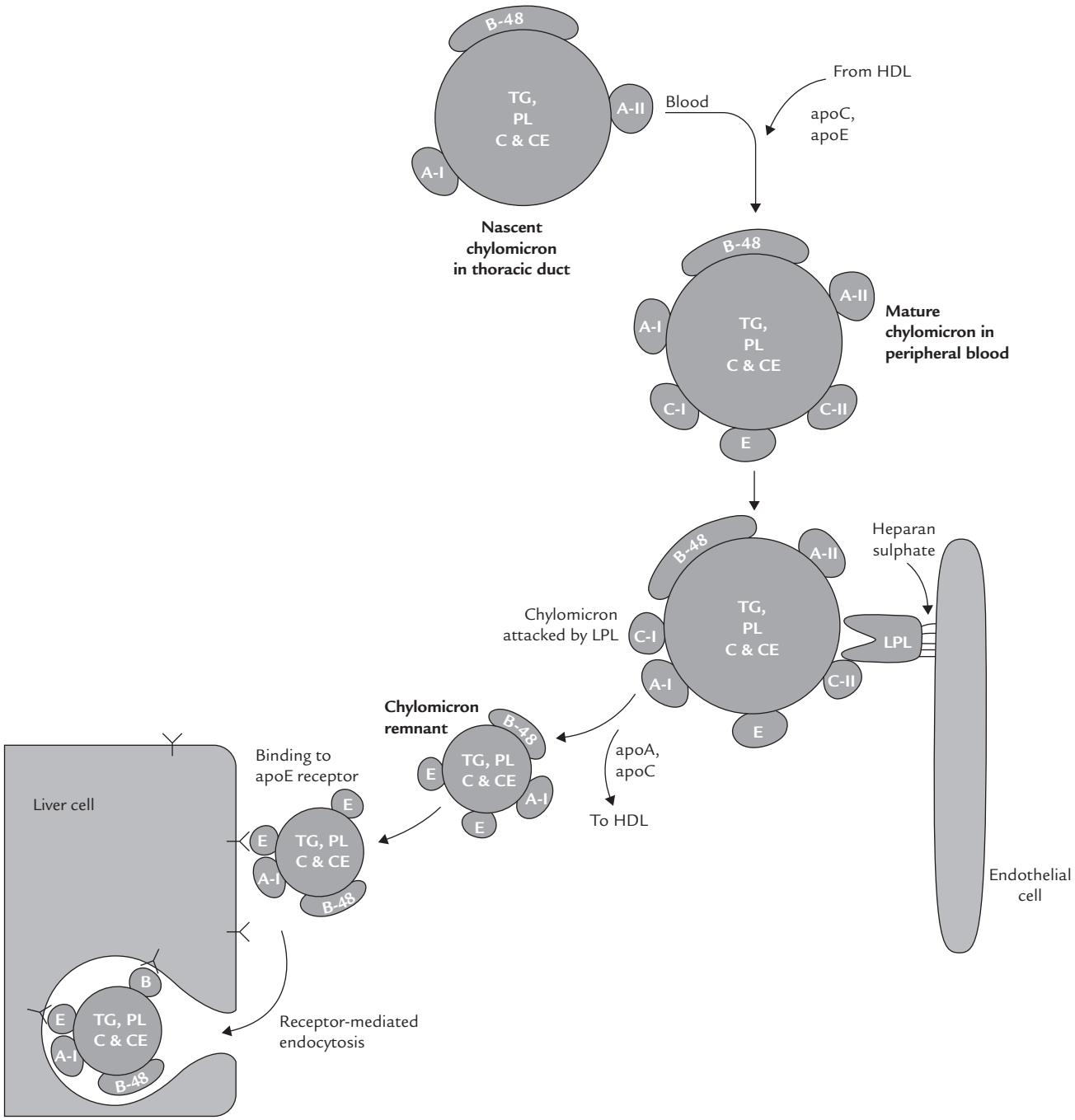


Fig. 12.3. Metabolism of chylomicrons showing their role in transporting dietery cholesterol to liver. (TAG = triacylglycerol, PL = phospholipid, C = free cholesterol, CE = cholesterol ester, LPL = *lipoprotein lipase*).

Decreased degradation of chylomicron triacylglycerols due to impaired LPL activity results in elevated levels of chylomicron even after overnight fast (**Case 12.1**).

Fate of Chylomicron Remnant

The chylomicron remnants formed by off loading of triacylglycerols are finally removed from blood circulation by liver. They bind to lipoprotein receptors on the surface of hepatocytes, including the *LDL receptor and the LDL-receptor-related protein (LRP)*. This binding requires **apoE**. After binding to one of these receptors, the whole remnant particle is taken up by hepatocytes (i.e. endocytosis). Intracellularly, the endocytosed vesicles are carried to the lysosomes where they are degraded to release the constituents. Hence, much of the cholesterol taken in the diet is delivered to liver, together with fat-soluble vitamins. The lifetime of a chylomicron, from its secretion by the intestinal cell to the uptake of remnant by the liver is less than **1 hour**.

Triacylglycerols in mature chylomicrons are extensively degraded by *lipoprotein lipase* on surface of the endothelium cells, and the resulting chylomicron remnants are transported in blood to the liver where they are taken up by receptor mediated endocytosis.

Very Low Density Lipoproteins (VLDL)

Synthesis and Secretion

Synthesis of **nascent VLDL** particles occurs in the hepatocytes by a similar set of reactions as described for the synthesis of nascent chylomicrons. It involves assembly of triacylglycerols, phospholipids, cholesterol and apoB-100. The biosynthetic process requires participation of several organelles:

- RER is involved in the production of apoB-100 and SER in its assembly with the lipid components.
- Glycosylation occurs in the Golgi to produce nascent VLDL.

For additional information on synthesis of apoB-100, refer to Box 12.1.

Secretion of the nascent VLDL particles occurs in the same way as in case of the chylomicrons. However, the nascent VLDL particles are released directly into blood circulation, where they obtain more of apoE and C-apolipoproteins from HDL to form **mature VLDL** particles (Fig. 12.4).

The TAG-rich mature VLDL exchanges TAGs for cholesterol esters with HDL, as discussed latter.

Metabolism of VLDL

Like chylomicrons, the mature VLDL is acted upon by *lipoprotein lipase* in capillaries of peripheral tissues. The enzyme is activated by apoC-II of mature VLDL. (Refer to Box 12.2 for details on action of LPL.)

VLDLs are synthesized in the liver and transport triacylglycerols, cholesterol and phospholipids to other tissues, where *lipoprotein lipase* hydrolyzes the TAGs and releases the fatty acids for uptake.

The VLDL triacylglycerols are hydrolyzed more slowly compared to those in chylomicrons with a residence time in blood of 15–60 minutes (5–10 minutes for chylomicron triacylglycerols). Half-life of VLDL in serum is **1–3 hours**.

Progressive stripping of triacylglycerols is accompanied by the transfer of apoC to HDL. Like the chylomicron remnants, the VLDL remnants possess apoE, which would mediate their uptake into the liver. The fatty acids released from the degraded triacylglycerols are taken up by mainly by peripheral tissues.

Processing of VLDL remnants to IDL and LDL: Initial hydrolysis of the VLDL triacylglycerols by *lipoprotein lipase* in peripheral tissues produces smaller particles called *VLDL remnants* (Fig. 12.4). The VLDL remnant is smaller and contains relatively more cholesterol and

BOX 12.1

ApoB-100 and apoB-48 are Closely Related Polypeptides

The major apolipoprotein of VLDL (apoB-100), a single polypeptide of 4536 amino acids, is encoded by the same gene as apoB-48, the prominent apolipoprotein of chylomicron. In fact, apoB-48 consists of the first 2152 amino acids of apoB-100, counting from the N terminal. In the intestine, the codon 2153 (CAA) which codes for glutamine is post-transcriptionally changed to UAA, a stop codon. Therefore, in intestine, translation stops at the 2153rd codon so that instead of a polypeptide of 4536 amino acids (i.e. apoB-100), a shorter polypeptide of 2152 amino acids (i.e. apoB-48) is synthesized. This is an example of *tissue-specific editing* of an RNA transcript.

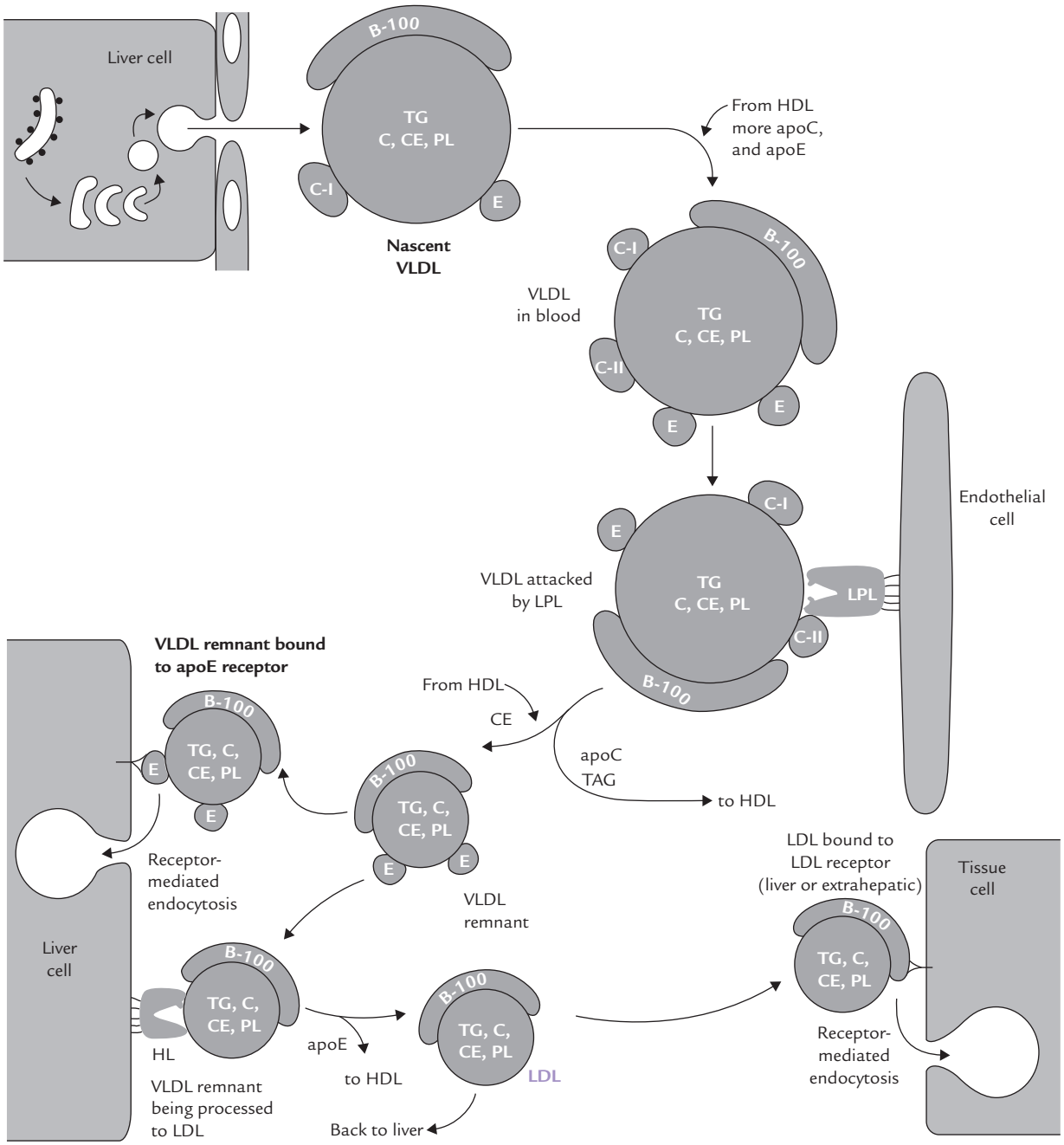


Fig. 12.4. Metabolic fate of VLDL and LDL (TAG = triacylglycerol, PL = phospholipid; C = free cholesterol, CE = cholesterol esters, HL = *hepatic lipase*).

apoE than VLDL. At this stage, apoE becomes responsible for further metabolism of the remnant particles. A proportion of these particles are internalized in the liver through mediation of receptors (that recognize the apoE) and the rest are further degraded by *hepatic lipase* (a liver specific enzyme present in space of Disse) *to yield intermediate density lipoproteins* (IDL), and then LDL. The transformation (IDL to LDL) is effected by the enzyme *hepatic lipase* (HL) by further loss of triacylglycerols as well as transfer of excess apolipoproteins to HDL. The HL, which has structural homology with LPL, is found on cell surfaces of hepatocytes, anchored to heparan sulphate proteoglycans. It hydrolyzes lipids of IDL (and HDL, as discussed later). *Finally, by the time IDL is converted to LDL, it has lost all apolipoproteins except apoB-100*. The LDL now acquires a conformation that enables it to recognize the LDL receptors on cell surfaces of peripheral tissues.

The VLDL remnants are transformed first to IDL and then to LDL as all of their apoproteins other than B100 are removed, and their triacylglycerols hydrolyzed.

BOX 12.2

Mode of Action of *Lipoprotein Lipase*

The fact that LPL is activated by apoC-II indicates that the enzyme must be located in the extracellular compartment. In fact, it is bound to the outer aspect of the plasma membrane of the endothelial cells of those capillaries which supply blood to adipose tissue and other peripheral tissues. The enzyme protein is bound to these cells by proteoglycan chains (heparan sulphates). The binding is disrupted by heparin, so systemic administration of heparin releases the enzyme into the circulation (heparin also increases the LPL activity). Thus, heparin injection causes greatly increased activity of LPL. The appearance of LPL on cell surface is under hormonal control: **insulin** increases and noradrenaline decreases the transfer of enzyme to cell surface. This explains the off-loading and uptake of fatty acids (and their storage as triacylglycerols) by adipose tissue by post-prandial release of insulin. (Insulin → transfer of LPL on cell surface → Enhances off-loading of TAGs from VLDL and chylomicrons) → uptake and storage.

LPL differs from *hormone-sensitive lipase* (HSL) of adipose tissue in some important aspects:

1. LPL is activated by apoC-II and phospholipids; HSL is activated by hormones through cAMP-dependent phosphorylation.
2. Insulin induces overall activity of LPL, but inhibits activation of HSL.
3. Finally, unlike LPL, the HSL is not bound to capillary walls by proteoglycan chains.

***Mode of action of** hepatic lipase: Hepatic lipase* is also released by heparin, like LPL. But unlike LPL the latter it is not activated by apoC and does not attack chylomicrons and VLDL. Its physiological substrates are triacylglycerols (and phosphoglycerides) in VLDL remnant, IDL and HDL. In addition to degrading TAGs in these lipoproteins, HL also promotes uptake of remnant particles into hepatocytes, as discussed later.

Lipids from endogenous synthesis in the liver are carried as constituents of very low density lipoproteins (VLDL) to other tissues where *lipoprotein lipase* degrades them into VLDL remnants, and *hepatic lipase* degrades the remnants into IDL and LDL-finally.

Low Density Lipoproteins (LDL)

Most of the LDL is derived from VLDL and IDL but a smaller amount is released directly from intestine. During its formation from VLDL, apoB-100 is the only apolipoprotein that is retained in LDL. Most of the triacylglycerols are also lost during this transition, so that LDL particles contain relatively higher concentration of cholesterol, and cholesterol esters. ApoB-100 serves as ligand for the LDL receptors.

The VLDL-IDL-LDL cascade results in the delivery of endogenous lipids from liver to peripheral cells.

LDL Receptors

These receptors are present in all cells but are most abundant on hepatocytes and adrenal cortex (Fig. 12.4). Helped by apoB-100, these receptors mediate uptake of cholesterol rich LDL molecules either by liver (approximately 70%) or by peripheral tissues (approximately 30%), especially fibroblasts, vascular smooth muscles and lymphocytes. It is a highly regulated process of **receptor-mediated endocytosis** occurring via clathrin-coated pits (Chapter 7). The LDL receptor is sometimes referred to as the **apoB-100/apoE** receptor because it has affinity for apoE also (but only at high concentration of apoE), and can mediate uptake of IDL particles as well. This step of *receptor-mediated endocytosis of LDL is an important facet of whole-body cholesterol homoeostasis*. The subsequent degradation of LDL (and IDL) in the lysosomal compartment of cells results in the release of cholesterol. The free receptor can then return to the cell surface for further uptake of LDL molecules. **Goldstein** and **Brown** were awarded Nobel prize in 1985 for their work on LDL receptor.

LDL is removed by receptor-mediated endocytosis both by the liver (two-third) and by extrahepatic tissues (one-third). The removal is initiated by binding of apoB-100 to the LDL receptor.

A summary of metabolic fate of chylomicrons, VLDL and LDL is shown in Figure 12.5. Defective LDL receptors lead to decreased clearance of circulating LDL particles by peripheral tissues. The resulting condition is called *familial hypercholesterolaemia* (**Case 12.2**).

Regulation of LDL Receptors

The number of LDL receptors present on the cell surface is regulated by the cellular needs. For example, when the

intracellular cholesterol level falls, it induces increased expression of receptors on the cell membrane. This process, known as **up regulation**, enables the cells to internalize more circulating LDL particles and utilize their cholesterol content. Conversely, when cell has adequate supply of cholesterol, number of receptors decrease to prevent any further uptake of LDL. This process is called **down regulation**, and is seen in the following conditions:

1. Raised intracellular cholesterol concentration, which occurs following cellular uptake of the circulating LDL particles.
2. Hypercholesterolaemia, when the blood circulation is flooded with cholesterol and excessive uptake is to be avoided.

Increased intracellular cholesterol concentration depresses transcription of the genes for LDL receptors (Fig. 12.5) thus reducing the number of receptors on cell surface.

Fate of the Internalized LDL

The internalized LDL is acted upon by *lysosomal enzymes*, and cholesterol esters are hydrolyzed to cholesterol and fatty acids (Fig. 12.5). The free cholesterol can then be used directly as a component of membranes (the fatty acids can have any of the catabolic or anabolic fates discussed earlier).

Cholesterol, not needed for membrane synthesis can have following effects:

1. It decreases cholesterol biosynthesis by inhibiting activity of the key enzyme, *HMG CoA reductase*. Decreased de novo synthesis of cholesterol prevents any further elevation of the cellular cholesterol content. This point has important implications: the dietary cholesterol suppresses the synthesis of cholesterol by the body, especially in tissues other than liver.
2. Free cholesterol in cell inhibits synthesis of LDL receptors. As a result, cellular uptake of LDL is inhibited, and the circulating cholesterol in bloodstream increases.
3. The intracellular cholesterol activates the intracellular enzyme *acyl CoA cholesterol acyl transferase (ACAT)*. This enzyme catalyzes transfer of an acyl group from a fatty acid derivative to cholesterol, resulting in the formation of esterified cholesterol. By this mechanism, the surplus cholesterol that is not immediately required for utilization by the cells is converted to cholesterol ester. The cholesterol ester is stored in the cell for subsequent use.

Some cholesterol is expelled from the cell and transferred to liver by HDL.

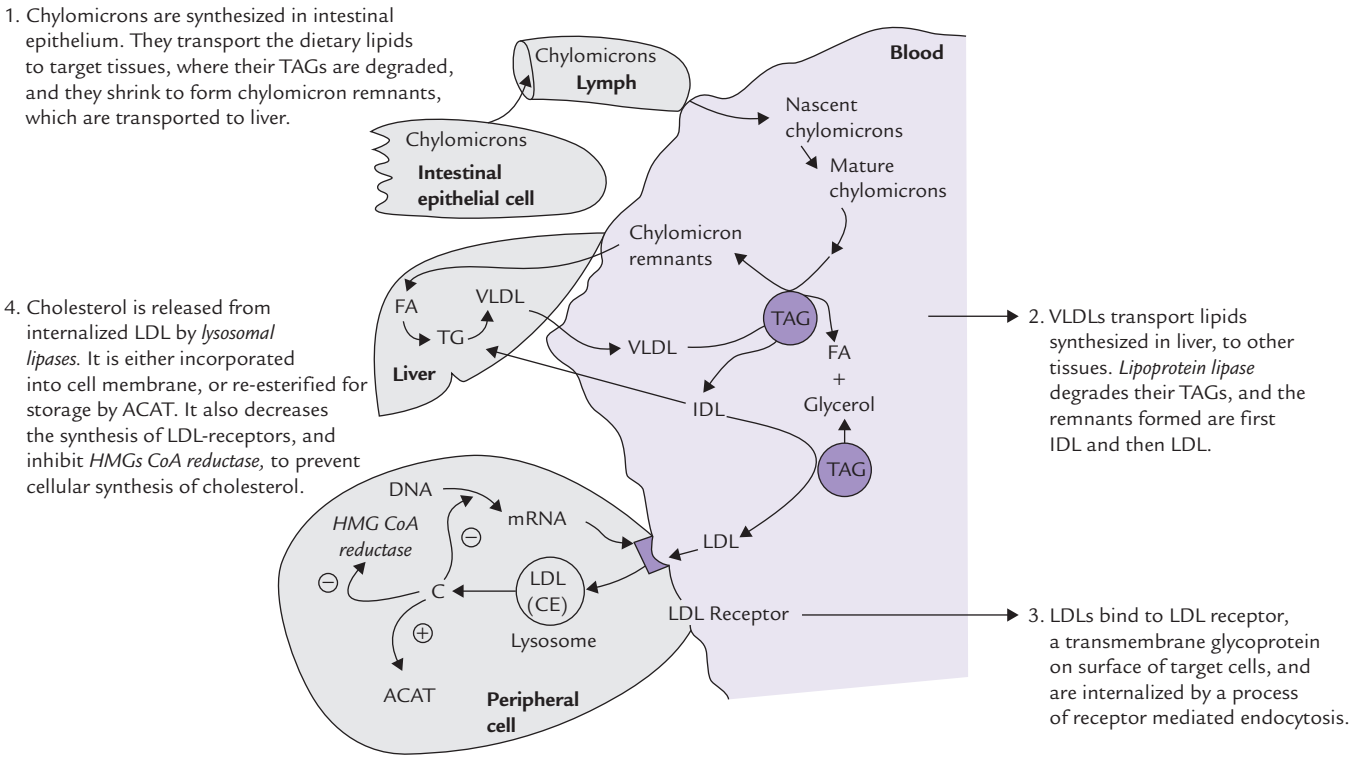


Fig. 12.5. Summary of metabolic fate of chylomicrons and very low density lipoproteins (VLDL), and regulation of low density lipoproteins (LDL) receptors (ACAT = *acyl cholesterol acyl transferase*, C = cholesterol, CE = cholesterol ester, FA = fatty acids, TAG = triacylglycerol, IDL = intermediate density lipoproteins).

LDL Receptors on Macrophages are not Down-Regulated

The circulating LDL infiltrates through arterial walls and taken up by macrophages through another type of receptors which are different from the LDL receptors. These receptors on macrophages readily take up the LDLs, especially when they are chemically modified, either by acetylation or oxidation of apoB-100. They are not down regulated by intracellular cholesterol content of the macrophages. Therefore, *uptake of cholesterol by macrophages continues even after intracellular cholesterol concentration is substantially built up*. Furthermore, the intracellular cholesterol content does not depress activity of *HMG CoA reductase* within the macrophages. Thus, macrophages are capable of unchecked and excessive uptake of circulating cholesterol. When level of circulating cholesterol rises (i.e. hypercholesterolaemia) the intracellular cholesterol concentration (of macrophages) builds up substantially. This causes transformation of the macrophages into "*foam cells*" which play an important role in development of the atherosclerotic plaques.

Macrophages play a role in atherogenesis because of unchecked and excessive uptake of circulating cholesterol.

LDL is bad cholesterol: The cholesterol rich LDL particles (about 75% of the serum cholesterol is contained in LDL) serve to transport cholesterol from liver to peripheral tissues. Therefore, elevated serum concentration of LDL is deleterious for having a positive correlation with cardiovascular disease. Hence, the LDL cholesterol is lethally dangerous, and commonly referred to as the "bad cholesterol".

LDL contains most of the total cholesterol (70%) in the fasting serum, and VLDL is the major triglyceride rich lipoprotein.

High Density Lipoprotein (HDL)

The HDL particles are referred to as scavengers because their primary role is to remove free (unesterified) cholesterol from the extrahepatic tissues, which is then excreted through bile. Evidently, this a crucial mechanism that prevents the inappropriate accumulation of cholesterol in peripheral tissues. Because accumulation of cholesterol in tissues is strongly associated with the development of atherosclerosis, the *level of HDL in serum is inversely related to the incidence of myocardial infarction*. Thus, HDL is cardioprotective or anti-atherogenic in nature, and is referred to as the "**good cholesterol**".

A key role in reverse transport of cholesterol from extrahepatic tissues to the liver is played by HDL, which also serves as circulating reservoir of apolipoproteins.

Metabolism

The intestinal cells and hepatocytes synthesize various components of HDL and release them into bloodstream. These newly assembled particles are called nascent HDL. They are found to be discoid in shape (Fig. 12.6). The disc consists of a phospholipid bilayer, associated with which are a number of apolipoproteins (A-I, A-II, E and C). *ApoA-I is the principal apolipoprotein initially, others are acquired later*. Once in circulation, the nascent HDL attracts free, unesterified cholesterol from cells. It (the nascent HDL) binds these cells via apoA-I or apoE receptors, and acquires cholesterol from them.

Nascent HDLs are released from liver and intestine and extract cholesterol from cell membranes.

Further fate of this cell-derived cholesterol depends on a soluble, plasma enzyme, *lecithin cholesterol acyl transferase* (LCAT). This enzyme binds to the disc (i.e. HDL) and gets activated by the apoA-I. The activated enzyme catalyzes transfer of fatty acyl residues from lecithin (phosphatidylcholine) to cholesterol to form lysolecithin and cholesterol ester:

$$\text{Cholesterol} + \text{Lecithin} \xrightarrow{\text{LCAT}} \text{Cholesterol ester} + \text{Lysolecithin}$$

The lysolecithin is transferred to albumin while the hydrophobic cholesterol ester sinks between the two leaflets of the bilayer of the discoid, thereby forcing them apart. The originally flat HDL now bulges into a spherical, mature HDL particles called **HDL_3** with a hydrophobic core of cholesterol esters. The latter may then be transferred to other lipoproteins, such as chylomicrons, VLDL or remnants, either alone or in exchange for triacylglycerols (Fig. 12.7). The exchange is effected by **cholesterol ester transfer protein** (CETP; or **apoD**). The triacylglycerols thus acquired increase the particle size and the smaller HDL_3 thus changes to larger HDL_2.

At the liver surface, HDL_2 particles bind to a receptor called SR-B_1. They (HDL_2) are substrate for *hepatic lipase* (HL) which removes triacylglycerols in the liver. Cholesterol and cholesterol esters are also off-loaded from the degrading particles and transferred into liver cells. The lipid depleted, apoA rich HDL particles now return to plasma and continue to extract cholesterol from all membranes. A minority of these HDL particles acquire multiple copies of apoE from other lipoproteins

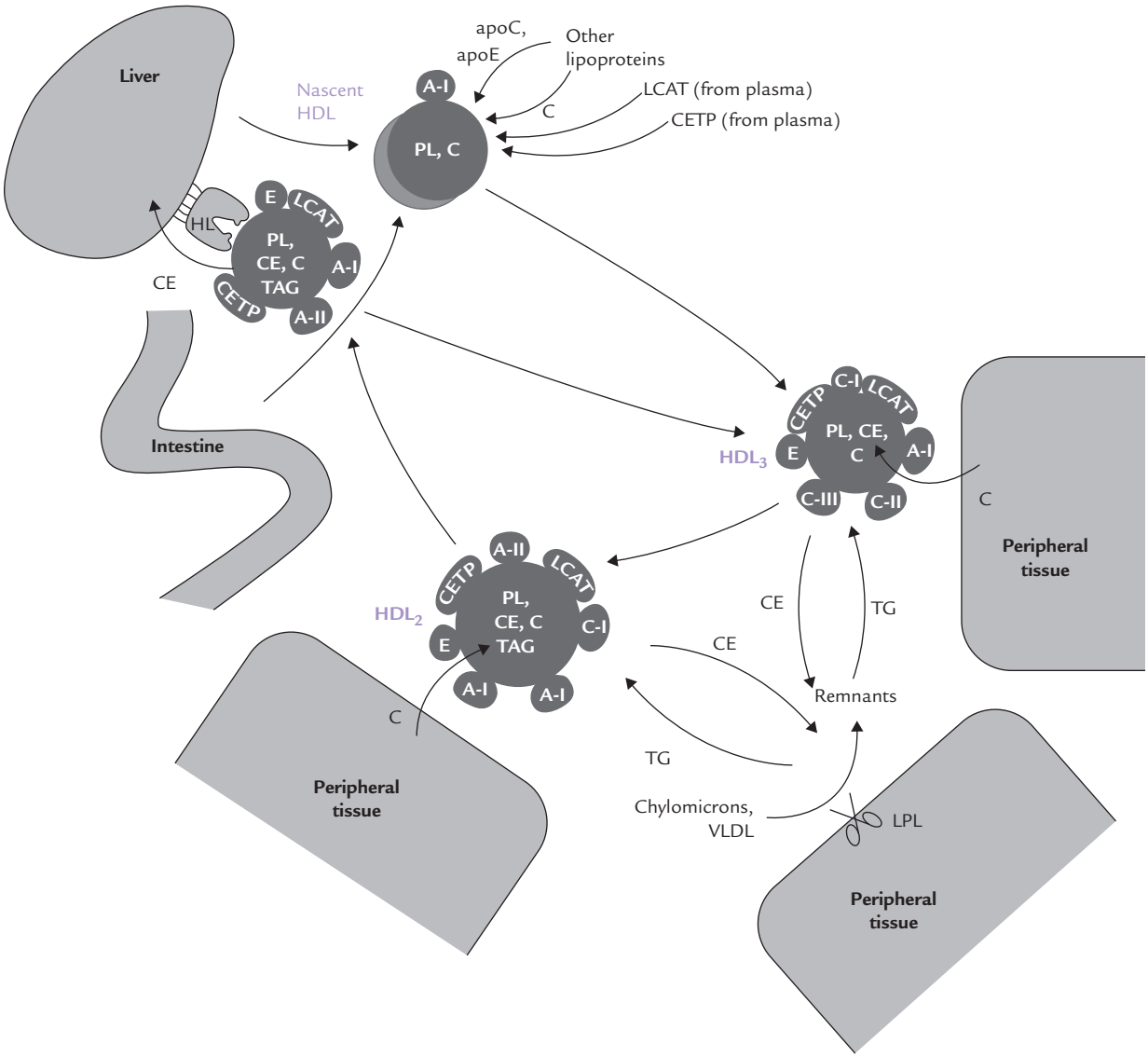


Fig. 12.6. Metabolism of HDL (TAG = triglyceride, PL = phospholipids, C = cholesterol esters, CETP = cholesterol ester transfer protein, LPL = lipoprotein, HL = *hepatic lipase*. All HDL apolipoproteins can be exchanged with other lipoprotein classes).

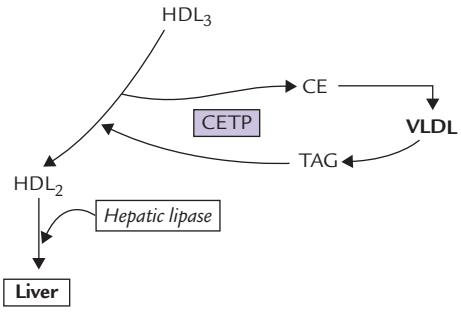


Fig. 12.7. Transformation of HDL_3 to HDL_2 by transfer of cholesterol esters to VLDL in exchange for triacylglycerols (CE = cholesterol esters, CETP = cholesterol ester transfer protein, HDL = high density lipoprotein, TAG = triacylglycerol, VLDL = very low density lipoprotein).

and are taken up by liver cells by apoE-mediated receptor mechanism for degradation.

Half-life of HDL is **5 days**.

HDLs extract cholesterol from tissues and convert it to cholesterol ester by action of LCAT. The HDLs then transfer this cholestrol ester to VLDLs, of which about half are taken up into liver cells, and their cholesterol excreted, mainly as bile salts.

Functions

1. HDL extracts cholesterol from cell membranes for their subsequent excretion. This process termed **reverse cholesterol transport** (Box 12.3), ultimately lowers the

BOX 12.3

Reverse Cholesterol Transport

HDL extracts cholesterol from extrahepatic tissues and carries them to liver for subsequent disposal; this is reverse cholesterol transport. There are two major routes for this:

1. By direct uptake of the HDL particles by hepatocytes.
2. By CETP-mediated transfer of cholesterol esters to lower density lipoproteins (mainly VLDL). These acceptors will ultimately undergo receptor-mediated endocytosis and hydrolysis in liver.

The latter mechanism is quantitatively more important in humans.

level of intracellular cholesterol (since the cholesterol esters stored in the cell will be mobilized to replace the cholesterol removed from the cell membrane).

2. HDL serves as the circulating reservoir of apolipoproteins, transferring apoC and apoE to the nascent VLDL and chylomicrons (Figs 12.3 and 12.4).

G. Lipoprotein Receptors

A number of receptors present on hepatocytes or peripheral cells play an important role in the metabolism of lipoproteins.

1. The **LDL receptor** protein is 839 amino acid long and spans the cell membrane. Its gene is located on chromosome 19. It binds with two apolipoproteins (apoE and apoB-100) and therefore participates in endocytosis of apoB-100-containing LDL particles, and apoE-containing IDL particles. ApoE binds to the LDL receptor with higher affinity than apoB-100. ApoB-48 does not bind to the LDL receptor, so chylomicrons uptake is not mediated by LDL receptor.
2. Chylomicrons use a different type of hepatic receptor which recognizes apoE and apoB-48 and belongs to LDL-receptor gene family.
3. A third lipoprotein receptor, **VLDL receptor** has a high degree of homology with the LDL receptor. It has been shown to bind apoE containing VLDL, though its exact function in humans is uncertain.
4. The fourth type of receptor binds apoE and mediates uptake of HDL in liver cells.

H. Lipoprotein(a)

Lipoprotein(a) has nearly identical structure as LDL. However, it contains an additional apolipoprotein molecule, called **apolipoprotein(a)**, which is covalently linked to apoB-100.

Presence of lipoprotein(a) in blood in excessive amount increases risk of coronary heart disease. This could possibly be due to structural resemblance of this apolipoprotein with plasminogen. These two proteins have about 80% sequence homology. Plasminogen is the precursor of a blood *protease*, **plasmin** that degrades fibrin, the major component of blood clots. Because of close structural resemblance, apolipoprotein(a) competes with plasminogen, thereby decreasing its conversion to plasmin.

Plasminogen ⊣► Plasmin

Reduced plasmin activity results in decreased degradation of clots. This, in turn, increases susceptibility of an individual to atherosclerosis.

Recent studies, however, do not substantiate this hypothesis. Clinical trials are underway to investigate this point.

I. Disorders of Plasma Lipoproteins (Dyslipoproteinaemia)

The disorders of lipoprotein metabolism are divided in two categories: hyperlipoproteinaemias and hypolipoproteinaemias.

Hyperlipoproteinaemia

There is elevation of circulating lipoprotein levels, which may be **primary** or **secondary** to some chronic disease (e.g. diabetes-mellitus, alcoholism, or hypothyroidism). In most cases, however, the hyperlipoproteinaemia is multifactorial, resulting from the interaction between a sedentary lifestyle and the genetic constitution. Most widely accepted **Fredrickson's classification** (later modified by WHO in 1970) catalogues hyperlipoproteinaemias into five types (Table 12.4), based on plasma levels of chylomicrons, VLDL, LDL, HDL, cholesterol and triacylglycerols after a 14-hour fast. These types are not diseases but rather phenotypes that occur in a variety of contexts, e.g. defect in apoproteins, receptors or

Table 12.4. The five primary hyperlipoproteinaemia phenotypes

Type	Name	Plasma lipids		Fraction	Incidence	Causes
		Triglyceride	Cholesterol			
I	Hyperchylo-micronaemia	↑↑↑	(↑)	Chylomicrons	Rare	Inherited deficiency of LPL or apoC-II, systemic lupus erythematosus, or unknown.
II	Hypercholes-terolaemia	(↑)	↑↑	LDL (and VLDL in IIb)	Common	(a) Primary: familial hypercholes-terolaemia; (b) secondary: obesity, poor dietary habits, hypothyroidism, diabetes mellitus, nephrotic syndrome.
III	Dysbetalipo-proteinaemia	↑	↑	Chylomicron remnants VLDL remnants	Rare	Homozygosity for $apoE_2$ (does not bind to hepatic apoE receptors), combined with poor dietary habits.
IV	Hypertrigly-ceridaemia	↑↑	↑	VLDL	Common	Diabetes mellitus, obesity, alcoholism, poor dietary habits.
V	Familial type V hyperlipo-proteinaemia	↑↑	↑	Chylomicrons, VLDL	Rare	Obesity, diabetes mellitus, alcoholism, oral contraceptives.

enzymes. All of these except the type I are associated with atheromatous vascular disease. The deposition of lipids in subcutaneous tissue leads to xanthomas. Cholesterol deposition in cornea leads to corneal arcus; it indicates hypercholesterolaemia.

Type I (Hyperchylomicronaemia)

It is a rare form (prevalence 1 in 10,000) in which *hydrolysis of chylomicrons is impaired* due either to

(a) deficiency of the enzyme *lipoprotein lipase,* or
(b) lack of apoC, which is required as an activator of LPL.

In other cases, no specific cause can be identified.

Chylomicron band is prominent in fasting plasma and plasma triacylglycerols are markedly elevated (close to 1000 mg/dL). The patients complain of abdominal pain, have eruptive cutaneous xanthomas, and at risk of developing dangerously lethal pancreatitis in case the severe hypertriglyceridaemia (exceeding 2000 mg/dL) persists.

Type IIa (Familial-hypercholesterolaemia)

Cellular uptake of *LDL is impaired* resulting in high LDL and **hypercholesterolaemia**. The underlying causes is *deficiency of functional LDL receptors in liver and extrahepatic tissues*. This pattern is a major risk factor for atherosclerosis and coronary heart disease (CHD). The inheritance is autosomal dominant. The receptor deficiency may be 50% of normal in affected heterozygotes, and in homozygotes very few receptors may be there or even absent altogether. Xanthomas, especially in the form of tendon xanthomas, develop in most patients, but more serious is coronary atherosclerosis, which results in premature death by CHD (see **Case 12.2**). Fortunately, homozygosity for familial-hypercholesterolaemia (FH) is rare: 1 in 10^6.

Recently, **four classes of mutations** have been defined in FH.

The class 1 defect (most common) results in loss of receptor synthesis (due to gene deletion—partial or complete).

The class 2 defect result in synthesis of receptor proteins that are poorly translocated to the plasma membrane.

The class 3 defect results in receptors that bind poorly to apoB-100.

The class 4 defects result in receptors that are not internalized properly after binding LDL.

Type IIb (Hypercholesterolaemia)

In contrast to major inherited abnormality of type IIa, this type includes multifactorial and secondary forms that are far more common, affecting 2% to 8% of the total population. There is combined elevation of LDL and VLDL, therefore, both cholesterol and triacylglycerols are elevated in plasma. Obesity, diabetes mellitus, hypothyroidism, nephrotic syndrome and high cholesterol diet and saturated fats are important causes of this pattern.

Type III (Dysbetalipoproteinaemia)

A genetic variant of apoE, termed $apoE_2$ is expressed, which cannot bind to the hepatic apoE receptors. This results in accumulation of remnant particles, e.g. IDL, chylomicron remnants, and IDL-like VLDL remnants in plasma, resulting in hypercholesterolaemia and hypertriglyceridaemia. Palmar xanthomas and tubo-eruptive xanthomas on knees, elbows and buttock are seen and there is high incidence of atheromatous vascular diseases.

The atherosclerotic lesions are seen in peripheral arteries and coronary arteries, and so risk of CHD is also increased.

Type IV (Familial endogenous hypertriglyceridaemia)

It is a very common type caused by overproduction of VLDL associated with glucose intolerance and hyperinsulinaemia. Serum triglyceride levels are markedly elevated but milder elevation of cholesterol is also seen, so that incidence of atherosclerosis is increased. These manifestations generally set in after the fourth decade of life. Weight gain and obesity, diabetes mellitus (type 2), impaired glucose tolerance, chronic alcoholism, oral contraceptives, and excess dietary carbohydrates (especially sugars) are all frequently associated with this condition.

Type V

There is combined increase of chylomicrons and VLDL (cause unknown), so triglyceride levels are increased. Mostly this pattern has a familial background, but may also be secondary to other causes, such as diabetes mellitus, obesity, alcoholism and renal disease.

Other Causes

Some other causes of hyperlipoproteinaemia, apart from those explained above, are as below:

1. *Familial combined hyperlipoproteinaemia:* A dominantly inherited gene that increases synthesis of apoB-100 is present in this condition. However, hyperlipoproteinaemia is multifactorial because, depending on a variety of genetic and environmental factors, either the type II or the type IV pattern is expressed in these patients. These two most common patterns may occur in different members of the same family.
2. *Familial LCAT deficiency:* Cholesterol cannot be esterified, so free-cholesterol : cholesterol-ester ratio is increased enormously in all classes of lipoproteins. The reverse cholesterol transport and consequently cholesterol excretion is impaired and a variable hypercholesterolaemia and hypertriglyceridaemia develops.
3. *CETP deficiency:* Reverse cholesterol transport is decreased in this benign condition. Affected homozygotes have approximately a fourfold elevation of HDL cholesterol, but LDL cholesterol is low or normal (40–150 mg/dL). The circulating HDL particles are oversized, with lot of cholesterol esters, but little triacylglycerols.
4. *Deficiency of hepatic lipase or cholesterol ester hydrolase in lysosomes* (***Wolman's disease***): It causes elevated VLDL levels in blood.
5. *Excess of Lp(a):* It often causes premature coronary heart disease (CHD).

Treatment of Hyperlipoproteinaemias

Dietary restrictions, hypolipidaemic drugs are used, and in secondary cases control of the underlying causes are aimed at. These have been described later in this chapter.

Hypolipoproteinaemias

In this group of disorders concentration of one or more lipoproteins in plasma is decreased. The commonest of these disorders are abetalipoproteinaemia, characterized by absence of betalipoprotein (LDL) because of defect in synthesis or secretion of apoB-100 containing lipoprotein; hypobetalipoproteinaemia in which inability to synthesize apoB-100 and apoB-48 is the underlying defect; and hypoalphalipoproteinaemia in which the underlying defect has not been clearly defined (**Case 12.3**).

Abetalipoproteinaemia

It is a rare autosomal recessive disorder in which the liver and the intestinal cells cannot incorporate lipids in the apoB containing lipoproteins, i.e. chylomicrons, VLDL, and LDL. These lipoproteins are either absent or greatly decreased in plasma. Extremely low plasma levels of cholesterol and triacylglycerols occur. Severe fat malabsorption, and accumulation of triacylglycerols in the intestinal mucosa and the liver occur. Absorption of fat-soluble vitamins is severely impaired resulting in degenerative changes in retinal (atypical form of retinitis pigmentosum), myopathy, and spiny projections on erythrocytes (*acanthocytes*). Many patients die in childhood.

Hypobetalipoproteinaemia

This is an autosomal dominant disorder. There is decreased synthesis of apoB due to apoB gene mutations (approximately 20 different mutations identified) and so the apoB-containing lipoproteins are synthesized at a lower rate. LDL concentrations are 10–20% of the normal, VLDL is slightly lower and HDL is normal. Plasma cholesterol and triacylglycerols are decreased. Clinical abnormalities are mild and the condition is compatible with life.

Hypoalphalipoproteinaemia

It is an autosomal recessive disorder in which HDL level is reduced to less than 5% of normal. It is also known as **Tangier disease** as it was first reported on the Tangier island. Although the primary genetic defect is not yet known, there is marked *deficiency of the major apolipoproteins of HDL* (apoA-I and apoA-II), probably because of accelerated catabolism. This results in decreased level of HDL in plasma and impaired reverse-transport of cholesterol. Cholesterol esters are accumulated in liver, spleen, tonsils, lymph nodes, Schwann cells, and the patients develop neuropathy, hepatospleenomegaly and

lymphadenopathy. An orange discolouration of the tonsils is the most obvious clinical sign.

Despite HDL deficiency and consequent impaired reverse cholesterol transport, there is only a mild tendency for early atherosclerosis because the decrease of LDL cholesterol reduce the need for reverse cholesterol transport (LDL levels are reduced to approximately one third of the normal).

Inherited defects in lipoprotein metabolism lead to deficiencies or excess of individual lipoprotein classes (hypo- or hyperlipoproteinaemias, respectively), which are severe diseases. Hyperlipoproteinaemias are grouped into five phenotypes.

II. Fatty Liver Syndrome

A. Overview

The total lipid content of a hepatocyte is around 50% of the total cell mass and triacylglycerols account for a third of it. When excessive amounts of triacylglycerols are accumulated in the liver to such an extent that it may be observed by naked eye, the condition is referred to as **fatty liver**. As the fat accumulation becomes chronic, fibrotic changes that later progress to liver cirrhosis are induced in the hepatic tissue.

Excessive accumulation of triacylglycerols may occur due to:

- Excessive synthesis of TAGs in liver.
- Decreased exportation of TAGs from liver.

Synthesis

In hepatocytes, the TAG molecules are synthesized from the fatty acids that are derived from two sources:

1. Hydrolysis of the TAG stores of adipose tissue by *hormone sensitive lipase*.
2. Hydrolysis of chylomicrons and VLDL triacylglycerols by *lipoprotein lipase*.

The synthesis of triacylglycerols in liver cells occurs within the cisternae of the endoplasmic reticulum.

Exportation

Subsequently, these triacylglycerol molecules are complexed with apolipoprotein B100 and other biomolecules to form very low density lipoprotein (i.e. VLDL) which then leave the liver cells to enter the blood circulation (Fig. 12.8).

Fig. 12.8. Balance between synthesis and exportation of triacylglycerols from hepatocyte. In fatty liver, either triacylglycerols are synthesized excessively or exported inadequately.

B. Causes of Fatty Liver

The TAG accumulation initially produces morphological changes in liver cells, called **fatty change**. Metabolic derangements and a host of agents, including chemicals as well as the bacterial or viral infections induce fatty change in liver cells.

As shown in Figure 12.8, the TAG accumulation may result from interference at some points in the metabolism in the liver cells: **either TAGs are synthesized excessively, or less TAGs are exported from hepatocytes**.

Excessive Synthesis of Triacylglycerols

It occurs when plasma fatty acid levels rise, and therefore, excessive amount of fatty acids enter the liver cells for TAG synthesis. For example, in *starvation* and *diabetes mellitus*, depot fats are mobilized and increased amount of fatty acids are brought to the liver where they are synthesized into triacylglycerols. *Intake of alcohol* causes elevation of plasma triacylglycerols and induces esterification of fatty acids; however, other mechanisms may also be involved.

Impaired Exportation of Triacylglycerols as VLDL

It may occur as a consequence of (a) diminished VLDL synthesis or (b) impaired release of VLDL into the circulation.

Major causes of impairment of the lipoprotein production are as here:

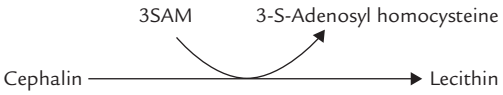
Inhibition of Protein Synthesis

Inhibition of protein synthesis in rough endoplasmic reticulum is caused by substances like puromycin, carbon tetrachloride, phosphorus, lead and arsenic. By inhibiting protein synthesis, these substances decrease availability of apolipoprotein B-100, which is the major apolipoprotein of VLDL. This results in decreased production of VLDL.

Impaired Formation of the Lipid Moiety

It occurs when there is a dietary deficiency of choline or its precursor methionine or betaine. These substances, known as the **lipotropic factors**, are required for the formation of phospholipids.

- *Choline:* It is an essential component of the phosphatidylcholine (lecithin), the major phospholipid of VLDL. Deficiency of choline, therefore, leads to diminished synthesis of VLDL.
- *Methionine:* Activated methionine, termed S-adenosyl methionine (SAM), participates in synthesis of lecithin by transmethylation reaction. It involves transfer of methyl groups from **active methionine** (S-adenosyl-methionine; SAM) to cephalin.



Therefore, methionine is actively involved in synthesis of lecithin and thus acts as a lipotropic substance.

Betaine which contains three methyl groups, also serves as a lipotropic factor.

Lipotropic factors are required for synthesis of phospholipid, and hence VLDL. Their deficiency leads to decreased VLDL synthesis—an important cause of fatty liver.

Deficiency of lipotropic factors leads to fatty liver by additional mechanism such as by inducing deficiency of phosphocholine and carnitine.

(a) *Phosphocholine deficiency:* Phosphocholine is required for stimulating glycosylation of apoB-100. Since this step is important for the production of nascent VLDL particles, its impairment leads to a decreased production of the latter.

(b) *Carnitine deficiency:* This compound contains three methyl groups and is required in β-oxidation. Its deficiency results in decreased utilization and hence increased cellular content of fatty acids, which generate excessive triacylglycerol deposits.

Fatty liver is an excessive accumulation of triacylglycerols in liver, which results because of imbalance in the rate of triacylglycerol synthesis (↑) and export (↓) from liver. Impaired export is due to decreased VLDL synthesis, which is mostly associated with the deficiency of lipotropic factors: choline, methionine, betaine.

C. Role of Chemicals and Other Agents

A number of chemicals and other agents are known to cause fatty liver. Substance like ***carbon tetrachloride, phosphorus, lead*** and ***arsenic*** interfere with the function of the ER, where assemblage of various components of VLDL occurs. They bring about disruption of the ER membrane through generation of free radicals. Protection against such damage is provided by antioxidants including *vitamin E*. When adequate *vitamin E* is not supplemented in the diet, it may lead to enhanced oxidative damage and thus fatty liver.

Carbon tetrachloride has an additional damaging effect: it causes an impaired secretion of VLDL from hepatocytes.

Ethionine acts by a competitive mechanism because of its structural similarity with methionine. It replaces methionine in S-adenosyl methionine (SAM). This traps the available adenine and thus prevents synthesis of ATP. Administration of ATP or adenine, therefore, reverses the inhibition caused by ethionine.

Orotic acid causes fatty liver by interfering with glycosylation of apoB-100. This inhibits the formation and release of nascent VLDL.

Clostridium diphtheriae exotoxin is also known to induce fatty liver. It acts by interfering with metabolism of carnitine.

***Deficiency of essential fatty acids (EFA)* can also lead to fatty liver** because EFA are needed for the synthesis of phospholipids. The second carbon position of the glycerol backbone is usually esterified with EFA. Deficiency of EFA therefore, inhibits formation of the phospholipids, which in turn inhibit synthesis of VLDL. Excess cholesterol also induces deficiency of EFAs. In this condition most EFAs are used for esterification of cholesterol and less are left for phospholipid synthesis.

Chronic alcoholism also causes fatty liver, which may later lead to cirrhosis. The triacylglycerols are accumulated due to increased endogenous synthesis of triacylglycerols following alcohol metabolism (Chapter 15).

Finally, the significance and interpretation of fatty change depends on the underlying pathogenic mechanisms. A milder form of fatty change may have no effect on the cellular function while it may be fatal if severe. In most cases the fatty change is reversible. However, if some vital intracellular process is irreversibly impaired, the consequences may be serious.

III. Metabolism of Cholesterol

A. Overview

Cholesterol (in Greek, *chole* means bile, *steros* means solid, *ol* means alcohol) is the most important sterol in animal cells. It is the only important membrane steroid in animals and humans. In a 70 kg man, approximately 140 g cholesterol is present, the majority of which is "free" (unesterified) cholesterol in the cellular membranes. Cholesterol is most abundant in those tissues that contain large amounts of membrane lipids, namely the nervous system; brain contains approximately 30 g cholesterol.

The condensed ring structure of cholesterol makes it rigid, and this is believed to alter the stiffness of membrane and its permeability.

Cholesterol molecule (27 carbons) comprises a non-aromatic ring system known as **cyclopentanoperhydrophenanthrene ring**, which is decorated with a hydroxyl group, two methyl groups and a branched hydrocarbon chain (Chapter 3). The hydroxyl group, present at C-3 of the ring, is esterified with a long-chain fatty acid, to form **cholesterol ester** (CE), which is used as intracellular storage form of cholesterol. CE is abundant in some steroid hormone producing tissues, particularly the adrenal cortex where the cholesterol ester stores may be so extensive that the cells resemble adipocytes in possessing visible lipid droplets. CEs are also prominent in the plasma lipoproteins, wherein approximately 70% of the cholesterol is esterified. Cholesterol esters make up about two third of the total circulating cholesterol.

Cholesterol is a product of animal metabolism. It is present in various foodstuffs of animal origin such as egg yolk, meat, and liver; *a vegetarian diet is essentially cholesterol free.* Plants contain other sterols, which are collectively called phytosterols. **Ergosterol** (mostly in fungi) and **beta-sitosterol** (in higher plants) are examples of phytosterols. They are poorly absorbed from dietary sources and therefore are present only in small amounts in the human body.

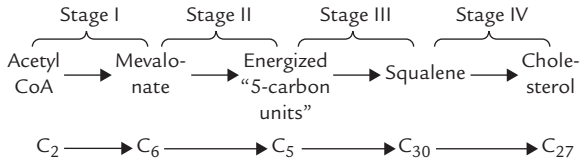
Cholesterol, the least soluble membrane lipid in animals and humans, is derived in part from dietary sources and in part from endogenous synthesis in liver and other tissues.

B. Biosynthesis of Cholesterol

All nucleated cells of the body can synthesize cholesterol, including liver, adrenal cortex, intestine, ovaries, testes, skin, arterial walls, etc., but **liver** is the most important site, accounting for at least 50% of the total. Some sterol hormone producing endocrine tissues, such as the adrenal cortex and the corpus luteum, have very high rates of cholesterol synthesis. Endogenous cholesterol synthesis amounts to 0.5–1 g per day, depending on the dietary supply. Low cholesterol diet enhances the synthesis, whereas high-cholesterol diet suppresses it.

Acetyl CoA is the source of all the 27 carbon atoms of cholesterol, as reported by **Konrad Bloch** in 1940. Formation of a 27-carbon molecule from a two-carbon precursor (i.e. acetyl CoA) involves a series of condensation reactions. These reactions require input of a large amount of energy. The energy is provided by the high energy thioester bonds, as also by the ATP molecules. These reactions also require reducing equivalents, which are provided by NADPH.

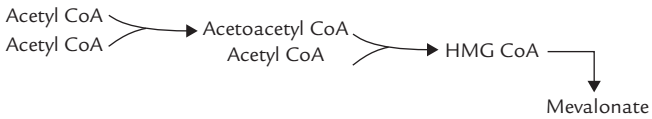
The enzymes of the biosynthetic pathway, which number close to 30, are found in the cytosol and the **endoplasmic reticulum** (ER). An outline of various stages of the pathway is as below:



Stage I: Formation of 6-C Compound, Mevalonate

By a sequential action of three enzymes, mevalonate is produced from acetyl CoA molecules. Two acetyl CoA molecules condense first to form a four-carbon compound the acetoacetyl CoA. The reaction is catalyzed by the enzyme *thiolase*. One more acetyl CoA molecule then condenses with the acetoacetyl CoA to form a six carbon compound called hydroxymethyl glutaryl CoA (HMG CoA). This reaction is catalyzed by *HMG CoA synthase* (Fig. 12.9). Liver parenchyma cells contain two types of *HMG CoA synthase: cytosolic* and *mitochondrial*. The cytosolic form participates in cholesterol synthesis, whereas the mitochondrial enzyme is involved in the synthesis of the ketone bodies.

In the next step, HMG CoA is reduced by the enzyme *HMG CoA reductase* to form mevalonate; two molecules of NADPH are used to provide the reducing equivalents. This step is ***rate-limiting*** for the pathway. Activity of *HMG CoA reductase* is regulated by a number of modulators, as discussed later in this chapter.



Like other biosynthetic reactions, these three reactions are also endergonic, which require input of energy. For this purpose, the precursors for these reactions are energized: acetyl CoA represents the energized form of acetate.

Stage II: Formation of Energized 5-C Units: Isopentenyl Pyrophosphate and Dimethylallyl Pyrophosphate

Mevalonate is further activated to form the energized *"5 carbon units"* which are capable of participating in the more complex endergonic reactions that follow. Formation of these 5-C units occurs in two phases:

- First, mevalonate is activated by three consecutive phosphorylation reactions to form mevalonate-3-phospho-5-pyrophosphate. These reactions are driven by the ATP energy.

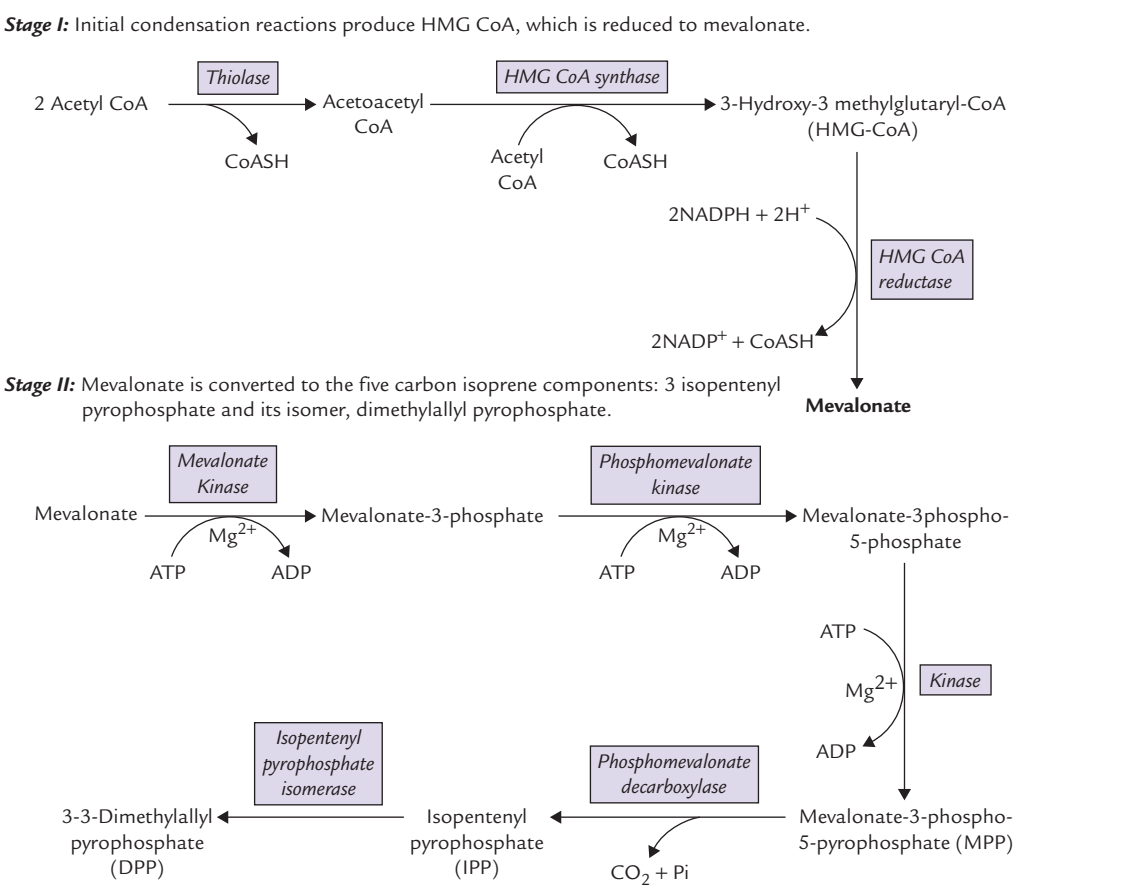


Fig. 12.9. Cholesterol biosynthesis.

- Decarboxylation of mevalonate-3-phospho-5-pyrophosphate follows to form the 5-carbon isopentenyl pyrophosphate (IPP). The reaction is again driven forward by energy provided by ATP. IPP is then isomerized to 3-3 dimethylallyl pyrophosphate (DPP).

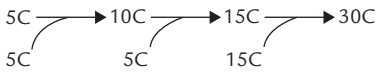
Both IPP and DPP are energized compounds and are therefore, highly reactive. They are referred to as the ***isoprene units***.

Stage III: Squalene Formation by Condensation of Six Isoprene Units

Formation of 30-C squalene from the 5-C (isoprene) units is evidently a complex process (Fig. 12.9). It involves successive condensation reaction as below:

1. IPP and DPP condense to form a ten carbon compound, **geranyl pyrophosphate** (GPP). The reaction is driven forward by pyrophosphate hydrolysis.

2. A second molecule of IPP condenses with GPP to form **farnesyl pyrophosphate** (FPP), a 15-carbon compound. This compound stands at branch point of various biosynthetic pathways; it can enter the pathways leading to biosynthesis of cholesterol, dolichol or ubiquinone.
3. Two molecules of FPP (15-C each) further condense to yield a 30-carbon compound, called squalene. *Squalene synthase* is the enzyme catalyzing this reaction. Two molecules of pyrophosphate are released in this step (one from each of the two molecules of farnesyl pyrophosphate), which ensures irreversibility of this complex reaction. Two NADPH molecules are used for providing the reducing equivalents.



Stage IV: Formation of 27-C Cholesterol from Squalene

A series of modifications in the ER produce cholesterol from squalene. In order to understand these modifications, it is essential to know the structural differences between these two compounds. These are:

1. Squalene contains 30 carbon atoms, whereas cholesterol is a 27-carbon compound.
2. Squalene is a linear molecule, whereas cholesterol has a cyclic ring structure.
3. Squalene contains six double bonds, whereas cholesterol contains only one.
4. Squalene does not contain a hydroxyl group, whereas cholesterol has one at the third carbon position.

Therefore, the following changes are introduced in the squalene molecule during its conversion to cholesterol:

- Loss of 3 carbon atoms by decarboxylation reactions
- Cyclization of structure
- Reduction of the double bonds by NADPH molecules
- Hydroxylation involving the third carbon

Squalene to cholesterol conversion is a multi-step process which has been proposed to include 19 different enzymatic reactions. All the enzymes catalyzing these reactions are located in the **endoplasmic reticulum**. This is in contrast to enzymes of the stage I, II and III reactions which are located in **cytosol**.

Some of the important intermediates of the pathway are lanosterol, zymosterol and desmosterol (Fig. 12.9).

The committed step in cholesterogenesis is formation of mevalonate from HMG CoA (enzyme: *HMG CoA reductase*). Mevalonate is processed to the branched-chain 5-C compounds (isoprene units), six of which condense to form the C-30 compound, squalene, which cyclizes to yield lanosterol—the steroid precursor of cholesterol.

C. Regulation of Cholesterol Biosynthesis

The rate limiting enzyme of cholesterol biosynthesis, *HMG CoA reductase*, is an intrinsic membrane protein in the endoplasmic reticulum. Its active site extends into the cytosol. Activity of this enzyme is regulated by a number of factors:

Feedback Inhibition

Activity of *HMG CoA reductase* is inhibited by cholesterol. Thus, amount of cholesterol in diet significantly influences the endogenous cholesterol synthesis; average diet contains 300–400 mg of cholesterol, and the synthesis increases when cholesterol in diet is low.

Hormonal Regulation

The cAMP-directed cascade, which is involved in regulation of glycogen metabolism controls the activity of *HMG CoA reductase* on short-term basis (Fig. 12.10). For example, glucagon increases intracellular cAMP level, which activates the enzyme *reductase kinase*. The latter in turn favours phosphorylation of the *HMG CoA reductase*. Since phosphorylated form of this enzyme is inactive, rate of cholesterol biosynthesis decreases by this sequence of reactions. In contrast, insulin favours dephosphorylation

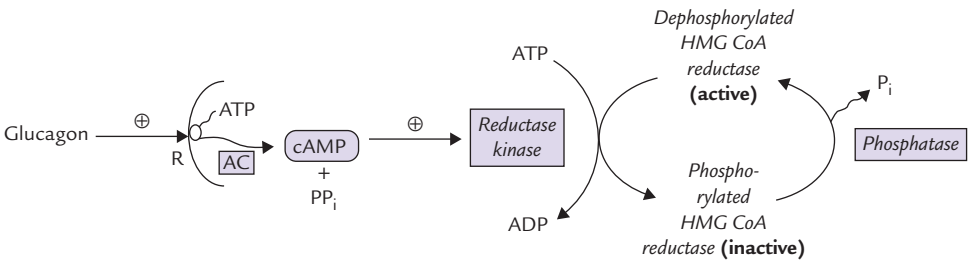


Fig. 12.10. Regulation of activity of *HMG CoA reductase* by cAMP cascade (AC = *adenylate cyclase*, R = receptor).

and hence enhances formation of the active (dephosphorylated) enzyme. This results in an increased rate of cholesterol biosynthesis.

Sterol-mediated Regulation of Transcription

Rise in intracellular cholesterol concentration inhibits cholesterol biosynthesis by decreasing the transcription of the *HMG CoA reductase* gene. The effect may also be mediated at post-transcriptional level. Thus, intracellular levels of cholesterol, when elevated, inhibit both the activity as well as the synthesis of *HMG CoA reductase*. Since HMG CoA has a lifespan of approximately 4 hours, a change in its rate of synthesis can effect cholesterol synthesis rather rapidly.

HMG CoA reductase, the rate limiting enzyme, is regulated through **short-term** control by phosphorylation, as well as through **long-term** control of its rate of synthesis.

Inhibition by Drugs

Some drugs inhibit cholesterol synthesis by acting as reversible competitive inhibitors of *HMG CoA reductase*. Examples include **mevastatin** and **lovastatin**. They are used as cholesterol lowering agents in patients having elevated serum cholesterol concentration. Action of various hypolipidaemic drugs is described later in this chapter.

D. Cholesterol Homeostasis in General

Despite popular belief, cholesterol is not a health-hazard, but an essential component of cell membranes and the basis of several steroid hormones. It creates a problem only if it is in excess and in this respect the maintenance of cholesterol homeostasis is of vital significance. Normally, the body has to deal with about **one gram** of cholesterol a day (Fig. 12.11), derived from both exogenous sources (i.e. **dietary intake**) and endogenous sources (i.e. **de novo synthesis**).

Normally, the de novo synthesis and dietary sources make roughly equal contributions of about 500 mg each. Total quantity (1 g a day) remains fairly constant because if dietary intake changes, compensatory changes occur in the biosynthesis of the sterol in the tissues.

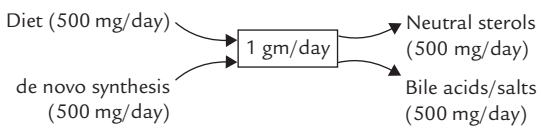


Fig. 12.11. Cholesterol can either be obtained in diet or synthesized in liver, to a total quantity = 1 gram/day. Bile acids (salts) and neutral sterols are the major excretory forms of cholesterol.

Homeostasis is maintained by the loss of cholesterol metabolites in faeces: as bile acids or neutral sterols.

1. ***Cholesterol from diet and de novo synthesis:*** Dietary cholesterol is absorbed in form of **chylomicrons**, which are delivered to blood via lymphatic system and then presented to the liver. Plant sterols decrease intestinal absorption of cholesterol. The absorption depends on the presence of bile salts and is influenced by the amount of dietary cholesterol: it may be above 60% on a low cholesterol diet, but an absorption of 30% is more typical for a high-cholesterol diet.

   As mentioned earlier, there is a tight control of biosynthesis so that it is responsive to changes in dietary supplies.
2. ***Excretion of cholesterol:*** Daily excretion of cholesterol amounts to about one gram, which corresponds to daily cholesterol load from diet plus endogenous synthesis. In humans the sterol ring of cholesterol cannot be cleaved. Therefore, the cholesterol has to be disposed off after some modifications in its side chain. These modifications result in formation of **neutral sterols** and **bile acids**.

Neutral sterols: Cholesterol present in bladder bile or that present in epithelial cells desquamated into intestinal flora serve as precursor for the neutral sterols. **Coprostenol** is the principal neutral sterol, formed from cholesterol in the lower intestine by the bacterial flora therein, and is eliminated in faeces. The fecal sterols account for elimination of 500 mg cholesterol each day.

Bile acids (salts): These are formed from cholesterol in hepatocytes and secreted actively into biliary canaliculi. They aid the digestion of dietary lipids. A majority of the secreted bile acids are recovered via the enterohepatic circulation, but daily faecal losses equivalent to 500 mg cholesterol occurs (Fig. 12.12).

Metabolism of Bile Acids and Salts

The 24-carbon bile acids are synthesized in hepatocytes and stored in the gall bladder, and secreted via biliary canaliculi into intestine where they play a vital role in lipid digestion. They are more appropriately called bile salts because they are present in the deprotonated form at pH of 7.0.

Synthesis

About half of the cholesterol in the body is ultimately metabolized to bile acids. The **primary bile acids**, which are synthesized in the liver include *cholic acid* and *chenodeoxycholic acid*, cholic acid being more abundant. They are synthesized from cholesterol by:

(a) *Hydroxylation* – addition of hydroxyl groups to the ring, (b) *Chain cleavage* – removal of a 3-carbon fragment from the hydrocarbon chain attached at C-17 of the steroid

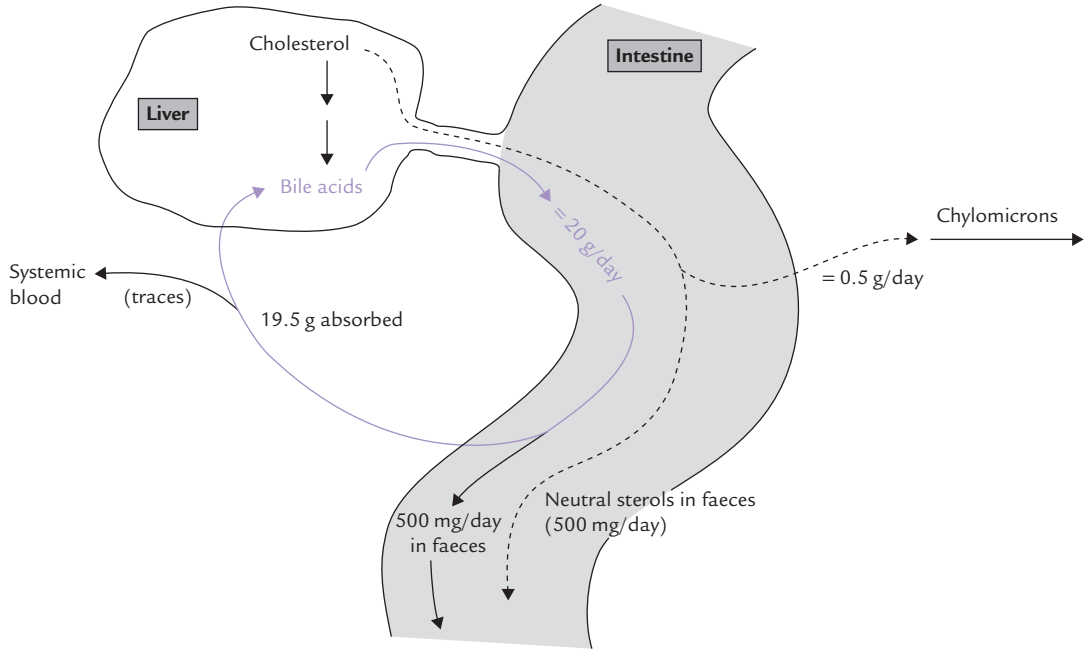


Fig. 12.12. Disposition of cholesterol in enterohepatic system of about 20 grams bile acids daily secreted into intestine, about 19.5 grams are recovered by enterohepatic circulation. Fecal loss equivalent to 500 mg cholesterol occurs as bile acids, and another 500 mg as neutral sterols.

nucleus, (c) *Reduction* of the B-ring of the steroid nucleus, and (d) *Oxidation* of the terminal carbon of the hydrocarbon chain to carboxyl group.

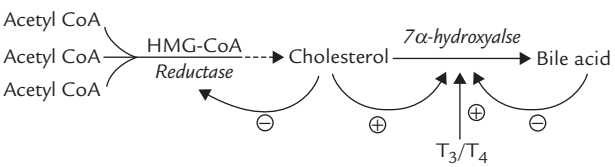
Liver releases the bile acids not in the free form but as conjugation products with glycine or taurine (Fig. 12.13).

The *primary bile acids*, cholic acid and chenodeoxycholic acids are conjugated with glycine taurine, and are excreted through the bile, where they exist as sodium or potassium salts of bile acids, called **bile salts.**

Structures of primary bile acids are shown in Figure 12.14.

Regulation of Biosynthesis

- Synthesis of *7-α-hydroxylase*, the rate limiting enzyme, is inhibited by bile acids (via reducing transcription of its gene).
- Dietary cholesterol induces the transcription of this enzyme.



Because cholesterol also exerts a stimulatory effect on *7-α-hydroxylase*, it not only inhibits further synthesis of cholesterol, but also channels the existing cholesterol for bile acid synthesis. These regulatory mechanisms maintain an adequate pool of free cholesterol in the liver.

Thyroid hormones (T_3, T_4) induce transcription of *7-α-hydroxylase*; thus *in patients with hypothyroidism plasma cholesterol tends to rise.*

Secondary bile acids: The primary bile acids are modified in the lower parts of small intestine by bacterial enzymes to form secondary bile acids, e.g. **deoxycholic acid** and **lithocholic acid**. Only a proportion of primary bile acids are converted into secondary bile acids, a process which requires:

1. Deconjugation by hydrolysis of the amide link to glycine or taurine.
2. Removal of the 7-α-hydroxyl group.

In these reactions, cholic acid is converted to deoxycholic acid and chenodeoxycholic acid to lithocholic acid (Fig. 12.15). They are mostly returned to the liver (98%) through the portal circulation where they are conjugated and secreted in the bile. Because of this enterohepatic circulation, discussed below, both primary and secondary bile acids are present in the bile.

Functions of bile acids: Bile acids act as detergents (because they contain polar and non-polar regions) assisting the emulsification of ingested lipids into very small globules; this aids enzymatic digestion of dietary fats. They form molecular aggregates called micelle, which respectively bring about the enzymatic digestion and absorption of dietary fat (Chapter 26).

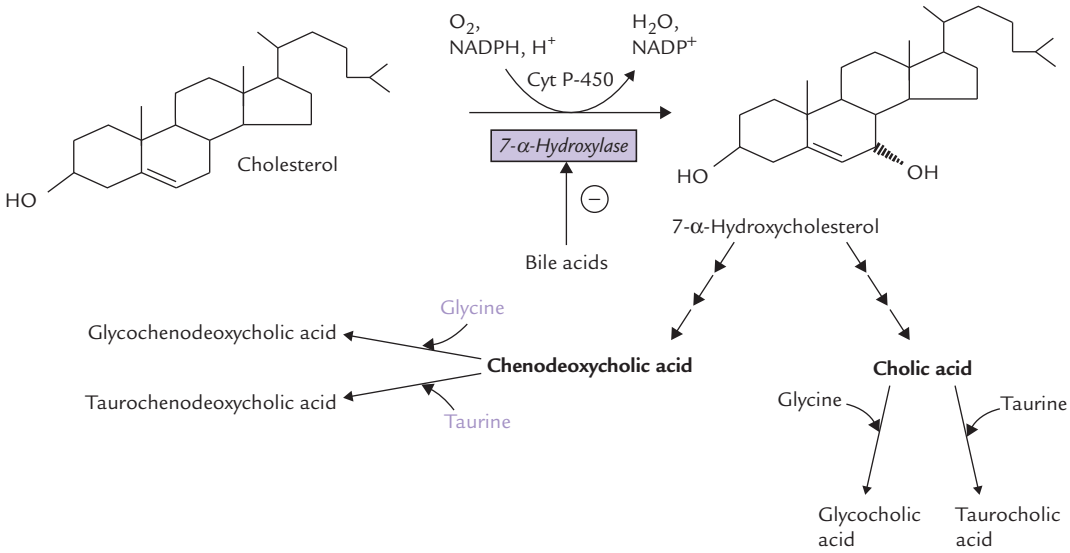


Fig. 12.13. Synthesis of the primary bile acids by ER-associated enzymes in hepatocytes. Cholesterol is converted into cholic acid, which combines with either glycine (to form glycocholic acid) or taurine (to form taurocholic acid). The other primary bile acid, chenodeoxycholic acid, also forms similar conjugated products, which have stronger detergent effect.

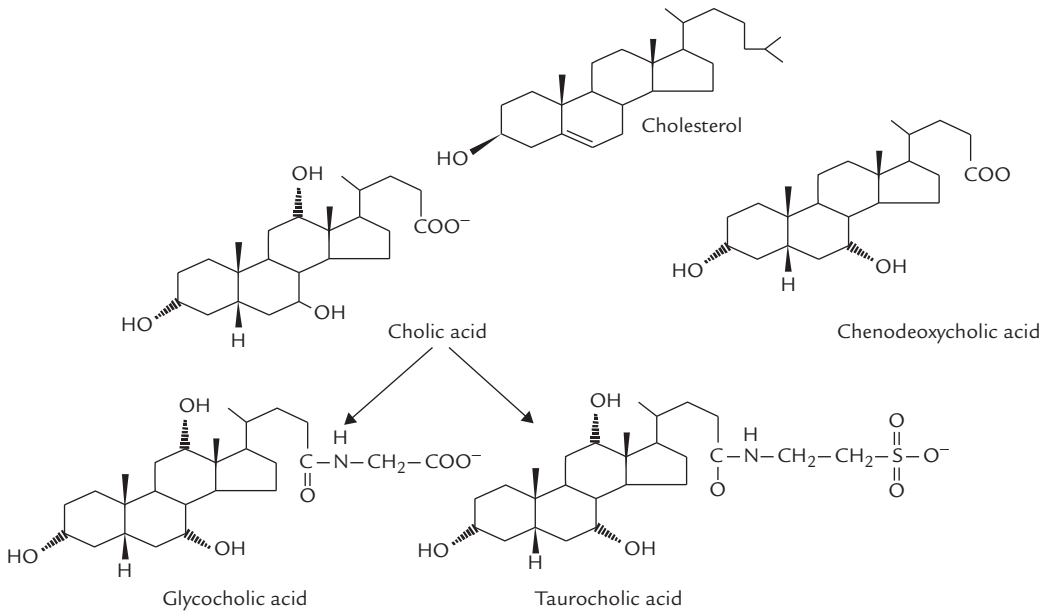


Fig. 12.14. Structures of primary bile acids compared with cholesterol.

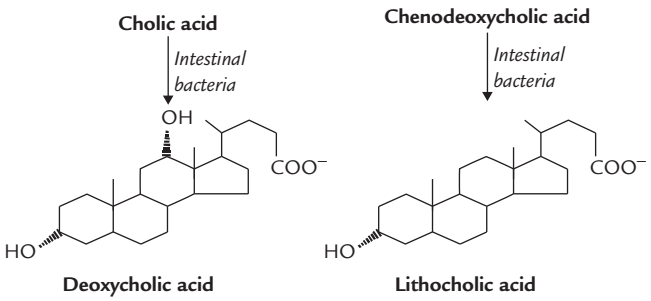


Fig. 12.15. Synthesis of the secondary bile acids by intestinal bacteria.

Enterohepatic circulation: Approximately 20 g bile acids pass from the bile duct into the intestine each day but only 2% of this (500 mg) is lost through faeces (Fig. 12.12). The rest (19.5 g) is deconjugated and reabsorbed. Passive reabsorption of bile acids occurs in the jejunum and colon but the majority is reabsorbed in the ileum by a process of active transport. The reabsorbed bile acid molecules are non-covalently bound to albumin and transported in blood via the portal vein to liver, and finally re-secreted through bile.

This enterohepatic circulation of bile acids allows large amounts of bile acids (20–30 g) to be delivered to

the intestine daily from a relatively small total bile acid pool (3–5 g). Only 500 mg is synthesized per day but an average bile acid molecule is recycled five to eight times every day, and it remains in the system for approximately a week before it is finally excreted. The daily loss and daily synthesis (500 mg each) balance each other.

Excess cholesterol can be excreted in bile, but approximately half of the total cholesterol is converted to bile acids and excreted. Bile acids are subject to extensive enterohepatic circulation.

E. Specialized Products Derived from Cholesterol

Cholesterol is not only an essential component of cell membrane, but also a precursor of the following biologically active compounds.

1. *Vitamin D:* Formation of vitamin D_3 occurs by ultraviolet irradiation of the skin. It converts 7-dehydrocholesterol into choecalciferol by spontaneous cleaving of the B-ring of the cholesterol.

2. *Adrenocortical hormone:* The steroid hormones of the adrenal cortex are produced from cholesterol by a series of changes (Chapter 30).

3. *Sex hormones:* The male sex hormone, testosterone (19-C), and the female sex hormone, oestrogen (18-C) are also derived from cholesterol.

4. *Bile acids and bile salts:* These have already been described.

IV. Atherosclerosis

Atherosclerosis is a condition in which arteries are blocked to a greater or lesser extent by the deposition of cholesterol plaques, leading most commonly to *coronary heart disease* by blocking of coronary arteries. Involvement of cerebral arteries causes *cerebral thrombosis*, and of arteries supplying blood to lower limbs to *peripheral vascular diseases*.

A. Role of Hyperlipidaemia in Atherosclerotic Diseases

Several studies throughout the world have provided substantial evidence that *raised plasma concentration of lipid parameters are linked to atheromatous vascular diseases* (although a vast number of individuals manifesting hyperlipidaemia do not suffer from one of these monogenic diseases). This provides a major stimulus for estimation of plasma lipids in clinical practice. As a preliminary to the discussion on the role of certain forms of hyperlipidaemias in causing atherosclerosis, it is useful to consider normal lipid profile (hereafter called atherogenic profile) and its interpretation.

B. Atherogenic Profile

The lipid parameters commonly estimated are: total cholesterol (TC), LDL-cholesterol, HDL-cholesterol, triacylglycerols and phospholipids. The sample should be collected only after 12–14 hours of fasting. **Total plasma lipids** are estimated to be **400–600 mg/dL**, of which, roughly speaking, one-third is cholesterol, one-third is triacylglycerols and one-third is phospholipids (Table 12.1). Several methods are available for the estimation of total cholesterol, such as Liebermann–Burchard reaction, Carr and Dructor method and, more recently *cholesterol oxidase* method. HDL-cholesterol is estimated after adding polyethylene glycol (PEG), which precipitates LDL and VLDL. VLDL is equivalent to one-fifth of the plasma triacylglycerols in the fasted state. LDL-cholesterol can be calculated by the following formula:

$$\text{LDL-cholesterol} = \text{Total cholesterol} - (\text{HDL-cholesterol} + \text{TAG}/5).$$

Cholesterol is present in plasma mainly in *LDL* (*70%*; the "**bad cholesterol**") and in smaller amounts in HDL ("the **good cholesterol**"). Normal values of total cholesterol and LDL- and HDL-cholesterol, and the associated risk of atherosclerosis in adults are shown in Table 12.5.

The values mentioned in Table 12.5 are recommended by the American Heart Foundation, but cannot be applied to Indians for Indians have higher incidence of other risk factors, such as central obesity with high waist: hip ratio and Lp(a). Lower values of total cholesterol (< 180 mg/dl) are therefore desirable.

In addition to the usual lipid parameters described above, measurement of apolipoproteins are carried out in more advanced set-ups for additional information.

Table 12.5. The usual lipid parameters and associated risk of atherosclerosis

Cholesterol level	Desirable	Moderate risk	High risk
Total	< 200 mg/dL	200–239 mg/dL	> 240 mg/dL
LDL	< 130 mg/dL	130–159 mg/dL	> 160 mg/dL
HDL	35–59 mg/dL	–	< 35 mg/dL
(Low risk: 60 mg/dL of HDL)			

It has been reported that plasma levels of apoproteins A-I and B are more informative than HDL and LDL. A **rise in A-I** *indicates a decreased risk of CHD whereas* **a rise in apoB** *indicates an increased risk.* A ratio of A-I : A-III exceeding 3.0 is interpreted as highly cardioprotective.

Age Related Changes in Atherogenic Profile

Age related changes are observed in the above parameters; for example, after the fourth decade of life plasma cholesterol levels tend to slowly rise in men and postmenopausal women. Serum level of LDL for adult males of 30–40 years = 80–175, and for 40–60 years = 90–200 mg/dL have been reported.

Pathological Variations

Serum cholesterol is subject to variations in a number of conditions. Increased concentration (**hypercholesterolaemia**) is characteristically seen in the following conditions:

1. ***Hypothyroidism:*** This is due to decreased synthesis of *7-α-hydroxylase* that causes conversion of cholesterol to bile acids (thyroid hormones induce synthesis of this enzyme). Moreover, there is decreased expression of HDL receptors on hepatocytes in hypothyroidism.

2. ***Diabetes mellitus:*** This is because of increased formation of acetyl CoA (by β-oxidation), the precursor of cholesterol synthesis.

3. ***Obstructive jaundice:*** Impaired excretion of cholesterol through bile leads to plasma accumulation.

4. ***Nephrotic syndrome:*** Urinary loss of albumin leads to compensatory increase in synthesis of globulins. When lipoproteins are thus increased, there is a corresponding elevation in cholesterol.

C. The Assessment of Cardiovascular Risk

Several epidemiologic studies and clinical trials have demonstrated that there is a relationship between the levels of lipids and lipoproteins in plasma and the incidence of atherosclerosis and coronary heart disease. Though total plasma cholesterol level is believed to be the prime risk factor, other lipid parameters also play a role. Furthermore, several studies have shown that a sustained reduction in plasma cholesterol levels leads to reduced mortality and morbidity from coronary heart disease.

1. ***Role of cholesterol:*** The risk of CHD is related to plasma levels of total cholesterol and LDL-cholesterol. **Framingham Heart Study** with 30 years of follow up reported a linear increase in coronary "risk" when the total cholesterol level crosses 200 mg/dL. Conversely, a 1% decrease in total plasma cholesterol predicts a 2% reduction in CHD risk. The Multiple Risk Factor Intervention Trial in the USA also suggests a link between prevalence of CHD and serum cholesterol levels.

2. ***Role of LDL and HDL:*** LDL carries about 70% the of total cholesterol, transporting them to tissues, hence enhancing the "risk" of CHD, and HDL has a beneficial, cardioprotective role. Recent studies have shown that at any given level of LDL the probability of CHD increases as the level of HDL decreases. The prevalence of CHD at 35 mg/dL plasma HDL-cholesterol is eight times the CHD rate of persons with HDL levels of 65 mg/dL or greater. In addition to removing cholesterol via the hepatobiliary route, the other **antiatherogenic effects of HDL** cholesterol include:

- Inhibition of conversion of LDL to oxidized LDL, which is preferentially taken up by the tissue macrophages to form fatty streaks.
- Prevention of adhesion of monocytes to the endothelium.
- Prolongation of the half-life of prostacyclin produced by endothelial cells; this promotes vasodilatory effects.

Finally, decreased HDL levels are associated with male sex, obesity, physical inactivity, cigarette smoking and hypertriglyceridaemia—the other risk factors for CAD, described later.

3. ***Role of triacylglycerols:*** Some of clinical studies suggest that an increased triacylglycerides level is associated with an increased risk for CHD.

Atherosclerosis is a multifactorial disease characterized by accumulation of atheromatous plaques, consisting of a core of cholesterol esters surrounded by an area of fibrosis. Lipoproteins are important risk factors: elevated VLDL, IDL or LDL are associated with atherosclerosis, whereas high levels of HDL have a protective effect.

4. ***Other contributing risk factors:*** Lipids are not the only factors that determine the development of atherosclerotic disease. A number of other risk factors are also involved, as summarized in Table 12.6. These factors have additive effect on the risk of CHD, meaning that presence of these risk factors increases the risk associated with dyslipidaemia. Therefore, the desirable cholesterol levels are lower in individuals with other risk factors, such as diabetes or hypertension.

Table 12.6. Risk factors for atherosclerotic diseases

Risk factor	Comment	Remarks
Male sex	Female hormonal background provides a measure of protection, so the difference in "risk" equalizes in post-menopausal women.	–
Sedentary life	Greatly increases risk. Beneficial effects of exercise are related to: lipid-lowering and physiological processes, e.g. dilatation of the arteries of heart, increased blood flow.	Lifestyle changes
Age	Increased risk with age may relate to age-related decline in physical activity, and other factors.	–
Smoking	Nicotine causes transient vasoconstriction of coronary and carotid arteries, and also enhances adipose lipolysis to increase free fatty acids and serum cholesterol.	Smoking cessation
Stress	Catecholamines mobilize free fatty acids from adipose tissue. Increase synthesis of VLDL.	Lifestyle changes
Diet	Dietary fat, particularly cholesterol and saturated fats are especially hazardous.	Diet low in saturated fat, and where appropriate, cholesterol-lowering drugs
Apo(a)	Independent risk factor for CHD.	–
Homocysteine	Increased level in blood is a risk factor for CHD (normal < 200 mg/dL).	–
Hypertension	Systolic > 160 mmHg is a major risk factor for stroke, and a risk for CHD. Roughly an increase by 10 mmHg reduces life expectancy by 10 years.	Control blood pressure, diet, drugs
Obesity	Greatly enhances risk of CHD.	Attain ideal body weight
Diabetes	TAG stores of adipose mobilized; dyslipidaemia common; glycosylation of LDL prolongs half-life, and oxidized LDL more readily taken up, thus enhancing risk of atherosclerosis. Cardiovascular disease is the main cause of death in diabetes.	

CHD = coronary heart diseases, LDL = low density lipoproteins, TAG = triacylglycerols, VLDL = very low density lipoproteins.

D. Pathogenesis and Consequences of Atherosclerosis

Atherosclerosis is a set of pathological changes in intima of large arteries that can lead to serious and often fatal complications. The characteristic lesion in the intima is atheromatous plaque having a core of soft lipid (*athere*), surrounded by an area of fibrous (*sclerosis*) component. The lipid core mainly consists of cholesterol esters which are derived from LDL particles. The latter enter intima by transcytosis where they are endocytosed by macrophages through the scavenger receptors. Since the scavenger receptor is not regulated, the macrophages become overloaded with cholesterol and become **foam cells**. The conglomerates of foam cells form **fatty streaks**, yellow patches visible in the arterial wall. Fatty streaks are an early reversible lesion: most of the fatty streaks regress spontaneously but some may develop into **atheromatous plaques** by fibroproliferative response.

The **fibroproliferative response** starts when surrounding smooth muscle cells and the endothelial cells start secreting a range of small peptides such as cytokines and growth factors that regulate cell growth. Examples include platelet-derived growth factor, interleukin-1, and tumour necrosis factor. These peptides stimulate smooth muscle cells to proliferate and to migrate toward the luminal side of the arterial wall. The proliferating muscle cells start synthesizing extracellular matrix, particularly collagen. These changes (migration of smooth-muscle cells and accumulation of collagen rich extracellular matrix) result in the formation of a cap that surrounds the lipid rich core. This is mature atherosclerotic plaque. The plaque protrudes into the arterial lumen, which impedes the blood flow and makes it turbulent. The turbulent flow increases the risk of clot formation. In due course of time the atheromatous plaques may get dislodged because macrophages and T-cells residing at the edges of the plaque secrete enzymes (called *metalloproteinases*) that degrade extracellular matrix. The degradation exposes collagen to the blood stream. This leads to adherence and aggregation of platelets to the exposed collagen and initiates the formation of blood clot. The resulting occlusion of a major blood vessel leads to ischaemic death of the affected tissue.

Consequences are serious; for example, block in coronary artery manifests as acute myocardial infraction, blockage in arteries supplying blood to brain causes stroke and block in arteries supplying to lower limbs results in peripheral vascular diseases.

E. Management of Hyperlipidaemia: Biochemical Basis

The management guidelines for hyperlipidaemia rely heavily on biochemical results. The aim is to reduce cholesterol

and triacylglycerols to **acceptable limits:** *total cholesterol* < 200 *mg/dL, LDL* < 130 *mg/dL, HDL* > 40 *mg/dL* and *TAG* < 140 *mg/dL.*

F. Dietary Management

Some of the management guidelines are:

1. **Reduce calorie intake** (coupled with exercises) to achieve ideal body weight. In addition to beneficial effect on lipid profile, this frequently improves glucose tolerance and lowers blood pressure.

2. **Reduce total fat intake** to provide only about 30% of total calories.

3. **Reduce saturated fat intake** to provide only about 30% of total fat intake, and instead take polyunsaturated fatty acids since they are required in esterification and final excretion of cholesterol, thereby reducing serum cholesterol level.

4. In essence, it is recommended that consumption of red meat, dairy products and fried food be reduced while that of dietary fibres (fruits and vegetables), pulses be increased. Fibres increase intestinal mobility and reduce reabsorption of bile salts. Increased excretion of bile acids leads to increased conversion of cholesterol to bile acids and an up-regulation of hepatic LDL receptors, resulting in reduction in plasma LDL levels. Vegetables also contain plant sterols (sitosterol) which decrease the absorption of cholesterol. Moderate alcohol may raise HDL level.

Drug Therapy

Drug therapy may be considered only if the patient does not adequately respond to dietary management and other lifestyle changes. Currently, a range of **lipid-lowering drugs** are available. More commonly (though erroneously) referred to as hypocholesterolaemic drugs, they are divided in three classes based on their action.

1. ***Bile acid sequestrant resins*** *(colestipol* and *cholestyramine):* They are insoluble, non-absorbable anion exchangers which bind bile acids in lumen of intestine. This blocks their reabsorption and interrupts their enterohepatic circulation, so that instead of returning to the liver, the bile salts are excreted in stools. Feedback inhibition of *7-α-hydroxylase* by bile salts is also released, so that conversion of cholesterol to bile salts is increased. The resulting depletion of the cellular cholesterol pool leads to an up-regulation of hepatic LDL receptors and a reduction of the LDL level in the plasma.

2. ***Inhibitors of HMG CoA reductase:*** These are the commonest drugs for the treatment of primary hypercholesterolaemia with least side effects, and are also called **statins**. By causing enzyme inhibition, they decrease endogenous cholesterol synthesis, so that the cells become dependent on the supply of cholesterol from plasma lipoproteins. Decreased endogenous synthesis also causes increased synthesis of LDL receptors. By this mechanism, the HMG CoA inhibitors reduce LDL cholesterol. **Lovastatin** and **mevastatin** are the oldest members of this class; the newer ones are pravastatin, simvastatin, fluvastatin, etc.

3. ***Fibrates:*** These are the commonest drugs for the treatment of primary hypertriglyceridaemia or combined hyperlipidaemia. Commonest drugs of this group are **clofibrate, fenfibrate** and **gemfibrozil**. They activate *lipoprotein lipase* to lower triglyceride, total and LDL cholesterol, and in some cases increase HDL cholesterol.

Nicotinic acid in large doses decreases the serum cholesterol level, though the mechanism is not clear.

Antioxidants

The oxidation of LDL facilitates its uptake through macrophage scavenger system, so antioxidants are expected to reduce lipid accumulation in tissue macrophages. Antioxidant vitamins, E and C, are not only non-toxic, but also less expensive than lipid lowering drugs.

V. Prostaglandins

A. Overview

Prostaglandins (PG) are a group of 20-carbon compounds derived from arachidonic acid, which act as **local hormones**. They elicit a vast range of biochemical effects, and are widely used as therapeutic agents.

Prostaglandins and related compounds—*prostacyclins, thromboxanes* and *leukotrienes* are collectively known as **eicosanoids** because they are all C-20 compounds (Greek: *eikosi* means twenty). Prostaglandins contain a five-membered cyclopentane ring. Prostacyclins (also called prostaglandin I) contain a five-membered oxygen containing ring, and the thromboxanes have a six-membered oxygen-containing ring instead of the cyclopentane ring. Eicosanoids act at very low concentrations and are involved in the production of pain and fever, and in the regulation of blood pressure, blood coagulation and reproduction.

They tend to act locally, close to the cells that produce them, and are metabolized rapidly (within seconds or minutes), which limits their effects on nearby tissues.

Prostaglandins (and other eicosanoids) have been often referred to as local hormones. However, they **differ from the true hormones** in some important aspects:

1. ***Site of production:*** Prostaglandins are produced by **almost all tissues** (except erythrocytes) in contrast to

hormones which are produced in specialized endocrine organs.

2. ***Site of action:*** Prostaglandins generally act locally unlike the hormones which are transported through blood circulation to the specific target tissues where their action is elicited.

3. ***Metabolism:*** Prostaglandins are not stored, and are metabolized to inactive produces near the site of their synthesis only; this is in contrast to the hormones which are metabolized at sites distant from the site of origin.

Prostaglandins are 20-carbon lipids that are derived from arachidonic acid and related polyunsaturated fatty acids. Most of them are short lived and act only locally in their tissue of origin.

B. Historical Perspective

Prostaglandins were discovered in 1933 by *Ulf von Euler*, a Swedish scientist. It had been observed that muscle contraction was induced in uterus by human semen, and von Euler speculated that this action was elicited by a group of compounds originally synthesized in prostate. Hence, he named them prostaglandins. Though it was discovered later that prostaglandins were synthesized in almost all tissues, the name prostaglandin was by then widely accepted, and hence continued to be used.

Biosynthetic pathway for prostaglandins was elucidated by *Bengt Samuelsson* (Nobel Prize, 1982). *Elias Corey* was awarded Nobel Prize in 1990 for chemical synthesis of prostaglandins.

C. Chemical Nature

Chemically, the prostaglandins are derivatives of the hypothetical parent compound, *prostanoic acid* which resembles a 20-carbon saturated fatty acid, but has a cyclic five membered ring (Fig. 12.16). This **cyclopentane ring** is formed by carbon atoms 8 to 12.

Numerous **types of prostaglandins** are known, which are designated by alphabetical letters (A, B, C, etc.) according to the nature of the substituent on the cyclopentane ring (Table 12.7). They are characterized further by subscript indicating the number of double bonds outside the cyclopentane ring (Fig. 12.17).

The different types of prostaglandins, as designated by the alphabetical letter, have different and sometimes antagonistic biological effects. The number of double bonds, on the other hand, affects the potency of the product but not the kind of effect on a particular target tissue.

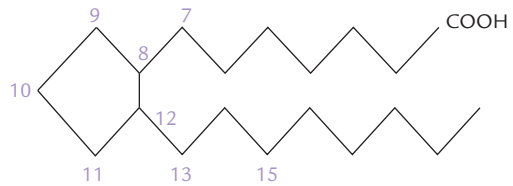


Fig. 12.16. Structure of prostanoic acid. The cyclopentane ring is formed by carbons 8 to 12.

Table 12.7. Different prostaglandins named according to substituent groups on their cyclopentane ring

Prostaglandins	Substituent group	Structure
PGE	Keto group at C-9; OH group at C-11	PGE, PGD
PGD	OH group at C-9, keto group at C-11	
PGF	OH groups at C-9 and C-11	PGF, PGG
PGG	Ring has two oxygen atoms bonded at C-9 and C-11, substituent hydroperoxide group at C-15	

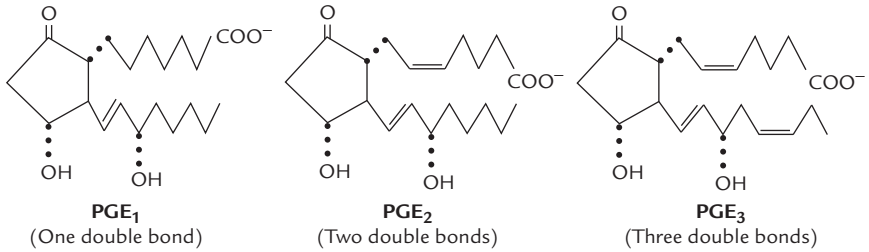


Fig. 12.17. Prostaglandins are characterized by subscript indicating number of double bonds outside the cyclopentane ring. Number of double bonds affects the potency of the product.

D. Metabolism

Synthesis

In humans, the most important precursor for prostaglandins is **arachidonic acid**, a polyunsaturated fatty acid with four double bonds (*eicosatetraenoic acid*). It is stored in cell membranes as the C2 ester of phosphatidylinositol and other phospholipids. The biosynthetic pathway, called *cyclooxygenase* **pathway**, occurs in the following steps:

1. ***Release of the arachidonate from the membrane bound phospholipids:*** It occurs by action of *phospholipase* A_2. This reaction is *rate-limiting* for the pathway and is promoted by epinephrine and bradykinin.

2. ***Oxidation and cyclization of arachidonic acid:*** It produces PGG_2, which is then converted to PGH_2 (Fig. 12.18). The reactions are brought about by the microsomal ***prostaglandin synthase*** **complex**, which consists of the *cyclooxygenase* and *peroxide components*. The overall reaction consumes two moles of oxygen and two moles of reduced glutathione. The enzyme is short lived and gets rapidly inactivated, which prevents excessive production of prostaglandins.

   PGH_2 is the first prostaglandin to arise in the *cyclooxygenase* pathway, hence called primary prostaglandin.

3. ***Conversion to primary prostaglandin to other eicosanoids:*** PGH_2 serves as the immediate precursor for a number of other prostaglandins, as also of prostacyclins and thromboxanes. The enzymes for the conversions of H-prostaglandin to the other prostanoids are *cell specific*. Each cell type produces only one or a few major prostanoids. Thus, **vascular endothelium produces PGE and PGI, and platelets produce thromboxanes.**

   Figure 12.18 shows various enzymes responsible for various eicosanoids from PGH_2.

Note: All naturally occurring prostaglandins derived from arachidonic acid belong to **series 2**. They have two double bonds outside cyclopentane indicated by the subscript 2 (e.g. PGE_2). In addition, two other series of prostaglandins are known:

- Those derived from linolenic acid belong to **series 1**, (having single double bond), and
- Those derived from eicosapentaenoic acid belong to **series 3** (having three double bonds).

Catabolism

Prostaglandins are short lived, they act locally and are rapidly inactivated mainly by the enzyme *15-hydroxyprostaglandin dehydrogenase* at the site of their production. Another enzyme, *13-prostaglandin reductase*, is also involved in PG catabolism, but is of lesser significance.

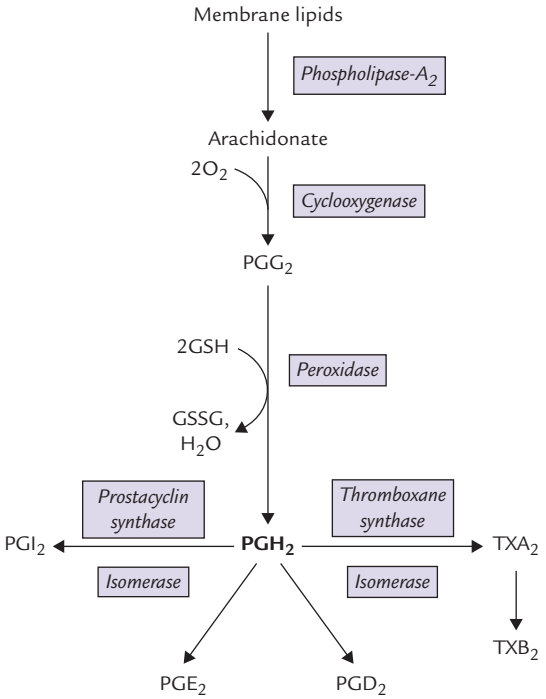


Fig. 12.18. Synthesis of prostaglandins (PGG_2, PGH_2, PGD_2, PGE_2, PGI_2) and thromboxanes (TXA_2, TXB_2) by *cyclooxygenase pathway*. The key enzyme is the microsomal *prostaglandin synthase* complex, which has *cyclooxygenase* and *peroxidase* components (PG = prostaglandin, TX = thromboxane, GSH = reduced glutathione, GSSG = oxidized glutathione).

E. Anti-inflammatory Drugs Inhibit Prostaglandin Synthesis

Two important classes of anti-inflammatory drugs are in current use: the steroids (glucocorticoids) and the non-steroidal anti-inflammatory drugs (NSAIDs). Both of these drug classes act on the synthesis of prostaglandins.

- The **steroids** inhibit the action of *phospholipase* A_2, which normally is rate limiting in eicosanoid synthesis, and
- **NSAIDs** are inhibitors of *cyclooxygenase* (Fig. 12.19).

The steroids reduce the synthesis of all eicosanoids, but the NSAIDs inhibit synthesis of only prostaglandins. Synthesis of leukotrienes (formed from arachidonate by the *lipoxygenase* pathway) is not affected and may even be increased because more arachidonic acid is available for the *lipoxygenase* pathway.

Aspirin Irreversibly Inhibits Cyclooxyganase Activity

Prostaglandins are potent mediators of inflammation, pain and fever; and aspirin (acetyl salicylic acid) is an inhibitor of prostaglandin synthesis. Therefore, aspirin is widely used as an anti-inflammatory, analgesic (pain relieving), and antipyretic (fever reducing) agent. Its use

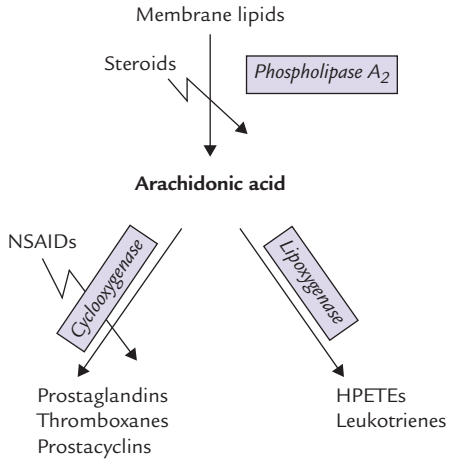


Fig. 12.19. Inhibitors of synthesis of prostaglandins and related compounds. NSAIDs (non-steroidal anti-inflammator drugs, e.g. aspirin, indomethacin, ibuprofen).

was prevalent much before its exact mechanism of action was elucidated in 1971 by John Vane. Aspirin acts by acetylating a specific serine residue of the *PG synthase*, which prevents the arachidonate from reaching the active site of this enzyme. This causes irreversible inhibition of *the cyclooxygenase* activity of the enzyme, thereby inhibiting prostaglandin synthesis.

Another useful effect of aspirin is decreased production of thromboxane A_2, which results in decreased vasoconstriction and platelet aggregation. This explains the utility of aspirin in prevention of heart attack.

However, since aspirin may enhance the *lipoxygenase pathway*, it is contraindicated in asthma, a disease in which bronchoconstriction is induced by leukotrienes rather than prostaglandins.

Prostaglandins are important mediators of inflammation, and agents that inhibit their synthesis are important anti-inflammatory drugs.

F. Biological Action and Uses

Prostaglandins participate in several physiological processes. They are involved in production of pain and fever, in regulation of blood pressure, blood coagulation, and reproduction. Almost every cell in the human body responds to one or several types of prostaglandins; often opposing actions are induced in the same tissue by different PGs. Only some of the more important functions of prostaglandins are discussed here.

1. *Control of blood pressure:* PGE and PGI produced by vascular endothelium are potent vasodilators. This lowers peripheral resistance to decrease the blood pressure. However, these PGs cannot be used as antihypertensives because of their rapid inactivation and numerous extravascular effects.

PGE_1 is used in infants with pulmonary stenosis to keep patency of the ductus arteriosus till surgical correction can be performed.

2. *Mediators of inflammation:* PGE_2 and PGE_1 are synthesized in macrophages and other cells as local mediators of inflammation (they produce inflammation mainly by increasing capillary permeability). Together with cytokines, histamine and bradykinin, they mediate the cardinal signs of inflammation.

3. *Platelet aggregation and thrombosis:* PGI (prostacyclins) inhibit platelet aggregation, and PGE_2 promotes it. Thromboxanes also promote platelet aggregation and blood clotting, which may lead to thrombosis, as described in the next section.

4. *Reproduction:* PGE_2 and PGF_2, which are produced in the endometrium, induce uterine contraction. Their levels increase massively at parturition, and together with oxytocin they participate in the normal *induction of labour.* They may be used for this purpose and for the *medical termination of pregnancy* via intrauterine, intravenous or intra-amniotic administration. Unlike oxytocin, the prostaglandins contract the uterus not only at term of pregnancy but also at all times, so they are suitable for the inducing abortion.

5. *Regulation of gastric secretion:* Synthetic prostaglandins of PGE_2 series inhibit gastric secretion, which makes them therapeutically useful in treatment of acid peptic disease. There is evidence that healing of the gastric ulcers is also accelerated by the prostaglandins.

6. *Relief of asthma:* Prostaglandins are bronchodilators and, hence may be of pharmacological use in bronchial asthma.

7. *Influence on renal functions:* Prostaglandins increase the tubular reabsorption of water and sodium.

8. *Effects on metabolism:* Prostaglandins inhibit lipolysis in adipose tissue.

9. *Control of fever response:* Fever is mediated by interleukin-1 (IL-1), which is released from stimulated monocytes and macrophages. IL-1 binds to vascular receptors in the preoptic area of the hypothalamus, where it induces the formation of PGE_2. The latter causes fever by a direct action on the thermoregulatory centre.

Mechanism of action of prostaglandins: The actions of prostaglandins are mediated through second messenger systems:

- The relaxation of vascular smooth muscles by *prostaglandins-E* and *-I* is mediated by increases in the *cAMP* level.

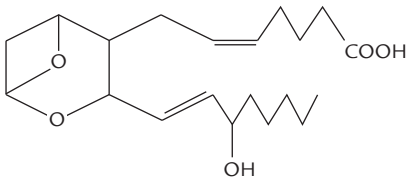


Fig. 12.20. Structure of thromboxane.

- Thromboxane-induced vasoconstriction is mediated by increased intracellular calcium and decreased cAMP.
- PGF acts by increasing intracellular cGMP level.

G. Thromboxanes, Prostacyclins and Leukotrienes

1. ***Thromboxanes A_2*** are structurally similar to prostaglandins, but have an oxygen atom in the cyclic ring (Fig. 12.20). Synthesis of TX_2, the major thromboxane, occurs in platelets from PGH_2 by the enzyme *thromboxane synthase* (Fig. 12.18). It has a half-life of about 30 seconds in blood; it is rapidly transformed into the inactive thromboxane B_2.

Its major effects are vasoconstriction and platelet aggregation. The effect on platelet aggregation is mediated via cAMP. Dipyrimadole, an inhibitor of *thromboxane synthase*, is a widely used therapeutic agent.

2. ***Prostacyclins*** are formed in vascular endothelium in aorta and probably in other artery walls. Action of the prostacyclins is mediated via cAMP. Their action is antagonistic to thromboxane and they are inhibitors of platelet aggregation. They are highly potent; as little as 1 mg/mL of PGI_2 is capable of preventing platelet aggregation. This has a protective effect on vessel wall against deposition of platelets, hence minimizing the risk of thrombosis. But any injury to vessel wall would inhibit the prostacyclin synthesis so that platelet aggregation occurs to promote thrombus formation at the site of injury.

In addition, prostacyclins cause relaxation of vascular smooth muscles. Because of these actions, the prostacyclins decrease the risk of myocardial infarction.

Thromboxanes and prostacyclins are antagonistic. Whereas thromboxane TXA_2 (as also prostaglandin E_1) promotes platelet aggregation, the prostacyclin PGI_2 inhibits platelet aggregation.

3. ***Leukotrienes (LT)*** are 20-carbon polyenoic fatty acids having a number of substituents. Depending on the substituents, they are divided into LTA, LTB, LTC, LTD and LTE. Each type is subdivided depending on the number of double bonds, which vary between one and five. The most important leukotrienes are LTB_4, LTC_4, LTD_4, and LTE_4 (each has four double bonds).

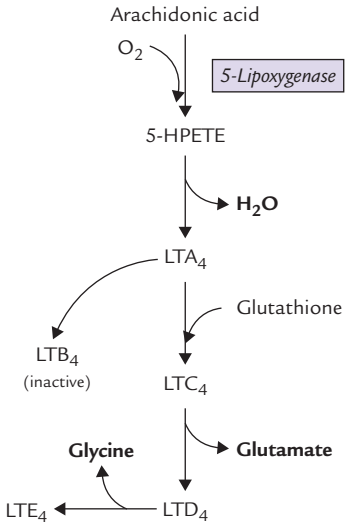


Fig. 12.21. Conversion of 5-HPETE, the product of the *5-lipoxygenase*, to leukotrienes.

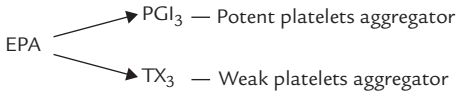
Leukotrienes are produced from arachidonic acid by an alternate pathway called *lipoxygenase pathway* (Fig. 12.19). The *lipoxygenases* are *dioxygenases* that produce hydroperoxy derivatives known as hydroxyperoxy eicosatetraenoic acids (**HPETEs**). Formation of various leukotrienes then occurs from these intermediates in *leukocytes, platelets, mast cells,* and *the vascular tissue of heart and lungs*. Synthesis of various leukotrienes from the intermediate 5-HPETE is shown in Figure 12.21.

Leukotrienes are powerful constrictors of bronchial and intestinal smooth muscles. They are mainly involved in *hypersensitivity reactions* and *inflammation*. Unlike the other eicosanoids, leukotrienes are relatively stable and persist for a few hours in the tissues. The **slow reacting substance of anaphylaxis** (SRS-A) has three leukotrienes as its active component (e.g. LTC_4, LTD_4, and LDE_4). It is far more potent than histamine (or prostaglandins) in stimulating allergic reactions. In addition, LTB_4 attracts leucocytes at the site of inflammation.

Leukotrienes, produced from arachidonic acid by *lipoxygenase* pathway, are involved in hypersensitivity reactions and inflammation.

Fish foods are cardioprotective. Certain fish oils are rich in an unsaturated fatty acid, namely eicosapentaenoic acid (EPA; 5 double bonds at carbons 5, 8, 11, 14, 17). The prostacyclin derived from EPA (PGI_3) is a highly potent anti aggregator of platelets the thromboxane (TX_3) derived from it is a weak aggregator. As a result the anti-aggregating effect gains predominance, which accounts

for lower incidence of myocardial infarction among Eskimos.



Moreover, both PGI_3 and TX_3 inhibit the release of arachidonate from phospholipids, thereby decreasing the formation of the aggregators, e.g. PG_2 and TX_2.

Exercises

Essay type questions

1. Describe synthesis of cholesterol and its regulation. Add a note on the significance of plasma cholesterol estimation.
2. Why do serum cholesterol levels depend on LDL receptor activity? Discuss the role of cholesterol in atherosclerosis and describe the mechanism of action of hypocholesterolaemic drugs.
3. Describe eicosanoids and their functions. Which drugs act by inhibiting eicosanoid synthesis?
4. Describe the synthesis and catabolism of various lipoprotein fractions. Explain role of various lipoproteins in atherogenesis.

Write short notes on

1. LACT
2. Enterohepatic circulation of bile acids
3. Hypolipidaemic drugs
4. Reverse cholesterol transport
5. Familial hypercholesterolaemia
6. Comparative features of hormone *sensitive lipase* and *lipoprotein lipase*
7. LDL-Receptor
8. Lipoprotein(a)
9. Lipotropic factors
10. Primary and secondary bile acids
11. Hypolipoproteinaemias
12. Fatty liver

CLINICAL CASES

CASE 12.1 Pinkish blood, milky serum, and yellow patches on trunk and limbs

A 10-year-old boy was admitted to the hospital emergency with severe abdominal pain. He had suffered the pain throughout the night. In the past, he suffered from diffuse abdominal pain very frequently. On examination, numerous yellow patches (xanthomas) were seen on his trunk and limbs. Blood sample was obtained for biochemical investigations. The blood was noted to be pink, the serum was milky and after centrifugation a thick band of 'cream' collected at the top.

Investigation test	Patient's reports	Reference range
Blood glucose	82 mg/dl	< 100 mg/dl
Serum cholesterol	272 mg/dl	150–250 mg/dl
Serum triacylglycerol	598 mg/dl	80–140 mg/dl

Electrophoresis of plasma lipoproteins suggested that chylomicron levels were highly elevated (confirmed by further tests, e.g. ELISA). However, there was deficiency of apolipoprotein C-II, as detected by the monoclonal antibodies specific for this apolipoprotein. A tentative diagnosis was made and for diagnostic confirmation free fatty acid levels were measured in plasma, 10 minutes and 20 minutes following intravenous administration of heparin. In contrast to normal response of a control subject, little fatty acids appeared in the patient's plasma.

Q.1. What is the most probable diagnosis in this case?
Q.2. What will be the effect of this condition on the other lipoprotein fractions?
Q.3. Interpret results of the heparin test.
Q.4. Will this child benefit from a diet containing short and medium chain fatty acids?
Q.5. In another child of the same age group, plasma chylomicron level was low. Given that this condition is also related to abnormality of apolipoprotein, metabolism of which apolipoprotein fraction would most likely be deficient in this child?
Q.6. If the apoB-48 were found elevated in intestinal cells of child, what is the most likely biochemical defect?

CASE 12.2 A 7-year-old child with fatty eruptions

A 7-year-old girl was brought to the hospital OPD with fatty eruptions on the right elbow and Achilles tendon. These were identified as subcutaneous xanthomas. Her mother informed that the child's father, had similar eruptions. The latter had died of heart attack at a young age of 32.

Examination showed bilateral corneal arcus, but there was no visual impairment. Blood sample was obtained for biochemical analysis.

Investigation test	Patient's reports	Reference range
Blood glucose (Fasting)	84 mg/dl	< 100 mg/dl
Serum cholesterol	486 mg/dl	150–250 mg/dl
LDL cholesterol	318 mg/dl	90–150 mg/dl
HDL cholesterol	25 mg/dl	30–70 mg/dl
Serum triacylglycerols	118 mg/dl	80–140 mg/dl
Electrophoresis showed excess of the β-lipoproteins.		

The child was put on the standard cholesterol free diet, with minimum saturated fatty acids. Six weeks later the above tests were repeated, but no fall in the elevated lipid parameters was observed. Soon after she suffered a heart attack.

Q.1. What is your the probable diagnosis? How do you confirm this diagnosis?
Q.2. Comment on the biochemical test results.
Q.3. The child was synthesizing 40 mg/kg bw (body weight) of apoB per day, which is double than that synthesized by a normal subject. Provide a biochemical explanation for this observation.
Q.4. Enumerate the risks associated with this condition.
Q.5. Discuss the rationale behind treating this type of disorder by the following measures; dietary restriction, lovastatin and ileal bypass.
Q.6. Liver transplant surgery was considered for a child. What is rationale behind performing the liver transplant surgery? How would this effect the apoB synthesis

CASE 12.3 Fat intolerance and dyslipidaemia in a 4-year-old girl

A 4-year-old girl was brought to medical OPD when she developed diarrhoea following a rich meal in a fast food restaurant. Examination showed that her gait was unsteady and there was generalized muscle weakness.

Test	Patient's reports	Reference range
Plasma glucose	80 mg/dl	< 140 mg/dl
Serum cholesterol	68 mg/dl	150–250 mg/dl
Serum triacylglycerol	90 mg/dl	40–140 mg/dl

Circulating levels of both the VLDL and the LDL fractions were low. A liver biopsy was obtained, cell-free extract was prepared and treated with polyclonal antibodies raised against apoB-100. Precipitation occurred but the amount of precipitate formed was much lower than that expected in a normal person.

Q.1. Identify the biochemical defect in this child.
Q.2. If polyclonal antibodies against apoB-100 were raised, do these antibodies react with apoB-48? Give reason.
Q.3. What is the effect of this disorder on the level of serum HDL?

CHAPTER

AMINO ACID AND PROTEIN METABOLISM

13

Amino acids are the building block molecules of body proteins. In addition, they serve as precursors of a variety of biologically important compounds including peptides, polyamines, purines, pyrimidines, etc. The neurotransmitters such as dopamine, epinephrine and norepinephrine are also synthesized from amino acids. Catabolism of amino acids provides carbon skeleton for gluconeogenesis, ketogenesis, as also for the energy yielding pathways.

In this chapter, various aspects of amino acid and protein metabolism are described. After going through this chapter, the student should be able to understand:

- Nitrogen metabolism; essential, non-essential and semi-essential amino acids, protein turnover and amino acid pool.
- Amino group metabolism; transamination, oxidative deamination, transport of ammonia, ammonia toxicity and urea cycle.
- Metabolism of carbon skeleton of amino acids; outline pathways of degradation of amino acids.
- Disorders of amino acid metabolism; phenylketonuria, alkaptonuria, maple syrup urine disease, isovaleric acidaemia, methyl malonic aciduria, histidinaemia, homocystinuria, cystinuria, cystinosis. Hartnup disease, blue diaper syndrome and familial renal iminoglycinuria.
- Biosynthesis of amino acids.
- Amino acids as precursors of specialized products.

There are 20 primary amino acids which are classified into three categories: non-essential, essential and semi-essential. There are eight primary amino acids (viz., *lysine, leucine, isoleucine, methionine, phenylalanine, tryptophan, valine* and *threonine*) which cannot be endogenously synthesized, and therefore, need to be obtained from the diet: they are classified as **essential** amino acids. The **non-essential** amino acids can be synthesized in body and hence their dietary intake is not essential. These amino acids can be produced from the essential amino acids, as also from intermediates of certain metabolic pathways, or even by interconversion among themselves. For example, tyrosine, a non-essential amino acid, is produced by hydroxylation of an essential amino acid, phenylalanine. Similarly glutamate can be formed by amination of α-ketoglutarate.

Two amino acids, *arginine* and *histidine*, are classified as **semi-essential** amino acids. Arginine can be synthesized in adequate amounts in adults, but in growing children additional dietary supplementation is required. This is because the body requirement for this amino acid is higher in the growing stage. Histidine is not synthesized in the human body and yet it is not essential in diet. This is because of its relatively lower requirement and a fairly high reserve in the body, especially in form of carnosine in muscles.

The lower living forms like microorganisms and plants are capable of synthesizing the essential amino acids. During evolution, animals lost the genes that encoded the enzymes necessary for the synthesis of these amino acids. This implies that the *essential amino acids are relatively less important for the body*, since animals could survive without synthesizing them.

I. Nitrogen Metabolism

A. Protein Turnover

Most proteins in the body are constantly and repetitively synthesized and degraded. In adults, the rate of synthesis is just sufficient to replace the protein that is degraded, so that the total amount of proteins in the body remains nearly the same. This process, called **protein turnover**, involves about 1–2% of body proteins per day (about 12 kg proteins are present in an adult of 70 kg body weight).

The rate of protein turnover shows wide variation for individual proteins. For example:

- Collagen and lens crystalline are degraded slowly, having very long half-life, whereas
- *HMG CoA reductase* has half-life of just about 2 hours only.

Protein turnover results in degradation of proteins to amino acids. About 80% of the amino acids so liberated are recycled, whereas the rest are catabolized. The average lifespan of an amino acid in plasma is only about 5 minutes.

B. Amino Acid Pool

The amino acids released as a result of protein degradation join the pre-existing amino acids distributed throughout the body. Together, they constitute the amino acid pool. The ingested (dietary) proteins, as also the endogenously synthesized (non-essential) amino acids, contribute to this pool (Fig. 13.1). Although this pool of about 100 g of amino acids appears trivial compared to the total amount of proteins in the body, it is of vital significance because of its dynamic nature. It can be channeled in various pathways depending on body requirements.

- A major proportion of the amino acids that join this pool every day are recaptured for the synthesis of body proteins and a variety of biologically important compounds (i.e. special products), as discussed later in this chapter.
- Amino acids corresponding to about 30 g of proteins are catabolized every day and their amino group nitrogen converted to **urea** for excretion.

Since the amino acids of this pool are readily available, deficiency of amino acids (especially the non-essential amino acids) does not manifest immediately. This is because amino acids are constantly replenished from the pool.

Note: Urea elimination represents a net loss of nitrogen from the body, which must be compensated by dietary intake of proteins.

Nitrogen Balance

Nitrogen balance is the difference between ingested nitrogen and excreted nitrogen. Because dietary proteins are an important source of nitrogen (100 g proteins contain 16 g of nitrogen), nitrogen balance is an important index of protein and amino acid metabolism. It is determined by comparing the nitrogen entering the body with that leaving it.

In healthy adults, a state of **nitrogen equilibrium** exists, meaning that nitrogen intake equals nitrogen excretion. In growing children and pregnant women in whom nitrogen intake exceeds excretion, **positive nitrogen balance** occurs. Negative nitrogen balance (excretion exceeds intake) is observed in dietary protein deficiency. In an adult, even if protein starved, at least 30–40 g of amino acids are degraded each day; this amount defines the dietary requirement. If the dietary supply drops below this limit, a **negative nitrogen balance** occurs and the body protein is lost. Essential amino acid deficiency has the same effect because relative deficiency of even a single essential amino acid results in corresponding decrease in protein synthesis.

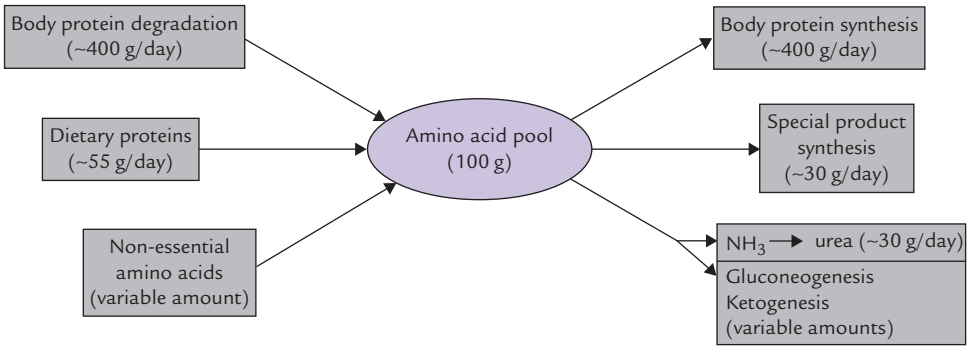


Fig. 13.1. Sources and fate of amino acids.

Nitrogen equilibrium:
Ingested nitrogen = Nitrogen excreted.
Positive nitrogen balance:
Ingested nitrogen > Nitrogen excreted.
Negative nitrogen balance:
Ingested nitrogen < Nitrogen excreted.

Negative nitrogen balance is also seen in a variety of non-physiological conditions including severe infections, metastatic carcinoma, burns, trauma and post-surgical stress. The negative nitrogen balance in these patients is related in part to increased secretion of ***stress hormones*** *cortisol* and *epinephrine*, which favour protein degradation over protein synthesis, and also stimulate the use of the liberated amino acids for gluconeogenesis by the liver. *Cytokines*, the biologically active products released by the leucocytes in a variety of disease states, also elicit metabolic effects similar to those of the stress hormones.

Nitrogen balance, determined by comparing the nitrogen entering the body with that leaving it, is an important index of amino acid and protein metabolism.

Nitrogen balance from nutritional viewpoint is discussed in Chapter 28.

II. Catabolism of Amino Group Nitrogen

The amino acids in excess of those needed by the body cannot be stored. They are catabolized to yield energy; about 5–10% of the total energy requirement of the body comes from catabolism of amino acids. Thus, amino acids are secondarily used as fuel molecules, while their primary role is in synthesis of body proteins and special products such as amines, porphyrins, nitrogenous bases of phospholipids, polyamines, etc. (Fig. 13.1). This is in contrast with carbohydrates and lipids whose primary role is to provide energy.

During catabolism, the fate of α-amino group is different from that of the rest of the carbon skeleton (Fig. 13.2). The α-amino group is released as ammonia and carried to liver where it is incorporated in urea by way of a cyclic pathway, called urea cycle. Urea is a waste product which is eliminated in urine. In this section, formation of urea by α-amino group catabolism has been described. Metabolism of the remaining carbon skeleton will be discussed later.

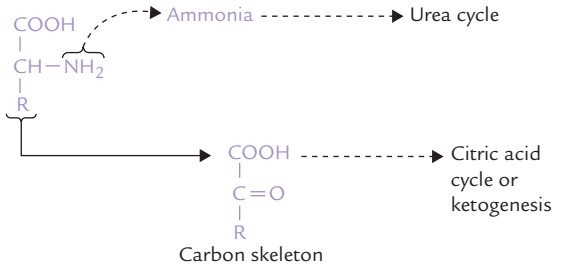


Fig. 13.2. Fate of the α-amino group and the carbon skeleton of an amino acid.

Amino acids are degraded by the removal of the α-amino group and its conversion to urea. The remaining carbon skeleton is converted to one or more metabolic intermediates.

The catabolism of α-amino group nitrogen occurs in the four stages: **transamination, oxidative deamination, ammonia transport,** and **ureagenesis** in a cyclic pathway of urea cycle.

A. Transamination

Catabolism of all α-amino acids begins with removal of the α-amino group by a process called transamination; the only exceptions are lysine, threonine, glycine, proline and histidine. The amino group thus removed is transferred usually to α-ketoglutarate forming glutamate. Other reaction product is a keto acid corresponding to the amino acid substrate.

$$AA_1 + \alpha\text{-Ketoglutarate} \rightleftharpoons KA_1 + \text{Glutamate}$$

[KA_1 = Keto acid corresponding to the amino acid substrate (AA_1).]

The enzymes catalyzing the transamination reactions are *aminotransferases* (*transaminases*). Liver, kidney, muscle and brain contain appreciable amounts of these enzymes. At least a dozen different *transaminases* have been identified in these tissues, each one being specific for different amino acid substrates. They are named after the amino acid that serves as the amino group donor. For example, the enzymes catalyzing the transfer of the amino group from alanine and aspartate are termed *alanine aminotransferase* (ALT) and *aspartate aminotransferase* (AST) respectively

$$\text{Alanine} + \alpha\text{-Ketoglutarate} \xrightleftharpoons{ALT} \text{Pyruvate} + \text{Glutamate}$$

$$\text{Aspartate} + \alpha\text{-Ketoglutarate} \xrightleftharpoons{AST} \text{Oxaloacetate} + \text{Glutamate}$$

AST and ALT have considerable diagnostic significance in cardiac and hepatic disorders (Chapter 6).

The coenzyme, **pyridoxal-5-phosphate (PLP)**, is required for the transamination reactions. PLP is derived from pyridoxine, which is vitamin B_6. PLP is bound to the active site of the enzyme both by electrostatic interactions and by a Schiff base bond with a lysine side chain of the apoprotein. The α-amine group of the original amino acid is firstly transferred to PLP to form pyridoxamine phosphate, and then transferred from pyridoxamine phosphate to α-ketoglutarate (Fig. 13.3). Thus, the reaction follows a **ping-pong mechanism:** the prosthetic group is modified chemically during the reaction, and the first product is released before the second substrate binds (Chapter 6).

Transamination involves removal of α-amino groups from amino acid to the α-keto acid acceptor, usually α-ketoglutarate, which results in formation of glutamate and the corresponding α-keto acid. The coenzymes for *transaminases* is pyridoxal phosphate (derived from vitamin B_6).

α-Ketoglutarate is the acceptor of the transferred amino groups of various amino acids. In the process, it is converted to *glutamate* which acts as a *collecting point* for the α-amino groups (Fig. 13.4).

The biological advantage of such an arrangement is that the subsequent steps are common, since it is only the glutamate nitrogen which has to be handled. Glutamate is deaminated and the ammonia released is funneled into ureagenesis, as discussed in the following sections.

B. Oxidative Deamination

The α-amino group of glutamate is released as ammonia by the following reaction:

$$\text{Glutamate} + NAD^+ + H_2O \xrightarrow{\text{GDH}} \alpha\text{-Ketoglutarate} + \text{Ammonia} + NADH + H^+$$

The enzyme catalyzing this reaction is *glutamate dehydrogenase* **(GDH)**. Thus, successive actions by *transaminases* and *glutamate dehydrogenase* effectively cause release of the α-amino group as ammonia (Fig. 13.5).

Glutamate dehydrogenase is a complex allosteric enzyme consisting of six subunits. It is unusual in being able to utilize both NAD^+ and $NADP^+$ as co-substrates. Its activity is allosterically regulated by several modulators. Guanosine triphosphate (GTP) and adenosine triphosphate (ATP) are *allosteric inhibitors*, whereas guanosine diphosphate (GDP) and adenosine diphosphate (ADP) are *allosteric activators.* Thus, lowering of the cellular energy (i.e. increased ADP or GDP) activates GDH and therefore, increases catabolism of amino acids. Conversely, increased cellular energy (increased ATP and GTP) inhibits this enzyme.

GDH can catalyze the reverse reaction also, i.e. amination of α-ketoglutarate to form glutamate. This reaction requires NADPH as coenzyme (cosubstrate).

$$\text{Glutamate} + NADP^+ \rightleftharpoons \alpha\text{-Ketoglutarate} + \text{Ammonia} + NADPH + H^+$$

The reverse reaction is activated by ATP and GTP and inactivated by ADP and GDP. Thus, when the cellular energy charge is high amination of α-ketoglutarate to glutamate is favoured and deamination of glutamate to α-ketoglutarate is impeded.

ATP and GTP are allosteric inhibitors of *glutamate dehydrogenase, and* GDP and ADP are allosteric activators. Therefore, the enzyme is activated with lowering of cellular energy. The reverse reaction—amination of α-ketoglutarate to glutamate, is activated with increased cellular energy.

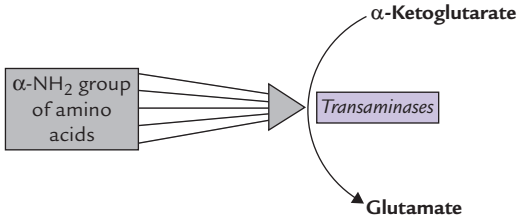


Fig. 13.4. α-Ketoglutarate serving as common acceptor of α-amino groups of various amino acids to form glutamate.

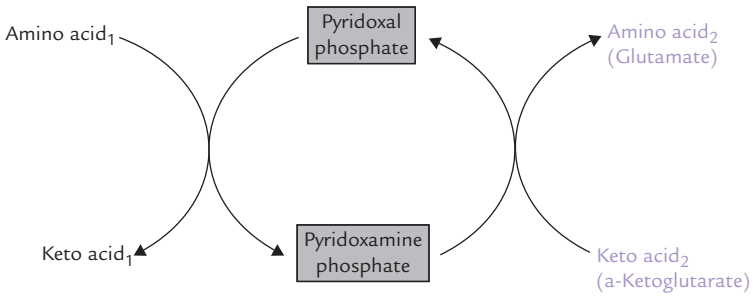


Fig. 13.3. Pyridoxal-5-phosphate (PLP), the coenzyme for all trasamination reactions, is transiently converted to pyridoxamine phosphate during transamination. Interconversion of pyridoxal phosphate and pyridoxamine phosphate is shown here.

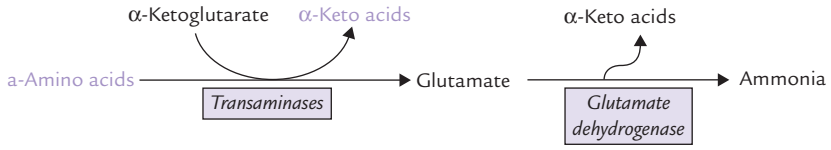
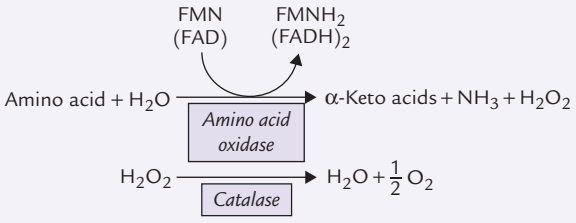


Fig. 13.5. The α-amino groups of various amino acids incorporated into glutamate during transamination reactions are liberated as ammonia by *glutamate dehydrogenase*.

BOX 13.1

Alternate Pathways of Deamination

Two other enzymes, besides *glutamate dehydrogenase*, provide alternative route for releasing nitrogen of α-amino groups as ammonia. They are *L-amino acid oxidase* (cofactor: FMN) and *D-amino acid oxidase* (cofactor: FAD). Both enzymes are present in liver and kidneys, localized in peroxisomes where they generate hydrogen peroxide. The latter is decomposed by *catalase*.



The *L-amino acid oxidase* can act on all amino acids except hydroxy amino acids and dicarboxylic amino acids. Activity of the D form is higher than that of the L form. It is not important for the degradation of common L-amino acids. Its main function is to degrade D-amino acids present in bacterial cell walls. Thus, *D-amino acid oxidase* functions during the degradation of D-amino acids of bacterial origin that are absorbed from the gut.

The utilization of two different coenzymes by *glutamate dehydrogenase*—NAD^+ for release of ammonia and $NADP^+$ for incorporation of ammonia—permits independent regulation of these two reactions. It is noteworthy that independent regulation is possible although both the reactions are catalyzed by the same enzyme.

In addition to *glutamate dehydrogenase, amino acid oxidase* provides an alternative pathway for deamination (**Box 13.1**).

To sum up, successive transamination and oxidative deamination reactions effectively release the α-amino group as ammonia. Because glutamate is the only amino acid that can undergo rapid oxidative deamination, most other amino acids form glutamate by transferring their amino group to α-ketoglutarate.

The ammonia so released is neurotoxic, and so safely transported, as discussed below.

C. Transport of Ammonia

Release of the α-amino group as ammonia, discussed above, is potentially a hazardous event because ammonia is neurotoxic even in low concentrations (above 4×10^{-5} M) in blood. Therefore, it is vital to deal efficiently with the ammonia generated. Aquatic animals can release ammonia as such through their gills (***ammonotelic)***, while birds, reptiles and insects convert it to uric acid **(*uricotelic*)**. In terrestrial vertebrates, including humans, it is transported to liver chiefly as glutamine and alanine, where it is detoxified to urea (humans are ureotelic). Urea is then transported to the kidneys and excreted in the urine. Kidneys also form some urea.

Ammonia is produced in most tissues and must be transported to liver without causing ammonia toxicity. **Glutamate, glutamine and alanine are the transport forms of ammonia** from peripheral tissues to liver.

1. ***Glutamate*** (α-ketoglutarate plus ammonia) may be considered as the major participant in such interorgan transport of ammonia. Concentration of glutamate in blood is about tenfold higher than other amino acids.
2. ***Glutamine*** is the transport form of ammonia from brain. Since brain is extremely sensitive to ammonia, it possesses a special mechanism for its immediate detoxification by combining it with glutamate to

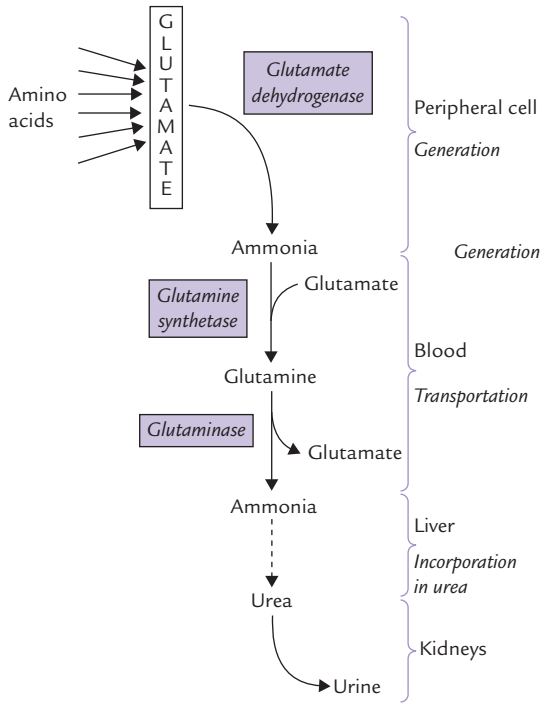


Fig. 13.6. Generation, transport and hepatic incorporation of ammonia into urea.

form glutamine (Fig. 13.6). Synthesis of glutamine is catalyzed by the enzyme *glutamine synthetase*

$$\text{Glutamate} + NH_3 + \text{ATP} \rightarrow \text{Glutamine} + \text{ADP} + \text{Pi}$$

Glutamine is released into the blood circulation and carried to liver, where it enters hepatocytes (because of its neutral nature, glutamine can readily move across the cell membrane by facilitated diffusion). *Glutaminase* converts it into glutamate and ammonia, and the latter is channeled into the urea cycle. Thus, *glutamine serves as the major transport form of ammonia from* **brain**, based on successive actions of *glutamine synthetase* and *glutaminases*.

Asparagine may also similarly transport ammonia (by successive action of *asparagine synthetase* and *asparaginase*). However, *asparaginase* is quantitatively of much less significance.

3. ***Alanine*** serves as the transport form of ammonia from *muscles*. The amine group of various amino acids is transferred to pyruvate in muscles to form alanine. The latter is transported to liver where the amine group is removed to reform pyruvate, which is used for glucose synthesis. This is glucose alanine cycle (or Cahill cycle), a major participant in interorgan transport of ammonia (see **Box 13.2**).

Ammonia Toxicity

Ammonia is a highly toxic substance, particularly to the brain. The glial cells surrounding the neurons are highly susceptible to ammonia. So high is its toxicity that injection of extremely dilute solutions of ammonia into the blood stream is capable of rendering the experimental animal comatose. Ammonia can readily diffuse through cell membranes and enter tissues so that little is left in blood circulation (its concentration in peripheral blood is 30–60 μg/dL).

Most of the ammonia is converted to ammonium ion because of its high pK′ value (less than 1% is left as ammonia) and even such small amount is capable of causing serious toxic effects. **Biochemical basis of ammonia toxicity is not clearly understood**. The following mechanisms have been proposed:

1. *Decreased glucose utilization and ATP generation:* In the brain cell mitochondria, excess ammonia may cause the reductive amination of α-ketoglutarate by *glutamate dehydrogenase*, forming glutamate. This reaction depletes α-ketoglutarate, a key intermediate of the TCA cycle, and leads to its impairment. As a result, there is severe inhibition of glucose utilization and fall of ATP generation.

   This theory is, however, not universally accepted.
2. *Glutamate depletion:* Ammonia exerts an inhibitory effect on activity of *glutaminase*, resulting in depletion of glutamate, an excitatory neurotransmitter, in the neuronal cells. Glutamine, synthesized and stored in glial cells, is the most likely precursor of glutamate (it is transported into the neurons and hydrolyzed by *glutaminase*). Ammonia inhibits *glutaminase* and depletes the glutamate concentration.

   Moreover, intracellular accumulation of glutamine causes osmotic shifts of water into the cell, resulting in oedema and swelling of astrocytes. This may aggravate the encephalopathy of hyperammonaemia.
3. *Neuronal dysfunction*: Hyperammonia increases permeability of the neuronal membrane to K^+ and Cl^- ions to cause neuronal dysfunction.
4. *Accumulation of excito-toxins*: Increased transport of tryptophan across the blood brain barrier and accumulation of its metabolites, which are excito-toxins, are also implicated in ammonia toxicity. Two of such tryptophan-derived metabolites are serotonin and quinolinic acid.

Other Sources of Ammonia

Although major source of ammonia is amino acids, additional sources are also known. These are:

1. Bacterial degradation of urea in the intestinal lumen.
2. Action of renal *glutaminase* on glutamine in renal tubular cells.
3. Action of intestinal *glutaminase* on glutamine in intestinal mucosal cells.
4. Release of amino groups of purines and pyrimidines as ammonia during catabolism of these nitrogen bases.

BOX 13.2

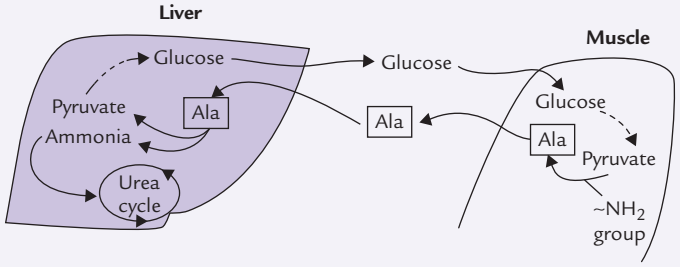
Glucose–Alanine Cycle

In muscles, pyruvate accepts the $\sim NH_2$ group from various amino acids to form alanine. The latter is also released into blood circulation, from where it is taken up by the liver. The following transformations occur in the liver cell.

1. The α-amino group of alanine is removed by transamination and channeled into urea cycle.
2. The remaining 3-carbon skeleton (i.e. pyruvate) is converted to glucose by gluconeogenesis.

The glucose formed by gluconeogenesis is circulated back to muscles where it again forms pyruvate by way of glycolysis. Pyruvate then accepts the $\sim NH_2$ group to form alanine which again moves to liver, thus repeating the cycle. In this way, the glucose–alanine cycle plays dual role:

- Transports ammonia to liver in a non-toxic form.
- Transports carbon skeleton to the liver for gluconeogenesis.



The amino group is removed from amino acids by transferring it to α-ketoglutarate to form glutamate (transamination). Glutamate undergoes deamination to liberate ammonia, a toxic compound, which is transported to liver for synthesis of urea, a non-toxic excretory product.

D. Urea Cycle

The only organ where urea synthesis occurs is **liver**. Urea is the major excretory product in humans, accounting for an average of **86%** of nitrogen eliminated. The rest of the nitrogen is eliminated as follows: 4.5% by creatinine, 2.8% as ammonium ions, 1.7% as uric acid, and 5.0% as other compounds. About *30 g urea is excreted per day*; the amount excreted is dependent on protein intake. Higher the protein intake, more is the urea synthesis and excretion.

Reactions of Urea Cycle

The sequence of reactions leading to urea synthesis was first proposed by *Krebs* and *Henseleit* in 1932, five years before the elucidation of TCA cycle. Urea cycle was the first cyclic pathway to be identified. All the reactions of this pathway are shown in Figure 13.7. *The first two reactions take place in the mitochondria, and the rest occur in cytosol.*

Reaction 1: Formation of Carbamoyl Phosphate

Carbamoyl phosphate is formed from ammonia and carbon dioxide in an energy-requiring reaction. Two ATP molecules are required to drive this reaction forward. The enzyme catalyzing this step, *carbamoyl phosphate synthetase-I* (CPS-I) is ***rate-limiting*** for the pathway. It is present in very high concentration in liver mitochondria and its K_m for ammonia (250 μM) is not much higher than the physiological ammonia concentration. These properties enable the enzyme to effectively remove ammonia quantitatively from its environment.

N-Acetyl glutamate (NAG) is an obligatory positive effector of CPS-I, as discussed later.

Formation of carbamoyl phosphate is the rate-limiting step of urea synthesis. Increased dietary consumption of proteins raises the level of N-acetylglutamate in liver, which stimulates carbamoyl phosphate synthesis and hence urea production.

In humans there are two immunologically distinct *carbamoyl phosphate synthetases*, one *mitochondrial* **(CPS-I)** and the other *cytosolic* **(CPS-II)**.

- **CPS-I** is involved in ureagenesis, uses NH_3 exclusively as the nitrogen donor and requires binding of NAG for its activity.

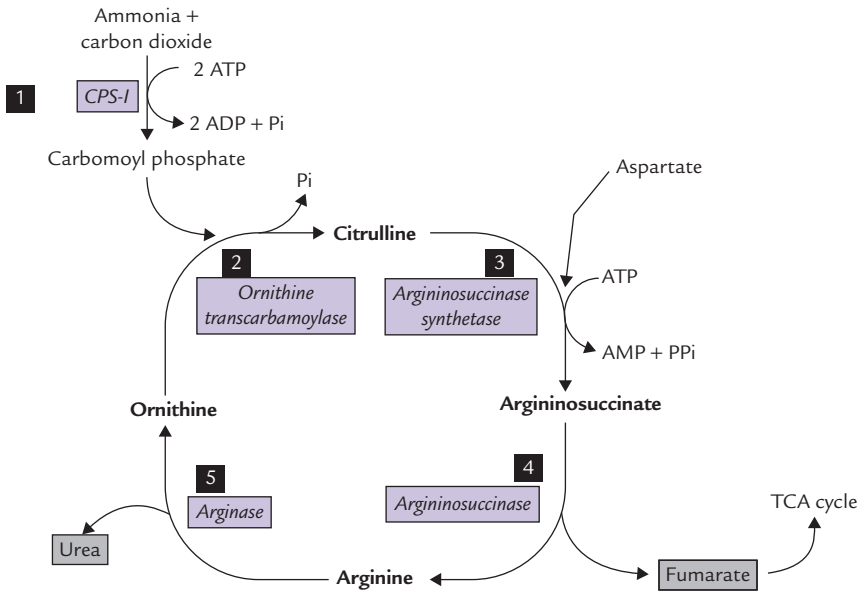


Fig. 13.7. Reactions of the urea cycle.

- **CPS-II** is used in pyrimidine synthesis, does not depend on NAG and uses glutamine as a substrate (Chapter 20).

These characteristics make possible compartmentalization of the two pathways, one of which is degradative to remove nitrogen from the body, and the other is synthetic to build the bases for DNA or RNA.

Reaction 2: Formation of Citrulline

Carbamoyl phosphate, a high energy mixed anhydride, condenses with ornithine to form citrulline. The reaction is catalyzed by a mitochondrial enzyme *ornithine transcarbamoylase*.

Citrulline diffuses out of the inner mitochondrial membrane so that *the subsequent reactions of the urea cycle take place in the cytosol.*

Reaction 3: Condensation of Citrulline with Aspartate

This is a complex condensation reaction between citrulline and aspartate to form argininosuccinate. The reaction is catalyzed by the enzyme *argininosuccinate synthetase*. Free energy is required in this reaction which is provided by pyrophosphate cleavage of ATP. The *pyrophosphate cleavage ensures irreversibility of the reaction*.

Reaction 4: Cleavage of Argininosuccinate

The carbon skeleton of aspartate is released as fumarate while its nitrogen remains in the cycle, forming one of the nitrogen side chains of arginine. Reaction is catalyzed by the enzyme *argininosuccinate lyase* (also known as *argininosuccinase*).

Reaction 5: Formation of Urea

In the last step, arginine is hydrolyzed by the enzyme *arginase* to form urea. The other product of this reaction, ornithine, enters the mitochondrial matrix to participate in the urea cycle again.

Urea cycle is linked to TCA cycle: Synthesis of fumarate (reaction 4) is important because it links urea cycle to the citric acid cycle. Fumarate is converted to malate which is in turn oxidized to oxaloacetate. Oxaloacetate can condense with acetyl CoA to form citrate, the first intermediate of TCA cycle.

Energetics of Ureagenesis

The stoichiometry of urea synthesis is as below:

$$CO_2 + \text{Ammonia} + \text{Aspartate} + 3\ \text{ATP} \rightarrow \text{Urea} + \text{Fumarate} + 2\ \text{ADP} + \text{AMP} + P_i + P_i$$

As shown, formation of one molecule of urea is powered by **three ATPs** and requires one molecule each of ammonia, carbon dioxide and aspartate. In all, hydrolysis of *four high energy phosphate groups is required in each cycle*: two are needed to drive the formation of carbamoyl phosphate and two for the formation of argininosuccinate.

However, the net energy expenditure may fall to only **one ATP** if fumarate (formed in the fourth step) is converted to malate. When this malate is oxidized to oxaloacetate, one NADH molecule is generated that can give rise to three ATP molecules through the electron transport chain. Thus, the energy expenditure becomes one (4–3) ATP molecule for each molecule of urea formed.

Control of Urea Cycle

Coarse Regulation

It occurs by **induction-repression** mechanism. The urea cycle enzymes are induced or repressed depending on the metabolic needs of the body. In **starvation**, urea cycle enzymes are induced. Their activities are elevated by 10–20-fold. It permits increased formation of urea in response to increased catabolism of the proteins that occurs in starvation. Moreover, cellular energy falls low in starvation, which activates the *glutamate dehydrogenase*. This results in increased production of ammonia which is channeled into urea cycle.

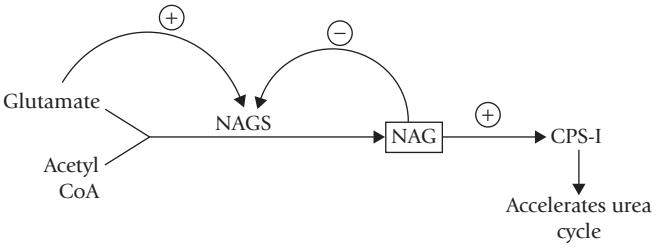
A **protein rich diet** also accelerates urea cycle through the activation of rate limiting enzyme, *carbamoyl-phosphate synthetase*.

Fine Regulation

It occurs by **allosteric-modulation**. The major regulatory enzyme of the urea cycle is *carbamoyl phosphate synthetase-I*, which is subject to allosteric activation by N-acetyl glutamate (NAG). Transfer of the acetyl group from acetyl CoA to glutamate by the enzyme *NAG synthase* (NAGS) forms this compound.

Glutamate + Acetyl CoA → N-Acetyl glutamate

NAG synthase is under positive allosteric modulation by arginine and product inhibition by NAG. A high glutamate level also leads to increased NAG synthesis; this situation can occur when more amino acids are degraded. High glutamate level leads to increased NAG, hence enhanced activity of CPS-I, and thereby increased rate of ureagenesis.



Defects of Urea Cycle

Synthesis of urea provides the major route for the removal of toxic ammonia from body. Blockage of any of the steps of urea synthesis, therefore, results in accumulation of ammonia in the blood; the condition is known as **hyperammonaemias**. The blockage may result due to genetic deficiency of an enzyme of the urea cycle (i.e. *familial hyperammonaemia*) or due to some acquired defect (i.e. *acquired hyperammonaemia)*.

Familial Hyperammonaemia

Genetic deficiency of each of the five enzymes of urea cycle have been described with an overall prevalence of about 1 in 30,000 live births. In addition, deficiency of *NAG synthase* has also been described. Inheritance pattern of the latter is not known, but the four of the urea cycle defects are autosomal recessive and *ornithine carbamoyltransferase* deficiency is X-linked. They are enlisted below, with name of the enzyme in parenthesis:

- Hyperammonaemia type I (*carbamoyl phosphate synthetase*).
- Hyperammonaemia type II (*ornithine transcarbamoylase*) – X-linked. (All others are autosomal recessive).
- Citrullinaemia (*argininosuccinate synthetase*).
- Argininosuccinic aciduria (*argininosuccinase*).
- Hyperargininaemia (*arginase*).

Feeding difficulties, lethargy or irritability, vomiting and poor intellectual development, and tendency for coma and death are common to all these disorders, though they vary in severity.

Earlier blocks: When the block is in one of the earlier enzymes, ammonia itself accumulates and the condition is more severe.

Later blocks: When the block is in later enzymes, accumulation of other intermediates occurs, which are less toxic, and therefore the condition is relatively mild.

Dietary proteins aggravate the symptoms and many patients develop spontaneous distaste for protein-rich foods.

Acquired Hyperammonaemia

Impaired detoxification of ammonia in advanced **liver cirrhosis** is a far more important cause of hyperammonaemia. This condition is end point of several disease processes, such as alcoholism, hepatitis or biliary cirrhosis, and is characterized by progressive loss of hepatocytes. The lost hepatocytes are replaced by fibrous connective tissue, which impairs blood flow through the liver, and so the portal blood is shunted directly into the systemic circulation without having access to liver. The detoxification of ammonia in liver is, therefore, impaired and its concentration in systemic blood rises.

Portal hypertension also develops, which is dangerous because it can result in fatal haemorrhage from the dilated lower oesophageal veins. These derangements lead to CNS involvement, i.e. **hepatic encephalopathy**, also known as portal-systemic encephalopathy (because the shunting of blood around the cirrhotic liver is important in its pathogenesis). It has poor prognosis, frequently progressing to coma and death. Although various other biochemical and pathophysiological abnormalities have also been implicated, hyperammonaemia is a major cause of the CNS disorder.

Diagnosis of hyperammonaemias calls for detection of increased ammonia levels in blood and/or urine. Besides

ammonia, glutamine level is also elevated because excess ammonia is diverted into glutamine synthesis. The immediate substrate of the deficient enzyme also is elevated in blood and urine in familial hyperammonaemia. Paper chromatography, paper electrophoresis, and ion-exchange chromatography are available for the quantitative determination of these compound amino acids in serum and urine. Final confirmation requires an assay-proven enzyme defect.

Treatment of hyperammonaemia involves a twofold strategy: *dietary manipulations* and *activation of alternative routes of nitrogen excretion*:

1. *Dietary protein restriction* is considered a mainstay of the long-term management of hyperammonaemia cases. The replacement of essential amino acids by their corresponding α-keto acids minimizes the requirement for nitrogen disposal, without precipitating the essential amino acid deficiencies. This is because transamination reactions generate most of the essential amino acids from their corresponding α-keto acids in diet.
2. The second treatment strategy promotes *nitrogen excretion in forms other than urea*. This is accomplished by activation of some latent biochemical pathway to bypass a genetic defect. The following examples illustrate this strategy:
   (a) The metabolic block resulting due to *argininosuccinase* deficiency can be partially bypassed by supplementing the diet with arginine (Fig. 13.8). The surplus arginine is metabolized into argininosuccinate (via steps 5, 2 and 3), in the process utilizing ammonia. Argininosuccinate is subsequently disposed. In this way argininosuccinate substitutes urea for eliminating nitrogen from the body.
   (b) A different approach is employed for the treatment of *carbamoyl phosphate synthetase* deficiency or of *ornithine transcarbamoylase* deficiency. In these conditions, excessive ammonia accumulates in glutamine (and glycine). The aim of treatment is to get rid of these two amino acids. This is accomplished by *supplementing the diet with large amount of benzoate and phenylacetate*. Benzoate is activated to benzoyl CoA which reacts with glycine to form hippurate. Likewise, phenylacetate is activated to phenylacetyl CoA which reacts with glutamine to form phenylacetylglutamine. These two conjugates act as substitutes for urea in the disposal of ammonia.

   Case 13.1 describes clinical presentation and treatment strategy of a urea cycle disorder.
3. Other treatment strategies for management of hyperammonaemia may require aggressive approach. For instance, any neonatal hyperammonaemia, irrespective of the cause, is medical emergency and requires a *rapid lowering of ammonia levels* to prevent serious effects on the brain. These include **haemodialysis, exchange transfusion, peritoneal dialysis,** etc. Gastrointestinal haemorrhage from dilated oesophageal veins in portal-systemic encephalopathy needs to be controlled. Sterilization of gut in hepatic encephalopathy is helpful, although there is a risk of over growth by drug-resistant bacteria, resulting in enterocolitis.

III. Catabolism of Carbon Skeleton of Amino Acids

Excess amino acids are metabolized by oxidative pathways to generate energy. As discussed earlier, the metabolism occurs in two stages:

- **Metabolism of the amino group**
- **Metabolism of the carbon skeleton** (Fig. 13.2).

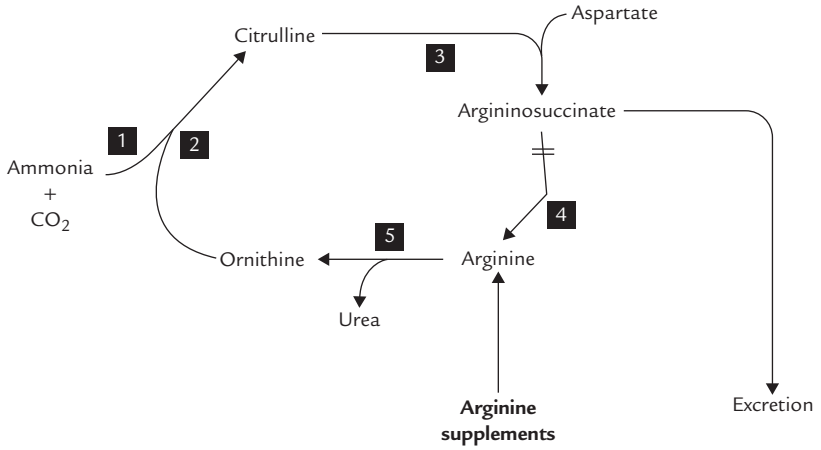


Fig. 13.8. Bypassing the metabolic block in *argininosuccinase* deficiency. Numbers 1 to 5 are steps of reactions discussed in text.

One or more of the following compounds are generated as catabolic (end) products from the carbon skeleton: acetyl CoA, acetoacetyl CoA, pyruvate, or some intermediates of TCA cycle (e.g. oxaloacetate, α-ketoglutarate, succinyl CoA or fumarate). Thus, carbon skeletons of different amino acids are converted into a limited number of compounds, which are further metabolized in TCA cycle. This provides an example of the *economy of metabolic conversions* (Fig. 13.9).

Further fate of these seven products depends on the cellular energy requirements. When the cell needs energy, these products undergo further degradation via TCA cycle to provide energy. On the other hand, when the cell is already supplied with adequate energy, they enter pathways of intermediary metabolism to yield glucose or ketone bodies. Evidently, *strategy of amino acid catabolism is to form major metabolic intermediates that can be oxidized via TCA cycle, or converted to either glucose or ketone bodies*.

Depending on the nature of the end product formed, the amino acids are classified in three categories (Table 13.1). They are designated as **glycogenic** if they can be converted to glucose, **ketogenic** if they can be converted to ketone bodies, and **both glycogenic** and **ketogenic** if they are convertible to both types of compounds. This classification is based on experiments performed by administration of each amino acid to experimental animals and determining whether there was an increase in glucose in the urine (glycogenic amino acid), an increase in ketone bodies in the urine (ketogenic amino acid), or both. Glycogenic amino acids are catabolized to pyruvate

Table 13.1. Classification of amino acids on the basis of catabolic end products

Convertible to glycogen (Glycogenic)	Convertible to fat (Ketogenic)	Convertible to glycogen and fat (Glycogenic and ketogenic)
Alanine	Leucine	Isoleucine
Glycine	Lysine	Phenylalanine
Histidine		Tyrosine
Glutamine		Tryptophan
Asparagine		
Cysteine		
Glutamate		
Aspartate		
Serine		
Threonine		
Valine		
Methionine		
Arginine		
Proline		

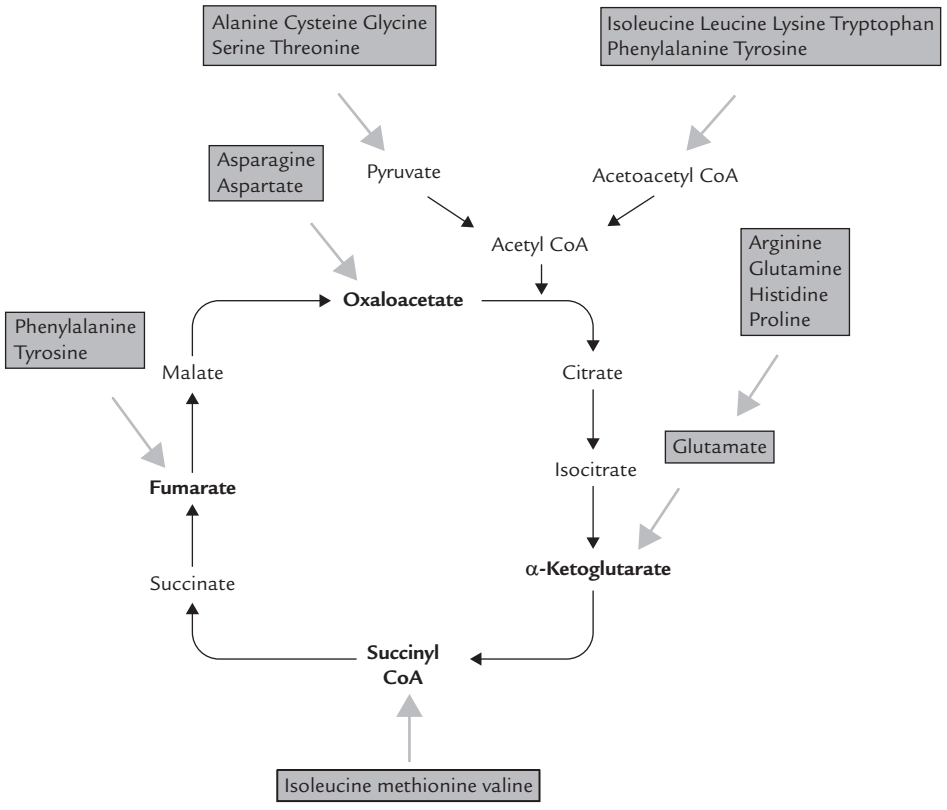


Fig. 13.9. Catabolic end products of primary amino acids.

or a TCA intermediate, which are potential precursors for glucose and glycogen. Ketogenic amino acids yield acetyl CoA or acetoacetyl CoA, or both. Some of the carbon atoms of the amino acids, which are both glycogenic and ketogenic, emerge in acetyl CoA and acetoacetate, whereas others appear in potential precursors of glucose. *Leucine* and *lysine* are the only amino acids that are regarded as being **exclusively ketogenic**. The aromatic amino acids and isoleucine (four amino acids) are both glycogenic and ketogenic, and the remainder (14 amino acids) can be considered as purely glycogenic. This classification is, however, not universally accepted.

Amino acids are designated as **glycogenic** if catabolized to pyruvate or a TCA cycle intermediate, **ketogenic** if convertible to acetyl CoA or acetoacetyl CoA, and **both glycogenic and ketogenic** if they can be converted to either type of compounds.

Catabolism of the primary amino acids, as discussed below, has been divided on the basis of nature of the catabolic end product. Generally, the oxidative metabolism of non-essential amino acids is simple, and that of the essential amino acids is relatively complicated.

A. Amino Acids that Form Pyruvate (Fig. 13.10)

The 3-carbon amino acids, alanine, serine and cysteine are directly convertible to the 3-carbon keto acid, i.e. pyruvate. The carbon skeletons of two other amino acids, glycine (2-C) and threonine (4-C), also enter the metabolic mainstream through pyruvate.

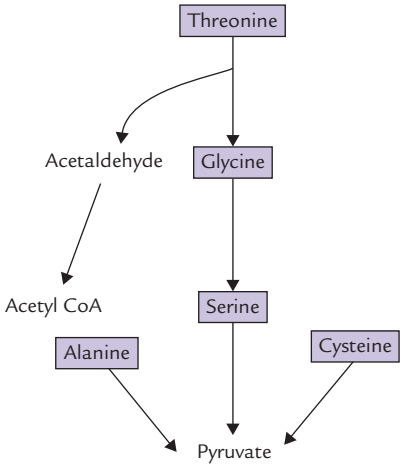


Fig. 13.10. Outline of the pathways of conversion of serine, cysteine, alanine, glycine and threonine to pyruvate.

1. **Alanine** can be directly converted to pyruvate by transamination catalyzed by the enzyme ALT.

   Alanine + α-Ketoglutarate ⇌ *Pyruvate* + Glutamate

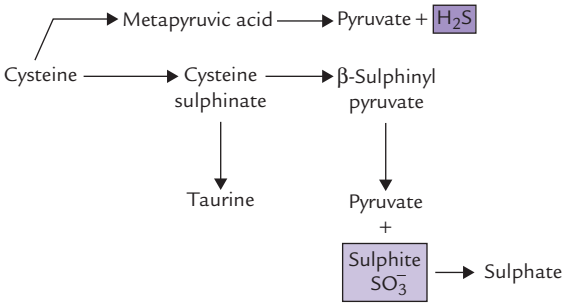
2. **Serine** undergoes dehydration followed by loss of α-amino group to yield pyruvate as depicted in Figure 13.11 (see Reactions 1 and 2). Both the reactions are catalyzed by *serine dehydratase*, a pyridoxal phosphate-dependent enzyme.

   Alternatively, serine can undergo a transamination reaction (3) to produce 3-hydroxypyruvate, which is reduced by an NADH-dependent enzyme *glycerate dehydrogenase* (4) to yield D-glycerate. The latter is phosphorylated to 3-phosphoglycerate (5), a glycolytic intermediate.

   Both these pathways produce a gluconeogenic intermediate that can also be metabolized by mainstream reactions.

3. **Cysteine** can be converted to pyruvate by loss of sulphur atom, which is released as H_2S or SO_3^-. The metabolism is considered intricate because of the multiple pathways cysteine can enter during sulphur metabolism:

   (a) *Loss of* SO_3^-: Cysteine reacts with oxygen to yield cysteine sulphinate, which undergoes transamination to yield β-sulphinylpyruvate. The latter then undergoes an exergonic hydrolysis to yield pyruvate and sulphite (SO_3^-), which is oxidized to produce sulphate by the enzyme *sulphite oxidase*. Cysteine sulphinate can be converted to taurine in a quantitatively minor pathway, which combines with cholyl-CoA to produce taurocholate, a bile salt.



   (b) *Loss of* H_2S: Cysteine may undergo isoergonic transamination to mercaptopyruvic acid, which liberates H_2S to form pyruvate.

   The conversion of *cysteine to pyruvate* by either pathway accounts for the glycogenic nature of this amino acid.

4. **Glycine** metabolism is intimately linked with tetrahydrofolate (THF).

   The following pathways are involved:

   (a) The **major pathway** is via the *glycine cleavage enzyme*, which catalyzes degradation of glycine

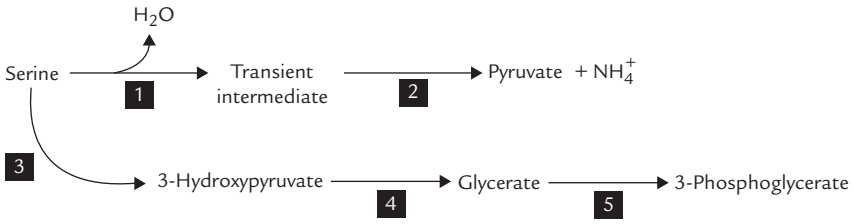


Fig. 13.11. Metabolism of serine. The numbers 1 to 5 are steps of reactions discussed in text.

to carbon dioxide and a THF-bound one carbon unit. The enzyme is also known as the *glycine synthase* for the reverse reaction

$$\text{Glycine} + \text{THF} + \text{NAD}^+ \rightleftharpoons \text{5,10-Methylene THF} + \text{CO}_2 + \text{NH}_3 + \text{NADH}$$

(b) **Hydroxymethylene transfer** reaction is another significant way to metabolize glycine. It bring about glycine (2-c) to serine (3-c) conversion by addition of one carbon (hydroxymethylene) group. The donor of one carbon group is 5,10-methylene tetrahydrofolate and the enzyme is *hydroxymethyltransferase*.

$$\text{Glycine} + \text{5,10-Methylene THF} \rightleftharpoons \text{Serine} + \text{THF}$$

Serine is further metabolized as discussed.

(c) A **minor pathway** of glycine metabolism is its conversion to glyoxylate via *glycine oxidase* or *D-amino acid oxidase*. Glyoxylate can be oxidized to oxalate and excreted in urine; the precipitation of soluble calcium oxalate is the most common cause of renal calculi.

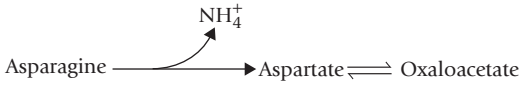
5. **Threonine**, the 4-C amino acid, is cleaved to 2-C products, glycine and acetaldehyde (Fig. 13.10). Glycine is then metabolized as above; and the other two carbons of threonine (that appear in acetaldehyde) form acetyl CoA. Threonine metabolism can also produce another TCA cycle intermediate, succinyl CoA, as discussed later.

B. Amino Acids That Form TCA Cycle Intermediates (Fig. 13.9)

Carbon skeleton of certain amino acids enter the metabolic mainstream via some intermediates of TCA cycle such as oxaloacetate, α-ketoglutarate, fumarate or succinyl CoA.

Oxaloacetate producing: The C-4 family of amino acids, i.e. **aspartate** and **asparagine** form oxaloacetate (Fig. 13.9). Asparagine degradation involves:

- An exergonic hydrolysis at the amide nitrogen by the enzyme *asparaginase* to produce aspartate and ammonia.
- Aspartate undergoes an isoergonic transamination reaction to yield oxaloacetate.



Asparaginase in treatment of cancer: Parenteral administration of the enzyme *asparaginase* has been used in various malignancies, in particular **leukaemias** in adults. The leukaemic cells lose ability to synthesize asparagine and depend for its supply from blood circulation. The injected *asparaginase* sufficiently lowers the plasma level of asparagine to decrease its uptake by these cells, thereby inhibiting their growth.

α-Ketoglutarate producing: End point of several amino acids of C-5 family, such as **glutamate**, **glutamine, arginine, proline** and **histidine** is α-ketoglutarate (Fig. 13.9). These pathways for α-ketoglutarate are discussed below:

1. **Glutamate** is converted to the corresponding 5-C keto-acid, α-ketoglutarate, by action of the enzyme *glutamate dehydrogenase*, as described earlier in detail.

$$\text{Glutamate} + \text{NAD}^+(\text{P}) \rightleftharpoons \alpha\text{-Ketoglutarate} + \text{NAD(P)H} + \text{H}^+ + \text{NH}_4^+$$

The other four amino acids of the C-5 family, i.e. glutamine, arginine, proline and histidine are first converted to glutamate which then yields α-ketoglutarate.

2. **Glutamine** loses amino group by action of the enzyme *glutaminase*. Glutamate so formed is then converted to α-ketoglutarate.
3. **Proline** undergoes a flavoprotein-dependent oxidation to form Δ^1-pyrroline 5-carboxylate. Water adds to this compound and the ring opens non-enzymatically to form glutamate γ-semialdehyde (Fig. 13.12). The semialdehyde is oxidized to the carboxylate state in an NAD^+-dependent reaction to produce glutamate. The conversion of aldehyde to a carboxylate is exergonic and irreversible because the carboxylate is stabilized by resonance.
4. **Arginine** is hydrolyzed by the enzyme *arginase* to form ornithine; urea is the other reaction product (Fig. 13.12). Ornithine is a five-carbon amino acid that undergoes transamination reaction to form

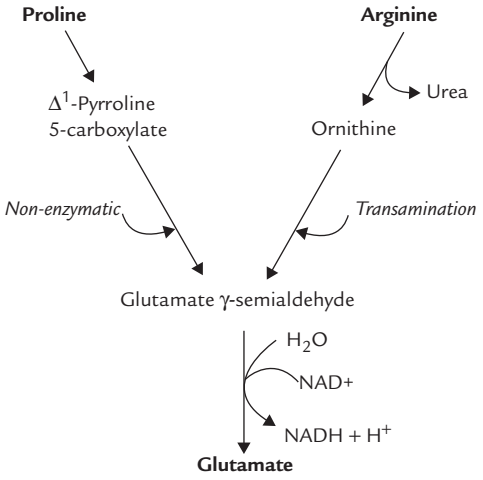


Fig. 13.12. Conversion of proline and arginine into glutamate.

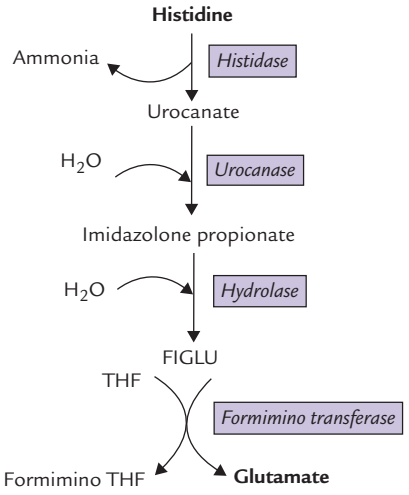


Fig. 13.13. Conversion of histidine to glutamate. THF (tetrahydrofolate) plays important role.

glutamate semialdehyde. The latter is oxidized to glutamate.

5. **Histidine**: One carbon of the histidine molecule is transferred to the one-carbon pool of tetrahydrofolate derivatives and the other five are converted to glutamate. The first step, catalyzed by the enzyme *histidase* involves a *lyase* reaction with the elimination of ammonia to form urocanate (Fig. 13.13). Urocanate undergoes a hydration and isomerization to form imidazolone propionate, which then undergoes an exergonic hydrolytic reaction to form N-formimino-glutamate **(FIGLU)**. The latter donates its formimino group to tetrahydrofolate, leaving behind glutamate. This reaction exemplifies the role of tetrahydrofolate as **one carbon unit pool (Box 13.3)**.

Folic acid is required for amino acid degradation. One-carbon fragments (formyl formimino, methylene, etc.) get attached to the reactive portions of this vitamin, from where they are transferred to several acceptor compounds.

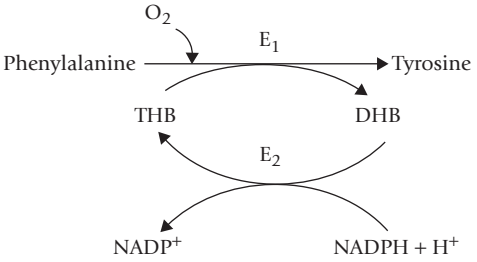
As folic acid is essential for metabolism of histidine, in cases of folic acid deficiency, metabolism of histidine stops at FIGLU. If histidine is orally administered to such patients, plasma concentration of FIGLU and its urinary excretion rises. Thus, folic acid deficiency is diagnosed by giving oral load of histidine and measuring urinary excretion of FIGLU; in case of deficiency, marked rise in the urinary excretion of this compound is observed. The test is termed the **FIGLU test**.

FIGLU test had been used earlier to assess folate deficiency: following an oral dose of histidine, individuals with folate deficiency excrete increased amount of FIGLU in urine.

Fumarate producing: *Phenylalanine* and *tyrosine* produce fumarate (and acetyl CoA; Fig. 13.9) They share a common pathway of degradation, which occurs in liver, and ultimately yields fumarate (and acetyl CoA).

The pathway starts with oxidation of phenylalanine to tyrosine. Phenylalanine is an essential amino acid, but tyrosine is not. Tyrosine in the diet, however, decreases the requirement for phenylalanine, a phenomenon called **sparing**.

Conversion of phenylalanine to tyrosine: The enzyme *phenylalanine hydroxylase* (E1) catalyzes the reaction between phenylalanine, oxygen and **tetrahydrobiopterin** (THB) to form tyrosine. Addition of a hydroxyl group to the aromatic ring of phenylalanine occurs in this reaction. The enzyme is a *monooxygenase* (also called *oxidase*) because in this reaction one atom of oxygen is incorporated in the substrate and the second atom is reduced to water. Tyrosine cannot be converted back to phenylalanine, thereby accounting for the essential nature of phenylalanine.



E_1 = *Phenylalanine hydroxylase*
E_2 = *Dihydroxybioprotein reductase*

During the reaction, THB is converted to its oxidized form, the dihydrobiopterin (DHB). Reconversion of DHB to THB requires reducing power of NADPH; and the reaction is catalyzed by the enzyme *DHB reductase* (E_2).

BOX 13.3

One Carbon Unit Pool

Several biosynthetic reactions involve transfer of 1-C unit to a metabolic precursor. The 1-C units exist in a variety of oxidation states, including methane, methanol, formaldehyde, formate, formimino and methylene. The carriers of these compounds are **folic acid** and **S-adenosyl methionine** (SAM); the 1-C pool refers to these carrier compounds.

The 1-C units can be transferred from these carrier compounds to specific acceptors that are consequently modified. SAM carries methyl groups and can act as a methylating agent. Folic acid is a more important carrier because it can carry 1-C units in various oxidation states.

As described in Chapter 18 its active form is tetrahydrofolate (THF), and 1-C units are covalently attached to THF at positions N^5, N^{10}, or both. These 1-C units, in different oxidation states, are interconvertible by enzymatic redox reactions.

THF acquires 1-C unit in conversion of serine to glycine by *serine hydroxymethyl transferase* (page 292), in cleavage of glycine (page 292) and in histidine breakdown (Fig. 13.13). The 1-C units carried by THF are used in the synthesis of methionine from homocysteine (Fig. 18.7) and in the synthesis of purine nucleotides (Fig. 18.11) as described in Chapter 18.

Biotin, a water-soluble vitamin, is a carrier of carbon dioxide, which is transferred to certain acceptor molecules. These reactions are referred to as the carboxylation reactions; for example, conversion of pyruvate (3-C) to oxaloacetate (4-C). However, biotin is not considered as a member of the 1-C pool.

Because the cellular $[NADPH^+]/[NADP^+]$ ratio is greater than $[NADH]/[NAD^+]$, $NADP^+$ is probably the physiologically important reductant.

Degradation of tyrosine (phenylalanine): Further metabolism of tyrosine and phenylalanine is considered together. Because phenylalanine is converted to tyrosine, a single pathway, depicted in Figure 13.14 is responsible for the degradation of both amino acids. The sequence of reactions from tyrosine onwards is described below:

1. Tyrosine undergoes **transamination** to form p-hydroxyphenyl pyruvate. The reaction is catalyzed by a PLP-dependent enzyme, *tyrosine transaminase*.
2. p-Hydroxyphenyl pyruvate is a substrate for an **irreversible** *dioxygenase* **reaction** in which both atoms of oxygen are incorporated into the substrate, forming homogentisate (Fig. 13.14). The enzyme catalyzing this reaction, *p-hydroxyphenyl pyruvate hydroxylase* is a *dioxygenase*. It contains iron and requires ascorbic acid for keeping the ferrous iron in the reduced state.
3. **Cleaving of the benzene ring** of homogentisate occur next to form 4-maleylacetoacetate. Molecular oxygen is required for opening the aromatic ring, and the enzyme, *homogentisate oxidase (*iron metalloenzyme*)* catalyzes this reaction.
4. Maleylacetoacetate undergoes an isoergonic **isomerization** to form fumarylacetoacetate (maleate is the *cis* isomer of the 4-C dicarboxylate, and fumarate is the *trans* isomer). The enzyme catalyzing this reaction, *maleylacetoacetate isomerase*, requires glutathione as a cofactor.
5. **Hydrolysis** of fumarylacetoacetate occurs next by the enzyme *fumarylacetoacetase*. *The reaction produces fumarate (glycogenic) and acetoacetate (ketogenic), accounting for the glycogenic as well as ketogenic nature of these amino acids*.

Inherited deficiencies of the enzymes of this pathway lead to the diseases such as *phenylketonuria, alkaptonuria* and *albinism*, discussed later in this chapter.

Succinyl CoA producing: *Methionine, valine, isoleucine* and *threonine* yield succinyl CoA as end product (Fig. 13.9).

1. **Methionine:** Catabolism of methionine is divided into two parts. The first part comprises six steps that lead to the formation of propionyl CoA (Fig. 13.15); the second part consists of three steps that brings about conversion of propionyl CoA to succinyl CoA.

Formation of propionyl CoA: Methionine is activated by the enzyme *methionine adenosyltransferase* to form S-adenosyl methionine (SAM). It is an unusual reaction in which all three phosphates of ATP are released (**Reaction 1**). *SAM is the most important methyl group donor in biological methylations*; its methyl group can be transferred to a variety of acceptor molecules (**Reaction 2**).

By donating its methyl group, the S-adenosylmethionine is converted to S-adenosylhomocysteine. Displacement of adenosine from the latter by water produces homocysteine (**Reaction 3**). Homocysteine combines with serine to form cystathionine as catalyzed by *cystathionine synthase*, a pyridoxal phosphate-dependent enzyme (**Reaction 4**).

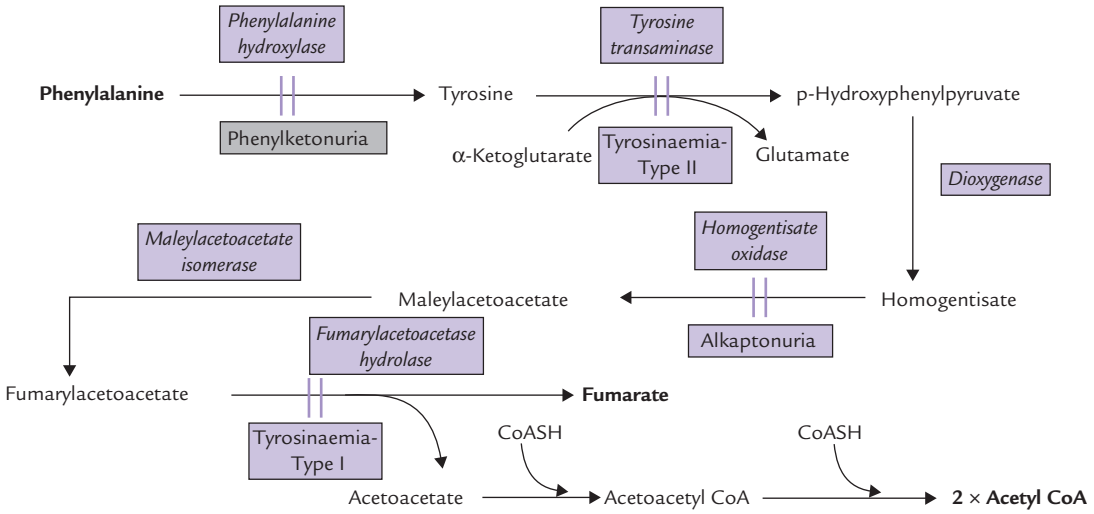


Fig. 13.14. Pathway for conversion of phenylalanine and tyrosine to fumarate and acetyl CoA.

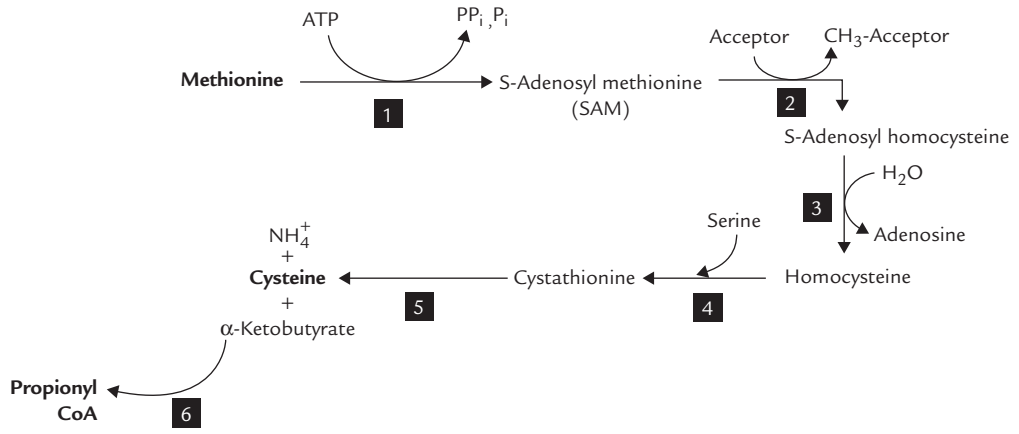


Fig. 13.15. Conversion of methionine to propionyl CoA.

Cystathionine lyase, also a pyridoxal phosphate-dependent enzyme, catalyzes hydrolysis of cystathionine to produce α-ketobutyrate, cysteine and ammonia (**Reaction 5**).

The α-ketobutyrate is then converted to propionyl CoA (**Reaction 6**) by a reaction sequence that parallels the conversion of pyruvate to acetyl CoA.

The first two reactions of this reaction–sequence illustrate role of SAM in transmethylation reaction. For more examples of transmethylation, refer to **Box 13.4**.

Conversion of propionyl CoA to succinyl CoA has been described earlier in Chapter 11. It should be noted that propionyl CoA derived from either amino acid degradation or odd chain fatty acid catabolism is converted to succinyl CoA by the same reaction sequence.

Reconversion of homocysteine to methionine: In reaction 3 homocysteine is produced, which must be reconverted to methionine. There exist two reactions for this purpose:

- The first involves transfer of a methyl group from *betaine*, a metabolite derived from choline (Chapter 12).
- The second reaction requires both folate (as methyltetrahydrofolate) and vitamin B_{12} (as methylcobalamin), as described in Chapter 18.

Methionine is an essential amino acid because humans cannot synthesize methionine from metabolites other than homocysteine.

During breakdown of methionine, serine is converted to cysteine by transfer of the sulphydryl group. Methionine is activated to SAM, which serves as a methyl group donor for synthesis of several biologically important compounds (choline, epinephrine, melatonin, creatine, etc.).

2. **Valine and isoleucine:** Valine and isoleucine are branched chain amino acids. Both yield succinyl CoA upon degradation (Fig. 13.16). Whereas valine yields only succinyl CoA, isoleucine yields acetyl CoA also (in addition to succinyl CoA).

BOX 13.4

S-Adenosylmethionine and Transmethylation Reactions

S-Adenosylmethionine is synthesized from methionine and ATP. It is capable of providing methyl groups to various acceptor molecules. These reactions, known as transmethylation reactions, are important for biosynthesis of a number of products.

Certain conversions where transmethylation is involved are:

1. Guanidinoacetate to creatine.
2. Phosphatidylethanolamine to phosphatidylcholine conversion.
3. Norepinephrine to epinephrine conversion.
4. Acetylserotonin to melatonin conversion.
5. Conversion of polynucleotides to the corresponding methylated polynucleotides.

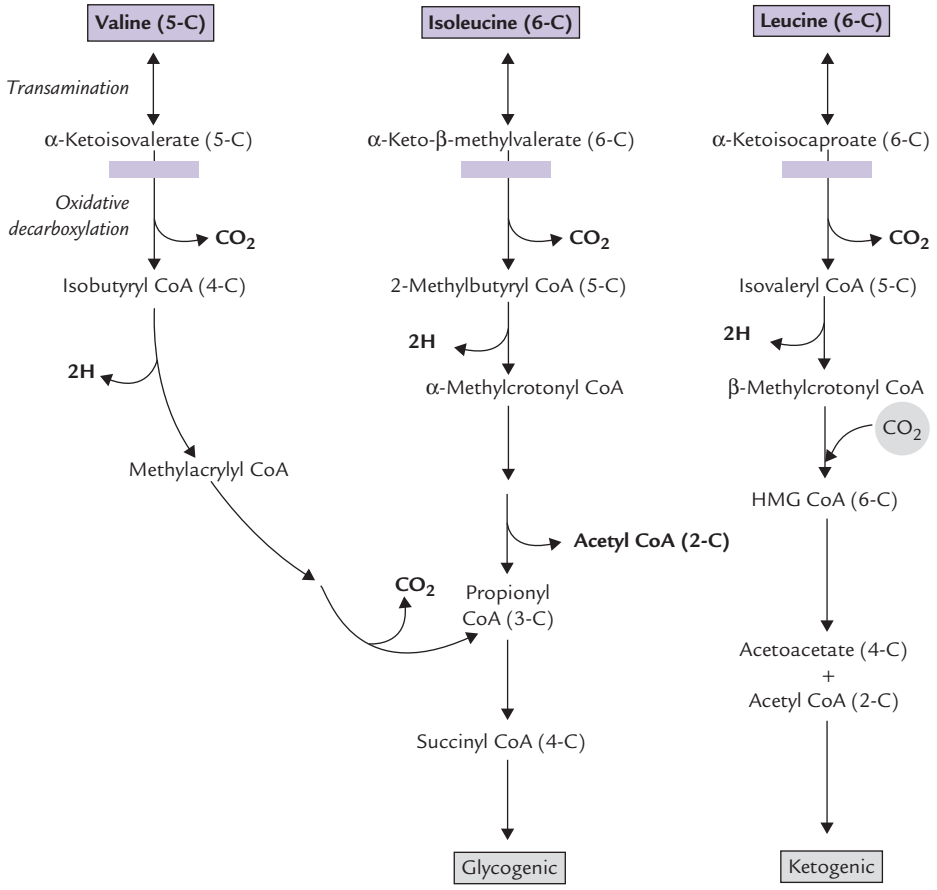


Fig. 13.16. Degradation of the branched amino acids. Blockage in the oxidative-decarboxylation step (▭) results in maple syrup urine disease.

Metabolism of these two amino acids, as also of leucine, shows a striking similarity. Therefore, the three branched chain amino acids (Valine, isoleucine and leucine) are considered together (Fig. 13.16), although leucine is purely ketogenic.

- The first reaction is an isoergonic transamination, occurring in muscle and other extrahepatic tissues. The resulting keto acids are transported to the liver, where they are metabolized further.
- The keto acids undergo oxidative decarboxylation by a single enzyme complex, the branched-chain *α-keto acid dehydrogenase*. This second reaction is analogous to *pyruvate dehydrogenase* and *α-ketobutyrate dehydrogenase* reactions. Similar to the *pyruvate dehydrogenase* reaction, five cofactors and three protein-activities participate in this reaction. The reaction products are: isobutyryl CoA from valine, 2-methylbutyryl CoA from isoleucine and isovaleryl CoA from leucine.

- The third reaction is a FAD-dependent dehydrogenation in which the bond between the α and the β carbons of the CoA-thioester is oxidized to a double bond by a flavoprotein. This reaction resembles the first step of β-oxidation (Fig. 11.3). This produces methylacrylyl CoA from valine, α-methylcrotonyl CoA from isoleucine and β-methylcrotonyl CoA from leucine.

The remaining reactions are intricate and complex, differing for the three amino acids, and it is unnecessary to commit them to memory. An outline of these is presented in Figure 13.16. Evidently:

– Valine is glycogenic (forms succinyl CoA).
– Leucine is ketogenic (forms acetyl CoA).
– Isoleucine is both glycogenic and ketogenic.

3. **Threonine:** Removal of a water molecule from threonine yields α-ketobutyrate. The latter is converted to propionyl CoA which then forms succinyl CoA. Threonine metabolism can produce pyruvate also, as discussed earlier.

C. Amino Acids that Form Acetyl CoA or Acetoacetyl CoA (Fig. 13.17)

There are six amino acids that yield acetyl CoA or acetoacetyl CoA, making them ketogenic. These are: two branched chain–amino acids (**isoleucine** and **leucine**), aromatic amino acids (**phenylalanine**, **tyrosine** and **tryptophan**), and **lysine**. Figure 13.17 presents a simplified overview of their metabolism, which is otherwise quite intricate. For example, eight steps are required for conversion of tryptophan to α-ketoadipate and five more steps are required for conversion of α-ketoadipate to acetoacetyl CoA.

1. **Leucine:** Its metabolism has been discussed earlier (Fig. 13.16).

2. **Tryptophan:** Metabolism of tryptophan is highly complex, and as mentioned, tryptophan is converted to acetyl CoA (ketogenic) and alanine (glycogenic).

The process is initiated by *tryptophan pyrrolase*, a haem protein, which oxidatively cleaves the pyrrole ring to form N-formylkynurenine (Fig. 13.18). The further pathway may branch off as follows:

(a) ***Glycogenic branch***: The portion of tryptophan molecule outside the indole ring (shown in colour) becomes alanine a glucogenic precursor.
(b) ***Ketogenic branch:*** The indole ring itself is converted to a variety of other products, of which the α-ketoadipate is quantitatively the most important (98%). After four more sequential reactions, it is finally converted to acetoacetyl CoA, and then acetyl COA.
(c) ***Kynurenine branch:*** While bulk of tryptophan in metabolized as above, a minor fraction (less than 2%) is converted into nicotinic acid, a vitamin, by the kynurenine pathway. About 60 mg tryptophan is equivalent of 1 mg of nicotinic acid by this pathway, also called the **nicotinic acid pathway** of tryptophan (chapter 18). It is diverted from the major pathway at the level shown in Figure 13.18.

Some tryptophan derived isoquinolines such as xanthurenate are not degraded further, but are excreted in urine. They are in part responsible for the yellow colour in urine.

3. **Lysine:** Catabolism of lysine is similar to that of tryptophan in that it comprises a lengthy pathway that forms *α-ketoadipate*, which finally yields acetyl CoA. This makes

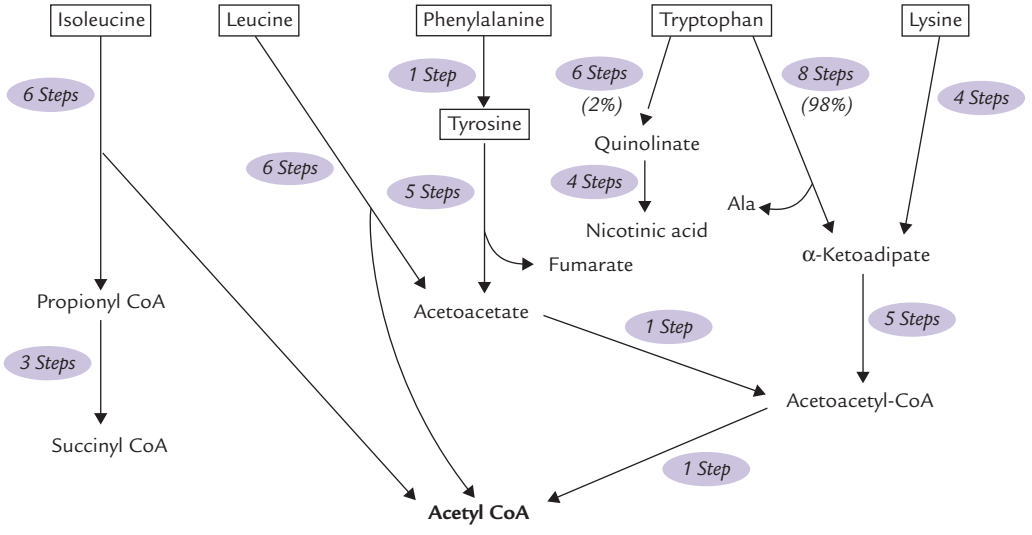


Fig. 13.17. Overview of the metabolism of ketogenic amino acids: the branched-chain amino acids (isoleucine and leucine), the aromatic amino acids (phenylalanine, tyrosine, and tryptophan), and lysine. Only leucine and lysine are exclusively ketogenic.

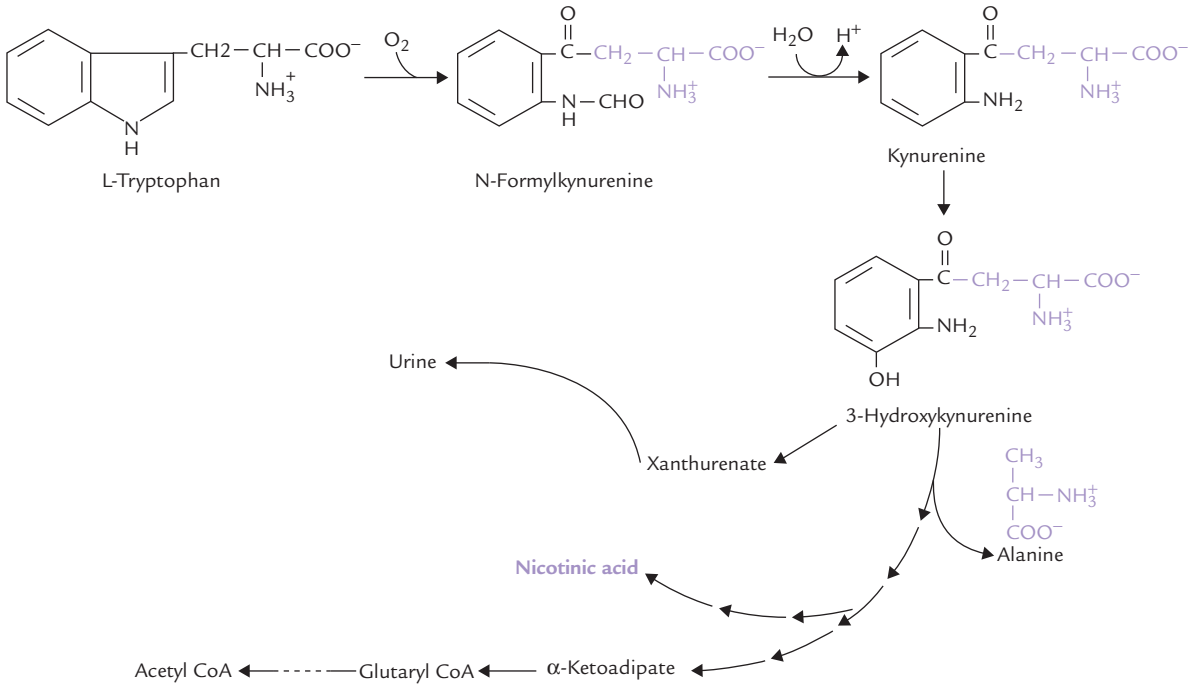


Fig. 13.18. Pathway of tryptophan catabolism. The portion of the molecule outside indole ring forms alanine (a glucogenic precursor), and the balance of carbons are ultimately converted to α-ketoadipate (a ketogenic precursor). This major pathway is diverted to form nicotinic acid, and vitamins.

lysine a ketogenic amino acid. However, reports from animal studies indicate that lysine is both glycogenic and ketogenic. This discrepancy indicates that we lack complete information on the metabolism of this essential amino acid.

4. **Phenylalanine and tyrosine:** Metabolism of these two amino acids produce fumarylacetoacetate, which undergoes an exergonic hydrolysis to form fumarate (glycogenic) and acetoacetate (ketogenic), thereby accounting for the glycogenic and ketogenic nature of these amino acids (Fig. 13.14).

5. **Isoleucine:** This is, likewise, both glycogenic and ketogenic since it produces both succinyl CoA and acetyl CoA.

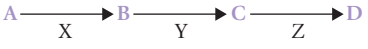
IV. Disorders of Amino Acid Metabolism

A number of inborn errors of amino acid metabolism are known. They occur due to *genetically determined deficiency or absence or modification of a specific protein*. The affected protein may be an enzyme or a transport protein.

- **Enzyme:** In majority of the cases, the deficient protein is an enzyme of a metabolic pathway. As a result, the metabolic pathway is blocked (i.e. *metabolic block*), which causes abnormalities in the normal metabolism. For example, when the enzyme *phenylalanine hydroxylase* that causes conversion of phenylalanine to tyrosine, is deficient, normal metabolism of phenylalanine is disrupted; the condition is called *phenylketonuria*.
- **Transport protein:** In some other cases, the transport proteins responsible for the renal or the intestinal absorption of amino acids, is defective. For example, renal tubular reabsorption of cystine is impaired in *cystinuria*; renal tubular and intestinal transport of the neutral amino acids is impaired in *Hartnup disease*.

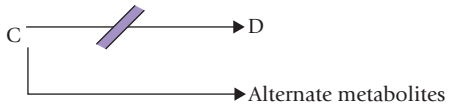
Such disorders are rare, yet they are of considerable significance. Unless early diagnosis is made and appropriate treatment is initiated, the consequences of many of such disorders are disastrous. However, if timely measures are taken, the patient can lead a nearly normal life. Most of these disorders are inherited in an autosomal recessive manner. Phenotypically, the heterozygotes are usually normal.

For a better understanding of such disorders, it would be worthwhile to review the *fundamental design of the normal metabolic pathways, effects of metabolic aberrations and the treatment strategies*. Consider the following pathways consisting of three sequential reactions, catalyzed by enzymes designated as X, Y and Z.



A metabolic block, where the enzyme Z is defective, produces the following effects:

- Decreased concentration of the reaction product, D
- Increased concentration of substrate, C
- Increased production of alternate metabolites from the accumulated substrate.



The treatment strategies of these disorders primarily aim at (i) restricting the substrate intake, and (ii) increasing the supply of the missing product.

For example, in phenylketonuria following measures are recommended:

- A restricted dietary intake of phenylalanine and generous intake of tyrosine.
- Providing the coenzyme of the defective enzyme or replacing the gene coding for the enzyme protein may also be helpful.

Finally, *elimination of the accumulated amino acids or alternate metabolites* by appropriate drugs ameliorates the patient's symptoms in some of these disorders. For example, in treatment of **cystinuria** (an inherited disorder characterized by defective renal tubular reabsorption of cystine and the dibasic amino acids, i.e. lysine, ornithine and arginine) penicillamine is used. This drug complexes with the accumulated cysteine and thus prevents formation of renal stones from cystine.

Some most common errors of amino acid metabolism are discussed in this section.

A. Disorders of Aromatic Amino Acid Metabolism

Phenylketonuria (PKU)

This autosomal recessive trait is the *most common inborn error* of metabolism. In its classical form (**type I**) PKU is caused by complete deficiency of the enzyme *phenylalanine hydroxylase*. In **type II**, the enzyme deficiency is partial. Many mutations of the *phenylalanine hydroxylase* gene (located on chromosome 12q) have been identified, such as missense, nonsense, insertions, deletions and duplications. The incidence of classical PKU is 1 in 10,000–20,000 live births, but shows considerable geographic variation: the incidence in Ireland is 1 in 4000, whereas the condition is rare among Asians.

Variant forms: Variant forms of PKU are also known, e.g. **type III** and **type IV**, which account for 2% of cases of hyperphenylalaninaemia.

- Type III is due to deficiency of *dihydropteridine reductase* or the reductant coenzyme, NADPH (page 15).
- Type IV is due to deficiency of one of the enzymes that catalyzes tetrahydrobiopterin synthesis from GTP.

Because tetrahydrobiopterin is required for hydroxylation of tyrosine and tryptophan also, synthesis of catecholamines from tyrosine and synthesis of serotonin and melatonin from tryptophan are also hampered. Therefore, these variant forms result in more severe clinical manifestations.

Biochemical Abnormalities

1. *Hyperphenylalaninaemia*: The effect of blocking of the *phenylalanine hydroxylase* reaction is accumulation of phenylalanine in blood.
2. *Phenylpyruvate production*: When the concentration of phenylalanine exceeds a certain limit (above 1200 μM; normal is 30–20 μM) conversion to phenylpyruvate by transamination reaction becomes the major metabolic fate of the accumulated amino acid (Fig. 13.19). Among unaffected individuals, such direct transamination is a very minor pathway of phenylalanine metabolism.
3. *Urinary elimination of phenylpyruvate:* Phenylpyruvate is excreted in the urine and accounts for the name "phenylketonuria".
4. *Other urinary metabolites*: In addition to phenylalanine and phenylpyruvate, other major urinary metabolites are formed in these patients due to opening of alternate pathways. These include phenyllactate, phenylacetate and phenylacetylglutamine, a conjugation product of phenylacetate.

Diagnostic tests: The affected infants are normal at birth, so diagnosis of PKU is missed unless screening tests are carried out. These tests in the newborn period are mandatory in the developed Western nations.

Ferric chloride test for urinary phenylpyruvate is widely used: transient blue colour is given by the phenylketonuric urine with this reagent.

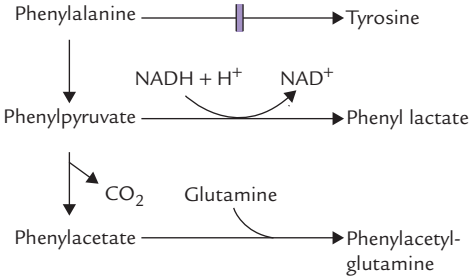


Fig. 13.19. Formation of alternate metabolites—phenylpyruvate, phenyllactate, phenylacetate and phenylacetylglutamine—from the accumulated phenylalanine in phenylketonuria.

Guthrie bacterial inhibition assay: In this test, the blood of phenylketonuric patient—but not the blood of normal individuals—supports the growth of phenylalanine-dependent bacterial strain (*Bacillus subtilis*). Ideally, the test should be done not less than 2 days after birth because false negatives are common within the first 24 or even 48 hours after birth.

Detection of hyperphenylalaninaemia: Measurement of blood phenylalanine level is reliable test; in PKU the level is 1200 μM or even more.

Amino acid analysis of blood by paper chromatography, showing elevated phenylalanine and normal tyrosine is highly diagnostic.

Prenatal diagnosis by the enzyme assay is difficult because *phenylalanine hydroxylase* is expressed only in the liver; but the amniotic cells, which are used for the prenatal diagnosis, do not express the enzyme. DNA-based diagnostic procedures are under study, but are mostly inconclusive because of the allelic heterogeneity in PKU. It means that *different patients have different mutations in the gene for phenylalanine hydroxylase.*

Clinical presentation: The most important clinical presentation is **mental retardation**. Phenylalanine and its metabolites are transferred from the mother to the fetus and impair the fetal brain development. However, PKU is not evident at birth and clinical manifestations appear a few days or weeks later. The child develops mental retardation, with IQ values typically between 25 and 50. The reasons for the development of mental retardation are not clear, some possible mechanisms are discussed in **Case 13.2.**

Neurological signs, such as hyperreactive deep tendon reflexes, hyperactivity and seizures are also present sometimes.

Carrier state: Because PKU is recessive, it manifests clinically only in homozygous state. Heterozygotes, which have approximately half of the normal enzyme activity, can metabolize the normal load of phenylalanine, but not when this amino acid is injected in a dose of 4 g. The transient increase in serum phenylalanine level, seen in unaffected individuals with this dose, is exaggerated in the heterozygotes (*phenylalanine tolerance test*).

Treatment

1. *Low phenylalanine diet*: PKU is treated effectively with a synthetic diet that is low in phenylalanine but which maintains normal nutrition (as an essential amino acid, phenylalanine cannot be omitted entirely from the diet). Since the developing brain is damaged by the biochemical abnormality, prompt initiation of the dietary restrictions is important. These restrictions can be tapered off in older children. If initiated within the first few weeks after birth, rigorously controlled and continued until 5 or 6 years of age, mental development is essentially normal. In some cases, however, continuation of this diet throughout the first decade, or for life, may be necessary.

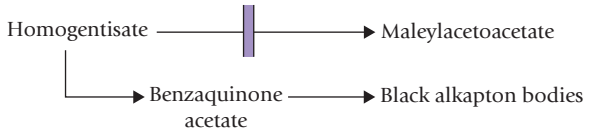
> Tyrosine becomes an essential amino acid for the affected (phenylketonuric) children, but dietary supplements of this amino acid are mostly not required.

2. *Oral tetrahydrobiopterin supplement*: In the variant disorders (biopterin and *biopterin reductase* deficiency) treatment consists not only of regulating the blood levels of phenylalanine but of supplying the missing form of coenzyme, mostly oral tetrahydrobiopterin.
3. *Gene replacement therapy* is under study with some encouraging results.
4. Successfully treated females who have reached reproductive age and become pregnant may expose their offsprings to excess phenylalanine. This will affect the brain development of the fetus and may cause congenital disorders or spontaneous abortions. An extremely strict dietary control throughout pregnancy may be required.

Case 13.2 gives relevant aspects regarding biochemical abnormalities, clinical features and management of a phenylketonuric infant.

Alkaptonuria

The enzyme deficient in this disorder is *homogentisate oxidase* (Fig. 13.14). The disorder is inherited as autosomal recessive trait. Deficiency of the enzyme prevents conversion of homogentisate to maleylacetoacetate, resulting in plasma accumulation and excretion of homogentisate, an uncoloured hydroquinone, in urine. The urine darkens upon exposure to air owing to oxidation of homogentisate to benzoquinone acetate, which is polymerized to black alkapton bodies.



Alkaptonuria was the first to be described as an inherited disease, caused due to an enzyme deficiency (Archibald Garrod, 1902). Along with cytinuria, albinism and pentosuria, it was included in ***Gerrod's tetrad***. The condition causes much anxiety to mother who notices dark nappies which become darker on washing in alkaline soaps and detergents. However, alkaptonuria is a harmless condition and as such does not need any treatment. The problem that the patient may face is development of arthritis in the middle age or later life. This is

due to deposition of alkapton in cartilage; the condition is called **ochronosis**.

However, the relationship between pigment deposition and arthritis is not understood.

Tyrosinaemia

Two different types of tyrosinaemia have been identified: type I and type II, caused by deficiencies of the cytoplasmic *fumarylacetoacetate hydrolase* and *tyrosine transaminase* respectively (Fig. 13.14).

Type I tyrosinaemia, also called *tyrosinosis* or *hepatorenal tyrosinaemia*, is more common, with an incidence of 1.5 per 1000 live births. Accumulation of fumarylacetoacetate and related organic acids causes a cabbage-like odour, abnormal liver function and renal tubular dysfunction. Anaemia and vitamin D resistant rickets are also observed to develop. An abnormal metabolite, succinyl acetone, derived from fumaryl acetoacetate, inhibits haem synthesis, resulting in porphyria-like neurological symptoms.

Type II tyrosinaemia, also referred to as *oculo-cutaneous tyrosinaemia*, manifests as painful corneal erosions and plaques, inflammation (from intracellular crystallization of tyrosine), keratosis of palmar surface and mental retardation. Low-tyrosine and low-phenylalanine diets are beneficial.

Transient tyrosinaemia of the newborn, particularly in premature infants, is the most common form of tyrosinaemia in infancy. It is caused by absence of the enzyme *p-hydroxyphenylpyruvate dioxygenase* (Fig. 13.14). It is mostly a benign condition and responds well to ascorbic acid.

B. Disorders of Branched Chain Amino Acids

Maple Syrup Urine Disease

The condition results due to deficiency of the *branched-chain* ***α-ketoacid dehydrogenase***, the second enzyme of the pathway, as shown in Figure 13.16. The enzyme is rather non-selective, acting on all three branched-chain α-keto-acids. The condition is termed maple syrup urine disease, the name derived from the fact that *odor of urine resembles that of maple syrup*. The prevalence in newborn infants is on the order of 1 in 200,000.

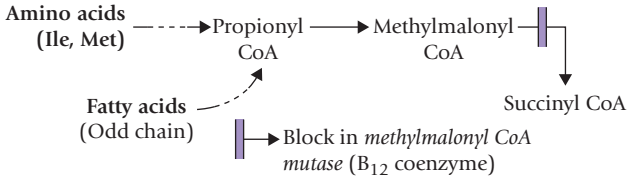
Diagnosis is confirmed by raised plasma levels and enhanced urinary excretion of the branched chain amino acids and their keto acids. If the diagnosis is made in the first week of life, and a diet low in branched chain amino acids is started, normal development is possible. Failure to initiate such measures results in serious consequences—severe **neurological lesions** develop with death occurring in a few weeks or months.

Isovaleric Acidaemia

The deficient enzyme in this condition is *isovaleryl CoA dehydrogenase*, the third enzyme of leucine catabolic pathway (Fig. 13.16). It results in plasma accumulation of isovalerate. The children suffering from this disorder emit a characteristic body odour. They have frequent episodes of vomiting, acidosis and coma, during which the *unusual body odour* is more noticeable. Treatment involves restriction of leucine in diet.

Methyl Malonic Aciduria

This disorder is due to inadequate metabolism of methylmalonyl CoA. This intermediate is obtained during metabolism of propionyl CoA, which is produced during catabolism of some amino acids (isolucine, methionine) or from odd chain fatty acids (Chapter 11).



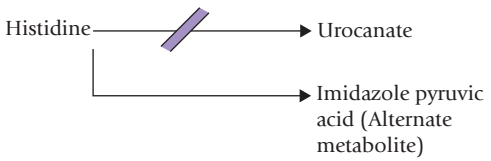
The defective enzyme is *methylmalonyl CoA mutase* which converts the methylmalonyl CoA to succinyl CoA. This enzyme requires deoxyadenosyl cobalamin, a **vitamin B_{12}** derivative, as a coenzyme.

Inadequate metabolism of methylmalonyl CoA may occur due to: (i) inborn error involving the enzyme protein, or (ii) to inadequate supply of the B_{12} coenzyme (**Case 13.3**).

C. Inborn Errors of Histidine Metabolism

Histidinaemia

It is inherited as an autosomal recessive trait. It results due to deficiency of the enzyme *histidase* which is required for the normal metabolism of histidine (Fig. 13.13). Elevation of blood levels of histidine and an alternate metabolite, imidazole pyruvic acid results.



Increased urinary excretion of these metabolites also follows. Incidence of this disorder in newborns is on the order of 1 in 200,000. Most reported cases have shown mental retardation and speech defects, but some remained symptom-free.

Diagnosis is readily made with **ferric chloride test** since the imidazole pyruvic acid gives blue–green colour with this reagent. Other diagnostic procedures include determination of *serum histidine, enzyme determination in skin biopsy*, and the *urocanate concentration in sweat*. (*Histidase* is present only in the skin and liver, and urocanate is a normal constituent of sweat).

D. Inborn Errors of Sulphur-Containing Amino Acids

Homocystinurias

These are autosomal recessive disorders of methionine metabolism. The metabolic defects known to lead to this group of heritable defects are deficiencies of *methionine adenosyltransferase, cystathionine synthase and cystathionine lyase*, that catalyze reactions 1, 4 and 5 respectively of the pathway shown in Figure 13.15. The first enzyme defect leads to *hypermethioninaemia*, and the third defect to *cystathionuria*, but both are benign conditions, not associated with any clinical abnormality.

The second enzyme defect (deficiency of *cystathionine synthase*) leads to plasma accumulation of methionine, homocysteine and homocystine (a dimer of two homocysteines), and urinary elimination of homocystine. In fact, *this condition is the best known cause of homocystinuria in humans*.

Homocystinurias lead to several clinical manifestations, discussed in following paragraphs, that appear to be initiated by the plasma accumulation of homocysteine, which gradually starts depositing in various tissues. The tissue accumulation interferes with the maturation of collagen and elastin, probably by binding copper, which is required for the activity of the enzyme *lysyl oxidase*. This results in skeletal deformities. The patient may also suffer from mental retardation, increased susceptibility to thrombosis and posterior dislocation of lens.

Restriction of dietary methionine *is the mainstay of treatment of homocystinurias*. Supplements of **vitamin B_{12}** and **folic acid** also are often given for enhancing the homocysteine to methionine reaction. **Betaine** (N,N,N-trimethylglycine), which is a methyl group donor in an alternative reaction for the synthesis of methionine from homocysteine, can be employed with the same aim of boosting the conversion of homocysteine to methionine.

Homocysteine as risk factor for CAD: Several studies have shown the relationship between homocysteine and altered endothelial cell function leading to thrombosis. Thus, elevated homocysteine in blood appears to be an independent risk factor for occlusive vascular disease. A rise of merely 6 μmol/L of homocysteine in plasma enhances the risk for occlusive vascular diseases as much as cholesterol increase of 20 mg/dL does. Five to ten per cent of the general population has mild hyperhomocysteinaemia.

Cystinuria

It is a disorder of renal and gastrointestinal tract **transport of cystine**, that also affects lysine, ornithine and arginine. The four amino acids share a common transport mechanism.

Because cystine is relatively insoluble, in cystinuric patients it may precipitate in the renal tubules and form **cystine calculi**. This is the major complication of the disease. To avoid this, **penicillamine** is used for the treatment. The cysteine-penicillamine complex being relatively more soluble, tends to be rapidly excreted.

Cystinosis

It is a familial condition, characterized by deposition of cysteine crystals in various tissues and organs. The crystals are deposited in lysosomes, because of an abnormality in transport of cysteine across cell membrane. Thus, cystinosis appears to be a lysosomal disorder. There is *generalized amino-aciduria* and the renal functions are seriously affected. Many patients die of renal failure at an early age.

Hereditary *Sulphite Oxidase* Deficiency

This may occur alone or along with *xanthine oxidase* deficiency. Both enzymes contain molybdenum. Patients with *sulphite oxidase* deficiency exhibit mental retardation, motor seizures, cerebral atrophy, and lens dislocation. Dietary deficiency of molybdenum can cause deficient activities of both *xanthine oxidase* and *sulphite oxidase*.

Several inborn errors in amino acid metabolism have been identified. Mostly these errors result due to enzyme defects, e.g. phenylketonuria (defect in *phenylalanine hydroxylase*), maple syrup urine disease (defect in *α-keto acid dehydrogenase*, homocystinuria defect in *cystathionine synthase*).

F. Other Inborn Errors

Hartnup Disease

Impairment of transport of neutral amino acids in the intestinal mucosa and renal tubules is the cause of this rare autosomal recessive disease. *Amino aciduria* involving large neutral amino acids (valine, isoleucine, leucine, tyrosine, phenylalanine and tryptophan) is a prominent feature. Tryptophan is the most remarkably affected amino acid. Its depletion results in features of pellagra

(dermatitis and dementia) because part of the niacin requirement normally is covered by endogenous synthesis from tryptophan.

In Hartnup disease, since the intestinal absorption of amino acids is decreased, the amino acids tend to accumulate in the intestinal lumen where they are degraded by the intestinal bacteria. Degradation of tryptophan in this way yields indole compounds, namely *indolyl acetic acid* and *indolyl acetyl glutamine*. These compounds have neurotoxic action, which accounts for the neurological symptoms of these patients.

Blue Diaper Syndrome and Familial Renal Iminoglycinuria

These are the other disorders resulting due to defective amino acid transport.

- Intestinal transport proteins for tryptophan are defective in blue diaper syndrome.
- Renal transport of glycine, proline and hydroxyproline is impaired in familial renal iminoglycinuria.

Hyperprolinaemias

Deficiencies of the *proline oxidase* and *Δ1-pyrroline 5-carboxylate dehydrogenase*, the first and the second enzyme, respectively of the metabolism of proline (Fig. 13.12) results in hyperprolinaemia type I and type II respectively. Both are clinically harmless autosomal recessive traits.

Non-ketotic Hyperglycinaemia

It is a rare recessively inherited disease (approximately 1 in 125,000 live births) of glycine metabolism. *Glycine synthase* is apparently absent, with the result that glycine levels in the blood are increased and glycine is excreted in the urine. Most infants die shortly after birth and some survive with profound mental retardation.

Primary Hyperoxaluria

It is a genetic disorder due to a deficiency of the enzyme *glycine transaminase* (converts glyoxylate to glycine coupled with impaired oxidation of glyoxylate to formate). Because of deficiency of *glycine transaminase*, there is accumulation of glyoxylate. The accumulated glyoxalate is channeled into production of oxalates which is excreted in urine. Enhanced urinary excretion of oxalate carries risk of developing calcium oxalate stone in the genitourinary tract.

V. Biosynthesis of Amino Acids

The bacterial cell is capable of synthesizing all the 20 primary amino acids. The human cell can synthesize 11 of them (i.e. non-essential amino acids) but lacks capacity to make the other nine (i.e. essential amino acids). It is important to synthesize amino acids in adequate quantities and desired proportions so as to meet the body requirements. *Failure to synthesize even a single amino acid results in negative nitrogen balance*. In this state, more proteins are degraded than are synthesized and excessive excretion of nitrogen occurs.

This section deals with biosynthesis of the non-essential amino acids in the human cell and the control devices that coordinate the biosynthetic pathways.

A. Nitrogen Fixation

Assimilation of nitrogen into amino acids starts with nitrogen fixation by certain microorganisms. These organisms bring about reduction of nitrogen to ammonia by the enzyme *nitrogenase*. The reaction is one of the most important reactions in cellular biosynthesis because it traps atmospheric nitrogen which can then be used for the synthesis of biomolecules. Thus, it serves as a link between the living and the non-living. Some of such nitrogen-fixing microorganisms, namely **blue-green algae** and certain soil-bacteria, invade roots of leguminous plants and form root nodules. It is in these nodules that nitrogen fixation occurs.

The next step in assimilation of the trapped nitrogen into biomolecules is its incorporation into amino acids. Glutamate and glutamine play a major role in this regard. Glutamate is synthesized by incorporation of ammonia into a keto acid (i.e. α-ketoglutarate); the reaction is catalyzed by *glutamate dehydrogenase*.

Another ammonia molecule is then assimilated into glutamate to form glutamine; the reaction is catalyzed by the enzyme *glutamine synthetase*, and is driven forward by ATP hydrolysis. Thus, nitrogen gets entry into the metabolic mainstream in the form of glutamate and glutamine.

In assimilation of ammonia, glutamic acid and glutamine are formed initially.

B. Biosynthesis of Non-essential Amino Acids

The carbon skeletons of the non-essential amino acids can be derived from glycolytic and Krebs cycle intermediates

(Fig. 13.20). Alanine and serine can be derived from glycolytic intermediates, pyruvate and 3-phosphoglycerate respectively, and glycine can be obtained from serine (discussed in the following sections). The carbon atoms of cysteine can be derived from serine via cystathionine as previously noted (Fig. 13.15). The carbon atoms of aspartate and asparagine can be derived from oxaloacetate, and the carbon atoms of glutamate, glutamine and proline are derived from α-ketoglutarate. Tyrosine is produced from phenylalanine, as discussed.

Carbon skeletons for the non-essential amino acids can be derived from glycolytic and Krebs cycle intermediates.

Synthesis by Transamination: Pyruvate, oxaloacetate and α-ketoglutarate gain an amino group by participating in the transamination reaction to form the corresponding amino acids. In this way, pyruvate (3-C) forms alanine (3-C), oxaloacetate (4-C) forms aspartate (4-C) and α-ketoglutarate (5-C) yields glutamate (5-C). Glutamate can also be synthesized by amination of α-ketoglutarate by the enzyme *glutamate dehydrogenase*, as discussed earlier.

Synthesis by Amidation: Glutamine and asparagine can be synthesized from glutamate and aspartate respectively, by formation of an amide linkage. This linkage is formed between ammonia and the carboxyl group(s) present in R group(s) of these dicarboxylic amino acids.

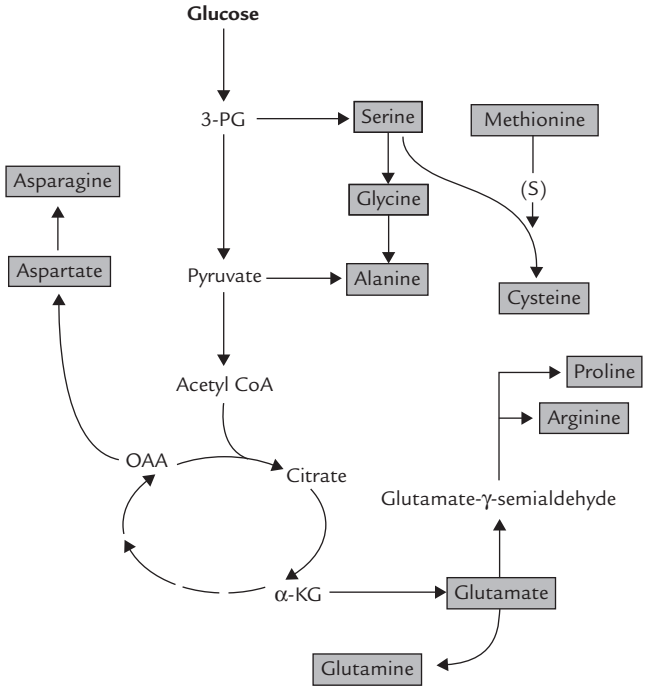
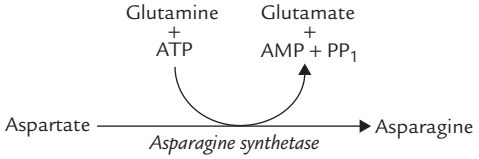


Fig. 13.20. Synthesis of non-essential amino acids (OAA = oxaloacetate, α-KG = α-ketoglutarate, 3-PG = 3-phosphoglycerate, S = sulfhydryl group).

Synthesis of glutamine is catalyzed by the enzyme *glutamine synthetase*, as discussed earlier. Synthesis of asparagine, catalyzed by *asparagine synthetase*, takes place by a similar reaction. However, the donor of amide group in this instance is glutamine. The reaction is driven forward by pyrophosphate cleavage of ATP.



Synthesis of Proline and Arginine

Proline can be obtained from glutamate via glutamate γ-semialdehyde (refer Fig. 13.12). The conversion of glutamate to the semialdehyde is intricate because free carboxylate groups are difficult to reduce. Formation of an energy rich intermediate (γ-glutamylphosphate) has been reported.

Glutamate → γ-Glutamylphosphate → γ-Glutamate γ-semialdehyde

Spontaneous cyclization of the semialdehyde with removal of a water molecule occurs next to yield Δ^1-pyrroline 5-carboxylate. The final reaction is to add hydrogen across a double bond, to form proline. This synthetic pathway occurring in intestinal mucosal cells is irreversible and different from the hepatic and renal pathway that catabolizes proline back to glutamate.

Glutamate-γ-semialdehyde can produce arginine also. First, this compound forms ornithine by transamination (Fig. 13.12), which then produces arginine via the last reactions (2, 3 and 4) of urea cycle (Fig. 13.7). The capacity for arginine synthesis is however limited, making this a *semi-essential amino acid*.

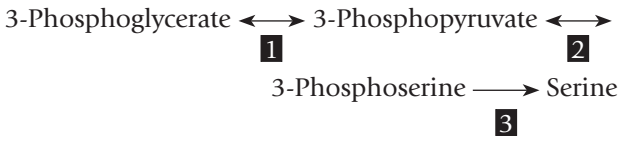
Synthesis of Serine, Glycine and Cysteine

Serine is synthesized from the glycolytic intermediate 3-phosphoglycerate (Fig. 13.20). The reactions are as below:

Reaction 1: An *NAD^+-dependent dehydrogenase* catalyzes the oxidation of the alcohol of 3-phosphoglycerate to form 3-phosphopyruvate. This is a simple oxidation-reduction reaction similar to the *lactate dehydrogenase* reaction and is bidirectional.

Reaction 2: 3-Phosphopyruvate undergoes a transamination reaction with glutamate to form 3-phosphoserine.

Reaction 3: A *phosphatase* catalyzes removal of the phosphate group to yield serine.



Glycine is formed from serine in one step by the *serine hydroxymethyl transferase reaction*, as noted earlier.

Cysteine is synthesized during catabolism of methionine (Fig. 13.15) obtaining its carbon atoms from serine and sulphur group from methionine.

Synthesis of Tyrosine

Tyrosine can be synthesized by hydroxylation of phenylalanine by the enzyme *phenylalanine hydroxylase*.

C. Biosynthesis of Semi-essential Amino Acids

Arginine can be synthesized in a limited amount, but histidine, the other semi-essential amino acid cannot be synthesized in humans. Nevertheless, histidine-free diet for several weeks does not produce any ill effects. This is probably due to its release from **carnosine** (α-alanyl-histidine), a dipeptide that is present in large quantities in muscle tissue. Carnosine also prevents alteration of pH during severe muscle contractions when large amounts of lactic acid are formed, thereby acting as a pH buffer.

VI. Amino Acids as Precursors of Specialized Products

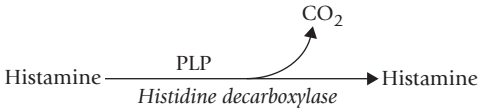
Amino acids serve as precursors of a large number of biologically important compounds; for example, amines, porphyrins, nitrogenous bases of phospholipids, creatine phosphate, polyamines, etc. Compounds like purines and pyrimidines are derived in part from amino acids (Fig. 13.21).

A number of useful products are obtained from amino acids, e.g. creatine, haem, glutathione and purines (from glycine), catecholamines, melanin and thyroxine (from tyrosine), histamine (from histidine), GABA (from glutamate), polyamines (from arginine), etc.

A. Histamine

Histamine is a *biogenic amine* produced from histidine. It is a major *mediator of anaphylaxis and several allergic processes*. It is released by circulating basophils and mast cells typically in response to antigenic stimulation of surface-bound IgE.

Biosynthesis: Histamine is produced from histidine by removal of its α-carboxyl group; the reaction is catalyzed by *histidine decarboxylase*, a pyridoxal phosphate (PLP) dependent enzyme.



Another enzyme, aromatic *α-amino acid decarboxylase*, can also bring about decarboxylation of histidine. However, specificity of this enzyme is broader since it can cause decarboxylation of aromatic amino acids as well.

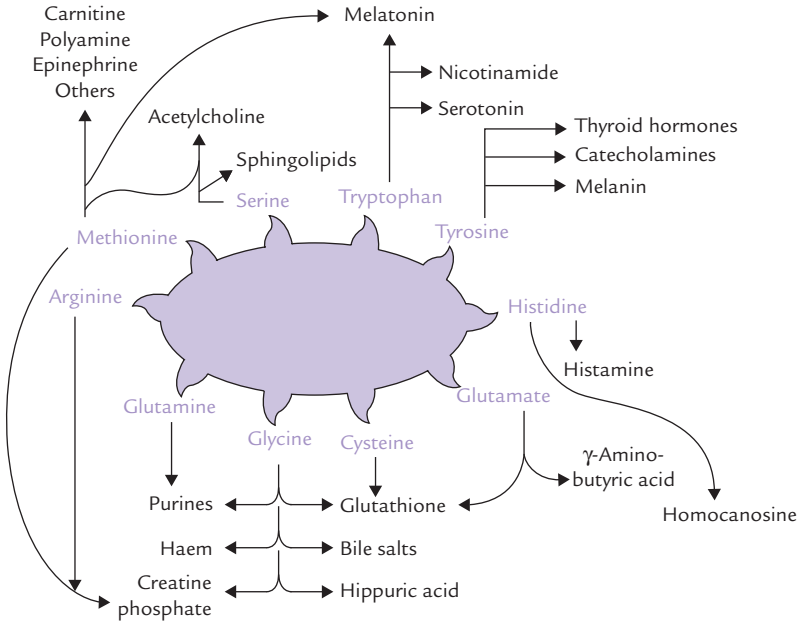


Fig. 13.21. Special products obtained from amino acids.

Functions: Histamine dilates small blood vessels, increases capillary permeability, contracts bronchial and intestinal smooth muscles, stimulates gastric acid secretion and nasal fluid discharge, and regulates cells of the immune system. These effects are elicited following its interaction with two different types of receptors: the **H_1 receptor** and the **H_2 receptor**.

- **H_1 receptors** are located on some smooth muscle cells. Interaction of histamine with these receptors accounts for the above stated effects on blood vessels and smooth muscles. The anti-allergic drugs like *diphenhydramine* and *chlorophenaramine* oppose action of histamine by occupying these receptors, hence blocking them from action of histamine.
- **H_2 receptors** are located on the gastric mucosa. Binding of histamine to these receptors results in increased secretion of hydrochloric acid. *Cimetidine*, a common drug used for the treatment of hyperacidity, occupies the H_2 receptor. This decreases the histamine-induced acid secretion.

Histamine is obtained by decarboxylation of histidine by *PLP dependent decarboxylase(s)*. It is mediator of anaphylaxis and severe allergic reactions.

B. Gamma (γ) Aminobutyric Acid

γ-Aminobutyric acid (GABA) is an *inhibitory neurotransmitter* in the brain.

Biosynthesis: GABA is produced by removal of the γ-carboxyl group of glutamate by the enzyme *glutamate decarboxylase*. The enzyme is present in the GABAergic nerve terminals in the brain, especially the grey matter. The insulin secreting beta cells of pancreas also form this enzyme and this has a bearing on the occurrence of diabetes mellitus.

Further metabolism of GABA occurs by a transamination reaction to form succinate semialdehyde which is then oxidized to succinic acid (Fig. 13.22).

The first two enzymes of this pathway require pyridoxal phosphate (PLP), but *glutamate decarboxylase* is more susceptible to drugs that interact with pyridoxal phosphate (probably because its Km for the PLP is higher). Example of such *antipyridoxal drugs* are the hydrazides, e.g. hydralazine, that are used for their vasodilatory effect on arteriolar smooth muscle. But overdosage with such antipyridoxals can cause convulsions as formation of the inhibitory GABA is prevented.

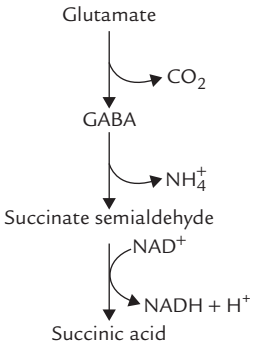


Fig. 13.22. Synthesis and metabolism of γ-aminobutyric acid (GABA).

Functions: GABA is concentrated in the synaptic vesicles where it enhances passage of chloride ions through the post-synaptic membrane. This action of GABA may account for its role as an *inhibitory neurotransmitter.* It is noteworthy, that glutamate (the precursor of GABA) acts as an excitatory neurotransmitter.

GABA and Huntington's disease: Various movement disorders arise as a result of defects in the basal ganglia of the brain, e.g. *parkinsonism* and *Huntington's disease*. (The former is discussed in a subsequent section, while consideration of the latter is relevant here because it involves degeneration of GABAergic neurons.) The disease is named after the scientist *George Huntington*, who described it first in 1872. Loss of the inhibitory pathways results, which leads to uncontrolled (*choreic*) movements that characterize this condition.

GABA, an inhibitory neurotransmitter, is synthesized by PLP dependent decarboxylation of glutamate, an excitatory neurotransmitter, in GABAnergic neurons. Degeneration of these neurons causes Huntingdon's disease, characterized by uncontrolled (choreic) movements.

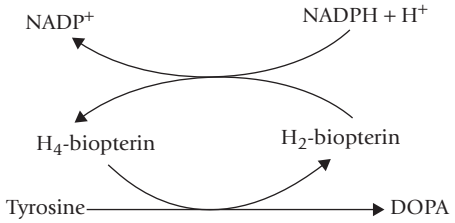
Note: GABA links covalently with histidine to form **homocarnosine**, present in the brain. However, function of this dipeptide is not clear.

C. Catecholamines

Epinephrine, *norepinephrine* and *dopamine* are collectively referred to as catecholamines. They are synthesized from the same precursor amino acid, i.e. tyrosine. In common with the other compounds containing amino groups such as histamine or serotonin, catecholamines are also known as biogenic amines.

Biosynthesis: The reaction sequence for the synthesis of catecholamines is shown in Figure 13.23. Synthesis of epinephrine requires four steps, whereas only three and two steps suffice for norepinephrine and dopamine respectively. The initial reactions up to dopamine take place in the *cytoplasm*, whereas the formation of norepinephrine is catalyzed by enzymes within the *synaptic vesicles* in neurons. Conversion of norepinephrine to epinephrine occurs in cytosol of the *chromaffin granules* of the adrenal medulla.

(a) The *first step* involves hydroxylation of tyrosine to dihydroxyphenylalanine (L-DOPA) by the enzyme, *tyrosine hydroxylase*, found in catecholaminergic nerve terminals. This is a *monooxygenase* that requires molecular oxygen, ferric ions and tetrahydrobiopterin as a reducing agent. The latter is formed from dihydrobiopterin at the expense of NADPH.



(b) In the *second reaction*, DOPA loses its α-carboxy group to form dopoamine. The enzyme catalyzing this step is *DOPA decarboxylase*. Dopamine is then taken up into synaptic vesicles where conversion of dopamine to norepinephrine would occur. This process is dependent on a vesicular *ATPase*.

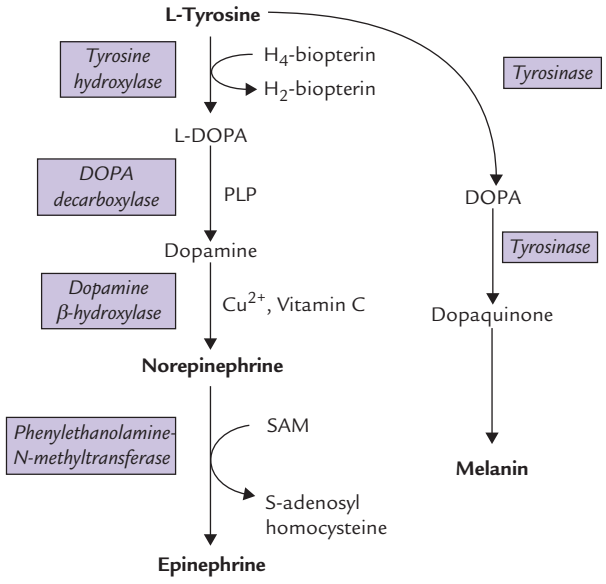


Fig. 13.23. Conversion of tyrosine to epinephrine, norepinephrine or melanin (DOPA = dehydroxy phenylalanine, PLP = pyridoxal phosphate).

(c) Within synaptic vesicles (of adrenergic neurons), the vesicular *dopamine β-hydroxylase* catalyzes the formation of noradrenaline. The enzyme is a copper containing *monooxygenase* that requires ascorbic acid and molecular oxygen.

(d) The *final step* involves formation of epinephrine. Though in majority of CNS, norepinephrine is the final product, but a few neurons in the brain, and chromaffin cells of adrenal medulla, epinephrine is produced by action of the cytosolic enzyme, *N-methyl transferase*. The enzyme catalyzes the transfer of a methyl group from S-adenosylmethionine (SAM) to form adrenaline.

$$\text{Norepinephrine} \xrightarrow[\text{Transmethylation}]{\text{SAM}} \text{Epinephrine}$$

Although the pathway to three catecholamines is common, the *end product formed in a given cell depends on its enzymatic outfit*. Some cells (e.g. dopaminergic neurons of nigrostriatal system) having only the first two enzymes synthesize dopamine, whereas the cells of adrenal medulla having all the four enzymes of the pathway synthesize both epinephrine (80%) and norepinephrine (20%).

Synthesis of three catecholamines: dopamine, norepinephrine, epinephrine, occurs by a comman biosynthetic pathway starting from tyrosine. The pathway operates partly in cytosol and partly in synaptic vesicles of neurons. Chromaffin cells of adrenal medulla and few neurons of brain can synthesize epinephrine (from norepinephrine) also.

Regulation of biosynthesis: The committed step of the biosynthetic pathway is the *tyrosine hydroxylase reaction*, which is *allosterically inhibited by dopamine and norepinephrine* when the adrenal medulla is unstimulated. Continuous stimulation of the adrenal medulla (as during prolonged stress) promotes *tyrosine hydroxylase* activity, primarily because the turnover of dopamine and norepinephrine is rapid.

Functions

- Norepinephrine is a *major transmitter in the sympathetic nervous system*; stimulation of these nerves is responsible for various features of the *'fight or flight'* response, such as stimulation of the heart rate, sweating, vasoconstriction in the skin and bronchodilatation. There are also norepinephrine containing neurons in the CNS, largely in the brain stem.
- Epinephrine is produced in the adrenal medulla; it is more active than norepinephrine on the heart and lungs.

- Dopamine is both an intermediate in the synthesis of norepinephrine and a neurotransmitter. It is a major transmitter in nerves that interconnect the nuclei of the basal ganglia in the brain and control voluntary movements. Damage to these nerves causes Parkinson's disease.

Parkinson's disease: This second type of movement disorder, Parkinsonism (the first is Huntington's disease, discussed earlier), was first described by James Parkinson in 1817. It involves loss of dopamine neurons resulting in deficiency of dopamine synthesis. The condition is relatively common after the fifth decade of life, with an incidence of 1 in 1000. Nerve transmission is affected in substantia nigra of the upper brain stem (because of dopamine depletion), which accounts for the clinical features of this disease, e.g. muscular rigidity, tremors, expressionless face and slow rhythmical movements.

Treatment involves administration of DOPA. This compound enters the brain cells where it is converted to dopamine. Administration of dopamine for the treatment is ineffective because dopamine is incapable of entering the brain cells.

D. Serotonin and Melatonin

Both these compounds are derived from tryptophan (Fig. 13.24).

Biosynthesis: Serotonin is found in enterochromaffin cells, brain and platelets. In the former two it is synthesized from tryptophan, whereas in platelets, serotonin is taken up from plasma. Synthesis involves the following steps: (i) Hydroxylation of tryptophan by tryptophan *5-hydroxylase* followed by (ii) decarboxylation by aromatic *L-amino acid decarboxylase* (Fig. 13.24). Hydroxylation, the *rate-limiting* reaction, is analogous to that of phenylalanine, and requires molecular oxygen and tetrahydrobiopterin.

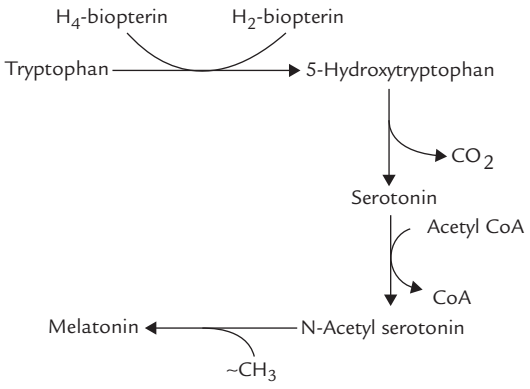


Fig. 13.24. Synthesis of serotonin and melatonin.

Synthesis of melatonin from serotonin occurs in *pineal gland* by N-acetylation followed by methylation.

Functions: Serotonin is a neurotransmitter in the brain and serves as a precursor of melatonin in the pineal gland. The serotoninergic neurons are concentrated in the upper brain stem, but project up to the cerebral cortex and down to the spinal cord. They are more active when the subject is awake than when he is asleep, and serotonin may control the degree of responsiveness of motor neurons in the spinal cord. In addition, it regulates various vegetative body functions, such as appetite, sleep, sexual activity and aggression. It is a potent smooth muscle constrictor and a vasoconstrictor. It may act as a transmitter in gastrointestinal tract to evoke release of peptide hormones. It might play a role in intestinal motility also.

Melatonin: is a biologically active compound present in pineal gland. It exerts an inhibitory effect on gonads. It blocks actions of melanocyte stimulating hormone (MSH) and adrenocorticotrophin. Synthesis of melatonin is sensitive to light and darkness. In light, the synthesis is stimulated whereas in dark it is diminished. These effects are thought to be mediated via cAMP.

Under normal circumstances, only about 1% of tryptophan is converted to serotonin; the rest either undergoes degradation or enters kynurenine pathway to form nicotinamide (Chapter 18). However, about **60%** or more of tryptophan is channeled towards serotonin formation in a condition called **malignant carcinoid** (argentaffinoma). Widespread tumour cells in the argentaffin tissue of the abdominal cavity characterize this condition. Symptoms of nicotinamide deficiency (i.e. *pellagra*) and negative nitrogen balance develop in this condition.

E. Melanin

Melanins are the dark pigment of skin, hair, iris and the retinal epithelial cells, derived from tyrosine. They are formed in organelles called melanosomes. The latter occur in pigment producing cells called melanocytes. Two types of melanins occur in humans: the brown-black **eumelanins** and the yellow-red **pheomelanins.** The latter are derived, in addition to tyrosine, from cysteine also.

Biosynthesis: A series of enzymatic and non-enzymatic oxidation and coupling reactions are required to form melanins (Fig. 13.23).

- Tyrosine is first hydroxylated by the enzyme *tyrosinase* to form dihydroxyphenylalanine (DOPA).
- The same enzyme then converts the DOPA to dopaquinone.

- Depaquinone then undergoes a series of enzymatic and non-enzymatic reactions, some of which are catalyzed by *tyrosinase* to form polymeric melanin molecules.

It should be noted that the enzyme that catalyzes formation of DOPA from tyrosine in nerve and adrenal medulla (*tyrosine hydroxylase*) is different from *tyrosinase*, although both form the same product.

Functions: Melanin is derived from a Greek world, *melan*, meaning black. It is a polymeric product with heterogenous molecular weight and poorly defined structure. It is present in skin, hair, choroid plexus, substantia nigra, and in retina and ciliary body in eye, imparting them their characteristic brown colour. Melanin protects our skin because it absorbs not only the visible light but also the UV radiation. The latter is a natural component of sunlight (wavelength 280–320 nm) and is dangerous because it induces DNA damage, a common cause of both sunburn and skin cancer (Chapter 21).

In the eye, melanin is present in pigment epithelium underlying the sensor cells of the retina where it absorbs stray light. This enhances visual acuity and prevents overstimulation of the photoreceptors.

Albinism: It refers to a variety of recessively inherited conditions that exhibit hypomelanosis based on metabolic defects in the melanocytes of the eye and skin. They occur with frequencies of approximately 1 in 50,000 in most parts of the world. *The classical form of albinism arises due to deficiency of the enzyme tyrosinase*. Such *tyrosinase* negative individuals lack detectable pigment in the skin, hair, or eyes. For this reason, this disorder is sometimes called **complete perfect albinism.**

Other forms of albinism called **complete imperfect albinism** occur in *tyrosinase*-positive individuals. These individuals have some detectable *tyrosinase* activity but fail to synthesize melanin in adequate amounts. This may because of a *transport defect*–decreased ability to transport tyrosine to a site where it will form melanin (**Case 13.4**).

Various clinical forms of albinism have been recognized. These are: ocular, oculocutaneous, and cutaneous, depending on the tissue involved.

The albinos are abnormally sensitive to light; even a slight exposure to light causes burns and can lead to skin cancer. Photophobia is a common feature.

F. Thyroid Hormones

The generic term "thyroid hormones" refers to the iodinated amino acid derivatives $\mathbf{T_3}$ (triiodo-L-thyronine) and $\mathbf{T_4}$ (tetraiodo-L-thyronine). They are synthesized from iodine and the tyrosyl residues present in **thyroglobulin**, a homodimeric glycoprotein (MW 669,000). It contains about 134 tyrosyl residues. Up to 25–30 of the tyrosine side chains become iodinated, but not more than 8–10 of these are processed to active hormones (Chapter 30).

G. Creatine Phosphate and Creatinine

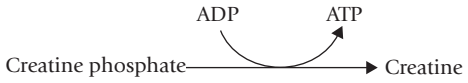
Creatine (*N-methylguanidinoacetate*) is an amino acid-derived product that is present in muscle tissue, and to a lesser extent in nervous tissue. It plays a pivotal role in the metabolism of high-energy phosphates. Creatinine is a dead end metabolite formed by spontaneous cyclization of creatine and creatine phosphate.

Biosynthesis: Creatine phosphate is synthesized from three amino acids: glycine, arginine and methionine. The biosynthetic pathway consists of three sequential reactions, one reaction each occurs in kidney, liver and muscle (Fig. 13.25).

Functions: In vertebrate muscles, creatine phosphate occurs as a reservoir of high-energy phosphate groups. It helps generation of ATP in exercising muscles by substrate-level phosphorylation. Nearly 1% of the weight of skeletal muscle is accounted for by this compound. Small amount is present in smooth muscles, testes, liver and kidneys as well.

Creatine phosphate possess an energy rich phosphate bond, with a standard free energy of hydrolysis of −10.3 kcal/mole.

This value is much higher than the standard free energy of ATP hydrolysis (−7.3 kcal/mole). Therefore, hydrolysis of creatine phosphate can be coupled with concomitant generation of ATP (i.e. **substrate level phosphorylation**).



During muscle contraction, when the ATP: ADP ratio declines, ATP is regenerated rapidly by the above reaction. This reaction, called the reversible *creatine kinase* reaction (Fig. 13.25) is the most important source of ATP during the first few seconds of muscle contraction. For more sustained muscular activity, however, ATP has to be regenerated by (anaerobic) glycolysis or oxidative metabolism.

The compound corresponding to creatine phosphate in invertebrates is **arginine phosphate**. These high-energy compounds of muscles are termed **phosphogens**.

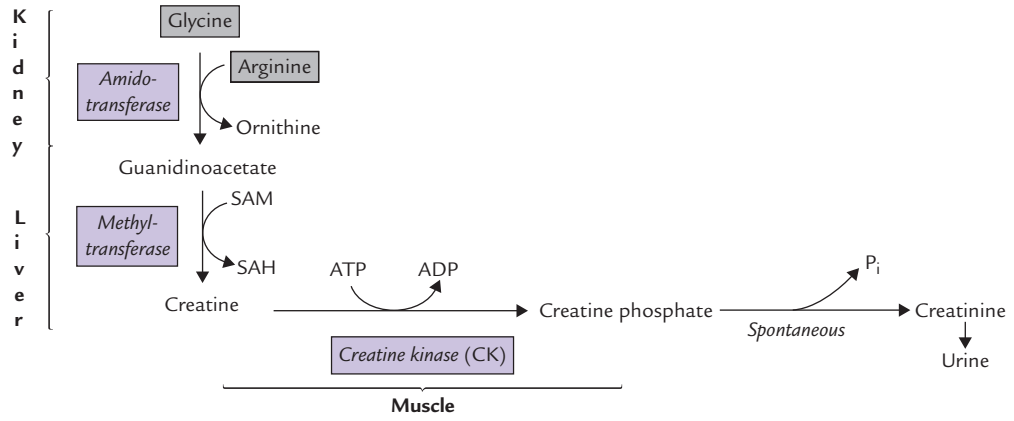
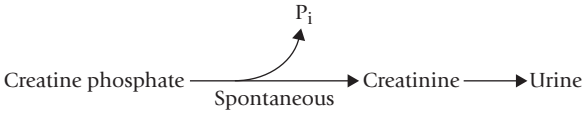


Fig. 13.25. The synthesis of creatine phosphate and its spontaneous conversion to creatinine (SAM = S-adenosyl methionine, SAH = S-adenosyl homocysteine).

Clinical implications: Creatine phosphate is relatively unstable at the pH prevailing in the sarcoplasma, and is non-enzymatically converted to creatinine in this organelle. This conversion takes place continuously in healthy muscle. The creatinine so produced is released in circulation transported to kidneys and excreted in the urine.



This has two important implications in clinical medicine:

1. The amount of creatinine excreted in the urine over a 24-hour period, which correlates with the muscle mass, is constant for a given individual (about 15 mg/kg of body weight). The quantity of creatinine is measured in the 24-hour urine specimens to validate that the collection was complete.
2. It is a useful kidney function test because its blood level is remarkably constant. It is neither secreted nor reabsorbed in the tubular system, and so its excretion during a specified time in a **creatinine clearance test** serves as a measure of glomerular filtration rate.

Small quantity of creatine is also excreted in urine. Muscle wasting due to any cause, such as muscular dystrophy, starvation, diabetes, fever and thyrotoxicosis results in increased excretion of creatine in urine.

H. Carnitine

Carnitine is a coenzyme that carries long-chain fatty acids into mitochondrion, thereby playing an important role in β-oxidation (Chapter 11). It is synthesized from trimethyllysine, which is obtained during normal turnover of proteins like myosin and histone.

I. Nicotinamide

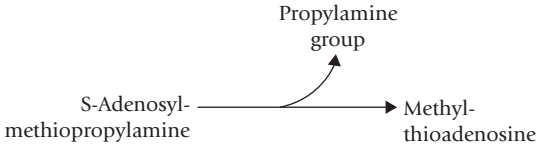
Nicotinamide is a water-soluble vitamin synthesized from tryptophan (Fig. 13.21). In humans, about 3% of tryptophan is converted to nicotinamide derivatives. **The pathway of its formation, termed the kynurenine pathway,** is discussed in Chapter 18. **Nicotinate**, a related compound, is also obtained from tryptophan by the kynurenine pathway. Its administration decreases the serum cholesterol levels. However, larger doses are toxic. Nicotinamide is free from these side effects.

J. Polyamines

There are *aliphatic amines possessing multiple amino groups*. Polyamines were originally identified in sperms, but now they are known to exist in a number of other tissues. Biologically important polyamines are **putrescine, spermine** and **spermidine**.

Biosynthesis: Polyamines are synthesized from *ornithine*, which is derived from arginine. The following steps are involved:

1. *Ornithine decarboxylase*, a PLP-dependent enzyme, catalyzes decarboxylation of ornithine to form putrescine (Fig. 13.26).
2. Another PLP-dependent enzyme, *S-adenosylmethionine decarboxylase*, catalyzes decarboxylation of S-adenosylmethionine to form S-adenosylmethiopropylamine, an activated donor of propylamine group.



3. The S-adenosylmethiopropylamine donates the propylamine group first to putrescine to form spermidine, and then to spermidine to form spermine.

Functions: Polyamines are basic in nature and possess multiple positive changes. Hence, they readily associate with negatively charged cellular macromolecules such as nucleic acids. They also interact with the membrane phospholipids and negatively charged residues of membrane bound proteins. Because of these interactions, polyamines carry out a variety of biological functions:

1. The interaction of polyamines with DNA and RNA may stimulate synthesis of proteins and nucleic acids and enhance cellular proliferation.
2. The polyamines may open certain channels because of their binding with anionic sites on the cell membranes. This may lead to enhanced transport of specific substances.
3. Polyamines may play a role in hormone mediated actions and in membrane fusion during exocytosis and endocytosis due to their association with membrane phospholipids.
4. Certain enzymes are inhibited by polyamines, e.g. *protein kinase*.

Biogenic Amines

These are a group of basic compounds of diverse origin and functions. They are synthesized by decarboxylation of amino acids (and their derivatives), e.g. histamine from histidine and taurine cysteine. Examples are listed below:

Histidine	→	Histamine
Tryptophan	→	Tryptamine
Tyrosine	→	Tyramine
Lysine	→	Cadoverine
Serine	→	Ethanolamine, choline
DOPA	→	Dopamine

K. Glutathione

Glutathione is a tripeptide of glutamic acid, cysteine and glycine with the sequence L-γ-glutamyl-L-cysteinylglycine. It plays a major role in the maintenance of proteins in their reduced forms (Chapter 27).

L. Acetylcholine

Acetylcholine is a *neurotransmitter of the parasympathetic nervous system and sympathetic ganglia*. Stimulation of the parasympathetic system produces slowing of heart rate, bronchoconstriction, and stimulation of intestinal smooth muscle; these effects are broadly opposite to those of the sympathetic system.

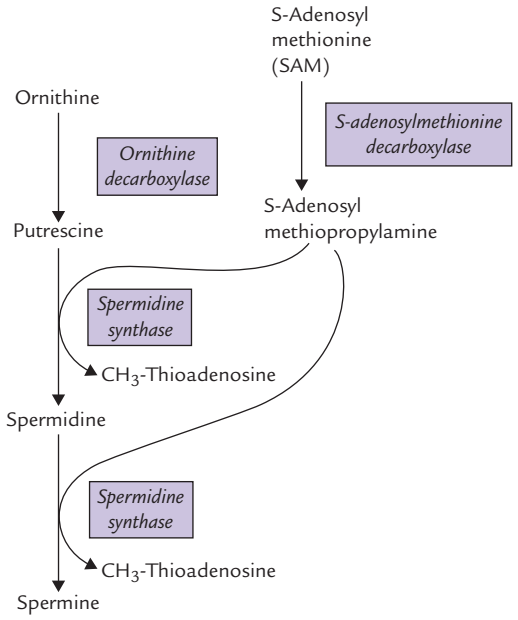


Fig. 13.26. Biosynthesis of polyamines.

Acetylcholine also serves as the *neurotransmitter of the neuromuscular junctions* where motor nerves contact skeletal muscle cells (i.e. motor end plate) and cause them to contract. Neurons containing this transmitter also exist in the brain, and they are thought to be involved in learning and memory.

Synthesis of acetylcholine occurs by the cytoplasmic enzyme *choline acetyl transferase*. The initial precursor is the amino acid, serine (Fig. 13.21), which undergoes decarboxylation to form ethanolamine. The latter receives three methyl groups from SAM molecules to form choline which combines with an acetyl group to form acetylcholine.

After acetylcholine is secreted into the synaptic cleft, it is largely degraded by the enzyme *acetylcholinesterase*. The breakdown products, choline and acetate, are taken up rapidly into the nerve terminal, where they are used for the resynthesis of acetylcholine.

M. Others

A number of other compounds, in addition to those just mentioned, are also obtained from amino acids. Glycine, the smallest of all amino acids, is a precursor for the largest number of such products (Fig. 13.21).

Special products of glycine: Glycine participates in generation of several important products:

1. *Hippuric acid:* It is formed by an amide linkage between the carboxyl group of benzoic acid and the amino group of glycine.
2. *Bile salts:* Glycine becomes conjugated with bile acids to form conjugated bile salts.
3. *Haem:* Glycine combines with succinyl CoA in the first step of the haem biosynthetic pathway.

4. *Creatine,* discussed earlier.
5. *Purine nucleotides* (Chapter 19).

Exercises

Essay type questions

1. Describe the transport of ammonia, and highlight how ammonia detoxication in brain differs from that in liver. Explain biochemical basis of ammonia toxicity.
2. Describe reactions in the synthesis of urea.
3. Draw a thumbnail sketch of the TCA cycle and indicate where each of the amino acids may enter this cycle.
4. Outline metabolism of phenylalanine. Name special products obtained from tyrosine and describe synthesis of any one of them.

Write short notes on

1. Amino acid pool
2. Nitrogen balance
3. Transamination
4. Deamination
5. Ammonia toxicity
6. One-carbon metabolism
7. Albinism
8. Polyamines
9. Alkaptonuria
10. Phenylketonuria
11. Hartnup disease
12. Maple syrup urine disease
13. S-Adenosyl methionine
14. Parkinsonism
15. Huntington's disease
16. Homocystinurias

CLINICAL CASES

CASE 13.1 Lapsing into coma 3 days after birth

A full term child, born after normal pregnancy, developed grunting respiration 36 hours after birth. After three days he had severe bouts of vomiting and became lethargic and unresponsive to stimuli. He was admitted in the hospital nursery, where a blood sample was obtained for analysis.

Test	Patient's reports	Reference range
pH	7.58	7.35–7.45
Urea	11 mg%	15–45 mg%
Ammonia	468 µg/dL	20–63 µg/dL

Chromatography of the plasma sample was carried out and compared with that of a normal child. The results indicated that plasma levels of citrulline, glutamine and alanine were grossly elevated, but the argininosuccinate level was decreased. Treatment was started immediately.

Q.1. Suggest a probable diagnosis.
Q.2. In view of the test results, identify the biochemical lesion.
Q.3. Comment on other biochemical test results.
Q.4. Why did this defect become apparent only a few days after birth?
Q.5. What treatment do you suggest for this child?

CASE 13.2 A 2-week old infant having convulsions

An infant of 2 weeks had convulsions. He was born after a normal pregnancy and had taken his feeds normally. His mother had observed a peculiar mousy odor in the child's urine. The urine was tested by ferric chloride test: characteristic green colour was observed which indicated the presence of phenylpyruvic acid.
Quantitative analysis of the blood and urine yielded values for phenylalanine and its metabolites.

Metabolite	Patient's reports	Reference range
Plasma phenylalanine	1.8 mmol/L	< 0.09 mmol/L
Urine phenylalanine	4.8 mmol/L	Trace
Urine phenylpyruvate	6.2 mmol/L	Absent
Urine phenyllactate	11.2 mmol/L	Absent

Similar results were obtained on repeating tests after few days. Liver biopsy was performed and cell-free extract was prepared from the hepatocytes. Radioactive phenylalnaine was added to the extract and amount of radioactive tyrosine produced from it was measured. It was found to be less than 1% of that in a normal subject. However, addition of tetrahydrobiopterin (THB) to the reaction mixture enhanced the tyrosine production. Addition of dihydrobiopterin (DHB) did not make any difference.

Q.1. Identify the biochemical defect.
Q.2. Comment on the biochemical test results.
Q.3. Explain the cause of convulsions.
Q.4. The children affected by this condition often have very fair skin. Explain why?
Q.5. Is this condition injurious to brain? Discuss the mechanism.

CASE 13.3 Recurrent vomiting in an infant

A 4-month-old-female infant was referred to the hospital OPD with recurrent vomiting, convulsions, respiratory distress and failure of adequate growth. Biochemical examination showed decreased plasma bicarbonate concentration. Urinary ammonium excretion was enhanced. Specialized analysis in the hospital laboratory found excessive methyl malonic acid in the urine sample. Injection of vitamin B_{12} markedly decreased the urinary excretion of this compound.

Q.1. Identify the metabolic block in this infant?

Q.2. Urine chromatography of the infant revealed presence of homocystine and cystathionine and plasma level of methionine was decreased. Comment on these observations.
Q.3. If an infant with similar clinical features and biochemical results does not respond to vitamin B_{12} injection, what could be the biochemical defect?
Q.4. Comment on the other biochemical test results, i.e. decreased plasma bicarbonate and enhanced urinary ammonium excretion.

CASE 13.4 A 3-year-old child with white hair and abnormally fair skin

A 3-year-old child was exceedingly fair skinned, had white hair, and red pupils with bluish iris. On examination, visual impairment (astigmatism) was detected. A sample of melanocytes was obtained by skin biopsy and a cell free extract was prepared from these cells. Radioactive tyrosine was added to the extract and formation of radioactive dihydroxyphenylalanine was measured. It was found to be markedly decreased as compared to that expected in a normal subject. Production of radioactive melanin was also found to be greatly diminished, being less than 1% of that expected in a normal person. Addition of terahydrobiopterin had no effect on production of either dihydroxyphenylalanine or melanin.

Q.1. Identify the biochemical defect in this child.
Q.2. Given that tyrosine to DOPA conversion is an important step in biosynthesis of catecholamines, do you expect any abnormality of catecholamine metabolism in this child?
Q.3. In an albino, addition of radioactive tyrosine to the cell free extract resulted in normal synthesis of melanin. However, when melanocytes were cultured and radioactive tyrosine was added to them, decreased synthesis of radioactive melanin occurred. Identify the biochemical defect in this patient.

ELECTRON TRANSPORT, OXIDATIVE PHOSPHORYLATION AND MITOCHONDRIAL MEMBRANE TRANSPORTERS

CHAPTER

14

Flow of electrons along the electron transport chain (ETC) is the final event in the cellular respiration which releases energy for generation of ATP. The electron transport chains are embedded in the **inner mitochondrial membrane.** Each chain consists of a series of electron carrying molecules. Electrons move from one carrier protein to another in a stepwise fashion; each carrier being capable of accepting electrons (either with or without accompanying protons) from the preceding carrier and donating them to the next one. At the end of ETC, the electrons reach oxygen, the final acceptor of electrons.

The electrons for the ETC are obtained during catabolism of foodstuffs, e.g. carbohydrates (particularly glucose), fatty acids and amino acids. As these substances undergo oxidation, they lose electrons to the reduction–oxidation cofactors NAD^+ and FAD (less commonly to others) to generate NADH and $FADH_2$, respectively. These reduced cofactors can give rise to ATP by transferring their electrons to the electron transport chain. Initially these electrons are energy rich, but as they flow down the ETC, much of their energy is lost. The lost energy is released in small packets, most of which are used for the generation of ATP, while the rest is dissipated (i.e. *entropy*). Flow of electrons along the ETC (i.e. *oxidation*) and generation of ATP (i.e. *phosphorylation*) are coupled processes, together referred to as **oxidative phosphorylation.** The phosphorylation follows the oxidation; if the latter is inhibited, the former is also similarly affected. Oxidative phosphorylation is described in this chapter followed by a discussion about other carrier proteins that are located in the inner mitochondrial membrane.

At the end of this chapter, the student should be able to understand:

- Electron transport chain: its localization, carrier proteins, redox couples, free energy changes, reactions and inhibitors.
- Oxidative phosphorylation, hypotheses for ATP generation during ETC and uncouplers.
- Various membrane transport systems located in the inner mitochondrial membrane.

I. Electron Transport Chain (ETC)

A. Sources of Electrons for ETC

The electrons for the ETC are released during catabolic pathways of biomolecules such as carbohydrates, fats, and amino acids (Fig. 14.1) by action of the enzymes known as ***dehydrogenases***. These electrons are then funneled into the ETC. For example, during glycolytic sequence a pair of electrons is removed from glyceraldehyde 3-phosphate. In TCA cycle, an electron pair is removed from each of the following substrates: isocitrate, α-ketoglutarate, succinate and malate by specific *dehydrogenases* (Table 14.1). These *dehydrogenases* remove a pair of hydrogen atoms initially. Since each hydrogen atom contains an electron, removal of two hydrogen atoms implies removal of an electron pair.

These electrons travel down the ETC, and combine with the last acceptor, i.e. oxygen, and two protons are also taken up from the surrounding medium. This results in the formation of water (Fig. 14.2).

Respiratory chain or electron transport chain located in inner mitochondrial membrane consists of a series of electron carriers. Electrons from the reduced coenzymes, NADH and $FADH_2$, pass through these before they reduce oxygen.

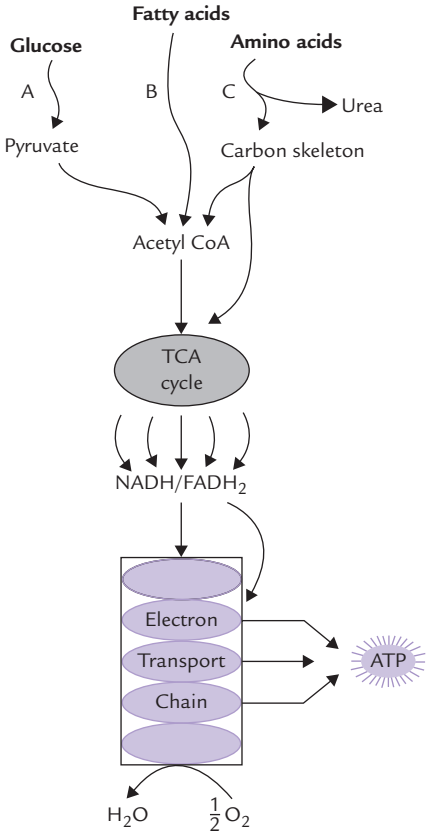
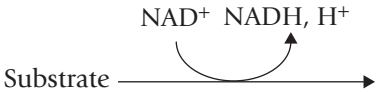


Fig. 14.1. Electron transport and oxidative phosphorylation re-oxidize NADH and $FADH_2$ and trap the energy released as ATP. The NADH and $FADH_2$ arise from all major catabolic pathways: glycolysis, (A); β-oxidation of fatty acids, (B); amino acid catabolism, (C); and TCA cycle.

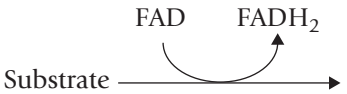
Coenzymes linked with dehydrogenases: The electron pairs removed from the substrate molecules do not enter ETC directly. They are first transferred to specialized coenzymes, *nicotinamide adenine dinucleotide* **(NAD$^+$)** or *flavin adenine dinucleotide* **(FAD)**.

- Certain *dehydrogenases*, termed ***NAD$^+$-linked dehydrogenases***, transfer the electrons to NAD$^+$-to form NADH.



(i)

- Other *dehydrogenases*, called ***FAD-linked dehydrogenases***, transfer electrons to FAD to form $FADH_2$.



(ii)

Subsequently the reduced coenzymes, NADH and $FADH_2$, transfer electrons to the ETC at different levels. Some NAD$^+$-linked and FAD-linked *dehydrogenases* present in catabolic pathways are given in Table 14.1.

During all major catabolic pathways, electron pair is removed from the substrate and transferred to NAD$^+$ or FAD. This forms reduced coenzymes—NADH and $FADH_2$, respectively. Indeed, formation of reduced coenzymes is common feature of all catabolic pathways.

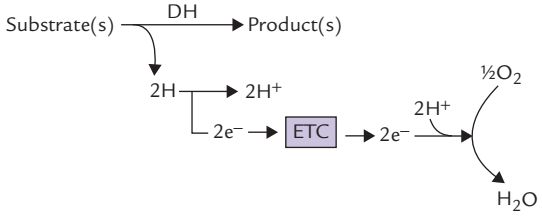


Fig. 14.2. The electron pair, removed from the substrate travels down the electron transport chain (ETC) to the final acceptor, oxygen (DH = *dehydrogenase* enzyme).

Table 14.1. Action of FAD-linked and NAD$^+$-linked *dehydrogenases*

Enzyme	Reaction	Location
NAD$^+$-linked		
Pyruvate dehydrogenase	Pyruvate → Acetyl CoA	M
Isocitrate dehydrogenase	Isocitrate →α-Ketoglutarate	M
Malate dehydrogenase	Malate → Oxaloacetate	M
α-Ketoglutarate dehydrogenase	α-Ketoglutarate → Succinyl CoA	M
Glutamate dehydrogenase	Glutamate → α-Ketoglutarate + Ammonia	M
Glyceraldehyde-3-P dehydrogenase	Glyceraldehyde 3-P→1, 3-Bisphosphoglycerate	C
Lactate dehydrogenase	Lactate → Pyruvate	C
FAD-linked		
Succinate dehydrogenase	Succinate → Fumarate	M
Fatty acyl CoA dehydrogenase	Fatty acyl CoA → Enoyl CoA	M

M = mitochondrion, C = cytoplasm.

B. Localization of ETC

Electron transport chains are located in the inner mitochondrial membrane (IMM). The catabolic pathways that yield electrons for ETC occur in the mitochondrial matrix. Thus, there is close proximity of catabolic pathways with ETC which ensures that the electrons obtained during these pathways are rapidly transferred to ETC (Fig. 14.3). Moreover, the ATP synthesizing system is also located in IMM. This permits a rapid utilization of the energy, which was released earlier during the electron flow, for the synthesis of ATP. Since all these processes concerned with oxidative phosphorylation take place in mitochondria, it is appropriate to examine the anatomy of mitochondria in some detail.

Biochemical Anatomy of Mitochondrion

Mitochondrion has two surrounding membranes: the outer mitochondrial membrane and the inner mitochondrial membrane. The **outer mitochondrial membrane** has a relatively simple structure. It contains special pores which make it permeable to most small molecules and ions. A few enzymes are also located in the membrane. The **inner mitochondrial membrane (IMM)** has a more complex and specialized structure. It contains a number of protein components such as ***enzymes, transport proteins, several sets of electron transport chains,*** and the ***ATP synthesizing system.*** A large surface area is required to accommodate all these components. For this purpose, the membrane structure is highly convoluted, being thrown into numerous folds, called *cristae,* which serve to increase surface area of the IMM several-fold (Fig. 14.3). Thus, the IMM of a single mitochondrion in liver may contain over 10,000 sets of electron transport chains and *ATP synthase* molecules.

In contrast to the outer mitochondrial membrane which is freely permeable, the inner mitochondrial membrane is **selectively permeable**, meaning that it is impermeable to most ions (H^+, K^+, Na^+) and small molecules (ADP, ATP), and permeable only to a few: only water, carbon dioxide and oxygen can freely move across the IMM. Movement of other substances across the membrane can take place only through mediation of specific transport proteins.

Presence of these protein components makes the IMM unusually **rich in proteins** which constitute 75% or more of the total membrane weight. Rest of the membrane structure is formed by lipids. This contrasts with the other cell membranes which contain less amount of proteins (Chapter 7).

Mitochondrial matrix is the space enclosed by the inner mitochondrial membrane. It contains a gel-like solution in which several *catabolic pathways,* e.g. *citric acid cycle, β-oxidation of fatty acids,* and *oxidation of amino acids,* occur. In addition to enzymes of these pathways, the matrix contains coenzymes such as NAD^+, $NADP^+$ and FAD, and components of the phosphorylation reaction, e.g. ADP, ATP and phosphate ions.

> Proximity of catabolic pathways (occur in matrix) with the electron transport chains and ATP synthesizing systems (located in IMM) ensures that (i) the electrons obtained during catabolism are promptly transferred to ETC and (ii) there is rapid utilization of energy for ATP synthesis.

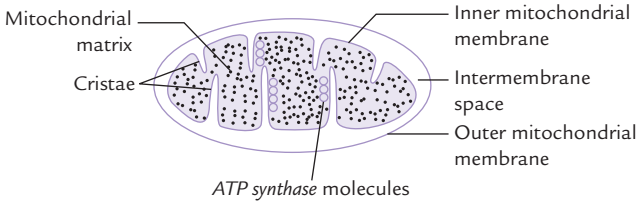


Fig. 14.3. Biochemical anatomy of a mitochondrion.

C. Electron Transport: An Overview

ETC consists of several electron carriers that are arranged in a sequence shown in Figure 14.4.

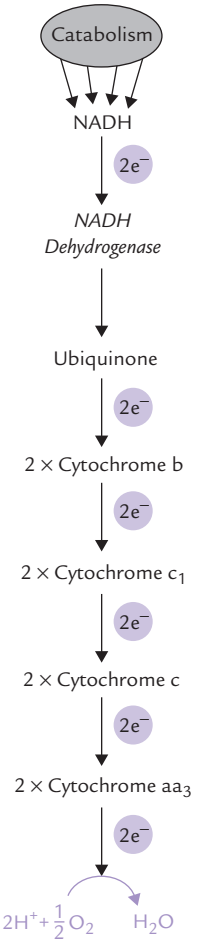
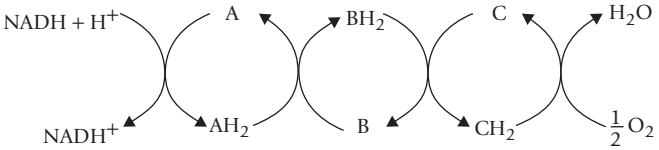


Fig. 14.4. An overview of electron transport chain.

The electrons are transported from one carrier molecule to the next either as *hydride ion* ($:H^-$) which bears two electrons (discussed later), or as hydrogen atoms or as free electrons.

The starting compound, that loses its electrons to the ETC is NADH and the final acceptor of the electrons is oxygen. NADH is formed by transfer of energy-rich electron pair removed from a substrate during catabolism, as mentioned earlier (equation i). NADH loses its electrons to the first carrier of ETC (*NADH dehydrogenase*). Thereafter, the electrons pass down from one carrier to another. During the electron transfer, the components of ETC accept or donate electrons, either with or without accompanying protons. A simple transport chain is depicted as below.



The sum of all these reactions is:

$$NADH + H^+ + \frac{1}{2}O_2 \rightarrow NAD^+ + H_2O \qquad \text{(iii)}$$

The electrons may travel, not only as free electron (or hydride) but also as hydrogen atoms. This is logical as a hydrogen atom consists of one electron and one proton.

Oxidations are broken down into several sequential reactions with smaller free energy changes. In each step the starting carrier is oxidized and the next carrier is reduced. This step-wise arrangement of electron transfer enables the body to capture the energy for work rather than simply dissipate it as heat.

D. Redox Couples and Redox Potential

Why do electrons flow unidirectionally from NADH (or $FADH_2$) to oxygen, and what is the thermodynamic origin of the high energy bonds being created? The answer may be obtained by looking at the redox potentials of various carriers. As noted, a given carrier oscillates between the reduced and the oxidized states. For example, ubiquinone (CoQ) interchanges between a reduced form, i.e. ubiquinol or $CoQH_2$ and an oxidized form, i.e. ubiquinone or CoQ. *The reduced and the oxidized forms of the same carrier are together referred to as redox couple.* Thus, the CoQ and the $CoQH_2$ constitute redox couple, as do the oxidized cytochrome b and the reduced cytochrome b.

Flow of electrons between these two redox couples is shown in an expanded form below:

$$CoQH_2 \longrightarrow CoQ + 2H^+ + 2e^-$$
$$2 \times \text{Cyt b (oxd)} + 2e^- \longrightarrow 2 \times \text{Cyt b (red)}$$
Sum: $CoQH_2 + 2 \times \text{Cyt b (oxd)} \longrightarrow CoQ + 2 \times \text{Cyt b (red)} + 2H^+$

The tendency of a redox couple to donate or accept electrons is determined by *standard redox potential* (E_0'). Every redox couple has a characteristic value of E_0'. Electrons always flow from the lower to the higher redox potential, so that redox systems with negative E_0' tend to donate electrons to redox systems with positive E_0'. For example:

- Standard redox potential of $NADH/NAD^+$ is –0.32V.
- That of $CoQH_2/CoQ$ is +0.10V.
- Therefore, direction of flow of electrons is from $NADH/NAD^+$ to ubiquinol/ubiquinone and never the opposite way.

For the same reason, electrons flow from $FMNH_2$/FMN (E_0' −0.30V) to $CoQH_2$/CoQ (E_0' +0.05V). In this way, all components of the respiratory chain are arranged to range from the most negative E_0' (−0.32V of NADH/NAD^+) to the most positive E_0' (+0.82V of H_2O/O_2) pair.

Standard redox potential of components of ETC are shown in Table 14.2.

Redox reactions are electron-transfer reactions requiring a reductant (e^- donor) and an oxidant (e^- acceptor). Together they constitute a redox couple. The standard redox potential is the most useful property of redox couples.

E. Free Energy Changes During Electron Flow

Just as all spontaneous transformations proceed in such a direction that there is a loss of free energy, the electron

Table 14.2. Redox potential of electron carriers

Redox couple	E_0' (volts)
$NADH/NAD^+$	−0.32
$FADH_2$/FAD (Protein bound)	−0.12
$FMNH_2$/FMN (Protein bound)	−0.30
Ubiquinol/ubiquinone	+0.10
Cyt b (Fe^{2+})/Cyt. b (Fe^{3+})	+0.08
Cyt c_1 (Fe^{2+})/Cyt. c_1 (Fe^{3+})	+0.23
Cyt c (Fe^{2+})/Cyt. c (Fe^{3+})	+0.22
Cyt a (Fe^{2+})/Cyt. a (Fe^{3+})	+0.29
Cyt a_3 (Fe^{2+})/Cyt. a_3 (Fe^{3+})	+0.55
H_2O/O_2	+0.82

transfer between the redox pairs is also accompanied by release of free energy. Amount of free energy released is directly proportional to a difference in the standard redox potentials of the redox pairs. A definite relationship exists between the two as below:

$$G^{0\prime} = -nF\,\Delta E_0'$$

G°′ = standard free energy change in kcal/mole.
n = number of reducing equivalents (i.e. electrons) transferred (its value is 2 in the present case, as the electrons are transferred in pairs).
F = Faraday's constant (23.062 kcal V^{-1} mol^{-1}).
E_0' = difference between the standard redox potentials of the electron-donor and the electron-acceptor redox systems.

The given relationship holds good under standard conditions, i.e. concentration of 1.0 M of both the electron donor and the electron acceptor, temperature 25°C and pH 7.0. It can be used to calculate the standard free energy change as a pair of electrons passes along the electron transport chain, from the first redox couple (i.e. NADH/NAD^+; $E^{o\prime} = -0.32$ V) to the last one (i.e. H_2O/O_2; $E^{o\prime} = +0.82$ V).

$\Delta G^{0\prime} = -2\ (23.062)\ [0.82 - (-0.32)] = -52.6$ kcal/mole.

This relationship can also be used to calculate the energy changes for individual segments of the electron transport chain.

Electrons flow from the redox couples with more negative redox potential to those with more positive redox potentials and release energy in the process.

F. Components of Electron Transport Chain

Five distinct types of components are present in ETC: (a) Nicotinamide nucleotides, (b) flavoproteins (containing flavin mononucleotide or FMN; and flavin adenine dinucleotide or FAD), (c) iron–sulphur centres, (d) ubiquinone, and (e) cytochromes. Their arrangement in the respiratory chain has been depicted in Figure 14.5.

Nicotinamide Nucleotides

Nicotinamide adenine dinucleotide (**NAD^+**) and nicotinamide adenine dinucleotide phosphate (**$NADP^+$**) are the coenzymes derived from the vitamin *niacin*. The former is more actively involved in the ETC. It is reduced to NADH by transfer of a pair of electrons (in form of hydride ion) from the substrate by action of various *dehydrogenases* (see NAD^+-linked *dehydrogenases*; Table 14.1).

$NAD^+ + \mathbf{H^+ + 2e^-} \rightarrow$ NADH (Hydride is a proton, H^+, bearing two electrons)

NADH subsequently loses its electrons to the initial component of ETC: *NADH dehydrogenase*, a large protein complex embedded in IMM.

$NADP^+$ is similarly produced by action of *NADP-dependent dehydrogenases*, but it is mostly involved in reductive biosynthesis of biomolecules.

Flavoproteins

Flavoproteins are important hydrogen carriers. Their prosthetic groups are flavins **(FMN** or **FAD)**, which are derivatives of the vitamin *riboflavin*. These coenzymes have

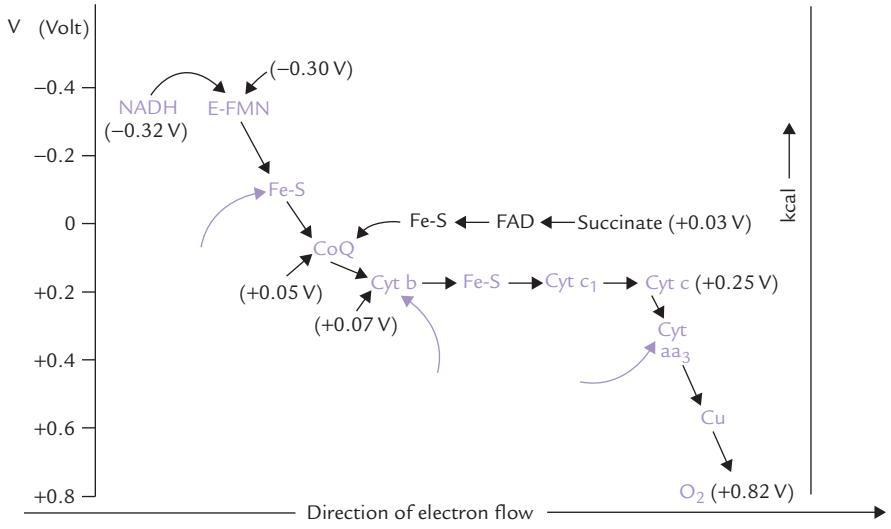


Fig. 14.5. The respiratory chain of mitochondria showing direction of flow of electrons and energy relationships. The arrows (→) indicate the sites of ATP generation (E = *NADH dehydrogenase*, Fe-S = iron-sulphur centres, CoQ = coenzyme Q or ubiquinone, cyt = cytochrome).

chemically related structures. Both have a flavin mononucleotide (FMN) unit, which contains the reactive site. FAD has an additional sugar group and an adenine base, which completes its structure (Chapter 18). FAD and FMN react with two protons plus two electron, in alternating between the reduced and the oxidized state:

$$FAD + \mathbf{2H} + \mathbf{2e^-} \rightarrow FADH_2$$

FMN is prosthetic group of *NADH dehydrogenase*. It is bound firmly to the enzyme protein and does not function as a diffusible co-substrate. FAD, on the other hand, is linked to another flavoprotein, *succinate dehydrogenase*, and functions as a diffusible co-substrate.

Iron-sulphur Centres

They are also known as **non-haem iron proteins** or the iron-sulphur clusters. Several types of iron sulphur (FeS) centres exist, but in each case the iron atoms are coordinated to inorganic sulphur atoms (and the sulphur of cysteine side chains in proteins). Within an FeS cluster, an electron is carried by the iron atom, which, upon accepting the electron, changes from the Fe^{+3} (ferric) state to the Fe^{+2} (ferrous) state. As the electron is passed to another electron carrier, the iron atom of the FeS cluster changes back again to the ferric state.

As explained in the next section, one group of iron-sulphur protein participates in the transfer of electrons from FMN to ubiquinone (CoQ), and the other from cytochrome b to cytochrome c_1 (Fig. 14.5).

Ubiquinone

Ubiquinone, also known as coenzyme Q, is a *mobile diffusible* hydrogen carrier, which can move from donor to acceptor molecules during electron transport. It is a benzoquinone with a hydrocarbon tail of (usually) 10 isoprene (branched 5-carbon) units (Fig. 14.6), which makes it strongly hydrophobic and confines it to the lipid bilayer of the inner mitochondrial membrane. It was also named ubiquinone because of its ubiquitous or widespread distribution.

The names coenzyme Q and ubiquinone and the abbreviations CoQ and UQ are used interchangeably. A subscript(n) indicates the number of isoprene units; for example, CoQ_6 or CoQ_{10} contain 6 or 10 isoprene units respectively. *CoQ_{10} is the homologue that occurs in humans.* **It can accept electrons from $FMNH_2$ or from $FADH_2$ and transfers them to cytochromes.**

Generally, ubiquinone carries two hydrogen atoms, but can also act in one-electron transfers by forming a free-radical intermediate, called semiquinone intermediate.

The reduced form is called *ubiquinol* (Fig. 14.6).

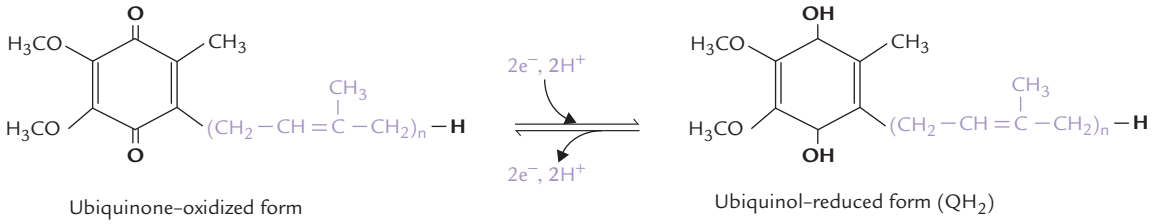
Cytochromes

Cytochromes are the final class of components that participate in electron transport. They are integral membrane proteins (with the exception of cytochrome c), and are named so because they are coloured cellular components (*cyto* cell, *chrome* colour). They are conjugated proteins, containing an *iron-porphyrin as a prosthetic* group. In most cases, this iron-porphyrin is the haem group, also found in oxygen transport proteins, e.g. haemoglobin and myoglobin. In contrast to these proteins, however, in cytochromes the haem iron cannot bind oxygen (or carbon monoxide or any other potential ligand), but rather acts as a reversible electron carrier. Iron in cytochromes (but not in haemoglobin or myoglobin) undergoes *physiological oxidation–reduction between the ferrous (+2) and ferric (+3) states, as electrons are shuttled from one protein to another.*

Why cytochromes can carry only electrons, but cannot bind oxygen? This is shown diagrammatically in Figure 14.7. In cytochromes, the iron group of the haem is anchored on both sides, on one side by histidine and on the other by methionine. In haemoglobin, it is only anchored on one side by histidine, enabling oxygen to interact with the free side of Fe.

The **main varieties of cytochromes** in mitochondria are the **a, b**, and **c types** of cytochromes, the classification being based on the respective absorption spectrum and the type of porphyrin structure present. The cytochrome of b type contains haem, whereas the others have porphyrins with different side chains. Cytochromes a and a_3 contain *"haem a"* rather than "ordinary haem". In this the porphyrin ring of haem is modified: (i) a methyl group of haem is oxidized to a formyl group, and (ii) a hydrophobic isoprenoid chain is attached to one of the vinyl groups. (See chapter 16 for haem structure.)

The **cytochrome families are subclassified** in terms of the historical order of their discovery (b_1, b_2, b_3 ...) or



Ubiquinone–oxidized form

Ubiquinol–reduced form (QH_2)

Fig. 14.6. Structure of ubiquinone, also called coenzyme Q (CoQ) or UQ. The isoprene units are shown in colour.

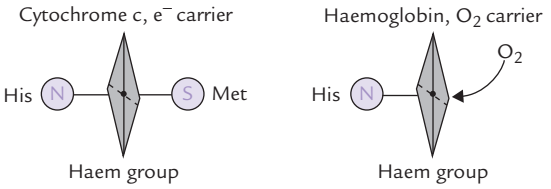


Fig. 14.7. Haem is prosthetic group in both cytochrome C and haemoglobin, but performs different roles. It serves as oxygen carrier in haemoglobin and electron carrier in cytochrome C (His = histidine; Met = methionine).

the wavelength (nm) of a characteristic spectral absorption peak (b_{562}, b_{566}). The electrons are transported from coenzyme Q to cytochromes in the **order**: **b, c_1, c, a** and **a_3**. *The last two can directly react with molecular oxygen.* Besides haem iron, the aa_3 are associated with **copper**, which acts as an electron carrier by a valence change between the cuprous (Cu^+) and cupric (Cu^{2+}). As described later, it (aa_3) participates in the last reaction and transfer of electrons to molecular oxygen.

Respiratory chain contains flavoproteins, iron-sulphur proteins, ubiquinone, cytochromes and protein-bound copper. Each component participates in electron transfers by changing oxidation state between the oxidized form and the reduced form.

G. Pathway of Mitochondrial Electron Transport

Only four components of the respiratory chain are freely diffusible (NADH, ubiquinone, cytochrome c and oxygen), and the rest are organized as constituents of large **protein complexes** (Fig. 14.8). There are four complexes (I to IV) embedded in the inner mitochondrial membrane. Exact detail of the complexes is not known. What is known is: (i) that *a complex contains many polypeptide or protein subunits and several iron centres*, (ii) that these components can be readily reduced or oxidized, and (iii) that they transfer electrons as they flip between the reduced state and the oxidized state.

Complex I

Oxidation of NADH begins with Complex I (also called ***NADH-ubiquinone reductase*** or ***NADH dehydrogenase* complex),** which comprises 28–41 protein subunits (depending upon the species), FMN as a prosthetic group and about 7-iron-sulphur (FeS) centres. It transfers electrons from NADH to ubiquinone via FMN and FeS centre:

$$\text{Substrate} \rightarrow \textbf{NADH} \longrightarrow \underset{\textit{Complex I}}{\text{FMNH}_2 \rightarrow \text{FeS}} \rightarrow \text{Ubiquinone}$$

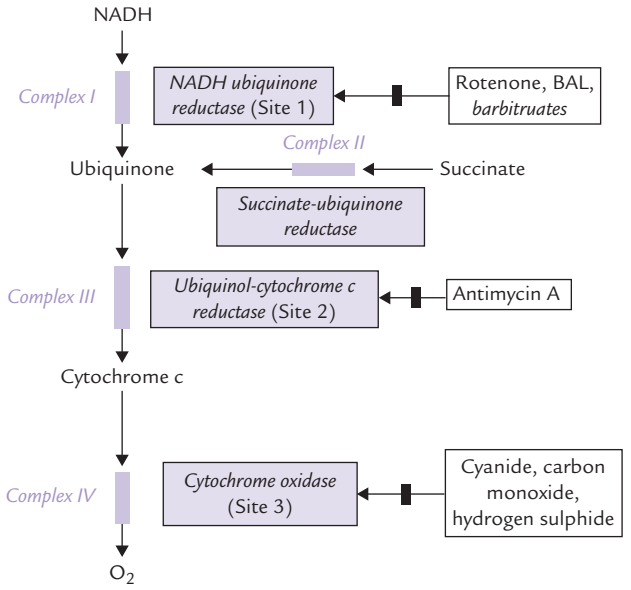


Fig. 14.8. The structured complex of functionally related electron carriers. Inhibitors of ETC (▬) are also shown. Sites 1, 2 and 3 refer to the sites of proton pumping ("phosphorylation sites").

As the electron pair flows from NADH to complex I, it is accepted together with a hydrogen ions, H^+, such that two electrons and two H^+ are accepted in total. As a result FMN is converted into $FMNH_2$. The electrons are then transferred within the complex I to iron-sulphur clusters, and then passed onto ubiquinone, which is thereby converted to ubiquinol (Fig. 14.6). *Ubiquinone has a long, flexible, lipid soluble arm, and so it can readily move through the inner membrane* to transfer the electrons to the next enzyme in the sequence, i.e. the Complex III (the Complex II will be dealt with later).

The, complex I contains various electron carriers that work sequentially to carry electrons down the chain to ubiquinone, which then transfers them to the complex III.

Complex III

Complex III (***ubiquinol-cytochrome c reductase* complex**) once again uses iron atoms to shuffle electrons within its structure. It contains cytochrome b, an iron-sulphur centre (called Rieske's protein), and cytochrome C_1.

$$QH_2 \rightarrow \underset{\textit{Complex III}}{\text{Cyt b} \rightarrow \text{FeS} \rightarrow \text{Cyt } c_1} \rightarrow \text{Cyt c}$$

Figure 14.9 elaborately depicts the electrons passing from ubiquinol, through the cytochrome b, FeS and cytochrome c_1 components of this complex, and the accompanying ferric-ferrous interconversions.

Note that *cytochrome c is a peripheral membrane protein* bound to the outer membrane surface, which transfers its electrons to Complex IV.

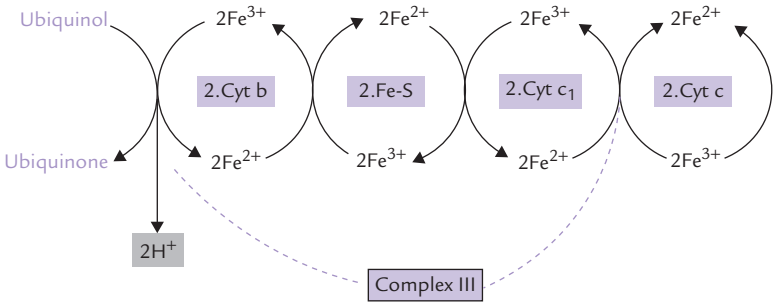


Fig. 14.9. Flow of electrons from ubiquinol to cytochrome c via the large, multiprotein complex III, consisting of cytochrome b, an iron sulphur protein and cytochrome c_1.

Large protein complexes are arranged asymmetrically in the membrane. Ubiquinone (Q) is present at junction of complex I and III and cytochrome c lies at junction of III and IV. Q is in the lipid bilayer and cytochrome c is peripheral membrane protein bound to outer surface of IMM.

Complex IV

Complex IV (***cytochrome oxidase* complex**) contains 13 different polypeptides, two haem groups, and two copper ions. The complex IV transfers electrons to O_2, the final acceptor, to form water:

$$\text{Cyt c} \rightarrow \underset{\textit{Complex IV}}{\text{Cyt aa}_3} \xrightarrow{O_2} H_2O$$

The complex contains two cytochromes (a and a_3), which are associated with haem iron and copper. Cytochrome a is paired with a copper atom, Cu_A and cytochrome a_3 is paired with a different copper atom, Cu_B. During electron transfer, the iron atoms of the haem cycle between the ferric and ferrous states while the copper atoms cycle between cuprous and cupric. The *cytochrome oxidase* reaction is complex: it transfers four electrons to molecular oxygen to form two molecules of water. Oxygen is tightly bound between haem a_3 and copper during its reduction, and it is released only after its complete reduction to water.

$$4e^- + O_2 + 4H^+ \rightarrow 2H_2O$$

Because of high affinity of *cytochrome oxidase* for molecular oxygen, the oxidative phosphorylation is near maximum even at low oxygen pressure.

Complex II

Complex II (***succinate-ubiquinone reductase or succinate dehydrogenase***) contains four polypeptide chains; the first two constitute the bonafide *succinate dehydrogenase* (SDH), a Krebs cycle enzyme, which catalyzes the following reaction:

$$\text{Succinate} + \text{FAD} \rightleftharpoons \text{Fumarate} + FADH_2$$

SDH contains iron-sulphur centres, in addition to the covalently bound FAD. Electrons from $FADH_2$ are transferred to ubiquinone (via cytochrome b_{560} and FeS) and reduce it to ubiquinol. The latter then proceeds to reduce Complex III.

$$\text{Succinate} \rightarrow \underset{\textit{Complex II}}{FADH_2 \rightarrow \text{Cyt } b_{560} \rightarrow \text{FeS}} \rightarrow \text{CoQ}$$

Evidently, *succinate dehydrogenase* (and other mitochondrial flavoproteins) bypasses the Complex I and pass their electrons directly to ubiquinone. This results in **generation of only two ATPs when $FADH_2$ is used as the substrate; three are generated if NADH is used as the substrate.** The reason for this is that *one ATP each is generated at Complex I, III, and IV*.

The respiratory chain contains four large multi-protein complexes. Coenzyme Q shuttles electrons from complexes I and II to Complex III; and cytochrome c shuttles electrons from Complex III to Complex IV, which then transfers them to oxygen, the final acceptor.

H. Inhibitors of ETC

A number of site-specific inhibitors are known to block transport of electrons along the respiratory chain (Fig. 14.8). This action inhibits oxygen consumption and secondarily halts synthesis of ATP from ADP and Pi. Some common inhibitors and their sites of action are as follows:

Rotenone: It blocks transfer of electrons from NADH to ubiquinone. It is a plant product that is extremely toxic and is used by the American Indians as a fish poison.

However, it is not very toxic to humans because of its poor intestinal absorption. Humans, therefore, can safely eat the poisoned fish.

Barbiturates: Besides rotenone, some barbiturates (e.g. *amytal*) inhibit electron flow through Complex I.

Antibiotics: An antibiotic, *British antilewisite* (BAL) also acts at Complex I; and another antibiotic, *antimycin A*, inhibits Complex III.

Cyanide, carbon monoxide azide, and hydrogen sulphide: These inhibitors act at complex IV. They bind to the iron protoporphyrin in cytochrome aa_3 whose sixth coordination position is not occupied by an amino acid side chain but is reserved for oxygen. *Cyanide* and *azide* bind to the ferric form of the iron; *carbon monoxide* binds to the ferrous form.

Hydrogen sulphide also is a potent inhibitor. *Carbon monoxide* not only inhibits electron flow, but also has high affinity for haemoglobin, which adds to its clinically important toxicity.

Electron transport inhibitors block oxidation–reduction reactions at specific locations along the ETC.

Inhibition of the electron flow by these agents creates **crossover points**. The electron carriers present before the block become reduced and those after the block become oxidized. Since absorption spectra of the oxidized and the reduced forms of electron carriers differ from each other, these changes can be detected by spectrophotometer.

II. Oxidative Phosphorylation

There are several hypotheses proposed to explain how energy released during flow of electrons along the respiratory chain is used for ATP generation.

A. Hypotheses for ATP Generation

Chemical Coupling Hypothesis

This hypothesis, proposed by *Edward Stater* (1953), postulates that the energy released from ETC causes formation of high energy covalent intermediates. These intermediates are subsequently cleaved to release their energy content, which is used for the synthesis of ATP.

Conformational Coupling Hypothesis

The conformational coupling hypothesis put forth by *Paul Boyer* (1964), proposes that the energy of electron transport is used for altering conformation of certain proteins that are located in the IMM. The proteins with altered conformation have high energy content, which is subsequently used for ATP generation.

Chemiosmotic Hypothesis

Proposed by the British biochemist *Peter Mitchell* (1961), this hypothesis is widely accepted. According to the hypothesis, oxidative phosphorylation occurs in two steps:

- Generation of electrochemical gradient across the IMM (***Step I***).
- Utilization of this gradient to fuel ATP synthesis (***Step II***).

Step I

The electrochemical gradient is built by pumping of protons out of the mitochondrial matrix. The *proton pumping is fueled by the energy released by exergonic redox reactions of ETC* (Fig. 14.10). This creates a proton gradient of approximately 1.4 pH units across the IMM (inside alkaline). It also builds a membrane potential of 100–200 mV_1 (inside negative). The concentration and electrical potential add up to a steep electrochemical gradient for protons.

Step II

The electrochemical gradient is a potential source of energy (4–6 kcal/mol) for generation of ATP. The enzyme responsible for utilizing this gradient for ATP synthesis is known as *ATP synthase.* It is also known as the *ATPase*, because the isolated enzyme is capable of catalyzing hydrolysis of ATP to ADP and inorganic phosphate.

$$ATP \xrightarrow{ATPase\ activity} ATP\text{-}P_i$$

Figure 14.11 gives a magnified view of the *ATP synthase*, which consists of two components: F_0 and F_1. The F_0 is embedded in the IMM, whereas the F_1 protrudes into the mitochondrial matrix (Fig. 14.11). **F_1 unit** (F_1 stands for "coupling factor 1") is a protein complex of subunit structure $\delta_3\beta_3\gamma\delta$ and a molecular weight of

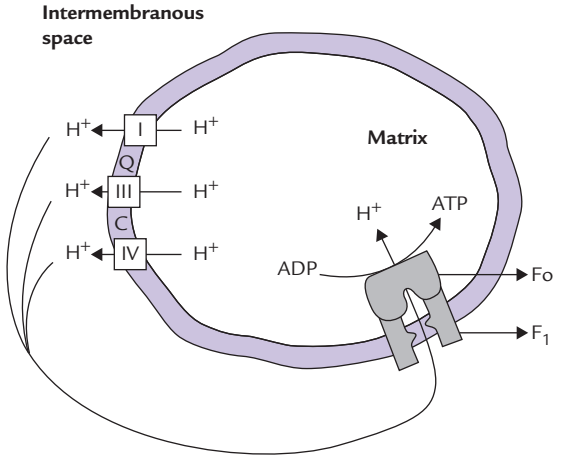


Fig. 14.10. Protons are translocated from the mitochondrial matrix to intermembranous space, fueled by the redox reactions of the respiratory chain in **step I.** The three complexes shown (I, II, IV) are actually sites of proton pumping (Q = ubiquinone, C = cytochrome c).

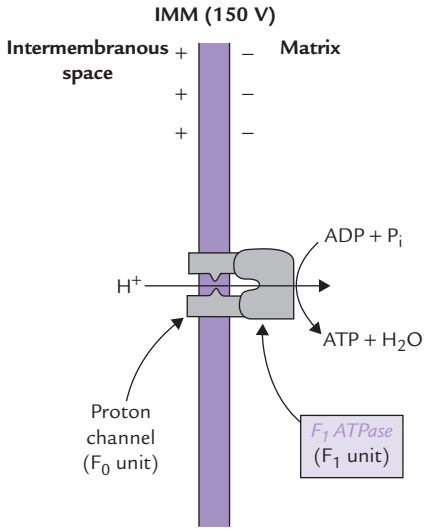


Fig. 14.11. Protons are allowed to flow back into the mitochondrion through a specific proton channel in F_0. This proton channel is coupled to an ATP synthesizing enzyme (the F_1 unit).

380 dDa. It can be visualized in the electron microscope as small buttons on the inner surface of the inner mitochondrial membrane. The F_1 unit is attached to the **F_0 unit**, an integral membrane protein having a pore called *proton channel*.

The proton channel is crucial for ATP generation. This is because the IMM is impermeable to protons, and so the extruded protons in the intermembrane space can re-enter the mitochondrial matrix through this proton channel. As these protons move inwards, the energy inherent in the electrochemical gradient is liberated as concentrated packets. These energy packets are readily used by the F_1-unit to fuel ATP synthesis.

The two major components of *ATP synthase* are F_1 (ATP synthesizing) attached to F_0, which is proton channel spanning the IMM. In mitochondria this complex uses the energy released by electron transport to drive ATP synthesis, but in isolation it hydrolyzes ATP (*ATPase* activity).

Action of the *ATP synthase* is inhibited by an antibiotic, **oligomycin**. The latter binds with F_0 portion of this enzyme and closes the proton channel. As a result, reentry of protons is blocked and the energy inherent in the electrochemical gradient cannot be used for ATP generation. As more and more protons are actively pumped out, the proton concentration in the intermembrane space rises. It becomes difficult to pump protons against the rising gradient and finally electron transport also stops. At this stage, the *ATP synthase* is turned off.

Success of chemiosmotic hypothesis: The chemiosmotic hypothesis has been widely accepted because it can provide explanations for the following phenomena:

1. An intact membrane is required for ATP synthesis by oxidative phosphorylation.
2. Respiring mitochondria generates a proton gradient.
3. Certain compounds (such as DNP) stop ATP synthesis without inhibiting electron transport from NADH, or succinate, to oxygen. These compounds prevent building up of the proton gradient, which stops/halts synthesis of ATP. Thus, they act as uncouplers of oxidative phosphorylation (discussed later).

The exergonic reactions of the electron transport chain cause translocation of protons out of the mitochondrial matrix to form an electrochemical gradient whose free energy drives ATP synthesis.

B. ATP Production and P : O Ratio

It is well established experimentally that a pair of electrons originating in NADH generates three molecules of ATP via the oxidative phosphorylation pathway, and that from $FADH_2$ generates two. **It is said that the P : O ratio is three for electrons from NADH and two for those from $FADH_2$**. P referring to a high-energy phosphate bond being synthesized and O referring to an atom of oxygen being reduced. Thus ***P : O ratio is*** defined as the number of ATPs *generated for each oxygen consumed*. Refer to **Box 14.1** for more information on relationship between ATP generation and oxygen consumption.

Sites of ATP Production

There are three energy conserving segments that release sufficient amount of energy to result in synthesis of an ATP molecule each. At each of these segments, called *phosphorylation* sites (indicated by arrows in Fig. 14.5), quantum of energy released exceeds 10 kcal/mole. Because ATP is synthesized only when protons flow, this results in halting ATP synthesis.

1. The **Site I** is between NADH and ubiquinone (specifically during electron flow through the FeS complex).
2. The **Site II** is between ubiquinol and cytochrome c (specifically during electron flow from cytochrome b to cytochrome c_1).
3. The **Site III** is between cytochrome c and oxygen (i.e. the cytochrome oxidation reaction).

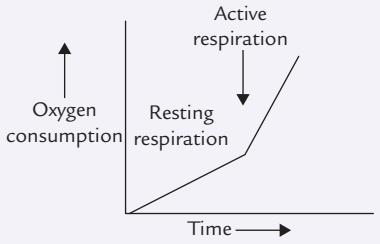
P : O ratio is defined as the number of ATPs synthesized per oxygen reduced, is three with NADH and two with $FADH_2$.

The three sites correspond to Complexes I, III and IV of the ETC. *Since Complex I is bypassed with $FADH_2$, only two ATPs are generated (P : O ratio-2).*

BOX 14.1

Acceptor Control of Respiration

Though ATP generation follows electron transport, the rate of electron flow is dependent on ATP generation itself. Under normal circumstances, as the electron transport proceeds, ADP and phosphate are removed from cytosol and intramitochondrial concentration of ATP builds up concomitantly. When the ADP is depleted from cytosol, the rate of electron flow (measured by oxygen consumption) falls. This is because of the limiting action of ADP on rate of oxygen consumption. This is known as "resting respiration". Conversely, when ADP concentration rises in cytosol (for example, after some energy requiring process which degrades ATP to ADP), phosphorylation of ADP increases. Rate of oxygen consumption increases in this instance because of **ADP-induced oxygen consumption**. This is followed by (coupled) phosphorylation of ADP to ATP. Such dependence of the rate of oxygen consumption on the intracellular concentration of ADP (the phosphate acceptor), is called **acceptor control of respiration.**



In the case of poisoning by certain uncouplers (e.g. pentachlorophenol), ATP generation is halted and concentration of ADP in cytosol rises. Rise in ADP concentration causes increase in oxygen consumption, which implies speeding up of the electron transport activity.

C. Energetics

If one starts from NADH, with its redox potential of –0.32 V, and ends with oxygen, redox potential +0.82 V, then one spans a total of 1.14 V (Fig. 14.5). As discussed earlier, this comes to a theoretical availability of about 52.6 kcal/mole (equation iv). Because three ATP molecules are synthesized for each pair of electrons channeled to oxygen, oxidative phosphorylation recovers 21.9 kcal (each ATP = 7.3 kcal) of the 52.6 kcal available as ATP, for an approximate efficiency of 48%.

$$\text{Efficiency} = \frac{21.9}{52.6} \times 100 = 48\%$$

The remaining 52% of energy is lost as heat.

Uncouplers: Being linked through a proton gradient, the oxidation and the phosphorylation are said to be coupled processes. They can be uncoupled from each other by certain compounds, called the *uncouplers*. Primary action of these compounds is to **increase permeability** of the inner mitochondrial membrane to protons. As a result, relatively free movement of protons across the IMM occurs, which prevents building of the electrochemical gradient. Since this gradient is essential for ATP generation, failure to build it stops ATP generation.

An example of uncoupler is **2,4-dinitrophenol** (Fig. 14.12). The compound can freely move across the IMM due to its lipophilic nature. It can carry a proton along with itself. As a result, free movement of protons across the IMM occurs,

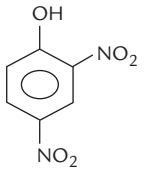


Fig. 14.12. Structure of 2,4 dinitrophenol, a highly diffusible chemical that uncouples oxidative phosphorylation by transporting protons across the IMM.

preventing the building up of proton gradient. Though an uncoupler impedes ATP generation, it has no effect on electron transport. The latter process continues as before, or is rather enhanced due to *acceptor control* of respiration, discussed later. The energy released by electron flow is mostly dissipated (i.e. entropy). Some portion of it generates heat: this explains *thermogenic effect of some uncouplers* **(Case 14.1).**

The other uncouplers include pentachlorophenol, dinitrocresol, and trifluorocarbonylcyanide. The latter dissipates the proton gradient 100 times faster than DNP.

The uncouplers allow electrons transport to proceed without ATP synthesis. They uncouple by carrying H^+ ions across the IMM and hence dissipate the proton (electrochemical) gradient.

Physiological uncouplers and ionophores: Some endogenous compounds, including ***free fatty acids, bilirubin*** and possibly **thyroxine**, also can act as uncouplers at concentrations well above the physiological range. They

are called *physiological uncouplers*. In case of bilirubin, which deposits in the basal ganglia of infants with severe hyperbilirubinaemia (see Chapter 16), a role of the uncoupling effect in the resulting brain damage is possible.

Valinomycin is a transport antibiotic or an ionophore that makes the inner mitochondrial membrane permeable to potassium (the term *ionophore* refers to compounds that promote the transport of ions across biological membranes; Chapter 7). It also acts as an uncoupler.

This is because the translocation of potassium ions dissipates the membrane potential, which is an essential component of the proton-motive force.

Significance of Uncoupling

Uncoupling under natural conditions plays an important role in hibernating animals, and in animals of polar regions, for maintaining their body temperature. These animals contain **brown adipose tissue** which is specialized to carry out oxidation uncoupled from phosphorylation. When fats are oxidized in the brown adipose tissue, the energy liberated is not trapped as ATP but rather released as heat.

Presence of active brown adipose in some individuals is believed to burn excess calories, thereby preventing them from gaining weight even after indulging in overeating.

III. Mitochondrial Membrane Transporters

The carrier proteins in IMM, also known as mitochondrial membrane transporters, are highly specific in their action. They permit movements of several molecules that are important for mitochondrial function, such as ATP, H^+, pyruvate, malate, citrate, etc. A few transport proteins (Fig. 14.13) and their actions are described here.

A. *Adenine Nucleotide Translocase (ANT)*

The *adenine nucleotide translocase* permits inward movement of ADP into the mitochondrial matrix, and a simultaneous outward movement of ATP into the intermembranous space. These movements are important because ADP (and phosphate) have to enter the mitochondrion as substrates for oxidative phosphorylation, and ATP has to pass from the mitochondrion to the cytosol, where it is used for energy-dependent processes. Structurally, ANT is an integral membrane protein that extends across the IMM.

Atractyloside is toxic glycoside which specifically inhibits the ANT system. Inhibition of ANT hampers the transport of ATP, which leads to serious consequences since ATP is required to drive a number of cellular activities.

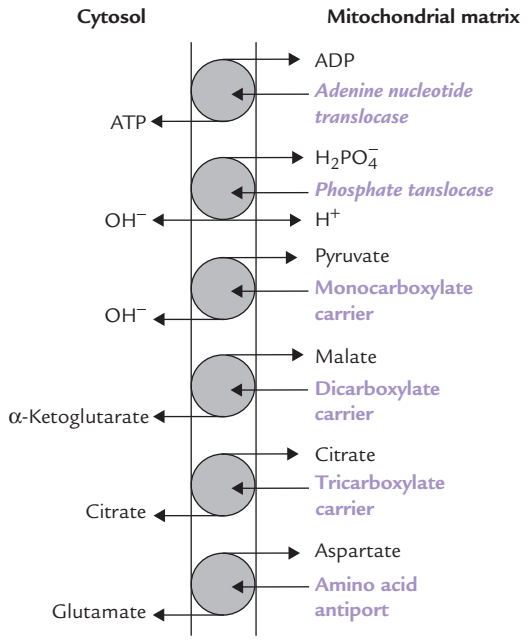


Fig. 14.13. Energy of electrochemical gradient being used for ATP generation.

B. *Phosphate Translocase System*

The *phosphate translocase* system is functionally related to the *adenine nucleotide translocase*. It cotransports a phosphate ion along with H^+ into the mitochondrial matrix (Fig. 14.13). Phosphate is then used for the generation of ATP from ADP. Thus, a combined action of these two *translocase* systems permits entry of phosphate and ADP into the mitochondrial matrix, where they combine to form ATP.

C. Monocarboxylate Carrier

This carrier transports pyruvate produced in cytosol, mainly through the glycolytic sequence, into mitochondria where it is further metabolized.

D. Dicarboxylate Carrier

It moves malate from site of its production (i.e. the mitochondrial matrix) into cytosol.

E. Tricarboxylate Carrier

It transports citrate, the first intermediate of TCA cycle, into cytosol. This carrier is very specific: it does not transport even a closely related molecule, i.e. isocitrate.

Combined action of the dicarboxylate and tricarboxylate carriers play vital role in lipogenesis (Chapter 11).

Specific transporters in IMM mediate the transmembrane movements of ADP, ATP, Pi and several metabolites.

F. Shuttle Systems

Certain shuttle systems are known to operate in mitochondria that **transport reducing equivalents from cytosol into mitochondria**. The reducing equivalents (in the form of NADH) are generated in the cytosol (during glycolysis). They have to be transported to mitochondrial matrix for oxidation and generation of ATP. However, the IMM is impermeable to NADH. Therefore, specific shuttle systems, namely **malate aspartate shuttle** and **glycerol phosphate shuttle**, accomplish their transport (Fig. 14.14).

Shuttle systems move the NADH produced in the cytosol into the mitochondrion.

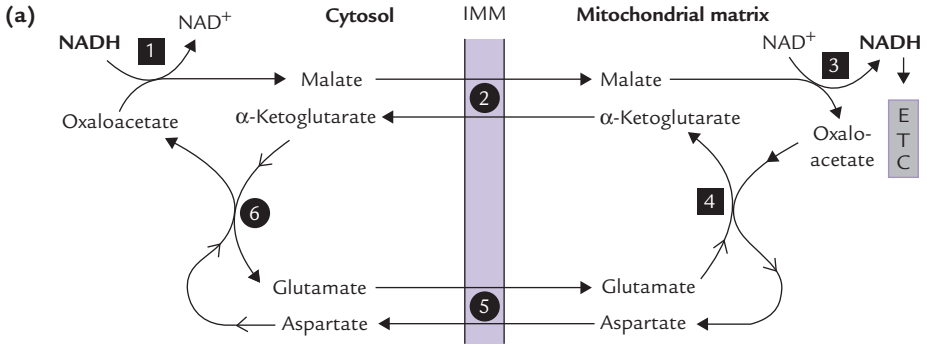
Malate aspartate shuttle: It operates in liver and heart muscles in the following steps:

1. The reducing equivalents are transferred from NADH to oxaloacetate to form malate. The reaction is catalyzed by the cytosolic enzyme ***malate dehydrogenase***.
2. Malate is then transported across the IMM by the dicarboxylate carrier.
3. Within the mitochondrial matrix, the reducing equivalents are transferred from malate to NAD^+ to form NADH. The enzyme that catalyzes this reaction is ***mitochondrial malate dehydrogenases***.

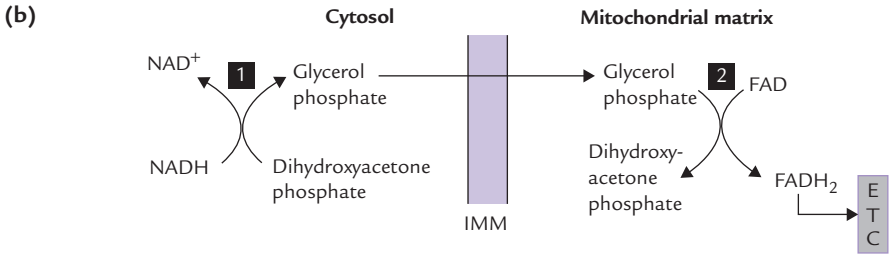
Thus, by the concerted action of the dicarboxylate carrier and two enzymes (i.e. *cytosolic malate dehydrogenase* and *mitochondrial malate dehydrogenase*), NADH reaches the mitochondrial matrix (Fig. 14.14a). The mitochondrial NADH now contains the reducing equivalents that were originally present in the cytosolic NADH.

Rest of the shuttle systems is concerned with transport of the oxaloacetate back into the cytosol. It requires participation of *amino acid antiport system* which exchanges aspartate for glutamate across the IMM.

Glycerol phosphate shuttle: In skeletal muscle and brain, another type of shuttle, called glycerol phosphate shuttle, is present. It transfers the hydrogen first to dihydroxyacetone phosphate forming glycerol phosphate (in cytosol), and then to the FAD prosthetic group of the *mitochondrial glycerol phosphate dehydrogenase* (Fig. 14.14b). The latter is an integral protein of the inner mitochondrial membrane, which regenerates its FAD by direct transfer of electrons to the respiratory chain. Only



1. *Cytosolic malate dehydrogenase*
2. Dicarboxylate carrier
3. *Mitochondrial malate dehydrogenase*
4. *Transaminase*
5. Amino acid antiport
6. *Transaminase*



1. *Cytosolic glycerol phosphate dehydrogenase*
2. *Mitochondrial glycerol phosphate dehydrogenase*

Fig. 14.14. Transfer of reducing equivalents from cytosol into mitochondrial matrix. **(a)** Malate aspartate shuttle, **(b)** Glycerol phosphate shuttle (IMM = mitochondrial matrix).

two ATPs are therefore produced for each pair of reducing equivalents transferred.

Malate aspartate shuttle (generates three ATPs) is more important in liver and heart; and the glycerol phosphate shuttle (generates two ATPs) is more important in muscle and brain.

IV. Enzymes Participating in Biological Oxidation

The enzymes involved in biological oxidation catalyze oxidation-reduction reactions, hence they belong to the class I, i.e. *oxidoreductases* (Chapter 6). They are subclassified as: (a) *dehydrogenases,* (b) *oxidases*, and (c) *hydroperoxidases*.

A. Dehydrogenases

They are most common of all *oxidoreductases*. They catalyze transfer of hydrogen between a substrate and a coenzyme, most commonly NAD^+, FAD, $NADP^+$ or FMN. These enzymes are named according to the substrate from which they remove the hydrogen (Table 14.1).

B. Oxidases

They catalyze transfer of hydrogen removed from a substrate to oxygen. Mostly this results in formation of water. *Cytochrome oxidase*, the terminal component of ETC, is an example.

C. Hydroperoxidases

They use hydrogen peroxide or an inorganic peroxide as a substrate. They are needed for degradation of hydrogen peroxide that may be produced in the reactions catalyzed by the *dehydrogenases*. Technically, ***catalase*** is also a *peroxidase*: it degrades hydrogen peroxide to molecular oxygen and water.

$$H_2O_2 \rightarrow H_2O + O_2$$

V. Mitochondrial Myopathies

Mitochondrial DNA consists of a circular DNA molecule of 16,569 base pairs which encodes 13 polypeptides, 22 tRNAs, and the RNAs of mitochondrial ribosomes (12S and 16S). The mtDNA is **maternally inherited** since mitochondria from sperm does not enter the ovum during fertilization. The polypeptide chains encoded by mtDNA include seven subunits of *NADH-Q reductase*, one subunit of *QH_2-cytochrome c reductase*, three subunits of *cytochrome oxidase* and two subunits of *ATP synthase*. All other mitochondrial proteins are encoded by nuclear genes and synthesized in cytoplasmic ribosomes.

Mitochondrial DNA is 10 times *more susceptible to mutations* than nuclear DNA. Some of these mutations result in absence of specific polypeptide chains. Faulty operation of the electron transport pathway results and the ATP production is decreased. These disorders are known as *mitochondrial myopathies* and they most commonly affect the tissues with high rate of oxidative phosphorylation. The mitochondrial myopathies are tissue specific; some affect the heart, others the skeletal muscles. Many are accompanied by lactic acidosis, because the inability to reduce NADH normally results in channeling of pyruvate toward lactic acid production (Chapter 9).

Leber's hereditary optic neuropathy (LHON) is a mitochondrial myopathy, characterized by sudden onset of blindness in adults, associated with degeneration of the optic nerve.

Mutations in mitochondrial DNA cause inherited disorders of oxidative phosphorylation, e.g. Leber's hereditary optic neuropathy.

Exercises

Essay type questions

1. What is redox potential and what is its significance? Describe various components of electron transport chain and discuss the oxidation of NADH and $FADH_2$.
2. What do you understand by the term P : O ratio? Indicate diagrammatically the sites of ATP generation in the mitochondrial respiratory chain and their inhibitors.
3. What is oxidative phosphorylation? Discuss in detail the chemiosmotic hypothesis of ATP synthesis and explain how the process of oxidative phosphorylation is arrested by various inhibitors.
4. What are uncouplers of biological oxidation? State their mechanism of action.

Write short notes on

1. Uncouplers
2. P : O ratio
3. Redox potential
4. ATP synthesis
5. Physiological uncouplers
6. Brown adipose
7. Ubiquinone
8. Cytochrome aa_3
9. Chemiosmotic theory
10. Malate aspartate shuttle

CLINICAL CASE

CASE 14.1 Labourer exposed to a toxic wood preservative

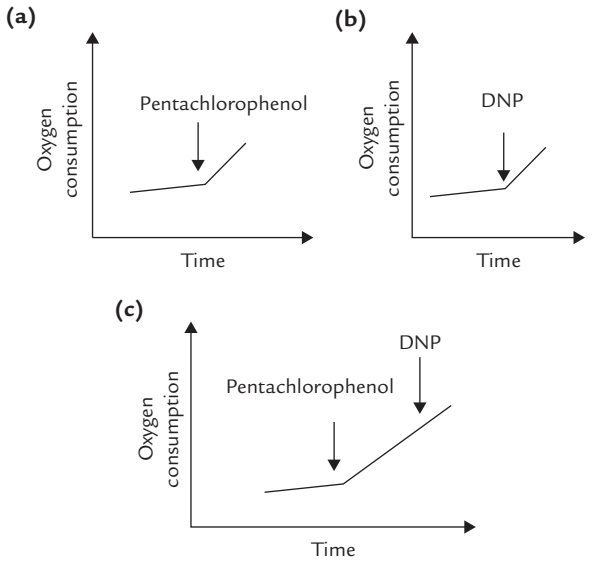
A labourer felt dizzy and feverish while spraying pentachlorophenol, a popular wood preservative that kills termites, fungi and bacteria. While being taken to the hospital he lost consciousness on the way. He was admitted in the hospital emergency. On examination, his body temperature was 39.9°C, and respiratory rate was 60 per minute. He died in a comatose state, after two days.

On autopsy, the subcutaneous and the omental fats were markedly depleted. There was pulmonary congestion, and the marrow examination showed evidence of enhanced erythropoiesis.

Exposure to the wood preservative was thought to be responsible for the patient's condition. Effect of the chemical was investigated by performing the following tests on a preparation of rat liver mitochondria. The mitochondria were suspended in a suitable buffer, to which ADP and succinate were added. The preparation was incubated in the oxygen electrode chamber and the rate of oxygen consumption was measured. Addition of pentachlorophenol to this preparation caused a marked increase in the oxygen consumption (Experiment a). A similar increase was observed following addition of 2, 4 dinitrophenol (Experiment b) to this preparation. However, when the dinitrophenol (DNP) was added following the addition of pentachlorophenol, no further increase in oxygen consumption resulted (Experiment c).

Experiment a–c. Effect of the pentachlorophenol and dinitrophenol (DNP) on oxygen consumption by rat liver mitochondria.

Note: Rate of electron flow is reflected by O_2 consumption rate.



Q.1. State the possible mechanism of action of the uncoupling effect of the pentachlorophenol.

Q.2. Interpret the test results and elucidate the action of pentachlorophenol on mitochondrial respiration.

Q.3. Explain the clinical and the autopsy findings of this patient in molecular terms.

CHAPTER

15

INTEGRATION OF METABOLISM

Anabolic and catabolic pathways for all major biomolecules such as carbohydrates, lipids and proteins are intricately linked and well coordinated. There are several **crossroads** among metabolic pathways, which help in the flow of key metabolites. While metabolic pathways of biomolecules are already discussed in previous chapters (9 to 13), this chapter deals with the metabolic interrelationships and the flow of metabolites from one pathway to another. These aspects are described under three metabolic states: **fed state** (high insulin : glucagon ratio), **starvation** and **diabetes** (low insulin: glucagon ratio). A special emphasis is given on the role of hormones in regulating metabolic activities as per the cellular requirements.

After going through this chapter, the student should be able to understand:

- Regulation of mainline metabolic pathways by hormones.
- Metabolic interconnections and tissue specific metabolism.
- Metabolic adaptations to effect variations in fuel metabolism in short-span fasting, prolonged fasting and starvation.
- Diabetes mellitus: metabolic alterations, biochemical basis of diabetic ketoacidosis and chronic complications of diabetes.
- Metabolism of xenobiotics: reactions and role of P-450 cytochromes.

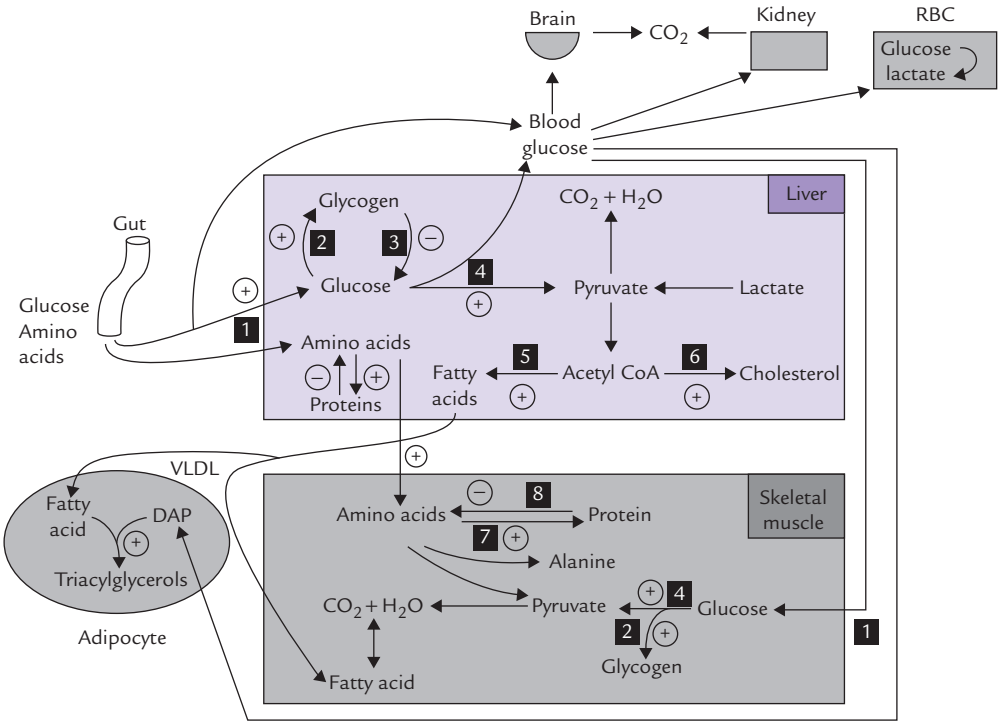
I. Hormonal Regulation of Major Metabolic Pathways

Insulin is the principal hormone which plays important role in regulation of various metabolic pathways. Glucagon, epinephrine, and the glucocorticoids are the other hormones which help in the integration of metabolic processes.

A. Insulin

Effect of insulin reaches peak in the fed state when blood glucose and serum insulin levels are high and glucagon levels are low. The high insulin : glucagon ratio is responsible for various metabolic adaptations that occur in the fed state. The major events in liver, skeletal muscles, adipose tissue, erythrocytes and other tissues have been depicted in Figure 15.1.

Insulin enhances *uptake of glucose* by peripheral tissues including skeletal muscles, by promoting carrier-mediated transport. Intracellularly, the glucose is channeled into **glycogenesis** because insulin-initiated cascade activates *glycogen synthase* by promoting conversion of the phosphorylated form of the enzyme to the dephosphorylated form. The same mechanism causes negative modulation of activity of *glycogen phosphorylase*, resulting in a concomitant inhibition of **glycogenolysis.** Insulin increases the transcription of *glucokinase gene* which, in turn causes increase in glycogen synthesis in liver. **Glycolysis** is also stimulated due to activation of the key regulatory enzymes of this pathway (e.g. *phosphofructokinase-1, pyruvate kinase*) by insulin.



Pathway	1 Glucose uptake	2 Glycogenesis	3 Glycogenolysis	4 Glycolysis	5 Lipogenesis	6 Cholesterol synthesis	7 Protein synthesis	8 Protolysis
Effect	↑	↑	↓	↑	↑	↑	↑	↓

Fig. 15.1. Metabolic adaptations in fed state brought about by insulin (actually, high insulin: glucagon ratio) (DAP = dihydroxyacetone phosphate, VLDL = very low density lipoprotein. ↑ = stimulates; ↓ = inhibits).

Normally the glycolytic sequence is followed by the action of ***pyruvate dehydrogenase*** (PDH) which yields acetyl CoA: the latter serves as substrate for lipids in the adipose and other tissues. *Insulin enhances the activity of PDH*, thus causing an increase in the production of acetyl CoA. Moreover, insulin influences various intercellular events which divert the excessive acetyl CoA into **lipid synthesis**, especially in adipocytes (Chapter 11). Activity of *HMG CoA reductase*, the principal enzyme of **cholesterol biosynthesis**, is also enhanced by insulin, therefore causing an increase in the rate of cholesterol biosynthesis (Fig. 15.1).

Insulin increases the **protein synthesis** and inhibits **proteolysis**. The action is exerted at different sites:

- Uptake of amino acids by tissues is promoted by insulin that facilitates the transport across the plasma membrane.
- Insulin enhances transcription, which is followed by increased translation.
- Increased synthesis of amino acyl-tRNA accounts for increased translation.

In muscles, insulin causes an increase in the number of ribosomes as well as their translational efficiency.

Synthesis of some proteins such as *glucokinase* is specially stimulated by insulin.

Insulin influences activities of a number of other enzymes as well, which results in *inhibition of gluconeogenesis, lipolysis* and *glycogenolysis* (Table 15.1).

B. Glucagon

Decreased blood glucose concentration is a potent stimulus for glucagon release from α-cells of pancreas. The *major target for glucagon is liver*; it is not able to act on the muscles because the muscle cells lack glucagon receptors. Glucagon stimulates **glycogenolysis** and inhibits **glycogenesis** in liver (Table 15.2). This is because glucagon converts the regulatory enzymes of these pathways from dephosphorylated form to phosphorylated form; these effects are mediated via cAMP (Chapter 9). Net result is increased production of glucose 6-phosphate which is converted to glucose (by *glucose 6-phosphatase*) and released into blood circulation.

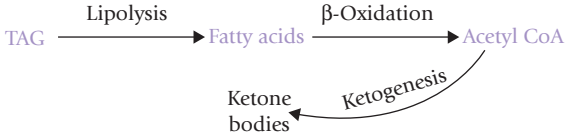
Triacylglycerol degradation (**Lipolysis**) is enhanced by glucagon because of the stimulation of activity of *hormone sensitive lipase*. This provides fatty acid substrate

Table 15.1. Effect of insulin on major metabolic pathways

Pathway	Overall effect	Regulatory enzyme affected
Carbohydrate metabolism		
Glycogenesis	↑	*Glycogen synthase* ↑ *Glucokinase* ↑
Glycolysis	↑	*Phosphofructokinase-1* ↑ *Pyruvate kinase* ↑ **Pyruvate dehydrogenase* ↑
Glycogenolysis	↓	*Glycogen phosphorylase* ↓
Gluconeogenesis	↓	*Pyruvate carboxylase* ↓ *PEP carboxykinase* ↓ *Fructose 1, 6-bisphosphatase* ↓ *Glucose 6-phosphatase* ↓
Lipid metabolism		
Lipogenesis	↑	*Acetyl CoA carboxylase* ↑ *Fatty acid synthase* ↑
Lipolysis	↓	*Hormone sensitive lipase* ↓
Cholesterogenesis	↑	*HMG CoA reductase* ↑
TAG synthesis	↑	*Fatty acyl CoA-glycerol 3-phosphate transferase* ↑
Lipoprotein degradation	↑	*Lipoprotein lipase* ↑
Amino-acid metabolism		
Transamination	↑	*Transaminases* ↑

PEP = phosphoenol pyruvate, HMG CoA = β-hydroxy-β-methylglutaryl CoA. ↑ = stimulates, ↓ = inhibits. **Pyruvate dehydrogenase* is neither a glycolytic nor a TCA cycle enzyme; it acts as a bridge between the two.

molecules for β-oxidation, and hence turns this pathway on. Production of acetyl CoA, the end product of β-oxidative pathway, is thereby greatly enhanced; the excessive amount of this compound is then channelled into ketone body formation.



Protein synthesis is decreased by glucagon, but protein and amino acid degradation continues as before (or increases in liver) for producing energy and for providing substrates for glucose production (i.e. **gluconeogenesis**).

With high glucagon : insulin ratio (during fasting), gluconeogenesis in liver speeds up to maintain normal blood glucose levels; fatty acids and ketone bodies serve as alternate substrates to meet the bulk of energy requirement.

Table 15.2. Effects of hormones on intermediary metabolic pathways

Pathway	Glucagon	Epinephrine	Glucocorticoids
Glycogenolysis	↑[a]	↑	
Glycogenesis	↓[a]	↓	↑[a]
Glycolysis			
Liver	↓[a]	↓	↓
Muscle		↑	
Gluconeogenesis	↑[a]	↑	↑
Lipolysis	↑[a]	↑	↑
Lipogenesis	↓[a]	↓	
Cholesterol biosynthesis	↓[a]	↓	
Protein biosynthesis	↓[a]	↓	
Protein degradation	↑[a]	↑	↑[b]

[a]Action in liver, [b]action only in muscles, ↑ = stimulates, ↓ = inhibits.

C. Epinephrine

Effect of epinephrine in the liver is same as that of glucagon. In addition, epinephrine is the *principal hormone that regulates muscle metabolism*.

Metabolism in muscle is distinct due to lack of the enzyme *glucose 6-phosphatase*. Glucose 6-phosphate is generated from glycogen due to stimulation of **glycogenolysis** by epinephrine; however, *glucose 6-phosphatase* cannot be converted to glucose in muscle due to the absence of *glucose 6-phosphatase*. Therefore, glucose 6-phosphate is channeled into glycolysis. Thus, muscle glycogen does not play any role in regulation of blood glucose; this is in contrast to liver glycogen, primary role of which is to liberate free glucose into blood circulation. Furthermore, activity of the glycolytic enzymes in muscle cells is enhanced by epinephrine (Table 15.2), causing an increase in glucose utilization.

During severe muscular exercise, epinephrine promotes muscle glycogen degradation to generate glucose 6-phosphate which is converted to lactate via **anaerobic glycolysis**. Lactate reaches liver where it is converted to glucose by gluconeogenesis. The concerted action of muscle and liver called **Cori cycle**, thereby ensure adequate supply of glucose fuel (Chapter 9).

Biological effects elicited by glucagon and epinephrine are similar in liver. Epinephrine has additional effects in muscle also, e.g. increased glycogenolysis, anaerobic glycolysis and protein degradation.

D. Testosterone

It is male steroid hormone which has potent anabolic effect. Its principal action is elicited on muscle cells, where it stimulates protein accumulation. The same effect is observed in other tissues as well, though the rate is lower.

E. Glucocorticoids

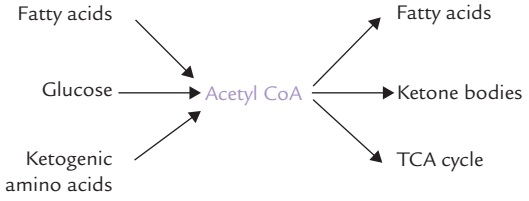
These hormones are liberated from adrenal cortex, more so in stressful state. Glucocorticoids increase protein catabolism in tissues (except liver; Table 15.2) and decrease transport of amino acids into the peripheral cells. The amino acids liberated from these tissues enter liver cells where they are channeled into the gluconeogenic sequence. Thus, the hepatic glucose generation increases, and therefore, glucocorticoids are said to act in concert with glucagon. In addition, the glucocorticoids step up lipolysis by stimulating activity of the *hormone sensitive lipase*.

Effect of glucocorticoids is elicited slower than that of insulin, glucagon or epinephrine.

Insulin is a satiety hormone, glucagon up-regulates blood glucose level, epinephrine mediates flight-or-fight response, and glucocorticoids are released in chronic stress.

II. Metabolic Interconnections and Organ Specialization

The metabolic processes involving various biomolecules are interconnected through certain **branch compounds** which lie at junctions of the major metabolic pathways (see Fig. 8.3). *Some important branch compounds are acetyl CoA, pyruvate* and *glucose 6-phosphate*. Acetyl CoA is the common degradation product of glucose, fatty acids, and ketogenic amino acids; its acetyl group can be used to synthesize fatty acids or ketone bodies, or can be oxidized via the TCA cycle and oxidative phosphorylation. Interconnections between major pathways via such branch compounds make interconversions of major foodstuffs possible. For example, pyruvate, the degradation product of carbohydrates and glycogenic amino acids can be converted to acetyl CoA, thereby providing its carbons for biosynthesis of cholesterol, fatty acids or other compound lipids.



Interrelationships and coordination of various organs is of crucial significance in maintaining metabolic homeostasis. Metabolic flow between these organs, each of which has a specialized role, occurs in well defined pathways. For example, following a meal, glucose, amino acids and fatty acids are directly available from the intestine. Later, when these fuels are exhausted, the liver and adipose tissue supply various organs with fuel molecules: liver provides glucose and ketone bodies and adipose tissue provides fatty acids. All these organs are connected via the blood stream.

Liver maintains proper levels of circulating fuels for use by other tissues, mainly brain and muscles. It is referred to as the *body's central metabolic clearing house*. It is uniquely suited to carry out this task because it is the first organ exposed to all nutrients (except fatty acids) absorbed from intestine, and to insulin and glucagon from pancreas. One of the liver's major functions is to act as a **blood-glucose buffer**; it imports and stores glucose following absorption of digested food; and generates and exports glucose into circulation when body needs it. Liver contains *glucose 6-phosphatase* which enables it to release glucose. Other important functions performed by liver are:

- It is active in cholesterogenesis and production of bile acids and triacylglycerols.
- It exports triacylglycerols and cholesterol as VLDL and takes up HDL (reverse cholesterol transport).
- Liver synthesizes proteins after meals and liver protein is degraded to provide precursors for gluconeogenesis. The major proteins synthesized in liver include albumin, most of the transport proteins and blood clotting factors. Liver is also the site of production of urea.
- It generates, but cannot utilize ketone bodies. It is responsive to insulin except for glucose transport.

Liver releases 180 g of glucose, 100 g of fat and 14 g of albumin into circulation daily. Evidently, liver has more metabolic versatility than any other organ (Table 15.3).

III. Metabolic Adaptations in Three Fasting–starvation States

Starvation refers to a state of severe and prolonged (> 10 days) food deprivation while **fasting** implies voluntary restrain from food. Elaborate biochemical mechanisms exist, as described by ***Cahil*** and his associates, which ensure maintenance of normal blood glucose level (Chapter 10) and adequate supply of energy during fasting period. These mechanisms are necessary for survival considering the fact that the body needs energy continually, whereas

Table 15.3. Organ metabolism

Organ	Properties
Brain	Constant and voracious user of fuels: uses 200 g of glucose per day and 20% of total ATP produced in resting adult. Glucose is used as exclusive substrate, but can use ketone bodies after few days of fasting.
Erythrocyte	Glucose is only substrate for metabolism. Lacks mitochondria and releases lactate into circulation.
Kidney	Active in gluconeogenesis, using glutamine. Produces ammonia from glutamine.
Adipose tissue	Stores or hydrolyzes triacylglycerol stores as per body requirement.
Muscle	Stores glycogen, but cannot release glucose into circulation because of lack of *glucose 6-phosphatase*. Can perform anaerobic glycolysis, uses fatty acids and ketone bodies for energy. Muscle proteins release amino acids (mainly glutamine and alanine) for gluconeogenesis during fasting and starvation.
Intestine	Releases digested carbohydrates into portal vein. Releases branched-chain amino acids and other essential amino acids, but little aspartate, asparagine, or glutamine. Forms chylomicrons for export of dietary cholesterol and triacylglycerols. Uses glutamine for energy production.

the food is taken intermittently. Fasting is divided in three broad stages:

1. ***Initial stage*** that lasts up to two days of low calorie intake.
2. ***Intermediate stage***, that lasts up to 24 days.
3. ***Advanced stage***, which begins after 24 days.

Body stores of energy: The main providers of energy in the body are hepatic and muscle **glycogen**, adipose tissue **triglycerides** and muscle **proteins** (Table 15.4). The body begins to derive energy from these sources after about 80 minutes of fasting. Till that time, energy for the basal metabolic requirement is met by the fuels available from plasma.

Liver glycogen is the first main provider of energy during a short-term fast. Liver uses this stored glycogen as a readymade source of free glucose. However, this is useful only to a limited extent as the glycogen is exhausted in about 36 hours only. Adipose tissue triglycerides and muscle proteins have much larger reservoirs and can provide energy for 1–3 months. In obese individuals, this duration is expected to be longer because the stores are up to three times as large.

Thus, an obese individual is adapted to survive for three times longer period during starvation. However, the actual duration of survival is shorter because the associated vitamin and mineral deficiencies and infections cause death much before the energy stores are exhausted. Moreover, only about 50% of the body proteins can be lost without any serious risk to body functions. Loss of lean body tissues such as respiratory muscles causes respiratory failure and death.

Initial Stage (Early Fasting)

In initial stages of fasting the metabolic changes are aimed at: (i) maintaining the blood glucose level within normal range and (ii) providing energy from alternate sources

Table 15.4. Energy reserves of a man weighing 70 kg

Provider of energy	Grams	Kilocalories
Plasma **glucose**	12	48
Hepatic **glycogen**	70	280*
Muscle **glycogen**	280	1120
Adipose **triglycerides**	13,000	117,000
Muscle proteins	10,000	40,000

*After a meal. Liver glycogen is 20% to 30% of this value after an overnight fast.

(fatty acids and ketone bodies). To maintain blood glucose level, the glycogen stores of liver undergo **glycogenolysis,** thereby serving as the major source of blood glucose. Glucose is also generated by **gluconeogenesis** from substrates such as lactate, amino acids and glycerol.

Utilization of alternate substrates starts increasing now (as the glycogen stores are fastly depleting). β-Oxidation of fatty acids is gradually speeded-up; the fatty acids are obtained by lipolysis of triacylglycerol stores of adipose tissue (Fig. 15.2).

Insulin : glucagon ratio falls low in this stage, which promotes hepatic generation of **ketone bodies** which are the readily utilizable metabolic fuels. The ketone bodies can be considered as water-soluble lipids and concentrated energy packets.

It is noteworthy that certain tissues do not use fatty acids and remain dependent on glucose for energy needs. These include red and white blood cells, the retinal cells and adrenal medulla. **Lactic acid** produced by these tissues is reconverted to glucose in liver.

Intermediate Stage

In the intermediate stage of fasting, the glycogen stores are mostly depleted and so cannot serve as source of blood glucose. This may have serious consequences since no fuel molecule is entering the body from outside. The

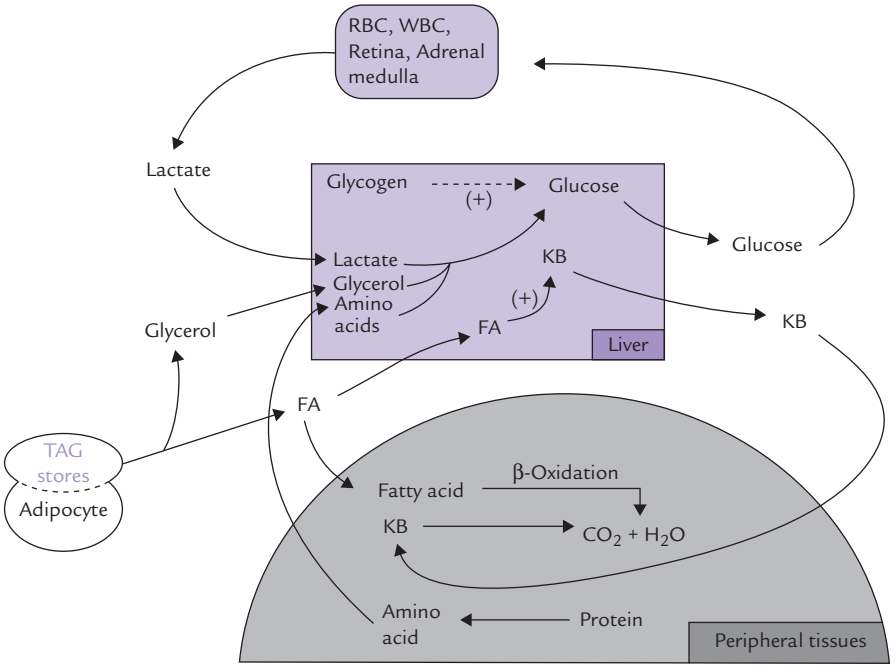


Fig. 15.2. Metabolic and fuel movements during early stages of fasting. Glycogenolysis and glyconeogenesis are speeded up to maintain normal blood glucose level, and ketone body synthesis is increased for use as fuel molecules (FA = fatty acids, KB = ketone bodies).

Table 15.5. Metabolic adaptations in various stages of fasting

Duration	Early stage	Intermediate stage	Late state
Pathways enhanced	Glycogenolysis ↑ Gluconeogenesis ↑ Lipolysis ↑ Ketogenesis ↑ Protein degradation ↑	Gluconeogenesis ↑↑ Lipolysis ↑↑ Ketogenesis ↑↑ Ketone body oxidation ↑↑	Ketone body oxidation By brain ↑↑↑ By other tissues ↓↓↓ Fatty acid utilization ↑↑ **Gluconeogenesis ↑↑**
Pathways slowed		Protein degradation ↓ Protein synthesis ↓	Glycolysis TCA Protein synthesis and degradation Glycogen synthesis and degradation

gluconeogenesis, therefore, assumes added significance for generating glucose and pouring it into blood circulation (Table 15.5).

Fatty acids and ketone bodies continue to serve as sources of energy for muscles, heart and kidneys. Significance of these two fuels increases as the availability of glucose falls. An important change at this stage is that decreased availability of glucose makes nervous tissues use ketone bodies as energy source. Protein synthesis and degradation decline in this stage.

Advanced Stage

As the advanced stage of fasting is reached, ketone bodies are used more and more by brain. Utilization of ketone bodies by other tissues such as muscle, heart and kidneys decreases drastically, so that the available ketone bodies are diverted towards brain. These tissues/organs (other than brain) now heavily depend on fatty acids for their energy needs. A small amount of available glucose is utilized by the brain at this stage. Some tissues, however, continue to depend on glucose, though availability of the latter is very low. These include white and red blood cells, retina and adrenal medulla.

Activity of several regulatory enzymes change during advanced stage of fasting and starvation:

- Activity of *glycogen synthase, phosphofructokinase, pyruvate kinase, pyruvate dehydrogenase* and *glucose 6-phosphatase* decline.
- Enzymes such as *pyruvate carboxylase, fructose 1,6-bisphosphatase, glucose 6-phosphatase* and *PEP carboxykinase* become more active. The result of these alterations

BOX 15.1

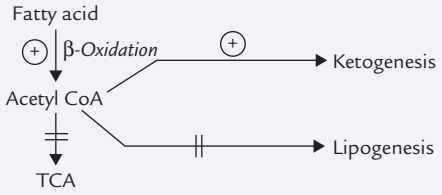
Ketone Body Metabolism in Starvation and Diabetes Mellitus

Low insulin : glucagon ratio in fasting, starvation and untreated diabetes leads to hydrolysis of stored triacylglycerols to release fatty acids. Fatty acids enter liver where they are degraded to acetyl CoA by β-oxidation. The *acetyl CoA cannot enter the TCA cycle or lipogenesis* because of the following metabolic derangements:

1. The NADH produced during β-oxidation converts oxaloacetate to malate, resulting in depletion of oxaloacetate.
2. The low insulin : glucagon ratio enhances gluconeogenesis, as does generation of excess acetyl CoA (acetyl CoA stimulates *pyruvate carboxylase*). Gluconeogenesis further exacerbates the dearth of oxaloacetate so that acetyl CoA is not able to enter the TCA cycle.
3. Fatty acid synthesis (lipogenesis) is decreased due to decreased induction and activity of *acetyl CoA carboxylase* (insulin induces this enzyme) and decreased concentration of malonyl CoA.

Under these circumstances, TCA cycle and lipogenesis are inoperative and the acetyl CoA must find another outlet. Therefore, it is channeled into ketone body production. Excessive amount of ketone bodies are released into circulation resulting in **ketosis.**

Besides substrate availability, enzyme induction is important for ketogenesis. Synthesis of *HMG CoA synthase*—the rate-limiting enzyme of ketogenesis—is enhanced in insulin deficiency. Fatty acids are also powerful inducers of the enzyme synthesis at the level of transcription.



Ketosis has serious consequences because ketone bodies are acidic in nature, and therefore can disturb the acid-base balance of the body resulting in **metabolic acidosis**. Ketone bodies displace bicarbonate ions so that the anion gap increases.

Large amount of ketone bodies are excreted in urine during spells of ketosis, and being anionic in nature they carry sodium, the major extracellular cation. Acidosis causes potassium and magnesium ions to move out of the cell and similarly lost in urine. This results in depletion of these ions in the body. Loss of circulating sodium ions may further aggravate acidosis because loss of Na^+ often results in a decrease in circulatory HCO_3^-. In severe cases, diabetic acidosis, hyperglycaemia and glycosuria induce a severe and prolonged osmotic diuresis, resulting in dehydration.

Dehydration, acidosis and fluid-electrolyte disturbances greatly impair cerebral metabolism and function, leading to coma. The condition is fatal if not promptly treated.

is that *only one metabolic pathway, i.e.* ***liver gluconeogenesis*** *is enhanced in advanced stage of fasting/starvation;* rest of the pathways are suppressed.

Major metabolic pathways are altered in starvation to provide glucose for brain and red blood cells. Fatty acids become the major energy source now, and are oxidized directly or after conversion to ketone bodies.

The above-stated metabolic adaptations are seen in diabetes mellitus (type I) as well. For example, metabolism of ketone bodies in starvation and diabetes mellitus shows marked resemblance (Box 15.1). In fact, *type I diabetes metabolically resembles starvation, although the metabolic consequences are much exaggerated.* Extensive research has been carried out in this field for the past several decades. As a result, the metabolic abnormalities associated with diabetes mellitus are understood better than those of any other disorder, as discussed in the next section.

IV. Diabetes Mellitus

Diabetes mellitus is an endocrine disorder caused by relative or absolute insulin deficiency. It is a major cause of suffering and morbidity, afflicting 20–30 million people of all social classes throughout the world. In a vast majority of cases, the diagnosis can be made with reasonable

certainty but the underlying aetiology and pathogenesis of this disorder have only been partially understood.

Diabetes mellitus is a syndrome rather than a single disease. For instance, in insulin dependent variety, known as **type 1 diabetes mellitus** there is total loss of insulin secretion because of degeneration of the pancreatic β-cells. These patients must receive exogenous insulin. In another major type, termed the **type 2 diabetes mellitus**, adequate insulin is present in circulation, but it fails to act on its target tissues. Decline in the number of insulin receptors in target tissues or blocking of interaction of insulin with its tissue receptors is the underlying cause in type 2 diabetes (**Case 15.1**).

Type 1 diabetes: This type has an early onset, usually in childhood or adolescence, and is caused by autoimmune destruction of pancreatic β-cells. The precipitating cause could be:

- Viral infection initiating a chain of autoimmune reactions, or
- Cytokine response to viral infection could attract monocytes and macrophages that infiltrate and destroy the pancreatic islets. Insulin is no longer produced and the patient requires lifelong insulin injections for survival.

Some patients with type 1 diabetes have antibodies against the β-cell proteins. These are often present before the diagnosis of diabetes.

Type 2 diabetes: It is a disease of middle age, usually developing in patients over 40 years of age and typically obese. *It is more common and less severe than the type I, and has more complex origin*. On lipid screen these patients usually show dyslipidaemia with elevated fasting triglycerides or a low HDL-c level. Impairment of insulin secretion or a reduced responsiveness of the target tissues to circulating insulin (**insulin resistance**), or both are involved in pathogenesis.

Patients with untreated diabetes have hyperglycaemia, disordered metabolism of carbohydrates, lipids and proteins and are at risk of developing **microvascular** and **macrovascular diabetic complications.** Deficiency of or resistance to insulin or both can be ameliorated by regulating diet and monitoring therapy. However, the standard treatment measures may not prevent the development of chronic complications of diabetes, which effect almost every major organ of the body, including eyes, kidneys, nerves, and arteries giving rise to pathogenic conditions such as retinopathy, nephropathy, neuropathy, atherosclerosis, etc.

Differences between the type 1 and the type 2 diabetes are highlighted in Box 15.2.

A. Metabolic Alterations in Diabetes Mellitus

In diabetes mellitus, various body tissues may be affected. Although diabetes is considered as a disease of glucose under-utilization, there are indications of *shunting of glucose from insulin responsive tissues to insulin unresponsive tissues* (such as kidney, erythrocytes, lens and retina). These tissues receive a large glucose load, and metabolize it by opening some alternate pathways such as polyol pathway. This forms basis for the chronic complications of diabetes. Thus, different tissues respond in different ways to insulin deficiency.

Insulin is a hypoglycaemic hormone. It stands alone against on assay of hormones that are ***hyperglycemic:*** *glucagon, growth hormone, glucocorticoid*s and *epinephrine*.

BOX 15.2

Comparative Features of Type 1 and Type 2 Diabetes

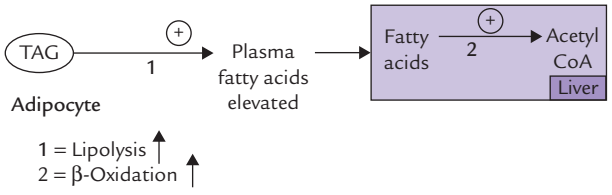
	Type 1	Type 2
Age of onset	< 20 years	> 40 years
Genetic locus	Chromosome 6	Unknown
Plasma insulin	Low to absent	Normal to high
Plasma glucagon	High, suppressible	High, resistant
Dyslipidaemia	Common	Very common
Acute complications	Ketoacidosis	Hyperosmolar coma
Insulin therapy	Responsive	Non-responsive
Sulphonylurea therapy	Non-responsive	Responsive
Pathogenesis	Beta cell destruction → absence of insulin secretion.	Low insulin of secretion, or insulin resistance, or both.

Insulin regulates a number of metabolic processes (Table 15.1). The control over these processes is lost in diabetes, resulting in a number of metabolic alterations described here.

1. ***Hyperglycaemia*** of diabetes mellitus is caused by both overproduction and underutilization of glucose. There is a relative excess of glucagon also. As a consequence, glucose is synthesized rather than consumed by liver, and glucose uptake into muscle and adipose tissue is reduced drastically (Table 15.1). All these lead to hyperglycaemia.

   As glucose is osmotically active, renal excretion of a large amount of glucose leads to loss of water (**osmotic diuresis**). Poorly controlled diabetic patients complain of having to drink large quantities of fluids (**polydipsia**) and of passing large volumes of urine (**polyuria**). The resulting fluid loss leads to **dehydration**.
2. ***Lipolysis* is greatly enhanced** in adipose tissue because of unopposed action of *hormone sensitive lipase* (HSL). As evident from Tables 15.1 and 15.2, low insulin : glucagon ratio triggers the enhanced HSL activity, as:
   - Insulin is an inhibitor of HSL.
   - Glucagon is a stimulator of HSL.

   The activated HSL greatly increases the degradation of TAG stores in adipocytes. Increased plasma free fatty acid concentration is a natural consequence of enhanced lipolysis. *Excess fatty acids are taken up by liver where they are degraded to acetyl CoA by β-oxidation.*



3. ***Excess acetyl CoA (over) produces ketone bodies:*** Acetyl CoA, generated as discussed above, can enter a number of pathways, e.g. lipogenesis, TCA cycle and ketogenesis. However, lipogenesis, and TCA cycle are shut down as described in Box 15.1. Excessive acetyl CoA molecules are, therefore channeled into ketogenesis in liver. These metabolic adaptation are **typical of the starving body** when adipose tissue and liver have to keep all of the other tissues alive by supplying them with glucose, fatty acids and ketone bodies.

TCA and lipogenesis being blocked (by death of OAA and by reduced activity of the key lipogenic enzyme), the excessive acetyl CoA is channeled into ketogenesis.

Excessive generation of ketone bodies in this way may occur in cases of near or total insulin deficiency in type 1 diabetes, leading to diabetic ketoacidosis with severe fluid and electrolyte imbalances and metabolic acidosis (**Case 15.3**). However, *ketosis does not occur* in *type 2 diabetes*. Reason for this difference is not known.

4. ***Increased influx of fatty acids in liver*** leads to another abnormal metabolic state, **fatty liver.** Decreased VLDL production and therefore, decreased exportation of the hepatic TAG, also contributes to fatty liver.

   Lipoprotein lipase activity is decreased in diabetes, resulting in decreased chylomicron and VLDL metabolism. This results in **hypertriglyceridaemia.**
5. ***Insulin promotes amino acid transport*** into cells and enhances protein synthesis; therefore, in diabetes, protein synthesis decreases and proteolysis increases. Muscle releases increased amounts of alanine, thereby providing a substrate for gluconeogenesis and promoting hyperglycaemia. Increased amino acid catabolism and increased urea production causes negative nitrogen balance and muscle wasting.

In diabetes mellitus, the metabolic pathways are deranged as in starvation.

B. Longstanding Diabetes Leads to Complications

In addition to the acute metabolic disturbances, just described diabetic patients suffer from slow development of many pathological changes in course of many years. These include abnormalities of small arteries (**microangiopathy:** diabetic retinopathy and nephropathy), large arteries (**diabetic macroangiopathy:** coronary heart disease and peripheral vascular disease), and **diabetic neuropathy**, which results from a combination of vascular and structural tissue changes. Though the longstanding complications develop slowly, they are as dangerous as the acute metabolic changes, markedly reducing life expectancy.

The longstanding complications mostly develop in **insulin independent tissues:** these tissues do not require insulin for entry of glucose, e.g. *brain, peripheral nerve tissue, kidney, intestine, lens, and red blood cells*. Consequently, in these tissues the intracellular glucose concentration parallels the blood glucose concentration. Intracellular glucose concentration, therefore, markedly rises in uncontrolled diabetes, and leads to longstanding diabetic complications by the following two biochemical mechanisms.

1. ***The polyol pathway*:** In, implying low affinity, hyperglycaemia, excess intracellular glucose can be reduced to

sorbitol by the action of *aldose reductase*, which is further oxidized by *sorbitol dehydrogenase* to fructose (Chapter 10).

- Because K_m of *aldose reductase* for glucose is very high, the pathway is not very active at normal glucose levels.
- In hyperglycaemia, however, glucose levels in insulin independent tissues increase, and consequently there is an increase in the activity of the polyol pathway. This enhances intracellular level of sorbitol. The high level of intracellular sorbitol exerts osmotic effect, and this plays a role in the development of diabetic cataract. In addition, accumulation of sorbitol and fructose interferes with the metabolism of inositol. The cellular levels of inositol have been found to be decreased in diabetic individuals, and the cellular metabolism gets deranged.

2. ***Protein modification by glucose***: The intracellular level of glucose is high in insulin independent tissues, and the glucose tends to attach, non-enzymatically, to protein molecules (protein glycation). This interferes with the normal function or turnover of proteins; for example, glycation of apolipoprotein-B slows down the rate of receptor dependent metabolism of low density lipoproteins (Chapter 12). This plays a major role in atherogenesis.

In fact, modification of proteins follows a sequence of non-enzymatic reactions, collectively known as the **Maillard reaction.** The reaction products (called Amadori products) transform further to form protein cross-links, known as **advanced glycation end products** (AGE). AGEs cause cross-linking of long lived body proteins such as tissue collagen or a nerve protein, myelin. It leads to thickening of basement membranes and stiffens the extracellular matrix, and decreases the elasticity of the arterial walls.

3. ***Enhanced oxidative stress***: Accumulation of AGE has been implicated as one of the factors responsible for generation of oxygen free radicals in diabetes (Chapter 27).

C. Assessment of a Diabetic Patient

Measurement of blood glucose level, an indispensable tool for diagnosis and monitoring of diabetic patients, has become part of a routine assessment of every patient admitted to hospital. Presently a broader range of tests are performed in diabetic patients.

Test for Diagnosis and Monitoring of Diabetes

1. *Urinalysis*: It is the most commonly employed and convenient test for screening diabetes mellitus. Generally, urinalysis is performed as a qualitative or semi-quantitative test for glucose and ketone bodies. Whenever the blood glucose level crosses *180 mg/dL* ***(renal threshold)***, glucose appears in the urine. However, there is no true renal threshold for ketone bodies and they are excreted in trace amounts at normal blood concentrations also. However, in uncontrolled diabetes, substantial quantities of ketone bodies are excreted in urine (**ketonuria**) together with glucose.

2. *Plasma glucose estimation:* It remains the mainstay for the initial diagnosis and follow up of diabetes. Diagnostic cutoff for diabetes mellitus is **126 mg/dL** in the **fasting sample** on two different occasions.

3. *Glucose tolerance test*: It consists of repeated measurement of the Plasma glucose level immediately before and at specific time (2 hours) after the ingestion of a (75 g) glucose solution. Diagnostic cut off between normal and diabetic individuals has been discussed in **Case 15.1**.

Diagnosis of diabetes based on values of: (i) fasting plasmo glucose (FPG) level, and (ii) 2-hour plasma glucose level after ingesting 75 g anhydrous glucose
FPG = 60–90 mg/dL : **Normal**
FPG = 112–126 mg/dL : **Impaired fasting glucose** (IFG)
FPG > 126 mg/dL : **Diabetes mellitus**
2h Plasma glucose < 140 mg/dL : **Normal**
2h Plasma glucose 140–200 mg/dL : **Impaired glucose tolerance** (IGT)
2h Plasma glucose > 200 mg/dL : **Diabetes mellitus**
IFG and IGT are pre-diabetic states

4. *Glycosylated haemoglobin*: It refers to glucose modified haemoglobin. It is formed because of a non-enzymatic glycosylation of haemoglobin: the glucose residue attaches to the free amino termini of the haemoglobin α- and β-chains. The glycosylated haemoglobins (GHb) make up about 4–6% of the total haemoglobin in normal red blood cells. In uncontrolled diabetes, because of persistent hyperglycaemia, increased formation and increased blood level of GHb results. Thus GHb serves as useful tool for estimating integrated/mean plasma glucose level over preceding few weeks.

The major fraction of the glycosylated haemoglobins is **HbA_{1c}**. It is formed when glucose is incorporated via N-glycosidic linkage into the N-terminal amino group of valine of each β-chain (Glc-CO-CH_2-NH-Val). Glycosylation is a continuous process occurring throughout the 120-day lifespan of the red cell, depending only on the prevailing plasma glucose concentration. Therefore, enhanced levels of HbA_{1c} occur in individuals with elevated plasma glucose levels, for example in diabetes mellitus. HbA_{1c} level may rise to as much as 20% of total haemoglobin in uncontrolled diabetes. However, unlike blood glucose, it remains elevated for the life of the red blood cell even if blood glucose levels decline to

normal. *HbA_{1c} thus reflects integrated plasma glucose levels over the last 4–6 weeks*, and thus identifies patients who do not comply with treatment. It is especially useful in childhood diabetes and during first trimester of pregnancy.

Estimation of GHb levels reflects history of plasma glucose levels over preceding 4–6 weeks. Being present in very small amounts, GHb is not pathological, but rather serves a useful purpose of monitoring patient's compliance in diabetic state or even in diagnosis of diabetes.

Fructosamine

Quantitation of glycosylated serum proteins, referred to as "fructosamine", reflects history of plasma glucose levels for a relatively shorter period of time (2–3 weeks). Fructosamine is a generic term applied to the stable condensation product of glucose with serum proteins, of which albumin is quantitatively the largest fraction. Because the circulating half-life of albumin is about 20 days, the serum fructosamine levels reflect glucose control over a period of the preceding 2–3 weeks. Fructosamine determination is done colourimetrically and determines the fructosamine-like adduct formed when glucose reacts with NH_2 residues of serum proteins.

Tests in Assessment of Long Term Complications of Diabetes

These are primarily tests of **renal function**, such as the measurement of the concentration of urea and creatinine in plasma. Urinary protein excretion (proteinuria) above 300 mg/day is the most important diagnostic test for diagnosis of diabetic nephropathy. Such overt proteinuria is preceded by presence of trace amount of albumin in urine (**microalbuminuria**). Microalbuminuria is only detectable by an assay, which is more sensitive than the conventional method for the measurement of albumin. It is used to predict the development of nephropathy.

Regular estimation of **serum lipid levels** evaluates risk of developing macrovascular complications.

Other Tests

Neurologic examination, ECG, and fundus (ophthalmic) examination detect complications, if any, at an earlier stage.

V. Metabolism of Xenobiotics

The body is exposed not only to nutrients but also to a variety of **non-nutritive chemicals,** collectively known as xenobiotics, the word meaning foreign to life.

Xenobiotics include *plant metabolites*, *therapeutic drugs*, *industrial chemicals* (soaps and detergents, dyes, bleaching agents), *food additives*, *pesticides* (DDT, aldrin), *cosmetics* and *bacterial toxins*. They may enter the body through various routes. Some of them are capable of causing biological damage, and so a prompt metabolism of xenobiotics is important.

Biological Damages Caused by Xenobiotics

1. Xenobiotics may bind to the body protein to alter its conformation, which may cause **functional inactivation of the protein.** Inactivation of enzymes or transport proteins in this manner may have deleterious effect on several biological functions.
2. Binding of xenobiotics to DNA may result in **mutation** often with serious consequences (Chapter 21).
3. Some xenobiotics exhibit structural similarity to a natural metabolite, therefore, act as **antagonists.** For example, dicumarol, a structural analogue of vitamin K, antagonizes action of this vitamin.
4. Lectins from plants **agglutinate** red blood cells.
5. Many bacterial products act as **neurotoxins.**

Humans in their inscrutable wisdom expose themselves to a variety of drugs, food additives, chemicals and intoxicants, collectively called xenobiotics, some of which are health hazards.

A. Metabolism of Xenobiotics

Water-soluble substances can be excreted in urine and bile. Indeed, some **water soluble xenobiotic** drugs like penicilline and amphetamine are disposed off by this simple route. **Lipophilic xenobiotics**, on the other hand, cannot be easily excreted and tend to accumulate in adipose tissue and other lipid-rich structures. A number of chemical transformations are required for metabolizing them to water-soluble products. These products are not only excretable, but also less toxic. However, there are some exceptions; for example, methyl pyridine derived from pyridine is more toxic than the parent compound. Similarly, benzpyrene and aflatoxins are converted to mutagenic epoxides, discussed later.

Liver is the most important organ of xenobiotic metabolism. It can handle an infinite range of molecules because of the low substrate specificity of the hepatic enzymes involved in xenobiotic metabolism. The metabolism proceeds in two phases:

- The ***Phase I*** includes reactions like hydroxylation, oxidation, reduction and hydrolysis; and
- The ***Phase II*** reactions include conjugation and methylation.

Polarity of the xenobiotic is increased in the Phase I because of introduction of functional group(s); and, the conjugated and methylated metabolites produced in the second phase are highly soluble. These metabolites can, therefore, be easily excreted.

Lipophilic xenobiotics are metabolized to water-soluble products (in two phases) that can be excreted.

Some xenobiotics are excreted after the Phase I, but the majority are eliminated after the Phase II. Mostly the final products are excreted in urine or bile, but some are also eliminated in expired air and faeces.

Phase I Reactions

Hydroxylation: It is the *most common* reaction type in phase I metabolism, which introduces hydroxyl group in the substrate molecule

$$\text{Substrate} + O_2 \xrightarrow[\text{NADP}^+]{\text{NADPH} + H^+} \text{Substrate-OH} + H_2O$$

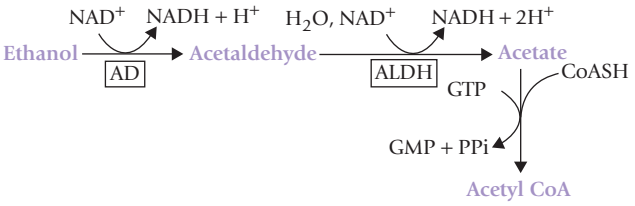
The reactions are *monooxygenase* reactions that require cytochrome P-450 as electron carrier. **Cytochrome P-450** is not a single protein but a whole superfamily of haem-containing proteins, each having 400–530 amino acids. The haem iron is liganded to both a conserved cysteine residue and a water molecule. All known P-450 cytochromes co-localize the microsomal (endoplasmic reticulum) membrane with (i) reduced nicotinamide adenine dinucleotide phosphate (NADPH) and (ii) *cytochrome P-450 reductase*. They hydroxylate a vast range of hydrophobic molecules (e.g. steroids, drugs, vitamin D intermediates, etc.) to more soluble products. The reaction sequences, as catalyzed by these enzymes, is shown in Figure 15.3.

Further information about cytochrome P-450 is given in next section.

Reaction sequence: The haem iron of cyto P-450 is normally liganded to a water molecule (shown by arrow), as discussed. When a substrate binds with cytochrome P-450, it displaces this water molecule, thus making space for binding of oxygen molecule to the haem iron. The bound oxygen is activated by transfer of an electron from *cytochrome P-450 reductase*, and after accepting electron, oxygen becomes highly reactive. It is readily incorporated in the substrate as hydroxyl group. It may be observed that one of the atoms of the molecular oxygen is incorporated into the substrate as hydroxyl group, and the other is reduced to water.

Oxidation: Immunosuppressant drugs like azathioprine and anticancer drugs like 6-mercaptopurine are metabolized by oxidative reactions. Azathioprine is first converted to 6-mercaptopurine, which is oxidized by the enzyme *xanthine oxidase* to 6-thiouric acid. *Xanthine oxidase* inhibitors, used in treatment of gout, therefore, impede metabolism of these drugs.

Detoxification of alcohol in liver also occurs by oxidative reactions catalyzed by two enzymes: *alcohol dehydrogenase* (AD) and *aldehyde dehydrogenase* (ALDH). The former oxidizes ethanol to acetaldehyde, which is further oxidized by *aldehyde dehydrogenase* to acetyl coenzyme A. Nicotinamide adenine dinucleotide is the cofactor for both these reactions, being reduced to NADH.



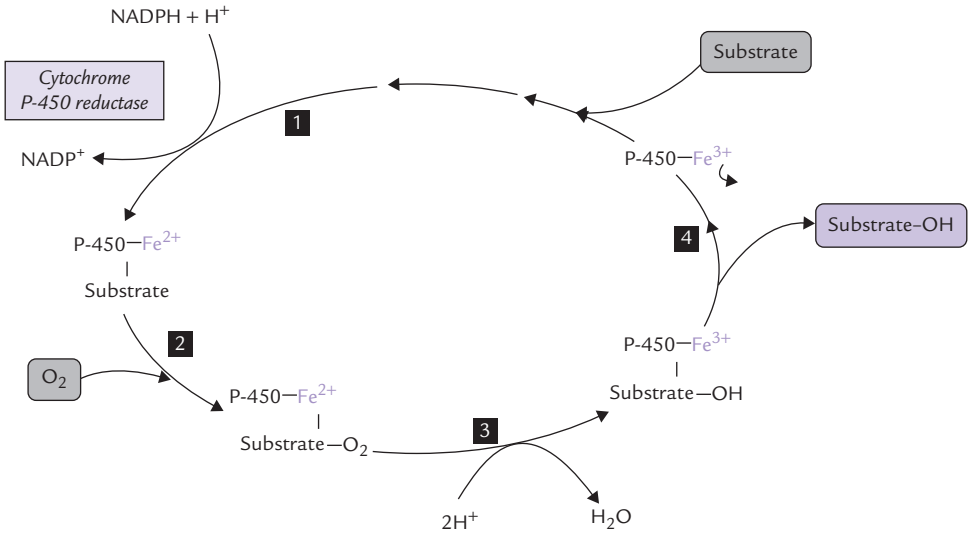


Fig. 15.3. Cytochrome P-450 reaction sequence.

The consequences of this pathway are discussed later in this chapter.

Methanol is also metabolized by the *alcohol dehydrogenase/ALDH* pathway, but it is converted into more toxic metabolites, formaldehyde and formic acid. Therapeutically, ethanol is used for the treatment of methanol poisoning.

Reduction: The major group of compounds which are metabolized by reductive reactions are nitro compounds. These are reduced to their amines, e.g. nitrobenzene converted to aminobenzene. Aldehydes and ketones are reduced to the corresponding alcohols.

Hydrolysis: Esters, amides and carbamates are subject to hydrolysis. Some important compounds metabolized in this way (i.e. by hydrolysis) are aspirin, procaine, clofibrate, indomethacin, di-isopropyl fluorophosphates (DFP), etc.

Phase I reactions oxidize the xenobiotic to effect change(s) in its biological properties. Mostly, this oxidation takes form of hydroxylation reaction, making it more polar.

The given reactions metabolize compounds to less toxic products, but sometimes have hazardous consequences, such as activation of a carcinogen (Box 15.3).

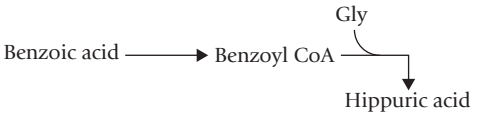
Phase II Reactions

In this phase, the metabolite of Phase I is conjugated with a hydrophilic molecule such as glucuronic acid, sulphate, glycine, glutamine or glutathione. Most of the reactions take place in liver and yield water-soluble products which are not only more water soluble and excretable, but also lose biological activity of the parent compound (Fig. 15.4). Conjugation may occur with the following compounds:

Glucuronic acid: It is the commonest reaction of phase II. Glucuronyl moiety of UDP-glucuronic acid is added to xenobiotics by *UDP-glucuronyl transferases*. Bilirubin is a good example of an endogenous compound excreted after glucuronidation. Some xenobiotics, e.g. phenol, acetaminophen, diazepam and morphine are also conjugated with glucuronic acid.

Sulphate: Steroids and methyldopa are examples of the drugs that are sulphated. Phosphoadenosine phosphosulphate (PAPS) transfers the sulphate group to xenobiotics.

Glutamine-glycine conjugate: Several drugs are converted to the corresponding thioester form and then conjugated with glycine/glutamine. For example, **benzoic acid**, after conversion to benzoyl CoA is conjugated with glycine.



Glutathione: Some carcinogens and acetaminophen are conjugated with glutathione: γ-glutamyl-glycinyl-cysteine. Conjugation occurs with the sulphydryl group of the cysteine residue; and the glutamate and glycine residues are split off. The cysteine residue is acetylated to form the mercapturic acid derivative of the xenobiotic, which is excreted.

Acetate: Sulphanilamide, caffeine and isoniazid are conjugated with acetate. *N-Acetyl transferase-2* transfers acetyl group from acetyl CoA to the xenobiotics.

In Phase II reactions, the xenobiotic (or its metabolite) is conjugated with a hydrophilic molecule (e.g. glucuronic acid, glycine, glutamine, glutathione) to form excretable product that has lost biological activity of the parent compound.

The **methyl conjugation** is discussed under methionine metabolism (Chapter 13).

BOX 15.3

Metabolic Activation of Carcinogens by P-450 Cytochromes

Benzpyrene is a polycyclic hydrocarbon in cigarette smoke. It binds with the specific receptors (called aryl hydrocarbon receptors), which induce the cytochrome P-450 system to bring about hydroxylation. Several sites of the benzpyrene molecules may be hydroxylated and several others undergo other oxidative transformations. These changes ultimately produce the final product, benz(a)pyrene-7,8-dihydrodiol-9,10-epioxide, which reacts spontaneously with DNA bases to cause point mutations. Individuals with genetically determined high activity of P-450 are at great risk of developing lung cancer if they smoke.

Metabolic activation of another carcinogen, aflatoxin B_1 (a toxin by the mold, *Aspergillus flavus*) occurs in liver to yield a highly mutagenic epoxide which reacts with guanine residues in DNA to produce point mutation. This may lead to liver cancer. Besides hepatitis B virus, aflatoxins are major risk factors for liver cancers in hot and humid climate of tropical countries.

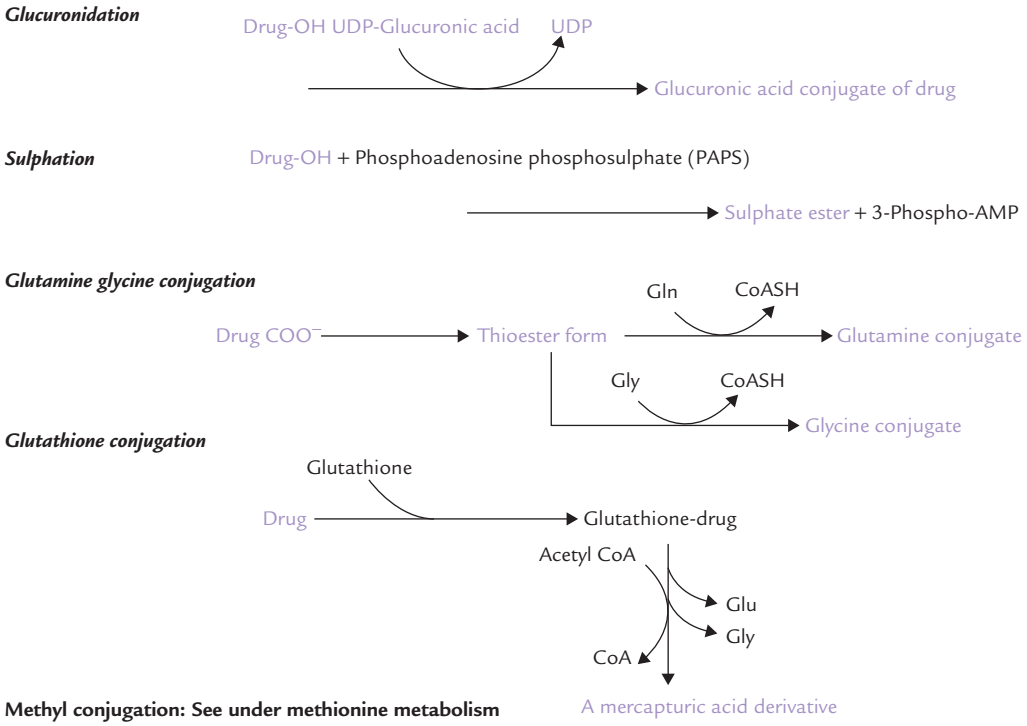


Fig. 15.4. Conjugation reactions in Phase II of xenobiotic metabolism.

B. The Most Versatile Biocatalysts: P-450 Cytochromes

The P-450 cytochromes are the most important enzymes of xenobiotic metabolism that catalyze hydroxylation reactions in phase I, using NADPH and oxygen. They are so named because they *absorb light of wavelength* ***450 nm****, when exposed to carbon monoxide*. Several hundred types of cytochrome P-450 are known, which occur as membrane bound proteins either in the micro-somes or in the inner mitochondrial membrane. They can be broadly divided into two varieties:

1. *Lipid metabolizing variety*: About a dozen of cytochrome P-450 participate in lipid metabolism, e.g. in ω-oxidation of fatty acids and in synthesis of steroid hormones. This variety of P-450 flavoproteins has a tight substrate specificity.

2. *Drug metabolizing variety*: These cytochromes are present in **microsomes** where they co-localize with NADPH and an enzyme called *cytochrome P-450 reductase*. The latter is a membrane flavoprotein. This variety of cytochrome P-450 has broad specificity, which enables them to produce a wide ranging efficacy in xenobiotic metabolism.

P-450 cytochromes have wide and overlapping substrate specificities which enable them to hydroxylate almost any xenobiotic.

Cytochrome P-450 are most versatile biocatalysts that can oxidize any conceivable foreign molecule. They are inducible by their own substrate.

Cytochrome P-450 are highly inducible: One of the most interesting aspects of cytochrome P-450 is that its substrates (and other xenobiotic agents) often induce the enzyme synthesis, usually by increasing the rate of transcription. Such **self-induction** provides an efficient mechanism of detoxification. However, it also results in rapid development of tolerance against drugs. For example, dose of the antiepileptic drug, phenobarbital has to be increased 3–4-fold within a week of starting the therapy in order to maintain the original therapeutic effect. Phenobarbital also induces the cytochrome P-450 that metabolizes other drugs such as dicumarol. This necessitates increase in the dosage of this anticoagulant if the patient is treated with phenobarbital.

C. Isoforms of Cytochrome P-450 (CYP)

A large number of isoforms of cytochrome P-450 (about 150) have been identified. Recall that they are encoded by at least 12 gene families. Three of these **gene families**—designated **CYP1, CYP2,** and **CYP3**—share responsibility

for xenobiotic metabolism. Members of a particular family show at least 40 per cent homology in amino acid sequence. Many families comprise **subfamilies;** for example, CYP2 has five subfamilies (CYP2A, CYP2B, CYP2C, CYP2D, CYP2E), and each of these subfamilies comprise individual members or isoforms.

All P-450 cytochromes → Families → Subfamilies → Isoforms

For example, CYP1A1 and CYP1A2 denote cytochromes that are members of family **1** and subfamily **A**, and are the first two isoforms (individual members) of that subfamily.

The CYP3A subfamily appears to be involved in the metabolism of most drugs, and CYP3A4 appears to be the most active member of this subfamily.

In human tissues, at least 12 families of cytochrome P-450 have been identified. Most families have many subfamilies, each of which comprise isoforms.

Mutations in CYP genes can lead to decreased activity of certain isoforms of CYP, which in turn, will also affect the pharmacologic activity of some drugs. The best described example is that of the CYP2D6 (family 2, subfamily D_i isoform 6th). It causes hydroxylation of debrisoquine, a now little-used blood pressure decreasing drug. CYP2D6 also metabolizes a significant number of other commonly used drugs (e.g. timolol, nortryptiline, perphenazine, fluxetine, dextremethorphan), so its deficiency remains a clinically significant condition.

Certain inhibitors (e.g. quinidine and haloperidol) cause inhibition of the CYP2D6, thereby affecting metabolism of these drugs.

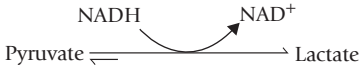
VI. Alcohol Metabolism

Alcohol is not only an intoxicant but also a nutrient. Oxidation of 1 g of ethanol provides about 7 kcal of energy, so a person consuming 100–120 g of alcohol per day covers about one-third of his energy needs. Excess intake of alcohol remains one of the most serious health concerns all over the world. As mentioned earlier, alcohol is oxidized in the liver, mainly by *alcohol dehydrogenase*, to form acetaldehyde, which is oxidized by *aldehyde dehydrogenase* to acetyl CoA. An isoform of P-450 cytochrome, CYP2E1, also contributes to ethanol metabolism. This route, however, contributes only for a small proportion of total alcohol metabolisms under most conditions.

Ethanol is oxidized in liver to acetyl CoA and, therefore, is a substrate for generation of metabolic energy (**empty calories**).

Biochemical consequences: Unlike carbohydrate and fatty acid oxidation, alcohol metabolism in liver is not subject to negative feedback controls, so the oxidation of alcohol takes preference over the oxidation of other nutrients. Because NAD^+ is the cofactor for both the reactions of the alcohol metabolism and is reduced to NADH, alcohol metabolism leads to an **increased NADH to NAD ratio**. This causes:

1. Inhibition of oxidation of lactate to pyruvate—a step that requires NAD^+ as a cofactor. Lactate accumulation results, which may lead to *lactic acidosis*.



2. *Inhibition of gluconeogenesis*, because hepatic gluconeogenesis requires pyruvate as a substrate. Therefore, alcoholism carries risk of hypoglycaemia.
3. Alcoholics have poor glycogen stores in liver, therefore the likelihood of *hypoglycaemia* is increased when they fast.
4. Increased NADH : NAD^+ ratio also inhibits β-oxidation of fatty acids and promotes *triglyceride synthesis*.

Polymorphism

Both the enzymes of alcohol metabolizing pathway—*alcohol dehydrogenase* and *ALDH*—are subject to genetic polymorphisms, and these contribute to individual differences in alcohol tolerance.

Three genetic variants of *alcohol dehydrogenase* with different pH optima and Vmax values have been described. The most interesting polymorphism, however, affects the mitochondrial *aldehyde dehydrogenase*, i.e. the ALDH2 allele, which encodes an enzyme with reduced catalytic activity. This leads to increased plasma concentrations of acetaldehyde after the ingestion of alcohol. As a result, an individual experiences unpleasant flushing with tachycardia and sweating, which discourages alcohol abuse. Indeed, alcoholism is rare among such individuals. Some 30–40% of Chinese, Japanese, Mongolians, Koreans, Vietnamese, and Indonesians, and also many South American Red-Indians have this ALDH2 allele.

Disulphiram (*an antabuse*), a drug that inhibits ALDH, also leads to these symptoms when alcohol is taken, and may be given for the treatment of alcoholics. But use of this drug requires strict medical supervision: fatal reactions have occurred when disulphiram was mixed into an unsuspecting alcoholic's drinks.

Exercises

Essay type questions

1. Explain briefly the biochemical changes in starvation and uncontrolled diabetes.
2. Describe the diverse biochemical effects of insulin. Mention the biochemical tests for diagnosis and monitoring of a case of diabetes mellitus.
3. Explain biochemical basis of development of complications in longstanding diabetes mellitus.
4. What is detoxification? Discuss two examples of metabolism of xenobiotics.
5. Discuss the pathway for metabolism of alcohol and the biochemical changes it leads to. Explain metabolic consequences of alcoholism.

Write short notes on

1. Cytochrome P-450
2. Alcohol metabolism
3. Glycosylated haemoglobin
4. Glucose tolerance test
5. Fructosamine
6. Metabolic adaptations in starvation

CLINICAL CASES

CASE 15.1 An obese, hypertensive woman experiencing generalized weakness

A 38-year-old woman complained of generalized weakness, lethargy, polydipsia, and loss of appetite for last 3 months. She was beings treated for hypertension with a calcium antagonist.

On examination, the patient was obese, weighing 112 kg (height was 5 ft 2 in). Possibility of diabetes mellitus was recognized and a blood sample was obtained for biochemical analysis. Fasting blood glucose level was 144 mg/dL and the fasting plasma insulin level was 28 μU/mL (normal is 6–24 μU/mL).

A request for estimation of lipid parameters in the fasting sample was also sent to the biochemistry laboratory. In view of the above test results, the patient was given an oral glucose tolerance test (OGTT) the next morning.

The OGTT results for (venous) plasma glucose (75 g of anhydrous glucose was orally given) were as below:

0 hr	146 mg/dL
1 hr	216 mg/dL
2 hr	214 mg/dL

Serum insulin level was 88 μU/mL at one hr, 84 μU/mL at two hours after glucose intake. The patient's lymphocytes bound 1.2% of the administered insulin per 56×10^6 cells, which is less than that expected in a normal subject. Results of the lipid parameters were returned from the laboratory after two days. These were:

Test	Patient's results	Reference range
Serum cholesterol	266 mg/dL	150–250 mg/dL
Serum triglycerides	170 mg/dL	< 140 mg/dL
HDL-cholesterol	28 mg/dL	30–70 mg/dL

The patient was advised to observe strict dietary restrictions and to reduce her body weight.

Q.1. Comment on the results of OGTT, and discuss whether the patient has diabetes mellitus?

Q.2. Discuss the biochemical basis of glucose intolerance in this patient.

Q.3. Why was the plasma insulin elevated before the weight reduction programme, and how would it be affected if the patient is successful in reducing weight by about 20 kg.

Q.4. Following weight reduction, what do your expect the results of IGTT and the other biochemical parameters to be?

CASE 15.2 Loss of weight despite increased thirst and appetite

A 14-year-old girl was brought to the primary health center with complaints of weakness, increased thirst, and loss of weight in spite of a ravishing hunger. These symptoms started about 4 weeks earlier and since then she had lost about 7 kg of weight. Her distress was aggravated by nocturnal polyuria: the frequency of micturation was increased and she had to get up 3–4 times at night to void. She thought that her father who died two years ago suffered from 'sugar in urine'. The Medical Officer in the health centre analyzed a urine sample and detected presence of reducing sugar (but no ketone bodies). The patient's mother was advised to take her to the teaching hospital immediately.

The patient was admitted by the resident doctor in the hospital emergency and a blood sample was urgently sent for estimation of blood sugar. On examination, the patient was found to be thin with dry skin and sunken eyeballs. There was nothing else of note upon examination.

Meanwhile the results were sent by the laboratory: the random venous plasma glucose level was markedly elevated (280 mg/dL).

The patient was diagnosed as having insulin-dependent diabetes mellitus (IDDM) and insulin therapy was started with daily administration of the intermediate-acting insulin. Blood glucose levels were monitored and the insulin dose was adjusted accordingly. She was discharged on the fourth day with a fasting plasma glucose of 110 mg/dL. Her mother was taught the basic skills of insulin injection, blood glucose monitoring, and urine ketone measurement.

Q.1. It is necessary to perform oral glucose tolerance test (OGTT) in this patient to confirm the above diagnosis?
Q.2. State the biochemical basis of symptoms of this girl?
Q.3. At the time of admission, what do you expect the patient's tissue glycogen levels to be?
Q.4. Why is it important to use the highly purified insulin in the management of IDDM?
Q.5. How does the response to insulin in an obese diabetic person compare with that in a non-obese diabetic person?
Q.6. What is C-peptide? State the significance of the C-peptide test?

CASE 15.3 Abdominal pain and dehydration in an insulin dependent diabetic patient

A 13-year-old boy was brought to the hospital emergency with complaints of abdominal pain and uncontrolled thirst. History was obtained from his parents who told that the child had insulin dependent diabetes mellitus. The condition was diagnosed about four years ago and since then he was on insulin therapy.

Over the last two days he had gone to a hill-station with his schoolmates for a weekend trip. After return, he started to pass large volumes of urine and developed abdominal pain, severity of which had considerably increased in the last 24 hours. His father thought that these problems were because the boy had been negligent; the boy admitted drinking a lot of soft-drinks and dining in fast-food fancy restaurants without taking insulin injections regularly. Presently, he was feeling nervous and drowsy, and was hyperventilating.

On examination, there was right quadrant tenderness and the bowel sounds were decreased. His blood pressure was 102/70 mmHg and the pulse was 98 beats per minute (supine). The skin was dry, thin, wrinkled and inelastic. Other signs of dehydration were also present, such as dry mucous membrane and sunken eyeballs. The patient was admitted with a preliminary diagnosis of dehydration, with possibility of an attack of acute appendicitis. Small bowel obstruction and urinary tract infection (UTI) were also considered.

Blood and urine samples were quickly sent to the biochemistry laboratory and the following results were obtained:

Test	Patient's results	Reference range
Glucose	290 mg/dL	60 – 100 mg/dL
Urine glucose	+++	–ve
Urine ketone bodies	+	–ve

Serum electrolytes (Na^+ 142 mmol/L; K^+ 5.8 mmol/L; Cl^- 108 mmol/L) urea nitrogen (18 mg/dL), and creatinine (0.8 mg/dL) were within normal range. Urine specific gravity was 1.025 and no proteins, haem, sediments, etc. were detected in the urine sample.

Soon after, the patient became drowsy and lapsed into coma. History, clinical picture together with laboratory evidence of hyperglycaemia, acidosis, glycosuria, and ketonuria commanded consideration of a diagnosis of diabetic coma. To assess severity of the metabolic derangement, an arterial blood sample was drawn for the estimation of blood-gas parameters.

Test	Patient's results	Reference range
pH	7.31	7.35–7.45
pCO_2	32 mmHg	35–48 mmHg
pO_2	108 mmHg	83–108 mmHg

Plasma ketone bodies were detected in moderate amounts in diluted plasma (nitroprusside test). Treatment was started immediately. Four hours later, the patient's condition showed considerable improvement. Two days later, he could be stabilized on his former dose of insulin and was discharged from the hospital.

Q.1. What is the most probable diagnosis?

Q.2. State the most likely cause of abdominal pain in this patient.

Q.3. Outline the treatment for this patient. State the biochemical basis for this treatment.

Q.4. Provide a biochemical explanation for the following events which appear paradoxical.

(i) Potassium concentration in the plasma tends to rise in spite of loss of this cation in urine.

(ii) Sodium is lost in urine, but its concentration in plasma is almost normal.

Q.5. What are the other causes of coma in diabetes mellitus?

CHAPTER

16 PORPHYRINS AND BILE PIGMENTS

Porphyrins are cyclic compounds formed by four covalently-linked pyrrole rings. They are metabolic intermediates related to the pathway of haem synthesis. Bile pigments, on the other hand, are produced by catabolism of haem. The pathways of synthesis and catabolism of haem are described in this chapter, and the associated pathological states are highlighted.

After going through this chapter, the student should be able to understand:

- Fundamental structure and general characteristics of porphyrins.
- Metabolic pathway, enzymes and intermediates, regulation of biosynthesis of haem and biochemical defects in different type of porphyrias.
- Pathway of haem catabolism that produces bilirubin; intermediates as well as the enzymes catalyzing various steps of bilirubin metabolism and the associated pathology.

I. General Characteristics of Porphyrins

Structure of a pyrrole ring is shown in Figure 16.1. The four pyrrole rings of the porphyrin are linked through methenyl (–CH =) bridges. This creates a cyclic structure, called **cyclic tetrapyrrole.** The pyrrole rings are named as I, II, III and IV; and the methenyl bridges as α, β, γ and δ. Since the tetrapyrrole contains alternate single and double bonds, structure is commonly referred to as the **conjugated ring system** (Fig. 16.2).

The double bonds of the cyclic tetrapyrrole absorb visible light, and so *porphyrins are coloured compounds*. The red colour of haemoglobin is due to the porphyrin ring of its haem component. Porphyrins also fluoresce intense reddish pink colour in ultraviolet light. This property is also accounted by double bonds.

Each of the pyrrole rings of porphyrin contains a nitrogen atom which can link with a metal atom. This permits chelation of several metal atoms by the porphyrin ring. Joining of a metal with a porphyrin ring in this way yields **metalloporphyrin.** Physiologically, the metalloporphyrins containing **iron** are most important ones; they are termed *iron-porphyrins.* Most abundant of these

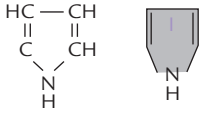


Fig. 16.1. Structure of a pyrrole ring.

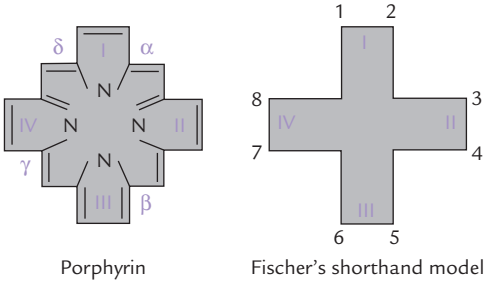


Fig. 16.2. Structures of porphyrin ring: I, II, III, IV are pyrrole rings, 1–8 are substituent positions and α, β, γ, δ are methenyl bridges.

is **haem**, which is present in haemoglobin, myoglobin, cytochromes, *catalase,* etc. Magnesium is present in **chlorophyll**, the metalloporphyrin that constitutes photosynthetic pigment of plants.

II. Biological Significance of Porphyrins

In nature, the metalloporphyrins are linked with proteins to form conjugated proteins termed **metalloporphyrino proteins.** Most important of these are the iron-porphyrin containing **haem-proteins,** which participate in oxygen transport (haemoglobin), oxygen storage (myoglobin), electron transport (cytochromes), hydrogen peroxide inactivation (*catalase*), hydroxylation, oxygenation and other processes.

1. ***Haemoglobin*** is an oligomeric protein consisting of four subunits; each subunit contains a globin, a polypeptide, joined to a haem molecule. Haemoglobin is the principal transporter of oxygen in blood. It also carries some amount of carbon dioxide, and serves as an important blood buffer (Chapter 17).
2. ***Myoglobin***, a monomeric protein, is structurally similar to a subunit of haemoglobin. It is present in muscles where it stores oxygen.
3. ***Cytochromes*** are electron transferring proteins that participate in oxidation–reduction reactions. Various types of cytochromes such as b, c, c_1, and aa_3 act in that sequence in electron transport chain (Chapter 14). Cytochrome P-450 participates in hydroxylation of aromatic and aliphatic compounds, such as steroids, alcohol, and many drugs.
4. ***Catalase*** is an iron-porphyrin enzyme present in animal cells, which catalyzes breakdown of hydrogen peroxide. The haem group forms a part of the active site of this enzyme.

   Activity of *catalase* is minimal in plants where another iron-porphyrinoprotein, ***peroxidase***, performs similar function.
5. ***Tryptophan pyrrolase*** is an iron-porphyrin enzyme that plays an important role in catabolism of tryptophan.
6. ***Erythrocruorins*** are iron-porphyrinoproteins occurring in blood and tissue fluids of some invertebrates. Their function corresponds to that of haemoglobin.

Most of the haem in our body is in the oxygen transport protein, haemoglobin. Other haem-proteins, e.g. myoglobin, *cytochrome, catalase,* etc. participate in a broad range of biological processes.

It is important to note that the prosthetic group, haem, is performing different roles in different proteins. The biological role of haem in a given protein is dictated by the three-dimensional structure of the protein.

III. Nomenclature

The porphyrins found in nature have side chains located at corners of the pyrrole rings. Figure 16.2 shows the possible sites of attachment that are denoted from 1 to 8. Different porphyrins vary in the nature and position of the side chains attached to these sites. The three most important porphyrins in humans are **uroporphyrin** (URO), **coproporphyrin** (COPRO), and **protoporphyrin** (PROTO).

- URO has four propionate and four acetate side chains.
- COPRO has four propionate and four methyl groups.
- PROTO has four methyl, two vinyl and two propionate groups (Fig. 16.3).

The side chain groups in porphyrins may be arranged in four different structural configurations (I to IV). In type I there is symmetric distribution of side chains, and in type III it is asymmetric (Fig. 16.4). For example, in uroporphyrin I, the acetate (A) and propionate (P) side chains of the pyrrole rings alternate regularly (1,3,5,7 and 2,4,6,8). In type III, on the other hand, there is asymmetric arrangement (1,3,5,8 and 2,4,6,7), meaning that the expected arrangement of side chains is reversed on ring IV. Only type I and type III series are found in nature and the porphyrins of the type III series are not only more abundant but also physiologically important in humans.

In case of PROTO, it is possible to arrange the two propionate, two vinyl and four methyl groups in 15 different configurations. Only the type IX isomer (Fig. 16.5) is produced in the human body.

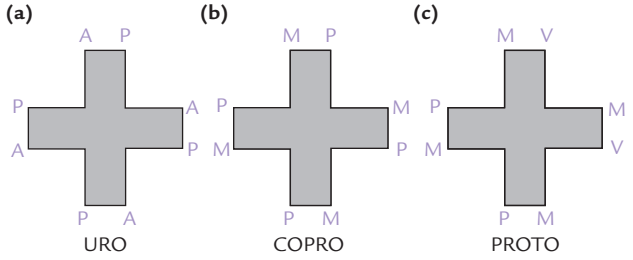


Fig. 16.3. Structures of **(a)** Uroporphyrin, **(b)** Coproporphyrin, and **(c)** protoporphyrin, showing various groups attached to the porphyrin nucleus (A = acetate, P = propionate, M = methyl, V = vinyl).

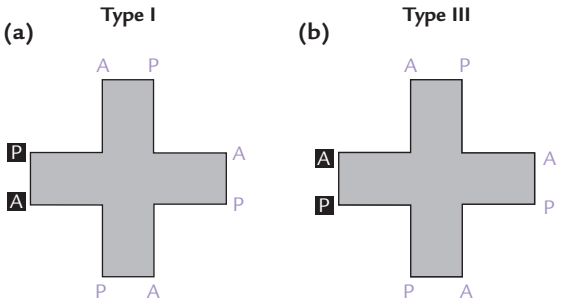


Fig. 16.4. Structure of **(a)** Uroporphyrin I with **symmetric distribution** of substituent groups, and **(b)** Uroporphyrin III with **asymmetric distribution** of substituent groups on the ring IV (A = acetate, P = propionate).

IV. Synthesis of Haem

Haem is a cyclic tetrapyrrole with a hexavalent iron (Fe^{2+}) atom at the centre that is coordinately linked with four pyrrole nitrogens. Haem is the final product of the porphyrin synthetic pathway, which starts from two relatively simple precursors: *glycine* and *succinyl CoA*. The two major sites of the biosynthetic pathway are *erythroid cells*, which synthesize approximately 85% of the body's haem groups, and the *liver*, which synthesizes most of the remainder.

Bone marrow and liver are the most important sites for the synthesis of haem, though some other tissues that synthesize haem-proteins can also produce haem in small amounts.

It requires participation of eight enzymes to synthesize haem. Of these enzymes, four (the first and the last three) are *mitochondrial* and the rest are *cytosolic* (Fig. 16.6). The reactions are all irreversible, occurring in 3 stages:

Stage I: Formation of Tetrapyrrole Ring

Step 1: Formation of δ-Aminolevulinic Acid

This step involves condensation of glycine and succinyl CoA to form δ-aminolevulinic acid (ALA). The reaction is catalyzed by δ-*aminolevulinic acid synthase* (*ALA synthase*), a mitochondrial enzyme, located on the matrix side of the inner mitochondrial membrane. Pyridoxal phosphate (PLP) is required as a coenzyme, hence anaemia may be manifested in PLP deficiency.

ALA synthase is an inducible, ***rate-limiting*** enzyme for the pathway. It is the principal regulated enzyme in the liver, as discussed later.

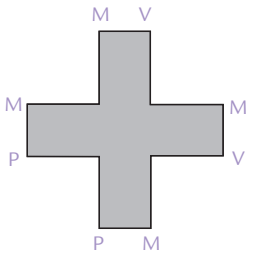


Fig. 16.5. Structure of protoporphyrin IX.

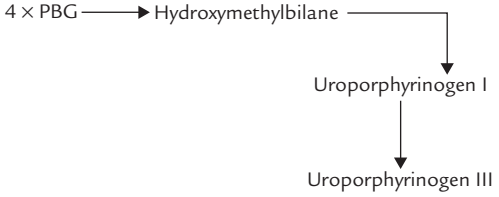
Step 2: Formation of Porphobilinogen

Two molecules of ALA are condensed by the cytosolic zinc containing *ALA dehydratase* to yield porphobilinogen. The enzyme is sensitive to inhibition by lead. This inhibition explains the elevation of ALA levels and anaemia in lead poisoning.

Step 3: Formation of Uroporphyrinogen III

Condensation of four molecules of porphobilinogen yields the first cyclic tetrapyrrole, uroporphyrinogen III. It requires successive action of two enzymes: *uroporphyrinogen I synthase* (also called *porphobilinogen deaminase*) and *uroporphyrinogen III cosynthase*.

1. *Uroporphyrinogen synthase* catalyzes condensation of four PBG units to form the linear tetrapyrrole, *hydroxymethylbilane*, in which the acetate and propionate side chains of the pyrrole ring alternate regularly. This unstable intermediate may cyclize spontaneously to form uroporphyrinogen I.
2. *Uroporphyrinogen III cosynthase* rearranges the orientation of propionate and acetate side chains on one of the pyrrole rings (ring IV), to produce the asymmetric, physiological type III isomer of uroporphyrinogen.



Uroporphyrinogen III is the first compound containing a porphyrin-like ring structure in the haem biosynthetic pathway (Fig. 16.6). It is also a key intermediate in the synthesis of vitamin B_{12} in bacteria, and of chlorophyll in bacteria and plants.

In stage I, simple precursors–glycine and succinyl CoA, condense to form a pyrrole ring. Four pyrrole rings condense to form first a linear pyrrole and then a cyclic tetrapyrrol, i.e. uroporphyrinogen-III.

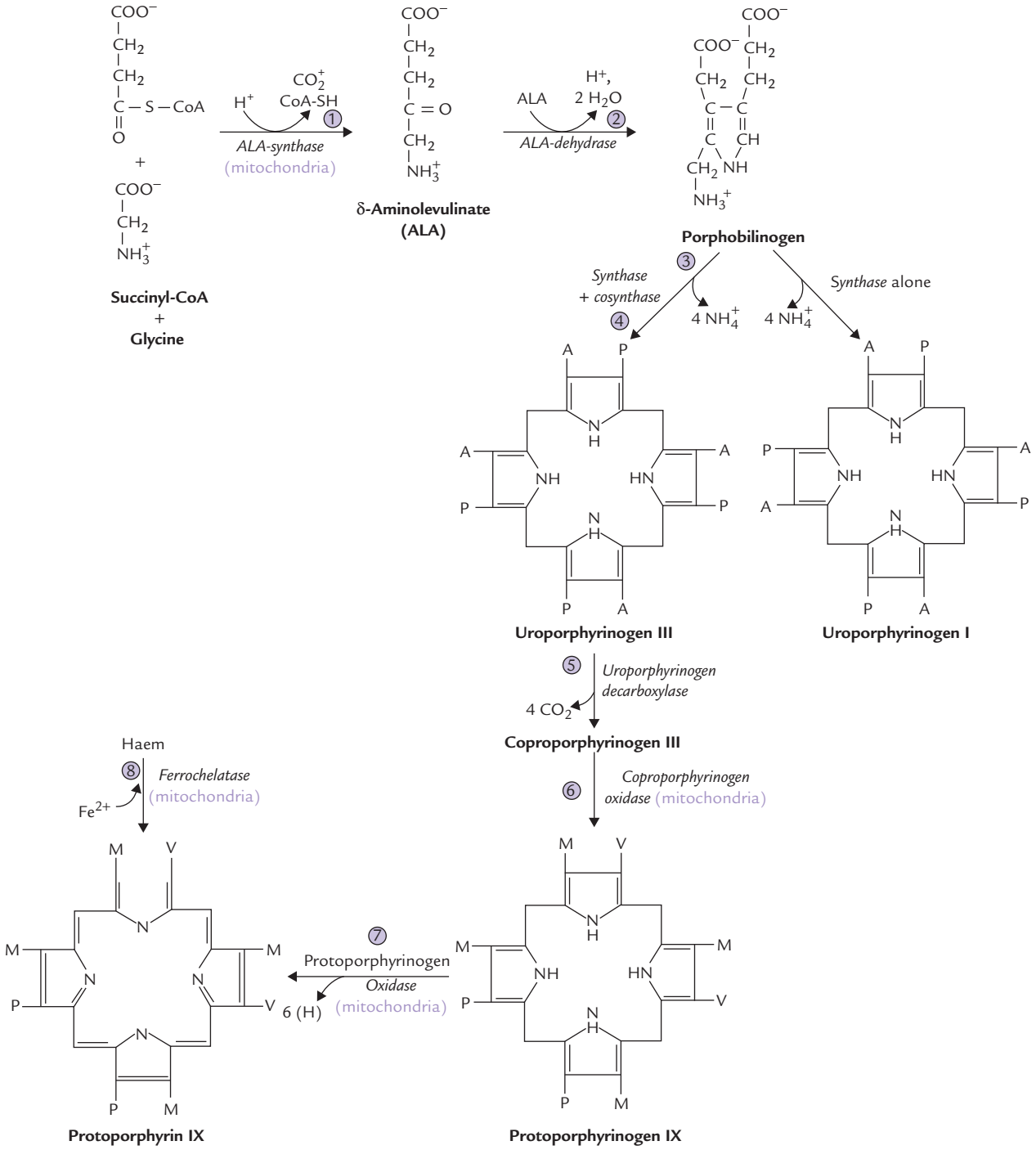


Fig. 16.6. Biosynthesis of haem.

Stage II: Processing of Uroporphyrinogen III to Protoporphyrin IX

Uroporphyrinogen III is processed to protoporphyrin IX by modification of its side chains and oxidation of porphyrinogen to porphyrin. It involves a series of reactions catalyzed by:

1. *Uroporphyrinogen decarboxylase* (step 5), decarboxylates all four acetate side chains (A) to form methyl (M) groups; the reaction product coproporphyrinogen III diffuses back into the mitochondrion.
2. *Coproporphyrinogen oxidase* (step 6), oxidatively decarboxylates two of the propionate side chains (P) from ring I and ring II to vinyl groups (V).
3. *Protoporphyrinogen oxidase* (step 7), oxidizes the methylene bridges ($-CH_2$) linking the pyrrole rings to methenyl bridges ($=CH-$). Protoporphyrin IX is thus formed.

Stage III: Formation of Haem

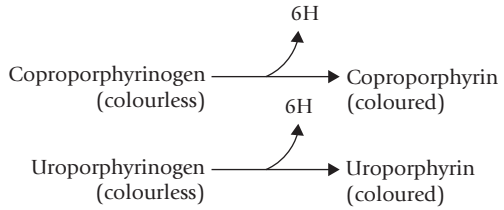
Finally, formation of haem is accomplished by *ferrochelatase* (or *haem synthase*), which incorporates Fe^{2+} in the centre of the porphyrin ring. The enzyme is inhibited by lead.

Uroporphyrinogen III is processed to haem by decarboxylation of the acetate and propionate groups, oxidation of the porphyrinogen to the porphyrin, and chelation of iron in centre of the molecule.

The following points about this biosynthetic pathway are noteworthy.

1. **Localization:** The pathway takes place in most mammalian tissues except mature erythrocytes which do not contain mitochondria. Major sites of biosynthesis include liver, which synthesizes a number of haem-proteins (most prominently, the cytochrome P-450), and the erythrocyte-producing stem cells of the bone marrow, which are active in haemoglobin synthesis.
2. **Polarity of intermediates:** The initial intermediates of this pathway are polar. Uroporphyrinogen is strongly polar since all its side chains (four acetate and four propionate) are predominantly polar.

   The subsequent intermediates become more and more non-polar.
3. **Porphyrinogens vs Porphyrins:** The porphyrinogens are reduced porphyrins. They are not coloured because their pyrrole rings are connected by methylene ($-CH_2$) bridges, and the double bonds are not conjugated over the whole system. The porphyrinogens are the actual intermediates in the haem biosynthetic pathway. Porphyrinogens can be converted to the corresponding coloured porphyrins on exposure to light, when they lose hydrogen atoms.



4. **Partly mitochondrial and partly cytosolic:** The pathway is initiated in the mitochondria, and the first intermediate, ALA, diffuses into the cytosol where next few reactions, up to formation of coproporphyrinogen III takes place. The latter compound diffuses into the mitochondrion where the last three reactions occur (Fig. 16.6).

A. Regulation of Haem Synthesis

The regulation revolves around the initial enzyme, *ALA-synthase*, and *haem is the principal regulator*. The regulatory effects **in liver** are multiple, as discussed here.

1. **Repression mechanism:** The *ALA synthase* has an unusually short biological half-life of 1 to 3 hours in the liver and its synthesis is suppressed very effectively by haem. Only free, non-protein bound **haem** can cause the repression. Haem (or its oxidation product **haematin**) activates a repressor protein that turns off *ALA synthase* biosynthesis at translation level.

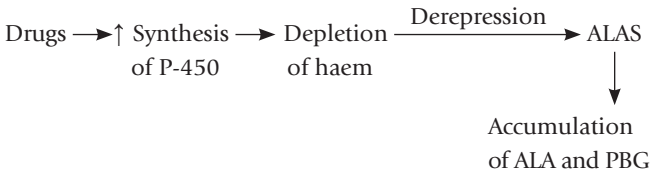
   Erythropoietin, a protein produced by the kidneys and found in larger than normal amounts in high-altitude dwellers, counteracts the effects of the repressor protein. Erythropoietin deficiency exists in chronic renal failure, which explains the observed anaemic condition in such patients.
2. **Allosteric inhibition:** The *ALA synthase* is allosterically inhibited by **haem, haematin** and **haemin.** The normal end product, haem, when present in excess, is oxidized to **haematin,** which contains a hydroxyl group attached to the Fe^{3+} atom.

   Replacement of the hydroxyl group by a chloride ion produces **haemin**. In addition to acting as allosteric inhibitor of *ALA synthase*, haemin represses synthesis of *ALA synthase* and also inhibits transport of cytosolic *ALA synthase* into mitochondria, where it acts.

The committed step of haem biosynthesis is the formation of ALA from succinyl CoA and glycine, and the enzyme catalyzing this step (*ALA synthase*) is subject to feedback inhibition by free, non-protein bound haem.

3. **Others**: Induction of *ALA synthase* is brought about by a variety of xenobiotics (environment pollutants), natural steroids and therapeutic drugs, such as barbiturates, phenytoin, griseofulvin, etc. These agents induce synthesis of cytochrome P-450. The latter belongs to a family of haem proteins that contain 65% of the total haem in the liver (Chapter 15). Increased synthesis of P-450 uses up the available haem, resulting in depletion of the small pool of free, unbound haem. The haem depletion leads to a derepression of *ALA synthase*.

Thus, administration of phenobarbital and other such drugs is contraindicated in porphyric patients, because they cause an inappropriate induction of *ALA synthase* and consequent accumulation of the offending intermediates. This may precipitate an acute porphyric episode, as discussed later.



ALA synthase is stimulated by 5-β-dehydro steroids (e.g. testosterone) in their reduced form.

In erythroid cells, the regulatory enzymes and effectors are different (**Box 16.1**).

V. Disorders of Porphyrin Metabolism: Porphyrias

Porphyrias are a group inherited and acquired disorders caused by abnormalities in the haem synthetic pathway (Fig. 16.7). They may affect the liver or the blood-forming tissues, and are characterized by excessive accumulation and excretion in urine/or faeces of porphyrins or porphyrin precursors. Porphyrins have a deep red or purple colour (Greek: *porphyria* = purple).

Unlike other disorders involving erythrocytes, anaemia does not usually dominate the clinical picture. The partial interruption of the haem biosynthetic pathway does not greatly diminish haem synthesis, but causes the intermediate metabolites to be present in excess. The excessive amount of porphyrins so formed are deposited in body tissues, to cause cutaneous, neurological or other clinical manifestations. Thus, *excessive intermediate metabolites, and not diminished haem production, is the offending factor in porphyrias.*

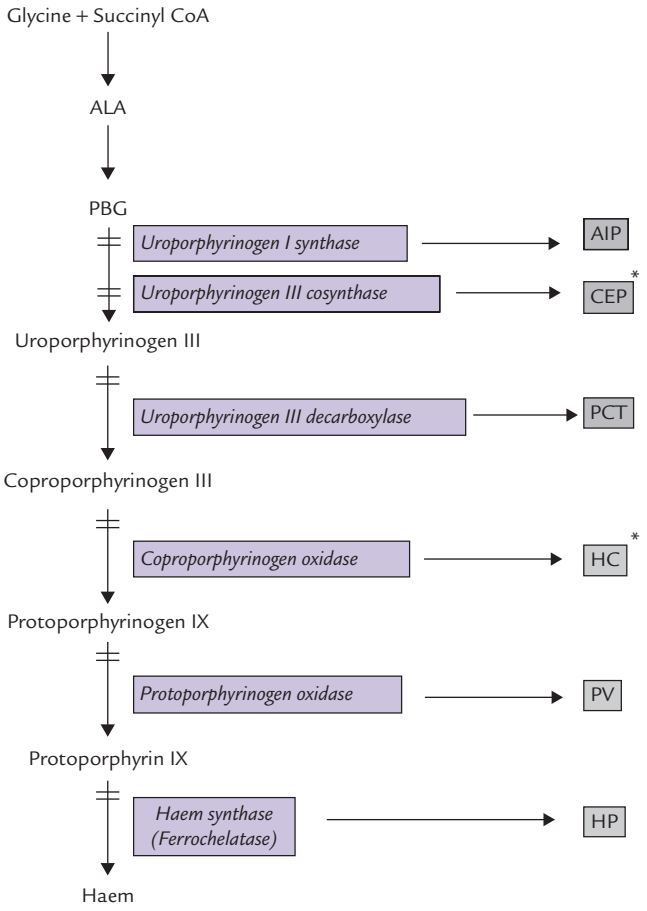


Fig. 16.7. Enzymatic blocks result in different types of inherited porphyrias (AIP = acute intermittent porphyria, CEP = congenital erythropoietic porphyria, PCT = porphyria cutanea tarda, HC = hereditary coproporphyria, PV = porphyria variegata, HP = hereditary protoporphyria. *CEP and HC are erythropoietic porphyrias, rest are hepatic porphyrias).

BOX 16.1

Regulation of Haem Biosynthesis is Different in Liver and Erythroid Cells

In liver the level of haem synthesis must be adjusted according to metabolic conditions; for example, the synthesis of the haem-containing cytochrome P-450 fluctuates with the need for detoxification. As discussed above, the principal regulator of the key enzyme *ALA synthase* is haem or its oxidation product (haematin or haemin), which controls the enzyme's activity through feedback-inhibition, repression-mechanism, and inhibition of transport of the enzyme molecules from cytosol into mitochondria. No such regulation is, however, required in erythroid cells where haem synthesis is a one-time event; it stops when the cell matures. In these cells, haem exerts quite a different effect: it speeds up rather than suppressing its own synthesis. Haem does so by causing stimulation of protein synthesis, including synthesis of the enzymes of haem biosynthesis. Haem also stimulates synthesis of globin, thereby ensuring that the haem and globin are synthesized in balanced amounts for assembly into haemoglobin. Moreover, the rate-limiting enzymes of haem biosynthesis in erythroid cells may be the *ferrochelatase* and *uroporphyrinogen I synthase* reactions rather than the *ALA synthase* reaction. This is desirable because when erythroid haem biosynthesis is switched on, all its steps must function at their maximal rates rather than any one step limiting the flow through the pathway.

Porphyrias are caused by partial deficiency of one of the haem synthesizing enzymes other than *ALA synthase*. Any one of the enzymes from *uroporphyrinogen I synthase* to *ferrochelatase* may be affected.

Salient Features of Porphyrias

1. All porphyrias are characterized by accumulation and excretion of the intermediate porphyrins of the haem biosynthetic pathway.
2. ***ALA synthase***, the key enzyme around which control of haem biosynthetics revolves, is **derepressed** in porphyrias because cellular haem levels are lower than normal.
3. **The route of excretion** of the accumulated porphyrins-urine or faeces-depends on their polarity. For example:
   - Uroporphyrin with its eight carboxyl groups is most water-soluble and is excreted almost entirely in urine.
   - Protoporphyrin, with only two carboxyl groups is excreted exclusively in faeces.
   - Coproporphyrin has four carboxylic groups and is excreted by either route.
4. Level of metabolic block determines **clinical features**:
   (a) **Early block:** If the metabolic block lies early in the pathway, so that synthesis of even the first tetrapyrrole ring does not occur, the defect may present as an **acute pain crisis**. This is because the accumulated metabolites (ALA and PBG) cause excitation of visceral pain fibres. ALA may block action of the inhibitory neurotransmitter (GABA) because of structural resemblance.
   (b) **Later block:** When the metabolic block lies beyond the formation of the first tetrapyrrole ring, **cutaneous manifestations** are the predominant features. The porphyrins accumulated in the patient's skin are responsible for these manifestations by the following sequence of events:
      - The conjugated ring system of porphyrins has the property of absorbing light near 400 nm (this distinguishing absorption band is called the **Soret band**) and then emit intense red light.
      - The emitted radiant energy generates superoxide radicals and other oxygen free radicals (OFRs).
      - The OFRs so formed are capable of damaging biological membranes and cause the release of destructive enzymes from lysosomes (Chapter 27).
      - The lysosomal enzymes in turn cause skin damage and scarring.

Treatment of porphyrias is aimed at alleviating the symptoms. Intravenous injection of haematin causes repression of *ALA synthase*, and therefore, prevents the accumulation of intermediates. In addition, the patient is advised to *avoid sunlight* and to increase dietary intake of *free radical scavengers* such as β-carotene and vitamin E.

The clinical manifestations of porphyrias are caused by abnormal accumulation of biosynthetic intermediates and may include abdominal pain, neuropsychiatric signs and cutaneous photosensitivitiy. According to the tissue in which the abnormality is expressed, hepatic and erythropoietic porphyrias can be distinguished.

Comparative properties of various types of porphyrias as listed in Table 16.1, depend on the molecular defect and the type of porphyrin that accumulates. Depending on the site where the enzyme deficiency is manifested, there are two major types: (a) erythropoietic porphyrias, and (b) hepatic porphyrias. (Porphyrias with both erythropoietic and hepatic abnormalities are also known.)

The two **erythropoietic porphyrias** (congenital erythropoietic porphyria and hereditary protoporphyria) are characterized by a build up of porphyrins in erythrocytes.

In **hepatic porphyrias**, liver is the primary organ affected and therefore the abnormality is expressed in this organ. Four types of hepatic porphyrias are known: *acute intermittent porphyria, porphyria cutanea tarda, hereditary coproporphyria* and *porphyria variegate* (Fig. 16.7).

All these disorders are inherited as *autosomal dominant trait*, except congenital erythropoietic porphyria, which is a genetically recessive disease.

A. Acute Intermittent Porphyria (AIP)

It is one of the most bizarre disorders a physician ever encounters. The defective enzyme is *uroporphyrinogen I synthase* and the metabolites accumulated are **ALA** and **PBG** (Fig. 16.7). The disease is marked by intermittent attacks of abdominal pain (which need to be differentiated from various surgical causes) and neuropsychiatric symptoms.

Clinical Features and Laboratory Findings

Accumulation of *ALA* and *PBG* in pharmacological amounts has been implicated in clinical manifestations. These metabolites act on the nervous system to cause a variety of signs and symptoms:

- **Abdominal pain** is caused by excitation of visceral pain fibres; constipation and cardiovascular changes

Table 16.1. Comparative properties of various types of porphyrias

	Acute intermittent porphyria (AIP)	Congenital erythropoietic porphyria (CEP)	Porphyria cutanea tarda (PCT)	Hereditary copro-porphyria (HC)	Variegate porphyria (VP)	Hereditary protoporphyria (HP)
Enzyme defect	*Uroporphyrinogen I synthase*	*Uroporphyrinogen III cosynthase*	*Uroporphyrinogen III decarboxylase*	*Coproporphyrinogen oxidase*	*Protoporphy-rinogen oxidase*	*Ferrochelatase*
Clinical findings						
Photosensitivity	No	Yes	Yes	Yes	Yes	Yes
Abdominal pain/ neurologic symptoms	Yes	No	No	Yes	Yes	No
Laboratory findings						
Urine	ALA, PBG ↑↑	*URO↑↑ *COPRO↑↑	URO ↑↑ COPRO↑↑	ALA, PBG↑↑ URO, COPRO↑↑	ALA, PBG↑↑ COPRO↑↑	Normal
Faeces	Normal	*COPRO↑	Normal	COPRO↑↑	PROTO↑↑ COPRO↑	PROTO↑↑
Red blood cells	Normal	*URO↑↑ *COPRO↑↑	Normal	Normal	Normal	PROTO↑↑ *COPRO↑

↑ = elevated; ↑↑ = highly elevated. *Type I isomers of uroporphyrin and coproporphyrin. Rest are type III isomers. HP and CEP are erythropoietic porphyrias, showing metabolic accumulation in RBC.

(including tachycardia and high blood pressure) are mediated by the autonomic nervous system.

- The involvement of peripheral nerves causes sensory or motor disturbances.
- Brain involvement is manifested by confusion, agitation or **seizures**.
- Excessive amounts of ALA and PBG are excreted in urine during and after acute attacks. However, the afflicted subject is **not photosensitive** because the metabolic block lies early in the pathway, and the porphyrin ring structure responsible for cutaneous photosensitivity is never formed **(Case 16.1)**.

Urinary excretion of ALA and PBG increases during acute porphyric attacks, but the urine sample is colourless (ALA and PBG being colourless compounds). But it increases in colour intensity on standing because of photo-oxidation of PBG to porphobilin.

Treatment

General treatment guidelines for porphyrias were discussed earlier. An appropriate treatment for AIP consists of withdrawal of any offending drug and infusion of **haematin**; the latter represses synthesis of *ALA synthase*. A **carbohydrate rich diet** benefits the patients by repressing (by an unknown mechanism) the synthesis of *ALA synthase*.

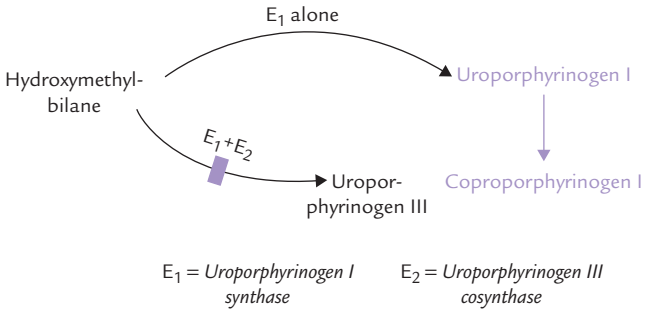
An acute porphyric attack is precipitated by starvation and by drugs, e.g. barbiturates, phenytoin and other drugs that induce synthesis of cytochrome P-450, such drugs therefore must be avoided.

King George III, who ruled England during the American revolution (1773–76), and who was widely portrayed as having fits of madness, had attacks characteristic of AIP. Some of his precipitate actions were taken during acute (porphyric) episodes, which might have infuriated the colonists, ultimately leading to American independence.

B. Congenital Erythropoietic Porphyria

Activity of *uroporphyrinogen III cosynthase* is severely depressed in this rare disorder, so that synthesis of type III isomers from hydroxymethylbilane is decreased. Instead, the hydroxymethylbilane is converted to uroporphyrinogen I, which, under the influence of *uroporphyrinogen decarboxylase*, may be converted to coproporphyrinogen I.

Both uroporphyrinogen I and coproporphyrinogen I are oxidized to their corresponding porphyrins, which are then accumulated in the tissues and excreted in the urine (Table 16.1). Their accumulation produces a pink to dark red colour in teeth, bones, and urine. Dark red urine (pot-wine appearance), and red-brown teeth (erythrodontia) are pathognomonic.



Extreme cutaneous sensitivity to light relates to accumulation in the skin of various type I porphyrins. Itching, burning, redness and swelling occur initially, and may lead to hyperpigmentation and scarring. The latter leads to a typical facial deformity, referred to as **monkey faeces.**

The deficient production of the type III isomer diminishes the regulatory effect on *ALA synthase*, which further increases levels of the type I isomers. Excessive amounts of porphyrins in erythrocytes may produce haemolysis.

C. Porphyria Cutanea Tarda

It is the most common porphyria with an incidence of 1 in 25000 (the second commonest form of AIP has a much lower incidence—1 in 50000). It is mainly a skin disorder caused by a decreased activity of liver *uroporphyrinogen III decarboxylase* (Fig. 16.7). The patient's urine contains large quantities of uroporphyrin III and some uroporphyrin I. Mild to severe photosensitivity and liver diseases are other clinical presentations. Acute episodes can be precipitated by *alcohol intake* or oestrogen therapy.

D. Others

Enzyme defect in hereditary coproporphyria (HC), variegate porphyria (VP) and hereditary protoporphyria (HP) and their clinical features and laboratory findings are shown in Table 16.1. Some important features of these conditions are:

1. Clinical manifestations in **HP** are similar to AIP, though much less in severity.
2. **VP** (incidence 1 in 100,000) is especially prevalent among white population in South Africans.
3. Hereditary protoporphyria (HP), generally a benign condition, is characterized by accumulation of protoporphyrin IX in maturing reticulocytes and young erythrocytes, and in liver; the latter may lead to severe liver damage. High levels of protoporphyrin are found in plasma, erythrocytes and faeces, but the urine is typically free of any porphyrin. When a smear of RBCs is exposed to fluorescent light, they exhibit red fluorescence.

Finally, it may be mentioned that porphyrias often make their appearance after puberty. This has been associated with the appearance of *5-β-steroid reductase*, as discussed (**Case 2**).

E. Acquired Porphyrias

Porphyrias can occur in situations other than the inherited conditions described above, such as **lead poisoning** and **iron deficiency.**

1. *Lead poisoning* is the commonest cause of acquired porphyria. Lead inhibits the uptake of iron by immature red cells and has direct inhibitory effect on two enzymes of the porphyrin pathway: *ALA dehydratase* and *ferrochelatase*. Consequently there is:
   - an increased excretion of ALA in urine (but not PBG);
   - an increase in urine and fecal coproporphyrins and protoporphyrins.

> The haem biosynthetic pathway is disrupted in porphyrias and related conditions, which can be caused by inherited enzyme deficiency (any one except *ALA synthase*), iron overload, or lead poisoning

2. *Iron deficiency:* Patients with iron deficiency can use zinc, instead of iron, as a substrate for *ferrochelatase*. Red cell lysates in such instances contain increased quantities of **Zn-haemoglobin**. In addition, red cells from iron deficient patients also contain increased amounts of protoporphyrin IX. *Both Zn-haemoglobin and protoporphyrin IX determinations are therefore used in the diagnosis of iron deficiency anaemia.*

   Coproporphyrinuria: A small (less than twofold) increase in urinary coproporphyrin is most commonly caused by problems unrelated to haem synthesis, such as *liver disease*, *acute illness* or *exposure to toxic chemicals*.

Diagnosis of Porphyrias

Porphyrias can be diagnosed biochemically by (a) measurement of metabolites of the haem synthetic pathway, (b) assay of the relevant enzymes, and (c) DNA-based testing.

(a) The **metabolite measurements** are performed in plasma, urine, faeces and erythrocytes (Table 16.1). α-Aminolevulinic acid and porphobilinogen are quantified by colourimetric tests, and, uroporphyrin

and coproporphyrin are determined fluorometrically. The most important test for diagnosis of acute porphyrias is the measurement of PBG in urine and plasma.

(b) Enzyme assay—only one enzyme of the synthetic pathway, *uroporphyrinogen I synthase*, is routinely measured by clinical laboratories. Its activity is decreased to about 50% of normal in the red cells of most individuals with acute intermittent porphyria, whether the disease is in latent or in an acute phase.

(c) DNA-based testing of porphyrias is more definitive than conventional biochemical tests. The test is possible now as all of the genes that encode enzymes of the haem synthetic pathway have been identified and their coding sequences determined. This has led to thwe discovery of mutations that cause each of the porphyrias.

VI. Haem Breakdown

End products of haem breakdown are **bile pigments**: **bilirubin** and **biliverdin**. About 300 mg of bilirubin is produced per day, a major fraction of which (250 mg) is derived during turnover of haemoglobin. Another 50 mg of bilirubin is formed from other haem-containing proteins, such as myoglobin and cytochromes.

Lifespan of the erythrocyte is 120 days. Old cells are removed from blood circulation and degraded in the reticulo-endothelial system, liberating the haemoglobin. The protein portion of haemoglobin is hydrolyzed to its constituent amino acids, and channeled into the body's amino acid pool. The iron so liberated is re-utilized, and the iron free porphyrin ring is converted to bilirubin (Fig. 16.8).

Note: About 85% of the haem degraded each day is provided by old erythrocytes; and the rest of the 15% comes from the immature erythrocytes and other haem-containing proteins.

A. Bilirubin Metabolism

Production of bilirubin and its metabolism, and the abnormalities of bile pigment metabolism are described in the subsequent sections.

1. ***Bilirubin production:*** Bilirubin, the end product of haem degradation, is produced in reticulo-endothelial cells in two steps:

(a) The microsomal *haem oxygenase* system in the reticulo-endothelial cells cleaves the porphyrin ring of haem (Fig. 16.9). Consequently, the cyclic tetrapyrrole ring structure of haem is converted to a linear tetrapyrrole. NADPH is required as a coenzyme in this step, and a ferrous iron and a carbon monoxide molecule are released. This results in the formation of **biliverdin,** a green pigment.

(b) Biliverdin is reduced to a red-orange pigment, **bilirubin,** by the enzyme *biliverdin reductase*; NADPH acts as donor of the reducing equivalents.

Changing colour of a bruise is due to appearance of biliverdin followed by bilirubin.

Further metabolism of bilirubin occurs in the liver and the intestine.

2. ***Transport of bilirubin to liver:*** Bilirubin forms a non-covalent complex with albumin in plasma, which is transported to the liver (Fig. 16.10). Formation of such a complex is necessary for the bilirubin transport since hydrophobic nature of bilirubin restricts free movement of the unbound-bilirubin in aqueous plasma. The complex has predominantly polar character (because of polar nature of albumin) which enables it to move freely in plasma.

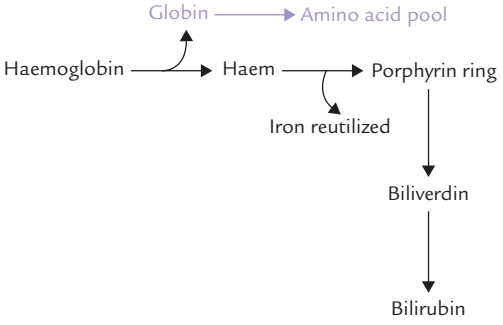


Fig. 16.8. Breakdown of haemoglobin in reticulo-endothelial cells of liver, spleen and bone marrow. (Total amount of haemoglobin in adult body is about 750 g and the daily turnover rate is around 6 g.)

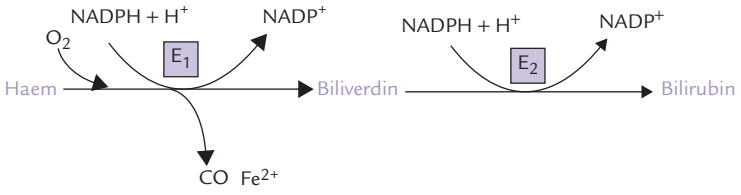


Fig. 16.9. Breakdown of haem. (E_1 = *haem oxygenase*, E_2 = *biliverdin reductase*).

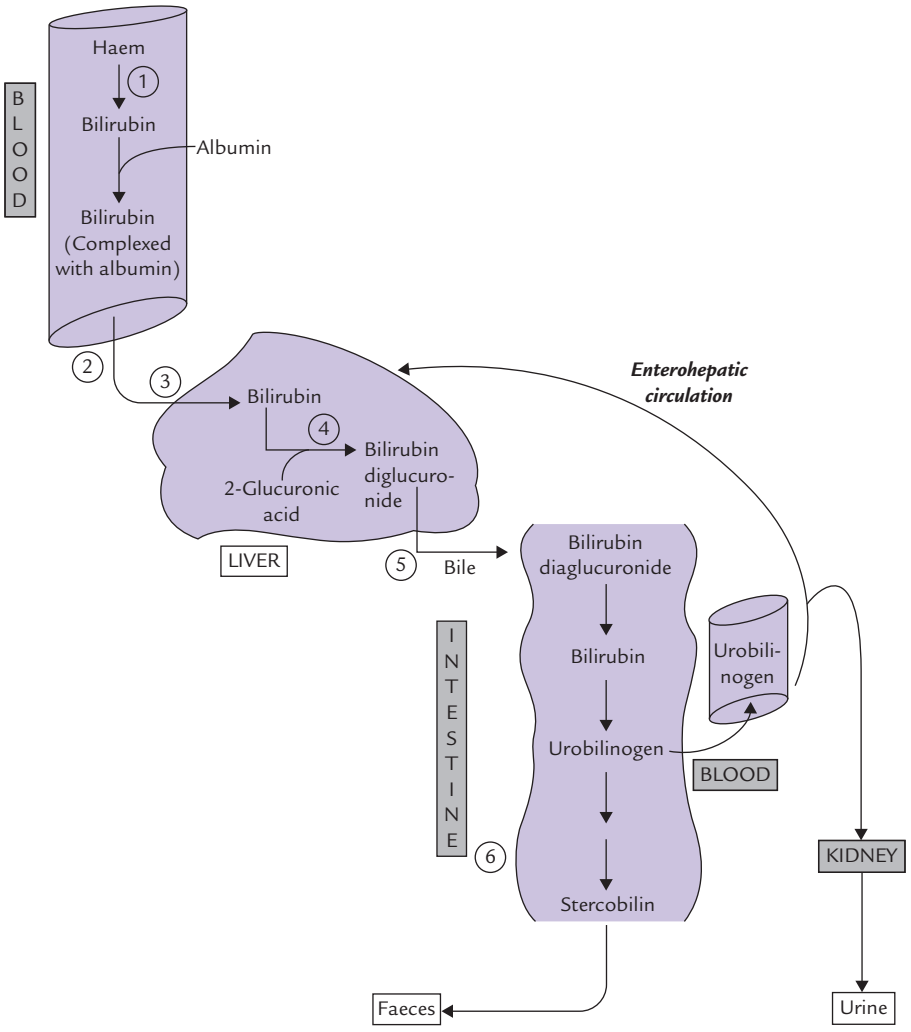


Fig. 16.10. Bilirubin metabolism. (1) Production in reticuloendothelial system. (2) Transport to liver. (3) Hepatic uptake. (4) Conjugation with glucuronic acid. (5) Excretion into bile. (6) Deconjugation and reduction in intestine.

On reaching the liver, bilirubin becomes free from its non-covalent association with albumin and enters the hepatocytes.

Albumin has two bilirubin-binding sites:

- a high-affinity site and
- a low-affinity site.

Bilirubin is first taken up by the high-affinity site, and only when all such sites are saturated, the excess bilirubin is bound to the low affinity site. In 100 ml plasma, albumin can take up to 20 mg bilirubin at its high affinity site. So, bilirubin associates with low affinity site only when serum bilirubin level crosses 20 mg/dL. Certain anionic drugs, such as *sulphonamides, penicillin* and *salicylates* can displace bilirubin from albumin, thus increasing the unbound-bilirubin concentration in plasma. This may have hazardous consequences in newborn infants because the unbound-bilirubin can cross the blood-brain barrier and enter the central nervous system. In the neurons, bilirubin has potential to cause neuronal tissue damage: the condition is known as **bilirubin encephalopathy.**

3. ***Hepatic uptake of bilirubin:*** When the albumin-bilirubin complex reaches the sinusoidal surface of the liver, the bilirubin is taken up by hepatocytes. The uptake is an **active** (*energy-requiring*) *process, which requires mediation* of the carrier protein. Within the hepatocyte, two proteins: **ligand** in **Y** and **Z** protein, bind the bilirubin.

4. ***Conjugation of bilirubin in hepatocytes:*** Inside the hepatocyte, two molecules of glucuronic acid are attached with bilirubin to form bilirubin diglucuronide, also called **conjugated bilirubin** (or **direct bilirubin**). UDP-glucuronic acid serves as activated donor of the glucuronic acid.

$$\text{Bilirubin (unconjugated)} + \text{UDP-Glucuronic acid} \longrightarrow \text{Bilirubin monoglucuronide} + \text{UDP}$$

$$\text{Bilirubin monoglucuronide} + \text{UDP-Glucuronic acid} \longrightarrow \text{Bilirubin diglucuronide} + \text{UDP}$$

Both these steps are catalyzed by *bilirubin glucuronyl transferase*. Because glucuronic acid is highly polar, its attachment with bilirubin imparts highly polar characteristics to the latter, so that it can mix with bile and subsequently get excreted through the hepato-biliary route (Fig. 16.10).

5. ***Excretion of conjugated bilirubin into bile:*** The hepatocytes secrete the conjugated bilirubin into the bile canaliculi, from where it reaches the bile duct (Fig. 16.10). Since concentration of bilirubin in bile is very high, the above transport occurs against a steep concentration gradient, and therefore, requires input of energy. This step is '***rate-limiting***' in the bilirubin metabolism. Being energy dependent, it is susceptible to impairment in the liver diseases.

The degradation of haem in reticulo-endothelial cells of the spleen or other organs involves oxidation to biliverdin first, which is reduced to bilirubin. The liver conjugates bilirubin to bilirubin diglucuronide and this product is actively secreted into bile.

6. ***Fate of conjugated bilirubin in intestine:*** The conjugated bilirubin enters the gut through bile, where it is acted upon by the intestinal bacteria.

- The first step is deconjugation by *bacterial glucuronidase*, which removes the polar glucuronate groups from the conjugated bilirubin.
- The free bilirubin so formed is further reduced to a colourless tetrapyrrole **urobilinogen** (URO).
- A part of urobilinogen (20%) is absorbed from the gut into the portal circulation. The rest remains within intestinal lumen, where it is reduced further.
- The reduction of the URO leads to the formation of **mesobilinogen** and **stercobilinogen**.

In small intestine, bilirubin is deconjugated and reduced to three substances, collectively called urobilinogens, that differ from each other in their degrees of oxidation.

The stercopbilinogen is mostly excreted in faeces (250–350 mg/day). Upon exposure to atmospheric air, it is oxidized to a coloured product, stercobilin. The *characteristic brown colour of stools is accounted by stercobilin*. Pale stools may, therefore, indicate biliary obstruction.

The enterohepatic circulation: The part of urobilinogen, which is absorbed into the portal circulation, can take two alternative routes:

(i) Some part of it is returned to the liver. Through hepatobiliary route, it reaches the intestine again. This is *enterohepatic circulation*, as depicted in Figure 16.10.

(ii) The remaining urobilinogen enters the systemic circulation and transported to kidneys. It is oxidized to a coloured product, **urobilin,** and excreted in urine. *The normal colour of urine is accounted by the urobilin*.

The reduced forms of bilirubin (urobilinogen, mesobilinogen and stercobilinogen) are colourless compounds. The oxidized forms—*urobilin*, *mesobilin* and *stercobilin*—are coloured (**bile pigments**). Colour of faeces becomes darker on standing because of further oxidation, which forms more of stercobilin. When intestinal bacterial flora is decreased by antibiotic therapy, bilirubin is not adequately reduced to urobilinogen. It is rather oxidized to biliverdin in large intestine, which imparts green colour to stools.

Plasma Bilirubin

The human bloodstream contains both unconjugated (albumin bound) and conjugated bilirubin. Their total concentration ranges from 0.2 mg/dL to 1.0 mg/dL, of which conjugated bilirubin accounts for 0.1–0.4 mg/dL.

The *conjugated bilirubin* is also known as **direct reacting bilirubin** because, being water-soluble, it gives purple colour immediately with Ehrlich's diazo reagent. The unconjugated bilirubin, on the other hand, is water-insoluble and has to be extracted with methanol or ethanol for reacting with the diazo reagent. Therefore, it is also called **indirect reacting bilirubin**.

The conjugated bilirubin is termed direct bilirubin and the unconjugated bilirubin is called indirect bilirubin.

Normally, little, if any, conjugated bilirubin is found in the urine. But a 24-hour urine sample may contain 0.5–4.0 mg urobilin. Faeces normally contains 250–350 mg of stercobilinogen and stercobilin per day.

VII. Jaundice

Rise in serum bilirubin level above normal (*hyperbilirubinaemia*), causes it to get deposited in tissues. A yellow discolouration of skin and sclera results and the condition is known as jaundice. Hyperbilirubinaemia exists when the plasma bilirubin concentration exceeds 1.0 mg/dL. Levels between 1.0–3.0 mg/dL indicate **subclinical jaundice** (*latent jaundice*). When the bilirubin level exceeds 3.0 mg/dL, it diffuses into tissues to produce yellowish discolouration of skin, sclera, conjunctiva and mucous membrane. This results in **icterus** (or **clinical jaundice**).

Elevation of serum bilirubin level, known as hyperbilirubinaemia, results in jaundice or icterus when it crosses 3 mg/dL.

Hyperbilirubinaemia occurs when there is imbalance between bilirubin production and excretion. There are three main reasons why this imbalance may occur:

1. In **haemolysis**, the increased haemoglobin breakdown produces bilirubin which overloads the hepatic conjugating mechanism.
2. In **hepatic diseases** there is important of the conjugating mechanism.
3. **Obstruction** in the biliary drainage.

Based on these underlying causes, the jaundice is conventionally classified as **haemolytic**, **hepatic** and **obstructive types**, respectively.

A. Haemolytic Jaundice

Liver has enormous capacity for conjugating bilirubin. In an adult, about 300 mg of bilirubin that is produced each day reaches liver, where it is conjugated. Liver is capable of handling about 10 times as much load of bilirubin (3 g/day). However, in case of **massive haemolysis**, there is an increased bilirubin production to such an extent that the liver is not able to conjugate it fully. A portion of the bilirubin remains unconjugated, which results in rise in serum concentration of unconjugated bilirubin, i.e. **unconjugated hyperbilirubinaemia**). Some other common causes of increased rate of haemolysis are given in Table 16.2. Unconjugated hyperbilirubinaemia, when rapidly rising in neonates, should be carefully monitored (discussed later).

Hyperbilirubinaemia of haemolytic diseases rarely exceeds 3–4 mg/dL because of enormous capacity of healthy adult liver to metabolize bilirubin. The biliary excretion of conjugated bilirubin is increased in proportion to the severity of haemolysis. Levels of the compounds derived from conjugated bilirubin (urobilin in urine and stercobilin in faeces) are thereby elevated (Table 16.3).

B. Hepatic Jaundice

Jaundice occurs due to diminished functional capacity of liver. Various causes for the same are outlined in Table 16.2.

Plasma levels of both the conjugated and the unconjugated bilirubin rises in this type of jaundice (Table 16.3). This is referred to as a "**biphasic rise**", the reasons for which are as below:

- The damaged hepatocytes have impaired capacity to take up the circulating bilirubin and to conjugate it. This leads to increased plasma level of free **(unconjugated) bilirubin**.
- The inflammatory oedematous cells compress intracellular biliary canaliculi and this produces an element of obstruction. Consequently, the **conjugated bilirubin**

Table 16.2. The major causes of jaundice

Type	Cause	Clinical example	Frequency
Haemolytic	Haemolysis	G6PD deficient	Uncommon
		Autoimmune	Uncommon
		Abnormal haemoglobin	Depends on region
Hepatic	Infection	Hepatitis A, B, C,	Common/very common
	Chemical/drug	Acetaminophen Alcohol	Common
	Genetic errors	Gilbert's syndrome	Uncommon: 1 in 20
		Crigler–Najjar syndrome	Very rare
		Dubin–Johnson syndrome	Very rare
		Rotor's syndrome	Very rare
	Genetic errors: specific proteins	Wilson's disease	1 in 200,000
		α_1 Antitrypsin	1 in 1000 with genotype
	Autoimmune	Chronic active hepatitis	Uncommon/rare
	Neonatal	Physiologic	Very common
Obstructive	Intrahepatic bile duct obstruction	Drugs	Common
		Primary biliary cirrhosis	Uncommon
		Cholangitis	Common
	Extrahepatic bile duct obstruction	Gall stones	Very common
		Pancreatic tumour	Uncommon
		Cholangiocarcinoma	Rare

Table 16.3. Comparative features of haemolytic, hepatic and obstructive jaundice

	Hemolytic jaundice	Hepatic jaundice	Obstructive jaundice
Blood, unconjugated bilirubin	Increased	Increased	Normal
Blood, conjugated bilirubin	Normal/mild increase	Increased	Increased
Urine, bilirubin	Nil	Present	Present
Fecal stercobilin	Increased	Very low	Nil
Urine, urobilin	Increased	Very low	Nil
Urine, bile salts	Nil	Nil	Present
Blood, ALP	Normal	Increased	Very high
Blood, *transaminases* (ALT, AST)	Normal*	Very high	Increased

*AST increases in case of haemolysis because of its high concentration in red blood cells. However, activity of ALT, the other major *transaminase*, remains normal. ALT = *alanine transaminase*, AST = *aspartate transaminase*, ALP = *alkaline phosphatase*.

formed in the hepatocytes may leak out of the damaged cells into the blood circulation. The conjugated bilirubin, therefore, appears in blood circulation in increased amount.

Since less conjugated bilirubin is produced in hepatocytes, the amount that enters the gut is also less, leading to decreased formation of urobilinogen and stercobilinogen. As a result, *urinary and faecal excretion of urobilin* and *stercobilin falls*.

Note: In some cases there is increased urobilin excretion in urine because the damaged hepatocytes are not able to internalize urobilinogen from blood stream (see enterohepatic circulation).

C. Obstructive Jaundice

In this case, jaundice results due to *obstruction of the intrahepatic or extrahepatic bile duct* (Table 16.2), though the bilirubin-conjugating capacity of the liver is more or less intact. The obstruction blocks the passage of the conjugated bilirubin into the intestine. It, therefore, accumulates in the hepatocytes and subsequently overflows into the blood circulation. As a result, concentration of the **conjugated bilirubin** rises in plasma. Since conjugated bilirubin is polar in nature, it can be excreted into the urine. Therefore, in obstructive jaundice bile pigments can be detected in urine. This accounts for the yellow-brown colour of urine.

Since amount of conjugated bilirubin carried by bile into the intestine is decreased, urobilinogen formation also decreases correspondingly. As a result, urinary urobilin and fecal stercobilin excretions fall. These pigments may be totally absent in case of complete obstruction (Table 16.3). So the stools no longer have brown colour, but appear clay coloured. Liver function is unimpaired in acute cases but longstanding cholestasis causes irreversible liver damage.

Conjugated bilirubin is excreted in urine due to its polar nature (obstructive jaundice), but unconjugated bilirubin is not excreted (haemolytic jaundice) because (i) It is non-polar, and (ii) it is complexed with albumin. This explains presence of bile pigments in urine in obstructive and hepatic jaundice, but their absence in haemolytic jaundice.

D. Physiologic Jaundice of the Newborn

In neonates, transient jaundice is common, particularly in premature infants, and it occurs due to immaturity of the enzymes involved in bilirubin conjugation. There is accumulation of unconjugated bilirubin in blood circulation, but the serum bilirubin level generally does not increase above 5 mg/dL. Normally, about 50% of the normal babies become jaundiced 48 hours after birth in this manner. *The jaundice normally disappears by the second week of life*. However, if it is prolonged for more than 2 weeks, the infant must be treated by phototherapy.

The unconjugated bilirubin is transferred from its loose association with albumin to other proteins, such as those in cell membranes. It is neurotoxic, and if levels rise too high (15–20 mg/dL) in neonates, permanent brain damage can occur. Since blood–brain barrier is not fully mature till the baby completes one year of life, the excess bilirubin can readily cross it to reach the brain. Within brain, bilirubin deposition in the basal ganglia, hippocampus, cerebellum, etc. can cause irreversible brain damage, a condition known as **kernicterus** *kern:* (German-nucleus). Fits, spasticity, and mental retardation are the predominant clinical features. It is often fatal and leaves behind permanent neurologic deficit if the baby survives.

Mild hyperbilirubinaemia (< 5 mg/dL) occurs in about 50% of newborn due to immature or sluggish hepatic enzymes, that cause bilirubin conjugation. It needs to be treated only when the serum bilirubin levels cross 15–20 mg/dL by phototherapy.

Role of Phototherapy and Phenobarbital

Treatment for neonatal jaundice by phototherapy involves putting the baby under an ultraviolet light lamp. This induces the isomerization of water-insoluble bilirubin (termed the Z-isomer) to the water soluble E-isomer. The latter can then be cleared by the kidneys. The administration of phenobarbital (to induce conjugating system of liver) is indicated when the bilirubin level stays dangerously high despite phototherapy.

Blood exchange transfusion is required in extreme cases, when sustained increase in serum bilirubin above 20 mg/dL occurs.

Haemolytic Disease of the Newborn

Any haemolytic condition in the neonatal period is potentially dangerous, but the most serious is **erythroblastosis fetalis** or "haemolytic disease of newborn". It occurs because of rhesus incompatibility: when a Rh negative mother conceives an Rh-positive baby. The Rh antigen crosses placental barrier to enter maternal circulation to induce synthesis of maternal IgG antibody. These antibodies are transferred across placenta and enter the fetal circulation, where they destroy red cells.

In Rh incompatibility, the first child often escapes, but in the second pregnancy the maternal antibodies enter fetus in large quantities to cause extensive red cell destruction even before birth. Sometimes the child is born with severe haemolytic disease the erythroblastosis fetalis. This may require exchange transfusion immediately after birth or even in the fetus before birth.

Conjugated hyperbilirubinaemia is typical for biliary obstruction (cholestasis); unconjugated hyperbilirubinaemia occurs in haemolytic conditions and in otherwise normal newborns; and a mixed hyperbilirubinaemia occurs in hepatocellular damage, e.g. in viral hepatitis and toxic liver damage.

E. Genetic Causes of Jaundice

There are a number of genetic disorders, referred to as **inherited hyperbilirubinaemia,** that impair bilirubin conjugation or secretion. Gilbert's syndrome, Crigler–Najjar syndrome, Dubin Johnson's syndrome and Rotor syndrome are some commonly described disorders. All except Gilbert's syndrome are extremely rare disorders; the latter affects 5% of the population.

Gilbert's syndrome: It is an autosomal dominant trait characterized by *mild unconjugated hyperbilirubinaemia* that is harmless and asymptomatic. Some cases appear to involve a defect in hepatic uptake, but more commonly there is a modest impairment in conjugating enzyme, *glucuronyl transferase*, activity. There are no other signs, symptoms or laboratory abnormalities, and the prognosis is excellent.

Crigler–Najjar syndrome: This results due to complete absence or marked reduction in bilirubin conjugation system. Two types are known:

Type I: It is a rare variety in which there is a complete absence of the *glucuronyl transferase* activity, resulting in *severe unconjugated hyperbilirubinaemia* that presents at birth. Most affected infants die from bilirubin-induced neurological damage before reaching 1 year of age.

Type II: Patients with type II disease have a partial deficiency of *glucuronyl transferase*. The serum bilirubin concentration is lower than in type I, and neurological complications are unusual.

Inheritance is autosomal recessive for the type I disorder and autosomal dominant for type II.

Dubin–Johnson and Rotor syndrome: These syndromes impair the biliary secretion of conjugated bilirubin, and therefore, cause *conjugated hyperbilirubinaemia*, which is usually mild. The Dubin–Johnson syndrome is distinguished by the dark pigment that accumulates in hepatocytes. Inheritance is autosomal recessive and the prognosis is excellent.

Exercises

Essay type questions

1. Describe haem synthesis and its regulation.
2. Describe the formation and excretion of bilirubin. Explain the changes in serum bilirubin in different types of jaundice.

Write short notes on

1. Acute intermittent porphyria
2. Thalassaemia
3. Methaemoglobin
4. Cyanmethaemoglobin
5. Haemoglobin S
6. Kernicterus
7. Crigler–Najjar syndrome

CLINICAL CASE

CASE 16.1 Severe abdominal pain and abnormal colouration of urine

As 22-year-old woman reported in hospital emergency with abdominal pain. The pain was poorly localized but of severe intensity lasting for several hours without relief. The pain was accompanied by nausea and vomiting. The patient had bilateral leg pain with paraesthesia, and a generalized weakness of such a degree that she found it difficult to stand. On examination, she appeared mildly dehydrated, irritable, and showed psychotic behavior. Abdomen was distended and showed scars of some surgery performed earlier. No localized abdominal tenderness or rebound pain was, however, observed. Neurological examination showed decreased motor strength in both legs and absence of reflexes. She was being treated with phenytoin for her neuropsychiatric complaints.

Further enquiry revealed that the patient had been suffering from acute painful episodes for about five years. Appendisectomy was performed three years earlier during one of such episodes, but her appendix (and also the ileum) was found to be normal. Typically, the pain would start about 4–7 days prior to menstrual period, would last for a few days, and ultimately resolve with the cycle. No gynaecological abnormality was, however, detected after carrying out detailed pelvic examination and other investigations, such as ultrasonography and laparoscopy. Blood sample was sent to the haematology and the biochemistry laboratories. All test results were normal, including the liver function tests (LFT) and serum activities of the pancreatic enzymes. A CSF sample was also obtained by lumbar puncture and analyzed. No abnormality was, however, revealed. Analysis of stool and urine sample also showed normal result. On electrophoresis, the γ-globulin fraction in plasma was reported to be mildly increased.

The urine was abnormally dark coloured; a peculiar finding was that it turned deep red upon standing at room temperature on the laboratory bench. Based on these observations, a tentative diagnosis was made and the phenytoin was withdrawn. Treatment with intravenous glucose and haematin was started. The patient showed remarkable improvement with these measures, which confirmed the above diagnosis.

Q.1. Name various conditions that cause abnormal coloration of urine. Which one of these could have affected the patient?

Q.2. How does effectiveness of the intravenous glucose and haematin therapy confirm the above tentative diagnosis?

Q.3. What is the rationale behind withdrawing phenytoin?

Q.4. Why was a sample of CSF analyzed?

Q.5. Which type of porphyria do you expect in this patient? What further investigations would you perform to confirm this?

Q.6. Provide biochemical explanation for the patient's symptoms.

CHAPTER

17

OXYGEN TRANSPORTERS: HAEMOGLOBIN AND MYOGLOBIN

Molecular oxygen is sparingly soluble in body fluids. This limits the amount that can be transported in physical solution to less than 30 g per day, whereas the body requires approximately 500 g of molecular oxygen per day. An effective oxygen transport is made possible by haemoglobin, the oxygen-binding protein, present in the erythrocytes. Because of the oxygen-binding properties of this protein, blood can dissolve approximately 70 times more oxygen than the plasma free of haemoglobin can do.

Just as haemoglobin is the oxygen-binding protein in erythrocytes, the myoglobin binds oxygen in muscle tissue. It permits storage of oxygen in muscles. The stored oxygen is released when it is required, for example, during vigorous muscular exercise. Binding of oxygen to haemoglobin and myoglobin is called **oxygenation.**

$$\text{Protein} + O_2 \rightarrow \text{Oxygenated protein}$$

It is a non-enzymatic, reversible process. Therefore, oxygen, when available in plenty, binds to the oxygen-binding proteins. Conversely, oxygen is released from these proteins when it is scarce. None of the constituent amino acids of these proteins has any binding affinity for molecular oxygen. Therefore, the oxygen-binding proteins have to employ a non-protein component to serve as the oxygen-binding site. This component is **haem** in haemoglobin as well as in myoglobin.

In this chapter, general characteristics and oxygen-binding properties of haemoglobin and myoglobin are described. After going through this chapter, the student should be able to understand:

- General properties of haemoglobin, such as structure of haem, organization of various globin chains, positioning of iron in the tetrapyrrole ring and mode of binding of oxygen; structural and functional peculiarities of various types of haemoglobins.
- Behaviour of haemoglobin as an allosteric oxygen binder, sigmoid shape of the oxygen saturation curve of haemoglobin, Bohr effect, and role of 2,3-BPG in oxygen off-loading.
- Aetiology and oxygen-binding peculiarities in various types of haemoglobinopathies.
- Oxygen-binding properties of myoglobin and its suitability to store oxygen in muscles.

I. Haemoglobin

Haemoglobin is a **tetrameric metalloprotein** consisting of four polypeptide chains, each with its own haem. In addition to carrying oxygen from lungs to the peripheral tissues, it carries carbon dioxide from actively metabolizing tissues to lungs. Haemoglobin possesses **buffering property** as well.

Upon oxygenation, the haemoglobin is converted to oxyhaemoglobin. This is not a simple metal–gas interaction, but is a dynamic process subject to various **allosteric controls**. Binding of other ligands such as carbon dioxide (and H^+) to haemoglobin is also regulated by allosteric mechanism. The allosteric mechanism is very well understood in case of haemoglobin, and discussed in detail in this chapter.

A. Basic Structure

Humans have several types of haemoglobins. All types are made up of two **α chains** and two **non-α chains** (*β*, *δ* or *γ*) with a total molecular weight of approximately 65,000. These chains, collectively termed globin polypeptides, differ in primary structures but have similar secondary and tertiary structures. The differences in tertiary structure among them are critical to the functional properties of each. A non-covalently bound haem group with a molecular weight of 616 is associated with each chain (Fig. 17.1). It should be noted that all globin polypeptides are **homologous proteins**, having arisen from the same ancestral gene.

Globin chains: The globin chains present in the major adult haemoglobin, called **HbA,** are α and β, and its subunit structure is $\alpha_2\beta_2$. The minor adult haemoglobin (**HbA_2**) and fetal haemoglobin (**HbF**) also have two α-chains but instead of the β-chains they have δ-chains and γ-chains respectively (Fig. 17.2).

The α-chain has 141 amino acids (MW 15,750) and the β-chain has 146 amino acids (MW 16,500). Like the β-chains, the other two non-α-chains (δ and γ) also have 146 amino acids and are structurally related.

The β- and δ-chain differ in 10 and the β- and γ-chains in 39 of their 146 amino acids. The α- and β-chains are not as closely related: they are identical only in 64 of their amino acids.

Secondary and tertiary structures of globin polypeptides: The secondary structure of all globin polypeptides consists largely of α-*helical segments* and short stretches of various bends and *interhelical segments*. α-chain has 7, and β-chain has 8 helical segments, labelled as A (N-terminal) through H. They are organized into a tightly packed, nearly spherical, globular tertiary structure, shown in Figure 17.1.

Like most soluble proteins, haemoglobin has a relatively hydrophilic surface and a hydrophobic interior. *Polar amino acids are located almost exclusively on the exterior surface* of the globin polypeptides, and contribute to their remarkably high solubility in cytoplasm of erythrocytes (5.2 mmol/L). This permits building up of a high

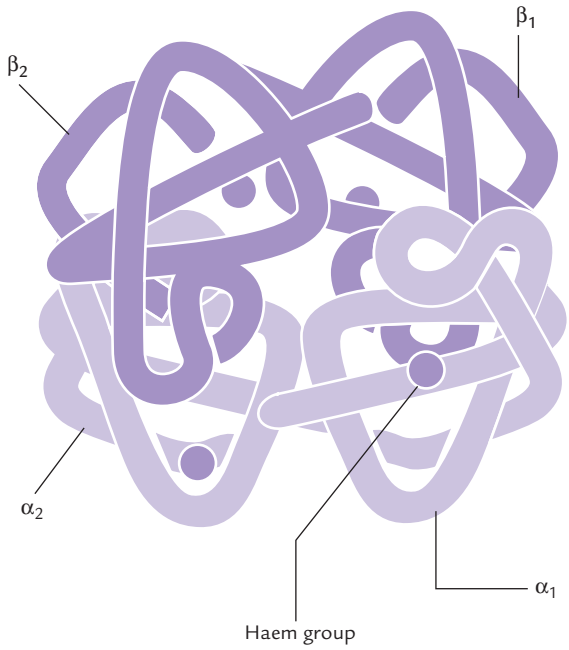


Fig. 17.1. Structure of HbA molecule consisting of two α-chains and two β-chains, each with a non-covalently bound haem group.

Type of haemoglobin	*Structure*	*Composition*	*Importance*
Major adult (HbA)	β, α, α, β	$\alpha_2\beta_2$	97% of adult haemoglobin
Minor adult (HbA_2)	δ, α, α, δ	$\alpha_2\delta_2$	2–3% of adult haemoglobin
Fetal (HbF)	γ, α, α, γ	$\alpha_2\gamma_2$	Major haemoglobin in second and third trimester of pregnancy

Fig. 17.2. Subunit structure and importance of various haemoglobin types. A subunit consists of a polypeptide chain and its haem group (● = haem, ◡ = polypeptide chain).

intracellular concentration of Hb within red cells, which appear as "**bags" filled with Hb**. Amino acids having both polar and non-polar side chains, such as tyrosine, threonine and tryptophan, are oriented with their polar functions toward the protein's exterior. *Hydrophobic amino acids are buried within the interior*, where they stabilize the folding of the polypeptide and binding of the iron-porphyrin prosthetic group.

Polar amino acids, located towards the exterior, make haemoglobin highly soluble, whereas the non-polar amino acids are buried in the interior and influence folding.

The only exception to this general distribution of amino acid residues in globins are the two histidine residues, termed **proximal**- and **distal-histidines,** which are oriented perpendicular to and on either side of the planar haem prosthetic group. They create a *hydrophobic pocket* in which the haem-group resides.

Haem group: A haem group is associated with each globin chain in the hydrophobic interior (tucked between the E and the F helices). Haem consists of a porphyrin ring system with various side chains: methyl, vinyl and propionate (Chapter 16). A ferrous iron, fixed in the centre of the ring through complexation to the nitrogen atoms of the four pyrrole rings, is functionally the most important element (Fig. 17.3).

The ferrous iron has an octahedral geometry and can be attached to six ligands:

- Four of these ligands are the pyrrole nitrogen atoms of the porphyrin ring system.

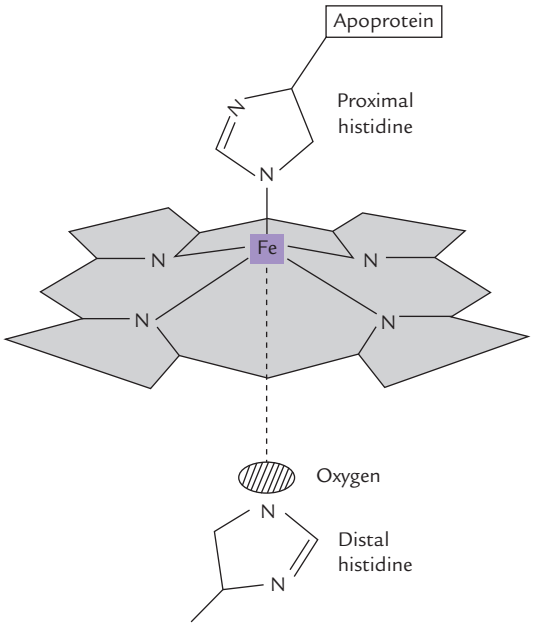


Fig. 17.3. The haem iron lying at centre of the tetrapyrrole ring.

- Fifth is an imidazole nitrogen of the so called proximal histidine.
- The sixth ligand is molecular oxygen bound between the iron and the imidazole side chain of distal histidine.

In deoxyhaemoglobin, the iron is coordinated with only five ligands since oxygen is not present.

Haem group is the oxygen-binding site. It consists of a porphyrin ring system with ferrous iron fixed in the centre.

Proximal and distal histidine: They are located on either side of the haem prosthetic group. Side chains of both these histidines are oriented perpendicular to the planar haem group. The proximal histidine is bound to iron, and the distal histidine guards the entrance on the opposite site (Fig. 17.3). Their presence ensures that the immediate environment of the haem iron is the same (i.e. hydrophobic).

Proximal and distal histidine, located on either side of planar haem, create a suitable hydrophobic environment for binding of oxygen to ferrous iron.

The ferrous iron in haem: The haem iron occurs in the ferrous state in haemoglobin (and myoglobin). During oxygen binding, it is not oxidized to the ferric form; *it becomes oxygenated but not oxidized*. Oxidation to the ferric state converts the haemoglobin to **methaemoglobin**, which is functionally inactive. It can no longer carry oxygen. Oxidation to ferric form imparts a net positive charge (haem is neutral), therefore, a negative counter ion must be associated with the molecule. If the counter ion is Cl^- then the compound is *haemin*; if it is OH^- then the compound is *haematin*.

Haemoglobin A is a tetrahedral complex of two identical α-globins and two identical β-globins. The minor adult haemoglobin and fetal haemoglobin also have α-chains, but instead of β-chains, they have δ-chains and γ-chains, respectively.

B. Quaternary Structure

Haemoglobin is a tetramer of four subunits, each subunit consisting of a polypeptide chain with a non-covalently bound haem group. Packing of these subunits relative to one another is referred to as the quaternary structure, as shown earlier in Figure 17.1.

The four polypeptide chains are visualized as comprising two identical dimers, each dimer having an α- and

β-chain. The dimers are designated $(\alpha\beta)_1$ and $(\alpha\beta)_2$, where the numbers refer to dimer 1 and dimer 2. Two chains within a dimer are held together tightly by the ionic bonds and the hydrophobic interactions, which prevent their movement relative to each other. However, the two dimers are linked with each other by weak polar bonds, so that movement at the interface of these two occurs more freely.

C. Allosteric Effects

Oxygen-binding characteristics of haemoglobin change by the following effectors:

- **pO_2:** The partial pressure of oxygen (pO_2).
- **pCO_2:** The partial pressure of carbon dioxide (pCO_2).
- **pH** of the surrounding medium.
- Presence of **2,3-bisphosphoglycerate.**

These effectors are collectively called **allosteric (*"other site"*) modulators** because their interaction at one site of haemoglobin molecule influences the binding characteristics of some other site within the same molecule.

The ligand-binding information is communicated within the haemoglobin molecule from one subunit to another. Such subunit interaction makes the haemoglobin well adapted for integrating the transport of oxygen, carbon dioxide and H^+.

Binding of Oxygen: Positive Cooperativity

Binding affinity of haemoglobin for oxygen changes with the state of oxygenation. As the first oxygen molecule binds with one of the haem groups, the binding affinity of the haemoglobin for the second oxygen molecule is enhanced. As a result, the second oxygen molecule binds much more easily. The same process is repeated for the third and fourth oxygen molecules.

Effectively, *binding of oxygen with one of the haem groups of haemoglobin increases the oxygen affinities of the remaining haem groups*. This effect is called **positive cooperativity.** It indicates a cooperative interaction between different subunits of haemoglobin (Box 17.1) because the ligand-binding information is transmitted from one subunit to another.

These characteristics are shown by the **sigmoid (*S-shaped*) curve** of oxygen dissociation of haemoglobin (Fig. 17.4). This curve describes the fractional saturation of the haem groups at various oxygen partial pressures. The initial flat region reflects low oxygen affinity of the deoxygenated haemoglobin for the first oxygen. With increasing oxygen partial pressure, however, the first haem group gets oxygenated. This facilitates subsequent binding of oxygen molecules. Therefore, the curve becomes steeper, indicating that subsequent oxygen molecules are bound with much higher affinity. In fact, *the fourth oxygen molecule binds haemoglobin 300 times as tightly as the first;* the binding affinities of the second and the third oxygen molecules are intermediate between those of the first and the last.

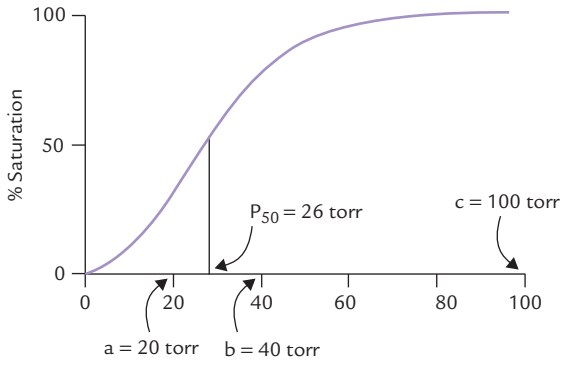


Fig. 17.4. The sigmoid curve of oxygen dissociation of haemoglobin. The arrows indicate: a = pO_2 in capillaries of working muscles; b = pO_2 in capillaries of resting muscles; c = pO_2 in alveoli of lungs, P_{50} indicates 50% saturation of haemoglobin when pO_2 equals 26 torr.

BOX 17.1

Hill Coefficient

The measure of cooperativity among ligand-binding sites is indicated by Hill coefficient(n), named after the scientist Archibald Hill (Nobel Prize in 1922). It reflects the extent to which the interaction of oxygen with one subunit influences the interaction of oxygen with other subunits. For fully cooperative interaction (meaning that binding at one site maximally enhances binding at other sites in the same molecule), value of Hill coefficient is equal to the number of sites. The theoretical maximum for haemoglobin would, therefore, be four. In practice, however, lower values are observed: the normal value for adult haemoglobin (n = 2.7) indicates that there is a substantial functional subunit-subunit interaction. In the absence of cooperative ligand binding, even with multiple sites, the Hill coefficient would be one. This is what is observed for those haemoglobin mutants that have lost cooperative ligand binding, as also for myoglobin.

Oxygen binding to a haem group in haemoglobin increases the oxygen affinities of the remaining haem groups; this effect is positive cooperativity and it leads to sigmoid oxygen-binding curve.

The oxygen dissociation curve is characterized by **P_{50} value**, the oxygen pressure at which the haemoglobin is 50% saturated with oxygen. Its value is **26 torr**. It should be noted that in proteins that are not allosteric and do not show coooperative oxygen-binding kinetics, have a simple **hyperbolic curve** with low P_{50}; this is what is observed with **myoglobin** (see Fig. 17.14).

Note: The term **ligand** refers to specific molecule (such as, oxygen in case of haemoglobin), that is bound by a protein. The word ligand comes from Latin, meaning *"bound entity"*.

Significance of Positive Cooperativity

The cooperative binding of oxygen enhances the efficiency of haemoglobin as an oxygen transporter. Without positive cooperativity, an 81-fold increase of the pO_2 would be required to raise the oxygen saturation from 10% to 90%. In case of haemoglobin, however, about fivefold increase is sufficient to do the same. This means that haemoglobin can rapidly bind oxygen in lungs (where pO_2 is high) and then readily liberate it in the tissue capillaries (where pO_2 is low).

This may be seen from the oxygen dissociation curve:

- Haemoglobin is 96% saturated with oxygen in lungs ($pO_2 = 100$ torr).
- In resting muscle ($pO_2 = 40$ torr) it is only about 64% saturated (Fig. 17.4). Thus, it delivers 32% (96% minus 64%) of its oxygen content to the resting muscle.
- In exercising muscles, further fall in pO_2 occurs because of rapid oxygen consumption. Since haemoglobin is only about 20% saturated at this partial pressure (Table 17.1), further liberation of oxygen occurs.
- Likewise, in the vigorously exercising muscles ($pO_2 = 10$ torr), where oxygen need is maximum, the saturated haemoglobin may release up to 90% of its oxygen content.

Note: **Torr** is a unit of pressure equal to that exerted by a column of mercury 1 mm high at O°C and standard gravity (1 mmHg). This unit is named after Evangelista Torricelli, the inventor of mercury barometer.

Haemoglobin binds oxygen efficiently in lung alveoli and releases it with similar efficiency to peripheral cells. This remarkable duality of function is achieved by cooperative interactions among the globin subunits of haemoglobin, which is reflected by a sigmoid oxygen-dissociation curve.

Table 17.1. Per cent saturation of haemoglobin at different partial pressures of oxygen

pO_2 (torr)	Per cent saturation of Hb
100 (in alveoli)	96
40 (in resting muscles)	64
20 (in working muscles)	20
10 (in vigorously exercising muscles)	10

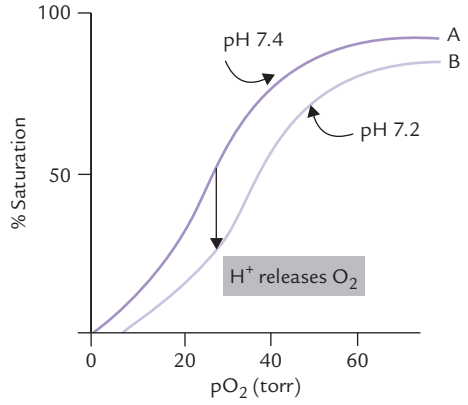


Fig. 17.5. Effect of pH on the oxygen affinity of haemoglobin. The decreased affinity in response to fall in pH is reflected by rightward shift of the oxygen dissociation curve.

Bohr Effect

The oxygen affinity of haemoglobin exhibits exquisite sensitivity to pH; the affinity decreases as the pH becomes more acidic (Fig. 17.5). This is known as the 'Bohr effect' after the discoverer Christian Bohr, father of Neil Bohr, the atomic physicist.

$$HbO_2 + H^+ \rightarrow H.Hb + O_2$$

Closely related to the Bohr effect is the ability of carbon dioxide to alter the oxygen affinity of Hb. Like the negative allosteric effect of H^+, increasing levels of CO_2 decrease affinity for oxygen. Thus, there exists an *inverse relation of oxygen binding affinity of haemoglobin vis-á-vis binding of CO_2 and H^+*.

Decreased affinity of haemoglobin for oxygen is reflected as **rightward shift** of the oxygen dissociation curve. It represents oxygen-offloading.

Effect of CO_2 on oxygen affinity is mediated via $[H^+]$ because in solution there exists the equilibria:

$$CO_2 + H_2O \underset{}{\overset{CA}{\rightleftharpoons}} H_2CO_3 \overset{CA}{\rightleftharpoons} H^+ + HCO_3^-$$

These reactions are enhanced in the red blood cells through the action of *carbonic anhydrase* (CA). Thus, rise in pCO_2 induces a decrease in pH, which then causes a loss of oxygen from oxyhaemoglobin.

Significance of Bohr Effect

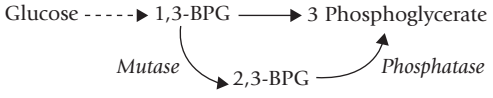
Bohr effect changes the oxygen-binding characteristics of haemoglobin according to the requirements of the tissue. Since an increase in pCO_2 and a decrease in pH are both characteristics of actively metabolizing cells, these cells promote the release of oxygen from oxyhaemoglobin. This is desirable since the actively metabolizing tissues require an adequate supply of oxygen.

A closer look at the curves A and B (Fig. 17.5) will illustrate this point further. At tissue $pO_2 = 40$ torr, haemoglobin will be saturated ~64% at physiologic pH (curve A), and ~50% at acidic pH (curve B). In lungs the haemoglobin is 96% saturated with oxygen. Thus, for curve A, 96 − 64% = 32% of the oxygen bound to haemoglobin would be released to the tissues, whereas for curve B the rate of delivery is 96 − 50% = 46%. Thus, a *rightward shift of the curve provides a mechanism whereby additional oxygen can be supplied to tissues.*

Oxygen-binding affinity of haemoglobin is reduced by low pH. This phenomenon, known as the Bohr effect, is important for delivering adequate oxygen in metabolizing tissue.

Effect of 2,3-bisphosphoglycerate (2,3-BPG)

This compound is an alternative intermediate in the glycolytic pathway, and serves as the third important modulator of oxygen affinity (others are oxygen and carbon dioxide/protons). It is formed in the human erythrocytes from 1,3-bisphosphoglycerate (1,3-BPG), where it is present at a concentration of approximately 5 mM, roughly equimolar with haemoglobin. Like H^+ and CO_2, *2,3-BPG is a negative allosteric effector of oxygen binding to haemoglobin:* it causes a marked increase in P_{50}, thereby causing **unloading of oxygen** from oxyhaemoglobin. Indeed, if it were not for the high erythrocyte concentration of 2,3-BPG, the oxygen saturation curve of Hb would approach that of myoglobin.



In hypoxic states, production of 2,3-BPG is increased which enhances release of oxygen from oxyhaemoglobin (Fig. 17.6).

Rapoport–Luebering Cycle

It is a shunt pathway in RBCs, that bypasses the *glycerophosphate kinase* reaction (step 6) of glycolysis. As diagrammed above, the 2,3-BPG is produced from 1,3-BPG, an intermediate of glycolysis, which is hydrolyzed to 3-phosphoglycerate. This is useful not only in oxygen-offloading, but also in dissipation of energy, as described earlier in Box 9.1.

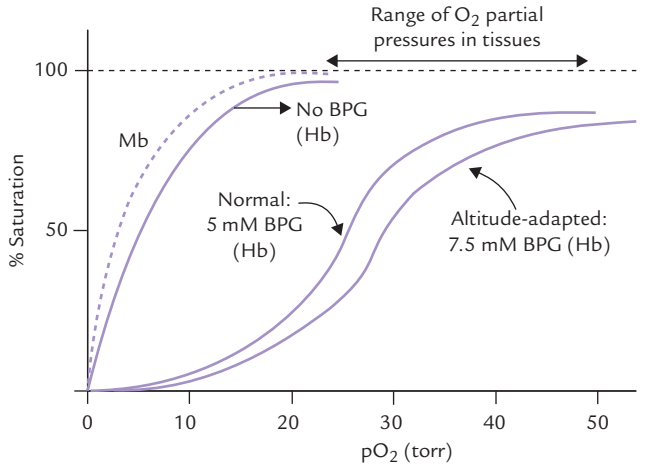


Fig. 17.6. Effect of 2,3-BPG on the oxygen-binding affinity of haemoglobin (Hb). Oxygen binding by myoglobin (Mb) is also shown (- - -).

Clinical Significance of 2,3-BPG

BPG has an indispensable physiological function. As noted, arterial blood, haemoglobin is ~96% saturated, and in venous blood (pO_2 = ~40 torr) it is only 64% saturated, and so unloads 32% (96–64%) of its bound oxygen in passing through capillaries. In the absence of BPG, little of this oxygen would be released, since haemoglobin's oxygen affinity will be increased, thus shifting the oxygen dissociation curve significantly towards the left (Fig. 17.6). The erythrocyte concentration of BPG is responsive to various physiologic and pathologic conditions: it increases when there is deprivation of oxygen, including lung diseases, severe anaemia and during adaptation to high altitude. This enhances unloading of oxygen in the tissues, thereby ensuring adequate oxygenation.

2,3-BPG accounts for oxygen-binding peculiarity of the fetal haemoglobin, discussed later.

Role in Stored Blood

2,3-BPG plays an important role in **blood** that has been stored for **transfusion**. The concentration of 2,3-BPG shows a drastic fall in the red blood cells from 2.2 mM to one-tenth of this concentrations within a few days of storage. This causes increased affinity of Hb for oxygen, so that the transferred blood cannot effectively release O_2 to peripheral tissues. This problem is solved by addition of inosine (ribose + purine base). Haemoglobin releases a significant amount of bound oxygen while passing through capillaries of metabolizing tissues. 2,3-BPG enhances this release to the medium, which is readily transported into red cells (2,3-BPG cannot move across cell membrane). Inside the cell, ribose moiety of inosine is split off from purine base and the former is metabolized (via a section of the HMP shunt and glycolysis) to produce 2,3-BPG.

Haemoglobin responds to body requirements and alters its binding properties accordingly. It can sense changes in the concentration of any of its four ligands—oxygen, carbon dioxide, H^+ and 2,3-BPG. The message is transmitted between different subunits through change in the conformation of the molecule.

Interactions between different ligands, such as BPG and oxygen, are called **heterotropic effects** and interactions between identical ligands, as in the cooperativity of oxygen binding, are called **homotropic effects.**

D. Molecular Mechanism of Allosteric Effects*

The allosteric effects depend on transmission of molecular signal from one subunit of the protein to another. The mechanisms underlying these effects were elucidated following the discovery that the subunits of haemoglobin can take on either one of the **two conformations**: the **R (relaxed)** and **T (tense) form**. The overall interactions between the subunits in the T form are stronger than in the R form, as discussed below.

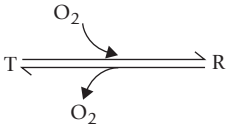
The Tense (T) Form

The deoxyhaemoglobin conformation is called the *tense (T)* or *taut form,* in which the subunits are held together by eight salt bonds between polypeptides in addition to a number of hydrogen bonds and other non-covalent interactions.

The Relaxed (R) Form

It is a less constrained form produced by tearing away of some of the salt linkages that stabilize the T form. Binding of oxygen to the T form causes rotation of an αβ pair relative to the other αβ pair by 15 degrees. This brings about rupture of some of the salt linkages to produce the R form.

The binding affinity of the R form for oxygen is 150 – 300 times greater than that of the T form. Hence, the R form has greater tendency for getting oxygenated. Thus, any allosteric modulator that favours formation of the R form (i.e. T to R shift) favours oxygen binding. Conversely, the allosteric modulators favouring R to T shift increase oxygen liberation.



With this information, it is now possible to understand the mechanism of oxygenation, the Bohr effect and to appreciate role of protons at molecular level.

*This topic relates to higher level learning and it is meant only for postgraduates. Undergraduates or MBBS students may refer to it for higher learning.

Mechanism of Positive Cooperativity in Oxygenation

Binding of oxygen to haemoglobin causes rupture of some of the salt linkages, thus favouring the T to R shift. Since the binding affinity of the R form for oxygen is higher than that of the T form, it accepts oxygen molecules readily. Thus, T → R transition explains why binding of the first oxygen enhances the binding of subsequent oxygen molecules to the other haem groups in the haemoglobin molecule.

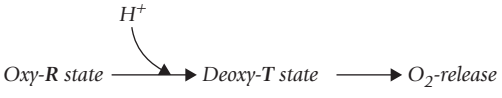
Haemoglobin has two conformational states: the T state (conformation of deoxyhaemoglobin) and the R state (conformation of oxyhaemoglobin). Oxygen binding to a haem group induces a T to R transition, which in turn induces conformational change in the entire haemoglobin molecule.

How does oxygen binding cause rupture of salt linkages for T to R transition? Detailed information about the molecular mechanisms responsible for this has been elucidated by comparative X-ray analysis. In the deoxygenated state, the iron atom is about 0.6Å out of the haem plane, due in part to stearic repulsion between the proximal histidine and the nitrogen atoms of the prophyrin ring (Fig. 17.7). Oxygen binding induces certain changes in the haem's electronic state, which shortens the Fe–porphyrin bonds by ~0.1Å. Consequently, the iron atom moves into the centre of the haem plane, where it can more tightly bind oxygen, and in doing so it drags the covalently linked proximal histidine (F8 His) and its attached F helix. A subtle conformational change is thereby induced, which strains other non-covalent bonds elsewhere in the subunit causing some of them to break. This is beginning of a series of reactions leading to the quaternary structure transition from the deoxy-T form to oxy-R form.

In defining transition from T to R form, a number of models have been developed (Box 17.2).

Mechanism of Bohr Effect

The principal residues involved in Bohr effect are the N-terminal amino groups of the α-chains, the C-terminal carboxy groups of β-chains and the imidazole side chains of His122α. They interact with protons to acquire more positive charge. The protonated, positively charged residues then form **electrostatic interactions** with negatively charged residues, thus converting the R forms into T form and thereby releasing oxygen.



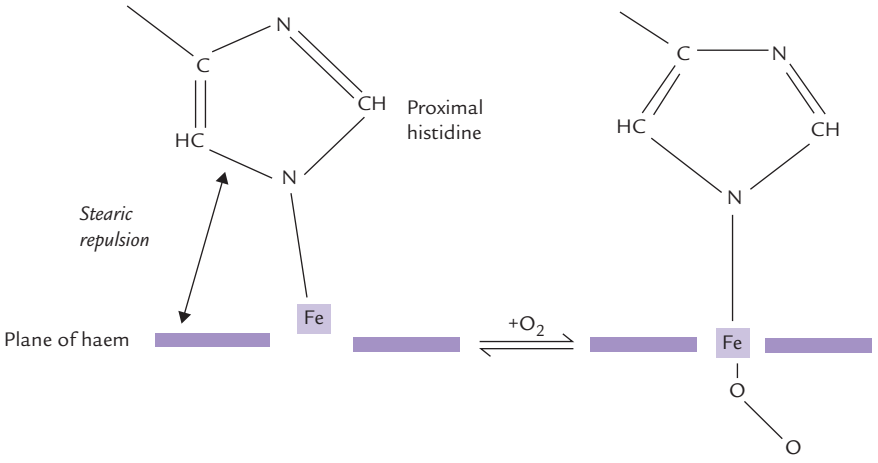


Fig. 17.7. Movement of iron atom into the plane of the haem upon oxygenation.

BOX 17.2

Sequential and Symmetrical Models

The **sequential (Aldair-Koshland) model;** proposed that as dexoyhaemoglobin (all T forms) bound the first molecule of oxygen and changed its conformation to R form; the conformational change would exert an influence on a neighbouring subunit to change its shape to the R form. The latter would then pick up O_2 more easily than the first subunit that. The sequential model thus proposed that each subunit sequentially responds to oxygen binding with a conformational change, so that partially oxygenated haemoglobin molecules would consist of hybrids of the R and T forms.

The **symmetrical (Monod) model** on the other hand proposed that any given haemoglobin molecule was in either all R or all T conformations and that all subunits switched together (concertedly, not one by one) as increasing quantities of oxygen got bound to haemoglobin. Hybrid states are forbidden in this model and oxygen binding causes concerted transition between T and R forms. The majority of evidence, however, points toward the sequential model.

For every equivalent of oxygen released from haemoglobin, about 0.3 equivalents of $[H^+]$ is bound. The bound H^+ are called **Bohr protons.** They are released in the lungs, where pO_2 is high and pCO_2 is low, and the T form changes back to R form.

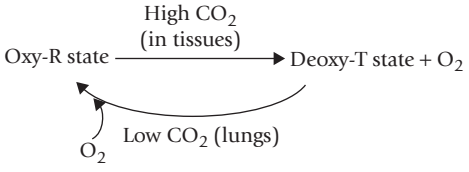
Binding of Bohr protons to certain amino acid residues favours formation of new bonds, thus causing R → T transition and off-loading of oxygen from oxy-haemoglobin.

Effect of CO_2: Lowering of oxygen-binding affinity of haemoglobin by carbon dioxide is effected by the spontaneous covalent binding of CO_2 with the N-terminal group (valine) of each of the globin polypeptides to give **carbamino complexes**.

$$R-NH_2 + CO_2 \rightarrow RNHCOO^- + H^+$$

Deoxyhaemoglobin because of its higher pI (7.6) has a greater affinity for CO_2 than oxyhaemoglobin (pI = 6.7). When the CO_2 concentration is high, as it is in the capillaries, the deoxy-T state is favoured, stimulating haemoglobin to release its bound oxygen. The H^+ produced in the above reaction further promotes oxygen release through the Bohr effect. Clearly, **effects of H^+ and CO_2 are often tightly associated, both promoting oxygen release from oxyhaemoglobin.**

It may be re-emphasized that *like the pH effect, the CO_2 effect ensures that oxygen is released in actively metabolizing tissues,* where it is most needed. On return to the lungs, blood is exposed to low pCO_2 and by law of mass action the carbamination reaction is reversed and T to R conversion occurs, therefore binding of oxygen is again favoured.



Mechanism of Oxygen Off-loading by BPG

The oxygen unloading effect of 2,3-BPG is accounted by the fact that it binds tightly to the deoxy-T state of

haemoglobin but only weakly to the oxy-R state. The result is stabilization of the T conformation, which has low oxygen-binding affinity. Moreover, the presence of 2,3-BPG also lowers intracellular pH (6.95). This **dual role** leads to the observed effect of oxygen unloading. **How does 2,3-BPG stabilize the deoxy-T state?** The X-ray structure of a BPG-deoxyhaemoglobin complex shows that BPG binds in the central cavity formed among the four subunits of Hb. The binding site for BPG is created at one end of this cavity by multiple positive charges, contributed by several residues, e.g. 2nd histidine, 143rd histidine, 82nd lysine in β-chain. BPG interacts electrostatically with these and the interaction in turn stabilizes the complex between the effector (BPG) and the Hb in its deoxy-T state.

BPG decreases haemoglobin's oxygen affinity by binding to the deoxy-T state, and stabilizing it.

II. Haemoglobin Variants

A. Fetal Haemoglobin (HbF)

Fetal haemoglobin differs from the adult haemoglobin in having γ-polypeptide chain, a variant of the β-chain (subunit composition $\alpha_2\gamma_2$; see Fig. 17.2). While it accounts for less than 1% total adult haemoglobin in adults, HbF predominates in fetus during the second and third trimesters of gestation. Its synthesis starts by seventh week of gestation and at birth it accounts for about 75% haemoglobin. There is rapid post-natal decline and within 4 months after birth HbF is almost completely replaced by HbA (Fig. 17.8).

Since the γ-chain differs from β-chain in 39 of the residues, the physicochemical properties of the HbF isoform are different: its electrophoretic mobility is slower and the deoxy HbF has higher solubility than HbA. The most striking difference between the two is: **HbF has lower affinity for 2,3-BPG** (because the residue 143 of the β-chain, which is histidine in HbA but is replaced by serine in HbF; the absence of this histidine eliminates a pair of interactions that stabilize the BPG-deoxyhaemoglobin complex). As a result, affinity of HbF for oxygen **(P_{50} = 19 torr,** vs 26 torr for HbA) is much higher and formation of the oxygenated R form is favoured. The oxygen dissociation curve in fetus consequently is shifted to left, a direct benefit of which is that there is a more efficient transplacental transfer of oxygen from maternal HbA to fetal HbF.

HbF is elevated up to 15–20% in individuals with mutant adult HbS, such as sickle cell disease where it has a compensatory effect. Children with anaemia and β-thalassaemia also show such compensatory response.

B. Embryonic Haemoglobins

The non-α-chains present in embryonic haemoglobins are the embryonic chains, i.e. epsilon (ε) and zeta (ξ). Expression of these chains is limited to early embryonic stage, from 3rd to 8th week of gestation (Fig. 17.8). The embryonic haemoglobin Gower-2 has two alpha and two epsilon chains, and the haemoglobin Gower-1 comprises only embryonic chains.

C. Minor Adult Haemoglobin (HbA_2)

It is a normal variant of adult haemoglobin, having δ-chains instead of β (subunit composition $\alpha_2\delta_2$; Fig. 17.2). In β-thalassaemia, concentration of this form of haemoglobin

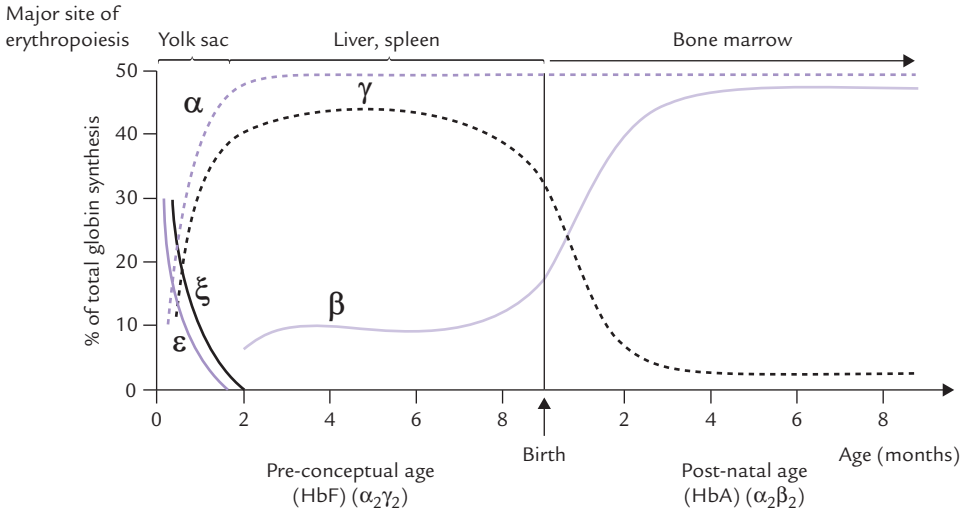


Fig. 17.8. Human haemoglobin polypeptide chains as a function of time during gestation and post-natal development.

is increased as a compensatory measure. Its pI value is 7.4, which is higher than that of HbA. Therefore, it shows slower anodic mobility on electrophoresis.

III. Haemoglobin Derivatives

Haemoglobin derivatives are formed by:

- Change in oxidation state of haemoglobin, e.g. *methaemoglobin*, a non-functional oxidized form of haemoglobin.
- Combination of different ligands with the haem part of the haemoglobin molecule. For example, *carboxy haemoglobin* is formed by binding of carbon monoxide to the haem iron; *sulph-Hb*, by its binding with hydrogen sulphide.

Haemoglobin derivatives can be detected by their characteristic absorption spectra.

A. Methaemoglobin

The haem iron of haemoglobin is present in the ferrous state. Its oxidation to the **ferric** form produces haemin from haem, and the haemoglobin molecule is now called methaemoglobin. Methaemoglobin is ineffective as an oxygen transporter.

Normally, about 1% of the circulating haemoglobin occurs in form of methaemoglobin. Certain chemicals cause rapid ferrous to ferric conversion, resulting in production of methaemoglobin. Examples include organic and inorganic nitrites, aniline dyes, and aromatic nitro compounds. The methaemoglobin production by these agents leads to conditions, collectively referred to as **acquired methaemoglobinaemias.** In contrast to the acquired forms, the **congenital methaemoglobinaemia** results mostly because of defect in the *cytochrome b_5 reductase* system, as discussed below.

Methaemoglobin is a non-functional oxidized form of haemoglobin formed by oxidizing agents that oxidize the ferrous haem iron to the ferric state.

Since the methaemoglobin formation results in loss of oxygen carrying capacity, it is hazardous and must be prevented. Erythrocytes have certain defensive mechanisms to prevent excessive formation of methaemoglobin. They are described below.

- Ascorbic acid and glutathione are reducing substances which destroy many oxidizing radicals, thus preventing them to react with haemoglobin (Chapter 27).

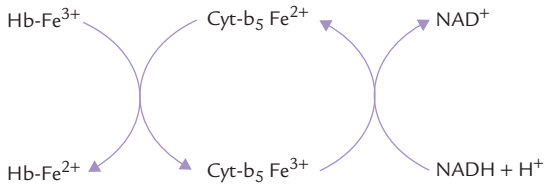


Fig. 17.9. The *methaemoglobin reductase* system.

- The methaemoglobin that is spontaneously produced can be reduced back to haemoglobin by the ***methaemoglobin reductase* system** present within the RBCs (Fig. 17.9).

  This enzyme system uses NADH as the reductant. Cytochrome b_5 and *cytochrome b_5 reductase* are the other essential components of this system. Impaired function of this system leads to *congenital methaemoglobinaemia* **(Case 17.1).**
- The polypeptide chains of haemoglobin contain a protected pocket each between the E and the F helices, in which the haem molecule is tucked in. These pockets offer protection against oxidation of haem. This is evident from the fact that when the haem molecule is dissociated from the polypeptide chain, it rapidly gets oxidized to haemin.

  The distal histidine residues appear to play an important role of creating a protective environment for haem. Replacement of this histidine by a tyrosine residue, for example, favours production of methaemoglobin.

Defect in the *cytochrome b_5 reductase* system, or structured abnormality in the polypeptide chain(s), result in congential methaemoglobinaemia. It is a mild condition and does not manifest clinically under normal circumstances. However, use of certain drugs, as mentioned earlier, results in excessive production of methaemoglobin and the oxygen deficiency symptoms appear.

Treatment of methaemoglobinaemia involves conversion of the ferric iron back to the ferrous state. Methylene blue is commonly used for this purpose.

B. Carboxy Haemoglobin

The toxic effects of carbon monoxide (CO) were explained first by John Scott Haldane on the basis of competition between carbon monoxide and oxygen for binding the haem iron. In fact, the affinity of carbon monoxide for Hb is 200 times more than that of oxygen.

$$Hb + CO \rightarrow CO.Hb \text{ (Carboxy haemoglobin)}$$

Since CO.Hb is unsuitable for oxygen transport, its binding renders the haemoglobin functionally inactive **(Case 17.2)**.

Haemoglobin can be poisoned by carbon monoxide that occupies the oxygen-binding site on the haem iron.

C. Glycated Haemoglobins

These are the haemoglobin derivatives that are produced by non-enzymatic post-translational modification (e.g. glycation) of the β-chains. They make up about *4–6% of the total haemoglobin* in normal red blood cells. Being present in very small amounts, the glycated haemoglobins are not pathological, but rather serve a useful purpose of monitoring patient's compliance in diabetic state or even in diagnosis of diabetes (Chapter 15).

D. Sulfhaemoglobin

It is a greenish compound produced by *covalent attachment of sulphur to the porphyrin ring* (**not iron** atom). Sulfhaemoglobin cannot combine with oxygen. Exposure to sulphur containing compounds, either occupationally or from air pollution, or depsone (used to treat leprosy) can produce this compound. In poisoning by phenacetin, acetanilide or sulfanilamides, sulfhaemoglobin is produced. These drugs produce methaemoglobin also, so that sulfhaemoglobin and methaemoglobin appear together in these patients.

E. Cyanmethaemoglobin

Haemoglobin cannot combine with cyanide, but *methaemoglobin reacts with cyanide forming cyanmethaemoglobin* (cyanmet-Hb). This property is used as a treatment modality in cyanide poisoning. First, the patient is given nitrite plus sodium thiosulphate, which converts haemoglobin to met-Hb. The latter then traps cyanide as cyanmet-Hb, which is a non-toxic compound.

IV. Haemoglobinopathies

Haemoglobinopathies are a family of disorders resulting due to mutations in genes for the haemoglobin chains. The mutation may cause synthesis of defective globin chain or may even stop the synthesis altogether. More than 400 different mutations have been described, which result in as many structural variants of the haemoglobin polypeptide chains. **Two types** of defects arise due to these mutations: the **qualitative defects** and the **quantitative defects**.

Qualitative defects lead to production of *structurally abnormal haemoglobin* molecules. These defects involve structural alteration in the polypeptide chain of the haemoglobin, leading to functional impairment. Replacement, addition and deletion of the amino acids are some common structural alterations.

A number of qualitative defects (over 300) have been identified, which may affect either of the two types of chains. For example, in **sickle cell anaemia**, the β-chain is affected, whereas in HbM Boston the α-chain undergoes structural alteration (Table 17.2).

Quantitative defects result in *synthesis of the normal haemoglobin molecules, but in insufficient quantities*. Prime example of this type is the **thalassaemias** in which either the α- or β-chains are underproduced. Rarely both these types of defects occur together in a single disorder.

A partial listing of haemoglobinopathies is given in Table 17.2. By convention, the haemoglobinopathies are usually named with a capital letter (e.g. haemoglobin S), a geographic location (e.g. haemoglobin Seattle), or both (e.g. haemoglobin M Boston or D Punjab).

Table 17.2. Some mutated human haemoglobins

Mutated haemoglobin	Affected chain	Residue	Substitution	Notes
S	β	6	Glu → Val	Decreased solubility of Hb. Sickling of RBCs.
C	β	6	Glu → Lys	Decreased solubility. Sickling.
E	β	26	Glu → Lys	
Zurich	β	63	His → Arg	Affinity of Hb for O_2 increased. Solubility decreased.
Seattle	β	70	Ala → Asp	
Hiroshima	β	146	His → Asp	High O_2 affinity. Salt bridges stabilizing T form impossible to make.
Kansas	β	102	Asn → Thr	Low oxygen affinity. Cyanosis common.
D Punjab	β	121	Glu → Gln	Migrates similar to HbS on electrophoresis. Severe condition.
M Boston	α	58	His → Try	Substitution in proximal and distal histidine results in increased tendency of haem to get oxidized to haemin. Cyanosis common.

Diagnostic Analysis of Normal and Mutant Haemoglobins

All significant Hb variants and many of the more common mutants, including HbS and HbC, can be separated by **electrophoresis.** The process is carried out at pH 8.6 at which these proteins are negatively charged, and therefore, move towards anode.

Change in amino acid composition of a globin chain may produce a haemoglobin with a different electrophoretic mobility. By examining the position and relative amount of the haemoglobin bands, it is possible to diagnose the most common haemoglobinopathies. For example, in sickle cell anaemia, the predominant haemoglobin (85–95% of the total) is haemoglobin S (HbS); its position on electrophoretogram is shown in Figure 17.10. The *HbS has lower anodic mobility because of removal of a negative charge by the sickle cell mutation* (Glu → Val). Likewise, any mutant exhibiting a net gain or loss of charged residues when compared with HbA can be detected by this method.

The volume of blood sample required is less than 100 µl, making this technique acceptable for even neonates. Quantification is performed by *scanning densitometry*. Another method, **isoelectric focusing,** also based on differences in pI, separates the proteins rapidly in a pH gradient generated by an electric field (see Chapter 4).

A. Sickle Cell Anaemia

It is the most commonly occurring haemoglobinopathy caused by an inherited structural abnormality in the β globin polypeptide. Originally described in 1910 by James Herrick, the disease was the *first inherited disorder shown to arise from a specific amino acid change in a protein.* Since then it has been so extensively studied, biochemically and biophysically, that it has virtually become a paradigm of a molecular disease.

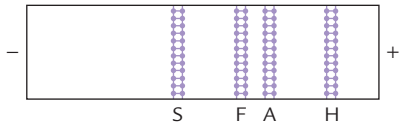


Fig. 17.10. The electrophoresis of haemoglobin.

Molecular Defect

The sickle cell haemoglobin (HbS) arises out of a substitution of the glutamic acid residue in position 6 of the β-chain with the valine residue (β6Glu → Val). As noted earlier, this substitution can be traced to an A → T transversion in the β-chain gene and results in formation of an abnormal haemoglobin of subunit composition $\alpha_2\beta^s_2$ (Fig. 17.11). Both, homozygous and heterozygous forms of disease have been described. The heterozygous form is also referred to as **sickle cell trait**.

Sickle cell anaemia results because of a point mutation (A → T transversion) in the β-chain gene. It leads to a Glu → Val substitution in the β-chain. The resultant haemoglobin S (HbS) is characterized by the subunit structure $\alpha_2\beta^s_2$.

Sticky Patches and Tubular Polymer Formation

Replacement of a negatively charged glutamic acid with a hydrophobic valine changes surface properties of the molecule, especially its solubility. It becomes less soluble in water in its deoxygenated form, although the oxygen-binding affinity remains unaffected. In a chain comprising of 146 amino acids, a single substitution may appear minor, but its effects are far from trivial. This is because the substituted valine forms a **hydrophobic pocket**, the so called *sticky patch*, on the surface of the molecule which is reactive to the nearby Hb molecules.

The X-ray structure of deoxyhaemoglobin S has revealed that *one mutant valine side chain in each HbS molecule fits into a sticky patch on the surface of another haemoglobin molecule.* This intermolecular contact allows HbS to form **linear polymers** of deoxy HbS. These polymers

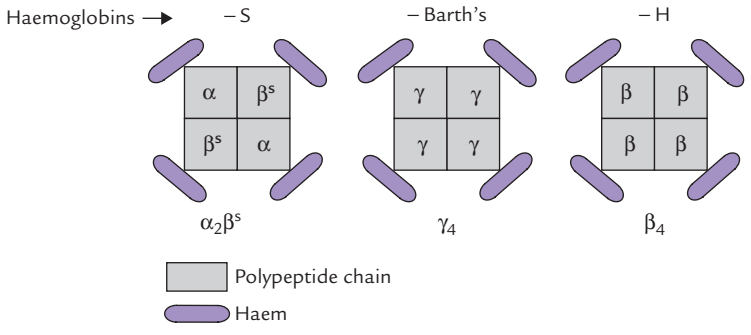


Fig. 17.11. Structure and subunit composition of different haemoglobins: HbS, HbH and Hb Barth's. HbS is the predominant haemoglobin type in sickle cell anaemia, whereas HbH and Hb Barth's occur in thalassaemia.

distort the RBCs into sickle shape, which in highly vulnerable to lysis.

The hydrophobic pocket on surface of the β-subunit is not seen in case of oxyhaemoglobin, therefore oxyhaemoglobin S cannot polymerize. This also explains why *danger of polymerization* is *greatest in hypoxic conditions*.

HbS has a normal oxygen binding affinity, but it has reduced water solubility in deoxygenated form. Deoxygenated but not oxygenated HbS forms an insoluble fibrous precipitate in the erythrocytes.

Clinical Features

Sickle cell anaemia is a severe-haemolytic, painful, debilitating, and often fatal disease characterized by erythrocytes of distorted shapes. The stiff polymers form precipitate in the erythrocyte, which causes distortion of shape of the RBCs into sickle shape (i.e. sickling). At first the sickling is reversible, but after some time the oxygenation–deoxygenation cycle permanently deforms the shape of the erythrocyte. The deformed cells increase the viscosity of blood and cause obstruction to blood flow in small vessels and capillaries. The resulting vascular occlusion may result in **tissue death** (*infarction*). Painful bone and joint infarctions are common; multiple renal infarcts can lead to renal failure; and many patients are crippled by recurrent CNS infarctions.

Recurrent attacks of sever pain in extremities, bones or abdomen, known as painful crisis or **sickling crisis** are experienced by most patients. Others may experience **aplastic crisis** (bone marrow failure), or **sequestration crisis** where sudden trapping of erythrocytes in the enlarged spleen occurs; the latter is associated with high mortality. Renal failure, cerebrovascular accidents, cardiac failure and liver diseases are the other important causes of death.

Haemoglobin S forms rigid fibres in the deoxy form within the erythrocyte and distorts its shape into sickle shape. The patient suffers from recurrent episodes of haemolytic and painful vaso-occlusive crisis.

Interestingly, the disease follows a milder course in many HbS homozygotes because they express relatively high levels of fetal haemoglobin, which contains γ-instead of defective β-chain.

People with ***sickle cell trait*** (*heterozygous form* with about 40% HbS) also form insoluble fibres, though at a lower rate. They are asymptomatic in normal circumstances. Their cells exhibit sickling only when pO_2 is low, such as at high altitudes or in chronic lung disorders. Therefore, sickle cell trait is a relatively benign condition as such not producing any clinical symptoms and compatible with a normal life span. The red cells of these patients contain approximately 70% HbA and 30% HbS (**Case 17.3**).

Diagnosis

Sickling test: A blood smear is exposed to low oxygen tension by adding a reducing agent, e.g. sodium dithionite. The red cells respond by changing to sickle shape. This is detected microscopically.

Electrophoresis: Electrophoresis remains the mainstay of diagnosis of sickle cell anaemia. Electrophoretic separation is possible because the sickle cell mutation (Glu → Val) removes a negative charge from the β-chain resulting in a decreased anodal mobility (**Fig. 17.10**).

Solubility test: A sample of haemolyzed red cells is exposed to anoxic conditions by adding a reducing agent. An opalescence is observed, which indicates presence of less soluble deoxy-HbS.

Treatment

1. Initial management aims at *avoiding hypoxia* and dehydration and *administration of cyanate*, which increases the oxygen affinity by covalent modification of the amino termini of the globin polypeptides.
2. An alternative approach, which involves manipulations to increase the synthesis of fetal haemoglobin, has been found useful. *Hydroxyurea* has been found to increase the fraction of cells containing fetal haemoglobin, although the mechanism whereby it acts is not known. Fetal Hb is useful because of its high O_2 affinity.
3. Repeated blood transfusions are required in severe anaemic cases.

Protection from Tropical Malaria

HbS gene occurs in India, Mediterranean region and Saudi Arabia. In some parts of Africa, up to 40 per cent of the population carries the sickle gene. These areas have highest prevalence of malaria as well. Presence of the sickle gene has a survival value since the native Africans with sickle cell genes show a much lower susceptibility to malaria. The mosquito borne parasite that invades the red blood cells finds less favorable conditions for its growth in the sickled cell compared to normal cells. Moreover, increased lysis cause premature destruction of the parasite.

Possession of the sickle gene confers significant survival value in the areas where malaria is endemic and is fatal in childhood.

HbC Disease

The HbC is a haemoglobin with a β6Glu → Lys substitution. Like HbS it is found in black population and is

related to sickle cell anaemia. Its symptoms, however, are less severe. Presence of HbC also appears to create a less hospitable environment for the malaria, perhaps even causing premature destruction of the parasitized cells.

It is now believed that HbS and HbC arose as a means of combating malaria.

B. Thalassaemias

Thalassaemias are a group of disorders characterized by **insufficient production of structurally normal α- or β-chains.** High prevalence of thalassaemias has been reported among the people from the regions around the Mediterranean Sea (*thalassa* is the Greek word for sea).

Insufficient production of one of the globin chains of haemoglobin results in lack of coordination between the synthesis of the α- and the β-chains. Normally, the subunit synthesis is coordinated in such a way that each newly synthesized β-chain readily pairs with an α-chain; and conversely, each newly synthesized α-chain pairs with a β-chain. The lack of coordination results in:

- Formation of *insoluble aggregates* by the polypeptide chains. The aggregaetes are damaging to the cell and reduce its lifespan.
- Impairment of the haemoglobin synthesis, so that the *erythrocytes are small* and *poorly filled with haemoglobin*.

Molecular Defect

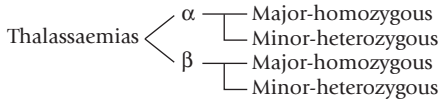
Underlying cause of the underproduction of a globin chain is a **gene mutation.** A number of mutations have been reported, including gene deletion, gene substitution, or deletions of one to several nucleotides in the DNA.

Some mutations lead to:

- Underproduction, abnormal processing and premature degradation of mRNA, or
- Increased proteolytic degradation of the α- or the β-chain.

Types of Thalassaemias

Depending on which chain is affected, the thalassaemias are divided into two major classes: α- and β-thalassaemias. In α-thalassaemia relative or absolute deficiency of the α-chain occurs, whereas in β-thalassaemia the β-chains are affected. Heterozygous forms of α- or β-thalassaemia are called **thalassaemia minor,** and the homozygous forms are termed **thalassaemia major**.



In thalassaemia minor, anaemia may be present but is very mild. In thalassaemia major, severe anaemia develops.

The thalassaemias are caused by underproduction of either the α- or the β-chain. The α-thalassaemia is caused most often by large deletions, and many different mutations can cause β-thalassaemias.

α-Thalassaemias

Genes for the α-chains are clustered on chromosome 16. Genome of an individual contains four copies (two copies from each parent) of the α-globin genes; two copies are located on each of the chromosome 16. *Most patients with α-thalassaemia have large deletions that remove one or both of the α-chain genes* from a chromosome. Figure 17.12 shows the possible deletion types. Four types of α-thalassaemias corresponding to deletion of one, two, three, or all four of the α-chain genes occur:

1. *Silent carrier state* (Fig. 17.12a) is due to deletion of one gene. It is a haematologically normal state; no clinical manifestations are observed in this condition.
2. *α-Thalassaemia trait* involves deletion of two genes from the same or different chromosomes (Fig. 17.12b and c). This results in mild physical manifestations, such as mild anaemia and slightly enlarged spleen.
3. *Haemoglobin H disease* involves deletion of three genes (Fig. 17.12d). The condition is characterized by mild to moderate haemolytic anaemia.
4. *Homozygous α-thalassaemia* (Fig. 17.12e) with four deletions is always lethal, either in utero or at birth. The condition is also called hydrops foetalis or **haemoglobin Barth's disease** (Box 17.3).

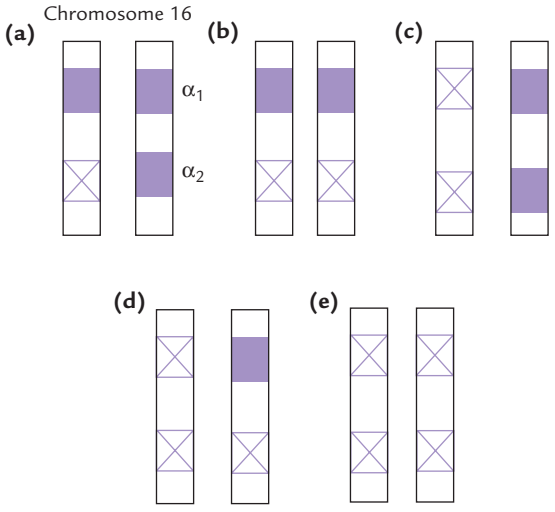


Fig. 17.12. One or more genes for the α-chains are deleted in α-thalassaemias. **(a)** Silent carrier, **(b** and **c),** α-Thalassaemia trait, **(d)** Haemoglobin H disease, **(e)** Haemoglobin Barth's disease.

BOX 17.3

Types of Haemoglobins in Thalassaemic Patient

Though synthesis of the α-chain is hampered in α-thalassaemias, the β-chain and the γ-chain synthesis proceeds as usual. As a result, there is accumulation of certain abnormal proteins (Fig. 17.11). These are:

(i) β-Tetramer, which is termed **HbH.** It is the predominant haemoglobin in patients with deletion of three α-chain genes. This aberrant haemoglobin is unstable; it gradually denatures to form inclusion bodies in the cells.

(ii) γ-Tetramer, termed **Hb Barth's,** have tenfold higher oxygen affinity than haemoglobin A and, therefore, cannot deliver the bound oxygen in tissues. γ-Tetramers cannot serve as effective oxygen carriers. Thus, tissue hypoxia is the major consequence, and oxygen deficiency symptoms rapidly appear.

β-Thalassaemias

Synthesis of β-chains decreases or stops altogether in the β-thalassaemias. In contrast to large deletions in α-thalassaemia, most patients with β-thalassaemia have **single base substitutions**. More than 90 such substitutions are known. Few cases of β-thalassaemia are caused by large deletions.

Normally, there are only two genes responsible for the β-globin chains, one from each parent. Each copy of chromosome 11 has one gene for β-globin chain. The mutation in β-thalassaemia may involve a single chromosome (i.e. the heterozygous state, called **β-thalassaemia minor**) or it is a homozygous state, involving both the chromosomes **(**called **β-thalassaemia major).**

1. *β-Thalassaemia minor* is benign and require any treatment. There is almost 50% decrease in β-chain synthesis. A compensatory increase in production of the δ- and the γ-chains occurs. These chains combine with the α-chains. Thus an increase in the amount of haemoglobin $A_2(\alpha_2\delta_2)$ and $F(\alpha_2\gamma_2)$ occurs, which is a distinguishing feature of β-thalassaemias.

   β-Thalassaemia minor is mostly asymptomatic and life expectancy is also normal since some amount of β-chains are normally synthesized by these individuals.
2. *β-Thalassaemia major* is the homozygous state, which is most severe among all congenital haemolytic anaemias. It is also called **Cooley's anaemia.** The affected individuals require regular blood transfusions for survival.

Clinical features of β-thalassaemia major: Since the β-chain begins to be synthesized at a later stage in fetal gestation, the physical manifestations of β-thalassaemia become apparent only after birth. The affected infants are normal at the time of birth because of abundance of fetal haemoglobin, but a severe anaemia develops during the first year of life and the haemoglobin level in most infants falls below 5 gm/dL. The marrow responds to anaemia and becomes over-active. Consequently, a massive expansion of the red bone marrow occurs, leading to facial deformities (*"chipmuk facies"*). Extramedullary erythropoiesis is seen in liver and spleen in some patients.

Some of β-thalassaemia mutations result in a complete absence of the β-chains in the homozygous state (β°-thalassaemia). Others cause a decrease in the β-chain synthesis (β^+-thalassaemia).

Treatment

Untreated, the patients with thalassaemia major suffer recurrent infections and severe anaemia. With regular ***blood transfusions***, life expectancy is approximately 20 years of age. Although blood transfusion is life saving, the cumulative effect of the procedure is iron overload, a syndrome known as ***haemosiderosis.*** It has a high mortality, usually between the ages of 15 and 25 years. For this reason, the patients are treated not only with regular blood transfusions but also with ***desferrioxamine***, an iron chelator that is best administered through a subcutaneous infusion pump. Desferrioxamine forms a soluble iron complex that can be excreted by the kidneys. ***Bone marrow replacement therapy*** has evolved as a highly successful mode of treatment since the last decade or so. It has improved life expectancy significantly.

V. Myoglobin

Myoglobin is a relatively small molecule (MW 167,00), consisting of a single polypeptide chain, attached noncovalently to a haem molecule (Fig. 17.13). It is mainly located in cardiac and skeletal muscles, where it serves as a **reservoir for oxygen**. In addition, it transports oxygen to the mitochondria. The haem group is responsible for the deep brown colour of myoglobin (and of haemoglobin as well). Myoglobin is particularly abundant in muscles of diving mammals such as whale and seal, whose muscles are so rich in myoglobin that they are brown coloured.

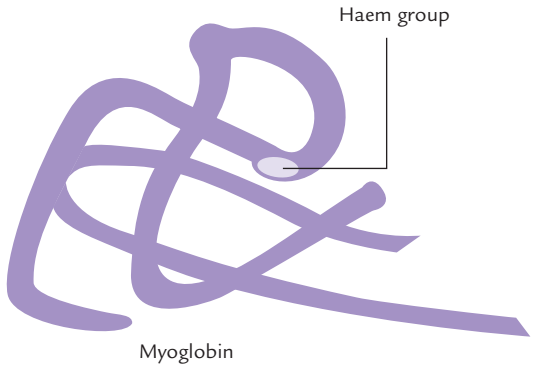


Fig. 17.13. Structure of myoglobin.

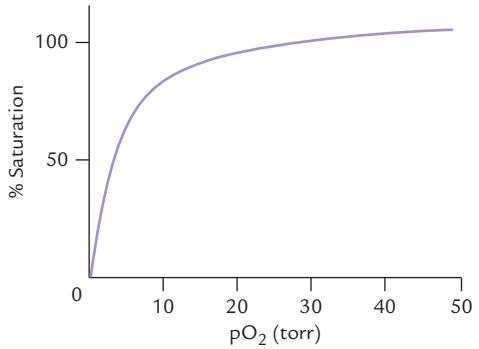


Fig. 17.14. Oxygen dissociation curve of myoglobin.

Storage of oxygen by muscle myoglobin permits these animals to remain submerged in water for long periods.

A. Basic Structure

The single polypeptide chain of myoglobin has 153 amino acid residues. About 75% of the chain is in a α-helical conformation. There are eight major α-helical segments, referred to as A, B, C, … H. Four of the these segments are terminated by proline residues, whose rigid R group does not fit within the straight stretch of α-helix. The non-helical segments lie in between these helical portions.

The polypeptide chain is so compactly folded that there is no space in its interior. Most of the hydrophobic R groups are in the interior of the molecule, hidden from water. They form a clustered structure stabilized by hydrophobic interactions. In contrast, the charged amino acids are located almost exclusively towards the exterior where they form hydrogen bonds with the surrounding aqueous medium.

The compact structure of myoglobin is stabilized by hydrogen and ionic bonds as well as by the hydrophobic interactions between the hydrophobic R groups.

The polypeptide chain of myoglobin is structurally similar to the individual polypeptide chains of the haemoglobin molecule. Thus, myoglobin provides a simpler model for the study of complex oxygen-binding properties of haemoglobin.

B. Oxygen-binding Characteristics of Myoglobin

Myoglobin has a very high affinity for oxygen; it is 50% saturated with oxygen at a partial pressure of just 1–2 torr. At about 20 torr, it is over 95% saturated. Binding of oxygen to myoglobin is directed by mass action of oxygen. When oxygen is plentiful, formation of oxygenated myoglobin occurs and when oxygen is very scarce, dissociation of the oxygenated myoglobin occurs.

The oxygen dissociation curve of myoglobin is a simple *hyperbolic curve* (Fig. 17.14) in contrast to the sigmoid shape of the oxygen dissociation curve of haemoglobin. *The curve is a molecular adaptation for its storage function.*

- High affinity of myoglobin for oxygen permits myoglobin to store oxygen even at the relatively lower partial pressure (40 torr) that exists in resting muscles.
- In exercising muscles, where the oxygen partial pressure falls to about 20 torr, myoglobin is still over 95% saturated. It unloads very little oxygen which makes it suitable for its role as a reservoir of oxygen.
- It is only during severe physical exercise—when pO_2 within the muscles falls to a low level (< 5 torr)—that myoglobin releases a significant proportion of its stored oxygen.

The myoglobin has high oxygen affinity, but does not show Bohr effect, cooperative effect and 2,3-BPG effect.

Myoglobin is a monomeric haem-containing muscle protein that reversibly binds a single oxygen molecule. Unlike haemoglobin it lacks allosteric properties, but has far greater affinity for oxygen.

Properties of haem are different in haemoglobin, myoglobin and other haem-containing proteins (Box 17.4).

VI. Anaemias

In our country, anaemia is one of the most common medical problem. In this condition, the haemoglobin concentration is reduced (normal is 13–16 g/dL in males; and 12–15 g/dL in females).

The commonest cause of anaemia in India is iron deficiency. For clinical manifestation of the iron deficiency

BOX 17.4

Haem Proteins

Haem proteins are a group of specialized conjugated proteins that contain haem as their prosthetic group. Most abundant haem proteins in humans are haemoglobin and myoglobin. Cytochromes and *catalase* are other examples. Haem is tightly bound to the apoprotein portion and performs different functions in different haem proteins. For example, haem group of myoglobin and haemoglobin reversibly bind oxygen; the haem group of the cytochromes functions as an electron carrier; whereas, that of the enzyme, *catalase*, forms part of the enzyme's active site that catalyzes breakdown of hydrogen peroxide.

anaemia (see **Case 19.1**). The causes of anaemia are given below:

Decreased production of erythrocytes: Defective synthesis of haemoglobin may lead to decreased production of erythrocytes. A number of cofactors and mineral elements are required for this process, including copper, iron, ascorbic acid, pyridoxal phosphate, and folic acid. Nutritional deficiency of these factors therefore results in anaemia.

Enhanced destruction of erythrocytes: Under normal circumstances a balance exists between the production and the destruction of erythrocytes. Integrity of the erythrocyte structure is affected in a number of disorders, both intracorpuscular and extracorpuscular. These disorders result in excessive destruction of the cells (haemolysis), resulting in (haemolytic) anaemia. Some of the common conditions that lead to haemolytic anaemia are:

(a) *Intracorpuscular defects* are haemoglobinopathies caused by enzyme deficiencies such as deficiency of *glucose-6-phosphate dehydrogenase* or *pyruvate kinase* (**Case 10.1**).

(b) *Extracorpuscular defects* such as infections (malaria or coxsackie virus infection), drug exposure (quinine or methyl dopa), autoimmune haemolysis or mismatched blood transfusion.

Haemorrhage: Chronic blood loss may lead to anaemia. The commonest causes are haemorrhoides, menorrhagia (i.e. heavy periods), bleeding peptic ulcer, or hookworm infection.

Exercises

Essay type questions

1. Draw a diagram of the sigmoidal- and hyperbolic-oxygen dissociation curves of haemoglobin and myoglobin respectively. Describe how haemoglobin effectively delivers oxygen to myoglobin in muscles.
2. Explain the structural and functional differences between haemoglobin and myoglobin.
3. Explain the structural basis for cooperative oxygen binding to haemoglobin. What is the physiological relevance of the Bohr effect?
4. How do BPG, carbon dioxide and pH value affect oxygen binding to haemoglobin?

Write short notes on

1. Thalassaemias
2. Haemoglobin electrophoresis
3. Methaemoglobinaemias
4. Haemoglobin S
5. CO poisoning
6. Bohr effect
7. Haemolysis in G6PD deficiency
8. Fetal haemoglobin
9. Sickle cell trait

CLINICAL CASES

CASE 17.1 Shortness of breath and cyanosis in a 1-year-old child

A 1-year-old had frequent episodes of headache, exertional dyspnoea and cyanosis. The venous blood sample, drawn for biochemical investigations, was chocolate-brown in color but turned red on shaking. Addition of a few drops of 10 per cent potassium cyanide to the blood resulted in rapid production of a bright red colour. A number of biochemical tests were performed as an initial screen; no abnormality was, however, detected.

Physical examination ruled out any cardiac or pulmonary disease and there was no history of any recent exposure to any drugs or chemicals. Oxygen therapy was started but the cyanosis was not as promptly alleviated with it, as would be expected in a normal subject.

Defective (or diminished) action of some enzyme protein was suspected and further investigations were conducted for measuring intracellular levels of some major proteins of erythrocytes. When antibodies to the enzyme *cytochrome b_5 reductase* were added, the precipitation was less than that expected in a normal subject. It was thereby concluded that the level of this enzyme protein within the erythrocytes was low.

Q.1. What is the probable diagnosis of this case?
Q.2. How do altered properties of the *cytochrome b_5 reductase* account for the symptoms of this patient?

CASE 17.2 Oxygen insufficiency following inhalation of automobile exhaust

A 48-year-old car mechanic was brought to the hospital emergency in a disoriented state. Earlier, he had felt confusion, throbbing headache and chest pain. These symptoms seemed to have made sudden appearance, for he had appeared his normal self at the time he reported for duty early in the morning.

On examination, his blood pressure was 108/70 mmHg, pulse was 112/min, and respiration was shallow and fast (34/min). Analysis of the blood sample showed increased concentration of carboxy-haemoglobin (39.6%).

The patient was put on oxygen therapy with a diagnosis of sub-acute carbon monoxide poisoning. Hyperbaric oxygen was given, to which he responded well.

Q.1. What is the rationale behind arriving at the above diagnosis? What further tests are required to confirm this diagnosis?
Q.2. What are the various factors that lead to tissue hypoxia in case of carbon monoxide poisoning?
Q.3. What is the rationale behind starting oxygen therapy immediately? Why was hyperbaric oxygen given?
Q.4. Has the smoking habit of the patient influenced development of the present condition?
Q.5. In spite of oxygen insufficiency, the patient does not present with any signs of cyanosis. Give reason.
Q.6. Why is CO poisoning more serious in infants?

CASE 17.3 A 10-year-old boy with breathlessness, pallor and persistent tiredness

A 10-year-old boy was brought to casualty with fever and breathlessness. Earlier he was seen by a general practitioner who had prescribed antibiotics with a diagnosis of upper respiratory tract infection. Presently the child appeared out of breath and was complaining of aches and pains, and tiredness. On examination, he was clinically anaemic, icteric and showed pallor and signs of retarded growth and development. His blood pressure was 98/68 mmHg and pulse was 100 beats per minute, with wide pulse pressure and hyperdynamic precordium. The sclera was yellow, abdomen was distended and the spleen was enlarged. Urine was analyzed by the staff-nurse in the side lab; presence of abnormally large amounts of urobilinogen was reported.

The child was admitted for the treatment of respiratory tract infection and for further investigations. Emergency blood sample was sent to the biochemistry and haematology laboratory. A sputum sample was also analyzed. The antibiotic was changed when pneumococci were isolated from the sputum sample. The boy responded well to the new antibiotic and the fever rapidly subsided. However, most other symptoms persisted as before. Results of some of the blood tests are as here.

Test	Patient's reports	Reference range
Haemoglobin	5.2 g/dL	13–16 g/dL
Red Blood cells	2 ($\times 10^{12}$)/L	4.5–6.5 $\times 10^{12}$/L
Platelets	190,000/μL	150,000–400,000/μL
Serum		
Bilirubin	4.8 mg/dL	0.1–1.0 mg/dL
Alanine transaminase	48 U/L	10–40 U/L
Alkaline phosphatase	90 U/L	40–100 U/L
Lactate dehydrogenase	386 U/L	100–300 U/L
Sodium	138 mmol/L	132–144 mmol/L
Potassium	5.5 mmol/L	3.6–5.0 mmol/L

The van den Bergh test for water soluble (conjugated) and water insoluble (unconjugated) bilirubin in the serum sample showed elevation of unconjugated bilirubin level. Bilirubin was not detected in the urine sample. However, the urine contained large amount of urobilinogen (i.e. urobilirubinuria), as stated above.

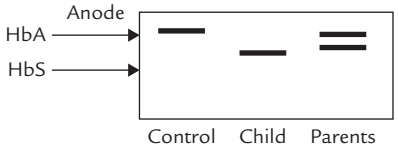
A fresh blood smear contained a few crescent-shaped cells. However, after 24-hour incubation in a sealed wet smear, nearly all the red cells assumed the shape of sickles. An increased reticulocyte count was also reported in this smear.

Electrophoresis of haemoglobin was performed to detect haemoglobin structural variant, if any. It revealed presence of HbS. Based on these tests, the child was diagnosed as having sickle cell anaemia.

Peptide maps (or finger prints) of the trypsin digestion of the haemoglobin of this child were obtained and compared with that of a normal healthy child of the same age group. The two were found to be remarkably different.

Q.1. How does the above diagnosis explain clinical features of the child?

Q.2. Paper electrophoresis of the haemoglobin obtained from the child's parents was also carried out; the results are shown below.



Interpret the results and discuss whether the child's parents also have clinical abnormalities like him?

Q.3. Comment on the quantitative blood picture. What are other biochemical features of this disorder?

Q.4. State principle of the peptide finger print test. How does the finger print test of this child differ from that of a normal subject? Give reason for your answer.

CHAPTER

VITAMINS

18

Vitamins are organic nutrients of low molecular mass (< 1500 Da) that play a central role in metabolism. Most vitamins are either not synthesized or are synthesized in inadequate amounts by the human organism; therefore, they must be obtained from an exogenous source, such as diet or bacterial flora in the gut.

Dietary vitamin deficiency may lead to diseases and even death. Vitamins and their utility for humans were known much before their structure and function were elucidated. It was observed that in some countries where the staple food of the people was rice, a disorder called "beriberi" was prevalent. In Sinhalese, beriberi means 'I cannot' (said twice), signifying that the patient is too ill to do anything. Diet of these patients was believed to be deficient in some essential organic nutrient which was subsequently purified by a Polish biochemist, Caimir Funk. Since this nutrient exhibited properties of an amine, it was called "vitamine", denoting that it was like an amine. Later, a number of other essential nutrients were discovered and it was revealed that all of them were not amines. Therefore, the suffix '-e' was dropped and the name "vitamin" was adopted. Presently 13 different vitamins are known to be required in the diet of the humans and many animal species for normal growth and function.

Vitamins are biosynthetic precursors of physiologically active forms called coenzymes. Coenzymes participate in many enzyme reactions. In addition, certain specialized functions in humans are performed by vitamins. For example, vitamin A plays an important role in visual process. The biology of vitamins may be examined from two viewpoints: nutritional and biochemical. The former is concerned with minimum daily requirements, dietary sources, bioavailability, and deficiency syndromes. The biochemist examines structure, conversion to coenzymes, mechanism of action, mode of transport and storage, metabolism and biochemical role. Both the aspects are covered in this chapter, though the emphasis will be on biochemical properties of vitamins.

After going through this chapter, the student should be able to understand:

- Sources, daily requirements and deficiency of vitamins.
- Water-soluble vitamins: structure, biochemical role and nutritional disorders of riboflavin, niacin, pyridoxine, thiamine, pantothenic acid, cobalamin, biotin and folic acid.
- Fat-soluble vitamins: structure, biochemical role and nutritional disorders of vitamins A, D, E, and K.

I. Classification and Nomenclature

Vitamins are divided into fat-soluble and water-soluble groups (Table 18.1). The **fat-soluble** vitamins are A, D, E, and K, whereas **water-soluble** vitamins include B group of vitamins (B_1, B_2, B_3, B_5, B_6, B_7, B_{12} and folate) and ascorbic acid (vitamin C).

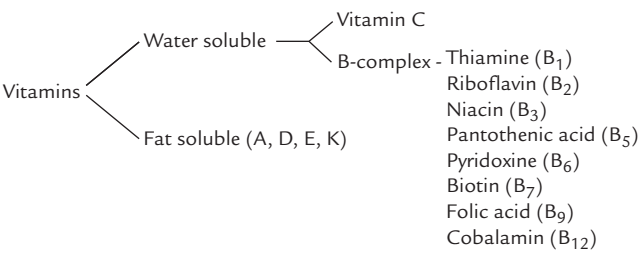


Table 18.1. Dietary sources and daily requirement of vitamins and the attendant deficiency diseases

Name	Daily requirement	Deficiency disease	Dietary sources
Water-soluble			
• Vitamin B_1 (Thiamine)	1.0–1.5 mg	Beriberi and Wernicke–Korsakoff syndrome	Legumes, pork, liver, nuts, the germ of cereals, yeast and outer layers of seeds.
• Vitamin B_3 (Niacin)	12–20 mg	Pellagra	Unrefined grains, yeast, liver, legumes and lean meat.
• Vitamin B_2 (Riboflavin)	1.1–1.5 mg	Ariboflavinosis (rare)	Milk, eggs, liver, and green leafy vegetables.
• Vitamin B_6 (Pyridoxine)	1.6–2 mg on a 100 g protein diet	Rare	Whole-grain cereals, wheat, corn, nuts, muscle meat, liver and fish.
• Pantothenic acid	5–10 mg	Rare	Yeast, liver and eggs
• Vitamin B_7 (Biotin)	5 µg/1000 kcal	Rare	Liver, kidney, milk, egg yolk, corn, and soya milk.
• Vitamin B_{12} (Cyanocobalamin)	3 µg	Pernicious anaemia	Liver, kidney, meats and milk
• Folic acid	400 µg	Megaloblastic anaemia	Liver, yeast and green vegetables.
• Vitamin C (Ascorbic acid)	45 mg	Scurvy	Citrus fruits, potatoes, particularly the skin, strawberries, raw or minimally cooked (green) vegetables and tomatoes; amla is the richest source.
Fat-soluble			
• Vitamin A	3500 IU for men 2500 IU for women	Night blindness; skin lesions	Liver, kidney, butter fat, oils, egg yolk, green leafy vegetables, fruits.
• Vitamin D	400 IU (10 µg vitamin D_3)	Rickets in children, osteomalacia in adults	Salt water fish, liver, egg yolk and butter.
• Vitamin E (Tocopherols)	10–30 mg	Liver atrophy, red blood cell haemolysis, neurological disorders	Vegetables, milk, seed oils, liver and eggs.
• Vitamin K	1 µg/kg	Bleeding tendency	Spinach, cabbage, egg yolk.

The classification in two broad groups—fat-soluble and water-soluble—has persisted, although the *compounds within a given group have widely different structures and functions*. For example, vitamin A, D, E, and K of the fat-soluble group have a limited structural resemblance with one another and perform vastly different roles. But **all fat soluble vitamins have same common properties:**

1. They are *soluble in fats,* and therefore handled in the same way with respect to their absorption from the gut, transport and distribution.
2. They may be stored in the body.
3. They may be *toxic* or even lethal if taken in excessive quantities.

Water soluble vitamins cannot accumulate to toxic levels in the body as the excessive intake results in their excretion in the urine. *Except for vitamin B_{12} there is no storage capacity for the water-soluble vitamins,* so their intake has to be more frequent than fat-soluble vitamins that are stored. A well-nourished adult, for example, may have three years' supply of vitamin A but only three months' supply of vitamin C.

Some vitamins of the B-complex group are known as the **energy-releasing vitamins** because they participate in the energy-yielding catabolic pathways. Examples include vitamins B_1, B_2, B_3, B_5, B_7, etc. Folic acid and vitamin B_{12} are loosely referred to as **haematopoietic**.

Vitamins are essential micronutrients that serve specific functions in metabolism. They can be subdivided into fat- and water-soluble.

II. Sources, Daily Requirements and Deficiency of Vitamins

Humans obtain vitamins from two sources:

1. ***Diet*:** Since vitamins cannot be manufactured by humans, they must be obtained from food (exception is vitamin D).
2. ***Intestinal microorganisms*:** Some vitamins can be synthesized by the intestinal microorganisms. However, the quantity synthesized in this manner may not be sufficient to meet the daily requirement. Biotin, however, is an exception; it is synthesized in larger quantities than required by the body.

Vitamins are referred to as **micronutrients** because their dietary requirements are in negligible quantities, amounting to few micrograms or milligrams per day. The dietary requirements for vitamins are specified in terms of a **recommended daily allowance** (RDA). The RDA defines not a minimal requirement but dietary intake that is considered optimal under ordinary conditions. Age, sex, body weight, diet and physiological status have a significant effect on the RDA. Thus, increased dietary intake of many vitamins are recommended during pregnancy and lactation. Table 18.1 summarizes the RDAs for most important vitamins for a 70 kg male.

A number of water-soluble vitamins are grouped into **B-complex** because they occur together in the same food source.

Dietary Vitamin Deficiencies

May result in pathology and even death, as noted earlier. Various causes of vitamin deficiency are as below:

1. **Inadequate dietary intake** due to faulty dietary habits or poverty.
2. **Inadequate intestinal absorption**, which may result from various gastrointestinal disorders, or from ***biliary obstruction***. The latter may lead to deficiency of the fat-soluble vitamins (A, D, E, and K) since bile is required for their absorption. Lack of *intrinsic factor*, a factor required for the intestinal absorption of vitamin B_{12}, results in deficiency of this vitamin which leads to pernicious anaemia. Some vitamins participate in enterohepatic circulation. Hence, deficiency of such vitamins is a natural consequence of impaired enterohepatic circulation.
3. **Inadequate utilization** by the target tissues may also lead to vitamin deficiency, even when dietary intake and absorption are normal. Causes of inadequate use include:
   (i) ***Lack of transport*** proteins that carry the vitamins to the peripheral sites where they are utilized.
   (ii) ***Defective uptake*** of the vitamins by the target tissues or impaired interaction of the (fat-soluble) vitamins with the receptors located on target tissues.
   (iii) Failure to convert the vitamin precursor(s) to activated form(s).
4. **Increased requirements** which occur during *pregnancy*, *lactation*, *growth*, *wound healing* and *convalescence*. In some cases, an increased requirement precipitates a borderline deficiency into a frank deficiency.
5. **Drug-induced deficiency** such as loss of vitamin synthesis in the gastrointestinal tract due to elimination of the microorganisms by antibiotics. Pyridoxine deficiency develops in patients receiving *isoniazid* for the treatment of tuberculosis (isoniazid forms a hydrazone with pyridoxal; the pyridoxal–hydrazone is rapidly excreted in urine, and the vitamin deficiency results).

Absolute or relative deficiencies of individual vitamins cause characteristic diseases or syndromes, whose symptoms can be traced to specific metabolic functions of the missing nutrient. They are a major public health concern, especially for such at-risk groups as infants, pregnant women, alcoholics and the elderly.

III. Water-soluble Vitamins

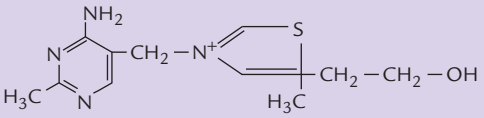
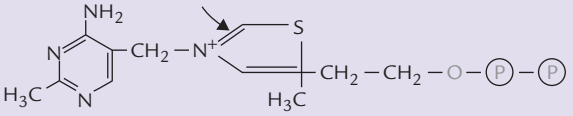
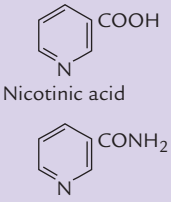
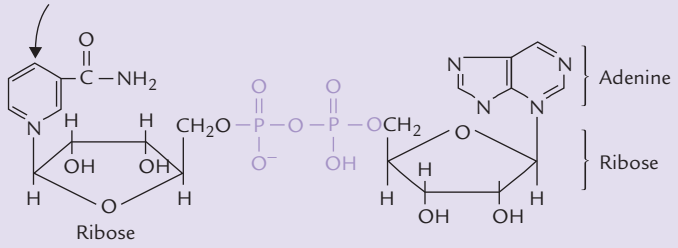
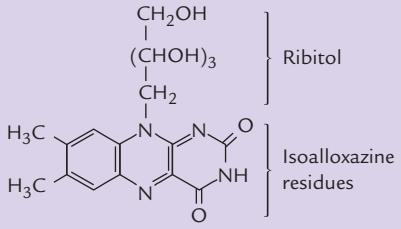
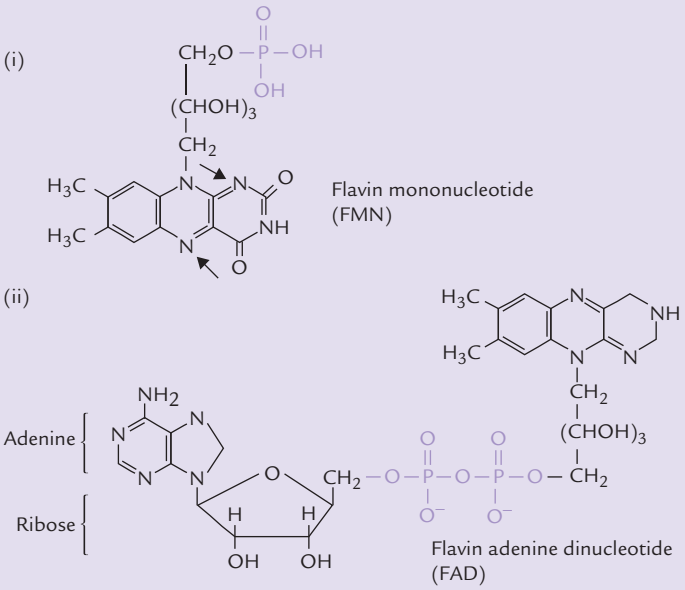
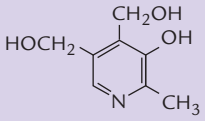
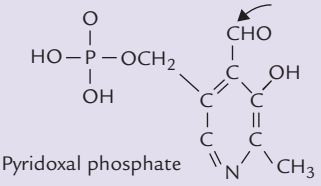
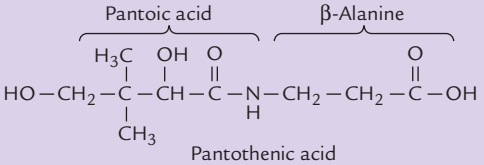
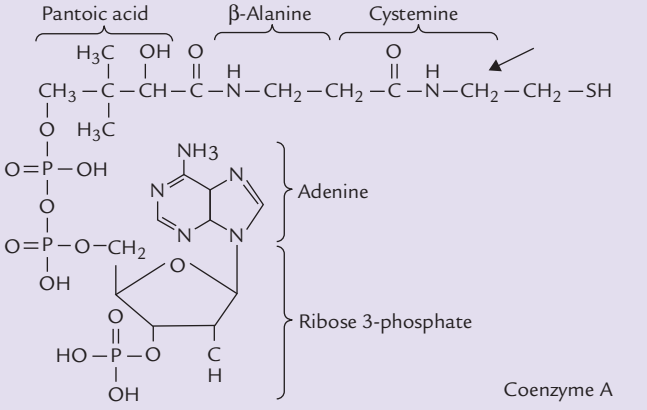
The *water-soluble vitamins are the largest class of essential micronutrients*, performing a broad range of biological functions. They are biosynthetic precursors of coenzymes for various enzymes, thereby playing pivotal role in cellular metabolism. Table 18.2 lists vitamins with their structures and coenzyme forms.

Most water-soluble vitamins are precursors of coenzymes: riboflavin, for example, is required for the synthesis of the flavin coenzymes, niacin for NAD^+ and $NADP^+$, thiamine for thiamine pyrophosphate, pantothenic acid for coenzyme A, and folic acid for tetrahydrofolate.

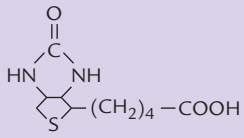
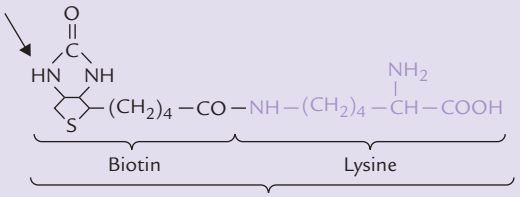
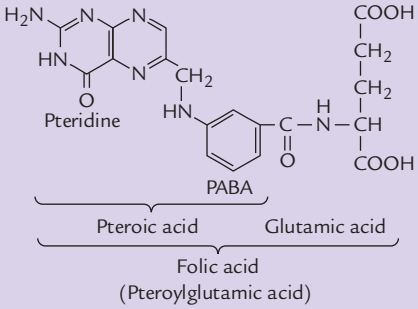
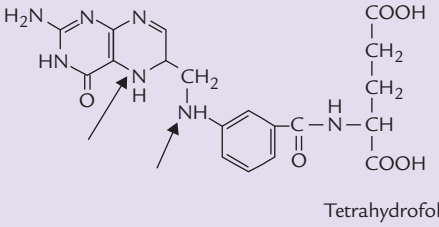
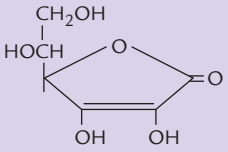
A. Thiamine (Vitamin B_1)

Thiamine, also called aneurine, was first recognized by a Dutch physician, *Christian Eijkman* in 1926 as the **antiberiberi factor**.

Table 18.2. Water-soluble vitamins: structures and coenzyme forms. The arrows indicate the portions which are important for biological activity

Name	Structure	Coenzyme structure
Thiamine		
Niacin		
Riboflavin		
Pyridoxine		
Pantothenic acid		

(Contd)

Name	Structure	Coenzyme structure
Biotin		
Folic acid		
Ascorbic acid		

*Biocytin is biotin bound to ϵ group of lysine.

Structure

As elucidated by *Adolf Windaus*, thiamine consists of a substituted pyrimidine residue linked to a thiazole (sulphur and nitrogen) ring through a methylene bridge. Thiamine-dependent reactions are *aldehyde transfers*, in which the aldehyde is bound covalently to one of the carbons in the thiazole ring, shown by arrow in Table 18.2.

Absorption and Transport

Dietary thiamine is absorbed readily from jejunum and proximal ileum and is transported to tissues where it is converted to the active form, **thiamine pyrophosphate**. Two different mechanisms are involved in absorption depending on the level of intake:

1. At levels below 5 mg/day, the absorption occurs by an **active, ATP-dependent process**, which is saturable at concentrations of 0.5–1.0 μmol/L.
2. At levels higher than this, the absorption occurs by **passive diffusion**.

Ethanol inhibits the active transport of thiamine, and this could be one of the causes of thiamine deficiency in alcoholics, whose thiamine intakes are usually also low because of their penchant for drinking but not eating.

Coenzyme Functions

Following absorption, thiamine is transported to tissues bound to albumin and, to a small extent, other proteins. After cellular uptake it is converted into its active form, **thiamine pyrophosphate** (TPP) in an ATP-dependent reaction catalyzed by *thiamine pyrophosphokinase*. Total amount of thiamine in the body is approximately 30 mg, about 80% of which is in the form of TPP.

$$\text{Thiamine} \xrightarrow{\textit{Thiamine pyrophosphokinase}} \text{TPP}$$

TPP serves as a coenzyme in several enzyme catalyzed reactions, which involve *aldehyde transfer*. These include oxidative-decarboxylation, transketolase reaction, decarboxylation, etc.

1. ***Oxidative-decarboxylation:*** TPP is a component of mitochondrial multienzyme complexes which bring about oxidative-decarboxylation of α-keto acids, e.g. pyruvate and α-ketoglutarate.

$$\text{Pyruvate} \xrightarrow{\textit{Pyruvate dehydrogenase}} \text{Acetyl CoA}$$

$$\alpha\text{-Ketoglutarate} \xrightarrow{\textit{α-Ketoglutarate dehydrogenase}} \text{Succinyl CoA}$$

Malfunction of TCA is thus an obvious consequence of thiamine deficiency, and results in defective energy metabolism.

Moreover, acetyl CoA is a precursor for the synthesis of the neurotransmitter acetylcholine, and also for the synthesis of lipids, including myelin, which may explain the importance of thiamine in the correct functioning of the nervous system.

2. ***Transketolase reactions:*** TPP is also a coenzyme for the *transketolase*, in the pentose phosphate pathway. The pathway generates ribose sugars and supplies NADPH necessary for a wide variety of redox and biosynthetic reactions.
3. ***Decarboxylation reactions:*** TPP serves as a coenzyme for decarboxylation reactions in the metabolism of branched chain amino acids. Patients with maple syrup disease suffer from a defect in the decarboxylation reaction during metabolism of these amino acids, and therefore, may be successfully treated with large doses of thiamine.
4. ***Others:*** Another thiamine derivative, *thiamine trisphosphate*, is known to be involved in nerve conduction, but the precise role of the vitamin in this respect is not clear.

Requirement and Dietary Sources

Thiamine is related to energy metabolism, therefore its dietary requirement depends on the caloric intake. The recommended daily allowance is **0.5 mg/1000 kcal** for adults with the recommendation for a minimum absolute intake of 0.5 mg/day. Dietary sources are given in Table 18.1.

Deficiency

Thiamine deficiency leads to **beriberi**. In rice-consuming areas of Asia, where it is customary to polish this staple grain to remove its coarse, thiamine-containing outer layer, the disease is most prevalent. Patients with history of malnutrition, chronic inflammation and parenteral nutrition are especially prone to develop thiamine deficiency.

Thiamine deficiency, however, is not common in developed western nations, except in chronic alcoholics.

Mild thiamine deficiency is associated with weakness and gastrointestinal disturbances. Peripheral neuropathy, mental confusion and ataxia develop in **moderate deficiency** conditions; and, severe deficiency presents with severe neuromuscular and cardiovascular disorders.

The **full blown deficiency**, known as *beriberi*, is categorized into the following types.

1. ***Dry beriberi*:** The *neuromuscular symptoms* predominate in this type. In longstanding cases, there is degeneration and demyelination of sensory and motor nerves and severe wasting of muscles.
2. ***Wet beriberi*:** This type develops in more severe cases of thiamine deficiency. The *cardiovascular manifestations* appear in this type and oedema is a notable feature. These signs and symptoms are largely accounted by accumulation of pyruvic and lactic acids **(Case 18.1)**.
3. ***Infantile beriberi*:** This type develops in breastfed infants, mostly between the second and the fifth month, in the areas where beriberi is endemic. Classical signs of beriberi, however, are present in only about 50% of the mothers.
4. ***Thiamine deficiency due to alcoholism:*** Chronic alcoholism does not result in beriberi but in **Wernicke-Korsakoff syndrome**. The acute stage of this disease, known as Wernicke's encephalopathy, is characterized by mental derangement, delirium and ataxia. In the chronic stage, known as *Korsakoff psychosis*, the patient has anterograde amnesia.
5. ***Thiamine deficiency due to thiaminase:*** Thiamine can be destroyed if the diet contains *thiaminases*, which cleave the pyrimidine ring from the thiazole ring. These enzymes are present in some ferns, raw fishes and sea foods, and they are thought to contribute to the incidence of beriberi in areas in Japan where raw fish (Sushi) is a common dietary constituent.

Assessment of Thiamine Status

Thiamine status is evaluated by estimating *urinary thiamine excretion* and *plasma levels of pyruvate and lactate*, particularly after an oral glucose load (these acids accumulate because of the decreased activity of *pyruvate dehydrogenase*). Determination of *erythrocyte transketolase activity*, which requires TPP as a coenzyme, confirms the deficiency **(Case 18.1).**

Toxicity

Excess intake of thiamine, like other water soluble vitamins, is usually not toxic because it is promptly excreted in the urine. However, it has been recorded that chronic intake of thiamine in excess of 3 g/day are toxic to adults, causing headaches, irritability and dermatitis, and in extreme cases may even cause death.

B. Redox Vitamins: Niacin (Vitamin B_3) and Riboflavin (Vitamin B_2)

Niacin and riboflavin give rise to coenzymes that participate in *redox reactions*.

Niacin is converted to the coenzymes, NAD^+ and $NADP^+$, and riboflavin to flavin mononucleotide (*FMN*) and flavin adenine dinucleotide (*FAD*).

Niacin

Niacin is a nutritional term used to refer to the *vitamers* (different structural forms of a vitamin with the same biological activity): **nicotinic acid** and **nicotinamide**. The word niacin is coined from letters of three words: nicotinic acid and vitamin. This was to avoid confusion with nicotine, a compound present in tobacco. **Pellagra** caused by niacin deficiency, was once endemic among poor peasants of

Latin America; these people subsisted chiefly on maize (American corn), which is deficient in tryptophan.

Because nicotinic acid can be formed in the body from **tryptophan** (*1 mg of the vitamin from 60 mg tryptophan*), deficiency of this amino acid leads to deficiency of nicotinic acid as well.

Structure

Structures of the two vitamers, nicotinic acid and nicotinamide are given in Table 18.2. Nicotinic acid is a pyridine-3-carboxylic acid. Niacinamide, the form present in tissues, is the acid amide.

Coenzyme Forms of Niacin

Niacin, in the form of nicotinamide, is incorporated into the structure of two coenzymes: *nicotinamide adenine dinucleotides* (**NAD**$^+$) and *nicotinamide adenine dinucleotide phosphate* (**NADP**$^+$). The niacin is attached to a ribose phosphate to form a mononucleotide, which is then attached to AMP to form the nicotinate adenine dinucleotide (Fig. 18.1). The latter reacts with glutamine to form its amide, i.e., NAD$^+$. The nitrogen atom of nicotinamide contains one positive charge, hence the structure is abbreviated with a positive sign, as NAD$^+$.

In the case of NADP$^+$, a phosphate group is attached to the 2′-hydroxyl group of ribose (of the AMP).

Coenzyme Activities

1. ***Oxidation and reduction:*** Both the coenzymes, NAD$^+$ and NADP$^+$, participate in a vast variety of oxidation-reduction reactions. **NAD**$^+$ is used largely in the *catabolic redox reactions* occurring in the mitochondrial matrix Table (14.1). NADP$^+$ in its reduced form, **NADPH,** is used largely in the *reductive biosynthesis reactions* (e.g. cholesterogenesis and lipogenesis) in the extramitochondrial compartment of the cell. Thus, NAD$^+$ and NADP are present in distinct cellular compartments and catalyze different reaction types. In line with this is the observation that in rat liver cytosol, NADPH : NADP$^+$ ratio is about 80, whereas the NADH : NAD$^+$ ratio is only 8×10^{-4}.
2. ***ADP-Ribosylation:*** NAD$^+$ functions as an ADP-ribose donor. Such reactions are called the ADP-ribosylation reactions, and they occur in the nucleus; for example, the eukaryotic post-translational modifications catalyzed by the enzyme *poly ADP-ribose polymerase*. The enzyme seems to play a role in DNA repair and other cellular responses to DNA damage.

Clinical Deficiency

The niacin deficiency disease is **pellagra** (derived from Italian words *pelle* = skin + *agro* = rough) involves the gastrointestinal tract, skin and central nervous system. The symptoms comprise **three D's**: *diarrhoea*, *dermatitis* and *dementia*. The fourth D, i.e. death, follows in the untreated cases. The dermatitis, which resembles sunburn, is seen in areas of the skin exposed to the sunlight. It is probably related to the role of NAD$^+$ in DNA repair reactions following damage caused by the UV light.

Pellagra is also seen in conditions where dietary amino acids, including tryptophan, are not properly absorbed,

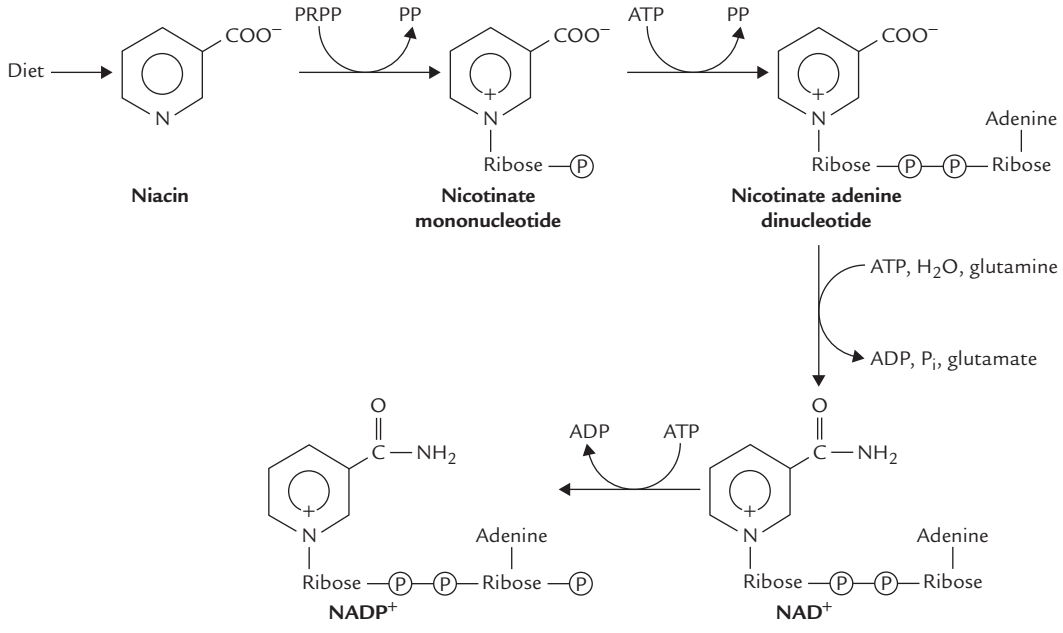


Fig. 18.1. Synthesis of nicotinamide adenine dinucleotide (NAD$^+$). Attachment of a phosphate group to a 2′–OH in the NAD$^+$ yields NADP$^+$.

as in *Hartnup's disease*, a genetic disorder of amino acid transport. Severe pellagra can be fatal.

Niacin (B_3) is a generic name for nicotinic acid or nicotinamide, either of which is an essential nutrient. Severe niacin deficiency produces dermatitis, diarrhoea and dementia.

Most cases of pellagra respond to nicotinamide in a dose of 100 mg every 4 to 6 hours: response is so rapid that within 24 hours the erythema of skin diminishes, the diarrhoea ceases and there is striking improvement in the patient's behaviour and mental attitude. Nicotinic acid is equally effective for the treatment.

Therapeutic Uses

Nicotinic acid is used in pharmacological doses (several grams daily) for the treatment of hyperlipidaemias. It is observed to lower total plasma cholesterol, LDL-cholesterol and VLDL-triacylglycerols in patients with hyperlipoproteinaemias. It also lowers the circulating free fatty acid levels by inhibiting *hormone-sensitive lipase,* thereby blocking adipose tissue lipolysis. Hypoglycaemic effects with large doses (1–2 g daily) have also been observed. *These effects are unrelated to the vitamin activity of nicotinic acid.*

The large doses necessary to produce the lipid- and glucose-lowering effects, have undesirable side effects, mainly vasodilatation and flushing. Cases of hepatic toxicity have also been reported in some patients.

Riboflavin

Structure

Chemically, riboflavin is 6,7-dimethyl-9D-ribitol isoalloxazine, consisting of an *isoalloxazine ring* with a *ribitol side chain* (Table 18.2). Ribitol is an alcohol of ribose sugar. The ring system of isoalloxazine accounts for the intense yellow colour of the vitamin and its emission of greenish-yellow fluorescence following exposure to UV light.

Coenzyme Activity

Riboflavin is a component of two coenzymes: flavin mononucleotide (**FMN**) and flavin adenine dinucleotide (**FAD**). They are known as **flavin nucleotides**. FMN is formed (in intestine) by attachment of a phosphate group to the ribitol side chain. FAD, which contains adenosine linked via phosphate group to FMN, is formed by adenylation of the latter in liver (Fig. 18.2).

The flavin nucleotides serve as prosthetic groups for the enzymes called *flavin-dependent* (*flavoprotein*) *enzymes*. These enzymes remove a pair of hydrogen atoms from the substrate and thereby participate in a number of oxidation-reduction reactions in metabolism.

FMN-dependent enzymes:

- FMN is a cofactor for *L-amino acid oxidase*.
- The enzyme *NADH dehydrogenase,* of respiratory chain contains FMN.

FAD-dependent enzymes:

- FAD is a constituent of the Complex II of the respiratory chain.
- FAD is a constituent of the *microsomal hydroxylase* system.
- FAD is a cofactor for several enzymes, e.g. *D-amino acid oxidase, succinate dehydrogenase, acyl CoA dehydrogenase, glycerol 3-phosphate dehydrogenase, xanthine oxidase, pyruvate dehydrogenase, α-ketoglutarate dehydrogenase,* etc.

During oxidation process, FAD accepts two hydrogen atoms from substrate and gets reduced to $FADH_2$. The two nitrogen atoms of the isoalloxazine ring accept the hydrogen atoms, as indicated by the arrow in Table 18.2. FMN is likewise reduced to $FMNH_2$.

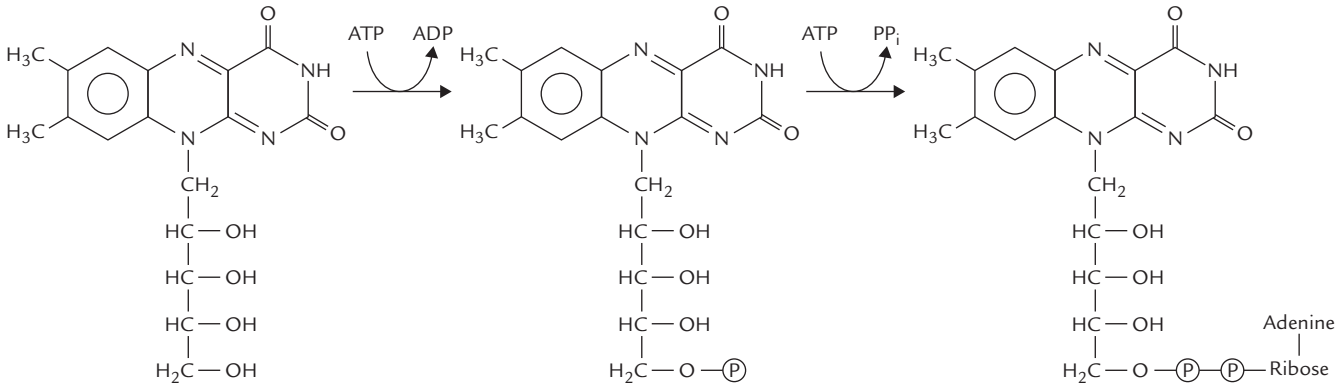


Fig. 18.2. Synthesis of FMN and FAD from dietary riboflavin.

Absorption, Transport and Storage

Riboflavin is ingested in form of flavoproteins. The FAD and FMN components are released from the protein complex in the stomach, and free riboflavin is released in the intestine, from which it is absorbed by an *active ATP-dependent process*. The activation of riboflavin via an ATP-dependent enzyme system occurs next, resulting in the production of FMN and FAD. The main storage form of the vitamin, found mainly in the liver, is FAD.

Clinical Deficiency

Dietary requirement of riboflavin is low (Table 18.1), and the intestinal flora synthesizes it. Therefore, *isolated deficiency of riboflavin is rare*. It is generally combined with other deficiencies, such as beriberi, pellagra and kwashiorkor. Owing to light sensitivity of riboflavin, phototherapy for physiological jaundice may induce transient riboflavin deficiency in infants.

Deficiency Manifestations

Symptoms are confined to skin and mucous membranes. *Angular stomatitis* (inflammation of the mouth), *glossitis* (inflammation of the tongue), *cheilosis* (reddening of the mucous membrane of lips), *seborrheic dermatitis* (rough and scaly skin around nasolabial and scrotal areas), corneal vascularization, and a form of peripheral neuropathy are the prominent features. The condition is referred to as **ariboflavinosis**.

The symptoms of ariboflavinosis are relatively mild and certainly not life threatening even though flavoproteins are essential for life. There are two main reasons for this. One is that riboflavin is associated with proteins in the diet and any diet providing protein will also provide a fair amount of riboflavin. The other reason is that the recycling of riboflavin released from FAD and FMN is extremely efficient and, therefore, only small amounts need be ingested.

Like thiamine deficiency, riboflavin deficiency can be seen in conditions such as *chronic alcoholism*, malnutrition, anorexia and malabsorption. *Drugs*, such as *barbiturates*, may also cause riboflavin deficiency by inducing microsomal oxidation of the vitamin.

Laboratory diagnosis of riboflavin deficiency is difficult. Serum and urine riboflavin fall low only in severe deficiency. Erythrocyte enzyme activity measurement ***glutathione reductase*** (a riboflavin-dependent enzyme) used as index of riboflavin status.

C. Pyridoxine (Vitamin B_6)

Vitamin B_6 plays a key role in amino acid metabolism, being required in several reactions for the synthesis, catabolism and interconversion of amino acids (Chapter 13). It consists of three different vitamers: *pyridoxine, pyridoxal*, and *pyridoxamine*, all of which can be phosphorylated and converted to the active form, **pyridoxal phosphate** (Fig. 18.3). Pyridoxine is the major form of vitamin B_6 in diet, occurring widely in both animal and plant tissues.

Structure

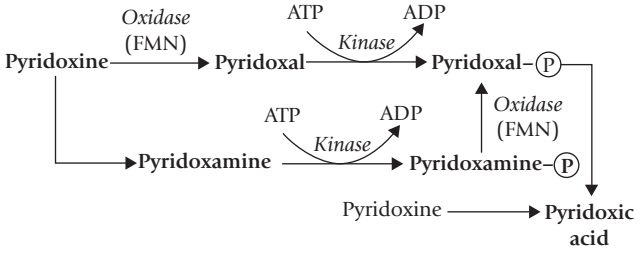
All forms of the B_6 vitamin are **pyridine derivatives**, differing from one another only in nature of the functional group attached to the pyridine ring.

The aldehyde group of PLP forms aldimine derivatives with primary amino groups of amino acids.

Absorption, Transport and Excretion

Vitamin B_6 is rapidly absorbed from the intestine by *passive diffusion*. The absorption is highly efficient: about 80% of the ingested vitamin is absorbed. Phosphorylated pyridoxine vitamers are hydrolyzed by intestinal membrane *alkaline phosphatase* and dephosphorylated forms are absorbed.

The major circulating form of vitamin B_6 is pyridoxal phosphate (PLP). It is produced from the absorbed pyridoxine and pyridoxamine. Pyridoxic acid is the principal excretory form of the vitamin in urine. Its formation is catalyzed by *aldehyde oxidase*.



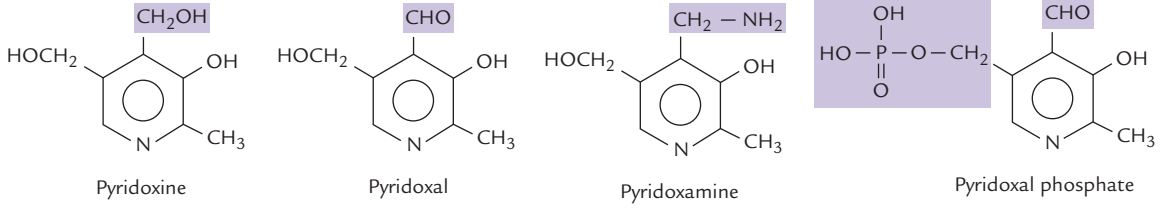


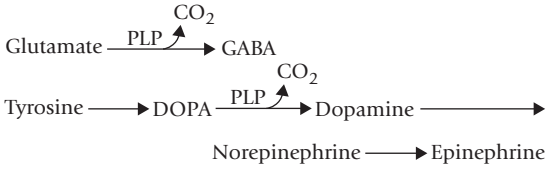
Fig. 18.3. Structures of B_6 vitamins. Pyridoxine is primary alcohol, pyridoxal is an aldehyde, and pyridoxamine is an amine.

Limited storage of pyridoxine is possible: most of the body stores are associated with the enzyme *glycogen phosphorylase*, for which it is a coenzyme. In fact, about 70% of total PLP content of the body resides in muscles, in association with this enzyme.

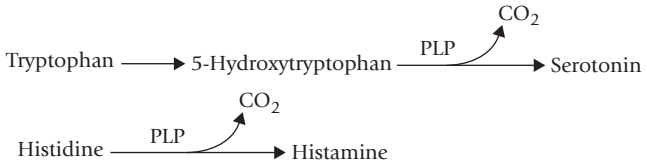
Coenzyme Functions

Pyridoxal phosphate serves as a coenzyme for a broad range of reactions in intermediary metabolism, especially of amino acid metabolism where it seems to play a central role (Chapter 13). For this reason, *requirement of B_6 also increases with protein intake*. Some of the more important PLP-dependent reactions are:

1. ***Transamination***: PLP is a coenzyme for *transaminases*, where it acts as an amino group carrier (Chapter 13).
2. ***Decarboxylation***: All decarboxylation reactions of amino acid metabolism require PLP. Particularly important in this respect are *glutamate decarboxylase* and *DOPA decarboxylase*. These enzymes are involved in the production of γ-aminobutyric acid (GABA) and catecholamine neurotransmitters (epinephrine, norepinephrine, dopamine) respectively in the nervous system.



Synthesis of the serotonin neurotransmitter also requires B_6, as does that of histamine.



3. ***Condensation***: The enzyme 8-*aminolevulinic acid synthase* that catalyzes joining of glycine and succinyl CoA in the haem biosynthetic pathway requires PLP as a cofactor (Chapter 16). This accounts for anaemia seen in B_6 deficiency.
4. ***Transulfuration***: PLP is a coenzyme for *cystathionine synthase* and *cystathionine γ-lyase*, which play important role in methionine and cysteine metabolism (Fig. 13.15). In these reactions transfer of sulphur from methione to serine occurs to produce cysteine. *In PLP deficiency, blood level of homocysteine (substrate for cystathionine synthase) is increased, which is correlated with occlusive vascular diseases.* Therefore, pyridoxine intake may be recommended in clinical practice to prevent CAD.
5. ***Tryptophan metabolism***: PLP is a cofactor for the enzyme *kynureninase*, which catalyzes release of alanine from 3-hydroxykynurenine to form 3-hydroxyan-thranilic acid, during catabolism of tryptophan (Fig. 13.18).

   In deficiency of PLP, 3-hydroxyanthranilic acid accumulates and is converted into an alternate metabolite, xanthurenic acid which is excreted in urine. *The urinary excretion of xanthurenic acid serves as an indicator of pyridoxine deficiency*: it increases in the deficiency state.
6. ***Others***: PLP is a cofactor for *glycogen phosphorylase* in both liver and muscle, and therefore, favours glycogenolysis. PLP also serves as coenzyme for the enzymes responsible for the specific deamination of serine, threonine and cysteine. It is also required for the synthesis of sphingosine, a component of sphingomyelin and sphingolipids. Cellular uptake of L-amino acids also requires participation of PLP.

Clinical Deficiency

Deficiency of vitamin B_6 is rare because of its widespread distribution in a variety of foodstuffs. Deficiency symptoms are most often seen in *alcoholics*, women taking oral contraceptives, and in infants fed with formula diet low in this vitamin. Probably the commonest cause of deficiency is drug antagonism. **Isoniazid**, an antitubercular drug, reacts with PLP to form a hydrazone that is biologically inactive, rapidly excreted in urine, and also inhibits *pyridoxal kinase*. Penicillamine, used in the treatment of Wilson's disease and rheumatoid arthritis, also combines with PLP to render it unavailable.

The major symptoms of B_6 deficiency include neural dysfunction and anaemia. The former is accounted by an impairment in the synthesis of neurotransmitters such as norepinephrine and serotonin. Anaemia is accounted by impairment of haem biosynthesis. Other features of B_6 deficiency include lesions of the skin and mucosa, *sideroblastic anaemia* and personality changes. In rare instances, vitamin B_6 can be toxic and cause convulsions at very high levels. Possible mechanism is by enhancing decarboxylation of L-DOPA in several tissues. In patients treated for parkinsonism, efficiency of DOPA is thereby reduced.

Assessment of pyridoxine status: As with other B-complex vitamins, assessment of pyridoxine status is based on the measurement of erythrocyte enzymes; the enzyme in this case is *aspartate aminotransferase*. **Tryptophan load test**, involving measurement of urinary excretion of xanthurenic acid following a load dose of tryptophan is the other important test: xanthurenic acid excretion is increased in B_6 deficiency. Measurement of urinary homocysteine and cystathionine after a methionine load, and urinary pyridoxic acid and PLP level in blood are the other important tests.

D. Pantothenic Acid

Pantothenic acid consists of pantoic acid and β-alanine (Table 18.2). It is a building block of coenzyme A and of the phosphopantotheine group in the *fatty acid synthase* complex (Chapter 11). The CoA plays a significant role in several body reactions, especially the biosynthetic pathways where the thiol (SH) group of CoA forms thioester bonds with many organic acids. This results in activation of the organic acids. Conversion of succinate to succinyl CoA and of acetate to acetyl CoA are some examples. Succinyl CoA is a precursor in haem biosynthetic pathway, and acetyl CoA in lipogenesis.

Deficiency of pantothenic acid in humans is unknown. In experimental studies, the isolated deficiency can be induced only under rigorously controlled experimental conditions. The RDA of 5–10 mg is readily met by ordinary diets.

Pantothenic acid is a building block of coenzyme A.

E. Biotin (Vitamin B_7 or H)

Biotin is a coenzyme for carbon dioxide fixation reactions. It serves as prosthetic group of ***ATP-dependent carboxylases***, including *pyruvate carboxylase*, *acetyl CoA carboxylase* and *propionyl CoA carboxylase*. It was first isolated by *Vincent du Vigneaud* in 1942 (Nobel Prize 1955).

Structure

Biotin consists of an imidazole ring that is *cis*-fused to a thiophene ring bearing a valerate side chain (Fig. 18.4). The valeryl carboxyl group is attached to the ε-amino group of a lysine residue in the apoenzyme.

Requirement, Sources and Causes of Deficiency

About *5 μg/1000 kcal* biotin is required daily. Biotin is widely distributed in foods. Liver, yeast, peanuts, kidney, and egg yolk are especially rich sources. Biotin is mostly bound to dietary protein by an amide linkage. The linkage is cleaved prior to absorption of biotin from intestine.

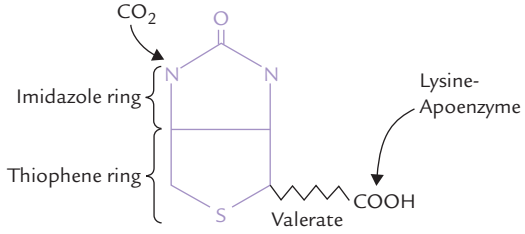


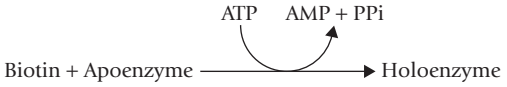
Fig. 18.4. Structures of biotin.

The intestinal flora synthesize biotin in quantities that far exceed the daily requirement. This is reflected in urinary and fecal excretion of biotin, which is 2–3 times its daily dietary intake. Thus, biotin deficiency is unknown, except in a small number of people with unusual dietary habit of consuming large amounts of uncooked eggs. Egg white contains the glycoprotein ***avidin***, a homotetramer of four subunits (MW 70,000), which binds the imidazole group of biotin tightly, thereby preventing its absorption from the intestine. However, consumption of more than 20 raw egg whites per day only is likely to induce biotin deficiency.

Sterilization of the intestine by antibiotics or administration of biotin analogues can also induce biotin deficiency.

Coenzyme Functions

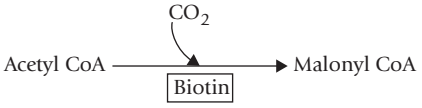
Biotin is covalently attached to the apoenzyme by an amide linkage with a lysyl ε-amino group, as noted earlier. This binding is catalyzed by the enzyme *holocarboxylase synthetase*:



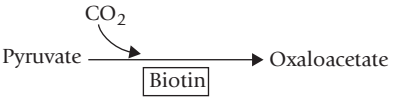
Biotin serves as a carrier of the activated carbon dioxide transferring it to various acceptor molecules, e.g. acetyl CoA, pyruvate and propionyl CoA, etc. Such reactions are termed as **carboxylation reactions** and the enzymes catalyzing them are called *carboxylases*.

The most important carboxylation reactions are:

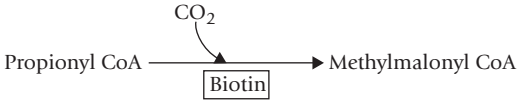
1. Conversion of acetyl CoA to malonyl CoA, catalyzed by *acetyl CoA carboxylase*, in fatty acid synthesis.



2. Pyruvate to oxaloacetate conversion, catalyzed by *pyruvate carboxylase*, in gluconeogenesis. It is an important anaplerotic reaction also.



3. Conversion of propionyl CoA to methylmalonyl CoA, by *propionyl CoA carboxylase*, in catalobism of branched chain amino acids and odd chain fatty acids



Biotin is a coenzyme in carboxylation reactions. It is synthesized by intestinal flora.

***Carboxylations not dependent on biotin*:** Not all carboxylation reactions, however, require biotin. Some examples are: (a) formation of carbamoyl phosphate by *carbamoyl phosphate synthetase*, (b) *vitamin K-dependent γ-carboxylation* of glutamyl residues of several of the clotting factors, (c) conversion of pyruvate to malate by *malic enzyme*, and (d) addition of C-6 to the purine ring.

Recycling: Proteolytic degradation of biotin-containing enzymes, both in the intestinal lumen and in the tissues, produces biotinyl-lysine or ***biocytin***. *It consists of biotin bound to the ε-group of lysine* (Fig. 18.4). *Biotinidase* cleaves biocytin to biotin and lysine. Thus, biotin is recycled.

$$\text{Biocytin} \xrightarrow{\textit{Biotinidase}} \text{Biotin + Lysine}$$

Biotinidase deficiency is an autosomal recessive disorder which disrupts the recycling. *Thus, biotinidase deficiency causes non-dietary biotin deficiency.*

Biotinidase deficiency is included in newborn screening programs for the treatable congenital diseases, together with other such disorders as phenylkenonuria, galactosaemia, maple syrup urine disease and congenital hypothyroidism in some countries.

Clinical Deficiency

Patients with biotin deficiency develop seborrhoeic dermatitis, anorexia, and alopecia with loss of hair follicles. The infants with *biotinidase* deficiency present with hypotonia, seizures, optic atrophy, dermatitis and conjunctivitis. In experimental biotin deficiency, produced by feeding excessive egg-white (at levels that provide 30% of the dietary energy), subjects suffered from muscle pain, hallucinations, depression, sleepiness as well as hair loss and dermatitis.

Toxicity due to excessive consumption of biotin is not known.

F. Cobalamin (Vitamin B_{12})

Vitamin B_{12} is important haematopoietic vitamin (together with folic acid) deficiency of which causes megaloblastic anaemic—large sized RBCs. The anaemia is associated with neurological deterioration, and called pernicious anaemia.

Structure

Cobalamin is chemically the most complex of all vitamins (Fig. 18.5). It is unique among vitamins as it contains an essential trace element, cobalt. Cobalt lies at the centre of a **corrin ring system** which consists of four pyrrole rings. The tetrapyrrole structure of corrin is similar

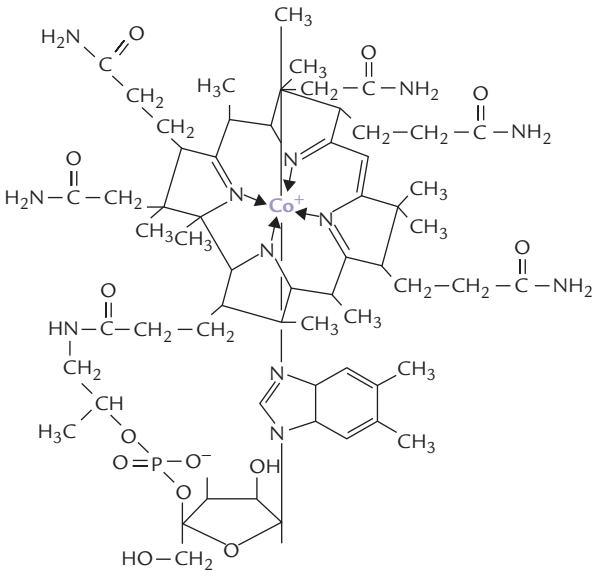


Fig. 18.5. Vitamin B_{12} (cobalamin) contains four pyrrole rings with a central cobalt atom.

to that of porphyrin, except that it is more hydrogenated and two of the pyrrole rings are linked directly rather than through a methenyl bridge.

Synthesis

Cobalamin is the only vitamin that is *synthesized by neither plants nor animals*, but only by a few species of bacteria. It is, therefore, absent from all plants but is concentrated in the liver of animals in three forms: *methylcobalamin, adenosylcobalamin* and *hydroxycobalamin*. Liver is thus a useful source of this vitamin.

Vitamin B_{12} when isolated from natural sources, is called *cyanocobalamin* because it contains a cyano group, which is essential for chelation of the cobalt ion. The cyano group needs to be removed for converting the vitamin to its active form. Interestingly, *the only known role of cobalt in mammalian systems is that it is a part of the vitamin B_{12}.*

Absorption, Transport and Distribution

Cobalamin is synthesized by microorganisms, but it enters the food chain only through foods of animal origin. It usually occurs in protein-bound forms and must be separated from these proteins by either acid hydrolysis in the stomach or trypsin digestion in the intestine. In stomach, dietary vitamin B_{12} is bound by a glycoprotein called the **intrinsic factor** which is produced by the stomach's parietal cells. The vitamin-intrinsic factor complex travels to the ileum, where the vitamin is absorbed, via a receptor in the intestinal mucosa (Fig. 18.6). The *rate-limiting factor* in this process is the number of ileal receptor sites. The complex then dissociates and the liberated vitamin B_{12} is converted to methyl cobalamin and released into the bloodstream.

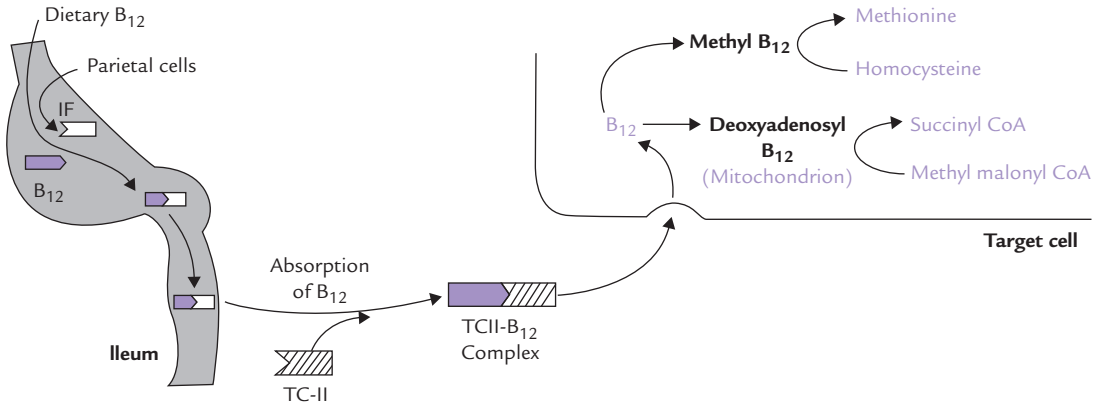


Fig. 18.6. Intestinal absorption of vitamin B_{12}, transport in bloodstream and utilization in the target cell (IF = intrinsic factor, TC-II = transcobalamin II).

BOX 18.1

Properties of Transport Proteins

Intrinsic factor, transcobalamins (TCs) and R proteins are various transport proteins for cobalamin. All these have certain common properties:

- Composed of a single polypeptide chain (340–375 amino acid residues).
- There is a single binding site for cobalamin.
- All are glycoproteins.

They do not, however, cross react with each other immunologically and are coded for by different genes.

Intrinsic factor vs R-proteins: R-proteins are cobalamin transport proteins, secreted by salivary gland and gastric mucosa. They have a higher affinity for cobalamin than IF at acidic pH. But they are partially degraded by *pancreatic proteinases* (at alkaline pH), whereas IF is resistant to these enzymes. This facilitates transfer of cobalamin from R-proteins to IF; the process being aided by the bicarbonate of pancreatic juice. Thus, in pancreatic diseases, R-proteins are not degraded and cobalamin is not available to be transferred to IF, resulting in loss of absorptive capacity for this vitamin. Vitamin deficiency is the obvious consequence.

Methyl cobalamin is the predominant circulating form. It is transported in a bound form with **transcobalamin-II** (TCII), a glycoprotein. The TCII-B_{12} complex enters virtually all cells of the body via a specific *cell-surface receptor*. The ligand receptor complex is internalized and appears in lysosomes, where TCII is degraded, thus freeing the cobalamin. The latter then passes into the cytoplasm through mediation of a specific transport mechanism.

Evidently, handling of B_{12} is far more complex than the other water-soluble vitamins, requiring participation of several transport proteins for its absorption and uptake by tissues.

Vitamin B_{12} Transport Proteins

1. *Intrinsic factor (IF)* is a highly specific glycoprotein, which accounts for absorption of about 97% of the ingested B_{12} (simple diffusion of the vitamin across intestinal membrane accounts for 3%).
2. *Transcobalamin*-I, *transcobalamin*-II, *transcobalamin*-III and *R-proteins* are the other transport proteins involved in the delivery or storage of the cobalamin. TC–II is a plasma transport protein, whereas TC-I and TC-III exist in liver also. Both provide storage forms for the B_{12} vitamin.

   R-proteins are secreted by the salivary gland and gastric mucosa, and together with transcobalamin-I and -III they are termed *cobalaphilins*.

For more information on transport proteins, refer Box 18.1.

Vitamin B_{12} can be Stored

Water-soluble vitamins cannot be stored, but vitamin B_{12} is an important exception for it is **stored in liver**. The *transcobalamin-I* and *-III provide excellent storage forms of the vitamin*. Whole liver contains about 2 mg of the vitamin which is sufficient for the requirement of 2–3 years: a situation unique in terms of water soluble vitamins, more akin to that of vitamin A and E.

Coenzyme Functions

Cobalamin is required in **two reactions** that lead to:

1. *formation of methionine* (from homocysteine), and
2. *succinyl CoA* (from methylmalonyl CoA).

1. *Homocysteine to methionine conversion:* Methylcobalamin donates a methyl group to homocysteine to form methionine; the reaction is catalyzed by the enzyme *homocysteine methyltransferase* (Fig. 18.6). After donating its methyl group, the methylcobalamin is reconverted to cobalamin. Methylcobalamin is re-formed from cobalamin by receiving methyl group of methyl-tetrahydrofolate (Fig. 18.7).

Thus, cobalamin and folate are involved together in this methylation reaction. An absence of cobalamin leads to cessation of the reaction and build up of methyltetrahydrofolate, known as the *'folate trap'*, discussed later.

2. *Formation of succinyl CoA*: Vitamin B_{12}, as deoxyadenosyl cobalamin, acts as a cofactor for the enzyme *methylmalonyl CoA mutase* (**Case 18.2**). This enzyme catalyzes conversion of methyl malonyl CoA to succinyl CoA.

As discussed earlier, methylmalonyl CoA is formed from (a) valine, isoleucine and methionine via propionyl CoA, and (b) fatty acids having odd number of carbon atoms. Succinyl CoA may be oxidized in Krebs cycle or converted to glucose.

Requirement and Dietary Sources

Dietary requirements vary:

Infants and children	1–1.5 μg/day
Adults	3.0 μg/day
Pregnancy/lactation	4.5 μg/day

Clinical deficiency manifests after a long time because normal requirement for cobalamin is very small (~3 μg/day) and the liver stores a 3- to 5-year supply of this vitamin.

Apart from liver, which is the largest source of cobalamin, kidney, meat, eggs, milk and cheese are good sources.

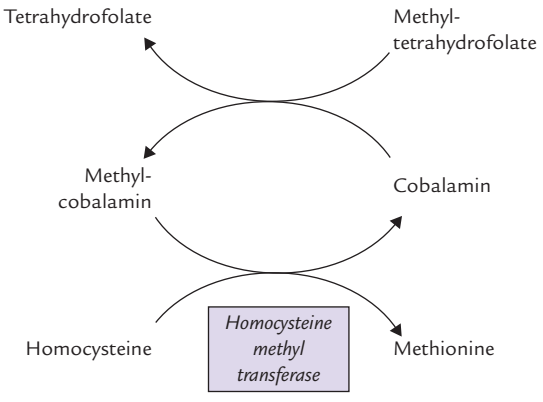


Fig. 18.7. Conversion of homocysteine to methionine requires vitamin B_{12} (as methyl cobalamin) along with folate (as methyl-THF).

Causes and Effects of B_{12} Deficiency

Vitamin B_{12} deficiency may arise due to decreased absorption or decreased dietary intake. The important causes are:

1. *Pernicious anaemia:* Cobalamin deficiency in developed countries is most commonly seen due to *deficiency of intrinsic factor* in the stomach and hence, decreased intestinal absorption. The condition, termed pernicious anaemia, is rare in India. It was named so because it was pernicious (fatal) when described first in 1849.

   Pernicious anaemia is an **autoimmune disease** with a strong familial background affecting people over 40 years of age. Antibodies against IF are generated leading to deficiency of IF and decreased absorption of cobalamin.

2. *Nutritional*: B_{12} deficiency is very common in India, especially among vegetarians of low socio-economic groups who cannot usually afford milk and milk products which are the only vegetarian sources of the vitamin. Interestingly, bacteria of the large bowel contain cobalamin, but humans are unable to absorb it because the bacteria are distal to the stomach (source of intrinsic factors) and the ileum (site of absorption). Interestingly, rabbits overcome this problem by habitually contaminating food with fecal matter.
3. *Gastrointestinal*: A situation similar to pernicious anemia will arise upon surgical removal of the ileum. Individuals with stomach cancer may develop pernicious anaemia because the diseased stomach fails to produce adequate amounts of intrinsic factors.
4. *Enterohepatic circulation*: Cobalamin is also secreted in the bile and there is a marked enterohepatic circulation of the vitamin. Disturbances in this circulation can, therefore, have major effects on vitamin B_{12} status.
5. *Others*: These include *iron deficiency anaemia* (associated with gastric atrophy), infection with *fish tapeworm* (the parasite binds B_{12}), or increased requirement in *pregnancy*.

Unlike other water-soluble vitamins, B_{12} can be stored in humans, and it is the only vitamin not synthesized by any plant or animal. Pernicious anaemia is caused by malabsorption of vitamin B_{12}, which occurs due to auto-antibodies against the intrinsic-factor.

Clinical Manifestations of B_{12} Deficiency

Major deficiency manifestations are of two types: *haematologic abnormalities* (megaloblastic anaemia, hypersegmentation of neutrophils), and demyelination in peripheral nerves and the spinal cord leading to *neurological deficiency*.

Haematological abnormalities: In initial stages, the predominant feature is macrocytic (*"large cell"*) anaemia: the

red cells are 25–50% larger than normal with fragile membranes and a tendency to haemolyze. *Macrocytosis is probably due to secondary deficiency of folate*. As discussed later, it is a consequence of the accumulation of methyltetrahydrofolate (*folate trap*), and hence cannot be distinguished from anaemia of folate deficiency.

Neurological disorders: The neurological damage may sometimes occur in the absence of haematological abnormalities and may be mistaken for diabetic neuropathy or neuropsychiatric disorders. This is known as subacute combined degeneration of the cord and is *unique to* B_{12} deficiency (*not associated with folate deficiency*). The lateral and posterior columns of the cord are affected, as are neurons of cerebral cortex, resulting in sensory as well as motor disturbances.

The anaemia of B_{12} deficiency is haematologically indistinguishable from that of folate deficiency, but a similar neurological disorder is not seen in patients with folate deficiency.

Laboratory diagnosis: Since cobalamin is required in only two reactions, its deficiency results in accumulation of methylmalonic acid (methylmalonic aciduria) and homocysteine (homocystinuria); both may be measured in laboratory. The plasma levels of cobalamin can be determined by bacteriological methods or radioimmunoassay.

G. Folic Acid

Folic acid is obtained from green leafy vegetables and its name reflects this—foliage, derived from the Latin word *folium*, meaning leaf. It serves as a **carrier of one-carbon (C_1) units** during several biosynthetic processes. Two other co-factors are also known to be involved in the addition of C_1 unit to a metabolic precursor, *biotin* in carboxylation reactions, and *S-adenosylmethionine* (*SAM*) as a methylating agent. However, folic acid is more **versatile** than either of these two because it can transfer the C_1 units in several oxidation states.

Structure

Folic acid consists of three components: a pteridine ring linked in sequence to para-aminobenzoic acid (PABA) and a glutamate residue (Table 18.2). Up to five additional glutamate residues are linked to the first glutamate via isopeptide bonds (between terminal carboxylate group and the amino group of the next glutamate residue) to form a polyglutamyl tail.

Synthesis

Pteridine and para-aminobenzoic acid are linked covalently by *pteroyl synthetase* to form pteroic acid. The latter is attached to a glutamate residue to form folic acid (Fig. 18.8). A polyglutamyl tail is built by addition of more glutamyl residues, which imparts multiple negative charges to the molecule and so it cannot traverse biological membranes by passive diffusion. Thus, ***polyglutamylation*** *serves to sequester folate in the cells in which it is required.*

Activation

Folate in the human organism must be doubly reduced to become an active coenzyme **tetrahydrofolate (THF)**. The reduction reaction is a stepwise one: folate to dihydrofolate and then to THF. A single NADPH-dependent enzyme *dihydrofolate reductase* (DHFR) catalyzes both the steps.

$$\text{Folate} \xrightarrow{DHFR} \text{Dihydrofolate} \xrightarrow{DHFR} \text{Tetrahydrofolate}$$

Folate antagonists: The *DHFR* reaction is inhibited by the antitumour agents (e.g. *methotrexate, amithopterin*, and *amenopterin*), which *competitively inhibit the DHFR*. This blocks the synthesis of tetrahydrofolate. Because THF is required for DNA biosynthesis and tumour cells have a very high level of DNA biosynthetic activity, even modest decrease in THF availability will inhibit tumour growth.

Mammals cannot synthesize folic acid, so it must be provided in the diet or by intestinal microorganisms. Many microorganisms can synthesize their own folate as long as PABA is present in the medium. Therefore, the **PABA analogues** (e.g. *sulphonamides*) can inhibit formation of folate (Fig. 18.8) in these organism, thereby inhibiting their

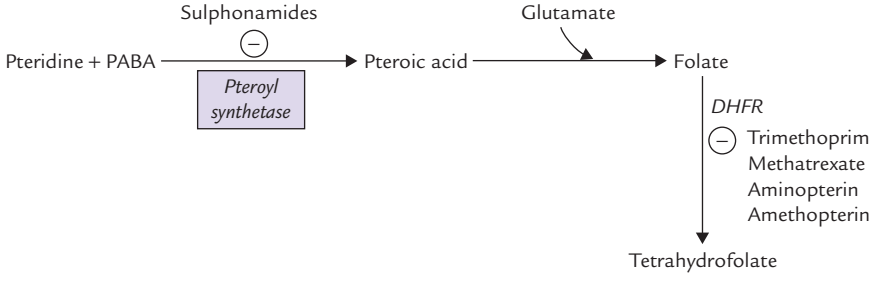


Fig. 18.8. Tetrahydrofolate synthesis and action of antifolates. Sulphonamides and trimethoprim are pharmacological inhibitors in bacteria. Methatrexate, aminopterin and amethopterin inhibit the *DHFR* reaction (PABA = para-amino benzoic acid).

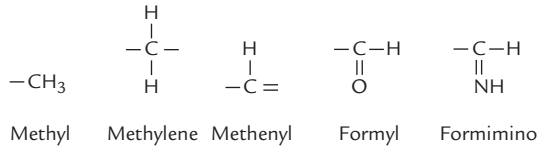


Fig. 18.9. The one-carbon groups transferred by THF.

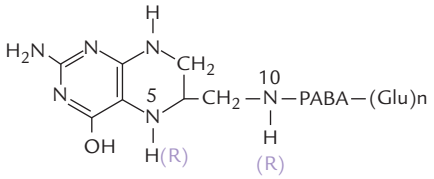


Fig. 18.10. The one-carbon units bound two N^5 and N^{10} of tetrahydrofolate.

growth. Sulpha drugs are effective in many infections, especially those of genitourinary tract. *Trimethopterin* inhibits *DHFR*, and hence acts synergistically with sulpha drugs.

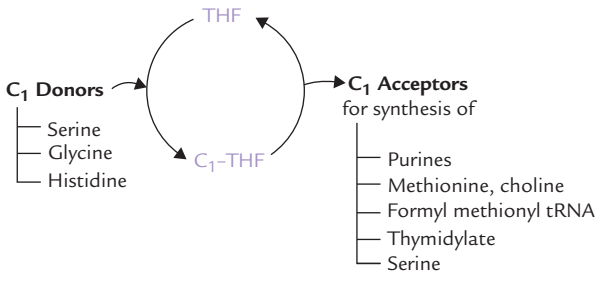
Absorption and Distribution

Folic acid is absorbed in the jejunum. In intestinal lumen, all but one of the glutamyl residues are removed by hydrolysis prior to absorption. This is achieved by two γ-*glutamyl hydrolases*. Following absorption, folic acid is transported in blood by two β-globulins. The major circulating form is methyltetrahydrofolate and the normal concentration range is 5–15 ng/mL. Once it arrives in the liver, the methyl derivatives are taken up by hepatocytes where various coenzyme forms are produced. Folic acid is not stored in tissues.

Coenzyme Functions

Tetrahydrofolate serves as **carrier of one-carbon units** at different oxidation levels (Fig. 18.9). These C_1 units are bound to one or both of the two nitrogens in the molecule, N^5 and N^{10} (Fig. 18.10).

THF receives the C_1 units from various donor molecules during catabolic reactions and can transfers them to specific acceptors for the synthesis of various compounds. Role of THF is thus vital in those reactions that require either addition or removal of C_1 units of various oxidation states.



1. ***Donors of* C_1 units:** THF acquires C_1 units from various donors during the following reactions:
   - (a) Serine to glycine conversion by *serine hydroxymethyltransferase*.

     $$\text{THF} + \text{Serine} \rightleftharpoons N^5\ N^{10}\ \text{Methylene THF} + \text{Glycine} + H_2O$$

   - (b) **Glycine breakdown** by the *glycine-cleavage enzyme*

     $$\text{THF} + \text{glycine} + NAD^+ \rightarrow N^5\ N^{10}\ \text{Methylene THF} + NADH + CO_2 + NH_4^+$$

   - (c) **Histidine breakdown**, where a C_1 unit from FIGLU is transferred to THF (Fig. 13.13).

     $$\text{FIGLU} + \text{THF} \rightarrow N^5\text{-Formimino THF} + \text{glutamate}$$

Once bound to THF, the C_1 unit can be oxidized or reduced enzymatically. *Thus, various oxidation states are interconvertible* (Fig. 18.11).

2. ***Acceptors of* C_1 units:** Some important one-carbon addition reactions in which C_1 unit is transferred to an acceptor are as here.
   - (a) *Synthesis of purine nucleotides*: The C-2 and C-8 of purines are contributed by formyl THF and methenyl THF, respectively (Chapter 20).
   - (b) *Conversion of homocysteine to methionine*: The methyl group required for the synthesis of methionine from homocysteine is provided by methyl THF (Fig. 18.7).
   - (c) *Synthesis of formylmethionyl-tRNA*: This is required for initiation of protein synthesis in prokaryotes and in mitochondria (Chapter 22).
   - (d) *Methylation of deoxyuridylate to thymidylate*: Methylene THF donates a carbon unit (C_1 unit) for the formation of deoxythymidylate (dTMP) (a nucleotide unit present in DNA) from deoxyuridylate (dUMP).

     $$\text{2′-Deoxyuridylate} + \text{Methylene THF} \xrightarrow{\text{Thymidylate synthase}} \text{2′-Deoxythymidylate} + \text{Dihydrofolate}$$

     This reaction is commonly referred to as *thymidylate synthase reaction*. It is unique as the tetrahydrofolate changes to dihydrofolate in this reaction. This is the only 1-C transfer reaction in which the redox state of methylene (the transferred 1-C unit) changes as the methyl group is added to dUMP.
   - (e) *Synthesis of serine*: The hydroxymethyl group required for glycine to serine conversion is provided by N^5, N^{10} Methylene THF.
   - (f) *Synthesis of choline*: Serine to choline conversion requires methyl group from N^5, N^{10} methylene THF.

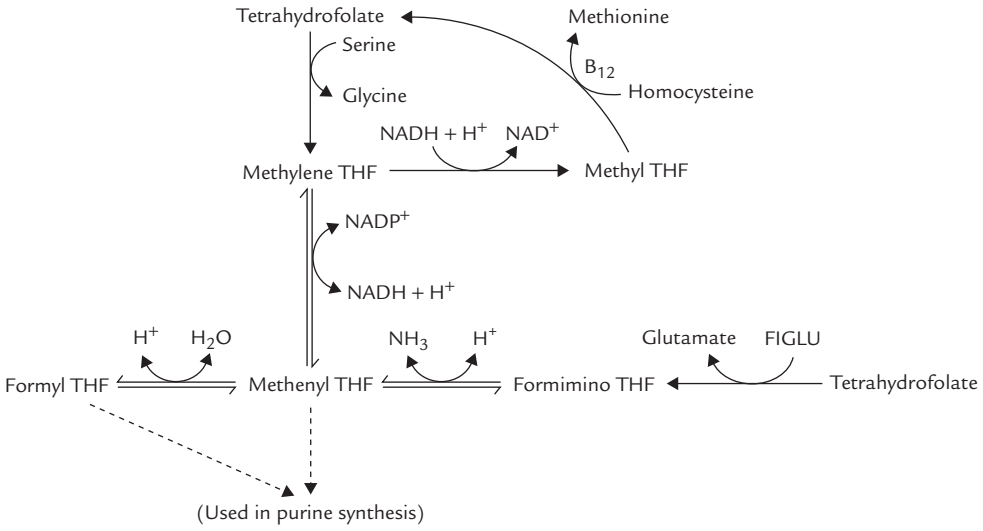


Fig. 18.11. Tetrahydrofolate as carrier of one-carbon units (FIGLU = formiminoglutamate).

Folic acid is involved in addition of one-carbon units to several metabolic precursors. It is more versatile than biotin (in carboxylation reactions) and S-adenoxylmethionine (methylating agent) as a carrier of one-carbon units.

Causes and Effects of Folate Deficiency

Deficiency of this vitamin is very common in India, particularly during pregnancy when the requirement is increased. Common causes are ***dietary*** deficiency and defective intestinal absorption. ***Chronic alcoholism*** is also a known cause.

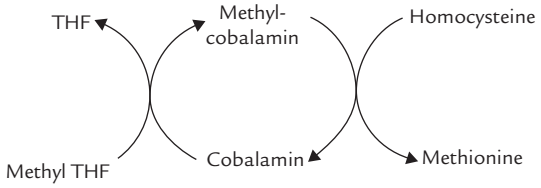
DNA synthesis indirectly requires folic acid because of its role in the synthesis of purines and in thymidylate synthesis. Hence, folic acid is needed in DNA replication and cell division. The cytoplasm of the folate-deficient cell grows at a normal rate, but DNA replication and cell division are delayed. This results in oversized cells that have an abnormally large amount of cytoplasm (i.e. megaloblasts). The cells most vulnerable to the deficiency state are the rapidly dividing cells of the bone marrow and intestinal mucosa, as they are most affected by slower than normal rates of cell division. The haematological changes may progress to a condition known as **megaloblastic anaemia** or **macrocytic anaemia,** where the red cells are large with diffuse nuclei (macrocytes; megaloblasts are oversized RBC precursors in the bone marrow). Formation of other blood cells, e.g. platelets and granulocytes are also compromised.

Megaloblastic anaemia seen in cobalamin deficiency is indistinguishable from the anaemia of folate deficiency. This is because in cobalamin deficiency, **folate is trapped in the form of 5-methyl-THF**. As a result, cobalamin deficiency leads to functional folate deficiency.

Folate Trap Hypothesis

Deficiency of cobalamin leads to functional folate deficiency by the following mechanism.

During the metabolism of one-carbon units, a small amount of methylene tetrahydrofolate is reduced irreversibly to methyltetrahydrofolate (Fig. 18.11). Since it cannot be used for the synthesis of purines or thymine, methyl-THF has to be converted back to one of the other coenzyme forms. The only reaction of methyl-THF is the methylation of homocysteine to methionine, which regenerates free THF. This reaction requires cobalamin, and therefore, methyl THF tends to accumulate in cobalamin deficiency.



Accumulation of methyl THF leads to depletion of the other coenzyme forms that are needed for nucleotide synthesis. Thus, folate trap hypothesis explains the anaemia of cobalamin deficiency but it cannot account for the neurological manifestations of pernicious anaemia.

H. Ascorbic Acid (Vitamin C)

King and Waugh in 1933, isolated from orange, an *anti-scurvy substance* having strong reducing nature. Its structure was established in 1938 by Howarth and it was named as ascorbic acid. Its reducing nature was due to its strong tendency to donate reducing equivalents.

Absorption and Storage

Vitamin C is absorbed from the small intestine by a *carrier-mediated process* at the luminal surface that requires a sodium gradient. The transport resembles the sodium-dependent transport of sugars and amino acids (Chapter 7). The efficiency of absorption is high (80–90%). Following absorption, the vitamin circulates in plasma, red cells and leukocytes. It is found in highest concentrations in the adrenals, the pituitary and the retina, in that order.

Structure and Synthesis

The structure of ascorbic acid resembles monosaccharides (hexoses) and it can exist as L- and D-isomers. Only the L form possesses the vitamin activity.

Vitamin C is synthesized by most of the plant and the animal kingdom in uronic acid pathway (Chapter 10). Only humans, higher primates, guinea pigs, and fruit-eating bats have lost the ability to synthesize the vitamin because of **lack** of the enzyme *L-gluconolactone oxidase*, that converts gluconolactone to ascorbic acid. Therefore, vitamin C is an essential nutrient in these animals.

Functions

Ascorbic acid functions as a **reducing agent** and a **scavenger of free radicals (antioxidant)**.

1. ***As a reducing agent*:** Ascorbic acid is promptly oxidized to its biological equivalent **dehydroascorbic acid**, which can be readily reduced to re-form ascorbic acid. *Mechanism of action of ascorbic acid relative to its many activities is explained by its ability to undergo such reversible oxidation and reduction reactions.* In a large proportion of reactions, the prime function of this vitamin is to maintain metal co-factors in their lower valence state, e.g. Fe^{2+} and Cu^{+}. Some of the important ascorbate-dependent reactions are as below:

   (a) During **collagen biosynthesis**, the *hydroxylases* causing post-translational hydroxylation of prolyl and lysyl residues require ascorbate.

$$\text{Proline} + \alpha\text{-Ketoglutarate} \xrightarrow[\text{Prolyl hydroxylase}]{\text{Ascorbic acid}} \text{Hydroxyproline} + \text{Succinate}$$

Thus, vitamin C plays a role in the formation of matrix of bones, cartilages, and connective tissue. In absence of vitamin C, newly synthesized collagen cannot form fibres properly, which accounts for the prominent connective tissue abnormalities of scurvy.

   (b) **Synthesis of norepinephrine** from dopamine by the enzyme *dopamine β-monooxygenase* depends on vitamin C (Fig. 13.23).

   (c) **Carnitine synthesis** requires two Fe^{2+}-containing *ascorbate-dependent dioxygenases*. Carnitine deficiency decreases mitochondrial fatty acid oxidation and thereby contributes to the fatigue, characteristic of scurvy.

   (d) During **bile acid synthesis** in liver mitochondria, the *7-α-hydroxylase* reaction requires ascorbic acid.

   (e) **Absorption of iron** is aided by vitamin C by converting ferric to ferrous ions (Fig. 19.2).

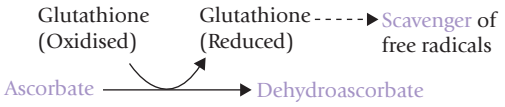
   (f) During **steroidogenesis**, ascorbic acid is thought to participate in several oxidation-reduction reactions. This may explain *highest tissue concentrations of ascorbate in the adrenal cortex*.

   (g) Ascorbate participates in **tyrosine catabolism** by serving as coenzyme for *4-hydroxyphenylpyruvate dioxygenase*.

   (h) Vitamin C also participates in bone mineral metabolism.

Vitamin C deficiency (scurvy) results in connective tissue problems because of impaired collagen synthesis.

2. ***As an antioxidant*:** Ascorbic acid is not only one of the strongest naturally occurring reducing agents known, but it can also serve as an antioxidant in several non-enzymatic reactions. It decreases oxidation of DNA and arrests protein damage, reduces lipid peroxidation and oxidation of low-density lipoproteins, and decreases production of extracellular oxidants from neutrophils. Because of these actions it provides several health benefits, especially in prevention of atherosclerosis and coronary heart disease.



An important area associated with its antioxidant properties is in the prevention and treatment of cancer. For example, it may suppress the formation of potentially carcinogenic nitrosamines from dietary nitrite and nitrate in the stomach, which may explain its protective effect in cancer. Epidemiological studies suggest that vitamin C exerts a synergistic effect with other dietary antioxidants, vitamin E and carotenoids, and this may have a significant role in the prevention of cancer, cardiovascular disease and cataract formation. The quantitative contributions of these components to the overall effect are not known.

Vitamin C (water-soluble antioxidant), acts together with membrane antioxidants (vitamins A and E) to limit the extent of free radical-mediated oxidative reactions. They prevent lipid peroxidation, and may have antimutagenic properties.

Clinical Deficiency

Scurvy, the vitamin C deficiency disease, first became prominent during the 15th century among sailors on long voyages whose diets were devoid of fresh foods. The introduction of limes to the diet of the British navy alleviated scurvy and led to the nickname *"limey"* for the British sailors. The disease is characterized by reduced cross-linking of collagen fibres, resulting in fragile blood vessels and **haemorrhagic diathesis**, which manifests in various forms. There is tendency to bleeding, especially in joints and under the skin. Gums become soft and spongy, teeth become loose and there is poor wound healing. Bones become weakened and anaemia and infections develop. If untreated, these infections may prove fatal.

Ascorbic acid saturation test is a highly sensitive test that can detect even ***subclinical cases of scurvy***. This test assumes that exogenously administered ascorbic acid is first taken up by tissues. The rest enters blood circulation from where it is eliminated in urine. In scurvy, the body stores of vitamin C are grossly depleted and therefore, any administered vitamin is promptly taken up by tissues; little or none is left for urinary excretion. Therefore, in a patient of scurvy, urinary excretion of ascorbic acid would be much lower compared to that in a normal subject.

Note: Though RDA for vitamin C is **45 mg/day**, a daily intake of 10 mg is sufficient to prevent scurvy.

IV. Fat-soluble Vitamins

The fat-soluble vitamins (A, D, E, and K or their precursor provitamins) are assembled from **isoprenoid units** by plants or bacteria. Despite limited chemical similarity, they share many properties:

1. **Absorption:** They are absorbed into the intestinal lymphatics along with other dietary lipids (or their digestion products).

2. **Transport:** The absorbed vitamins are delivered to liver (by chylomicrons), from where they are transported to other organs; the interorgan transport is affected by either plasma lipoproteins or specific transport proteins.

3. **Coenzyme role:** Though they participate in a variety of biological processes, their coenzyme functions are unknown (with possible exception of vitamin K).

4. **Storage:** Because they are predominantly non-polar, they cannot be excreted by the kidneys, and so tend to be stored in the body. Deficiency disease (except in the case of vitamin K) is difficult to produce in adults because of the storage in **liver** (A, D, K) or **adipose tissue** (E).

5. **Toxicity:** Excessive intake leads to accumulation of these vitamins to toxic levels in lipid laden structures of the body, which may prove even lethal (as in case of vitamins A and D).

Ample reserves of fat-soluble vitamins are stored in the tissues as they are not readily absorbed from food. Excessive accumulation in lipid laden structures may lead to toxicity.

Table 18.3 represents structures and function of fat-soluble vitamins. Daily requirements, dietary sources and deficiency diseases are given in Table 18.1.

A. Vitamin A

The term vitamin A refers to some polyisoprenoid compounds found only in animals. However, their precursors, carotenoids are found in plants.

Chemistry and Nomenclature

Vitamin A refers to three biologically active vitamers: **retinol** (an alcohol), **retinal** (an aldehyde) and **retinoic acid** (an acid), all of which are found only in animals. They are polyisoprenoid compounds comprising two distinct components: (a) a *cyclohexenyl ring*, and (b) a side-chain made up of several *isoprene units*, which is attached to the cyclohexenyl ring (Fig. 18.12).

The term **retinoids** has been used to define these three compounds (and other associated synthetic compounds with vitamin A-like activity). The *provitamin carotenoids* are also included in the vitamin A family. They are present in a variety of plants. The most important, quantitatively, is β-*carotene* (Table 18.3) (responsible for the orange colour of carrots).

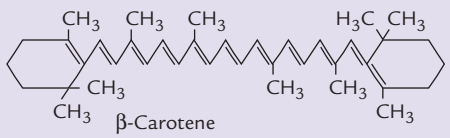
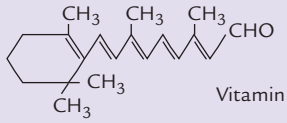
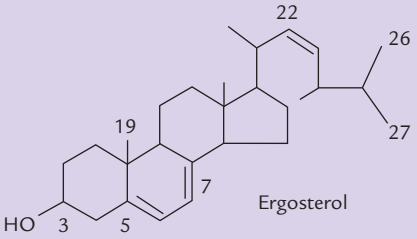
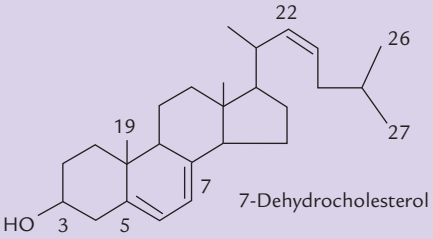
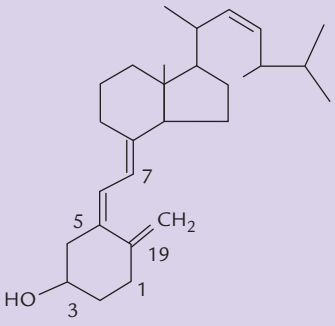
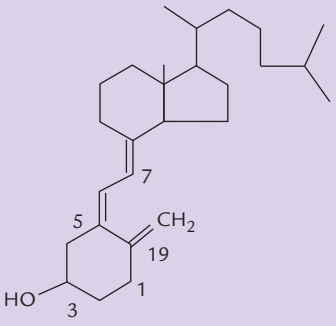
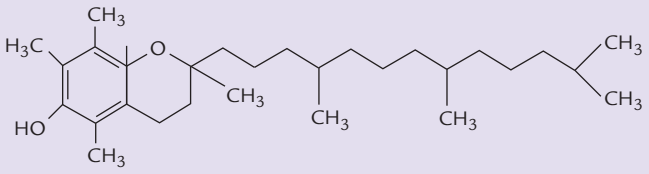
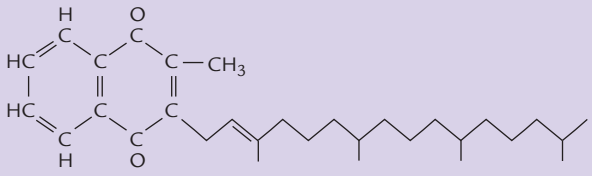
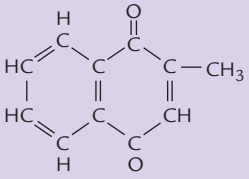
Vitamin A is found only in foods of animal origin as retinol esters, but the provitamins occur in plant foodstuffs, mainly as β-carotene. The latter consists, effectively, of two molecules of vitamin A joined end to end.

Absorption

β carotenes and retinol esters are handled differently.

(a) **β-carotene** is absorbed in small intestine and enters mucosal cells where it is cleaved into two molecules of *trans*-retinal (all-*trans* variety) by a *dioxygenase* and molecular oxygen; bile salts facilitate the reaction. Retinal is reduced to retinol (all-*trans* variety) by an NADH or *NADPH-dependent retinal reductase*. Retinol is esterified with a fatty acid, incorporated

Table 18.3. The fat-soluble vitamins: structures and functions

Name	Structure	Function
Vitamin A	 β-Carotene  Vitamin A_1 (all-*trans* retinol)	Visual cycle, bone formation, epithelial cell differentiation
Vitamin D	 Ergosterol  7-Dehydrocholesterol  Vitamin D_2 (ergocalciferol)  Vitamin D_3 (cholecalciferol)	Calcium absorption and metabolism
Vitamin E (tocopherols)	 5,7,8-Trimethyltocol (α-tocopherol)	Antioxidant in the lipid phase; scavenger of free radicals
Vitamin K	 Vitamin K_1 (phylloquinone)  Menadione (Vitamin K_3, methyl-naphthoquinone)	Blood coagulation, biosynthesis of calcium-binding proteins

* Vitamin D_2 differs from vitamin D_3 in having a double bond in the side chain and an extra methyl group (C_{28}).

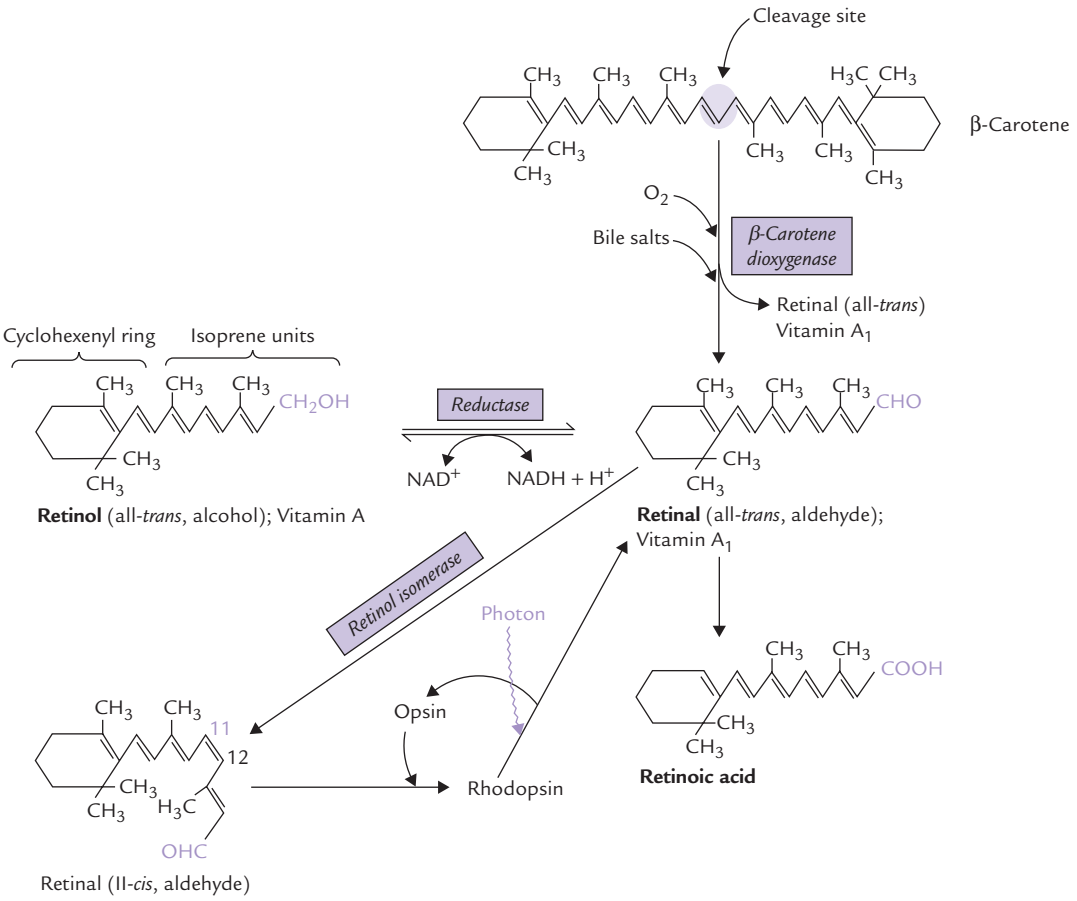


Fig. 18.12. Structures and interconversions of various members of the vitamin A family.

into chylomicrons together with dietary lipids, and secreted into lacteals (Fig. 18.13).

(b) The **retinol esters** of animal foods are handled differently. They are hydrolyzed in the intestinal lumen by a pancreatic enzyme (*carboxylic ester hydrolase*) and the free retinol generated is transferred across the intestinal mucosal cell. It is then esterified, incorporated into chylomicrons and secreted into lacteals.

The presence of lipids in the intestine ensures efficient absorption of retinol, up to 80% of the intake. Carotenoid absorption is less efficient, about 40% of the intake. Its activity is 1/6th that of retinol.

Transport

Chylomicrons carry the absorbed retinol into the circulation and then to liver parenchymal cells where the retinol esters are hydrolyzed. The retinol so released is reversibly bound to *retinol-binding protein* (RBP) in a one-to-one proportion and released into the circulation for transport to other organs. The target cells take up the retinol-RBP by a RBP-specific *receptor-mediated process*, and the retinol is bound to cellular *retinol-binding protein* (*CRBP*). Subsequently it may be oxidized to retinal or retinoic acid (Fig. 18.13).

RBP in human plasma is a monomeric polypeptide which has a single binding site. Being of low molecular weight (MW 20,000), RBP can be cleared by kidneys, so it circulates after being reversibly complexed with a plasma protein, **transthyretin** (T_4-binding prealbumin).

Retinoic acid is transported in association with albumin.

Storage

The hepatocytes not only dispatch retinol with RBP, they can also store the surplus in the form of retinol esters. More than 90% of the body's supply of vitamin A is usually stored in the liver cells, which contain a year's supply of the vitamin.

Functions

The role of vitamin A in vision has been known since the 1940s mainly through the studies of **G. Wald** (Nobel Prize, 1967). Its importance in various other cellular functions, unrelated to visual process is also being recognized:

- **Retinal** is involved in vision.
- **Retinoic acid** is involved in growth and cellular differentiation.
- **Retinol** is necessary for the reproductive system.

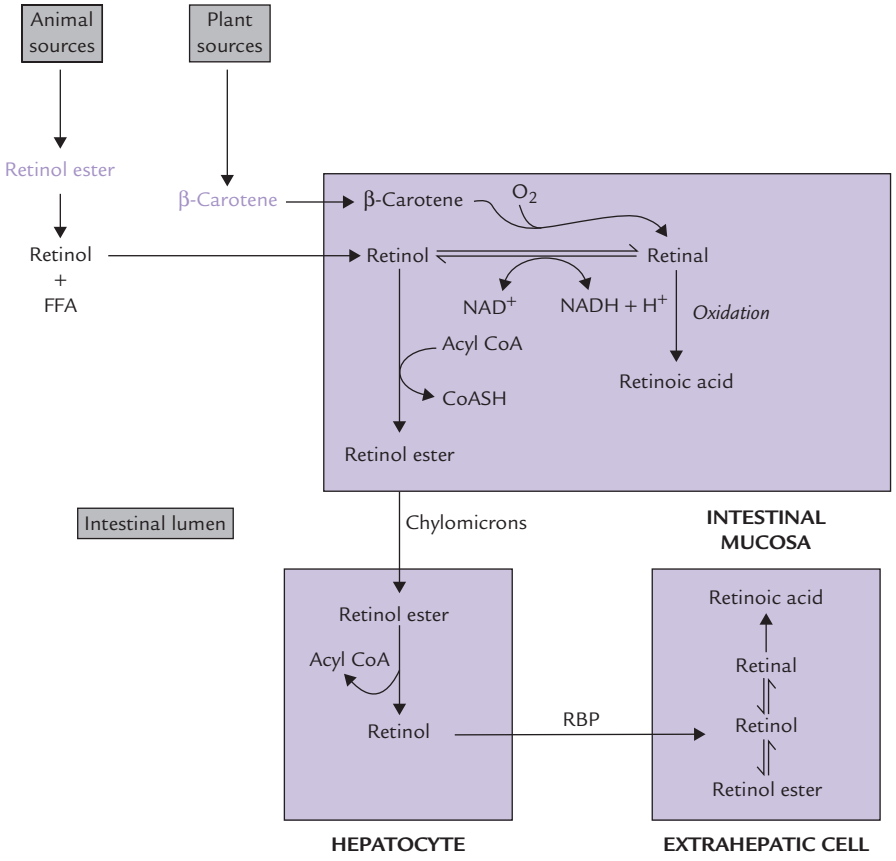
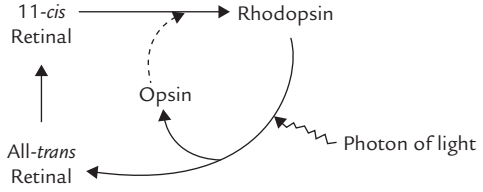


Fig. 18.13. Absorption, transport and metabolism of retinoids (RBP = retinol-binding protein, FFA = free fatty acid).

- **Carotenoids** have an antioxidant role per se (not through their conversion into vitamin A).

Vitamin A and Vision: Vitamin A is the prosthetic group of rhodopsin, the light sensing protein in retinal rod cells. Rhodopsin is located in a membrane system in the outer segment of the rod cell.

(a) ***Rhodopsin synthesis*:** Rhodopsin is made up of the protein opsin and 11-*cis*-retinal. The 11-*cis* retinal serves as the light absorbing part of rhodopsin, and opsin is an integral membrane protein with seven transmembrane helices. In the pigment epithelium of retina, all *trans*-retinal is isomerized to 11-*cis* retinal. The aldehyde group of the 11-*cis*-retinal then spontaneously links with a lysyl residue in the apoprotein (by a protonated, schiff-base bond) to form rhodopsin. This rhodopsin reaction places the 11-*cis*-retinal at the centre of the molecule, interacting with the membrane spanning α-helices. (In the cone cells, opsin is somewhat different and the rhodopsin equivalent is iodopsin.)



(b) ***Wald's visual cycle*:** Rhodopsin is a photosensitive pigment, which plays a *central role in the dim light vision*. Exposure to light of wavelength centred around 500 nm on the rod cells induces isomerization of 11-*cis*-retinal (Fig. 18.14). This photo-isomerization switches rhodopsin to a series of unstable intermediates in the following sequence:

Opsin + 11-*cis*-Retinal: **Rhodopsin** → Bathorhodopsin → Lumirhodopsin → Metarhodopsin-I → Metarhodopsin-II γ and finally, all-*trans*-retinal + Opsin.

The last one (**metarhodopsin-II**) is an activated "photoexcited conformation" referred to as **active rhodopsin (R*)**. The Schiff-base between the all-*trans*-retinal and the apoprotein of R* then hydrolyzes, and all-*trans*-retinal dissociates from the apoprotein. Before this bond is hydrolyzed, the photoexcited R* does something useful: it activates the G-protein (**transducin**)-based cascade. Activation of this cascade is implicated in generation of nerve impulse that is transmitted in optic nerve and is perceived by the brain as light.

A complex series of biochemical events are involved in generation of the nerve impulse, which are described later (*see Mechanism of vision*).

(c) ***Regeneration of 11-cis-retinal*:** Conversion of all-*trans*-retinal (released following dissociation of R*) back to 11-*cis*-retinal is the final event of the Wald's visual cycle.

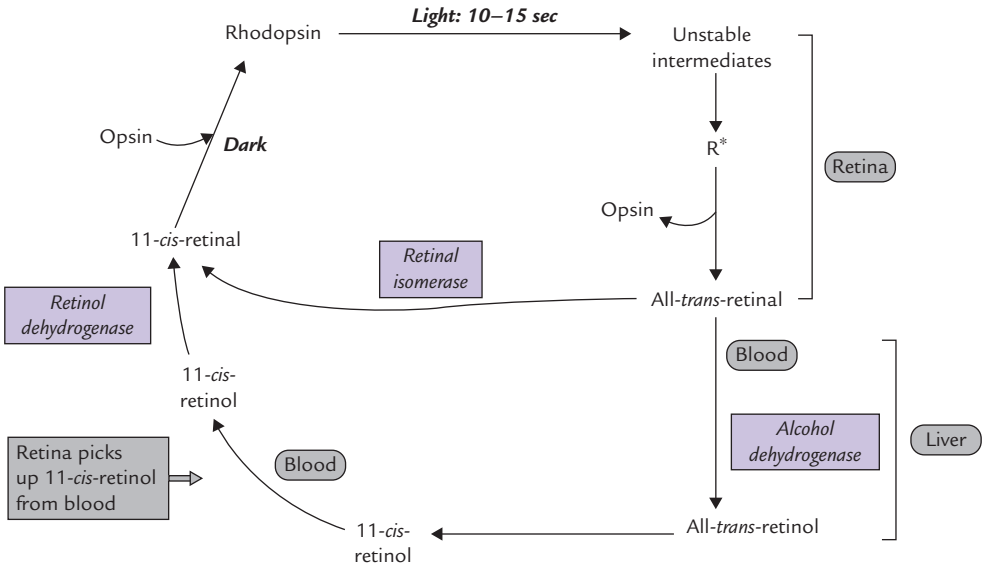


Fig. 18.14. Wald's visual cycle (rhodopsin cycle); R* activates transducin based cascade.

This reaction can occur in the retina itself, or it may require participation of an enzyme system located in hepatocytes (Fig. 18.14).

- *In retinal cells*: The *trans*- to *cis*-isomerization reaction is catalyzed in the dark by the enzyme *retinal isomerase*. The 11 *cis*-isomer then combines with opsin to regenerate rhodopsin, as discussed earlier.
- *In liver:* The all-*trans*-retinal is released into blood circulation and transported to liver. Following uptake by hepatocytes, it is reduced to all-*trans*-retinol by *alcohol dehydrogenase*, an NADH-dependent, zinc containing, enzyme. The all-*trans*-retinol is isomerized to *11-cis-retinol* and carried to retina by blood where it is oxidized to *11-cis-retinal*. This completes the Wald's visual cycle.

(d) ***Dark adaptation mechanism***: The time taken for regeneration of rhodopsin (following light induced depletion of rhodopsin) is known as dark adaptation time. It is a common experience that when a person shifts from bright light to dark, there is difficulty in seeing (e.g. in cinema hall), and after a few minutes the vision improves. During these few minutes rhodopsin is resynthesized, and the time taken is referred to as the dark adaptation time. It depends on vitamin A status of the person. It is prolonged in the vitamin A deficiency state.

Mechanism of Vision

Rhodopsin is a transmembrane protein located in rod cells, bound to membranous structures in the outer segment.

Rod cell physiology: The rod cell consists of outer and inner segments connected to each other by a narrow cilium (Fig. 18.15). The outer segment is packed with membranous disc structures and the inner segment is rich in

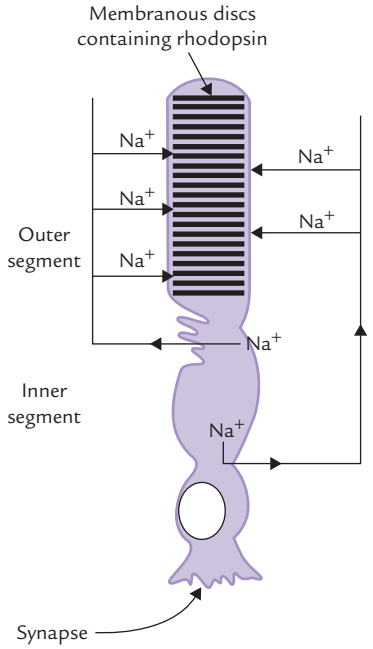


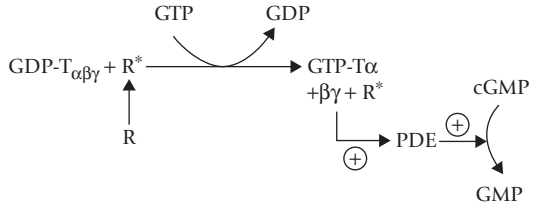
Fig. 18.15. A rod cell showing the dark current: the sodium ions enter the outer segment through cGMP-sensitive channels and are actively pumped out from the inner segment by an *ATPase*.

mitochondria. Next comes the cell body with its characteristic axon leading to the nerve ending and synapse. The discs in the outer segment are closed membranous structures, some 16 nm thick, and there are about 1000 packed into each segment. The *light-sensitive rhodopsin is bound to these discs*. The rod plasma membrane enclosing the outer segment contains **sodium channels** that admit Na^+ into the cell. Sodium is later pumped out of the cell by an *ATPase* located in the inner segment. Such movement of sodium ions is called the **dark current** because it takes place in dark (absence of light). *The sodium channels are kept*

open by the intracellular nucleotide, **cGMP**. When light strikes the rhodopsin, these sodium channels close, and the dark current ceases. This leads to polarization of the cell membrane and consequent generation of an electric current which results in a visual impulse. *What causes closure of the sodium channel, and how is it related to light exposure? This intricate problem has been solved and is outlined here.*

Active rhodopsin (formed on exposure of rhodopsin to light) plays a key role in the events that lead to closure of sodium channels. A G-type protein, ***transducin***, also participates in this complex series of events. In general, the G-type proteins function by binding GTP or GDP, and by triggering the second messenger phenomenon (Chapter 29).

- Transducin consists of three subunits: α, β and γ, and it binds with GDP.
- Active rhodopsin (R*) interacts with transducin and triggers the following changes in its composition: (a) GDP is replaced by GTP, and (b) the whole complex dissociates into the βγ complex and the α-GTP complex (GTP-Tα).
- GTP-Tα activates a *phosphodiesterase* (PDE), which causes hydrolysis of cyclic GMP.
- This leads to depletion cGMP. *Since cGMP is needed to keep the sodium channel open, its depletion leads to closure of these channels.*



- Closure of Na^+-channels prevents influx of Na^+ resulting in **hyperpolarization**.
- A single photon can cause hydrolysis of up to 100,000 cGMP molecules and thereby prevent the influx of more than a million sodium ions to cause hyperpolarization of the cell by approximately 1mV. Thus, **the transducin-based cascade has an extraordinary amplifying power**.

The light absorbed by the rods causes the plasma membrane to hyperpolarize by causing closure of the Na^+ channels. Photoexcited rhodopsin (R*) and a G-protein play key role in closing of the Na^+ channels.

The final question that is yet to be resolved is how the hyperpolarization of the rod plasma membrane results in depolarization of the ganglion cells, the axons of which project in the optic tract. The student is advised to refer to a textbook of physiology for the same.

Other Functions of Vitamin A

1. *Regulation of gene expression:* **Retinoic acid** is an important regulator of gene expression. It is transported to the nucleus bound to intracellular proteins: *cellular retinoic acid-binding protein* (CRABP)-I and -II. In nucleus, retinoic acid binds to and activates two families of nuclear receptors, which regulates gene expression by binding to response elements on the DNA.

Action of retinoic acid resembles that of steroid hormones, and the retinoic acid receptors (like the thyroid and steroid hormone receptors) belong to a large super-family of ligand-regulated transcription factors (Chapter 29).

2. *Growth and differentiation:* Because retinoic acid regulates gene expression, several processes associated with growth and differentiation depend on **retinoic acid**. These include maintenance of healthy epithelial cells, cell differentiation in spermatogenesis, and the differentiation of epithelial cells, among others. Severe vitamin A deficiency leads to production of abnormal epithelial tissues (*keratinization*).

3. *Glycoprotein synthesis:* Retinyl phosphates, obtained from **retinol** play a role in glycoprotein synthesis. Its hydrophilic portion serves as an anchor for growing oligosaccharide chains. This function appears analogous to that of dolichol phosphate (Chapter 5).

4. *Role in reproduction:* **Retinol** is required in reproduction. This function is mediated by control of expression of certain genes by retinol bound to cellular retinol-binding protein.

Retinol, and to a lesser extent the retinal, support spermatogenesis in males and prevents fetal resorption in females.

5. *Antioxidant role of vitamin A*: The antioxidant properties of the **carotenoids** at low oxygen partial pressures in tissues have been reported. It is in this role that carotenoids are thought to prevent the development of diseases in which the action of free radicals is implicated, such as cancer and cardiovascular diseases.

6. *Others:* Retinoic acid exerts a number of metabolic effects on tissues, such as **control of the biosynthesis of proteoglycans,** particularly sulphation of the latter.

Since retinol and retinal are interconvertible and can also be converted to retinoic acid, they can perform all functions of vitamin A. However, retinoic acid cannot perform other functions of vitamin A.

Requirements and Dietary Sources

Vitamin A requirement is difficult to calculate because of the different forms in which it is present. It is usually expressed in terms of international units (IU: one IU is the activity present in 0.3 μg of retinol, 0.344 μg of retinol acetate or 0.6 μg of β-carotene).

Only animal foods contain vitamin A and the provitamins are found in foods of plant origin (Table 18.1).

Clinical Deficiency

Vitamin A deficiency may be primary (dietary) or secondary. Causes of secondary deficiency include:

1. Fat malabsorption.
2. Failure to synthesize chylomicrons into which vitamin A is normally incorporated after absorption (mostly due to inability to synthesize apoB_{48}).
3. Failure to cleave β-carotene because of an enzyme defect.
4. Impaired storage in hepatic cells in liver disease.
5. Failure to synthesize retinol-binding protein (RBP), thus impeding transport from liver to target tissues.

The deficiency leads to the following clinical manifestations:

1. *Effect on vision:* Retinal is an essential component of the pigment rhodopsin (visual purple) on which the dim light vision depends. Therefore, lack of retinal may result in impairment of 'dark adaptation'. Inability to see in dim light (i.e. **night blindness**) results because of elevation in the visual threshold. Prolonged deficiency leads to an irreversible loss of visual cells. Pathological dryness of the conjunctiva and cornea also follow prolonged deficiency since retinoic acid is required for growth and maintenance of epithelium. This condition is known as **xerophthalmia** ("dry eyes"). If untreated, the condition progresses to corneal ulceration and consequent scarring. The ultimate result of these changes is blindness. This hazardous outcome is preventable with a timely supplementation of vitamin A. The ophthalmological manifestation of vitamin A deficiency is a major health problem in the developing countries. The worst sufferers are children.

Vitamin A (retinal) is necessary for vision mediated by the rod cells, so deficiency often presents as night blindness, dry eyes, and more serious consequences if untreated.

2. *Failure of bone remodelling* occurs leading to thick, solid bones in the skull with an increase in the cerebrospinal fluid pressure. This may lead to hydrocephalus.

3. *Gonadal dysfunctions* occur in deficiency of the vitamin A. Testicular degeneration in males and an increased incidence of miscarriage, or malformed offspring have also been reported.

4. *Follicular hyperkeratosis:* Transformation of columnar epithelium into heavily keratinized squamous epithelia, a process known as **squamous metaplasia**. Follicular hyperkeratosis (gooseflesh) is an important sign, developing early in the disease process.

5. *Nerve lesions,* often occurring with bone lesions, are frequently seen. Certain forms of skin diseases are also common in vitamin A deficiency.

Vitamin A Toxicity

The first recorded cases of vitamin A toxicity were among the arctic explorers because the polar bear liver, which they consumed, has high content of the vitamin. However, with normal foods it is virtually impossible to develop the vitamin A toxicity unless pharmacological amounts are prescribed or when the patient self-medicates. A large single dose of more than 300 mg, or more commonly, chronic ingestion of amounts grossly in excess of requirements, leads to toxic effects; the condition is called **hypervitaminosis**. It is characterized by dry and pruritic skin, hepatomegaly and raised intracranial pressure, which sometimes mimics the symptoms of a brain tumour. Excessive ingestion can cause congenital malformation in the growing fetus and is teratogenic. Pregnant women should be particularly cautioned against these risks.

B. Vitamin D

Vitamin D is a group of sterol compounds that are required as accessory food factors in the individuals not exposed to sunlight. They are also required in the diet of growing children and pregnant women, but normal adults receiving even mild sunshine do not need a dietary source because they can photochemically synthesize these compounds in skin.

Historical Perspective

The vitamin deficiency disease, *rickets*, has been known in the European countries, notably England, for many hundreds of years. During Industrial revolution in the early 1900s, it was observed that the disease could be treated successfully by either incorporating fish, particularly ***fish oil***, into the diet or by exposing the children to ***sunshine***. An essential factor was later discovered to be missing in rickets, which was named vitamin D (the fourth vitamin to be discovered after A, B and C). Its structure

was subsequently elucidated by *Otto Diels* and *Kurt Alder* (Nobel Prize 1950).

Vitamin D is not usually required in the human diet.

Synthesis

Vitamin D is synthesized by irradiation of 7-dehydrocholesterol, an intermediate of a minor pathway of cholesterol biosynthesis, present in all tissues. Largest amount of this steroid is present in malpighian layer of skin. It is photochemically cleaved by radiation in the ultraviolet range (290–315 nm), which *causes scission of the B-ring of the steroid nucleus of* ***7-dehydrocholecalciferol***. An unstable intermediate is produced which undergoes spontaneous, but slow, rearrangement to form **cholecalciferol** or vitamin D_3 (Fig. 18.16). Thus *7-dehydrocholesterol serves as a provitamin.*

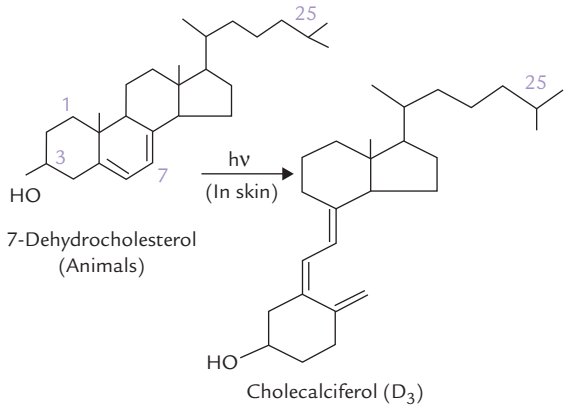


Fig. 18.16. Photochemical cleavage of B-ring of 7-dehydrocholesterol by ultraviolet rays to form cholecalciferol.

In addition to the naturally produced form of the vitamin (cholecalciferol; D_3), there is an artificially produced form D_2, or **ergocalciferol,** differing slightly in the structure of its side chain (Table 18.3). It is produced in the laboratory by irradiation of the plant sterol, **ergosterol**, and is the form readily available for pharmaceutical use. The biological actions of the two forms are comparable.

Requirements, Dietary Sources, Absorption and Transport

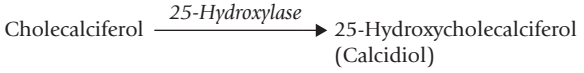
Daily requirement of the vitamin is **5–10 μg (400 IU)**. It is higher (12 μg) in children, and during pregnancy and lactation.

Exposure to sunlight is the major source of vitamin D, but it can also be obtained from food sources. Fish liver oil, egg yolk and fish are good sources, and milk contains moderate amount.

Dietary vitamin D is absorbed in the duodenum and jejunum from bile salt micelles and appears in the circulation as a constituent of chylomicrons. It is transported to the liver in chylomicron remnants.

Transport of the cholecalciferol formed in the skin to liver occurs in tight non-covalent binding to a vitamin D-binding globulin.

Activation: Cholecalciferol is activated in two steps in liver and kidneys. It is initially transported t**o liver** by cholecalciferol-binding globulin or a vitamin D-binding globulin. The hepatic enzyme ***25-hydroxylase*** brings about hydroxylation at the 25th position (C-25) to form **25-hydroxycholecalciferol** or **calcidiol** (Fig. 18.17).



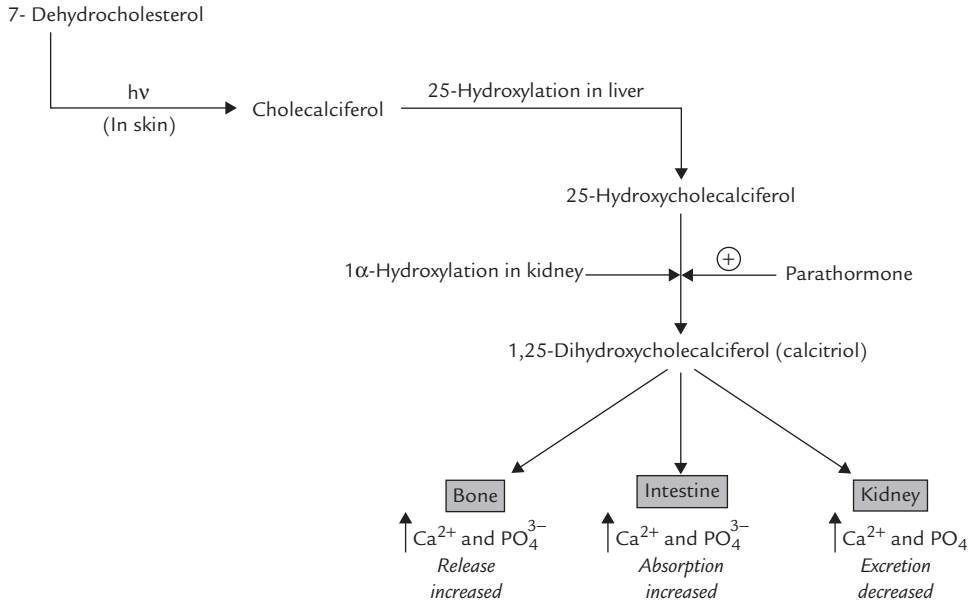


Fig. 18.17. Biosynthesis of biologically active vitamin D compound, calcitriol, from the pro-vitamin precursor.

Calcidiol thus formed is the major transport form of vitamin D. It is transported to kidneys by an α_2-globulin where it is further hydroxylated by *1-α-hydroxylase* to form 1,25-dihydroxycholecalciferol (DHCC or calcitriol). The enzyme is found in the inner mitochondrial membrane of the cells lining the proximal convoluted renal tubules. Note that DHCC is also termed **calcitriol** since it contains three hydroxyl groups at 1, 3 and 25 positions. It is the physiologically active form of the vitamin and is considered a hormone.

$$\text{Calcidiol} \xrightarrow{\textit{1-α-Hydroxylase}} \text{Calcitriol}$$

In kidneys, calcidiol can be hydroxylated at C-24 by *24-hydroxylase*. Like the *1-α-hydroxylase*, the *24-hydroxylase* is also a mitochondrial mixed-function *oxidase.*

Calcitriol is a Hormone

Calcitriol is thought of as a hormone rather than a vitamin because of its several hormone-like properties:

- It can be synthesized in the body, is released in the circulation and has distinct target organs.
- Further, its *mechanism of action resembles the Group II hormones* (Chapter 29). It binds with a nuclear receptor, termed **vitamin D receptor** (VDR), which is a ligand-regulated transcription factor belonging to the same superfamily as the receptors for steroid hormones and thyroid hormones. The binding of calcitriol to VDR is analogous to initiation of biochemical action by steroid-thyroid hormones (Chapter 29).
- The ligand-VDR complex then binds to **vitamin D response elements** in the promoter regions of target genes. The binding results in either an increase or a decrease in gene expression, depending on the gene in question. It is known, for example, that the expression of genes coding for calcium binding proteins, such as calbindin and osteocalcin, is stimulated, whereas the expression of the gene coding for parathyroid hormone is inhibited.

The activated form of vitamin D, calcitriol, is a hormone-like substance. Its receptor is a ligand-regulated transcription factor that belongs to the same superfamily as the receptors for steroid hormones and thyroid hormones.

Biochemical Effects

The predominant target organs for calcitriol are intestine, bone and kidney. It regulates serum calcium (normal = 9–11 mg/dL) and phosphate (2.5–4.5 mg/dL) concentrations by stimulating:

- Absorption of calcium and phosphate from intestine.
- Reabsorption of calcium and phosphate from the renal tubules.
- Mobilization of calcium and phosphate from bones.

Evidently, calcitriol plays a major role in **calcium homeostasis** (Fig. 18.17).

In addition, pancreas, pituitary and thymus are also target organs for calcitriol.

(a) ***Effect on intestine:*** The intestinal effect is considered the **most important effect** of vitamin D on calcium homeostasis, as it allows the entry of calcium into the body and thereby increases the total amount of available calcium. Actions on bone and kidney, on the other hand, involve rearrangement of already existing calcium pools. As noted, calcitriol-receptor complex interacts with DNA and causes transcription of the genes that code for ***calbindins***, the *calcium-binding proteins*, that **increase intestinal absorption of calcium**, thus elevating serum calcium levels.

Phosphate absorption is also stimulated by calcitriol.

(b) ***Effect on kidneys:*** In distal convoluted tubule, events similar to those described above take place to stimulate reabsorption of calcium and phosphate.

(c) ***Effect on bones:***

- *Bone resorption:* In bone, calcitriol facilitates the resorption of calcium and phosphate by stimulating osteoclasts. This is the best established action of vitamin D on bone, and in this instance, calcitriol acts cooperatively with parathormone to cause *bone erosion*.
- *Bone mineralization:* Vitamin D promotes bone mineralization as well, which seems opposite to the above stated effect. Increased mineralization may be

  1. Direct effect of the vitamin on bone, or
  2. Consequence of the increased levels of calcium and phosphate in the plasma (achieved because of vitamin D action on intestine).

It had been generally believed that the latter was the case but some more recent studies on calcium-binding proteins in bone, especially **calbindin 9K**, indicate that vitamin D may have a more direct effect on bone mineralization. Synthesis of this protein is increased by calcitriol and its binding with calcium affects bone mineralization. Other effects of the vitamin that promote mineralization are:

- Synthesis of vitamin K-dependent calcium-binding protein, **osteocalcin**, is increased by calcitriol.
- Cross linking of collagen and synthesis of various other calcium binding proteins (**osteopontin,** and the third component of **complement**) are also known to be increased.

Since cytoplasmic receptors play a vital role in action of vitamin D, any defect in these receptors results in rickets.

In addition to increasing the synthesis of the intestinal calcium-binding protein, calcitriol may also enhance conversion of this protein to its activated phosphorylated form. The intestinal absorption of phosphate is also stimulated by calcitriol, although the exact mechanism of action is not well defined.

Calcitriol acts primarily on three tissues—bone, kidney, and intestine—to increase serum calcium and phosphate levels.

Regulation of Calcitriol

Serum calcium and phosphate levels exert important influence on biosynthesis of calcitriol by controlling hydroxylation at position-1 (Fig. 18.18).

1. **Decreased serum calcium (hypocalcaemia)** induces parathyroid glands to secrete the parathormone (PTH) which is a potent activator of *1-α-hydroxylase*. This results in enhanced conversion of calcidiol to calcitriol (Fig. 18.17). Consequently the serum concentration of calcitriol is increased, which tends to replenish serum calcium through its effects on intestine, bones and renal tubules, described earlier in this chapter.

$$\downarrow Ca^{++} \longrightarrow PTH \longrightarrow \underset{\text{of calcidiol}}{\text{1-Hydroxylation}} \longrightarrow \text{Calcitriol} \longrightarrow \uparrow Ca^{++}$$

Thus, serum calcium participates in a **self-regulatory loop**: hypocalcaemia tends to elevate serum calcium levels, acting through parathormone and calcitriol.

2. **Decreased serum phosphate (hypophosphataemia)** directly stimulates the calcitriol formation by stimulating the *1-α-hydroxylase* (Fig. 18.18). The calcitriol increases intestinal absorption of phosphate, and enhances reabsorption of this ion from renal tubules. These effects bring back the serum phosphate level to normal.

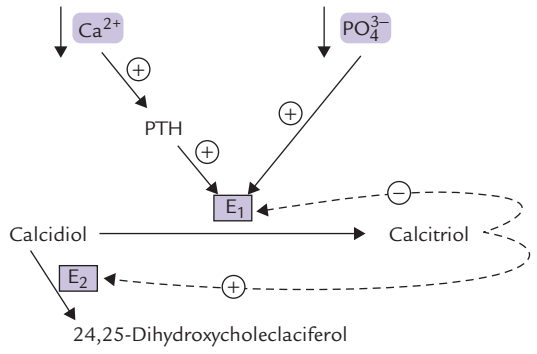


Fig. 18.18. Regulation of calcitriol production by serum calcium and phosphorus (E1 = *1-α-hydroxylase*, E2 = *24-hydroxylase*, PTH = parathormone).

In addition to the above-stated action on serum calcium and phosphate levels, the calcitriol has following effects on hydroxylation reactions:

- It **inhibits** the activity of the ***1-α-hydroxylase***, thus inhibiting its own production. As a result, excessive production of calcitriol is prevented.
- It **activates** ***24-hydroxylase*** so that level of 24,25-dihydroxycholecalciferol rises. The latter promotes bone mineralization.

Other Vitamin D Metabolites

1. ***24,25-dihydroxycholecalciferol*** is produced by the enzyme *24-hydroxylase* from 25-hydroxycalciferol. It **decreases** the serum concentrations of both **calcium** and **phosphate**. It has pronounced *mineralization promoting effect on bones*. Its production is stimulated by decreased PTH levels.

2. A ***26-hydroxylated metabolite*** is produced from calcitriol by hydroxylation of calcitriol at C-26 in course of its oxidative degradation. Hepatic microsomal mixed-*oxidases* are involved in these reactions.

3. *Other* polar metabolites (e.g. **23-hydroxylated derivatives**) are also produced in course of the oxidative reactions. These metabolites possess little of vitamin D activity. The *oxidases* that produce these metabolites are induced by numerous pharmacological agents, including phenytoin and phenobarbital. This may account for development of rickets in children receiving anticonvulsant therapy since the latter enhances formation of these metabolites **(Case 18.3)**.

Causes and Manifestations of Vitamin D Deficiency

- ***Inadequate exposure to sunlight*** among people living in congested slums may result in vitamin D deficiency (*i.e. rickets*) in children. The adult counterpart of rickets is called *osteomalacia*. There is now evidence that with advancing age, capacity of the skin to synthesize cholecalciferol diminishes, resulting in the age-related disorders of calcium metabolism.
- ***Other causes of vitamin D deficiency are:***
  (a) Inadequate dietary intake
  (b) Decreased intestinal absorption, and
  (c) Impaired activation: It occurs in hepatic and renal disorders since liver and kidneys are involved in the activation. The conditions are termed *hepatic rickets* and *renal rickets*, respectively.

The immediate effect of the deficiency is a decrease of the plasma calcium concentration, which is promptly restored by parathormone (PTH) at the expense of bones (PTH mobilizes calcium from bones). Thus, bones are tapped as a calcium reservoir and are gradually depleted

of their mineral content, resulting in **soft, poorly mineralized** and **pliable bones** that bend rather than break under stress (especially in weight bearing areas). The classical features of **rickets**, therefore, are bone deformities such as *bow legs, knock-knee, Harrison's sulcus, rickety rosary, bossing of frontal bones and pigeon chest*. In **osteomalacia**, the patients suffer from bone pains and are likely to suffer fractures. Abnormalities of biochemical parameters in both conditions are *low serum calcium* and *phosphate,* and *elevated serum alkaline phosphatase* (*bone isoenzyme*) *activity*. Serum calcitriol level is low.

Note: Decreased responsiveness of the target tissues to the vitamin D may also result in signs and symptoms of the vitamin D deficiency. Decreased responsiveness results because of defect in the tissue receptor so that its affinity with calcitriol is decreased.

Study of **Cases 18.5** and **31.1** will show that apart from classical vitamin D deficiency rickets, several other disorders affecting kidneys, liver and end organs may cause rickets.

Hypervitaminosis D: Prolonged intake of high doses of vitamin D (above 155 units/day) may results in toxic symptoms such as polyuria, intense thirst, confusion, difficulty in speaking and weight loss. Biochemical alterations are *hypercalcaemia, hypokalaemia* and *metabolic alkalosis*.

C. Vitamin E (α-Tocopherol)

Vitamin E is an important antioxidant. It was isolated from wheat germ oil and its structure was determined in 1936 by Paul Karrer (Nobel prize 1937). Vitamin E was earlier referred to as **anti-sterility vitamin**; in fact the word tocopherol is derived from Greek words, *tokos,* meaning childbirth and *pherin*, to bear, because it was believed that tocopherol is required for fertility. However, such anti-sterility effect has been observed only in some animals, not in human beings.

Chemical Nature

At least eight naturally occurring plant compounds with vitamin E activity are presently known, the most important being **α-, β-**, **γ-,** and **δ-tocopherol**. All are viscous, light yellow oils that are heat stable but readily degraded by oxygen or ultraviolet light. They have a *substituted chromane nucleus* with a *polyisoprenoid side chain* of variable length: usually three carbon atoms (Table 18.3). Of these vitamers, the most abundant and potent is α-tocopherol.

Absorption, Transport and Metabolism

The richest sources of the naturally occurring vitamin E are vegetable oils and nuts (Table 18.1). It is absorbed as free tocopherol along with other lipid components. The absorption requires bile salts: in their presence the vitamin is absorbed with efficiency of 20–40%. It has no specific binding proteins and is found in circulation in association with plasma lipoproteins and erythrocytes (plasma concentration is 0.5–1 mg/dL). Tissue distribution also shows little selectivity: vitamin E is **enriched in all lipid-laden structures**, with highest concentrations in fat depots of adipose tissue. It is mainly excreted in faeces via hepato-biliary route, after the chromane ring is oxidized followed by its conjugation with glucuronic acid.

Functions

1. ***Antioxidant role:*** Vitamin E is accepted as nature's most potent and most abundant biological antioxidant, in particular a **membrane antioxidant**. In cellular and subcellular membranes, it acts as a first line of defense against free radicals by acting as chain-breaking antioxidants (Fig. 18.19) as described in Chapter 27. As such, it is associated with the membrane lipid structure and so can promptly protect the membranes from attack by endogenous and exogenous free radicals. Vitamin E may be located near enzyme complexes that produce free radicals, such as *NADPH-dependent oxidase* systems.

Vitamin E is enriched in fat depots and lipo-proteins also, where it scavenges free radicals that are formed during lipid peroxidation, thereby interrupting free radical chain reaction. Further details are presented in Chapter 27.

2. ***Antiatherogenic role:*** Several epidemiological studies suggest an inverse relationship between vitamin E intake and the incidence of morbidity and mortality from coronary artery disease. Supplements of 400 IU of vitamin E for about 2 years result in about 40% reduction in the incidence of heart attack (1 mg of α-tocopherol is equivalent to one IU of vitamin E). The following properties of the vitamin may contribute toward its antiatherogenic role:

(a) It retards oxidation of LDL, thereby decreasing production of the pro-atherogenic oxidized LDL.
(b) It impedes various cellular signalling pathways (e.g. *protein kinase-C* initiated pathways), which results in inhibition of proliferation of smooth muscle cells,

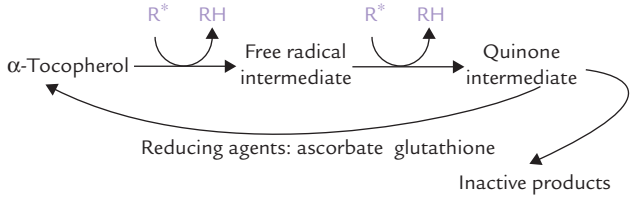


Fig. 18.19. The action of α-tocopherol as a scavenger of free radicals (R* = free radical).

platelet adhesion and aggregation, and function of adhesion molecules.

(c) Vitamin E may also decrease the synthesis of leukotrienes and increase synthesis of prostacyclins by upregulating *phospholipase* A_2 and *cyclooxygenase*.

3. ***Interrelationship with selenium metabolism:*** Selenium alleviates some symptoms of vitamin E deficiency, probably through its role as a cofactor for *glutathione peroxidase*, an important enzyme that oxidizes and destroys the free radicals (Chapter 27). Vitamin E also decreases requirement of selenium, and vice versa, thus they appear to act synergistically.

4. ***Others:*** Vitamin E boosts immune response, protects RBC from haemolysis, keeps structural and functional integrity of all cells, slows ageing process and appears to offer protection against Alzheimer's disease. These effects appear to be by virtue of its antioxidant role. Glutathione and ascorbate prolong action of vitamin E by (a) reducing the initially formed free radicals, and (b) reducing the quinone intermediates back to active form. Claims that high doses of the vitamin E are beneficial for treatment of skin disorders, fibrocystic breast disease, sexual dysfunction, cancer and baldness are yet to be substantiated.

High doses of vitamin E and the other antioxidant vitamins (A and C) provide protection from atherosclerosis (presumably by inhibiting the oxidation of LDL) and possibly from cancer (by preventing somatic mutations).

Clinical Deficiency

Vitamin E deficiency is not uncommon in the premature infants but is rarely seen in full term infants (in spite of poor placental transport) because the breast milk is a good source (Table 18.1). It occurs rarely in adults because of its widespread distribution in foods and because the body vitamin stores can meet the requirement for several months. A **daily intake** of **10–30 mg** is considered adequate and is provided by most diets. Defective lipid absorption or transport may sometimes cause the deficiency.

Vitamin E is often called a "vitamin in search of disease"; human deficiency does occur sometimes. It is seen in infants who are born with low tissue stores and who have poor intestinal absorption for several weeks after birth.

A genetic defect in the formation of hepatic α-tocopherol transfer protein (which facilitates incorporation of α-tocopherol in nascent VLDL) leading to the vitamin deficiency has been described. The deficiency may also be acquired—associated with defective lipid absorption or transport.

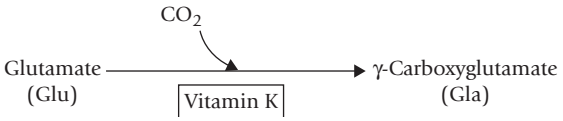
Haemolysis due to lack of protection for RBCs against peroxides, and creatinuria due to increased muscle breakdown are the features of vitamin E deficiency.

Hypervitaminosis E: High doses of vitamin E (more than 400 IU/day) depress coagulability. Patients with bleeding disorders and those receiving warfarin, should be cautioned against its use. Otherwise, vitamin E is least toxic of all fat soluble vitamins—even 50 times the recommended intake has been reported non-toxic.

D. Vitamin K

Vitamin K was initially recognized as an antihaemorrhagic dietary factor. ***It is the only fat-soluble vitamin that acts as a coenzyme.*** Its major function is to synthesize γ-carboxyglutamate by incorporating carbon dioxide into specific glutamyl residues of certain proteins such as prothrombin, osteocalcin and certain clotting factors.

The designation "**K**" derived from the initial description of the vitamin as the "*koagulation* vitamin".



Chemistry

Vitamin K activity is present in a group of structurally related compounds. All have a 2-methyl-1,4-naphthoquinone nucleus, but vary in the number of isoprenoid units in its side chain (Fig. 18.20). Two naturally occurring forms are **menaquinone** (K_2) and **phylloquinone** (K_1), present in animals and plants respectively. Menaquinone is synthesized by intestinal microorganisms and stored in liver. Some water soluble synthetic forms with vitamin K activity **e.g., menadione** have also been prepared for use in clinical practice. Sodium menadiol diphosphate and menadione sodium bisulphate are some other examples.

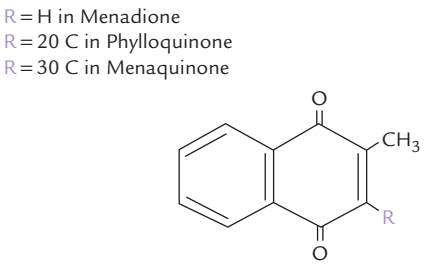


Fig. 18.20. Structure of vitamin K.

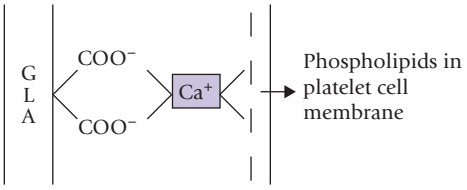
Absorption and Storage

The intestinal absorption of vitamin K is dependent on appropriate fat absorption and it requires bile salts. It may be obtained from diet or intestinal bacterial synthesis. But colonic bacteria do not make a significant contribution because vitamin K is absorbed only in the small intestine. Chylomicrons carry the absorbed vitamin to liver where it is stored. From liver, it is released into blood circulation where it is transported in association with β-lipoproteins. Unlike the other fat-soluble vitamins, the body stores of vitamin K are insignificant (50–100 μg), so *vitamin K is the first fat-soluble vitamin to be deficient in acute fat malabsorption.*

Functions

Vitamin K plays an important role in **blood coagulation** for it is *required for the post-translational processing of several proteins required in the coagulation cascade* (e.g. factor-II, -VII, -IX and -X) in the ER of liver cells. All these protein clotting-factors are initially synthesized as inactive precursors in the liver. Formation of mature clotting factors requires that the glutamyl residues of the precursor proteins be converted to **γ-carboxyglutamate** (Gla) residues by addition of carboxylate group. This reaction is dependent on vitamin K, which serves as a coenzyme (Fig. 18.21).

Role of Gla in clotting: The Gla residues so formed serve as binding site for calcium ions; each Gla contains two negative charges which chelate the positive calcium ion. In prothrombin molecule, for instance, there are 10 such carboxylated Gla residues and all of these are required for this protein's specific chelation of calcium ions. The calcium then binds with the negatively charged phospholipids present on the platelet cell membrane. In this way, *bridging of the phospholipids to the Gla residue of prothrombin occurs via calcium ion.*



Thus, prothrombin becomes anchored to the surface of the activated platelets. This provides a high local concentration of the prothrombin, which accelerates its activation by about 50- to 100-fold, an essential requirement for blood clotting.

Functional activation of a number of other proteins also requires the vitamin K-dependent carboxylation reaction. Such proteins include **osteocalcin**, **C-reactive protein** and structural proteins of kidneys, lungs and spleen. Osteocalcin is a small-sized protein (40–50 amino acids), synthesized by osteoblasts, which binds to hydroxyapatite crystals of bone. Vitamin K creates calcium binding sites on osteocalcin, which helps it to retain calcium.

Clinical Deficiency

Vitamin K deficiency is rare. It may be induced in following ways: (a) due to treatment with antibiotics that eliminates normal intestinal flora, (b) in fat malabsorption syndromes, (c) in liver diseases, and (d) by vitamin K antagonists, such as **dicoumarin** or **warfarin**. They act by competitively inhibiting the gamma carboxylation system due to structural similarity with vitamin K.

Deficiency of vitamin K increases the time for blood coagulation, and so *bleeding tendency* is the prominent deficiency manifestation. Even a minor cut may cause prolonged bleeding. The only important deficiency sign is *increase in prothrombin time (PT)*, and it is the most important laboratory test for the evaluation of vitamin K status. Note that PT may rise in liver damage also due to inability of liver to synthesize prothrombin. To distinguish between these two conditions, vitamin K is administered parenterally. PT remains unaffected in vitamin K deficiency, but returns to normal in liver diseases.

Newborn infants with their inadequate vitamin K stores may suffer from **haemorrhagic disease of the newborn**. It is the *most common nutritional deficiency in newborns*.

Toxicity

Vitamin K is non-toxic even in large amounts. However, some special cases merit attention. Large doses (5 mg) of

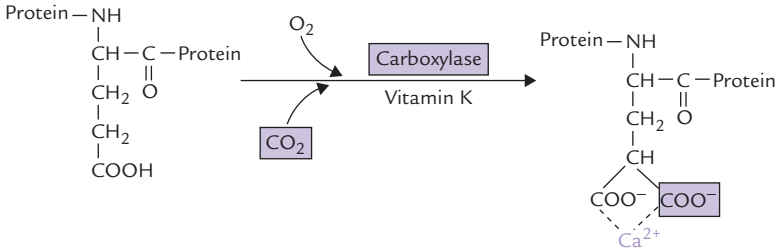


Fig. 18.21. Vitamin K as cofactor in synthesis of Gla residue.

menadione and its water-soluble derivatives are known to cause haemolytic anaemia and kernicterus in infants; the premature infants are more at risk. Haemolytic reactions may also occur in adults having *glucose 6-phosphate dehydrogenase* deficiency.

Exercises

Essay type questions

1. Explain the biochemical basis of anaemia due to deficiency to: vitamin B_6, vitamin C and vitamin B_{12}.
2. Mention biologically active forms of folic acid. How is it involved in one carbon metabolism. Explain action and clinical significance of folate antagonists.
3. Why is vitamin D considered as a hormone? Explain the role of calcitriol in regulation of serum calcium level.
4. What is the active, form of vitamin D and how is it formed? What is renal rickets and hepatic rickets? Explain biochemical basis of the clinical manifestations seen in vitamin D deficiency.
5. How is *cis*-retinal synthesized in the body? Explain its role in vision.
6. What are the biochemical roles of vitamin C in human body? What the is biochemical basis of anticancer role of this vitamin?

Write short notes on

1. Renal rickets
2. Folate trap
3. Functions of vitamin K
4. Antioxidant vitamins
5. Intestinal absorption of vitamin B_{12}
6. Coenzyme role of vitamin B_{12} and folic acid
7. Beriberi
8. Carboxylation reactions

CLINICAL CASES

CASE 18.1 A malnourished man with warm extremities and swollen feet

A 35-year-old man reported in the hospital OPD with complaints of lack of appetite, listlessness, and weakness of legs. Often he had a feeling of heaviness and numbness of legs with "pin and needle" sensation. He appeared malnourished and came from poor socio-economic background.

Examination showed warm extremities and loss of sensation in the skin over tibia. There was moderate oedema, mainly involving legs and face. Tachycardia and marked distension of the neck was noticed. X-Ray chest showed cardiac enlargement. Laboratory investigations showed raised plasma pyruvic acid level. *Transketolase* activity in the erythrocytes was low, but it showed a rise by about 40% after administration of thiamine pyrophosphate (TPP).

Q.1. What is the most probable diagnosis?
Q.2. Mention the cause of the neurotic symptoms in this patient.
Q.3. Explain the biochemical changes in this disorder that lead to (a) oedema, (b) dilatation of neck veins, and (c) cardiomyopathy.
Q.4. Discuss the treatment for this patient.

CASE 18.2 A 6-month-old child with feeding problems and recurrent vomiting

A 6-month-old male infant was referred to the teaching hospital for investigation of feeding difficulties, recurrent vomiting, respiratory distress and seizures. On examination, his body weight and head circumference were very small for his age. Analysis of blood revealed that serum levels of both the D- and L-forms of methylmalonic acid were elevated. High urinary excretion of these two compounds and of other compounds related to methionine metabolism (e.g. cystathionine and homocysteine) were detected. Vitamin B_{12} deficiency was suspected since this vitamin is required by certain key enzymes in metabolism of the above-mentioned compounds. However, serum level of this vitamin was well within the normal range of 150–600 pg/mL. Analysis of mother's milk was carried out since decreased B_{12} content of milk is a major cause of deficiency of this vitamin in infants. It was 1650 pg/mL, which is well within the normal range of 1000–3000 pg/mL. In view of these findings, deficiency of B_{12} vitamin as the likely cause was ruled out.

To further investigate the underlying defect, the following tests were conducted. Cell-free extract was obtained from the cultured fibroblasts. Addition of radioactive methylmalonyl

CoA to it showed that conversion of this compound to succinyl CoA was hampered (experiment 1). However, addition of cobalt of +1 state (which means valency of cobalt is +1) to the reaction mixture made this conversion possible (experiment 2).

Expt.1 - Methylmalonyl CoA ⟶ ↓ Succinyl CoA

Expt.2 - Methylmalonyl CoA $\xrightarrow{Co}$ ↑↑ Succinyl CoA

Q.1. Identify the biochemical defect in this patient.
Q.2. Comment on the urinary report of this patient.

CASE 18.3 Listlessness and persistent bodyache in an 8-year-old boy

An 8-year-old boy was brought to the orthopedic OPD with back injury, following a fall. X-Ray spine revealed a fracture in the first lumbar vertebrae. The other significant radiological finding was diffuse haziness at the metaphyseal border and inadequate mineralization of primary spongiosa. He had been feeling mild fatigue, bodyache, upper abdominal discomfort and fatty food intolerance. During the past three months, he had developed slight abnormality of gait and found walking a tiring experience. Examination showed diffuse bony tenderness, especially over tibia, mild liver enlargement, and proximal muscle weakness.

Investigations test	**Patient' report**	**Reference range**
Serum bilirubin	2.8 mg/dL	0.1–1.0 mg/dL
Alanine transaminase	88 U/L	10–35 U/L
Aspartate transaminase	102 U/L	10–40 U/L
Alkaline phosphatase	128 U/L	40–100 U/L
Serum calcium	8.2 mg/dL	8.5–10.5 mg/dL
Serum phosphate	2.3 mg/dL	2.5–4.5 mg/dL

Serum parathormone level was moderately raised. Circulating level of 25-hydroxycholecalciferol (i.e. calcidiol) was 3 ng/mL (normal 8–55 ng/mL). Analysis of a bone biopsy sample revealed a marked fall in the mineral content.

Q.1. What is the probable diagnosis?
Q.2. Identify the underlying cause of vitamin D deficiency in this child.
Q.3. What would you expect the circulating calcitriol levels to be?
Q.4. Enumerate various factors that lead to development of rickets in hepatobiliary disorders.
Q.5. What accounts for the development of rickets in children receiving anticonvulsant therapy?
Q.6. What is renal rickets?

CHAPTER

19 MINERALS

A large number of mineral elements are present in nature, of which only a few are essential for the human body. They can be divided into five groups:

1. The *first group* includes carbon, hydrogen and nitrogen. The body obtains these elements from dietary fats, carbohydrates, proteins, and also from water.
2. The *second group* includes calcium, phosphorus, magnesium, sodium, potassium, chloride and sulphur. These elements are nutritionally important and are required in relatively larger amount in the diet (100 mg/day). They are called the **major elements** or **macro elements.**
3. The elements of the *third group*, known as the **trace elements**, are required in diet in much smaller amounts (< 100 mg/day). Chromium, cobalt, copper, iodine, iron, manganese, molybdenum, selenium and zinc are examples of such elements. Fluorine is usually considered to be a trace element, although its role in humans is not clearly understood. Its deficiency is associated with tooth decay.
4. The *fourth group* contains arsenic, cadmium, nickel, silicon, tin and vanadium. These elements have well-defined role in animals but have *no essential (or known) function in humans.*
5. The *final group* consists of certain **toxic elements**, such as lead and mercury.

In this chapter, the metabolism and functional significance of some biologically important minerals are described. After going through this chapter, the student should be able to understand:

- Absorption, transport and excretion of minerals.
- Mineral deficiency or excess: causes and consequences.
- Sources and daily requirements of minerals.
- Major elements: physiological role of sodium, potassium, magnesium.
- Trace elements: physiological role and deficiency disorders of iron, copper, magnesium, zinc, selenium, and molybdenum.

I. Absorption, Transport and Excretion of Minerals

Intestinal absorption of most dietary minerals is difficult, except for sodium and potassium. This is because of the tendency of most minerals to form complexes with fibres or phytates in intestine, which are relatively insoluble. This hinders their absorption, and therefore, most of the ingested minerals are excreted in faeces. Mediation of specific carrier proteins is required for their absorption; for example, zinc absorption requires a mucosal binding factor and iron absorption requires a specific intracellular carrier molecule. Synthesis of these (carrier) proteins, therefore, serves as an important mechanism for control of mineral levels in the body.

Few minerals circulate in free, unbound form, whereas transport of others requires specific binding proteins-β_1 e.g. transferrin for iron.

Storage of minerals also requires binding to specific proteins; for example, *apoferritin* protein is required for storage of iron. Excretion of minerals occurs through the renal or the hepato-biliary route.

II. Mineral Deficiency or Excess

Total body content of some minerals is given in Table 19.1. Deficiency or excess of these is potentially hazardous. Because circulating levels of a mineral represents net result of its absorption, utilization, storage and excretion, loss of control over these processes (besides inappropriate dietary intake) causes mineral deficiency, or excess. For example, iron deficiency may occur due to any of the following causes:

1. reduced dietary intake,
2. decreased intestinal absorption
3. excessive loss due to bleeding.

Since minerals are required for performing specific functions, their deficiency leads to defined clinical syndromes. Iron deficiency, for instance, results in microcytic hypochromic anaemia, described later.

Excess content of almost all minerals causes toxic symptoms. Excess iron causes functional impairment of liver and pancreas (condition called **haemochromatosis**); copper overload results in hepato-lenticular degeneration (condition called **Wilsons's disease**), and excessive molybdenum results in several toxic manifestations, collectively referred to as **molybdenosis.**

Decreased intake, impaired absorption, and excessive loss are examples of conditions that could result in deficiency of a trace element, which in turn is associated with impairment of functions. Excess concentration may be toxic.

III. Sources and Daily Requirements of Minerals

The essential minerals are widely distributed in a variety of foodstuffs such as whole-grain cereals, meat, fish, vegetables, fruits and dairy foods. Since concentration of minerals in most foods is very small, it is necessary to consume sufficient quantities of varied foodstuffs to meet the daily requirements of all minerals. Daily requirements and occurrence in enzymes of important minerals are given in Table 19.2.

Importance of trace elements is of far greater magnitude than their concentrations in vivo suggest.

IV. Major Elements

A. Sodium

Sodium is the principal cation of the extra-cellular compartment. About *50 mmol/kg body weight* of sodium is present in human body, of which about 50% is present in bones, 40% in extracellular fluids and the remaining 10% in soft tissues. The plasma concentration of sodium

Table 19.1. Total body content (in moles) of some minerals in adult male of 70 kg body weight

Sodium	4.2	Iron	0.07
Potassium	3.0	Zinc	0.004
Chloride	0.0014	Copper	0.002
Magnesium	0.0012	Manganese	0.0002
Calcium	34	Iodine	0.0005
Phosphorous	23		

Table 19.2. Daily requirement and association with enzymes of certain essential minerals in the human being

Mineral	Daily requirement	Enzyme(s) in which the mineral is found
Ca	800 mg	Coagulation factors, e.g. prothrombin, second messenger
Mg	300–400 mg	Most ATP-requiring enzymes, e.g. *hexokinase, pyruvate kinase*
Fe	1–2 mg	*Proline hydroxylase, catalase diphosphoribonucleoside dehydrogenase, peroxidases*
Cu	2 mg	*Cytochrome oxidase, tyrosinase, lysyl oxidase, superoxide dismutases, monoamine oxidase*
Se	50–200 µg	*Glutathione peroxidase, deiodinase*
Zn	16 mg	Over 200 enzymes, e.g. *carbonic anhydrase, carboxypeptidases, ALA synthase*
Mo	0.15–0.5 mg	*Xanthine oxidase, sulphite oxidase, aldehyde oxidase*
Cr	50–100 µg	Glucose tolerance factor (GTF)
Mn	2.5–5 mg	*Superoxide dismutases*

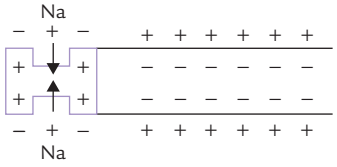
is 135–145 mmol/L. Other extracellular fluids are also rich in sodium, whereas the intracellular fluid concentration is only about 35 mmol/L (Table 19.3).

Biochemical Functions

Sodium is involved in maintenance of irritability of excitable tissues, regulation of osmotic pressure and pH of body fluids, and in membrane transport.

1. *Neuromuscular excitability:* Sodium, together with potassium (the other monovalent cation), increases the neuromuscular excitability; this is counterbalanced by the effect of the divalent cations, calcium and magnesium. The cations are distributed across the cell membrane of nerve fibres in such a way that the membrane exterior is slightly positive compared to the interior. This sets up a potential difference, known as the *resting potential*.

When a stimulus is applied, the localized area becomes permeable to sodium, resulting in inward movement of these ions. Such an influx of positive ions results in the interior becoming slightly electropositive in relation to the exterior (i.e. *action potential*).



This generates a nerve impulse, which is further propagated by the same mechanism, i.e. influx of sodium ions along the entire length of the nerve fibre.

2. *Fluid balance:* Being the major cation of extracellular fluid, sodium plays an important role in maintenance of osmotic pressure, and thus helps to retain water in ECF.

3. *Acid-base balance:* In form of sodium bicarbonate, sodium is a component of the bicarbonate buffer. The latter is the chief buffer system in the extracellular fluid (Chapter 1). The sodium-potassium exchange in renal tubules helps to acidify urine.

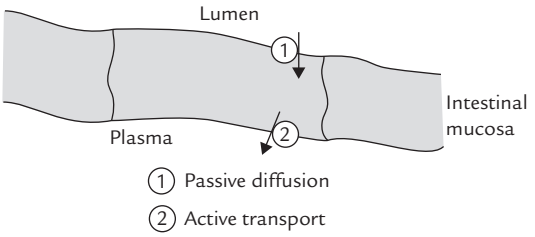
4. *Membrane transport:* Active absorption of a number of substances across the membranous barrier requires sodium cotransport (Fig. 7.16). This is termed the secondary active transport.

Absorption and Elimination

Sodium absorption occurs throughout the small and the large intestine. Because sodium concentration in the intestinal fluid (145 mmol/L) is several folds higher than that in the intestinal mucosal cells (10 mmol/L), sodium passively diffuses into the cell. The intracellular sodium is then actively moved into the plasma by sodium pump through expenditure of ATP energy.

Table 19.3. Distribution of Na^+ and K^+ (cations) in body fluids. Concentration of these cations are given in mmol/L (mean values)

	Na^+ (chief EC cation)	K^+ (chief IC cation)
Extracellular (EC) concentration	140 mmol/L	5 mmol/L
Intracellular (IC) concentration	35 mmol/L	145 mmol/L



Sodium is eliminated from the body via urine. It is a meticulously regulated process, in which ***mineralocorticoids*** (e.g. aldosterone) play a major role: they act at distal convoluted tubules to cause retention of sodium and loss of potassium from the body. ***Glucocorticoids*** and ***sex hormones*** also have the same effect, but they are far less potent than aldosterone.

Requirement and Dietary Sources

Normal Indian diet contains about 5–10 g of sodium, mainly in form of **table salt** (sodium chloride), though meat, fish eggs, milk, cheese, cauliflower, spinach, legumes and nuts are also good sources. The daily requirement in a tropical country like India is **4–6 g.** Excessive intake must be guarded against since it has been shown to lead to hypertension.

Disturbances of Serum Sodium

Hyponatraemia: In this condition the serum sodium level falls below normal. It can be caused by (a) excess sodium loss or it may be (b) secondary to excessive water retention (*dilutional hyponatraemia*).

Hypernatraemia: It is a condition in which serum sodium level is elevated. It may occur as a result of excessive loss of water relative to the sodium loss in the body. Some more common conditions are Cushing's disease, hyperaldosteronism, prolonged cortisone therapy, dehydration and nephrogenic diabetes insipidus.

Raised blood pressure and blood volume are important manifestations of hypernatraemia.

B. Potassium

Potassium is the chief intracellular cation. Body content of potassium is about 40 mEq/kg body weight, nearly

98% of which is located intracellularly. The extracellular concentration is much smaller (about 5 mmol/L) than the intracellular concentration of 145 mmol/L (Table 19.3).

Biochemical Functions

Functions of potassium in the body include regulation of neuromuscular excitability, contraction of the heart, intracellular fluid volume and hydrogen ion concentration.

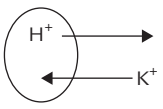
Neuromuscular excitability: Together with sodium, potassium helps maintenance of normal excitability of nerves and ensures smooth conduction of nerve impulses.

Contraction of heart: The potassium concentration in the ECF has a major influence on the contraction of cardiac muscles. A high concentration leads to slowing of heart rate, electrocardiographic abnormalities and possibly cardiac arrhythmia. These may be due to lowering of the membrane potential, which decreases the cell's action-potential intensity.

A low concentration of K^+ increases the membrane potential, decreases irritability, and produces other ECG abnormalities and muscle paralysis. The heart may cease to contract in extreme cases.

Intracellular fluid volume: Potassium in the cell maintains intracellular osmotic pressure and hence, intracellular fluid volume. Nearly half the ICF osmolarity is due to potassium.

Hydrogen ion concentration: The potassium concentration has a significant influence on hydrogen ion concentration in the blood. Movement of K^+ into a blood-cell is normally counter-balanced by movement of H^+ out of the cell. In hypokalaemia (low serum potassium), these counter-balanced movements are decreased, and less H^+ moves out of cell. The hydrogen ion concentration is, therefore, decreased in serum, resulting in alkalosis.



Secondary active transport: Like sodium, potassium is also involved in secondary active transport of many substances. The sodium pump that is involved in active transport of glucose, galactose amino acids, etc. is actually a sodium-potassium pump: it causes active efflux of sodium and influx of potassium.

Dietary Sources, Absorption and Elimination

Potassium is abundant in foodstuffs of plant and animal origin since it is the principal intracellular cation. Therefore, its daily requirement of about 4 g per day can be easily met and so dietary deficiency is rarely seen. Some important food sources are meat, fish, cereals, vegetables, oranges and peaches. Potassium absorption occurs by passive diffusion along a concentration gradient in both small and large intestine. Excretion mainly occurs through renal route.

Disturbances of Serum Potassium

Hypokalaemia: Fall in serum potassium levels ***below 3 mmol/L*** is called hypokalaemia. It may occur when excess potassium is lost from the body, when there is a reduced dietary intake, or when potassium is redistributed within the body.

(a) Loss of potassium may occur via gastrointestinal route, in renal tubular acidosis, K^+-losing nephritis, hyperaldosteronism, etc.
(b) Decreased dietary intake occurs with chronic starvation or anorexia nervosa.
(c) Hypokalaemia due to redistribution may be caused by insulin: insulin induces potassium to move into the cells. Magnesium deficiency may also lead to potassium deficiency.

Manifestations of hypokalaemia: Hypokalaemia manifests as muscular weakness, confusion, irregular heart beat, tachycardia, and altered ECG pattern (flattening of ECG waves).

Hyperkalaemia: Although less common, hyperkalaemia (***serum potassium > 5.5 mmol/L***) is dangerous because of its effect on cardiac muscles. It may be caused by:

(a) ***Failure of the kidneys to excrete potassium***, as in Addison's disease, in which low aldosterone production prevents potassium excretion. Renal failure can cause hyperkalaemia when daily urine output drops below 400–500 mL per day.
(b) ***Redistribution of potassium***, which occurs in acidosis and crush injuries. These conditions lead to hyperkalaemia even though total body potassium is not increased. In acidosis, potassium moves out of the cells as hydrogen ions move in, and in crush injuries, the damaged tissues release their potassium content into blood circulation.
(c) ***Others:*** Dehydration, massive blood transfusion and indiscriminate potassium therapy are other important causes of hypokalaemia.

Clinical manifestations: The symptoms of hyperkalaemia are similar to those of hypokalaemia in that myocardial irritability, irregular heartbeat, ECG changes, and muscle weakness may occur in either condition. *The greatest danger, however, is the possibility of cardiac arrest at levels greater than 7.0 mmol/L.*

C. Chloride

It is the major anion in the body. Normal serum chloride levels (**with sodium levels**), in general, undergo parallel alterations. The chloride in the CSF is about 125 mmol/L, which is higher than in any other body fluid. Total body content of an average adult is about 80 g.

Biochemical functions: Chloride is the major extracellular anion, and so, along with sodium, it is involved in *regulation of osmotic pressure of extracellular fluids.* It helps maintenance of pH of blood (see chloride shift, Chapter 1), and is also involved in the *formation of hydrochloric acid in stomach*.

Absorption: Chloride in food is almost completely absorbed, mostly in the proximal ileum by passive diffusion. In distal ileum and colon, chloride absorption occurs in exchange for bicarbonate ions.

Requirement and dietary sources: The daily requirement of chloride is 5–10 g. Most important source is sodium chloride, the **table salt.** Foods that provide sodium also provide chloride, e.g. meat, fish, cheese, cereals, eggs, etc. Thus, intake of sodium and chloride runs parallel.

Alterations of serum chloride levels: With a possible exception of metabolic alkalosis, serum chloride level undergoes parallel changes with sodium concentration.

Hyperchloraemia (increased chloride concentration) occurs in conditions of (a) dehydration, (b) adrenocortical hyperactivity: adrenal steroids cause increased reabsorption from renal tubules, (c) severe diarrhoea, leading to loss of bicarbonate and compensatory retention of chloride, and (d) respiratory alkalosis, metabolic acidosis and renal tubular acidosis.

Hypochloraemia is seen in (a) severe vomiting: loss of HCl decreases chloride concentration, and so compensatory rise in bicarbonate occurs, (b) Addison's disease: mineralocorticoid depletion causes impaired reabsorption of chloride from renal tubules, (c) prolonged gastric suction, and (d) respiratory acidosis and metabolic alkalosis.

D. Calcium and Phosphorus

These elements are discussed in Chapter 31.

E. Magnesium

Magnesium, an *essential activator of over 300 enzymes in humans*, is the fourth most abundant cation in the body after sodium, potassium and calcium (Table 19.1). In intracellular fluid, it is the second most abundant cation after potassium. It is also an important constituent of **chlorophyll**, just as iron is an essential component of haemoglobin. *An average adult contains about 25 g of magnesium, about 70% of which is present in skeleton.* The remaining 30% occurs in soft tissues (mainly liver and muscles) and body fluids. Only about 1% of the body magnesium is in the blood and ECF. The skeletal and ECF magnesium pools exchange freely with each other but not with the intracellular pool.

Serum concentration of magnesium is **1.6–2.3 mg/dL** (0.7–1.0 mmol/L), of which, about 20–25% is protein bound. In a manner similar to calcium, the unbound (ionized) portion of serum magnesium is the biologically active fraction. The intracellular magnesium concentration is much higher (about 10 mmol/L) than that in ECF, most of which is bound to organelles. Inside RBCs, the concentration is about 20 mmol/L.

Biochemical Functions

Enzyme activator: Magnesium activates a number of enzymes, particularly those in which ATP^{4-}–Mg^{2+} (a complex of ATP and magnesium) is a substrate. The magnesium ion is chelated between the β- and the γ-phosphates of ATP, thus diminishing the anionic character of the latter (Fig. 19.1). This permits smooth interaction of the ATP with specific protein sites. The enzymes needing magnesium are involved in intermediary metabolism, transcellular ion transport, muscle contraction, and oxidative phosphorylation. Some examples are *hexokinase, phosphofructokinase, pyruvate kinase, thiokinase, squalene synthase, glutamine synthetase, adenylate cyclase,* etc.

Neuromuscular excitability: It is decreased by magnesium. A low magnesium level results in increased nerve excitability.

Glucose tolerance: Magnesium improves glucose tolerance, probably by enhancing insulin dependent uptake of glucose.

Others: Magnesium binds to other nucleoside phosphates and nucleic acids and is required for DNA replication, transcription and translation. The DNA helix is stabilized by binding of histones and magnesium to the exposed phosphate groups. In green plants, chlorophyll—a magnesium porphyrin complex similar to haem—is vital for photosynthesis.

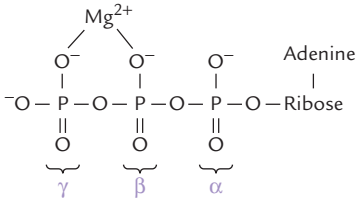


Fig. 19.1. Mg^{2+} complex of ATP. It decreases the anionic character of ATP.

Absorptions and Elimination

Magnesium is mostly absorbed from the small intestine but a small amount is also absorbed from colon. About 50% of the ingested magnesium is absorbed on an average diet, and the extent of absorption decreases as the dietary content of this element increases. Parathormones appear to increase the absorption.

Kidneys are the main organs responsible for maintaining plasma magnesium concentration within normal limits. Most of the magnesium (about 98%) filtered at glomeruli is reabsorbed. Urinary excretion varies with plasma magnesium concentration. PTH enhances the tubular reabsorption of magnesium, whereas aldosterone decreases it.

Requirement and Dietary Sources

Daily requirement is about 350 mg in adult men, 300 mg in adult women and 150–250 mg in children. In pregnancy and during lactation about 450 mg/day is needed to maintain positive magnesium balance.

Major food sources are nuts, cereals, beans and fish; almonds are a particularly rich source. Leafy green vegetables are a good source because of the high chlorophyll content.

Disturbances of Serum Magnesium

Serum magnesium level may fall (*hypomagnesaemia*) in a number of clinical settings and contributes to a poorer prognosis for the patient, especially when below 1.7 mg/dL. Some important causes are diabetes, chronic alcoholism, liver cirrhosis, hyperparathyroidism, aldosteronism, protein calorie malnutrition and drug therapy. Hypomagnesaemia leads to neuromuscular hyperirritability, tremors, increased vascular resistance, coronary vasospasm and hypertension.

Increase in serum magnesium level (*hypermagnesaemia*) is commonly seen in renal failure.

V. Trace Elements

Almost a dozen trace minerals are known to be essential in diet. Most of these are toxic at higher doses. Therefore, they are complexed with specific intracellular binding proteins, which not only store the mineral but also prevent toxicity.

Iron, copper, selenium, zinc, molybdenum, chromium, manganese, iodine and fluoride are some important trace elements, discussed below.

A. Iron

Iron is a very important element in medical practice. Though it is one of the most abundant elements on earth, yet only trace amounts are present in the living cells. A normal adult of 70 kg body weight contains about **3–4 g** of iron. Daily turnover is high (30–40 mg) since iron is continually circulated in various metabolic pathways. About **65–70%** of the total iron in human body is found **in haemoglobin** and approximately **10% in myoglobin** and other iron-containing enzymes and proteins. The remaining 20–25% consists of storage pool of iron.

By comparison, the average adult woman has only 2–3 g of iron in her body. This difference is attributable to (a) lower haemoglobin concentration, (b) smaller vascular volume, and (c) lower iron reserves in women.

Iron distribution is summarized in Table 19.4.

Biochemical Functions

Iron is essential for several biological processes, and therefore, biologically indispensable.

1. ***Tissue respiration:*** Iron can change readily between the ferrous and the ferric states, and can therefore function in *electron transfer reactions*. Both the haem iron of the cytochromes and non-haem iron in iron-sulphur proteins (e.g. *NADH dehydrogenase* and *succinate dehydrogenase*) are used in this way.

2. ***Transport of gases:*** Iron is able to bind a variety of ligands, including molecular oxygen and carbon dioxide. This property is exploited by oxygen-binding proteins, such as haemoglobin and myoglobin for its transport and storage.

3. ***Oxidative reactions:*** Iron is a component of various *oxido-reductase enzymes* (Table 6.1), and so plays a vital role in a number of oxidative reactions in the body.

4. ***Others:*** Iron is required for an effective immune response due to a variety of reasons. For instance, it is required for activity of the ***lysosomal peroxidase***, which helps in phagocytic and bactericidal activities of neutrophils.

Table 19.4. Distribution of iron in the body

Compound	Per cent of total	Function
Haemoglobin	65	Oxygen transport, blood
Myoglobin	10	Oxygen storage, muscle
Ferritin*	10	Iron storage
Haemosiderin*	9	Iron storage
Transferrin*	0.2	Iron transport
Enzymes (*catalase/peroxidase*)	<1.0	H_2O_2 decomposition, oxidation
Iron-sulphur*		Electron transfer
Cytochromes	<0.5	Electron transport
Other	<5	Oxidation-reduction reactions

*Non-haem iron compounds.

Requirements and Dietary Sources

Adult men	10 mg/day
Pre-menopausal women	18 mg/day
Post-menopausal women	10 mg/day
Pregnant women	40 mg/day
Children	10 mg/day

About ***1 mg iron is eliminated each day*** from the human body, mainly by shedding of skin epithelial cells and cells lining the urinary tracts, and a to a smaller extent in urine and sweat. *This much amount must be replaced by dietary intake.* An average Indian diet provides about 10–20 mg of iron, of which only about 10% is absorbed. This balances the average daily iron loss in men and post-menopausal women.

However, **requirements are higher in the women** since blood is lost in each *menstrual cycle.* Each cycle drains 20–40 mg of iron, and therefore, an additional 1 mg of iron per day is required. The diversion of iron to the fetus during pregnancy, blood loss during delivery, and subsequent breast feeding of the infant consumes 900 mg or iron on average. This increases daily iron demand to 3 mg or 4 mg in pregnant and lactating women.

Dietary sources include meat, fish, eggs, cereals and green leafy vegetables. *Milk is a poor source.* Iron obtained from animal sources is more readily available than that obtained from vegetarian food. This is because the latter is complexed with certain inorganic molecules, such as phytates, which impede its absorption.

Absorption

Intestinal absorption is especially important in case of iron. It is a regulated process and the efficiency of absorption increases or decreases depending on the body requirement. *The iron stores in the body are regulated by intestinal absorption.* Exact mechanism of absorption is still unknown, a working hypothesis is as below:

The dietary iron, which exists mostly in the ferric form, is converted to the more soluble ferrous form, which is readily absorbed. The ferric to ferrous conversion is brought about by action of HCl and reducing agents such as ascorbic acid, cyseine and –SH groups of proteins.

$$Fe^{3+} \xrightarrow[\text{Vitamin C, Cysteine}]{\textit{Ferric reductase}, HCl} Fe^{2+}$$

Entry of Fe^{2+} into the mucosal cell may be aided by an enzyme on the brush-border of the enterocyte (the enzyme possesses *ferric reductase* activity also). The ferrous iron is then transported in the cell by a divalent metal transporter (**DMT-1**), as shown in Figure 19.2.

Iron in Intestinal Cells

Within the intestinal cell, the iron may be either **stored** or **transported out**.

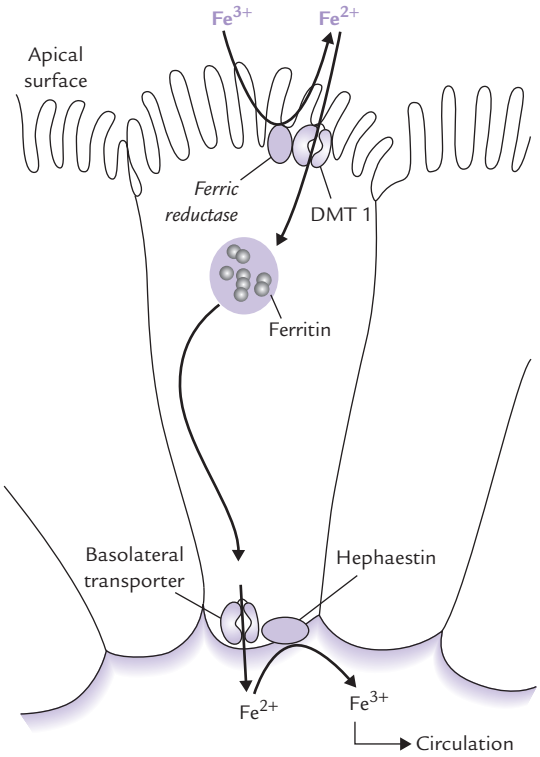


Fig. 19.2. Mechanism of iron absorption (DMT = divalent metal transporte).

(a) ***Stored*** by incorporation into **ferritin** in those individuals who have adequate plasma iron concentration. A *ferroxidase* converts the absorbed ferrous iron to the ferric form, which then combines with apoferritin to form ferritin.

$$Fe^{2+} \xrightarrow[\textit{Ferroxidase}]{} Fe^{3+} \xrightarrow[\text{Apoferritin}]{} \text{Ferritin } (Fe^{3+})$$

(b) ***Transported*** to a transport protein at basolateral cell membrane and *released into the circulation.* However, the ***basolateral-transport-protein*** has not yet been identified. It is believed to work in combination with *hephaestin*, a copper-containing protein, which oxidizes Fe^{2+} back to Fe^{3+}.

The intestinal cells internalize more iron than the amount that eventually will enter the circulation. The surplus, incorporated into ferritin for storage, is subsequently mobilized, if necessary. The ferritin stores are gradually built-up, but most are lost when the mucosal cells are shed. New cells take their place, and the cycle of iron build-up starts again.

Factors Affecting Iron Absorption

1. Gastric acid and dietary components that form soluble-ferrous-chelates (such as *ascorbic acid, sugars* and *amino acids*) keep iron in solution and **increase** its absorption.

2. Substances that form insoluble complexes with iron, such as **oxalates** and **phytates** (in vegetables), phosphates (in milk and eggs), and tannates (in tea), **decrease** iron absorption.
3. When body needs for iron increase, such as in iron deficiency, pregnancy and accelerated erythropoiesis, iron absorption is stimulated three-folds or more. On the other hand, the absorption is reduced after the consumption of iron or iron poisoning.
4. Marginal decrease in iron absorption occurs by tea and eggs.

The iron balance is controlled by changes in absorption, which speeds up or slows down depending on the body needs. *No other nutrient is regulated in this manner.* The overall effect is to prevent the absorption of excess iron (excess iron is toxic because it can bind to and disrupt structure of many proteins) while maintaining an adequate supply for current needs.

Absorption of haem iron: The above description relates to non-haem iron. Absorption of the haem iron, which is present in non-vegetarian foods, is a simpler and more rapid process. After haem is released from the polypeptide chain, it is absorbed intact by the intestinal cells, where porphyrin ring is split and iron is liberated. This process is not only more efficient, but also is not affected by dietary factors.

Serum iron concentration is low but total body iron content is approximately 4 g. Absorption of iron from the intestine is the primary means of regulating the amount of iron in the body.

Transport in Plasma and Cell-uptake

The iron absorbed from intestine (or that liberated from the ferritin of mucosal cells) is converted to ferric form by a copper-containing enzyme *ceruloplasmin*, which possesses *ferroxidase* activity. Ferric iron then binds with ***transferrin***, a liver derived β-globulin, which binds essentially all the iron in plasma (Fig. 19.3). Need for a specific carrier protein arises because free, unbound iron is quite toxic due to two reasons:

- It can disrupt native conformation of several biologically active proteins, and
- Initiates oxidative damage by forming reactive hydroxyl and peroxide radicals in the presence of molecular oxygen.

Transferrin is a single chain glycoprotein with a molecular weight of 78 kDa and a plasma concentration of approximately 300 mg/dL. Each transferrin molecule has two high affinity binding sites for ferric (but not ferrous) iron. Only approximately one-third of the binding sites are occupied. This is expressed as an iron saturation of 33%. In the laboratory, the transferrin concentration is measured as the **total iron-binding capacity** (**TIBC**).

The transferrin delivers iron to the cells with specific surface receptors for this protein. *These receptors bind with the iron loaded transferrin, and the transferrin-receptor complex is then taken into the cell by endocytosis.* The endocytotic vesicle becomes acidified to a pH of approximately 5.0. At this pH, the iron dissociates from transferrin and transported into the cytoplasm. The receptor-transferrin complex is then returned to the cell surface, where it undergoes additional rounds of iron transport and uptake.

Iron Storage

After the iron is delivered to the peripheral cells, it can either be used for various biological activities, or stored. The storage occurs in two proteins: **ferritin** and **haemosiderin**, which abound in liver, spleen, bone marrow, intestinal mucosa, pancreas, myocardium and other tissues. *Storage is important because it serves to package and isolate iron atoms from the intracellular environment, thus preventing any toxic action on cell constituents.*

Ferritin: It consists of a multi-subunit protein shell, known as *apoferritin* (MW 500,000). This shell with an external diameter of 13 nm and an internal cavity of 6 nm across surrounds a core of up to 4500 ferric ions. Ferritin is a readily mobilized form of storage iron.

Haemosiderin: It is derived from ferritin that loses some of its surface protein and undergoes aggregation to form micellar complexes. These complexes, now called haemosiderin, have higher iron content than ferritin, but release iron more slowly.

Substantial quantities of iron can be stored (up to several grams) in some older males. About one-third of the storage iron is present in liver, one-third in bone marrow and the remainder in spleen and other tissues. On the other extreme are many children and menstruating women, in whom the storage iron is nearly absent.

The iron (Fe^{3+}) is transported in plasma by transferrin, which delivers it to the tissue for use. It (Fe^{3+}) is stored by incorporation into ferritin in tissues such as liver, spleen and bone marrow.

The principal pathways of iron metabolism are outlined in Figure 19.3. Significance of RBCs is highlighted in Box 19.1.

Pathological Conditions

Iron overload: In iron overload, the iron molecules combine with the available apoferritin resulting in excessive production of ferritin. As mentioned earlier, surplus iron may cause

excessive formation and accumulation of haemosiderin; the condition is termed **haemosiderosis.** Various parameters in iron overload (and deficiency) are given in Table 19.5.

Iron overload is symptom-less initially. Gradually, however, massive deposits of haemosiderin may develop, which may cause functional impairment of the involved organs. The condition is then called **haemochromatosis.** It may be genetic (e.g. *primary* haemochromatosis) or acquired, called *secondary* haemochromatosis (Box 19.2).

Signs and symptoms of haemochromatosis: These depend on the organ system involved. Accumulation in liver can result in cirrhosis. In the pancreas, the excess iron can damage the β-cells to cause diabetes mellitus (seen in two-thirds of patients). The latter condition is termed **bronze diabetes.**

Cardiomyopathy, hyperpigmentation of the skin, and arthralgia (joint pain) are the other frequent signs.

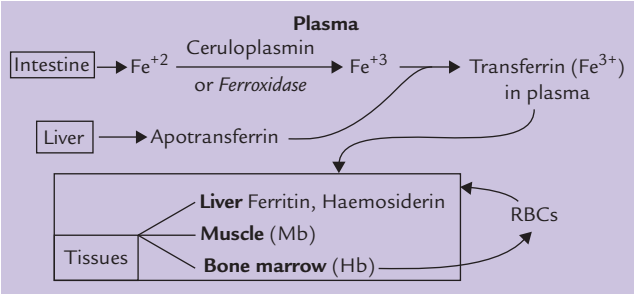


Fig. 19.3 Principal pathways of iron metabolism (Hb = haemoglobin, Mb = myoglobin).

Treatment: Haemochromatosis is treated by repeated *phlebotomy.* This is indeed the only disease in which the traditional use of leech (*Hirudo* sp.) is the treatment of choice. Iron-chelating agents, e.g. *desferrioxamine,* are used to remove excess iron. They form soluble iron complex that is rapidly excreted in urine.

Iron deficiency: Iron deficiency is a serious health problem. With possible exception of protein calorie malnutrition, iron deficiency is the *second most prevalent nutritional deficiency worldwide,* with a total of at least half a billion people affected. The deficiency may be caused by inadequate dietary intake of iron, impaired absorption, or by chronic bleeding due to piles, peptic ulcer, hookworm infestation or menstrual irregularities. Growing children, adolescent girls, pregnant and lactating women are the worst sufferers. Iron deficiency results in a **microcytic hypochromic anaemia** (microcytic = small RBCs; hypochromic = low haemoglobin content per cell). Biochemical indices of the deficiency are summarized in Table 19.5 and other relevant details are presented in (**Case 19.1**).

Iron deficiency causes microcytic, hypochromic anaemia. Excessive iron stores may be caused by primary haemochromatosis, a genetic defect, or a variety of secondary conditions, e.g. repeated blood transfusions, long term intake of supplement iron.

BOX 19.1

Red Blood Cell Turnover

Daily turnover of iron is **30–40 mg.** The absorbed iron (1 mg) represents only a small portion of it; the rest comes from destruction of old erythrocytes by tissue macrophages, primarily in the spleen. Within these cells, *haem oxygenase* cleaves the porphyrin ring to release iron, which is transferred to plasma transferrin and thereby carried to the bone marrow for haemoglobin synthesis. In this manner, the reticuloendothelial system continuously recycles iron from old red cells to new ones.

Table 19.5. Iron parameters in iron overload and deficiency

Parameter	Normal range	Iron deficiency	Iron overload
Iron store (g)	0.5–1.5	None	Up to 50 g
Iron absorption (mg/day)	1–2	10–20	↑ or ↓ (depending on cause)[a]
Serum total iron (μg/dL)	120	15	250
Serum total iron-binding capacity (μ/dL)	350	475[b]	263[c]
Serum ferritin (μg/mL)	12–250	< 10	> 1000

[a]↑ in idiopathic haemochromatosis, ↓ in most other conditions of iron overload.
[b]reflects increased serum transferrin levels.
[c]almost equal to total serum iron.

BOX 19.2

Haemochromatosis: Massive Iron Overload with Impairment of Function

Primary haemochromatosis: Patients with primary haemochromatosis, also called idiopathic haemochromatosis, absorb 4 mg or more of iron per day, even with a normal diet. Over a period of decades this leads to accumulation of 20 g or more of iron in the body. Since considerable time period is required for iron to accumulate, the disorder is uncommon before 40 years of age.

Men are affected 5–10 times more because the pre-menopausal women are protected by menstruation and child bearing (but women homozygous for the gene defect can develop clinical signs of iron overload after menopause). Inheritance is autosomal recessive, but for the reasons that are not clearly understood, only a fraction of homozygotes develop the full-blown disease.

The gene responsible for most cases of haemochromatosis encodes a protein known as HFE that appears to interact with transferrin and the transferrin receptor to regulate the amount of iron absorbed by cells. Patients with specific mutation in the HFE gene (C 282Y) absorb more iron.

Secondary haemochromatosis is caused by (a) impaired utilization of iron, (b) increased breakdown of haemoglobin, and (c) large amount of dietary iron. More than one of the above-mentioned factors are responsible for developing excess iron in thalassaemia. The patients do not incorporate iron normally in this disorder and the picture is further complicated by multiple blood transfusions (since the transfused RBCs have a lower life span).

Chronic use of wine enhances risk of secondary haemochromatosis due to high iron content of these drinks. Alcohol appears to increase iron absorption. The Bantu tribes of Africa develop haemochromatosis because of excessive amount of consumption of alcoholic beverages brewed in iron vessels.

B. Copper

Copper primarily functions as a component of metalloenzymes or proteins that participate in redox reactions. Adult human body contains **60–100** mg of copper which is widely distributed in a number of tissues, but largest amounts are present in muscles (30–50 mg), bones (10–20 mg), and liver (10–15 mg).

Biochemical Functions

1. **Copper is an essential component of several *oxidases*.** As a rule, these enzymes use either molecular oxygen or an oxygen derivative as one of their substrate. Examples include:
   - *Cytochrome oxidase*, which transfers electrons to oxygen, the ultimate acceptor of electron transport chain.
   - *Lysyl oxidase*, which catalyzes cross linking in collagen and elastin by converting certain lysine residues to allysine.
   - *Tyrosinase*, which is necessary for the synthesis of melanin; *ALA synthase*, which is required for haem synthesis; and several others, e.g. *superoxide dismutase, ascorbic acid oxidase, uricase, dopamine hydroxylase and monoamine oxidase.*
2. **Ceruloplasmin**, a copper-containing enzyme in plasma possesses *ferroxidase* activity; it converts Fe^{2+} to Fe^{3+}, the form in which the transferrin iron is transported in plasma.
3. **Certain non-enzyme proteins** also require copper, although their functions are not clearly known. These include *hepatocuperin* (in liver), *cerebrocuperin* (in brain) and *haemocuperin* (in red cells).

Requirements and Dietary Sources

Only about **2 mg/day** copper is required by adults (Table 19.2) and about half this amount by children. The average cereal-based Indian diet provides about 4 mg copper each day, which can meet the body's requirements. *Cereals, legumes, raisins, nuts, green leafy vegetables, liver, kidney, meat, and egg yolk* are common dietary sources. Milk is a poor source.

Absorption

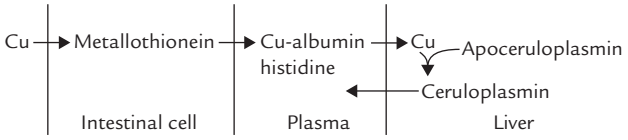
Intestinal absorption of copper occurs mainly from duodenum. Less than one-third of the dietary copper is absorbed from an average Indian diet, because dietary constituents (e.g. phytates, fibres, Zn, Mo) reduce its absorbability (from western diet 50% copper content is absorbed). Moreover, the *intestinal absorption is a sluggish process because of highly insoluble nature of copper.* Mechanisms, therefore exist to enhance its solubility. For example, in the intestinal lumen, copper complexes with a low molecular weight substance secreted by saliva and gastric juice, which keeps it soluble at the intestinal pH. Precise nature of this substance is still unknown.

The absorptive process consists of three phases:

1. ***Mucosal uptake:*** The ingested copper enters the intestinal epithelial cells by moving across luminal aspect of the cell membrane.

2. ***Intracellular transport:*** Inside the cell, copper is transported by a low molecular weight, binding protein, called **metallothionein.**
3. ***Extrusion:*** The intracellular copper is then moved across the serosal aspect of the mucosal cell membrane by an enzyme *copper-binding P-type ATPase*. The enzyme, present in intestinal mucosa (and some other cell types), effectively transfers copper into portal circulation from where it reaches the liver.

Within hepatocytes, an intracellular protein facilitates incorporation of copper into *apoceruloplasmin*, an endogenously synthesized α_2-globulin. This results in formation of ***ceruloplasmin***, which contains six to eight atoms of copper. Half of these are in cuprous state and the half in the cupric state. Failure to synthesize this protein is implicated in pathogenesis of Wilson's disease, discussed later.



Transport and Excretion

Plasma concentration of copper is 100–200 mg/dL, about 95% of which is bound to ceruloplasmin. *Ceruloplasmin is an enzyme (ferroxidase) and an antioxidant, but not a transport protein* since the copper is so tightly bound to it that it is not readily exchangeable with other molecules. The major carrier of copper is albumin, which loosely binds with copper. A small fraction of plasma copper is complexed with amino acids, mainly histidine.

Excretion of copper occurs mainly in the **bile,** which eliminates 1.5–2.0 mg/day. The urinary excretion is normally < 400 μg/day. A transport protien, the *copper binding ATPase* moves copper from lines cells into bile.

Disorders of Copper Metabolism

1. ***Copper deficiency:*** It is most likely to occur from reduced intake or excessive loss, e.g. in renal dialysis. Deficiency manifests as microcytic (small erythrocytes) hypochromic (pale erythrocytes) anaemia that is resistant to iron therapy. There is decreased number of leukocytes in the blood (neutropenia), degeneration of vascular tissue with bleeding (due to defects in elastin and collagen production), and skin de-pigmentation (because melanin synthesis requires *tyrosinase*, a copper-containing enzyme). Demyelination of neural tissue, myocardial fibrosis, and grey hair are the other manifestations.

Copper is a component of several enzymes or proteins involved in redox reactions. Copper deficiency causes microcytic hypochromic anaemia, associated with low concentration of ceruloplasmin, a copper-containing protein.

2. ***Menkes' disease:*** It is an X-linked genetic defect of deficient production of the *copper-binding P-type ATPase* in intestinal mucosa (and most other tissues). Copper accumulates in intestinal mucosa but cannot be released into circulation. Decreased copper in plasma and urine results, and in untreated cases more severe manifestations, such as mental retardation, abnormal collagen formation, greying of hair, and temperature instability results.
3. ***Wilson's disease (hepatolenticular degeneration):*** It is a rare, autosomal recessive disorder in which excessive *accumulation of copper occurs in tissues*. The possible causes are:
   (a) The gene encoding the ***copper-binding ATPase*** in hepatocytes, which is required for excretion of copper from liver cells, may be defective. In its absence copper is not eliminated (in bile) and accumulates.
   (b) Some workers suggest **failure to synthesize ceruloplasmin** or an impairment in the binding capacity of copper to this protein as the possible cause. Ceruloplasmin level less than 20 mg/dL is considered pathognomonic of Wilson's hepatocellular degeneration.

Excess copper accumulates in liver and brain, leading to liver damage or neurological degeneration, or both. Copper deposition in kidney may cause renal damage, with consequent excretion of amino acids, glucose, peptides and haemoglobin in urine. Copper deposits in bone marrow cause **haemolytic anaemia** and those in cornea form green or golden pigmented ring, the **Kayser–Fleischer ring.** These and some other manifestations of the disease are given in **Case 19.2**. D-Pencillamine, a copper chelator, is used in treatment. It forms a soluble copper complex, and permits its urinary excretion.

In Wilson's disease, patients develop copper overload in brain and liver, resulting in cirrhosis of the liver, brain lesions, and sometimes bronze-diabetes.

4. ***Copper toxicity:*** Copper is generally non-toxic, but in large doses it accumulates in tissues and can interfere with other metal ions, specifically iron and zinc. Acute toxicity manifests as *blue-green stools* and *saliva, acute haemolysis*, and *abnormalities of kidney functions.*
5. ***Indian childhood cirrhosis:*** Chronic excessive intake of copper and consequent deposition in liver occurs in this disorder. Feeding of milk boiled in brass utensils to infants is probably the cause of the excess.

C. Other Trace Elements

A number of other trace elements serve as prosthetic groups or cofactors in various enzymatic reactions and are,

therefore, required in our diet, though the daily requirement is very low. Examples include selenium, chromium, cobalt, iodine, manganese, and molybdenum, zinc.

The daily requirements and biological roles of these elements are outlined in Table 19.2.

Selenium

Total selenium content in adult human is only about **6 mg.** Skin, liver, pancreas, and renal cortex contain relatively large amounts of this mineral. It is known that in plants selenium is present as *selenomethionine* (in animals as *selenocysteine*).

Biochemical Functions

1. ***Antioxidant role:*** In the form of selenocysteine, this element occurs in *glutathione peroxidase* (GP), an enzyme with strong antioxidant properties. It prevents peroxidation of lipids and other compounds by hydrogen peroxide.

$$H_2O_2 + \underset{2 \cdot GSH}{\text{Glutathione (red)}} \longrightarrow \underset{GS\text{-}SG}{\text{Glutathione (oxd)}} + 2H_2O$$

   The above reaction may be considered as a detoxification reaction in which selenium acts as a non-specific antioxidant (being a component of *glutathione peroxidase*). This protects the tissues against potentially toxic effect of hydrogen peroxide and permits the life forms to survive in an oxygen atmosphere.
2. ***Synergism with vitamin E:*** Selenium prevents lipid peroxidation and offers protection against the free radicals. Vitamin E also has the same effect (antioxidant), though by a different mechanism (Chapter 18). *Thus, selenium and vitamin E supplement each other.* Availability of vitamin E reduces the selenium requirement, and conversely, in selenium-deficient tissue, vitamin E content is depleted.
3. ***Others:*** The other possible roles of selenium are that it is part of coenzyme Q, cofactor for the enzyme ***deiodinase,*** and plays a role in immune reactions. It binds with certain toxic, heavy metals (Cd and Hg) and protects the body against their harmful effects.

Requirements and Dietary Sources

Daily requirement of selenium is 50–200 mg for an adult (Table 19.2). It is present in sea-foods and organ meats such as kidney and liver.

Deficiency and Toxicity

Both selenium deficiency and toxicity are well known. **Keshan disease,** an endemic cardiomyopathy (occurs in Keshan Province of China), is caused by low selenium content of locally grown foodstuff. In fact, deficiency of this mineral has been reported when soil is deficient in it. For instance, certain mountainous regions are deficient in selenium, which is reflected in the low selenium content of the food plants grown in these soils (**Case 19.3**).

Selenium toxicity, called **selenosis,** occurs by accidental ingestion of metal polishes or anti-rust compounds. It manifests as garlicky odor in breath, hair and weight loss, diarrhoea and falling of nails.

Zinc

Zinc is the intracellular cation present in all body tissues and fluids and, next to iron, is the *second most abundant* of the trace metals in humans. Prostate, liver, kidneys, muscles, heart, skin, bones, teeth, etc. are particularly rich in zinc. Total zinc content of the adult body is about **2 g**, out of which 60% is in muscles, 28% in bones and 0.5% in blood. Erythrocytes contain 75–88% of the blood zinc. In the plasma, about 18% of the zinc (normal range 700–1200 μg/L) is tightly bound to an α_2-macroglobulin, 80% is loosely bound to albumin, 2% is bound to transferrin, ceruloplasmin, or the amino acids, and a small fraction is present as free zinc.

Biochemical Functions

1. Zinc in the form of **zinc fingers,** is an important constituent of the regulatory proteins that control transcription.
2. **More than 300 metalloenzymes require zinc** for their catalytic activity, including the enzymes critical for nucleic acid and protein synthesis, e.g. ***DNA* and *RNA polymerases,*** and ***reverse transcriptase.*** Thus, zinc is a necessary component for gene expression and cell replication. Other important zinc containing enzymes are *carbonic anhydrase, extra-mitochondrial superoxide dismutase, carboxypeptidases A and B, ethanol dehydrogenase, lactate dehydrogenase, malate dehydrogenase, alkaline phosphatase* and *glutamate dehydrogenase.*
3. Insulin when stored in the β-cells of pancreas contains zinc that is probably required for stabilization of the hormone molecule and its release.
4. Along with other trace metals (Cu and Mn), zinc is associated with a group of enzymes called *superoxide dismutases*, which scavenge the superoxide anion. The latter are by-products of various redox reactions or the electron transport system and may give rise to the very destructive hydroxy ($^{\bullet}OH$) radicals (Chapter 27).
5. Zinc is also required for immunological functions.
6. Zinc induces synthesis of metallothionein, a small protein of 61 amino acids (MW 7000). Both Zn and Cu are stored in tissues bound to this protein. The protein binds copper more firmly than zinc and forms

an unabsorbable complex in the gastrointestinal tract, hence reducing copper absorption. Therefore, dietary zinc supplementation is known to significantly lower the absorption of Cu and may elicit Cu deficiency symptoms.

(Zinc fingers, defined as domains of zinc-binding proteins that also bind to DNA, are involved in the gene expression of metallothionein.)

Zinc is a cofactor for over 300 enzymes (*carbonic anhydrase, alcohol dehydrogenase, alkaline phosphatase, DNA polymerase*), which are essential to growth, wound healing, reproductive function, and protection from free radical damage.

Requirement and Dietary Sources

Infants	5 mg/day
Children and adolescents	10 mg/day
Adults (males)	15 mg/day
Adults (females)	12 mg/day

In pregnant and lactating females, the daily requirement increases to 15 mg/day and 19 mg/day, respectively.

Seafood, red meat, fish, eggs and milk are good sources of zinc. Although vegetables contain appreciable amounts of zinc, the fiber and phytate in them bind zinc and hence diminishes its bioavailability. Hence, vegetarians are at greater risk for zinc deficiency.

Metabolism

Zinc is absorbed from the duodenum. Its absorption is incomplete because of the presence of certain interfering substances (e.g. fibres and phytates) in the food, which form insoluble complexes with zinc to impair absorption. *Copper, iron, cadmium and calcium also interfere with zinc absorption.*

In the plasma, there is no specific binding protein, and zinc is transported mainly in association with albumin. It is stored in many tissues in the form of *metallothionein*, which also binds copper and some other heavy metals. Metallothionein protects the cells from the toxic effects of the free, unbound metal ions. Zinc is present in pancreatic secretions, and the stools are the major route for its excretion. Urine, sweat and seminal fluid are the excretory routes of minor importance.

Deficiency

Acute zinc deficiency in humans, especially in growing children, manifests as skin lesions, testicular atrophy, poor growth, delayed sexual development and increased susceptibility to infections. Neuropsychiatric impairments and decreased taste acuity are the other prominent features.

As noted above, red meat being prime source of bioavailable zinc, vegetarians are at much greater risk for zinc deficiency.

Zinc deficiency may occur in **acrodermatitis enteropathica,** a rare, recessively inherited disease characterized by dermatitis, diarrhoea, and alopecia (hair loss). It is caused by an impairment of intestinal zinc absorption. High doses of orally administered zinc are curative.

Toxicity

Foods stored in galvanized containers have been reported to result in high zinc in-take. This causes epigastric pain, diarrhoea and vomiting. Supplements as little as 25 mg of zinc have been observed to cause diminished absorption of copper, probably because of competition.

Molybdenum

Biochemical Functions

Very small amounts of this element are present in the human body, mainly in liver and kidneys. It is known to function at active sites of at least three enzymes: *xanthine oxidase, aldehyde oxidase* and *sulphite oxidase*. The last one is a haem-protein similar to cytochrome C. The plant enzyme *nitrite reductase*, required for nitrogen fixation, also contains molybdenum.

Molybdenum is incorporated into several *oxidases: xanthine oxidase, aldehyde oxidase*, and *sulphite oxidase*.

Requirement, Dietary Sources and Absorption

The dietary requirements for molybdenum are not known. It is present in a variety of foodstuffs and is adequately absorbed (60–70%) in the small intestine. Therefore, deficiency of Mo is rarely seen in adults in natural conditions. However, among infants with pertinent genetic lesions, the deficiency may occur, resulting in neurological pathologies.

Toxicity

Excess consumption of molybdenum may cause toxic manifestations, such as impairment in growth, diarrhoea and anaemia. The condition is called **molybdenosis.** Copper deficiency may also develop in this condition because molybdenum is known to decrease mobilization and utilization of copper in the body.

Cobalt

The only known role of cobalt is that it is a **component of vitamin B_{12}**. Inorganic cobalt is not absorbed from the alimentary tract, so it must be provided in the diet as vitamin B_{12}.

In high doses for a prolonged period it may produce polycythaemia (increased RBC count), probably through stimulating synthesis of erythropoietin.

Cobalt as a constituent of vitamin B_{12} is involved in folate metabolism and erythropoiesis.

Chromium

An average adult male contains about **6 mg** of chromium. Concentration in the blood is 20 μg/dl. It plays important role in metabolism, but excess chromium can be toxic.

Biochemical Functions

1. Chromium is a component of "***glucose tolerance factor***", a poorly characterized complex of trivalent chromium, nicotinic acid and glutathione (MW 1500), which *potentiates action of insulin* on its target tissues. Brewer's yeast is a good source of the glucose tolerance factor.
2. Chromium is essential for the normal metabolism of carbohydrates and lipids. It lowers the total serum cholesterol and the LDL levels, and elevates the HDL level, thereby providing health benefits.
3. It is postulated that chromium binds to DNA, RNA and nuclear proteins, and is thereby involved in maintenance of the structural integrity of the nuclear strands and in the *regulation of gene expression*.
4. Amino acid transport into the cells (liver and heart) may require participation of chromium.

Absorption

An average adult consumes about 10–100 μg of chromium each day. But the intestinal absorption is sluggish: less than 1% of the ingested element is absorbed. Stainless steel utensils contain chromium which can be readily absorbed.

Requirement and Dietary Sources

Requirements for chromium are not clearly known. It is present in limited food materials, e.g. cereals, grains, brewer's yeast, cheese, and meat.

Deficiency

Deficiency of chromium produces an abnormal glucose tolerance curve resembling that of insulin-dependent diabetes. Glycosuria, elevated blood glucose, glucose intolerance, and elevated circulating insulin are the important features of chromium deficiency.

Manganese

The total manganese content in adult humans is **12–20 mg,** of which 25% is in skeleton. Liver, pancreas and kidneys also contain relatively higher amounts of this element. It is found mainly in the nuclei, where it is believed to impart stability to the nucleic acid structure.

Biochemical Functions

1. Manganese is required for **glycoprotein** and **proteoglycan synthesis**, and so plays important role in formation of the matrix of bones and cartilages.
2. Manganese **stimulates** the activity of **several enzymes**, such as *glucosyl transferases, arginase, RNA polymerase, pyruvate carboxylase* and *superoxide dismutase* (mitochondrial). However, in most cases it can be replaced by magnesium.
3. Normal functioning of the central nervous system, physical growth and development of reproductive functions requires manganese.

Requirements, Dietary Sources and Absorption

A daily intake of **2.5–5.0 mg** is recommended for an adult, though exact requirement is not known. Cereals, leafy vegetables, fruits and tea are good sources of manganese. Absorption occurs in the small intestine (3–4% absorbed), and is inhibited by iron and the other divalent cations.

Deficiency

Growth retardation, bone deformity, sterility and fatty infiltration in hepatocytes are some of the manifestations of manganese deficiency. Diminished activity of the β-pancreatic cells results in decreased insulin secretion.

Toxicity

Excess manganese is excreted through bile and pancreatic secretions; only a small amount is eliminated in urine. Excessive accumulation in body is toxic, causing psychosis and parkinsonism ("**manganese madness**").

Iodine

Iodine, atomic weight 127, is a group VII halogens in the periodic table. Although it is widely distributed throughout the earth's surface, the sea is the major source of this element. The iodide concentration of sea water is 50 μg/L, which is approximately similar to that of human serum. In human body, iodide is present in small amounts (45–50 mg), of which 10–15 mg is present in thyroid gland, 25 mg in muscles, 5 mg in skin, 3 mg in skeleton and 2 mg in liver.

Biochemical Functions

Iodine is a constituent of the hormones secreted by thyroid gland, **thyroxine** (T_4) and **triiodothyronine** (T_3).

These hormones are synthesized by iodination of tyrosine and are essential for healthy growth, differentiation, and development (Chapter 30).

Requirement and Dietary Sources

Daily iodine requirement varies with age as here.

Infants	40–50 µg
Children	70–90 µg
Adults	100–150 µg
Pregnancy	175 µg
Lactation	200 mg

Uptake of iodide by foodstuffs is directly proportional to the soil and water content of this element. *Marine fish, lobsters and seaweed*, growing in iodine rich sea-water, are therefore the best sources. Drinking water is also an important source. Prolonged consumption of crops growing in areas where soil is depleted of iodide leads to iodide deficiency.

Absorption Transport and Excretion

Iodine is absorbed mainly from the small intestine. Normally, about 30% of the dietary iodine is absorbed. Skin may also absorb iodine, so toxic manifestations may result from prolonged use of iodine containing skin ointments. The absorbed iodine is released into blood circulation where it is mostly (90%) present in the form of thyroid hormones bound to protein (**protein bound iodine**). Only about 10% is present in the form of inorganic iodide. Excretion of iodine occurs mainly through urine, and to a lesser extent through bile and saliva. Urine inorganic iodide correlates with plasma level.

Deficiency

Iodide deficiency is an important cause of hypothyroidism in certain areas of the world, where the soil and the plants grown in it, are deficient in this mineral. Such areas constitute the "**goiter belt**" and the sub-Himalayan regions are a part of this belt.

The follicular cells of thyroid gland undergo hypertrophy and hyperplasia in iodine deficiency because decreased thyroid hormone synthesis leads to over-stimulation of the gland by TSH (TSH release from the anterior pituitary is dis-inhibited whenever the levels of circulating thyroid hormones are abnormally low). This phenomenon is known as **simple goiter** (or diffuse goiter).

Use of **iodized salt** (potassium iodate mixed in common salt in the ratio of 1:10,000 to 20,000) in the goiter belt has considerably reduced the incidence of simple goiter.

Toxicity

Prolonged excess of iodine intake (more than 2 mg/day) results in **iodine goiter** and **myxoedema**.

Fluoride

Like iodine, fluorine (atomic weight 19) also belongs to group VII halogens of the periodic table of elements. Only about 2.5 mg of this element is present in an average adult body, of which, more than 95% is present in bones and teeth.

Biochemical Functions

1. Fluorine makes the teeth resistant to cavities (dental caries) and bones to development of osteoporosis in later life by improving quality of hydroxyapatite crystals in teeth and bones. The fluoride anion may substitute for the hydroxyl ion in the hydroxyapatite crystal structure to produce a "harder" crystal. This is believed to account for the protective effect against dental caries (cavities), caused by certain bacteria, which are normal inhabitants of the oral cavity, and which act on dietary carbohydrates and convert them into lactic acid. The latter corrodes the enamel of the teeth and produces cavities.
2. Hardening of hydroxyapatite by fluoride also explains its therapeutic use, either alone or in combination with vitamin D, in therapy of osteoporosis.
3. Fluoride is a potent inhibitor of activities of certain enzymes. Fluoroacetate inhibits *aconitase* (of TCA cycle) and sodium fluoride inhibits *enolase* (of glycolysis).

Fluoride strengthens teeth and bones by getting incorporated in their crystalline structure.

Requirements and Dietary Sources

The daily requirement of fluoride is 1–2 mg per day. In warm Indian climate, where the water intake is usually high, the drinking water, having a fluo-ride content of 0.5–0.8 ppm easily meets this meager daily requirement (ppm = parts per million; 1ppm = 1 gram of fluoride in million gram of water, which equals 1 mg/1000 ml).

Traces of fluoride are present in most foods; seafood and cheese are particularly rich. Drinking water is also a major source; fluoride can be present in drinking water, either naturally or because of artificial supplementation. In fact, a directly inverse association between the incidence of dental caries and fluoride concentration of < 1 mg/dL in drinking water has long been recognized.

Deficiency and Toxicity

Decreased fluoride intake is associated with **dental caries** and excessive intake with **fluorosis.** When drinking water contains less than 0.5 ppm of fluoride, dental caries results.

A prolonged high intake of fluoride is potentially hazardous. Fluorine levels more than 5 ppm cause **dental fluorosis** characterized by discoloured and mottled

teeth. In more severe cases (fluorine levels > 20 ppm) bones are also affected, leading to alternate areas of osteoporosis and osteosclerosis and calcification of tendons and ligaments. The condition is called **skeletal fluorosis.** Fluorosis is mostly seen in geographical areas where water content of this mineral is high.

Iodine is an important component of thyroid hormones. Fluorine is required for preventing dental carries, but in larger amounts causes mottling of teeth and skeletal abnormalities (fluorosis).

Exercises

Essay type questions

1. Describe metabolism of copper and zinc. Discuss metabolic importance of ceruloplasmin.
2. Discuss biochemical significance and disease states associated with trace elements.
3. Describe intestinal absorption of iron and various factors involved in its regulation.
4. What is the nutritional role of selenium? How do vitamin E and selenium supplement each other?
5. What are the biochemical roles of trace elements in humans. Illustrate your answer with two examples.

Write short notes on

1. Wilson's disease
2. Factors regulating iron absorption
3. Haemochromatosis
4. Antioxidant role of selenium
5. Ceruloplasmin
6. Menkes' disease
7. Fluorosis
8. Iron deficiency
9. Iron compounds and their importance
10. Iron homeostasis in body

CLINICAL CASES

CASE 19.1 A 42-year-old woman with palpitations and breathlessness

A 42-year-old woman visited general practitioner because of lethargy, dizziness, and excessive tiredness. In the past few months her physical performance had significantly decreased. She was out of breath with little walking, and felt as if her heart was thumping against chest. A school teacher by profession, she was deeply religious and had always stuck to a restricted, vegetarian diet. For the past few weeks she had been feeling that the food stuck to her throat after meals. Apparently she did not suffer from any nutritional deficiency. On examination, however, she was clinically anaemic, and so the general practitioner collected her blood sample for analysis. The results are as below:

Test	Patient's report	Reference range
Haemoglobin	8.2 g/L	12–15 g/L
Blood glucose (random)	96 mg/dL	< 140 mg/dL
Serum proteins	7.6 g/L	6–8 g/L
Red cell count	4.0×10^{12}/L	$3.8–5.8 \times 10^{12}$/L

Blood smear showed marginally diminished numbers of red cells that appeared microcytic (smaller size) and hypochromic (abnormally pale). In view of these findings the general practitioner urgently referred the patient to the teaching hospital for complete assessment and treatment.

The attending Medical Officer in the hospital found that the patient had brittle hair, thinning and spooning of the nails (koilonychia), sore tongue, angular stomatitis and inelastic skin. She was referred to the haematology section, where a blood sample was obtained for detailed analysis. In addition, fecal sample was sent for occult blood analysis. A bone marrow sample was stained for iron. A tentative diagnosis was made based on the reports of the haematology parameters.

The patient was referred to the gynecologist also because she suffered from menstrual irregularities and had heavy periods for the past several months.

Meanwhile other test results were obtained. Test for occult blood was negative, serum ferritin level was low (6.8 ng/mL) and the bone marrow sample, when stained for iron, showed minimum staining.

Q.1. What is the condition affecting this patient? Give reason for decreased haemoglobin level.
Q.2. In view of clinical history and test results, give an explanation for iron deficiency in this patient.
Q.3. Outline the treatment for this patient.
Q.4. Iron therapy, if prolonged, is hazardous. Give reasons.
Q.5. Supposing a patient with similar clinical and blood picture fails to respond to iron therapy but shows dramatic improvement with copper. What is the possible explanation for this observation?

CASE 19.2 A 13-year-old boy with slowly progressing dysarthria

A 13-year-old boy was referred to the neurological OPD with complaints of slowly progressing dysarthria, tiredness, abdominal pain and backache. The abdominal pain had developed about one month ago, and it was still persisting, without much remission. Prior to this, he was well although he was finding it difficult to keep up with the classwork. He had a younger sister who is apparently well. However, his mother told the neurologist that her elder daughter had died in her thirties from a liver problem. While taking history, it was noticed that the child often lost control over his emotions.

On examination, a yellow colouration was noticed in sclera; the urine was dark coloured and stools were pale. A moderate enlargement and tenderness of liver was detected. X-Ray abdomen and spine were normal. Liver function tests were suggestive of hepatocellular damage (i.e. elevated *transaminases*, increased serum bilirubin levels, both of direct and indirect types, and urobilirubinuria). Provisional diagnosis of hepatitis was made but serum sample was negative for HBs antigens as also for antibodies to HBs (i.e. anti-HBs antibodies).

Ophthalmic (slit-lamp) examination showed presence of greenish-brown discolouration of the corneal margins. The case was referred to ophthalmologist, who identified these as the Kayser–Fleischer rings.

Based on this report, tentative diagnosis was made, and blood and urine samples were obtained for further investigations. The results are as below:

Test	Patient's report	Refirence range
Serum copper	90 μg/100 mL	100–128 μg/100 mL
Serum ceruloplasmin	16 mg/100 mL	25–45 mg/100 mL

Analysis of a 24-hour urine sample was reported as showing moderate generalized aminoaciduria. The child's condition worsened, and soon he died of fulminant hepatic failure. Liver copper content was estimated, and was found grossly elevated. Liver copper = 1886 μg/g dry weight (mean normal = 35 μg/g dry weight; range 20–5 μg/g).

Q.1. Suggest a likely diagnosis.
Q.2. What is the biochemical defect in this disorder?
Q.3. Explain the biochemical basis of the signs and symptoms of this patient.
Q.4. Would you advise any biochemical investigations for the younger sister of this patient?
Q.5. Outline the treatment for this patient.

CASE 19.3 A 34-year-old man with weight gain and persistent tiredness

This case refers to a sub-Himalayan rural area of Nepal where iodide deficiencies are common. Moreover, the soil in these regions is poor in selenium. A 34-year-old school teacher reported in the Primary Health Centre (PHC) after he developed the following symptoms: loss of appetite, constipation, tiredness and weight gain. Lately, he had been feeling very weak, and found it tiresome even to walk to the school which was less than a kilometer away from his house. No abnormality was detected upon physical examination. Analysis of blood and urine sample for routine biochemical and haematological parameters (including haemogram, ESR, blood, and urine sugar) also did not show any abnormal result.

The patient was referred to the teaching hospital in the nearby town where the following biochemical investigations were done:

Investigations test	Patient's report	Reference range
Serum triiodothyronine (T_3)	56 ng/100 mL	70–180 ng/100 mL
Serum thyroxine (T_4)	7.2 μg/dL	5.5–12 μg/dL
Serum cholesterol	322 mg/dL	156–255 mg/dL

These results were suggestive of hypothyroidism. The patient was taking iodized salt for several years. This appeared to rule out iodine deficiency (which is so common in these regions) as the likely cause of hypothyroidism.

Q.1. Could hypothyroidism in this patient be related to selenium deficiency? Explain why?
Q.2. State the mechanism by which deficiency of the selenium cofactor affects activity of the enzyme, *deiodinase*?

NUCLEIC ACID CHEMISTRY AND NUCLEOTIDE METABOLISM

CHAPTER

20

Nucleotides are structural units of nucleic acids consisting of a pentose sugar, a nitrogenous base and one or more phosphate groups. They are widely distributed and perform a variety of functions of vital significance. Covalently linked nucleotide monomers form the nucleic acids, *DNA* and *RNA*. Most nucleic acids are very large. They often contain thousands (in case of DNA, millions) of nucleotide units held together by phosphodiester bonds. Nucleotides play an important role in **biosynthetic reactions** and in **energy metabolism.**

This chapter deals with important structural features and biological significance of nucleotides and nucleic acids. After going through this chapter, the student should be able to understand:

- *Chemistry*, *properties* and *functions* of purine and pyrimidine bases, nucleosides and nucleotides.
- The ***de novo*** synthesis and ***salvage pathways*** of purine nucleotides, and the purine degradation pathway; regulation of these pathways, the positive and negative effectors, enzymes controlling rate of metabolism, and the attendant pathologies.
- Pyrimidine synthesis and degradation; aetiology of various disorders related to these pathways; steps (of the above stated purine and pyrimidine pathways) inhibited by **antineoplastic drugs** and the sulpha drugs, and the mechanisms of inhibition.

I. Nucleotides: Chemistry and Biological Significance

A. Basic Structure

A nucleotide has three structural components: a heterocyclic compound termed **nitrogenous base**, a five carbon **(pentose) sugar**, and at least one **phosphate group**. The nitrogenous bases found in nucleic acids belong to one of the heterocyclic groups, either purines or pyrimidines. When the nitrogenous bases are combined with a pentose sugar, they are known as ***nucleosides***. Phosphate group can be attached either at the 5′-position or the 3′-position of the pentose, and the nucleoside-phosphate thus formed is known as ***nucleotide***.

Nucleoside = Base + Sugar
Nucleotide = Base + Sugar + Phosphate(s)

Nucleic acids are polymers of several nucleotide units, hold together by covalent linkages.

Nitrogenous Bases

The bases are derivatives of purines or pyrimidines. Structures of the parent compounds of **purine** and **pyrimidine** bases (which do not occur in nature) are shown in Figure 20.1, with the atoms numbered. Purines are numbered in anti-clockwise direction, and the pyrimidines are numbered in the clockwise direction.

Major Bases

The major purines of both DNA and RNA are guanine (G) and adenine (A). In DNA, the major pyrimidines are thymine (T) and cytosine (C), while in RNA the major pyrimidines are uracil (U) and cytosine (C). *Presence of T instead of U in DNA is important for preventing mutations* (Chapter 21). Structures of the major bases are depicted in Figure 20.1.

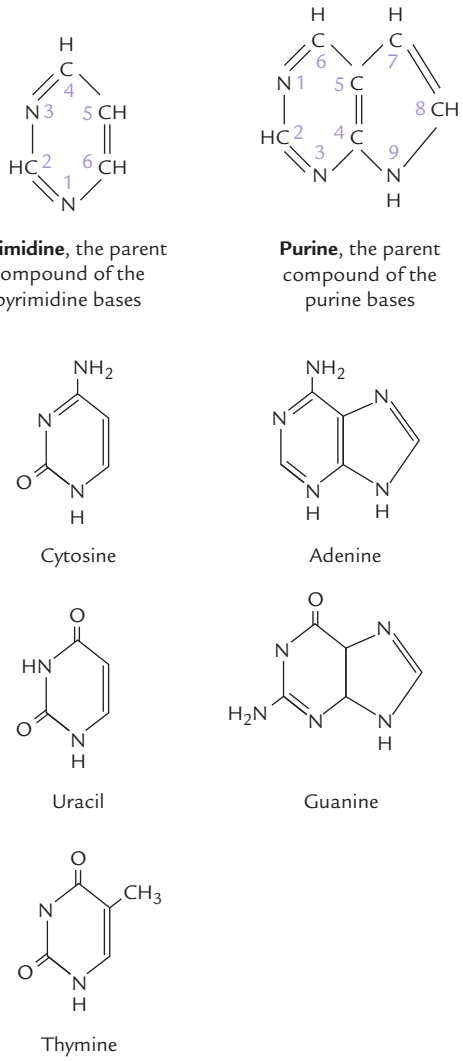


Fig. 20.1. The five major purine and pyrimidine bases of nucleic acids (DNA and RNA).

There are five major bases:

- Adenine, Guanine (purines)
- Cytosine, Uracil, Thymine (pyrimidines)

Thymine is present only in DNA, and uracil only in RNA.

Other Purine and Pyrimidine Bases

In addition to the five major purine and pyrimidine bases in nucleic acids, some minor or unusual bases are also found in nature. *Hypoxanthine*, *xanthine* and *uric acid* (Fig. 20.2) are present in **free state** intracellularly. The former two are the metabolites of adenine and guanine, which are ultimately converted to uric acid—the end product of purine catabolism.

Some minor or **unusual bases** are found in nucleic acids, mostly in RNA. These include *5′-methylcytosine, pseudouracil, N^6-methyladenine* and *N^7-methylguanine* (Fig. 22.5, Chapter 22). Finally, some methylated purine bases are found **in plants**. These include alkaloids like *theophylline* (1,3-dimethylxanthine), *caffeine* (1,3,5-trimethylxanthine) and *theobromine* (3,7-dimethylxanthine) (Fig. 20.3). Theophylline is found in tea, and is used therapeutically in asthma; caffeine is a component of coffee beans and tea, that elevates intracellular cAMP level; and theobromine is found in chocolate, and is a diuretic, heart stimulant and vasodilator.

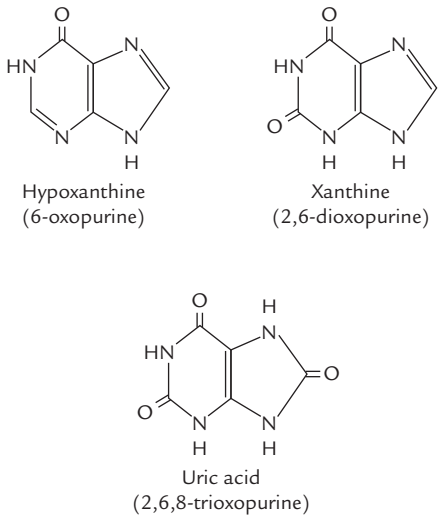


Fig. 20.2. Structures of some minor purine bases.

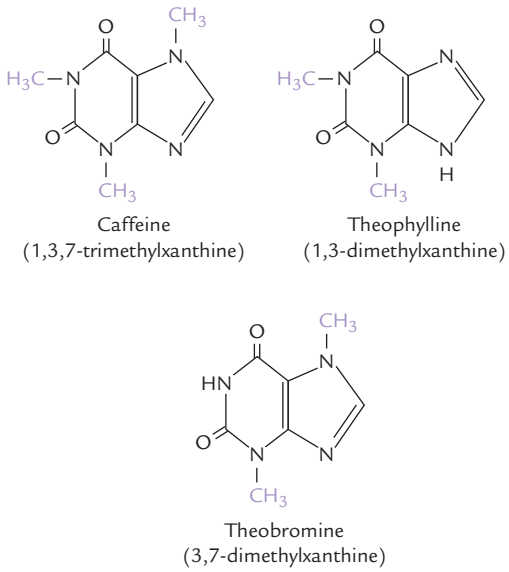
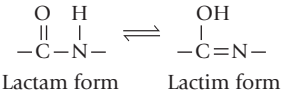


Fig. 20.3. Structures of some methylated bases present in plants.

Chemical Properties of Bases

Nitrogenous bases include highly conjugated double bond systems within the ring structures. For this reason, nucleic acids have a very strong *absorption maximum* at about **260 nm**, which is used for nucleic acid quantitation. Moreover, the heterocyclic rings of both purines

and pyrimidines have oxo (−C=O−) functional group. Therefore, a given base can exist in **two tautomeric forms**, the *lactam* or keto form and the *lactim* or enol form.



Lactam form ⇌ Lactim form

In DNA and RNA, the *keto forms are by far more predominant*, and this property makes it possible for the bases to form intermolecular hydrogen bonds (Chapter 21).

Sugars

The 5-carbon sugar found in nucleotides is either *D-ribose* or *D-2′-deoxyribose*. Each occurs in furanose form, the configuration at C-1 being β (see Chapter 2 for nomenclature). **D-Ribose** is present in **RNA**, while **D-2′-deoxyribose** is present in **DNA**. Deoxyribose, as the name suggests, differs from ribose in having one oxygen less (at C-2).

Nucleosides

A nucleoside is a compound of **base** and **sugar**; the latter two are linked by an *N-glycosidic linkage* (a carbon-nitrogen bond). The bond is formed between C-1 of sugar and nitrogen atom 1 of pyrimidine base or nitrogen atom 9 of purine base.

If the sugar is *ribose,* the compound is **ribonucleoside** and if the sugar is *deoxyribose*, the compound is **deoxyribonucleoside**. Adenosine (adenine + ribose), guanosine (guanine + ribose), cytidine (cytosine + ribose) and uridine (uracil + ribose) are the ribonucleosides of A, G, C and U, respectively. Similarly, deoxyadenosine, deoxyguanosine, deoxycytidine and deoxythymidine are the deoxyribonucleosides of A, G, C and T, respectively.

Base	Ribonucleoside	Deoxyribonucleoside
Adenine (A)	Adenosine	Deoxyadenosine
Guanine (G)	Guanosine	Deoxyguanosine
Cytosine (C)	Cytidine	Deoxycytidine
Uracil (U)	Uridine	–
Thymine (T)	–	Deoxythymidine

Nucleotides

As already stated, **nucleotide is nucleoside plus phosphate**. The phosphate group is esterified to one of the hydroxyl groups of the sugar component of the nucleoside. Mostly it is the hydroxyl group at the C-5 (less commonly C-3) that is phosphorylated.

To avoid confusion between the numbering of various atoms in the base and carbon atoms of the sugar, the latter are represented with an associated prime. Thus, carbons of sugar are numbered 1′, 2′, 3′, etc.

Nucleotide comprises a nitrogenous base, a sugar (ribose or deoxyribose) and one or more phosphate groups. Base and sugar are linked by an N-glycosidic linkage. Phosphate is linked mostly to C-5 of sugar.

B. Nomenclature

The nucleotides are named as phosphate derivatives of nucleosides. Thus, adenosine linked covalently with phosphate produces adenosine monophosphate (AMP). It is a *mononucleotide*. Likewise, other mononucleotides, e.g. guanosine monophosphate (GMP), cytidine monophosphate (CMP), and uridine monophosphate (UMP) are formed by phosphorylation of guanosine, cytidine and uridine, respectively.

The phosphate group can be attached either at the 5′-position or the 3′-position of the pentose, and accordingly the mononucleotides formed are designated as 5′ or 3′-monophosphates, respectively. However, attachment at the 5′-position is far more common, and therefore 5′ is usually omitted in the abbreviation. Thus, AMP means adenosine 5′-monophosphate, CMP means cytidine 5′-monophosphate, and so on. However, for adenosine3′-monophosphateorcytidine3′-monophosphate, the abbreviations 3′-AMP or 3′-CMP respectively are used.

A mononucleotide is also called *nucleoside monophosphate*.

Nucleoside-Diphosphate and Triphosphate

Additional phosphate groups may be attached to a mononucleotide. Addition of a second and a third phosphate to a nucleoside monophosphate yields a nucleoside-diphosphate, and triphosphate, respectively. AMP thus yields adenosine diphosphate (ADP) or adenosine triphosphate (ATP), respectively (Fig. 20.4). The subsequent phosphate groups are attached to the preceding ones by acid anhydride bonds. These bonds possess **high free energy of hydrolysis**, yielding more than 7 kcal/mole of free energy (Chapter 8).

C. Functions

Nucleotides are central to maintenance and propagation of life. They are most versatile of all biomolecules, being involved in a number of reactions including the energy transfer reactions.

Some important functions of nucleotides are:

1. The ribonucleotides such as ATP, GTP, CTP and UTP are important **coenzymes** and provide the **basic unit of RNA**. The deoxyribonucleotides, dATP, dGTP, dCTP and dTTP are required for **DNA replication** and **DNA repair**.

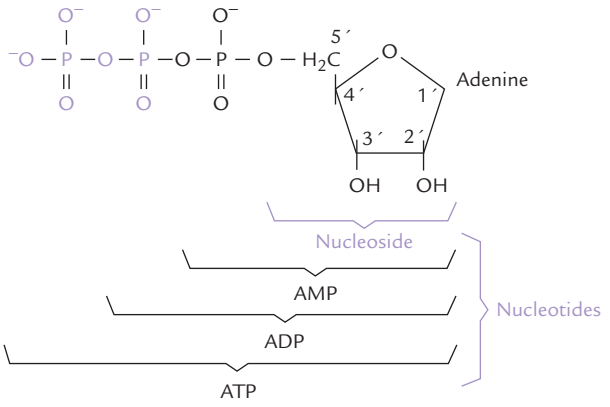


Fig. 20.4. Nucleoside and nucleotide structures.

The ribonucleotides are present in millimolar concentrations in the cell, and the deoxyribonucleotides are present in micromolar concentrations intracellularly.

2. Nucleoside sugars are activated precursors which are used in **biosynthetic reactions**. For example:
   - UDP-glucose is used in glycogen synthesis
   - UDP-galactose is used in synthesis of ceramides
   - CTP-choline is used in phospholipid synthesis.
3. **Nucleotides are the basis of high-energy compounds**. The best known example is adenosine triphosphate (ATP), referred to as the **currency of free energy** in the body, which provides energy for all types of cellular activities. GTP is used as an energy source in protein synthesis.
4. Regulation of enzyme activity involves certain nucleotides, such as cAMP and cGMP. They are signal-conducting molecules, acting as **intracellular messengers** for certain hormones (Chapter 29).
5. The nucleic acids, DNA and RNA, are involved in **storage and decoding of genetic information**.
6. The **coenzymes**, NAD^+, FAD and coenzyme A have nucleotides as essential components of their structures.

Although dietary nucleotides and nucleic acids are digested to nucleosides and free bases, the degradation products are poorly absorbed. The small amounts of absorbed products are not delivered to tissues but rather degraded within intestinal mucosa to form uric acid. Thus, ***dietary nucleic acids are not used for the synthesis of tissue nucleic acids***.

II. Purine Metabolism

A. Synthesis of Purine Nucleotides

Human body is capable of synthesizing the purine and the pyrimidine rings **de novo** *(anew)* and also by **salvaging** *(recycling)* the nitrogenous bases arising from degradation products of nucleic acids.

- In **de novo synthesis** the elements of purine ring system are added step by step, using C-1 of ribose 5-phosphate as a primer. It provides the cell with the capacity to *construct the purine ring afresh*, with carbon and nitrogen atoms coming from various sources.
- In the **salvage pathways**, the free purine bases or nucleosides (arising from degradation of preexisting nucleic acid bases) are **reutilized** for the synthesis of new nucleotides. The salvage reactions reduce the requirement for the energetically expensive de novo biosynthesis and prevent wastage of raw materials. They are the principal source of nucleotides for many tissues, most notably **brain**, in which the de novo pathway is either poorly developed or absent.

Brain, RBCs and polymorphonuclear leucocytes cannot produce purines by de novo synthesis.

The nucleotides arising from endogenous synthesis are sufficient to meet the body's requirements. Therefore, humans are said to be **prototrophic** for purines and pyrimidines for not being dependent upon dietary sources.

The de novo Pathway

The de novo synthesis of purines is most active in *liver*, though it may occur in several other tissues. Liver also exports the surplus bases and nucleosides to other tissues. All reactions of the de novo pathway are *cytosolic*.

The first clue to the de novo synthesis was provided in 1948 by *John Buchman*. By feeding of a variety of isotopically labelled compounds to pigeons and chemically determining the positions of the labelled atoms in their excreted uric acid (a purine), it was demonstrated that various atoms of purine ring arise from different sources. Many compounds contribute to purine ring (Fig. 20.5).

- N-1 arises from the amino group of *aspartate*;
- C-2 and C-8 originate from a single carbon moiety carried by *tetrahydrofolate*;
- N-3 and N-9 are contributed by the amide group of *glutamine*;
- C-6 comes directly from CO_2.
- C-4; C-5 and N-7 are derived from *glycine* (suggesting that glycine is wholly incorporated into the purine ring).

Purines are built on foundation of ribose 5′-phosphate. The heterocyclic ring systems are assembled from simple precursors: glycine, glutamine, aspartate, formate and carbon dioxide.

The initially synthesized purine derivative is inosine monophosphate **(IMP)**, the nucleotide of the base hypoxanthine. It is loosely referred to as the **parent nucleotide**. It can then serve as precursor for both AMP and GMP.

The de novo synthesis occurs in two stages:

1. Synthesis of IMP, the parent nucleotide.
2. Conversion of IMP to AMP and GMP.

Stage I: Synthesis of IMP: It involves a series of eleven reactions (Fig. 20.6) described below:

Reaction 1

Purine ring is built directly onto ribose 5′-phosphate, which is first converted to 5′-phosphoribosyl pyrophosphate (PRPP) by the enzyme *PRPP synthetase* (Fig. 20.6). The pyrophosphate group that gets attached to the ribose phosphate is provided by ATP. PRPP is also a precursor in the biosynthesis of pyrimidine nucleotides and the amino acids, histidine and tryptophan. Thus, *PRPP lies at crucial biosynthetic crossroads and so its intracellular concentration is subject to meticulous regulation*.

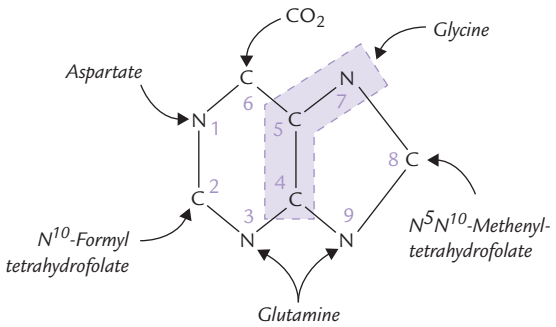


Fig. 20.5. The purine ring: origin of constituent atoms.

Reaction 2

Displacement of the pyrophosphate group (of PRPP) by amide nitrogen of glutamine occurs next. To form 5-phosphoribosylamine; the reaction is catalyzed by the enzyme *PRPP amidotransferase*. This is the first reaction unique to purine biosynthesis (i.e. **committed step)** and is also ***rate-limiting*** for the pathway.

Reactions 3–11

A series of nine reactions that follow construct the parent nucleotide, inosine monophosphate (IMP); these are shown in Figures 20.7 and 20.8. Some of these reactions are energy demanding: *ATP is required at each of the synthetase and ring closure steps*. A summary of these reactions is given here.

- Two carbons and a nitrogen are added in a single step, when glycine is attached to amine group of 5-phosphoribosylamine (**Reaction 3**). The reaction, catalyzed by the ATP-dependent enzyme, *GAR synthetase*, yields glycinamide ribonucleotide (GAR).
- A formyl (–CHO) group is then attached to the GAR to form formylglycinamide ribonucleotide (FGAR); the formyl group comes from N^{10}-formyl tetrahydrofolate (**Reaction 4**).
- In **Reaction 5**, another amino group is added onto FGAR to form formylglycinamidine ribonucleotide (FGAM). As for other amino group additions, the donor is glutamine.

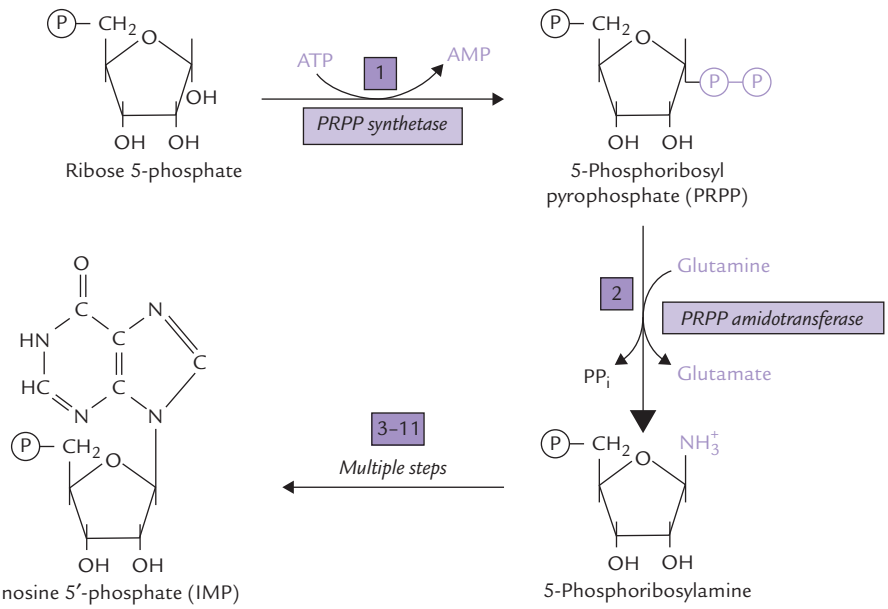


Fig. 20.6. Biosynthesis of IMP, the parent purine nucleotide, by the de novo pathway (reactions 1–11).

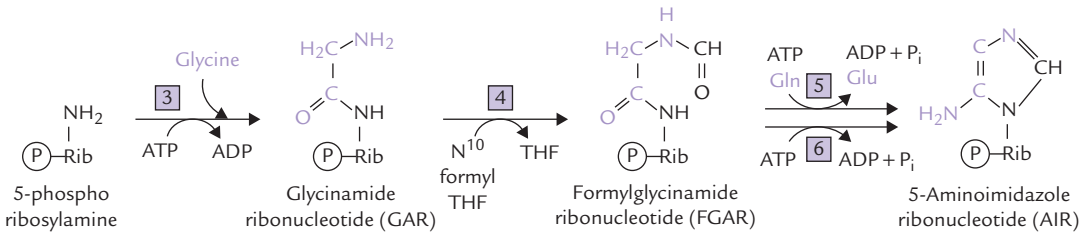


Fig. 20.7. Expansion of reactions 3–6 of IMP biosynthesis. The numbers also correspond to the enzymes for various reactions (3 = *GAR synthetase*, 4 = *GAR transformylase*, 5 = *FGAM synthetase*, 6 = *AIR synthetase*).

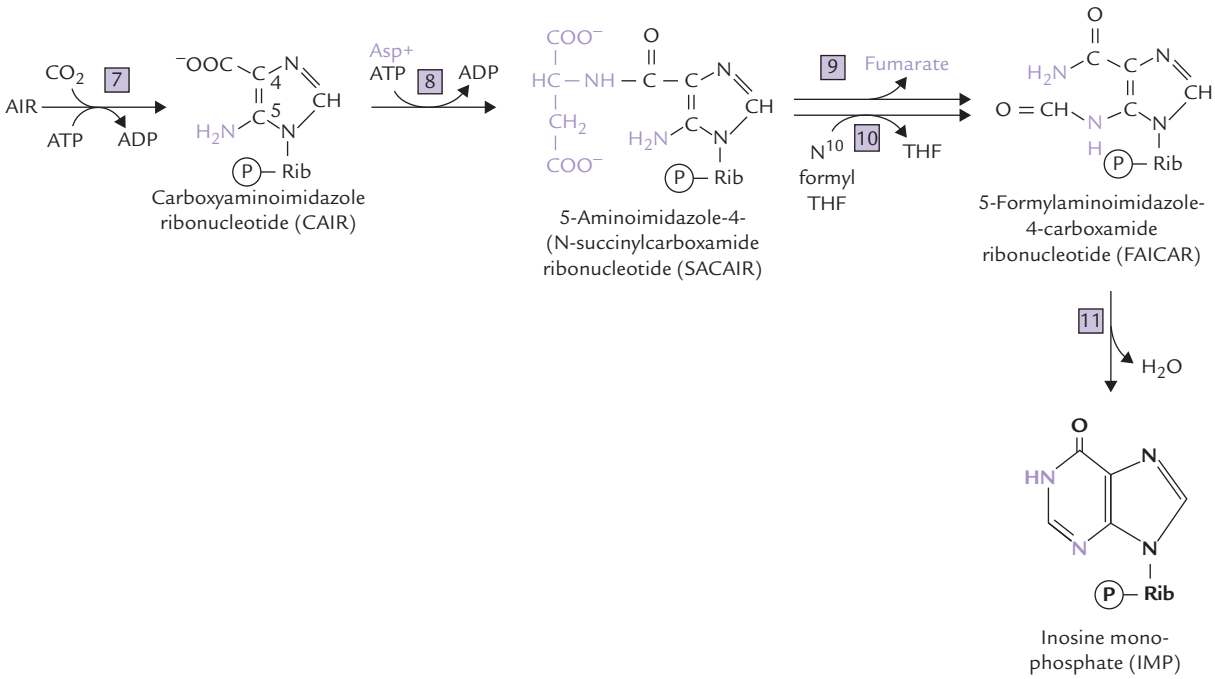


Fig. 20.8. Expansion of reactions 7–11 of IMP biosynthesis. The numbers also correspond to the enzymes for various reactions (7 = *AIR carboxylase*, 8 = *SACAIR synthetase*, 9 = *Adenylosuccinate lyase*, 10 = *AICAR transformyltase*, 11 = *IMP synthase*).

- The imidazole ring of the purine is then closed in an ATP-dependent reaction to yield 5-aminoimidazole ribonucleotide (AIR) by action of *AIR synthetase* (**Reaction 6**).

The second ring is now built by stepwise addition of the final three atoms (*Fig. 20.8*).

- The first atom to be added is the C-6 carbon (**Reaction 7**). It comes from CO_2 which is incorporated to form carboxyaminoimidazole ribonucleotide (CAIR). The reaction is an unusual one because the carbon atom arises directly from CO2, without involvement of biotin.
- The nitrogen atom at position 1 of the mature purine ring is now added in two steps (**Reactions 8 and 9**). The first step is the formation of an aspartate-CAIR covalent intermediate, termed succinylcarboxamide ribonucleotide (SACAIR), catalyzed by *SACAIR synthetase* (**Reaction 8**). The four-carbon dicarboxylic acid, fumarate is then cleaved off by the action of *adenylosuccinate lyase* and *only the amino group of aspartate is retained* to yield 5-aminoimidazole-4-carboxamide ribonucleotide (AICAR).
- The last purine ring carbon is added as a formyl group from N^{10}-formyl tetrahydrofolate, by *AICAR transformylase* (**Reaction 10**); the product is 5-formylaminoimidazole-4-carboxamide ribonucleotide (FAICAR). With this reaction, all the carbon and nitrogen atoms of the purine ring have been contributed by respective sources.
- In the final step (**Reaction 11**), *IMP synthase* catalyzes closure of the second purine ring to form IMP.

Finally, several reactions of the de novo pathway require input of ATP energy. One ATP each being required in reactions 1, 3, 5, 6, 7 and 8, a *total of* ***6 ATPs*** *are used up in synthesis of IMP*.

In animals, as *Mary Ellen Jones* demonstrated, a single polypeptide catalyzes reactions 3, 4 and 6. Likewise, a different polyprotein is required for each set of reactions 7 and 8 and 10 and 11. This arrangement increases the efficiency of the pathway.

The de novo pathway of purine synthesis is complex, consisting of 11 steps, and requiring six ATPs till the parent nucleotide, IMP. Transfer of an amide from glutamine to phosphoribosylpyrophosphate is the committed and rate-limiting step in purine synthesis.

Stage II: Conversion of IMP to AMP and GMP: IMP is readily converted to AMP and GMP by a two-step process in each case (Fig. 20.9).

For *IMP to AMP conversion*, aspartate's amino group is first linked to IMP to yield adenylosuccinate. This step is driven by hydrolysis of GTP. In the second step, *adenylosuccinate lyase* eliminates fumarate from adenylosuccinate to form AMP.

For *IMP to GMP conversion*, the C-2 of IMP is oxidized by *IMP dehydrogenase* (an NAD^+-dependent enzyme) to form xanthosine monophosphate (XMP). XMP is then converted to GMP by *GMP synthetase* using glutamine as the nitrogen donor. The reaction is powered by ATP hydrolysis.

6-Mercaptopurine inhibits conversion of IMP to AMP by inhibiting the enzyme *adenylsuccinate lyase* and to GMP by inhibiting *IMP dehydrogenase*.

Conversion of Nucleoside Monophosphate to Nucleoside Diphosphate and Triphosphate

The monophosphates are the forms synthesized de novo, but the di- and triphosphates are the most commonly used forms. Therefore, monophosphates must be converted to

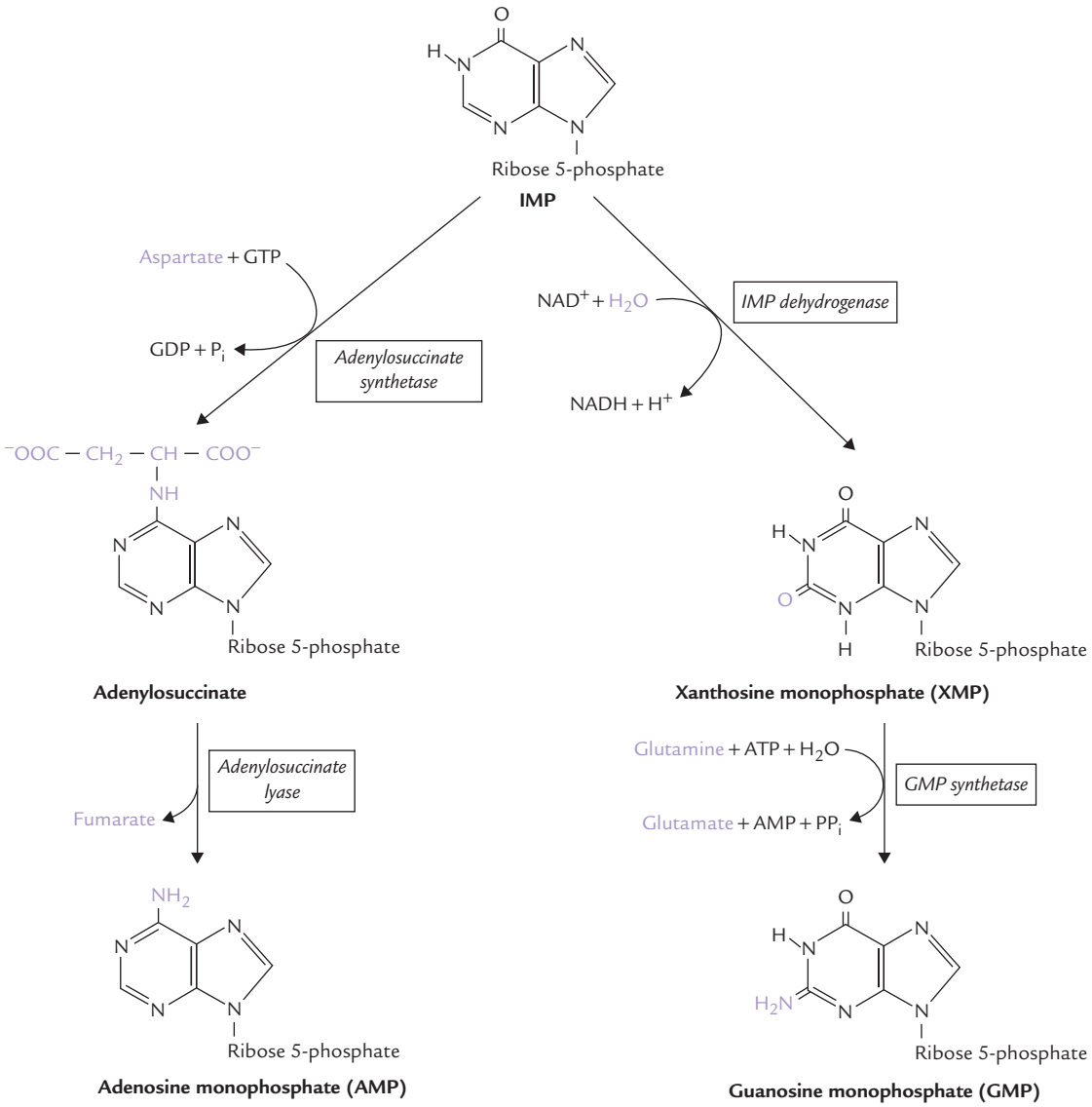
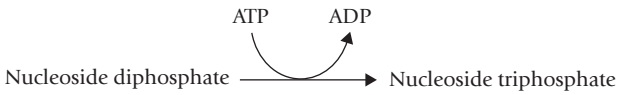


Fig. 20.9. Conversion of IMP to AMP or GMP in separate two-reaction pathways.

the corresponding di- and triphosphates, which is brought about by sequential action of the following enzymes:

(a) *The nucleoside monophosphate kinases* convert nucleoside monophosphates to the corresponding diphosphates. ATP is generally the source of energy, being present in relatively higher concentration intracellularly than the other nucleoside triphosphates. These enzymes are specific to nitrogenous base; for example, *adenylate kinase* converts AMP to ADP while GMP is converted to GDP by the action of *guanylate kinase*.

(b) Conversion of the nucleoside diphosphates to the corresponding triphosphates is catalyzed by a **single enzyme** called *nucleoside diphosphate kinase.* Evidently, this enzyme has *broad specificity.*



Conversion of Ribonucleotides to Deoxyribonucleotides

The compounds synthesized by the de novo pathway discussed so far are ribonucleotides, containing the ribose sugar. These can be used for RNA synthesis, but not for DNA synthesis. For synthesizing DNA, the deoxyribonucleotides are required, which differ from the ribonucleotides in having deoxyribose sugar instead of the ribose sugar. These compounds are synthesized by the enzyme *ribonucleotide reductase* (Fig. 20.10).

The ***ribonucleotide reductase*** is specific for a variety of nucleoside diphosphates: purines (ADP, GDP) as well as the pyrimidines (CDP and UDP), converting them to the corresponding deoxyribonucleoside diphosphates (dADP, dGDP, dCDP and dUDP). The enzyme requires a peptide coenzyme, *thioredoxin*, which is the donor of the reducing equivalents; the latter are required for reducing the ribonucleoside diphosphates (Fig. 20.10). During the process, the thioredoxin (reduced) is converted to the corresponding oxidized form. Subsequently, it is reconverted to the reduced form by the reducing equivalents of NADPH. Thus, in this series of reactions, NADPH is the ultimate donor of reducing equivalent.

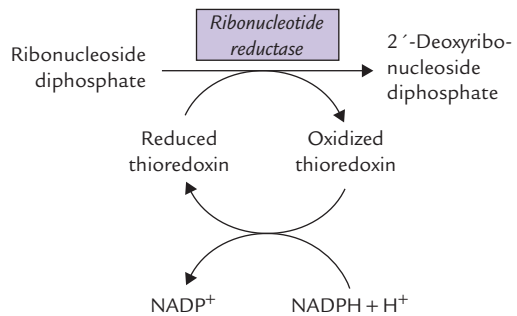


Fig. 20.10. Role of various components of the reaction system involved in reduction of ribonucleoside diphosphate to 2'-deoxyribonucleoside diphosphate.

The ribonucleotides are synthesized first, and they are the precursors of the corresponding 2′-deoxyribonucleotides, the substrates for DNA synthesis.

Regulation of the de novo Pathway

There are two important levels of regulation: **First** is the IMP pathway up to synthesis of IMP from ribose 5-phosphate, and the **second** is immediately below the branch point leading from IMP to ATP and GTP.

The first level regulation controls the total amount of purine nucleotides synthesized for nucleic acid synthesis, while the second one controls relative amount of adenine and guanine nucleotides synthesized. Thus, *an important feature of the regulatory network,* Figure 20.11, *is that the pathways synthesizing IMP, ATP and GTP are individually regulated* (Fig. 20.11).

The first level regulation: The principal regulatory enzymes of the IMP pathway up to IMP synthesis are the first two enzyme, i.e. *PRPP synthetase* and *PRPP amidotransferase*:

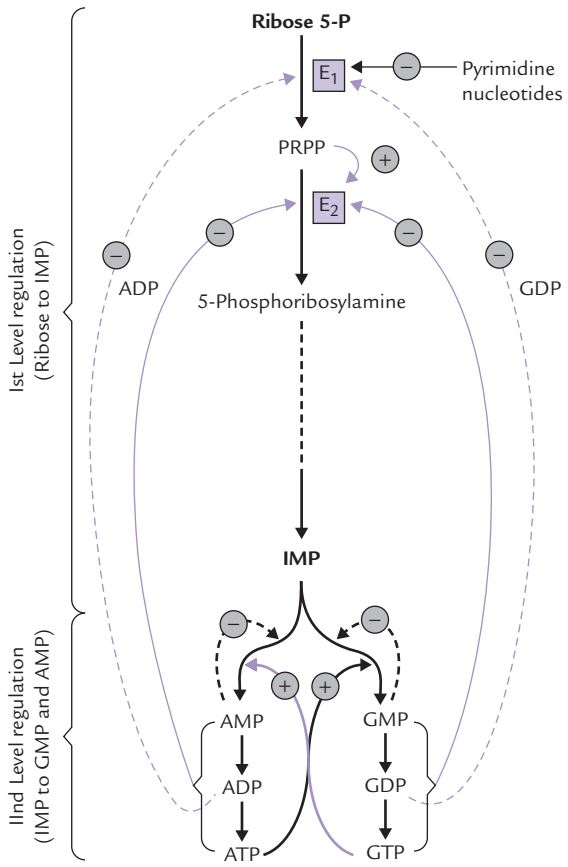


Fig. 20.11. Regulation of de novo purine biosynthesis (E_1 = *PRPP synthetase*, E_2 = *PRPP amidotransferase*, (–) = negative feedback, (+) = feed forward activation).

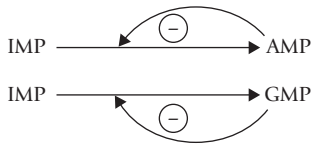
- *PRPP synthetase* (E_1) is inhibited by both ADP and GDP.
- *PRPP-amidotransferase* (E_2), the committed and the rate-controlling enzyme, is likewise subject to feedback inhibition. It is a dimeric protein to which ATP, ADP and AMP bind at one inhibitory site, and GTP, GDP and GMP at the other.

The rate of IMP production is, therefore, independently but synergistically controlled by the levels of adenine nucleotides and guanine nucleotides. Moreover, the enzyme activity is allosterically stimulated by **PRPP (*feed-forward activation*).** It has been reported that the enzyme is apparently inactive in dimeric form. Increased amounts of PRPP stimulate its dissociation into two subunits, which acquire catalytic activity.

The first two reactions of the biosynthetic pathway are feedback-inhibited by nucleotides. ADP and GDP inhibit the *PRPP amidotransferase*; whereas *PRPP synthetase* is inhibited by a number of nucleotides and is stimulated by PRPP allosterically.

The second level of regulation

- Feedback inhibitory mechanism is involved in conversion of IMP to AMP or GMP (Fig. 20.11).



Both AMP and GMP are competitive inhibitors of IMP in their own synthesis, which balances the production of each of these nucleotides (they are roughly required in equal amounts).

- The IMP to GMP conversion is stimulated by ATP, and of IMP to AMP conversion gets stimulated by GTP.

ATP is the energy source for guanosine monophosphate synthesis and GTP is the energy source for adenosine monophosphate synthesis. This reciprocity meticulously coordinates the rates of production of AMP and GMP and prevents excessive build-up of either.

The Salvage Pathways

Purine bases that are obtained from the normal turnover of cellular nucleic acids can be used to resynthesize the corresponding nucleotides. *Pathways of base to nucleotide conversion are referred to as the **salvage pathways***. A given base reacts with PRPP to yield the corresponding nucleotide; for example, guanine forms guanosine monophosphate (GMP).

$$\text{Guanine} + \text{PRPP} \xrightarrow{HGPRT} \text{GMP} + PP_i$$

The enzyme catalyzing this step is *hypoxanthine guanine phosphoribosyl transferase (HGPRT)*. This enzyme also converts another base, hypoxanthine, to its corresponding nucleotide, inosine monophosphate (IMP).

$$\text{Hypoxanthine} + \text{PRPP} \xrightarrow{HGPRT} \text{IMP} + PP_i$$

Adenine phosphoribosyl transferase (APRT), the other important enzyme of salvage pathway, is specific for adenine.

$$\text{Adenine} + \text{PRPP} \xrightarrow{APRT} \text{AMP} + PP_i$$

HGPRT is quantitatively a more important salvage enzyme. It is competitively inhibited by IMP and GMP. APRT is inhibited by AMP.

Compared to the de novo pathway, the salvage pathways account for production of smaller fraction of the total purine nucleotides in the human body. But they do have certain advantages for being:

- Simpler and more cost-efficient way of producing purine nucleotides.
- Preventing wastage of raw materials and are particularly important in **brain** and other such tissues where de novo pathway is slow or absent.
- Finally, in contrast to the de novo pathway, which is virtually identical in all cells, salvage pathways are diverse in character and distribution.

Free purines can be salvaged (recycled) for (re)-synthesis of the nucleotides. This requires PRPP-dependent salvage enzymes, *APRP* and *HGPRT*. Interestingly, PRPP, the key element of salvage pathway, is the starting point of the de novo synthetic pathway as well.

Inhibitors of Purine Synthesis

The inhibitors of purine biosynthesis can be divided in the following categories:

1. ***The PABA analogues***, which structurally resemble para-amino-benzoic acid (PABA), inhibit synthesis of folic acid from PABA in bacterial cells.

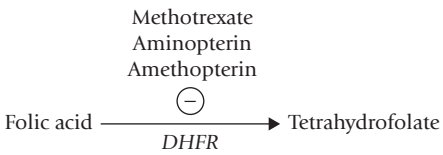
This decreases availability of tetrahydrofolate (THF), a derivative of folic acid. Since the de novo

purine synthesis requires THF (Reactions 4 and 10) decreased THF level inhibits this pathway. Diminished synthesis of purine nucleotides in the bacterial cell results, which in turn impairs bacterial cell growth and proliferation.

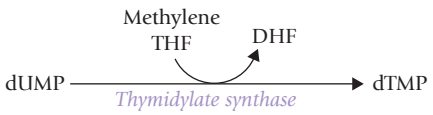
The above mechanism accounts for the *bacteriostatic effect* of **sulphonamides**, a PABA analogue (Fig. 18.8).

Humans lack capability to synthesize folic acid from PABA, but obtain this vitamin from dietary sources. Therefore, sulpha drugs have no deleterious effect on the human cell growth, so this drug can be safely used in humans.

2. ***Antifolates***, such as *methotrexate amethopterin, aminopterin,* and related compounds interfere with the action of folate cofactors, and are widely used in the treatment of cancer. They competitively (and irreversibly) bind to *dihydrofolate reductase* (DHFR) with about 1000-fold more avidity than dihydrofolate, the natural substrate of this enzyme. This causes inhibition of the enzyme DHFR activity, thus blocking the synthesis of *tetrahydrofolate*.



Decreased availability of tetrahydrofolate blocks synthesis of dTMP from dUMP by *thymidylate synthase* (because methylene THF is required in this reaction).



Further, these compounds (antifolates) are competitive inhibitors of several other THF-dependent reactions (4 and 10) used in the biosynthesis of purines.

These events interfere with the synthesis of new nucleotides and, therefore, of DNA and RNA. Consequently, cell proliferation is inhibited. This accounts for *antineoplastic effect of methotrexate and other antifolates*. A number of other purine analogues are also known to act as antineoplastic agents, as discussed below.

Development of nucleic acid analogues as antineoplastic drugs: An understanding of metabolism of nucleotides has helped development of several antineoplastic drugs. The toxic effects of these drugs are accounted by their ability to interfere with nucleotide metabolism. However, they suffer from a serious limitation, i.e. development of drugs with selective toxicity for tumour cells is difficult because most biological activities of the neoplastic cells are same as those of normal cells from which they are derived. *Drugs that are toxic to tumour cells are, therefore, likely to be nearly as toxic to normal cells*.

The tumour cells, however, do have a higher mitotic rate, and therefore, a higher requirement for DNA synthesis than normal cells. Therefore, they are more likely to be affected by antagonists of nucleotide synthesis. With this in mind, several drugs have been developed as antagonists of nucleotide synthesis. Some commonly used drugs and their modes of action are as here:

- *Glutamine antagonists*, e.g. azaserine, are useful as antineoplastic drugs. They inhibit those steps in purine and pyrimidine metabolism in which glutamine donates a nitrogen: the incorporation of N-3 and N-9 into the purine ring, and conversion reactions—IMP to GMP and UTP to CTP.
- *Structural analogues of bases of nucleosides* act by inhibiting individual reactions in nucleotide metabolism or through their incorporation into DNA or RNA. But such DNA or RNA is functionally inactive. This arrests cell division, and hence useful in treatment of cancers. A uracil analogue, 5-fluorouracil, exemplifies this type (discussed in **Case 20.4**).
- *Antifolates* are best exemplified by methotrexate, as discussed earlier.
- *6-Mercaptopurine* inhibits conversion of IMP to AMP and GM, as mentioned earlier.

Many antineoplastic drugs are inhibitors of purine or pyrimidine metabolism. These include glutamine antagonists, structural analogues of bases of nucleosides, antifolates and 6-mercaptopurine.

B. Catabolism of Purine Nucleotides

Catabolism of purine nucleotides is a continuous process, balancing the biosynthesis of these compounds. As a result, **the nucleotides of every cell are constantly revised**. Both adenine and guanine nucleotides are catabolized in a similar fashion (Fig. 20.12).

Catabolism of Adenosine 5′-Phosphate

First, AMP loses an amino group through the action of enzyme *AMP deaminase* to produce inosine 5′-phosphate. The phosphate group of the latter is then removed by the enzyme *5′-nucleotidase* to produce the corresponding nucleoside, inosine. Through phosphorolytic cleavage, ribose phosphate is liberated from inosine, leaving behind the hypoxanthine. The reaction is catalyzed by the enzyme *purine nucleoside phosphorylase*. Hypoxanthine is then converted to xanthine by the enzyme *xanthine oxidase*. Further conversion of xanthine to uric acid is catalyzed by the same enzyme.

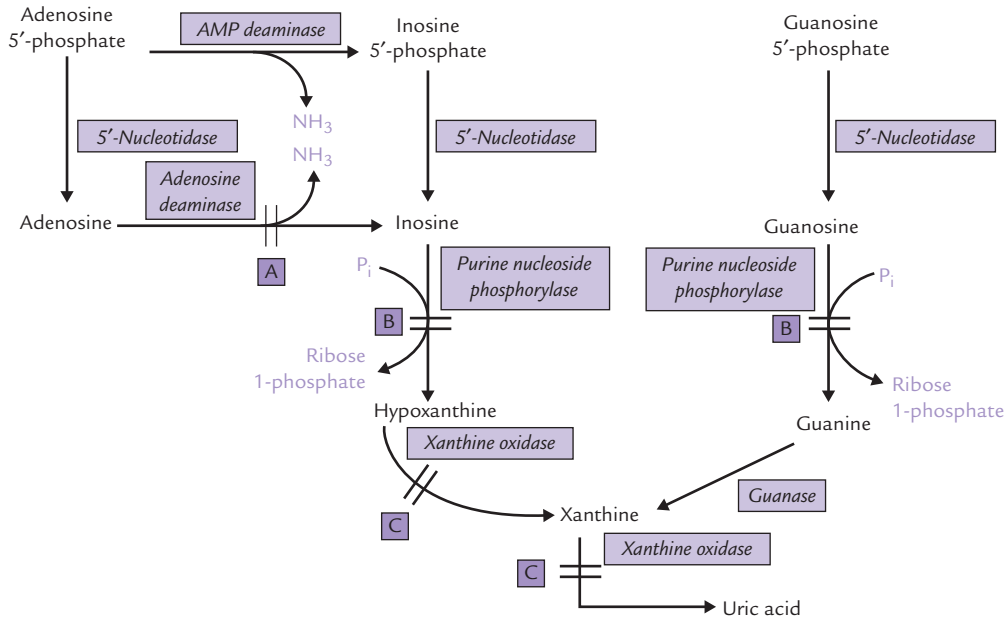


Fig. 20.12. Catabolism of AMP and GMP to uric acid (A = defect in severe combined immunodeficiency disease, B = defect in *purine nucleoside phosphorylase* deficiency, C = defect in xanthinuria. P_i = phosphate).

Inosine can be produced by an **alternate pathway** as well, in which the phosphate group of AMP is removed by *5′-nucleotidase* first to form adenosine. The latter then loses its amino group through action of *adenosine deaminase* to form inosine. (Adenosine, being a poor substrate for *purine nucleoside phosphorylase*, is not converted to adenine by this enzyme.)

Catabolism of Guanosine Monophosphate

It proceeds along the same lines. Sequential actions of 5′-*nucleotidase* and *purine nucleoside phosphorylase* produces guanine. The latter is deaminated by the enzyme *guanase* to form xanthine which is converted to uric acid.

Thus, xanthine is the common product in both adenine and guanine catabolic pathways, which is finally converted to uric acid by xanthine oxidase and excreted in urine. Thus uric acid is the final excretory product of purine catabolism. In addition, it can serve as an important antioxidant in primates by getting converted to allantoin. In animals other than primates an enzyme *uricase* converts uric acid to allantoin.

Purine ring cannot be broken down. To excrete the purine ring it must be modified to a form that can be excreted, i.e. the uric acid, which also acts as a natural antioxidant. Pyrimidine ring can be broken down.

Note: Most of the dietary purines are converted to uric acid in the intestinal mucosal cells only. Intestinal bacterial flora is involved in degradation of the rest of dietary purines that remain unabsorbed.

Table 20.1. Diseases of purine metabolism

Hyperuricaemia (gout):
- Overproduction of uric acid
  - Enzymatic defect **(primary gout)**
  - Secondary to other diseases **(secondary gout)**
- Renal disorder causing elevated threshold for uric acid excretion
- Unrecognized defects

Hypouricaemia:
- Diminished production of uric acid:
- Enzymatic defects in
  - *Adenosine deaminase*
  - *Purine nucleoside phosphorylase*
  - *Xanthine oxidase*
- Enhanced renal excretion of uric acid
  - Familial renal hypouricaemia
  - Fanconi's syndrome

C. Diseases of Purine Metabolism

Abnormalities in purine metabolism can lead to a number of disorders; the situation is akin to the disorders of amino acid metabolism. These disorders arise due to *enzymatic defects in the pathway of purine degradation or biosynthesis*. They may lead to altered serum uric acid concentration, i.e. **hyperuricaemia** or **hypouricaemia** (Table 20.1).

Hyperuricaemia and Gout

Hyperuricaemia refers to increased serum uric acid concentration. Gout is a clinical syndrome, associated with

hyperuricaemia. Interestingly, all patients who develop gout would have had hyperuricaemia at some point in the development of the disease. However, not all patients with hyperuricaemia develop gout: only a minority of them do so. The reasons for this are not known.

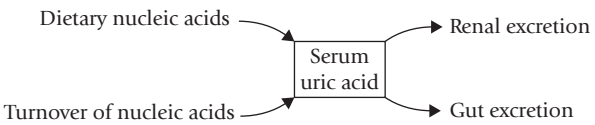
Gout is a clinical syndrome which is characterized by **hyperuricaemia** and **recurrent acute arthritis**. Renal disease (**nephropathy**) is also a common complication.

History

Gout has a remarkable history: it has been known to humans since ancient times and was defined by Hippocrates, the "father of medicine" in fifth century BC. Some of the eminent people including Newton, Martin Luther, Charles Darwin, and Benjamin Franklin suffered from gout. The word gout comes from the Latin word *gutta* which means drop, for it was earlier thought that a poison dripping into the affected joints was the cause of this disorder. Subsequently, hyperuricaemia was shown to be the underlying cause.

Causes and Consequences of Hyperuricaemia

Normal concentration of uric acid in serum is **3–7 mg/dL** (lower in women by about 1 mg/dL). Hyperuricaemia may result from **increased production** or **decreased excretion** of uric acid.



Some **unrecognized defects** may also lead to hyperuricaemia (Table 20.1).

In about 25% cases over-production of uric acid causes hyperuricaemia, and in the rest, decreased renal excretion is mostly the underlying cause.

Uric acid is barely soluble in plasma, so even a moderate rise in its concentration leads to precipitation of uric acid crystals. This accounts for the development of characteristic features of gout, which include subcutaneous deposits of sodium urate crystals called **tophi**, arthritis and renal impairment (**Case 20.1**).

Hyperuricaemia and its consequences are the major pathological processes in primary gout.

Gout is of two types: primary and secondary.

- In **primary gout** an inborn enzymatic defect leads to hyperuricaemia.
- **Secondary gout** refers to a state where some other primary disorder leads to the increase of uric acid level.

Over 90% of the hyperuricaemia patients are males; women suffering from this disorder are usually post-menopausal.

Primary Gout

Any disorder in which purine (de novo) synthesis is increased, causes enhanced production of purine nucleotides. Since purine nucleotides cannot be stored in human body, increased production is regulated by increased degradation. This results in enhanced uric acid production.

Some of the **enzymatic defects** leading to increased purine synthesis are as below:

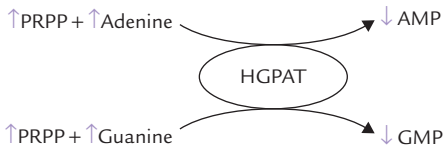
Increased activity of PRPP synthetase: Some variants of this enzyme have altered kinetic properties, such as increased V_{max} or low K_m for ribose 5-phosphate, or resistance to feedback inhibition. These properties permit this enzyme to bring about increased transformation of ribose phosphate to PRPP, leading to enhanced purine biosynthesis.

Note: The *PRPP synthetase* overactivity is inherited as an X-linked trait.

Glucose 6-phosphatase deficiency: The condition is called **von Gierke's disease**. Deficiency of this enzyme leads to intracellular accumulation of glucose 6-phosphate that forms pentose (ribose) phosphate through the stage II of hexose monophosphate pathway. Increased intracellular concentration of ribose phosphate leads to increased rate of purine synthesis.

Deficiency of hypoxanthine guanine phosphoribosyl transferase (HGPRT): It is a purine salvage enzyme; its deficiency is inherited as an X-linked trait causing **Lesch-Nyhan syndrome**. It is associated with increased purine synthesis by a 3-fold mechanism.

- Firstly, recycling of the substrate bases (hypoxanthine and guanine) is impaired, they are metabolized to uric acid.
- Secondly, accumulation of the PRPP, the other substrate for HGPRT, occurs, which is channeled into the de novo purine biosynthetic pathway. The pathway is stimulated because of feed-forward activation effect of PRPP.
- Finally, the defect in the salvage pathway leads to decreased levels of AMP and GMP, which decreases the feedback inhibition of the rate-limiting enzyme, *PRPP amidotransferase* (Fig. 20.11). This further speeds up purine biosynthesis.



Most cases of primary gout, however, fall into a less defined category in which the specific defect and the mode of inheritance is not known.

Secondary Gout

It refers to a state where hyperuricaemia is consequence of some other primary disorder as noted earlier. *Haematologic maliganancies* such as **leukaemia** and **multiple myeloma** lead to increased cell turnover, which is accompanied by increased degradation of excess nucleic acids to uric acid. **Chronic renal failure** is associated with an impaired excretion of uric acid with consequent retention of abnormally large amounts of this compound. Various drugs, particularly **diuretics** and **alcohol** also minimize the excretion of uric acid in urine, thus leading to hyperuricaemia **(Case 20.1)**. Urinary excretion of uric acid falls in starvation as well.

Uric Acid Pool in Gout

The miscible uric acid pool has been estimated to be around 1200 mg in normal subjects. It is increased to 3000 mg or more in patients suffering from gout. Uric acid may exist in the **dissociated** form (i.e. ionized form called *urate*) or the **undissociated** form (called *uric acid)*. At pH of 5.75, only about half of the total molecules are in urate form (since pK′ of the most acidic group of uric acid is **5.75**). *At body pH of 7.4, almost all molecules are in the urate form* because the body pH exceeds the pK′ value. Therefore, in all body fluids and blood, the urate predominates while uric acid accounts for a small proportion. In ***urine***, where pH may fall below 5.75, the *predominant form is uric acid*. The solubility of these two forms differs vastly: uric acid is about 13 times less soluble than urate. Solubility of the sodium salt of the latter is still lower.

Solubility of uric acid = 15 mg/dL
Solubility of urate = 150–200 mg/dL
Solubility of sodium urate = 7 mg/dL

Low Solubility: The Offending Factor in Gout

The low solubility and consequent precipitation of uric acid crystals is the cause of gouty arthritis and uric acid nephropathy. When concentration of serum urate is increased, its sodium salt (sodium urate) readily forms in the body. Its precipitation is a constant risk since limits of solubility (7 mg/dL) are exceeded with even moderate rise of serum urate level (normal serum urate is 3–7 mg/dL). Focal deposits of these precipitates, called *tophi*, are asymptomatic initially, but may lead to **arthritis** or **nephropathy.**

Gauty arthritis: Accumulation of trophi in joints triggers the **inflammatory response** of gouty arthritis by the following mechanism: the crystals are first ingested by leukocytes. These cells subsequently rupture, releasing the lysosomal enzymes that degrade articular tissue and induce the inflammatory response.

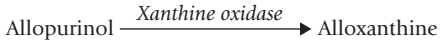
The crystals can be deposited in subcutaneous tissues also, most commonly in the external ear, over the knees and elbows, and along the tendons. However, a sustained hyperuricaemia for several years is required to bring about such deposition of tophi.

Nephropathy: Sodium urate or uric acid may precipitate in kidneys and ureters also to cause renal damage, termed **uric acid nephropathy**. As noted, formation of uric acid from urate is favoured as the pH falls. Therefore, with acidification of urine in renal tubules **uric acid** is generated. *Being of limited solubility, it tends to crystallize as urinary stones*. Sharp abdominal pain, obstruction to urine flow and damage to organ are the consequences.

Treatment of Gout

Is aimed at reducing joint inflammation (by *anti-inflammatory drugs*), increasing uric acid excretion (by *uricosuric drugs*), and decreasing uric acid production (by *allopurinol*). The most important drug treatments are:

- The *anti-inflammatory drug*, **colchicine**, is the classical treatment for the acute attack, though it is not effective in other forms of arthritis.
- **Probenacid** is a typical example of *uricosuric agent*.
- **Allopurinol**, an inhibitor of the *xanthine oxidase*, is a hypoxanthine analogue with interchanged N-7 and C-8 positions (Fig. 20.13). *Xanthine oxidase* hydroxylates allopurinol (as it does to hypoxanthine), yielding alloxanthine, which remains tightly bound to the enzyme, thereby inactivating it. Allopurinol is therefore classified as a **suicide inhibitor** of *xanthine oxidase* **(Case 6.1)**.



Decreased production of uric acid is accompanied by accumulation of hypoxanthine and xanthine. They are more soluble than uric acid, and hence readily excreted in urine.

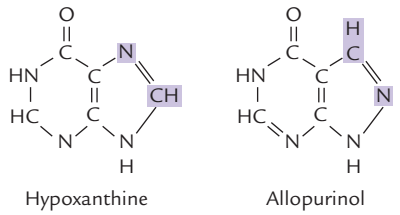


Fig. 20.13. Allopurinol, a purine analogue (resembles hypoxanthine), is initially oxidized to alloxanthine, which inhibits *xanthine oxidase* by tight, non-covalent binding.

Gout is a metabolic disease caused by the overproduction or under-excretion of uric acid. Due to its poor solubility, excess urate in the blood crystallizes and forms deposits in soft tissues such as kidneys (causing nephropathy) and in the toes and joints (causing gouty arthritis). Allopurinol, a purine analogue, is used for the treatment of gout due to its suicide inhibition effect on *xanthine oxidase*.

Besides *drug therapy,* **low protein diet** and **restriction in dietary intake of purines** is found beneficial in gout. Low protein diet is advised because amino acids increase de novo purine synthesis. Plenty of **oral fluids** and **abstinence from alcohol** is also recommended; the latter is harmful because of the associated dehydration and because the increased lactate levels during alcohol intoxication may impair renal excretion of uric acid. Acute attacks are precipitated by an alcoholic binge (**Case 20.1**).

Pseudogout

Deposition of calcium pyrophosphate crystals in joints leads to pseudogout. Clinical manifestations of this condition are similiar to those in gout but serum uric acid is not elevated.

Lesch–Nyhan Syndrome

The biochemical defect in this disorder is **deficiency** of the salvage enzyme **HGPRT**, as explained earlier. Partial deficiency of this enzyme causes only hyperuricaemia, but a complete deficiency results in Lesch-Nyhan syndrome. The disorder is named after the investigators who first described it in 1948. Since it is an X-linked trait, only males are affected.

Lesch-Nyhan syndrome (deficient *HGPRT*) presents as a vast array of neurological manifestations, superimposed on a strong tendency to develop gout.

Neurological abnormalities characterize Lesch-Nyhan syndrome and these appear after several months of life. Usually the initial presenting feature is delayed motor-development. Mental retardation is a predominant abnormality. The *brain has an extremely low capacity for de novo biosynthesis, therefore, a complete deficiency of salvage enzymes damages the brain by depriving it of purine nucleotides.* A unique feature of this disorder is the **self-mutilating behaviour** of the affected children who frequently bite their lips and fingertips, often violently; or jam their hands into spokes of their wheel-chair. Most patients die at an early age of uric acid nephropathy.

Hypouricaemia

Decreased concentration of serum uric acid (below 2 mg/dL) represents hypouricaemia. This is mostly due to enzymatic defects (in *adenosine deaminase* or *xanthine oxidase*), and less commonly due to some renal defect, e.g. Fanconi syndrome (Table 20.1).

Adenosine Deaminase Deficiency

Adenosine deaminase catalyzes deamination of adenosine to inosine, and of deoxyadenosine to deoxyinosine (Fig. 20.12). Deficiency of this enzyme results in **accumulation of adenosine** and **deoxyadenosine** (obtained from breakdown of DNA). It is believed that increased amount of deoxyadenosine is toxic to lymphocytes and suppresses immune function. This is because:

- The accumulated deoxyadenosine is converted to dAMP, dADP and dATP.
- The last one (dATP) is a powerful inhibitor of *ribonucleotide reductase*.
- This deprives the cell of precursor deoxyribonucleotide molecules for DNA synthesis. The selective impairment of lymphocytes but not of other cell types remains unexplained.

The enzyme depletion is associated with **severe combined immunodeficiency disease (SCID)**. In this inherited disease, the B-cells as well as the T-cells are defective.

Xanthinuria

Deficiency of *xanthine oxidase* is the underlying cause of this rare disorder (Fig. 20.12). Since this enzyme is involved in the production of urate, its deficiency causes the purine metabolism to stop at xanthine and hypoxanthine. As a result, *urate levels in blood and urine are exceedingly low, while xanthine and hypoxanthine become the major end products of purine metabolism*. In most cases this condition is not associated with any clinical symptoms and goes unnoticed. Diagnosis is usually made after the incidental finding of low serum urate (**Case 20.2**).

Purine Nucleoside Phosphorylase Deficiency

Purine nucleoside phosphorylase converts inosine to hypoxanthine and guanosine to guanine. Deficiency of this enzyme impedes conversion of the above nucleosides to their respective bases (Fig. 20.12). *Plasma accumulation of inosine and guanosine occurs, followed by increased urinary excretion of these compounds.* Plasma concentration of uric acid is decreased because uric acid is the ultimate catabolic end product of inosine and guanosine.

Purine nucleoside phosphorylase is responsible for degradation of purine deoxyribonucleosides also. Therefore, accumulation of deoxyribonucleosides also occurs in this disorder.

Adenine Phosphoribosyltransferase Deficiency

Adenine phosphoribosyl transferase (APRT) is a salvage enzyme that converts adenine to AMP. Deficiency of this enzyme impedes the salvage pathway resulting in plasma accumulation of adenine. The series of events that follow ultimately result in development of **renal stones (Case 20.3)**.

Familial Renal Hypouricaemia

The filtered uric acid is not adequately reabsorbed by renal tubules, resulting in increased urinary elimination of this compound. This, in turn, leads to hypouricaemia.

III. Pyrimidine Metabolism

A. Synthesis of Pyrimidine Nucleotides

Pyrimidine synthesis differs from purine synthesis in one important aspect. **The pyrimidine ring is synthesized before being attached to ribose 5-phosphate;** *in contrast, the purine ring is constructed on a pre-existing ribose 5-phosphate.* The constituent carbon and nitrogen atoms of pyrimidine are derived from various sources, namely aspartate, glutamine and carbon dioxide (Fig. 20.14).

Reactions of Pyrimidine Synthesis

The de novo reactions can be divided in two stages:

- Synthesis of the parent pyrimidine nucleotide (UMP).
- Conversion of UMP to cytidine and thymidine nucleotides (Fig. 20.15).

Stage I: Synthesis of UMP

The **first reaction**, the *committed step* of the pathway, involves *synthesis of carbamoyl phosphate* from glutamine and carbon dioxide. The enzyme catalyzing this step is *carbamoyl phosphate synthetase* II (CPS-II); 2 ATPs are required to drive forward the reaction.

It may be recalled that production of carbamoyl phosphate occurs by the mitochondrial enzyme, *carbamoyl phosphate synthetase I* (CPS I) during urea synthesis. Differences between the CPS I and CPS II are summarized in Table 20.2.

In contrast to the different enzymes in eukaryotes, prokaryotes contain only one *carbamoyl phosphate synthetase*, which is responsible for the biosynthesis of arginine and pyrimidines.

In the **second step,** carbamoyl phosphate and aspartate condense to form N-carbamoyl aspartate, catalyzed by the enzyme *aspartate transcarbamoylase* (*ATCase*). The **third reaction** is an *intramolecular condensation* leading to pyrimidine ring closure by the enzyme *dihydro-orotase.* The reaction product is dihydro-orotate, which is then oxidized to orotate in the **fourth reaction,** catalyzed by *dihydro-orotate dehydrogenase* in which NAD^+ serves as the coenzyme.

Orotate may be visualized as the parent pyrimidine base. Orotate is converted to the corresponding nucleotide, orotidine 5′-phosphate (OMP) by coupling with a phosphoribosyl group donated by PRPP. Thus, *attachment with phosphoribosyl group is postponed till the penultimate step of the pathway.* The **final reaction** involves decarboxylation *of OMP* to yield uridine 5′-monophosphate (UMP).

Stage II: Conversion of UMP to Cytidine and Thymidine Nucleotides

It occurs by a series of reactions shown in Figure 20.16.

Conversion of UMP to cytidine nucleotides: UMP is phosphorylated by two consecutive reactions, catalyzed by the enzymes *UMP kinase* and *nucleoside diphosphokinase* to produce UTP. The latter serves as a substrate for the cytidine nucleotide, cytidine 5′-triphosphate (CTP). The enzyme, *CTP synthetase* catalyzes this conversion by bringing about amination of the UTP; the source of amino group is glutamine.

Formation of thymidine nucleotides: The UMP is first converted to dUMP by the reaction sequence shown in Figure 20.16. The dUMP is then converted to dTMP by the enzyme *thymidylate synthase* which utilizes N^5N^{10}-methylenetetrahydrofolate as the source of the methyl group.

This is an unusual reaction where tetrahydrofolate donates not only a carbon unit but also undergoes change in its own oxidation state (from THF to DHF). Inhibitor of this reaction enzyme, 5′-fluorouracil (a thymine analogue), serves as an anti-tumour agent **(Case 20.4).**

Table 20.2. Differences between *carbamoyl phosphate synthetase* (CPS-I and CPS-II)

	CPS I	CPS II
Pathway involved	Urea cycle	Pyrimidine synthesis
Cellular location	Mitochondrial matrix	Cytosol
Source of nitrogen	Ammonia	Glutamine

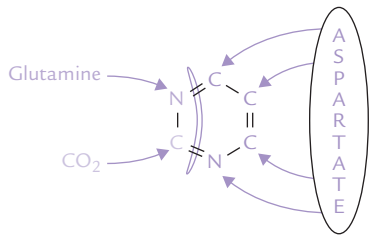


Fig. 20.14. The pyrimidine ring: origin of constituent atoms.

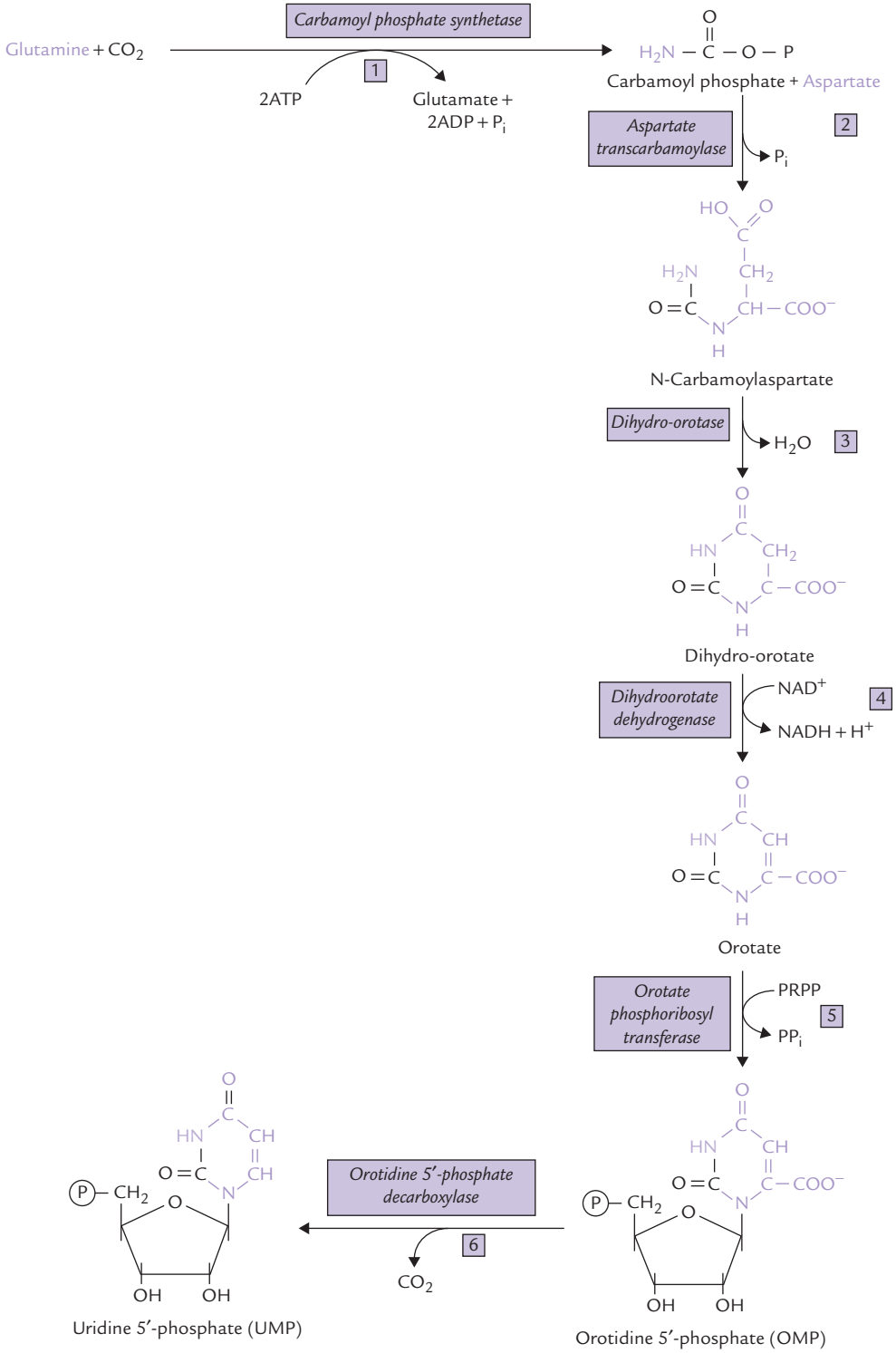


Fig. 20.15. Biosynthesis of pyrimidine nucleotides: the de novo pathway. *Dihydro-orotate dehydrogenase* is the only mitochondrial enzyme; all the others are cytosolic. Note that the pyrimidine ring is fully synthesized and then attached to ribose phosphate (in contrast purine ring is constructed on pre-existing ribose phosphate).

Regulation of pyrimidine synthesis

1. In mammalian cells, the main regulatory enzyme of the de novo biosynthetic pathway is *carbamoyl phosphate synthetase II*, which exists in a polyprotein (Box 20.1). The enzyme is *inhibited by UTP* and *activated by ATP* and PRPP, thus providing good examples of feedback inhibition and feed-forward stimulation, respectively.
2. In prokaryotic cells, the second reaction catalyzed by the enzyme *aspartate transcarbamoylase* is a regulated

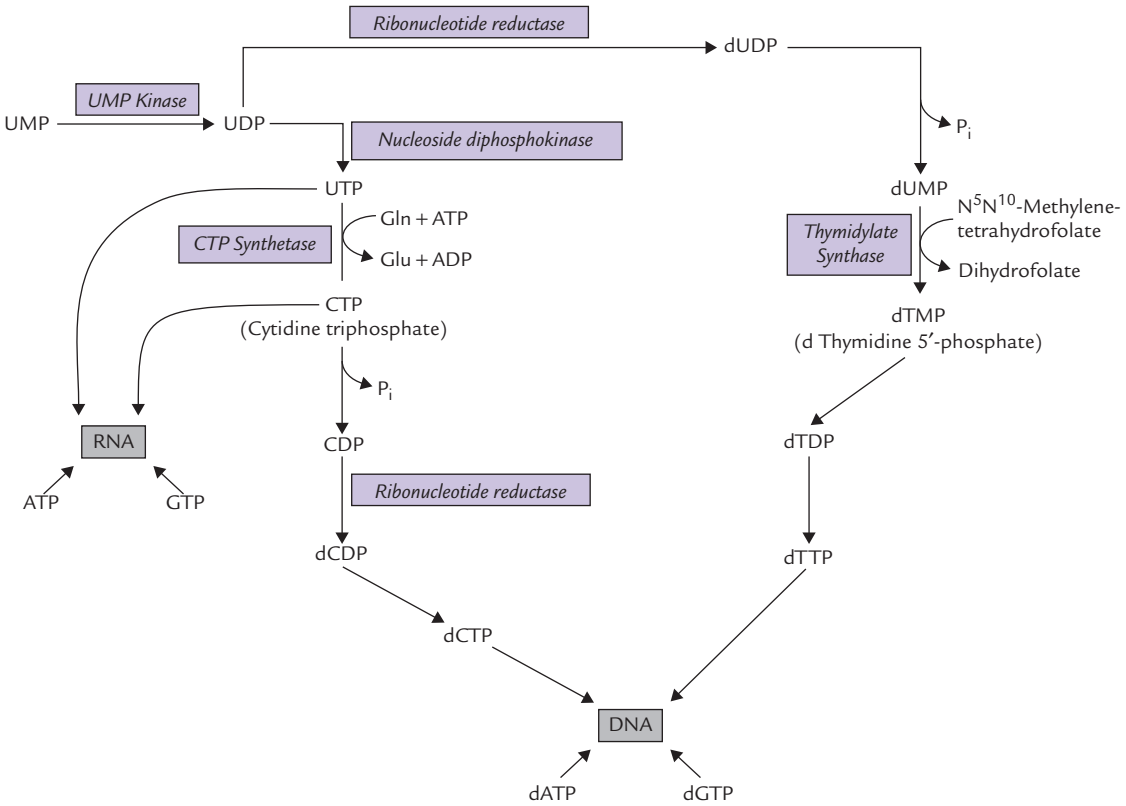


Fig. 20.16. Synthesis of cytidine and thymidine nucleotides from uridine nucleotides and their incorporation in nucleic acids.

BOX 20.1

Role of Polyproteins in Eukaryotes

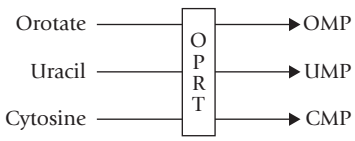
In bacteria, the six enzymes of UMP biosynthesis occur as independent proteins. In animals, however, the first three enzymatic activities of the pathway, namely *CPS II*, *ATCase* and *dihydro-orotase*, occur on a single 210 kD polypeptide chain. Similarly, the last two reactions (5 and 6) are catalyzed by a single polypeptide. The intermediate products of these multifunctional enzymes are not readily released to the medium but are channeled to the succeeding enzymatic activities of the pathway. This results in overall increase in the rate of these multistep processes.

step. It is *inhibited by CTP* and *stimulated by ATP*. In certain bacteria, UTP also inhibits *ATCase*.

3. *OMP decarboxylase* is inhibited by UMP and CMP. This also keeps pyrimidine formation under tight feedback control.

Salvage Reactions of Pyrimidine Synthesis

Orotate phosphoribosyl transferase (OPRT) that catalyzes phosphoribosylation of orotate (Reaction 5), also salvages other pyrimidine bases, such as uracil and cytosine, by converting them to their corresponding nucleotides.



B. Catabolism of Pyrimidine Nucleotides

Animal cells degrade pyrimidine nucleotides to their component bases, which are then broken down in the liver to amino acids, namely β-alanine and β-aminoisobutyrate. The latter is further metabolized to certain intermediates of fatty acid metabolism; thus, catabolism of pyrimidine nucleotides contributes, to a limited extent, to the energy metabolism of the cell.

1. The major enzyme involved in pyrimidine degradation is pyrimidine specific *5′-nucleotidase*. It brings about liberation of phosphate group from the pyrimidine nucleotides to form the corresponding nucleosides.

2. The next enzyme of the pathway, *nucleoside phosphorylase*, releases ribose 1-phosphate, leaving behind the corresponding free base. Conversion of uridine to uracil and of deoxythymidine to thymine takes place in this manner (Fig. 20.17).
3. The terminal steps involve a series of reactions (such as **ring splitting**, **reduction** of carbon bond, and **hydrolytic cleavage**), which ultimately result in the formation of β-alanine or β-aminoisobutyrate. Both compounds are transaminated and eventually metabolized to carbon dioxide although the exact pathway is not known.

Figure 20.17 summarizes the reactions involved in catabolism of cytidine 5′-phosphate and uridine 5′-phosphate. Degradation of the deoxythymidine 5′-phosphate to β-aminoisobutyrate by parallel pathway is also shown in this Figure.

A pyrimidine nucleotide is synthesized by firstly forming the pyrimidine ring and then attaching it to ribosyl phosphate. It is degraded to amino acids, namely β-alanine and β-aminoisobutyrate, which are then further metabolized.

The patients having certain defects in the catabolic pathway of pyrimidines need to be evaluated carefully. Though asymptomatic normally, they are at increased risk of developing severe neurologic reactions following administration of certain drugs **(Case 20.4)**.

Orotic Aciduria

Orotic aciduria is an autosomal recessive disorder, characterized by the *urinary excretion of large amounts of orotic acid*, retarded growth and severe anaemia.

Biochemical Defect

Orotic aciduria results from deficiency of either or both of the last two enzymes: *orotate phosphoribosyl transferase* and *OMP decarboxylase* (Fig. 20.15).

Clinical manifestations: Orotic acid is not inherently toxic, but excessive urinary level of this substance (more than 1 g/day; normal, 1.4 mg/day) leads to its crystallization and consequent *obstruction to urine flow*. Moreover, block in pyrimidine synthesis occurs in orotic aciduria, which leads to shortage of pyrimidine nucleotides for incorporation into DNA and RNA. Insufficient genetic material decreases generation of new cells. The rapidly proliferating cells such as red cells are most prominently affected, which results in **anaemia**. The erythrocyte precursors in the bone marrow become both abundant and unusually large. The diagnosis is confirmed by assay of the above enzymes in erythrocytes, or fibroblasts.

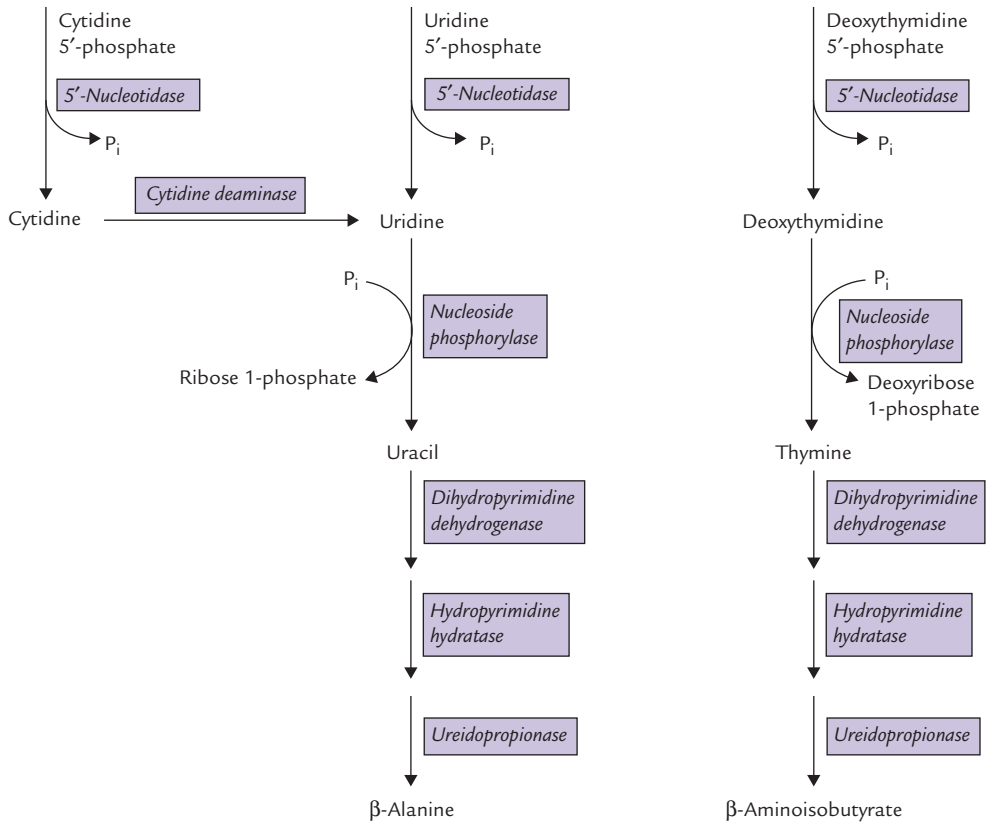


Fig. 20.17. Catabolic pathways of pyrimidine nucleotides.

Treatment: Intake of **diet rich in uridine** (or **cytidine**) is an effective treatment for orotic aciduria. Uridine provides enough UMP through phosphorylation, and also inhibits its CPS II so as to decrease the rate of orotic acid synthesis.

Orotic aciduria is an autosomal recessive defect in pyrimidine synthesis, treated by diet rich in uridine or cytidine, or by replacement therapy.

(Enzyme activities of both *orotate phosphoribosyl transferase* and *OMP decarboxylase* are present on a single protein as domains (bifunctional enzyme). Mutation in the *phosphoribosyl transferase* impairs its association with the other (*decarboxylase*) domain, resulting in loss of activity of both. This produces the **type 1** situation (i.e. *orotic aciduria type 1*). In **type 2**, mutation in the *decarboxylase* occurs, but it does not affect its aggregation with *phosphoribosyl transferase*. In this case, only *decarboxylase* is inactive, whereas activity of *phosphoribosyl transferase* is not affected.)

Rey's Syndrome

Excessive production of orotic acid may also occur in *deficiency of ornithine transcarbamoylase*, a urea cycle enzyme. It leads to accumulation of carbamoyl phosphate, which is diverted for enhanced synthesis of orotic acid. Increased urinary excretion of this compound follows, hence this defect is considered as a **secondary orotic aciduria,** also termed *Rey's syndrome*.

No other genetic deficiency in pyrimidine nucleotide biosynthesis is known because such defects are lethal *in utero*.

Exercises

Essay type questions

1. Review the nomenclature of bases, nucleosides, and nucleotides.
2. Compare the de novo pathways of purine and pyrimidine nucleotide synthesis with respect to (a) precursors, (b) energy cost, (c) acquisition of the ribose moiety, and (d) number of enzymatic steps.
3. Describe the salvage pathways for purines and pyrimidines. How do PRPP levels influence these? Describe an inherited disorder caused by deficiency of a purine salvage enzyme.
4. Explain biochemical defects in orotic aciduria, gout and Lesch-Nyhan syndrome.
5. How are folate cofactors involved in nucleotide metabolism? What antifolates serve as effective drugs?
6. What are the various mechanisms for regulating de novo purine synthesis? How does the cell balance the production of (a) purine and pyrimidine nucleotides, and (b) ribonucleotides and deoxynucleotides.
7. Differentiate between the primary and the secondary gout. Explain the biochemical basis of clinical abnormalities in primary gout and their treatment.

Write short notes on

1. Purine analogues
2. Allopurinol in treatment of gout
3. Secondary gout
4. Orotic aciduria
5. *Adenosine deaminase* deficiency
6. Synthesis of deoxyribonucleotides

CLINICAL CASES

CASE 20.1 A 40-year-old man with excruciating pain in big toe

A 40-year-old man was brought to casualty by the traffic police in a state of inebriation. He was involved in car accident while returning with his wife from a business party, late at night. On admission, he was disoriented, drowsy and was smelling of alcohol. Blood sample was sent to the biochemistry lab for estimation of plasma alcohol level and for performing other relevant investigations. History was obtained from his wife who had fortunately escaped without sustaining any major injury. She admitted that her husband had been a heavy drinker for the past 20 years, drinking about three bottles of whisky per week. Moreover, he had frequent episodes of excruciating pain in his big toe for the past several months.

Results of the biochemical investigations were returned after one hour.

Investigations test	Patient's report	Reference range
Glucose (random)	66 mg/dL	< 140 mg/dL
Urea	30 mg/dL	15–45 mg/dL
Urate	7.8 mg/dL	3.0–7.0 mg/dL
Na^+	140 mmol/L	135–145 mmol/L
K^+	4.6 mmol/L	3.6–5.0 mmol/L
Lactate	3.8 mmol/L	0.4–1.4 mmol/L

The House-Officer admitted the patient with a preliminary diagnosis of lactic acidosis. Early in the next morning, the patient had an attack of spasmodic pain in big toe. Emergency investigation showed that his plasma lactate had returned to the normal range by that time, but urate concentration was still elevated. On examination, the joints of big toe were highly inflamed and tender. On palpation, nodules could be felt in the ear lobes. Treatment with allopurinol was started and serum urate level was monitored. On the third day of the treatment, the urate level decreased and the patient was discharged.

Q.1. What is the probable diagnosis of this case?
Q.2. What is the origin of plasma lactate in this patient?
Q.3. How are patient's signs and symptoms related to ethanol abuse?
Q.4. Blood glucose level should be meticulously monitored in this patient. Explain why?
Q.5. Fructose accelerates metabolism of ethanol. What is the possible mechanism?

CASE 20.2 A 17-year-old asymptomatic boy with low serum urate

A 17-year-old boy was required to undergo routine medical check-up for appointment to a job in government service. He was apparently in good health and did not have any specific physical complaints. During biochemical analysis of his blood sample, serum urate level was found to be very low (0.9 mg/dL; normal, 3–7 mg/dL). Urinary excretion of uric acid was also markedly diminished. The tests were repeated after 4 weeks. No change/improvement was, however, observed.

24-hour urinary excretion of hypoxanthine plus xanthine (known collectively as oxypurines) was 540 mg. This was much higher than daily excretion of 6 mg xanthine and 10 mg hypoxanthine in normal people. Most of the excreted oxypurines in this patient was xanthine.

Q.1. What is the most probable diagnosis?
Q.2. Mention the test required to confirm the diagnosis.
Q.3. What are the complications associated with this condition?
Q.4. The major proportion of oxypurines excreted is xanthine and not hypoxanthine. Give reason.

CASE 20.3 A 6-year old boy with blood in urine

A 6-year-old boy complains of pain in lower abdomen for last few weeks. The pain was of severe intensity and caused great discomfort. Recently his mother had noticed blood in urine. The child came from a geographical region where the inclement climate makes growth of crops difficult and the people mainly subsisted on non-vegetarian foods that are mostly uncooked.

Physical examination showed no abnormal findings. Blood sample was sent to the biochemistry laboratory for a number of routine investigations, including the renal function tests. Radiological examination revealed stones in the right ureter.

Results from the biochemistry laboratory were returned the next day. No abnormal test results were, however, obtained. The child was operated, the stones recovered and sent to the biochemistry laboratory for determining the chemical composition of stones. Based on the results of the chemical analysis, it was decided to send a request to the laboratory to measure activity of *adenine phosphoribosyl transferase* (APRT) in erythrocytes. The enzyme activity was reported to be markedly reduced.

Q.1. How does the above biochemical defect lead to the formation of renal stones?
Q.2. Speculate the composition of the stones recovered from this child.
Q.3. Did the dietary habits of the child favoured stone formation? Would vegetarianism help him?
Q.4. Would treatment with allopurinol help this child? Give reason for your answer.

CASE 20.4 A 40-year-old woman with lump breast

A 40-year-old woman was investigated for a lump breast. Detailed investigations revealed it to be carcinoma. Following the surgical removal, the oncologist suggested palliative therapy with 5'-fluorouracil (5-FU) along with other chemotherapeutic agents. The first course of intravenous 5-FU appeared uneventful initially, though the patient developed mild anorexia, diarrhoea and a slight fall in white blood cell count. But soon she developed severe neurological reactions. Detailed investigations were carried out to elucidate the cause of the neurological reactions. Urinary levels of uracil and thymine were found to be markedly elevated.

Q.1. What is the rationale behind the use of 5'-fluorouracil in this patient?
Q.2. Identify the biochemical defect in this patient and provide an explanation for the above urine test results on the basis of your diagnosis.
Q.3. Is this biochemical defect responsible for the neurological reactions as well. Give reasons for your answers.

CHAPTER

21

MOLECULAR BIOLOGY I: DNA STRUCTURE, GENETIC ROLE AND REPLICATION

Genetic information stored in deoxyribonucleic acid (DNA) is inherited and expressed. Inheritance involves transfer of genes from parent cells to daughter cells; a **gene** is the length of DNA that directs synthesis of a polypeptide. At molecular level, inheritance requires replication whereby the parent DNA directs synthesis of daughter DNAs. The latter then passes into the daughter cells during cell division.

Gene expression determines characteristics of an individual. At molecular level, it involves DNA directed synthesis of proteins. This process occurs in two stages: transcription and translation. During **transcription,** which takes place in the nucleus and is directed by DNA, messenger RNA (mRNA) is synthesized. Structure of mRNA mirrors the image of a segment of DNA. In **translation,** the structure of the mRNA specifies the amino acid sequence of a polypeptide. In prokaryotic cells, translation follows immediately on transcription, whereas in eukaryotic cells these processes are separated both in time and place; transcription occurs in the nucleus and translation in the cytoplasm.

This chapter deals with the structure and genetic role of DNA. Causes of DNA damage and repair mechanism are also explained in detail. After going through this chapter the student should be able to understand:

- Structure of DNA: polynucleotide chains; size of DNA; and double helix.
- Physical properties of DNA: denaturation; buoyant density and supercoiling.
- DNA as genetic material: extra-chromosomal DNA; gene sequencing; architecture of human genome.
- Replication: semiconservative; initiation, elongation and termination in DNA replication: *DNA polymerases*.
- Mutations: Types and causes of mutations; repair systems for mutated DNA.

I. Structure of Deoxyribonucleic Acid (DNA)

DNA is the largest macromolecule in the body, consisting of millions of nucleotide units that are linked covalently. DNA is the storehouse of the genetic information which is translated into perceptible traits like the colour of skin, hair texture, height, etc.

A. Basic Chemistry

DNA is an antiparallel dimer of nucleic acid strands. Each strand is a **linear deoxyribose-phosphate chain** with purine and pyrimidine bases *attached to the* 2′-deoxyribose subunits. The deoxyribose differs from the ribose in that it lacks the hydroxyl group at the 2′-position. Carbon-3′ and -5′ of the deoxyribose subunits are involved in ester linkages with an inorganic phosphate to form a 3′,5′-**phosphodiester bond** (Fig. 21.1). Thus, alternating phosphate and 2′-deoxyribose units form the backbone structure of a DNA strand. The C-1 of the deoxyribose forms a **β-N-glycosidic bond** with nitrogen-1 of a pyrimidine base or with nitrogen-9 of a purine base.

The bases in DNA are adenine, guanine, cytosine and thymine; they lie flat in the interior of the dimer (Fig. 21.2). The phosphate groups are strongly acidic and, therefore, DNA molecule carries multiple negative charges at the physiological pH.

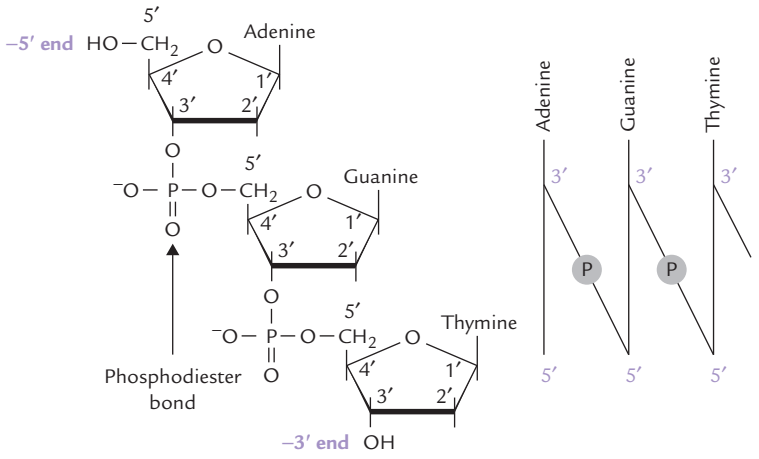


Fig. 21.1. Structure of a segment of DNA and formation of a phosphodiester bond between adjacent nucleotides.

Antiparallel strands: By antiparallel it implies that the two chains of a double helix have opposite polarity, i.e. they run in opposite directions. The polarity of a DNA strand is defined by two distinct ends:

- ***The 5′-end:*** This end has a free 5′-OH group on the deoxyribose as shown in Figure 21.1. The 5′-OH group is not connected to another nucleotide, though ***sometimes*** it may have a phosphate group esterified to it.
- ***The 3′-end:*** This end has a 3′-OH group that is not connected to another nucleotide (though it may also have a phosphate group on it).

Figure 21.1 shows a segment of DNA with the 5′ end at the top and the 3′ end at the bottom.

Polarity of the DNA chain is analogous to that of the polypeptide chain with an amino terminus and a carboxy terminus. Since the *carbons of 2′-deoxyribose are numbered by a prime to distinguish them from the carbons and nitrogens of the bases, the ends are designated as 5′ and 3′ ends*. A closer look at this simple oligonucleotide reveals that **variability of DNA structure depends on sequence of bases.** Since DNA consists of four bases, it is possible to construct 64 (4^3) different trinucleotides with them.

DNA is a large, double-stranded molecule with an alternating sequence of 2'-deoxyribose and phosphate in each strand. It contains heritable characters of the cell laid down in its base sequence.

B. Size of DNA

DNA molecules are extremely large. Contour length (the end-to-end length of a stretched-out native molecule) of a DNA molecule is several times more than the dimensions of the cell that accommodates it. For example, the

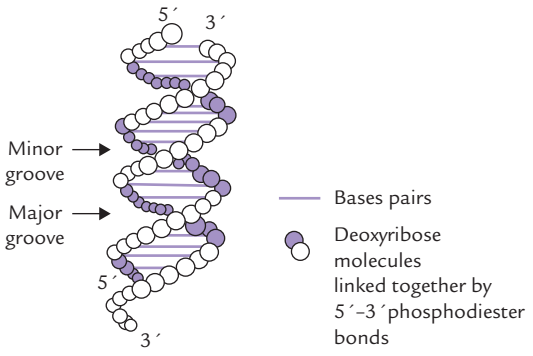


Fig. 21.2. Watson–Crick's double helix.

Table 21.1. Genomic size of living systems

System	Nucleotide pairs
Bacteria	$5 \times 10^5 - 2 \times 10^9$
Fishes	$1 \times 10^9 - 1 \times 10^{11}$
Mammals	$2 \times 10^9 - 6 \times 10^9$
Plasmids	$2 \times 10^3 - 4 \times 10^5$
Viruses	$5 \times 10^3 - 8 \times 10^5$
Humans	3.1×10^9

DNA molecule of an *Escherichia coli* cell contains 4.7 million base pairs. Its contour length is 1.4 millimeter, which is 700 times greater than the diameter of the bacterial cell. Size of the genome of various other living systems and viruses is shown in Table 21.1.

Viruses contain DNA which amounts to about 10% of DNA present in bacterial cells, which is consistent with the fact that viruses do not contain sufficient information for independent growth. On the other extreme are the **eukaryotic cells** which contain far more DNA than prokaryotes. An individual cell of a slime mold, one of the lowest eukaryotes, contains over 10 times more DNA than the *E. coli* cell. Length of the **human** genome (all the chromosomes taken together) totals one to two meters; there are about 10^{13} cells in a human being and if we add up the total

length of DNA in a person, it comes out to 2×10^{10} km: an astronomical length, of the order of the diameter of the solar system. Clearly, more complex the organism, greater would be the DNA content of its cell, but exceptions do exist. For example, a broad bean cell has more DNA than a human cell, and among vertebrates amphibia have the most. The apparent anomalies in DNA content are related to the fact that a vast majority of DNA is not in the form of functional genes, as described later.

Even the shortest DNA is very long in terms of the number of bases, and the longest is gigantic.

C. The Double Helix

Based on X-ray diffraction photographs of DNA taken by **Rosalind Franklin**, a two chain higher order structure was proposed by **James Watson** and **Francis Crick** in 1953 (Nobel Prize, 1962). The prominent features of this model of DNA, known as Watson–Crick's **double helix** (now known as B-DNA) are as here.

1. DNA is composed of two poly-deoxyribonucleotide (or simply deoxynucleotides) strands of opposite polarity, wound around each other on a common axis in a right-handed, helical structure (Fig. 21.2). Each strand is composed of monomeric units, of **deoxynucleoside 5′-monophosphates** namely
   (a) deoxyadenylate (dAMP),
   (b) deoxyguanylate (dGMP),
   (c) deoxythymidine (dTMP), and
   (d) deoxycytidine (dCMP).

2. The more hydrophilic deoxyribose subunits and phosphate groups are on the superhelix exterior, in contact with the aqueous environment; whereas the planar bases are stacked in the interior, where the environment is hydrophobic. The resulting helix has a **spiral staircase appearance**—the deoxyribosylphosphate residues act as the backbone, and the bases, which are oriented perpendicular to the superhelix axis, act as steps.

3. The bases on one strand interact with the bases on the other strand to form **base pairs**, which are planar and oriented nearly perpendicular to the axis of helix. (Fig. 21.3). The composite strength of the **hydrogen bonds** formed between the bases of the opposite strands is responsible for holding the two chains together and maintaining the double-helix structure.

4. Each base pair is formed by hydrogen bonding between a purine and a pyrimidine. To form hydrogen bonds, the tautomeric forms of the bases must be of the **keto type** (Chapter 20).

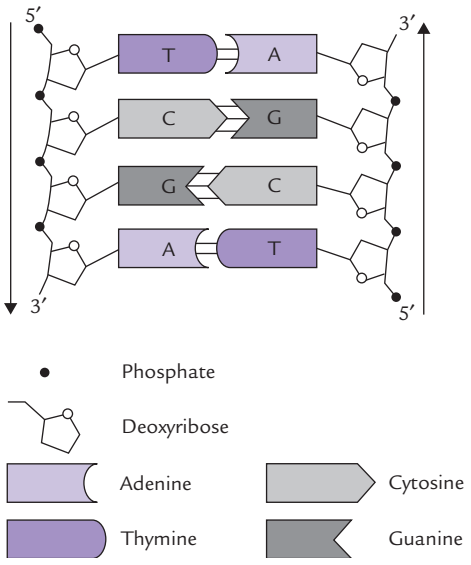


Fig. 21.3. DNA segment showing orientations of different components of the nucleotide subunits of DNA.

5. Because of the specificity of interactions between purines on one strand and pyrimidines on the opposite strand, the two strands are said to possess structures complementary to each other. **Complementary base pairing** occurs between A and T (A = T), and between G and C (G = C). As a result, the double-stranded DNA molecules always contain equal (molar) amounts of A and T, and of G and C. *Each A-T base pair is held together by two hydrogen bonds and each G-C by three.* The complementary base pairing proves the **Chargaff rule** (refer to Box 21.1).

Alteration of base pairing is potentially a disastrous event for the cell; it may lead to abnormal cell growth or even cancer (**Case 21.1**).

6. The bases lie flat in the interior, stacked on top of one another, each base being turned about 35° with respect to the next. They interact through *van der Waals* forces and *hydrophobic interactions*. Collectively these two forces are known as **base-stacking interactions** because of their contribution to the stacked arrangement of bases in DNA.

7. Each strand is wound into a **right-handed helix**, and each turn accommodates 10 base pairs. Since adjacent base pairs are stacked 3.4 A° apart, *pitch of the helix* is $3.4 \times 10 = \mathbf{34\ A°}$.

8. The ribosylphosphate backbones of the two strands are slightly offset from the centre of the helix and so two types of grooves of unequal width, the **major groove** and the **minor groove**, run the length of the DNA molecule. The major groove is more open (12 A° wide) and deep, and the minor groove is narrow (6 A° wide) and shallow. The unequal size is because the N-glycosidic bonds are not exactly opposite each other. Within these grooves,

BOX 21.1

Chargaff's Rule

Erwin Chargaff and his associates discovered in 1940s that in different DNAs (of even different species) quantity of purines is the same as that of the pyrimidines **(A + G = T + C);** the quantity of adenine was always equal to that of thymine and the quantity of guanine was equal to that of cytosine (A = T; G = C). This precisely fits the double-helix concept of Watson and Crick. Different DNAs vary in their percentages of A + T and G + C because their base compositions are different, reflecting their different genetic information.

the bases are exposed and hence access is possible for molecules, which need to interact with base pairs (e.g. the regulatory proteins). The latter interacts with DNA at these grooves without affecting properties of the double helix.

9. The double helix has a stiff, extended native conformation because of **electrostatic repulsion** between phosphate groups. However, it can be bent and twisted to a limited extent without major distortions of the regional structure.

The double helical structure is stabilized by Watson–Crick base pairing and by base-stacking interactions (i.e. van der Waals forces and hydrophobic interactions) between stacked base pairs.

Sense and antisense DNA strands: DNA transcription is normally initiated at specific sites on the DNA template and involves small, single-strand portions of the genome. The strand of the duplex DNA that serves as template during transcription is known as the antisense- or non-coding-strand since its sequence is complementary to that of the transcribed RNA. The other DNA strand is its sense or coding-strand. It has the same nucleotide sequence and orientation as the transcribed RNA (except for the replacement of U with T). *Two strands of DNA in an organism's chromosome can therefore contain different genes.*

D. Alternate Higher Order Structures of DNA

The double helix may assume a number of shapes in three-dimension. Although the vast majority of the DNA in living cells exists in the form described above (B-form), **five alternate forms** have also been recognized (**A-, C-, D-, E-** and **Z-DNA**). Among these, A and Z forms are important and may help to regulate gene expression. Their comparative features are listed in Table 21.2.

Table 21.2. Comparison of A-DNA, B-DNA and Z-DNA

Property	A-DNA	B-DNA	Z-DNA
Helix sense	Right-handed	Right-handed	Left handed
Base pairs/ turn	11	10	12
Rise/base pairs	2.6 A°	3.4 A°	3.7 A°
Helix pitch	28.6 A°	34 A°	44.4 A°
Widest diameter	25.5 A°	23.7 A°	18.4 A°
Major groove	Narrow and very deep	Wide and deep	Wide and flat
Minor groove	Wide and shallow	Narrow and deep	Very narrow

A-DNA

First reported by *Richard Dickerson*, this form is more squat in shape with its bases tilted. In the laboratory, at low salt concentrations and a maximum degree of hydration, the B-form predominates; but, when the salt concentration is increased and degree of hydration lowered, a reversible B- to A-DNA transition occurs. The A-helix is right-handed with 10.7–11 base pairs per helical turn and a rise of only 2.6 A° per base pair. The base pairs are tilted 19 A° relative to the helical axis. These features cause the **helix to be thicker and more compact than B-DNA.**

Small stretches of A-DNA are found within the conventional B-DNA strands. Double-stranded DNA-RNA hybrids also tend to assume the shape of A-DNA, because the –OH groups in the 2′ position prevent the B-conformation.

Z-DNA

The Z-form, first reported by *Alexender Rich*, is named so (Z stands for "zigzag") because location of the phosphodiester groups on the outer surface of the Z-helix can be connected by a broken (zigzag) line, rather than a smooth spiral as in D-DNA. The **helix is left-handed, and is thinner and more stretched out than B-DNA or A-DNA.** It has 12 pairs per turn and a 3.7 A° rise per base

pair. While formation of Z-DNA is favoured at high ionic concentration, it can also be induced at normal ionic concentration by DNA methylation. The control of gene expression caused by DNA methylation may be mediated by B- to Z-transition.

Base Composition Dictates Higher Order Structure

Formation of the alternate forms in a given segment of DNA is determined by its base composition:

(i) the polypurine segments become A-like,
(ii) the polypyrimidine segments are B-like, and
(iii) the segments in which purine and pyrimidine bases alternate (especially when G and C follow each other and C is methylated in 5′ position), the Z-DNA is formed.

All DNAs have grooves, which can interact with proteins such as **histones.** Such grooves have different dimensions, and therefore different protein-binding affinity in various forms of DNA. It is possible that the presence of the A- or Z-DNA within the normal B-DNA strand is a means of regulating its biological activity through changes in binding of regulatory proteins. This may help to regulate the gene expression.

II. Some Physical Properties of DNA

A. Denaturation

The term denaturation refers to disruption of native conformation of a biomolecule, so that it does not retain its higher-order structure. In case of DNA, denaturation refers to *separation of the double strands of the DNA into two component strands*. When temperature of the medium containing a DNA molecule is raised, the hydrogen bonds linking the complementary base pairs tend to break. As a result, separation of the two polynucleotide chains occurs.

Melting Temperature

Denaturation, also called melting of DNA, occurs over a narrow range of temperature. **Melting temperature (T_m)** indicates the temperature at which half of the double-stranded structure is lost. At the physiological pH and ionic strength, the melting temperature of DNA is **between 85°C and 95°C**. Since G and C are linked more strongly by three hydrogen bonds, in contrast to hydrogen bonds linking A and T, the DNA molecules having higher GC content have relatively higher melting temperature. For every 10% increase in GC content, the melting temperature increases by 5°C. One simple formula is T_m (°C) = 2 [number of AT base pairs] + 4 [number of GC base pairs].

Monitoring of Strand Separation

Denaturation affects properties of DNA and these can be used to monitor strand separation.

1. ***Hyperchromic effect:*** The heterocyclic DNA bases absorb ultraviolet light of wavelength 260 nm. Absorbance of this light increases by 40% upon denaturation. This is called **hyperchromic effect.** Thus, melting of DNA is monitored by the increase of UV absorbance: the melting temperature is characterized by a sharp increase in absorbance.
2. ***Viscosity:*** Viscosity of the DNA solution decreases on melting because the single strands are far more flexible than the stiff, resilient double helix.

Denatured DNA can Undergo Renaturation

Renaturation, also termed *reannealing*, is the process of formation of DNA double helix from two separated strands of DNA. For example, with gradual cooling to 5°–20°C below the melting temperature, the separated strands of the denatured DNA tend to join and form double-helix because of reformation of base pairs. This process, known as *annealing* (or renaturation), is useful in determining genetic similarity between DNAs derived from different organisms. It is a rapid process; even the large DNA molecules may take few seconds to minutes for renaturation.

The Watson–Crick double helix being unstable, the DNAs can undergo denaturation (into random coils), and can also reanneal under appropriate conditions.

In addition to **heat**, **alkali treatment** and decrease of **salt concentration** also cause denaturation. A variety of chemical agents that interfere with hydrogen bonding or base stacking are effective denaturants as well. Formamide, for example, is used as an alternative to heat denaturation in the laboratory. However, acid and ethanol cause DNA precipitation.

B. Buoyant Density

The maximum buoyant density of DNA is 1.70 ± 0.01. The DNA molecules having higher GC content have more compact structure and hence greater buoyant density. For every 10% increase in the GC content, the density rises by 0.12 units.

C. Hybridization

This refers to the pairing between RNA and complementary base sequences of a strand of DNA to form DNA-RNA hybrids that are slightly less stable than the corresponding DNA double helices.

D. DNA Supercoiling and *Topoisomerases*

Many naturally occurring prokaryotic DNA molecules are circular. The **circular DNA is compacted by supercoiling** which forms an important aspect of the DNA tertiary structure. In the relaxed form, the two polynucleotide chains of a DNA molecule are wound around a common central axis. Supercoiling implies twisting of the central axis of DNA upon itself (Fig. 21.4). It is not a random process but is precisely regulated by action of enzymes called ***topoisomerases***. These enzymes cause under-winding of the double-helix by removing one or more helical turns in its relaxed state (10 base pairs per turn). If one turn is removed by these enzymes, only seven turns would remain for the 84 base pairs. Each turn now accommodates 12 base pairs instead of 10. This is a deviation from normal and, therefore, it induces a **thermodynamic strain**. This strain can be relieved either by partial strand separation or by supertwist of the duplex around its own axis—much as a telephone cord twists around itself. This kind of supertwist permits the neighbouring bases to assume positions which closely resemble position of bases in a double helix.

The supercoiling mentioned above arises due to under-winding and is called **negative supercoiling.** The opposite situation arises due to over-winding and is called **positive supercoiling**. Most bacterial DNAs are supertwisted negatively, having approximately 5% to 7% fewer turns than expected from the number of base pairs. This is helpful because:

(a) Negative supercoiling counteracts, to some extent, the positive supercoiling upstream of the unwinding DNA (Chapter 21).

(b) Supercoiling also aids in the process of DNA packaging, discussed later. The negative supercoiling shown in Figure 21.4 is introduced by an enzyme in prokaryotic cells, called ***gyrase***, with the expenditure of ATP energy. *Gyrase* is a member of a group of *topoisomerases*.

Topoisomerases: The *topoisomerases* generate supercoils, but their more important role is to relieve the supercoils during the process of DNA replication and transcription. They perform this role by cleaving the *phosphodiester* bond on one or both DNA strands (*nuclease activity*) and then resealing the "nick" (*ligase activity*) after rotation of the DNA strand(s) to relieve tension. Various types of *topoisomerases* (Type I to IV) have been recognized. *Topoisomerase I* breaks only one strand, and *topoisomerase II* breaks both. *Gyrase* is *topoisomerase II*. *Topoisomerases* are thus integral components of DNA replication and transcription mechanisms (see Table 21.4).

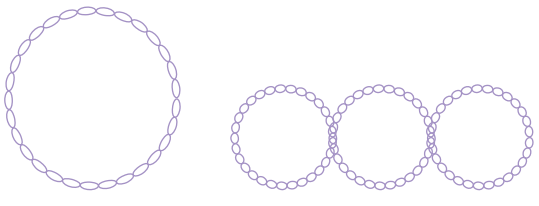


Fig. 21.4. Relaxed and supercoiled DNA.

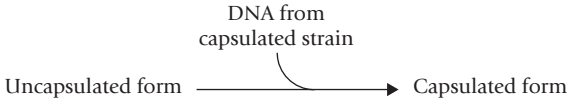
Most cellular DNAs are negatively supercoiled (under-wound). *Topoisomerases* are **molecular swivel** that relax supercoils by cleaving one or both strands of the DNA, passing the DNA through the break, and resealing the broken strand(s).

III. DNA as Genetic Material

DNA carries the blue-print of heritable traits in living systems. Simply stated, DNA is the genetic material in all prokaryotic and eukaryotic organisms. The genetic material must fulfill the following basic requirements:

1. It must serve as bearer of genetic information and be able to transmit this information with a high degree of fidelity to the daughter cells.
2. It must be stable so that the stored information remains protected.
3. A limited degree of variation must be allowed so as to provide raw material for natural evolution.

Miescher first isolated DNA in 1869, but the first experimental evidence to indicate that it is the genetic material came much later from the persuasive experiments of *Avery*, *Macleod* and *McCarty* (1944) on bacterial transformation. These workers demonstrated transformation of an uncapsulated strain of the bacterium *Pneumococcus* into a capsulated form by the addition of DNA extracted from a capsulated colony.



The DNA preparation used was shown to be free of proteins, and the DNA cleaving enzyme, *deoxyribonuclease*, was shown to destroy its activity. These observations provided evidence that the specific alteration of a heritable function (uncapsulated to capsulated transformation) was brought about by transfer of DNA.

Hershey and *Chase* in 1952 proved that viral DNA was infectious. These workers infected *E. coli* by T-2 bacteriophage. DNA of the latter was labelled with P-32 and the protein coat was labelled with S-35. When *E. coli* was infected with this doubly labelled phage, only P-32 labelled nucleic acid entered the bacterial cell. It was thus observed that

only DNA could enter the cell, none of the other viral constituents (e.g. proteins or carbohydrates) had this ability.

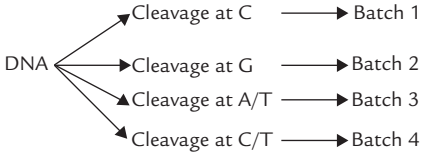
A. Extrachromosomal DNA

In eukaryotic cells, small amount of DNA is present in mitochondria. The **mitochondrial DNA** (mt DNA) undergoes replication and directs protein synthesis, independent of chromosomal DNA. Precise role of the mt DNA is still unclear. Probably it represents vestiges of the foreign DNA (bacterial) that infected the cells early in evolution (Chapter 24).

In bacterial cells, extrachromosomal DNA is present in structures called **plasmids**. Each plasmid contains circular DNA material. The plasmid DNA codes for proteins that confer resistance to antibiotics. Plasmids can be transferred from one cell to another and are, therefore, called **vector molecules** (Chapter 25).

Gene Sequencing

Several techniques have been developed for determining base sequence of DNA. One of these, called **Maxam–Gilbert method** employs selective chemical cleavage of a single-stranded DNA sample. The sample is divided into four batches and subjected to selective chemical cleavage at specific bases. Batch 1 is cleaved on 5′ side of cytosine; batch 2 is cleaved on 5′ side of guanine; batch 3 cleaved on 5′side of purines (A or T); and batch 4 on 5′ side of pyrimidines (C or T).



Smaller sets of DNA fragments are obtained (one set from each batch), whose length identifies position of a particular base. For example, a 20-nucleotide fragment in the batch 2 (G only) reaction mixture identifies G at position 21. Similarly, a fragment containing 23 nucleotide in batch 4 (C or T) reaction, but not in batch 1 (C only) reaction, indicates that T is at position 24.

The length of DNA fragments is determined by polyacrylamide gel electrophoresis, a technique that can separate fragments differing in length even by one nucleotide.

Another sequencing technique is the **Sanger method** (developed by Fred Sanger), which is based on controlled interruption of DNA synthesis by *dideoxynucleoside triphosphates* (ddNTP). The latter can be incorporated in the new strand by *DNA polymerase*, but halts further chain growth (Fig. 21.5). Enzymatic DNA synthesis is performed in four test tubes. Each tube contains all the four deoxyribonucleoside triphosphates (dATP, dGTP, dCTP, dTTP) but only one of the four ddNTP. For example, in tube 1 ddATP is added, in tube 2 ddGTP is added, etc. In each test tube a single-stranded, DNA template is copied.

A series of radioactive DNA fragments are generated, differing in length. The length depending upon:

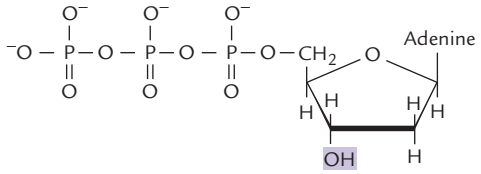
(a) Type of ddNTP added: in tube 1, chain termination occurs at A, in tube 2 the chain terminates at G.
(b) Stage at what ddNTP was incorporated into the chain.

These fragments are separated by polyacrylamide gel electrophoresis according to their size. The sequence of the DNA fragment can be directly deduced from the pattern of radioactive bands in the gel.

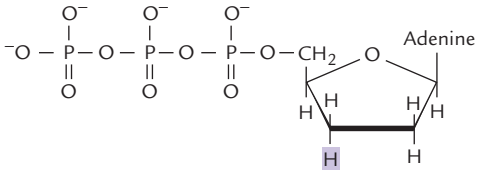
Determination of base sequence of the entire DNA of an organism, called **gene mapping** is now possible with tools of recombinant DNA technology (Chapter 25).

IV. Architecture of the Human Genome

In human cell there are 46 chromosomes, as already mentioned, with a total length of DNA per cell of 1–2 meters. It has to be packed into a nucleus millions of times smaller in size (dimensions, $6 \times 9\,\mu m$). DNA has responded to this challenge over the course of evolution through a very elaborate packing procedure by associating with a group of basic proteins called **histones**. The nucleoprotein complex so formed contains 50% DNA and 50% protein, and is called **chromatin**. It occurs in two forms: *euchromatin*, which has a loose structure and *heterochromatin*, which is more condensed and stains more deeply with basic dyes.



2′-Deoxy ATP
(normal substrate)



2′, 3′-Dideoxy ATP (ddATP)
(causes chain termination
during DNA replication)

Fig. 21.5. Structure of a dideoxynucleoside triphosphate. *DNA polymerases* can incorporate a dideoxynucleotide into a new DNA strand, but further chain growth is prevented for lack of a free 3′-hydroxy group.

Only euchromatin contains actively transcribed genes. Evidently, the access to information is not interfered with (due to tight DNA packaging) in chromatin.

Eukaryotic genomes are composed of several linear DNA duplexes that are organized into separate chromosomes. Each chromosome has a single DNA, stabilized by interaction with histones, the basic proteins.

Chromatin stains with basic dyes, such as haematoxylin and fuschin, and so is conspicuous in histological preparations. Chromatin from the resting interphase is usually studied, wherein it is visualized as amorphous and randomly distributed material. But in dividing cells (metaphase), the chromatin condenses and assembles into specific number of well-defined chromosomes.

A. Histones

There are five histones associated with DNA, essential features of which are listed in Table 21.3. They are rich in **arginine** and **histidine**, giving them a positive charge, that forms ionic bonds with the negative charges on the phosphate groups of the double helix. The amino acid sequences of eukaryote histones are highly conserved throughout the phylogenic tree. (This suggests that histones were invented early during evolution to precisely combine with an equally invariant structure, i.e. DNA). Thus, the histones, H3 and H4, from pea seedlings and calf thymus differ in only four and two amino acid positions respectively, and the changes are very conservative (valine for isoleucine, lysine for arginine). However, such extreme conservation is rarely seen in H1, which shows variation in structure in different species, and even in different tissues of the same organism.

Note: Mitochondrial DNA and bacterial DNA do not contain histones.

B. Nucleosomes

Clusters of histones consisting of two molecules each of H2A, H2B, H3 and H4 form an octamer protein complex, called a **nucleosome core**, around each of which the DNA wraps two turns (about 146 base pairs) in a left-handed orientation. The octamer with its DNA is a nucleosome, a small disc-like structure with a diameter of 10 nm. The nucleosome is the structural unit of chromatin: in electron microscopy it shows like beads on string (Fig. 21.6a), where the beads are nucleosomes.

The nucleosomes are connected by an intervening stretch of DNA of a variable length of 50–60 base pairs. It is called **linker DNA** and is associated with the H1 histone (Fig. 21.6b).

Table 21.3. Properties of histones

Type	No. of residues	Mol wt (kD)	Copies per nucleosome
H1	215	21.0	1
H2A	129	14.5	2
H2B	125	13.7	2
H3	135	15.3	2
H4	102	11.3	2

H1 is located in linker and the rest in nucleosome core.

C. Packaging of DNA

There are several levels of organization that result in tight packaging of DNA into the nucleus.

1. ***Winding around nucleosomes:*** The wrapping of DNA around the nucleosome core results in 7-fold compaction of DNA. The nucleosomes, being of about 10 nm diameter each, form the 10 nm fibres shown in Figure 21.6b.

2. ***Nucleosomes packed into a 30 nm fibre about three nucleosomes wide:*** The 10 nm fibre is condensed further to form a 30 nm fibre illustrated in Figure 21.6c. The nucleosomes are packed in the 30 nm fibre in an as yet undetermined manner, though it is often depicted as a helical or solenoid arrangement of the nucleosomes. There are six nucleosomes per turn of the helix; this would yield a fiber three nucleosomes wide, which is indeed the diameter observed (30 nm).

The condensation of DNA achieved so far is about 40-fold, which may be sufficient for the dispersed chromatin of the interphase nucleus. However, further 200-fold compaction is required for the formation of fully condensed chromosome.

3. ***The 30 nm fibres form loops, thousands of nucleosomes long attached to central scaffolding:*** The further compaction into the metaphase chromosome is achieved by attachment of the loops (of 30 nm fibre) to the *central protein scaffold* of the chromosome (Fig. 21.6d). Each loop of 30 nm fibre from one scaffold attachment to the next measures approximately 0.4–0.8 μm and contains 45,000–90,000 base pairs. It is believed that these loops are further condensed, forming yet other coils and/or folds and achieve a 10,000-fold packing of the original DNA.

V. Some Commonly Used Terms Related to Molecular Genetics

A. Gene Families and Pseudogenes

Most protein-coding genes are present in a single copy in the haploid genome. Occasionally, however, two identical

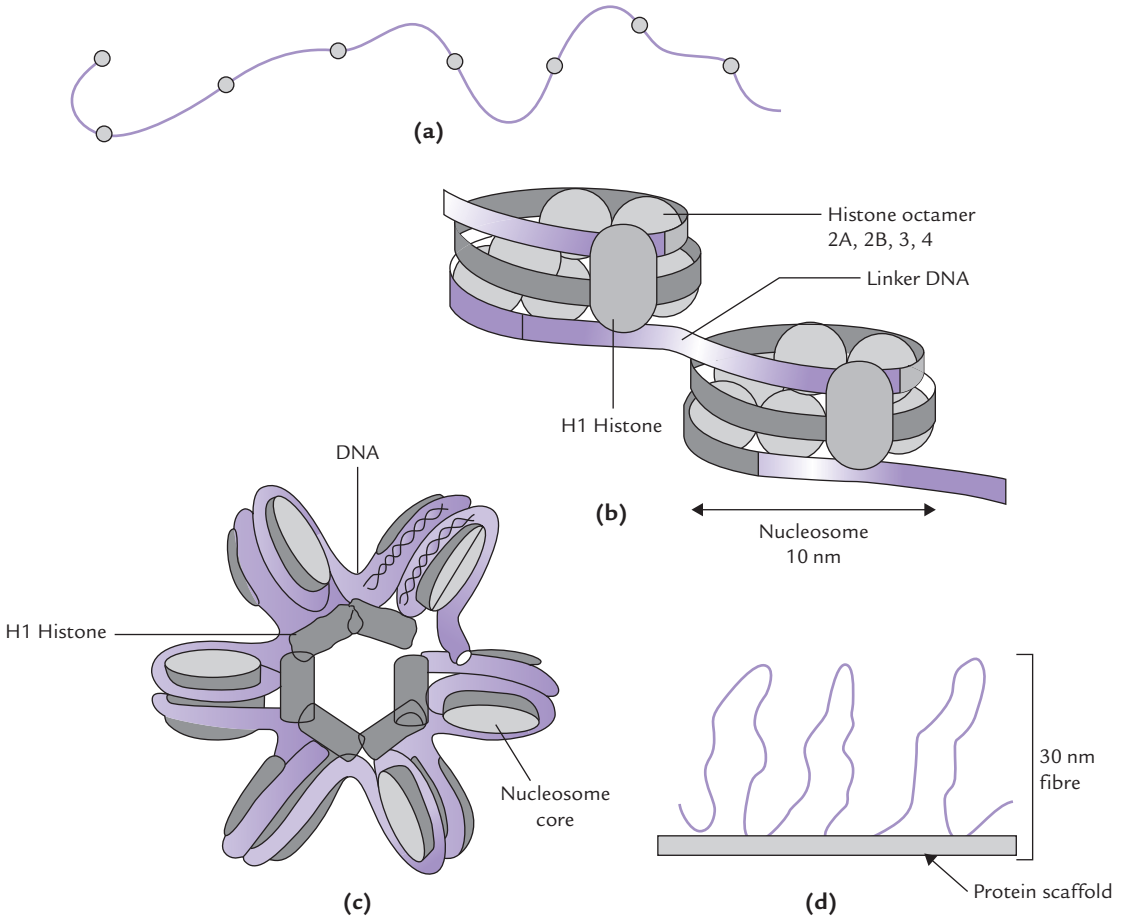
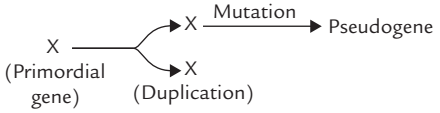


Fig. 21.6. Chromosome structure. **(a)** Swollen fibre of chromatin, **(b)** Nucleosome connected by linker DNA, **(c)** Formation of helix or solenoid: there are 6 nucleosomes per turn in the solenoid (30 nm fibre), **(d)** Fibres on central protein scaffold of chromosomes.

or near identical copies of a gene are present together on the same chromosome. They are called **duplicated genes**. They originate by duplication of a primordial gene.

Some genes that code for very abundant RNAs or proteins are present in *multiple copies*. In most cases identical (or near identical) copies of the gene are arranged in tandem over long stretches of DNA separated by untranscribed sequences. Examples in humans include rRNA genes (200 copies), the 5S rRNA genes (2000 copies), the histone genes (20–50 copies) and most of the tRNA genes. In some other cases, gene duplication of a primordial gene is followed by divergent evolution of the duplication products. Consequently, we find that two or more similar but not identical genes are present in the genome, usually together on the same chromosome. Such structurally related genes with a common evolutionary origin constitutes a **gene family**. The α- and β-like globin gene clusters are the classic examples of such gene families.

Pseudogenes are unexpressed DNA sequences that exhibit remarkable sequence homology to some functional gene. They originate by gene duplication, followed by crippling mutation in one of the duplication products that renders it defunct. Pseudogenes are generally located close to their functional counterpart on the chromosome.



B. Jumping Genes

These genes, described first by McClintock, *do not appear to have a fixed position within the genome, but rather move from one location to another.* They can jump to any place in the chromosome—there are no specific insertion sites for them. Such *mobile DNA sequences in prokaryotes include* ***insertion sequences*** *and* ***transposons***.

Insertion Sequence

An insertion sequence is a mobile segment of DNA (750–1500 base-pair long) that contains a single gene and is framed by inverted repeats (IR) of 9–41 base pairs. The gene codes for a protein called ***transposase*** that catalyzes the movement of insertion sequence to the new

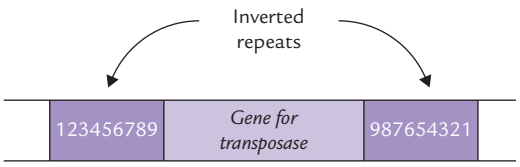


Fig. 21.7. The insertion sequence consisting of gene for *transposase*, flanked by inverted repeats of a base pairs (1–9 indicate sequence of base pairs).

genomic positions (i.e. *transposition*). From one point of view, insertion sequence (IS) might even be regarded as a futile element—doing nothing but code for its own movement from place to place. Most insertion sequences are present in 5–30 identical (or near identical) copies in the cell (Fig. 21.7).

Transposition is a rare event, occurring once every 1000 to 10,000-cell generations. It can occur in two forms: ***conservation transposition*** and ***duplicative transposition***.

- In **conservative type**, the insertion sequence vacates its old place and settles to a new position.
- In **duplicative transposition**, the insertion sequence replicates and one of the replication products moves to new site, whereas the other one stays at the old position.

Transposons

In common with insertion sequences, the transposons are flanked by inverted repeats and contain gene for their own *transposase* (the latter catalyzes their movement into new genomic locations). A distinctive feature of transposons is that (unlike IS) they contain useful genes, for example, antibiotic resistance genes.

Jumping genes are genetic elements that move within a genome by a mechanism involving their replications. Possibly, the chromosomal rearrangement caused by these mobile elements is of significance in the evolutionary process.

C. Silent Genes

These are mutant forms of genes which do not have a phenotypic effect. They may play a role in gene regulation.

D. Complex Genes

These are transcribed together to yield more than one RNA transcripts. In prokaryotic genome such genes are frequently arranged in tandem along a single strand (eukaryotic genes, by contrast, are transcribed individually).

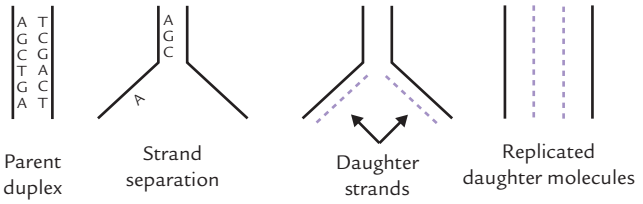


Fig. 21.8. Semiconservative nature of DNA replication. Each new daughter molecule consist of a parent strand and a daughter strand, which are complementary to each other.

E. Cell Interaction Genes

These genes play an important role in immunological recognition. They belong to major histocompatibility complex (Chapter 33).

Oncogenes, antioncogenes and *protooncogenes* are described in Chapter 32.

VI. Replication

Every time a cell divides, the entire content of its chromosomal DNA must be duplicated so that a complete complement of DNA can be given to each daughter cell. DNA is able to duplicate, or as is more commonly stated, replicate, by separating and copying the parent strands to produce two daughter molecules that are identical to the parent DNA. Each separated parent strand serves as template for the synthesizing a new strand, so that each of the two daughter molecules contains a parent strand and a new daughter strand (Fig. 21.8). It is noteworthy that the *template strand is copied in the complementary sense, which is dependent on Watson–Crick base pairing.*

Division of the cell into two daughter cells (mitosis) follows replication of chromosomal DNA into two daughter cells and each daughter cell gets an identical copy of the replicated DNA. In this way the genetic information is duplicated and transmitted to the next generation.

A. Replication is Semiconservative

The method of replication shown in Figure 21.9 is a 'semiconservative mechanism' because *each replicated duplex contains one parental strand and one newly synthesized strand*. Messelon and Stahl confirmed this mechanism at the molecular level by the following experiment.

1. They grew *E. coli* for several generations in a medium containing ^{15}N-ammonium chloride as the only source of nitrogen so that any newly synthesized DNA incorporated this heavy isotope of nitrogen (i.e. ^{15}N-DNA). This is **heavy** DNA.

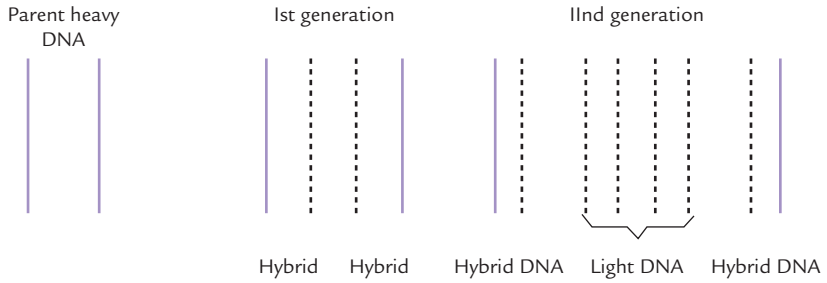


Fig. 21.9. Messelon–Stahl experiment proves semiconservative nature of DNA replication.

It is possible to separate (by density gradient centrifugation in cesium chloride) this ^{15}N-DNA (heavy) from ^{14}N-DNA (**light**) and from the **hybrid** (^{14}N/^{15}N-DNA) of intermediate density.

2. Fully ^{15}N labelled cells were then abruptly transferred to a medium containing ^{14}N-ammonium chloride, so that all subsequent DNA chains synthesized would be 'light'.
3. DNA was isolated at various generation times and its density monitored by density gradient centrifugation (Fig. 21.9).

Generation 1: There was a single band of DNA halfway between those of ^{14}N-DNA and ^{15}N-DNA, indicating its hybrid nature (^{14}N/^{15}N-DNA).

Generation 2: Half the DNA was hybrid DNA and the other ^{14}N-DNA.

In succeeding generations: Proportion of light strands increases, but hybrid molecules persist.

These observations rule out the conservative mode. For the **conservative model** to be correct,

(a) all new DNAs must have normal density,
(b) some heavy DNAs must be present at each generation, and
(c) no hybrid DNAs should never be formed.

On the other hand, the *semiconservative replication* predicts all hybrid DNAs at the first generation and after two generations equal amounts of hybrid and ^{14}N-DNAs; with no ^{15}N-DNA ever being formed. In view of the results discussed above, the conservative mechanism can be ruled out, thereby confirming the semiconservative mechanism.

Each daughter duplex consists of one parental strand and one newly synthesized strand (semiconservative mode).

B. Three Phases of Replication

The three phases of replication are *initiation*, *elongation* and *termination*. *E. coli* cells are mostly used for the study of replication. Fundamental mechanism and the participating enzymes of *E. coli* are similar to those of other living systems, including humans. Replication in prokaryotes is described in this Chapter; the comparative features of eukaryotic replication are discussed in Chapter 24.

Initiation

Initiation is a highly specific and precisely regulated stage of replication. It is at this stage that regulation of replication occurs; once the initiation phase gets started, replication of whole of the DNA must complete.

This phase starts with the unwinding of the chromosome at a single origin site, termed the **oriC locus**. It consists of a unique sequence of 245 base pairs that is very rich in A-T pairs, presumably to facilitate strand separation (A-T pairs having two hydrogen bonds are less tightly bound than G-C pairs which have triple hydrogen bonding). A complex of up to 30 subunits of the 52-kD **DnaA protein** binds to oriC and causes a segment of DNA to melt open. The melting process requires *ATP hydrolysis*, and is facilitated by the AT rich nature of the DNA segment. Further strand separation and subsequent unwinding of the duplex DNA are needed to produce the single-stranded DNA template required for DNA replication. A class of proteins termed *helicases* accomplishes separation.

Helicases

Strand separation by *helicases* is brought about through dissolution of hydrogen bonds holding the two DNA strands together. The process is ATP dependent; (it is likely that the hydrolysis of ATP could produce a cyclic change in shape of the *helicase*, which enables it to move unidirectionally along the DNA strand and disrupt the hydrogen bonds).

At least **nine *helicases*** have been described in *E. coli*. The DNA-binding proteins A, B and C (abbreviated as **DnaA**, **DnaB** and **DnaC** respectively) are the most important of these. Initial separation at the initiation site by *DnaA protein* is continued further by the hexameric *DnaB protein*, which is *the major strand separating protein*. It separates the DNA strands in both directions, resulting in the formation of two **replication forks** which move in opposite directions: one moves clockwise and the other counterclockwise, till they meet at the opposite side of

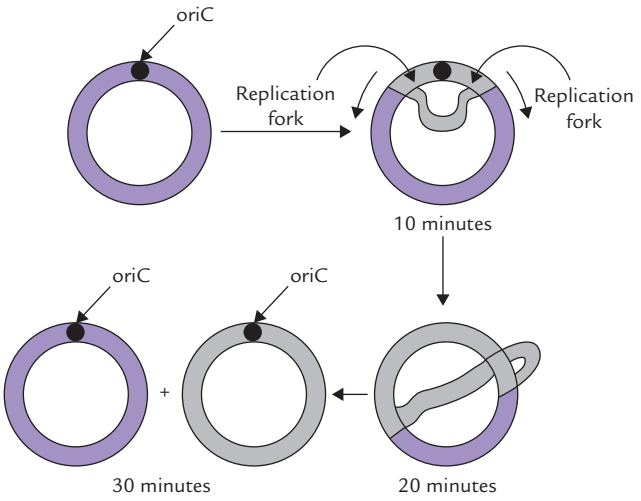


Fig. 21.10. Replication of the circular DNA of *E. coli*. Replication proceeds bidirectionally from a single replication origin (oriC).

chromosome (Fig. 21.10). *DnaC* is required for loading the DnaB at origin.

Single-stranded Binding Proteins (SSB Protein)

Once separated, the single-stranded regions associate with SSB protein, which *prevents reannealing*. This protein has high affinity for single-stranded DNA but has no base sequence specificity—it can bind anywhere on the separated strand. The regions of single-stranded DNA stabilized by SSB protein are rigid and semi-extended, which is an appropriate configuration for its role as a template for DNA replication. The SSB exists as tetramer, and binding of one tetramer facilitates binding of another to an adjacent region of single-stranded DNA (i.e. cooperative binding).

As the *helicase* moves in advance of the replication fork, the SSB proteins move on and off the DNA and are recycled for use on another site.

DNA Gyrase

The strand separation at the replication fork applies a turning force to the double helix, which may soon result in positive supercoiling (overwinding) of the remaining unseparated DNA and cessation of further separation. This is prevented by *DNA gyrase*, a type II *DNA-topoisomerase* that *relaxes positive supercoils passively* and *introduces negative supercoils by an ATP-dependent mechanism*.

Gyrase, SSB and *helicases* are collectively known as **unwinding proteins;** together they create single-stranded DNA for replication. A diagrammatic representation of their action is given in Figure 21.11 and summarized in Table 21.4.

Role of each of these proteins is of vital significance. Disrupted action of a protein may sometimes halt the prokaryotic replication altogether, thereby endangering growth and survival of the cell (**Case 21.2**).

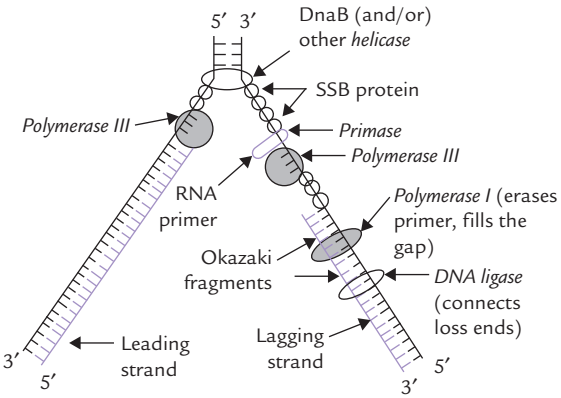


Fig. 21.11. Activities at the *E. coli* replication fork.

Table 21.4. *Escherichia coli* DNA replication proteins

Protein	Function
Dna A	Recognizes and binds specifically at replication origin (leading strand).
Dna B	Major strand separator (lagging strand).
Dna C	Required for loading Dna B at origin.
SSB	SSB protein prevents reannealing.
Primase	RNA primer synthesis.
DNAP III holenzyme	Progressive chain elongation.
DNAP I	Gap filling, primer excision.
Topoisomerase I	Removes negative supercoils.
Topoisomerase II	Forms negative supercoils, ATP dependent.
Ligase	Connects Okazaki pieces, ATP-dependent.

SSB = single-stranded binding, DNAP = *DNA polymerase*.

To initiate replication, parental strands are first melted apart at a specific region and further separated in an ATP-dependent process that requires a *helicase*. SSB prevents reaggregation of the single strands and the *gyrase* relaxes positive supertwists.

Thus, a coordinated action of several proteins ensures specific initiation at the precise location. Regulation of initiation is important because replication occurs only once in a cell cycle.

The following points merit re-emphasis at this stage.

- Separation of the parental strands and synthesis of the new strands occurs simultaneously during replication. The site of simultaneous separation and synthesis is known as *replication fork*.
- *Replication is bidirectional*, which means that it moves simultaneously in both directions from the initiation site. In other words, there are two replication forks: one moves clockwise and the other anticlockwise.

Because of the small size of the bacterial genome and fast speed of replication (up to 3 billion base pairs may be accumulated in a few hours), the two replication forks are sufficient to duplicate the DNA in only about 30 minutes.

These concepts were first demonstrated by *Cairns;* using techniques, discussed in Box 21.2.

Elongation

The principal enzymes for the synthesis of new strands are ***DNA polymerases*** (DNAP) which catalyze polymerization of deoxyribonucleotides under direction of DNA template. These enzymes catalyze the stepwise addition of deoxyribonucleotide residues to the free 3′-hydroxyl end of a preexisting DNA or RNA primer strand; thus the replication is propagated in the **5′ → 3′** direction. The overall reaction is

$$(\text{Deoxyribonucleotide})_n\text{-3′-OH} + \text{dNTP} \longrightarrow (\text{Deoxyribonucleotide})_{n+1}\text{- 3′-OH} + \text{PPi}$$

dNTP = Precursor deoxyribonucleotide; PPi = Pyrophosphate

DNAP needs participation of the following:

DNA template: Each separated parental polynucleotide strand serves as template and copied by *DNAP* (in complementary sense) according to Watson–Crick base pairing rules.

Precursors: Any of the four deoxyribonucleotides-5′-triphosphate (dATP, dGTP, dCTP and dTTP) can serve as a precursor molecule for DNA synthesis. In the above equation, dNTP represents a precursor and PPi represents pyrophosphate cleaved from the dNTP.

Primer: *DNAP* is incapable of assembling the first nucleotide of a new chain. It needs a pre-existing oligonucleotide sequence called *primer*, containing a 3′-hydroxy group at end (Fig. 21.12). This group attacks the innermost phosphate of the incoming precursor molecule and forms a phosphodiester bond at the end of the growing chain. This complex reaction is driven forward by two factors:

(a) ***Pyrophosphate cleavage:*** The second and third phosphates of the precursor nucleotide form pyrophosphate, which is cleaved rapidly to inorganic phosphates by cellular *pyrophosphatases*. The hydrolysis of pyrophosphate increases the negative $\Delta G^{0'}$ value for the synthesis, thus helping to drive the reaction.

(b) ***New hydrogen bond*:** Incorporation of a nucleotide into the new strand of DNA involves the formation of new hydrogen bonds with its template partner. There occurs liberation of free energy in this process, which adds to the thermodynamic drive of the process.

At the cost of over-emphasizing a point (for it is important): DNA is always elongated in the **5′ → 3′ direction**. The direction of elongation refers to polarity of the new strand and not of the template strand, which has an opposite polarity. The polarity of DNA strands has been explained earlier.

DNA polymerases bring about template-directed synthesis of the new DNA strand stepwise, nucleotide by nucleotide, in the 5′ → 3′ direction.

***Processivity*:** *DNA polymerase* has capacity to remain associated with the template, catalyzing addition of several

BOX 21.2

Autoradiography and Denaturation Mapping

Autoradiography: It is a process in which a photographic emulsion or X-ray film records the position of radioactive entities, such as proteins or nucleic acids, that have been immobilized in the matrix (e.g. nitrocellulose membrane or electrophoretic gel). It is possible to visualize the DNA through this technique by synthesizing the molecule that contains the radioactive isotope in place of the normally occurring isotope. *E. coli* cells are first grown in a medium containing thymidine labelled with the radioactive hydrogen isotope, **tritium (^{3}H)**. In this way, the DNA in these cells is made radioactive because after a few generations, the bacterial DNA incorporates the tritiated thymidine.

DNA is carefully isolated from such cells and is overlaid with photographic emulsion so that an image of the DNA is produced on the latter. The resting DNA produces a circular image on the emulsion. Image of a replicating DNA resembles the Greek letter Θ, which has circular shape with a loop. This structure shows that the duplex molecule maintains its circular shape during replication. Long stretches of single-stranded DNA are not seen in the theta structure, which suggests that parental strands are not unwound completely before serving as template for synthesis of new DNA. Rather new DNA synthesis is closely coupled to unwinding of parental DNA.

Denaturation mapping: This technique proves that replication is always initiated at a unique point (e.g. *oriC*). When temperature of the medium was increased, the A-T base pairs throughout the duplex molecule start getting separated because of their lower melting point. The separated regions were visualized as single-stranded "**DNA bubbles**" upon autoradiography. Since the DNA bubbles are formed at the sites of A-T bases, they form a reproducible pattern. They may serve as reliable reference points. Using this technique, it was found out that replication always initiated at specific origin only.

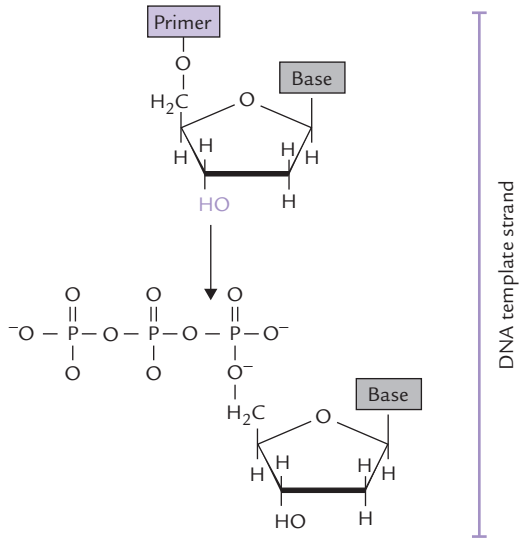


Fig. 21.12. Chain elongation reaction directed by the template strand and catalyzed by *DNAP.*

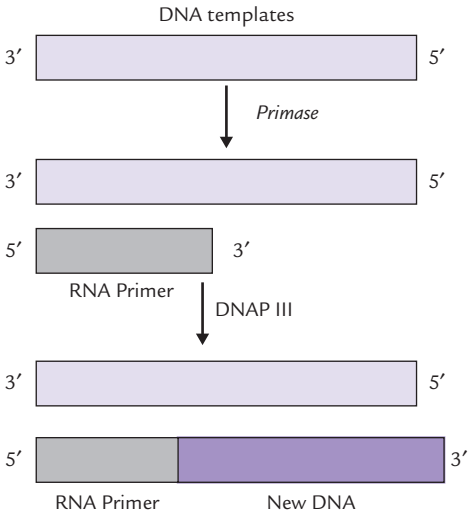


Fig. 21.13. *DNA polymerase-III* cannot assemble the first nucleotide of a new chain. This task is left to *primase*, which synthesizes a short sequence of RNA that serves as the primer for *DNAP III* (DNAP = *DNA polymerase*).

(nearly 100) nucleotide units to the growing daughter strand per binding event. This property, called processivity, enhances efficiency of replication several-fold.

***Role of primase*:** Despite the fact that *DNA polymerase* can read the base sequences of their template and incorporate "correct" bases with incredible accuracy, it cannot proceed until an RNA primer with a 3′-OH terminus is formed. A specific class of enzymes called *primases* accomplishes this task. (Table 21.4). A *primase* forms a short stretch (approximately 3–5 ribonucleotide units) of RNA in the replication fork. This small RNA, base paired with the template strand, is extended by the DNAP III (Fig. 21.13).

The primer is hydrolyzed soon after by *DNAP I* (discussed later). The same enzyme then fills the gap so created with a new deoxyribonucleotide (DNA) sequence. ***DNA ligase*** then joins the loose ends.

***Leading and lagging strands*:** A major problem, which remained unanswered for long, concerned the direction of DNA synthesis. It arises from the following observations: that DNAP can synthesize a new chain only in the 5′ → 3′ direction, reading the template in 3′ → 5′. If synthesis always proceeds in 5′ → 3′ direction, then how can simultaneous synthesis be possible in two strands of a DNA molecule which are antiparallel?

This problem was answered by *Okazaki and colleagues* in 1960s. These workers found that in only one of the strands, continuous synthesis takes place in 5′ → 3′ direction, which is also direction of movement of the replication fork. This strand is termed as the **leading strand.** In the other strand, known as the **lagging strand,** synthesis takes place in short DNA fragments, named **Okazaki fragments.** As the replication fork moves, the *primase* comes into action repeatedly, each time synthesizing short oligonucleotide sequences which serve as primers for DNAP III. DNA fragments, each approximately 1000 nucleotides long, are then added by DNAP III at the 3′ ends of these primers. The RNA primer is removed soon by *5′-exonuclease* activity of DNAP I and the gap is filled by its *polymerase* activity: it synthesizes a fragment of DNA that replaces the RNA primer.

Finally, the enzyme ***DNA ligase*** catalyzes formation of a continuous strand of DNA by forming phosphodiester bonds between loose ends (Fig. 21.11). The loose ends refer to DNA synthesized by *DNAP III* and the small fragments of DNA synthesized by *DNAP I*. The bond formation requires energy provided by pyrophosphate cleavage of ATP.

At each replication fork, the leading strand is synthesized continuously in direction of unwinding while the lagging strand is synthesized in opposite direction as discontinuous segments (Okazaki fragments) that are later joined.

Primosomes for repeated initiation of Okazaki fragments: To function effectively, *primase* needs to be complexed with various other proteins to form a *primosome (~600 kD)*. Various components of a primosome work cooperatively to prime for the repeated initiation of the Okazaki fragments. The complex includes a *helicase* (DnaB) and *primase*. It moves processively in the 5′ → 3′-direction on the lagging strand to repeatedly displace SSBs, recognize start loci and synthesize RNAs.

A *primase*-containing-primosome synthesizes a RNA primer: a primosome is required to initiate each Okazaki fragment.

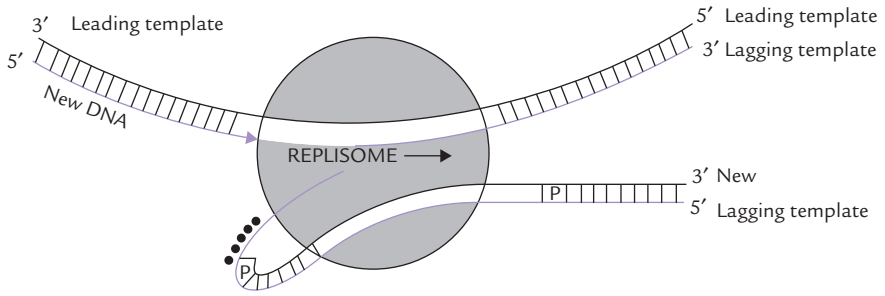


Fig. 21.14. Replisome, which contains a dimeric *DNA polymerase* holoenzyme, synthesizes both the leading and the lagging strands. The lagging template must loop around to permit the holoenzyme to extend the primed lagging strand (P = RNA primer, •••• = single strand binding proteins).

Replisomes for simultaneous synthesis of leading and lagging strands: Though the leading strand is synthesized continuously in the direction of unwinding, and the lagging strand in an opposite direction, it has been proposed that more or less simultaneous synthesis of both the strands occurs at replication fork. A ***dimeric*** *DNA polymerase III* holoenzyme, a primosome, a *helicase*, and replication proteins may form a complex termed **replisome.** Simultaneous synthesis of both the strands occurs by the action of a replisome if the DNA of the lagging strand template is looped backward so that it is in the same orientation as the leading strand template during polymerization.

As evident from Figure 21.14, the orientation of the lagging strand template (passing through the *polymerase* site in one subunit of the dimeric *DNA polymerase III* holoenzyme) is in the same direction as the leading strand template in the other *polymerase* subunit. After the addition of about 1000 nucleotides, the *polymerase* would release the lagging strand template to allow formation of new loop. *Primase* would synthesize a short RNA primer, and the process would be repeated. In this way, the lagging strand would not be too far behind the leading strand synthesis.

Termination

The two replication forks moving in opposite directions from the origin meet at the opposite side of the circular DNA. This is the terminal event in DNA replication. However, it is still not clear what causes the ultimate separation of the daughter strands from the parent strand. In *E. coli*, possibly *topoisomerase IV* is involved in this process. Precise mechanism of the termination is still under investigation.

Replication forks move bidirectionally from the origin of replication until adjacent forks fuse at the opposite side of the chromosome, and thus the replication is completed.

Replication is a fast process. *DNAP III*, the principal enzyme of chain elongation, adds about 600–1000 nucleotide units per second. Replication of whole length of circular DNA of *E. coli*, which is 1.3 mm long—approximately 1000 times the diameter of the cell—takes just about 30 minutes.

Single-stranded organisms such as bacteriophages also undergo replication process, but by a different method, called **rolling-circle replication mode.** The details of this process are given in Box 21.3.

C. Inhibitors of DNA Replication

Inhibitors of DNA replication are classified into the general categories:

1. Inhibitors that interact directly with DNA.
2. Inhibitors of deoxyribonucleotide synthesis.
3. Inhibitors that interact with replicative enzymes involved in DNA synthesis.

A large number of compounds in the first category bind between the stacked bases of the DNA duplex, disrupting the normal structure of DNA and causing some unwinding. This inhibits the DNA from serving as template for both replication and transcription. Examples of such compounds include acridine, ethidium and actinomycin. Mitomycin cross links adenine and guanine on DNA strands and prevents unwinding.

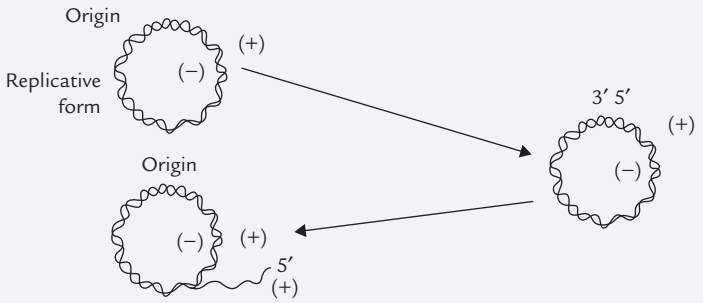
In the second category are compounds acting at various sites in the pathways for purine and pyrimidine biosynthesis. Actions of some of these, such as 6-mercaptopurine (inhibitor of purine synthesis) and 5-fluorouracil (inhibitor of thymidylate synthesis), and their role in treatment of cancer and infectious diseases have been discussed earlier in Chapter 20. Some compounds such as hydroxyurea and sulfonamides inhibit the synthesis of both purine and pyrimidine nucleotides.

In the third category are such compounds that inhibit *DNA polymerase* III (e.g. arylhydrazinopyrimidines) or *DNA gyrase*, e.g. norfloxacin and ciprofloxacin that are used as antibiotics. Others, such as ethylmaleimide, aphidicolin, butyl-phenyl-dGTP, act on enzymes of eukaryotic replication (Chapter 24).

BOX 21.3

Other Modes of DNA Replication

Several bacterial and viral genomes are replicated by the leading-lagging strand mechanisms. In addition, other modes of DNA replication for small circular chromosomes are also known. For example, some bacteriophages contain a single-stranded circular DNA, the (+) strand. A complementary strand to it, the (–) strand, is synthesized by a mechanism resembling conventional leading strand synthesis to form a circular duplex DNA. The latter is known as the replicative form. Additional copies of the (+) strand can be synthesized by the **rolling-circle replication mode.** This process initiates at a single strand break in the (+) strand and uses the (–) strand as a template. As the new (+) strand is synthesized, the existing (+) strand gets displaced by it. The replicating structure now resembles the Greek letter, sigma (σ), and therefore, this mechanism is commonly known as the replication mode σ. Multiple rounds of rolling-circle replication can generate a large number of tandemly linked (+) strands. These are later separated by an *endonuclease* for packaging into individual phage particles.



D. Types of *DNA Polymerases*

Prokaryotes have three different *DNA polymerases* (I, II and III), each designed for a different task.

DNA Polymerase I was the first DNAP enzyme to be isolated and was initially thought to be principal enzyme for replication. However, soon after its isolation in 1955, evidence began to accumulate that it alone cannot account for the observed speed of replication, and that other types of *DNA polymerases* must be present. The reason for this being, first, the rate at which this enzyme accumulates nucleotides (20 per second) is much slower than the observed rate of replication (3 billion base pairs added in a few hours). Second, it has lower processivity. Third and most importantly, when activity of this enzyme was altered in some bacterial strains, the affected bacterium still retained ability for an independent existence.

This observation by *John Cairns* intensified search for other *DNA polymerases*, which soon led to discovery of *DNA polymerase II* and *DNA polymerase III* in the early years of 1970s. Further, these studies helped to define the exact role of the DNAP I. Because of its low processivity, DNA I tends to fall off its template after polymerizing only a few dozen nucleotides. ***Its primary role is in DNA repair****, playing an accessory role in DNA replication, where it erases primers and fills gaps* (Fig. 21.11).

DNA Polymerase II has intermediate processivity. Though its exact role is not known, it is believed to participate in **DNA repair**.

DNP Polymerase III is the major enzyme of DNA replication. Its processivity is high: it can catalyze polymerization of thousands of nucleotides at a rate of almost 1000 nucleotides per second.

Comparative properties of the three *DNA polymerases* have been shown in Table 21.5.

E. *Exonuclease* Activities of *DNA Polymerases*

Nuclease refers to an enzyme that catalyzes hydrolysis of a phosphodiester bond in a nucleic acid. The *nucleases* that cleave the internal phosphodiester bonds are termed ***endonucleases*** and those that cleave bonds at ends are called ***exonucleases***. Some *exonucleases* cleave only at the 3′ end (the *3′-exonuclease* activity) while others cleave at the 5′ end (the *5′-exonucleases* activity).

The 3′-exonuclease activity: All the three types of bacterial DNAPs possess the *3′-exonuclease* activity (Table 21.5), meaning that they can cleave the phosphodiester bond starting from the 3′ end of the chain. This activity provides a means for *proofreading*: any wrong base, mistakenly incorporated by DNAP, is promptly removed. It is important to correct such mistakes since wrongly incorporated base can lead to permanent change in the genetic information, which can have disastrous consequences. The *3′-exonuclease* activity prevents any such eventuality by removing the mismatched base from the 3′ end (Fig. 21.15). The DNAP

Table 21.5. Comparison of *DNA polymerases*

	DNAP I	DNAP II	DNAP III
Subunits	1	4	8
3′-Exonuclease	Yes	Yes	Yes
5′-Exonuclease	Yes	No	Yes
Speed (bases per sec)	16–20	<7	600–1000
Processivity	3–200	~10,000	~500,000
Molecules per cell	400	100	15

then replaces the mismatched base with the correct one, and the replication continues.

It is noteworthy that DNAPs are very accurate enzymes rarely making mistakes; the error rate being 1 in 10^4. The proofreading mechanism can reduce this rate to about 1 in 10^7.

The 5′-exonuclease activity: *DNA polymerase I* can cleave phosphodiester bond of DNA, starting from the 5′ end of a chain. This is *5′-exonuclease* activity, which is different from the *3′-exonuclease* activity. First, the cleavage can occur at a bond several residues from the 5′ end. Second, the active site for 5′ → 3′ hydrolysis is different from that for 3′ → 5′ hydrolysis.

The *5′-exonuclease* is commonly referred to as the *error correcting activity* since it is required for removing the damaged DNA during DNA repair. For example, excision of pyrimidine dimers formed by exposure to ultraviolet light, as discussed later. The *5′-exonuclease* activity plays a key role in DNA replication by removing RNA primers.

DNA polymerases have *exonuclease* activities, which enable them to carry out their physiological functions of excising RNA primers to replace them with new DNA and participating in the repair of damaged DNA.

VII. Mutations

Mutations are *permanent, heritable changes in the base sequence of genomic DNA*. The most common changes are **substitution**, **addition**, **rearrangement**, or **deletion** of one or more bases. Mutations can either occur spontaneously or result from

(a) damage to the DNA by such external agents as radiation and chemicals,
(b) from errors during repair of damaged DNA, or
(c) even from replication-errors.

Although mutations provide the raw material for natural evolution, they may lead to severe and debilitating

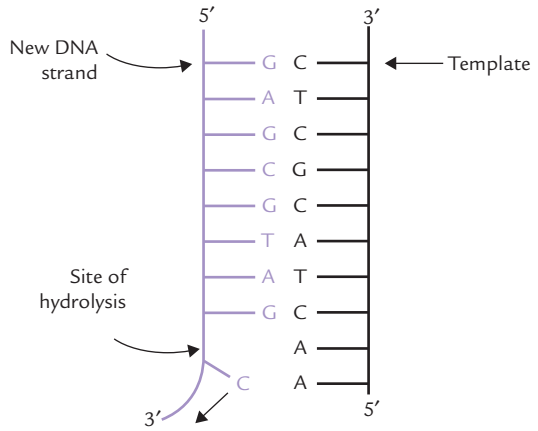


Fig. 21.15. The *5′-exonuclease* activity of *DNA polymerase*.

diseases including cancers, and contribute to normal aging process. They lead to a genetic disease when they occur in the germline and to a defective cell clone when they occur in a somatic cell. The former (*germline mutations*) are perpetuated in a heritable manner in the genome of an organism, and the latter (*somatic type*) results in production of cells with reduced viability or impaired functions.

Mutations are rare and random events as far as individual organism is concerned. The probability of a mutation occurring in lifetime of an *E. coli* cell is only one chance in 10^9. For a human cell, the probability is greater (1 in 10^5). Mutations of more than 2500 human genes have been detected to date, many of them impair biological functions or are even lethal. For instance, a single base alteration produced by certain carcinogens in cigarette smoke may lead to cancerous transformation of the affected cell (**Case 21.1**). However, not all mutations give rise to a mutant phenotype; some remain silent and some are beneficial. Also, some mutations (Box 21.4) suppress harmful effects of some other mutations. In the following sections, the molecular basis of mutagenesis, and the DNA repair mechanisms that keep the mutation rate within tolerable limits are discussed.

A. Types of Mutations

Following types of mutations are known to occur.

Point Mutations

These are *changes of a single base pair* of the DNA and are most common of all mutations. **Transitions** are point mutations in which one purine is substituted for another purine, or one pyrimidine is substituted for another pyrimidine. **Transversions** are point mutations in which a purine is replaced by a pyrimidine, or vice versa. When such base substitutions occur on the coding region of

BOX 21.4

Mutations in Anticodons of tRNAs

Mutations in anticodons of tRNAs have been observed among prokaryotes and lower eukaryotes, which may alter their base pairing characteristics so that a different codon is recognized. For instance, if anticodon of tRNA for tyrosine is mutated from GUA to CUA, this mutant tRNA would recognize UAG (instead of UAC) as a codon for tyrosine. This may suppress harmful effect of a point mutation on a structural gene, e.g. UAC (tyrosine) changing to UAG (stop). Normally, this mutation would result in premature chain termination, but a mutated tRNA would not recognize UAG as a stop codon, but rather respond by incorporating tyrosine. In this way, the mutated tRNA neutralizes a hazardous mutation in a structural gene. It is referred to as a *suppressor tRNA* and the mutation as *suppressor mutation*.

DNA, it can change the amino acid sequence of the encoded protein and the following manifestations are possible:

1. *Missense mutation* leads to a changed codon, which encodes a different amino acid, so that a single amino acid change appears in the encoded polypeptide. For example, replacement of A by C in the triplet UAU results in replacement of tyrosine by serine:

   5′-UAU-3′ ⟶ 5′-UCU-3′
   Tyr Ser

   Depending on the amino acid that is changed, missense mutations can have either no effect or very serious consequences. Sickle cell anaemia, a severe haemolytic disease, is an example of a missense mutation with serious consequences.
2. *Silent mutations* arise due to degeneracy of genetic code, discussed later (Chapter 22). In this instance, the point mutation produces a codon that still codes for the same amino acid. For example, replacement of T by C in the first position of the triplet (TAA) does not produce any change in the amino acid encoded:

Genetic code	5′TAA-3′	5′-CAA-3′
mRNA codon	3′-AUU-5′	3′-GUU-5′
Amino acid	Leucine	Leucine

The point mutations occurring in the non-coding regions of DNA are also silent.

3. *Nonsense mutations* are the ones where single base substitution causes generation of a stop codon, UAA, UAG or UGA. As a result, premature termination of translation occurs with complete loss of function of the truncated protein.
4. *Favourable mutations* are the ones which may produce useful proteins. Such mutations are made use of in agriculture to yield better varieties, such as tryptophan-rich maize (normal maize is deficient in tryptophan).

Insertion or Deletion of Base Pairs

Insertion and deletion are mutations in which bases are added to the normal sequence or removed from it. Both may lead to **frameshifts** (if the number of base pairs is not a multiple of 3), **changing the reading frame** of all codons beyond the point of insertion or deletion (Fig. 21.16). This changes the amino acids beyond the insertion/deletion so that the encoded polypeptide chain is completely abnormal. Frameshifts not only change the amino acid sequence but also generally lead to **premature termination** (or rarely elongation) of the encoded polypeptide. This occurs when stop codons are generated or removed by the frameshift. The insertion or deletion of three base pairs, or any multiple of three, does not result in a frameshift, and the resulting protein is more likely to retain part or all of its biological activity.

Frameshift mutations may be induced by some chemicals including **acridine** (Fig. 21.17) and proflavin, which have planar, fused ring structures that intercalate (insert themselves) between the adjacent stacked DNA bases. Figure 21.17 shows structure of acridine, which has some resemblance to purine base. When replication occurs in the region of an intercalated molecule, one or both daughter strands are synthesized that either lack one or more nucleotides or have additional ones.

Frameshift mutations may also result from insertion of large chunks of genetic material, such as transposons.

Triplet Expansion

It is a newly discovered type of mutation that leads, by an unknown mechanism, to a great increase in the number of triplet repeats. Once triplet expansion reaches a critical size, they begin to interfere with gene function and then result in clinical syndrome. Some genetic diseases are caused by such trinucleotide amplification within the confines of an important gene (Chapter 24).

The site where a mutation occurs has a significant effect on the protein product. Mutations in a promoter or enhancer leaves the structure of polypeptide intact, but

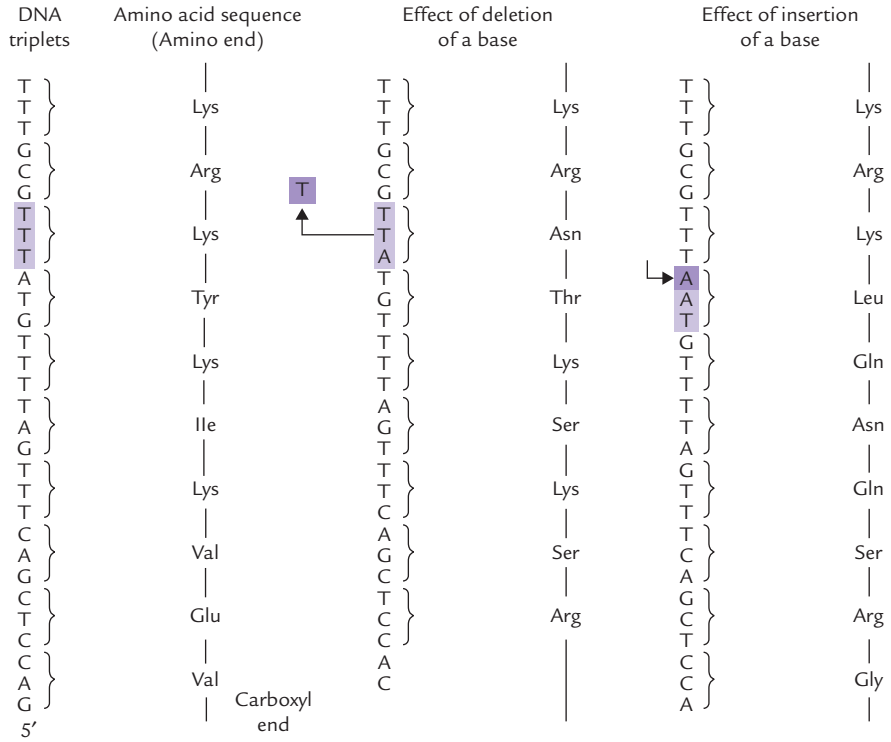


Fig. 21.16. Frameshift mutations caused by a deletion or an insertion of a base. Note the change in amino acids from the site of deletion/insertion and onwards.

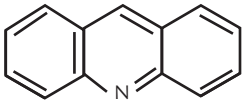


Fig. 21.17. Acridine, a mutagenic agent, produces frameshift mutation by virtue of its flat, ring structure that resembles a purine base.

can decrease (or sometimes increase) its rate of synthesis. Mutations in an intron-exon junction results in abnormal splicing and hence production of a non-functional protein (Chapter 24).

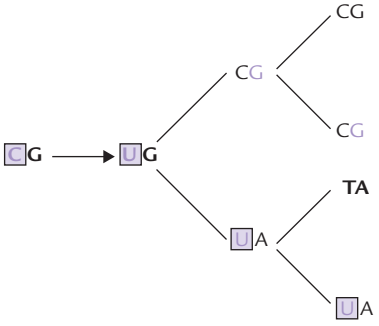
Finally, **translocation** (transfer of genetic material from one chromosomal location to another) may disrupt an important gene to cause disease.

B. Mutagens and Mutagenesis

Mutations may occur following exposure to physical agents or chemical reagents, collectively called ***mutagens***. For example, hydroxylamine reacts with a DNA base, changing its chemistry and hydrogen bonding properties, and therefore, is a mutagen. **Mutagenesis** is the process of producing a mutation. It may occur spontaneously or may be induced by mutagens. If it occurs spontaneously in nature without the action of a known mutagen, it is called ***spontaneous mutagenesis*** and the resulting mutations are *spontaneous mutations*, also called *basal mutations*. If a mutagen is used, the process is called ***induced mutagenesis***.

Basal Mutagenesis (or Spontaneous Mutations)

Several bases lose their exocyclic amino groups spontaneously, resulting in amino- to keto-group conversion. Thus cytosine is converted to uracil (U) which can form the base pair UA. In a later round of replication **CG to TA mutation** may follow.



- Adenine is likewise converted spontaneously to hypoxanthine (H), which can form the base pair, HC. This may lead to error of replication (**AT to GC mutation**) in a later round.
- Likewise, spontaneous deamination of guanine to xanthine (X), followed by XC pairing may lead to mutation in a later round.

Presence of thymine (instead of uracil) in DNA minimizes deleterious effects of spontaneous mutations (Box 21.5).

BOX 21.5

Why does DNA Contain Thymine Instead of Uracil?

Three of the ribonucleotide bases present in RNA (adenine, guanine and cytosine) occur as deoxynucleotide bases in DNA also. The fourth ribonucleotide base, uracil (U), does not occur in DNA. Instead, a different base, thymine (T) is present. Synthesis of T occurs from U through expenditure of a large amount of metabolic energy. Since T has identical base pairing properties as U, why do cells bother to synthesize it at all? This enigma was solved by the discovery that cytosine readily undergoes deamination (either spontaneously or by nitrites) to form uracil. A mismatched base pair GU is thereby formed from G-C. If uracil were normal DNA base, the deamination of cytosine would be highly mutagenic because there would be no indication of whether the mismatched G-U base pair had initially been G-C or A-U. Since thymine is normal base in DNA any U in DNA is certainly a deaminated cytosine and can be removed by *uracil N-glycosylase*.

Uracil N-glycosylase has another important function in DNA replication: it corrects a mistake which might have been committed by *DNA polymerases*. The latter cannot discriminate between dUTP and dTTP very well, both of which can base pair with a template A. Since dUTP is present in all cells in small amounts, newly synthesized DNA contains an occasional U. The 'U's are rapidly replaced by 'T's through excision repair.

Induced Mutagenesis

Radiations

Approximately 10% of all DNA damages caused by non-biological agents are because of ultraviolet or ionizing radiation.

(a) **Ultraviolet (UV) radiations** (wavelength 200–400 nm) from sunlight or tanning lamps induce dimerization of adjacent pyrimidines, particularly adjacent thymines, along one strand of the DNA to form (*cis*-syn cyclobutane) *pyrimidine dimers* and other photo-products. This distorts the DNA structure, inhibits transcription and disrupts replication.

(b) **Ionizing radiations,** including X-rays and radioactive decay, possess sufficient energy to displace electrons from their orbitals, creating highly reactive intermediates that react with DNA bases to cause strand breaks. They also generate reactive oxygen species that damage DNA. Single-stranded breaks are more common, though rarely double-stranded breaks also occur. Ionizing radiations are capable of penetrating deep and therefore cause both somatic and germline mutations.

Chemicals

Several chemicals alter DNA bases or the structure of DNA, which often leads to mutations. Types of chemical mutagens are:

1. *Deaminating agents*, particularly nitrous acid and the compounds that can be metabolized to nitrous acid or nitrites (Fig. 21.18a) convert amino groups to keto groups by oxidative deamination. Thus, cytosine, adenine and guanine are converted to uracil, hypoxanthine and xanthine respectively, which results in AT-to-GC or GC-to AT transitions, as discussed on page 458.
2. *Alkylating agents* (Fig. 21.18b) can cause chemical alterations in the bases so that the base pairing is changed. For example, *dimethylsulphate* can methylate a guanine residue to yield 6-methylguanine, which is unable to base pair with cytosine of the complementary strand. Alkylating agents are also capable of cross-linking the bases in the same (or opposite) strand.
3. *Base analogues* can erroneously get incorporated into DNA in place of a normal base. Some lead to inhibition of replication. Others, such as *bromouracil (BU)*, are mutagenic because they lead to mispairing. BU (Fig. 21.18c), an analogue of thymine, usually pairs with adenine. But through the influence of its electronegative Br atom, it frequently assumes a tautomeric form that base pairs with guanine instead of adenine. This may induce an AT-to-GC transition in subsequent round of replication. In this way a mutation is caused by a spontaneous tautomeric shift (keto to enol). Occasionally, BU is incorporated into DNA in place of cytosine, which generates a GC-to-AT transition.
4. *Intercalating agents*, such as *acridine* and *proflavin*, tend to cause frameshift mutations, discussed earlier.

Hydroxylamine is an interesting chemical mutagen, which reacts specifically with cytosine and converts it into modified base that pairs with adenine. This ultimately leads to GC-to-AT mutation.

Oxidative Radicals

DNA within the cell suffers from many environmental insults, the most serious of which is caused by reactive oxygen metabolites. In presence of these metabolites, the metal ions such as iron and copper catalyze the oxidation of DNA, which may lead to mutations. The damage sustained by DNA may range from oxidation of sugar

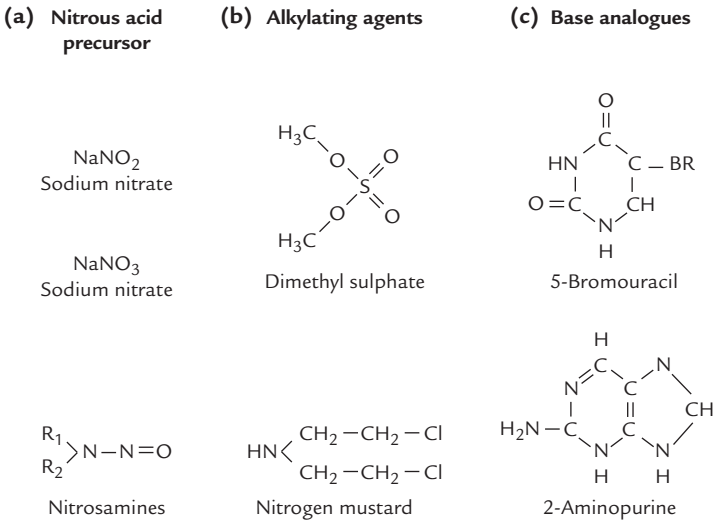


Fig. 21.18. Some chemical agents that may alter the structure of DNA or its base-pairing.

and base moieties of nucleotides to breaking of strands. Oxidative damage is increased in aging and in age-related degenerative diseases, including cancer. Approximately 20,000–60,000 modifications of DNA occur per cell per day.

C. Mutagenesis and Carcinogenesis

Most mutagens can cause cancer, and therefore, they are carcinogenic as well. Some mutagens that exist in the carcinogenic form, directly damage the DNA, whereas other mutagens are **pro-carcinogens**. Pro-carcinogens in their native form do not damage DNA, but are converted to carcinogens in the liver or in other tissues via a variety of detoxification reactions (e.g. oxidation by cytochrome P450). This process is called *metabolic activation* (see benzo(a)pyrene; **Case 21.1**).

Ames test for estimating carcinogenicity: The test, developed by Bruce Ames, an American biochemist, is based on the high correlation between carcinogenesis and mutagenesis. Ames constructed special test-strains of the bacterium *Salmonella typhimurium* that are *his*–, meaning that they cannot synthesize histidine. Therefore, they cannot grow in absence of histidine in the medium. The DNA of these histidine auxotroph strains contain nucleotide substitutions or deletions that prevents the production of *histidine biosynthetic* enzymes. Mutagenesis in these strains is indicated by their reversion to the ***his*** + **phenotype** (can synthesize histidine).

About 10^9 of the *his*– bacteria are spread on a culture plate lacking histidine and the suspected mutagen is added to the plate. Occasionally, the mutagen causes reversal of the histidine mutation to the *his* + phenotype; the latter can now synthesize histidine, and therefore, grow in the absence of histidine. This is detected by growth into visible colonies after 2 days at 37°C. Counting the number of such colonies that have reverted to the *his* + phenotype scores the mutagenicity of a compound. There is a significant correlation between the results of the Ames test and direct tests of carcinogenic activity in animals.

Addition of a small amount of liver homogenate, (which contains the liver P450 detoxification enzymes), in the Ames's test medium test may detect procarcinogens.

Mutations in nucleotide sequences arise spontaneously from replication errors, or induced by radiation and chemicals.

VIII. Repair Systems for Mutated DNA

An elaborate system exists in the body, which can repair most damages to DNA, thus preventing alteration in the stored genetic information. Despite its inherent stability, a number of chemical changes are occurring in DNA all the time, which may be spontaneous or triggered by 'insults' by a variety of agents. These changes may add up to result in large number of mutations per day in each cell if there were not constant repair systems. Repair is possible because usually only one strand of DNA is damaged at a given place, and there is always the other strand to act as template and direct the repair of the damaged part.

High fidelity of DNA replication is achieved (in addition to 3′ → 5′ proofreading) by DNA repair mechanism. Some forms of DNA damages, e.g. alkylated bases and pyrimidine dimers may be repaired in a single step, others are more complex.

The principal classes of repair systems and their essential features are discussed here:

Base Excision Repair and AP Site Repair

Some chemical changes, such as deamination convert normal bases into abnormal ones. These are removed from the DNA and replaced with the correct base by the base excision repair mechanism. Deamination of cytosine, for example, creates uracil, which is hydrolyzed off by *DNA glycosylase* through cleavage of the N-glycosidic bond (between the base and 2′-deoxyribose). This leaves *apurine* or *apyrimidine* (**AP**) *sites*, in which the deoxyribose has no base attached to it. Repair of the AP site involves:

- nicking of the polynucleotide chain adjacent to the lesion by AP *(apurinic)-endonuclease,*
- replacement of the section containing the nick by *DNA polymerase* I, and
- final sealing by *ligase* (Fig. 21.19).

Spontaneous depurination: AP sites can also be formed by spontaneous hydrolysis of the purine-deoxyribose link, which is somewhat unstable. It is repaired as above by successive actions of *AP endonuclease*, *DNAP-I*, and *DNA ligase*.

Mismatch Repair

The proofreading mechanisms (discussed earlier in this chapter) ensure that DNA is replicated with the almost incredible degree of accuracy, but in a system involving such vast numbers of nucleotide additions, mistakes invariably occur. Such mistakes (mostly base mismatches) cause some distortion of the duplex chain, illustrated in Figure 21.20.

The mismatches are repaired by the mismatch repair system, which has been described here in prokaryotes (and appears equally important in eukaryotes).

Strand discrimination: An important point is that the repair system must be able to discriminate between the parent (template) strand and the new daughter strand because it is the base on the daughter strand that is incorrect and so needs to be excised. Clearly, excising the newly synthesized base would preserve the genetic information, whereas excising the base on the parental strand would permanently alter the DNA, producing a mutation. Therefore, the challenge is to recognize the newly synthesized strand and remove the mismatched base on that strand.

How is such strand discrimination made? The repair enzymes identify the parent strand because it is methylated at the palindrome GATC (Fig. 21.21); a cytoplasm enzyme, *Dam-methylase*, methylates the adenine of this palindromic sequence. This *methyl group on the parent strand does not affect the base pairing or the DNA structure but serves as a tag by which the repair system identifies it as the* ***parent strand.*** When a new strand is synthesized, it takes some time to methylate it, and so, for this brief period it remains unmethylated. The repair enzymes recognize the methylated parent strand and leave it, but they act on the not-yet methylated new strand to remove the mismatched portion. The gap created by its removal is then filled by synthesis of new DNA and ligation.

Being guided by methylation, this system is referred to as *methyl directed mismatch repair system*. It correctly repairs mismatches that may be located as much as 1000 base pairs distant from a methylated palindromic (GATC) sequence. Further information on beneficial effects of methylation is given in Box 21.6.

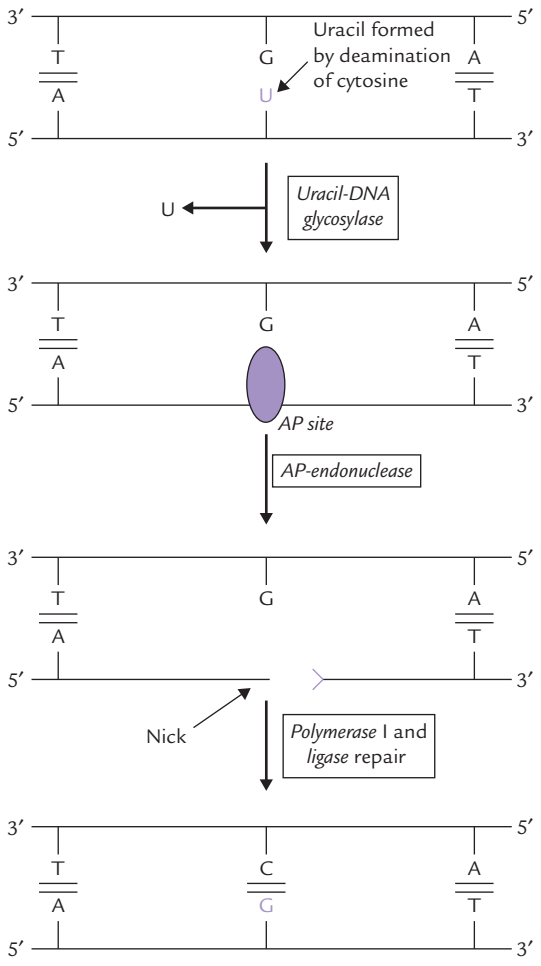


Fig. 21.19. Illustration of AP site formation and repair.

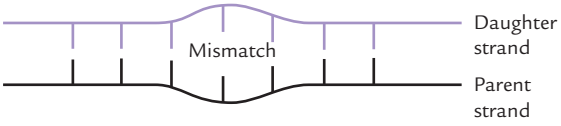


Fig. 21.20. Base mismatch in a double-stranded DNA, causing distortion of the duplex chain.

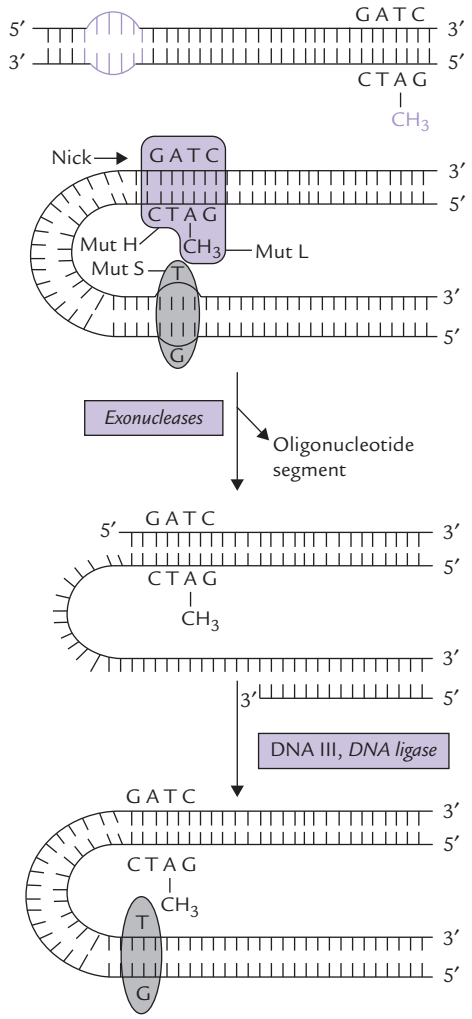


Fig. 21.21. Mechanism of mismatch repair. The mispaired base in newly synthesized DNA strand is shown by arrow.

Mechanism of repair: Exact mechanism of the mismatch repair is not clearly understood. Some proteins identified in this process are **Mut S**, **Mut H** and **Mut L**:

- **Mut S** protein recognizes the mismatch distortion in the helix.
- **Mut H** binds to an unmethylated GATC.
- **Mut L**, links the above two proteins to form a complex (Fig. 21.21).

Identifying the unmethylated strand as being newly synthesized, the Mut H nicks it. The entire section of DNA from the nick to beyond the error is removed by combined action of *exonuclease,* followed by the synthesis of new DNA (by *DNAP III*) and ligation (by *DNA Ligase*).

Removal of a segment, which may comprise thousands of base pairs, occurs just for correcting a single base. This increases fidelity of replication (by 1000-fold), which is needed to ensure continuity of life.

Nucleotide Excision Repair

This system repairs the bulky lesions that cause large distortions in the double helix, such as:

(a) T-dimers and other photoproducts,
(b) Adducts formed by covalent linking of some foreign molecules to DNA, and
(c) Some alkylated bases.

A short stretch of nucleotides, including the lesion is excised, which is followed by its correct replacement, the opposite strand serving as template for this (Fig. 21.22).

BOX 21.6

DNA Methylation

Any modification in the DNA bases can change the stored information and is, therefore, potentially hazardous. However, not all modifications are detrimental; for example adenine (A) and cytosine (C) of DNA may be methylated in a species specific pattern. Mostly only a few per cent of the susceptible bases are methylated in this manner. But in plants, this figure may rise to 30% of the C-residues. The methylated bases serve several useful purposes in different organisms, as discussed below.

In repair of the mismatched base pairs of the prokaryotic DNA, methylation serves as a useful marker of the parental strand. Distinction between the parent strand (that has correct bases) and the newly replicated daughter strand (having incorrect base) must be made so that the latter strand only is subjected to repair. The distinction is made by methylation.

Methylation in bacterial cells serves as markers that prevent the corresponding restriction *endonuclease* from degrading the DNA.

Eukaryotes

- Methyl groups project in the major groove of DNA, where they interact with the binding proteins.
- Methylation plays important role in control of gene expression. It turns off the gene expression, particularly when it occurs in the control regions upstream of the gene transcribed sequence. Thus, globin genes are less methylated in erythroid cells than in the non-erythroid cells, so that they are expressed more in the former.

The key enzyme in *E. coli*, an unusual *endonuclease* is termed Uvr *ABC endonuclease* complex, after the genes coding for the enzyme. It binds at the site of lesion and makes two nicks on the damaged strand. It is an ATP-dependent reaction which cleaves the lesion-containing strand at the seventh and fourth phosphodiester bonds on the 5′ and 3′ sides of the lesion, respectively. The oligonucleotide segment between the two nicks is then excised by the same enzyme. The excision gap is then filled by action of *DNA polymerase* I and ligated by *DNA ligase*.

The *ABC endonuclease* scans the DNA, recognizes the lesion, and removes the portion of damaged DNA. The resulting gap is filled by *DNA polymerase* and *DNA ligase*.

Further information on action of the *ABC endonuclease* is given in Box 21.7.

Direct Repair

Direct repair refers to those types where the repair occurs without removal of a base or a nucleotide. For example, repair of the **pyrimidine dimers** (produced by dimerization of adjacent thymines by UV rays) is effected by reversing this damage by *photoreactive photolyases*. The *photolyase* binds to the dimer and cleaves it after exposure to visible light. In fact, light exposure elevates the two cofactors present in it (FADH and methenyl-tetrahydrofolate) to a higher energy state. The excited form of FADH is capable of transferring an electron to the pyrimidine dimer, which breaks it.

Another example of direct repair involves the alkyl groups on bases, for example, the *O^6 alkyl derivatives of guanine*, (which are formed by some mutagenic agents). These may be directly repaired by transfer of alkyl group to cysteine side chain of a repair enzyme. In doing so the enzyme destroys its own action, hence acting as a **suicide enzyme.**

Recombination Repair

The recombination repair system may act when *damaged DNA undergoes replication before the lesion has been repaired*. It is an essential repair process for the rapidly dividing cells because a replication machinery promptly arrives at the damaged site, often before its repair (by nucleotide excision system).

The replication machinery halts when it encounters a lesion, such as pyrimidine dimer, and resumes polymerization at some point beyond the dimer block. The result of this process is that the daughter strand has large gap opposite the pyrimidine dimer (Fig. 21.23 a, b).

The gap is repaired by the process of recombination, which involves:

(i) Excision of an undamaged DNA segment from the parental DNA strand.
(ii) Insertion of this DNA into the gapped strand (Fig. 21.23c).

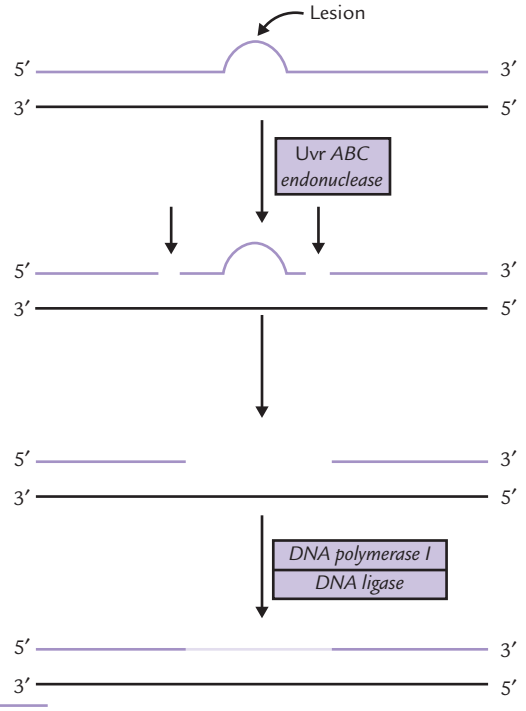


Fig. 21.22. Nucleotide excision repair by Uvr *ABC endonuclease* in *E. coli*.

BOX 21.7

Molecular Mechanism of Action of Uvr *ABC endonuclease*

The three constituent **subunits of the *ABC endonuclease*—*Uvr A*, *Uvr B*** and ***Uvr C***—perform distinct roles, which are well understood at molecular level. The **Uvr A** recognizes the lesion and unwinds the DNA at that site. It also causes conformational change in the Uvr B. Alteration in the conformation promotes tight binding of the Uvr B at the site of lesion. At this stage, the Uvr A leaves the *Uvr B-DNA complex*, which makes the complex a target for the Uvr C. **Uvr B** then makes a nick on 3′ side of the lesion, and **Uvr C** makes another nick on 5′ side of the lesion. The excised oligonucleotide segment is then released, and *DNA polymerase I* displaces Uvr B and fills the gap so created. *DNA ligase* completes the repair by a final seal.

Release of the excised nucleotide segment is facilitated by another subunit, **Uvr D**.

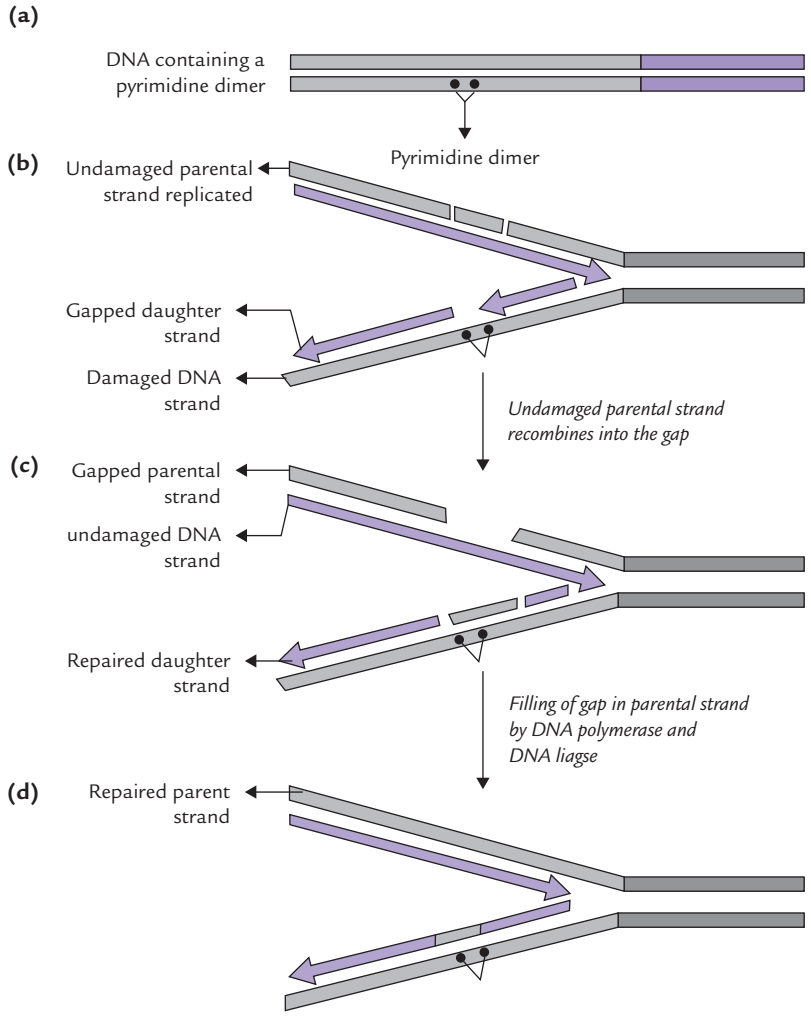


Fig. 21.23. Post-replication repair (•• indicates the site of the lesion in parental strand B).

(iii) A combined action of *polymerase I* and *ligase* joints this inserted piece to adjacent regions, thus filling in the gap.

(iv) This creates a gap in the parental strand, but it can be easily repaired since it is placed opposite an undamaged strand, and so can be filled-in according to the template, and then sealed (Fig. 21.23d).

(v) Finally, the pyrimidine dimer, associated with an intact complementary strand, can then be eliminated by nucleotide excision repair or photoreactive *photolyase*.

Since recombination repair occurs after DNA replication, it is often called **post-replication repair.** It, however, fails if two dimers in opposite strands are very near one another because no undamaged segments are available.

SOS Repair Response

Agents that extensively damage DNA induce a complex series of cellular changes in *E. coli,* known as SOS response. These changes result in production of proteins and enzymes that allow DNA chain growth across the damaged segments, though at the cost of fidelity of replication. The chain synthesis is possible because the editing system is relaxed, which means loss of the ability to remove incorrect bases added to the growing strand. As a result the process is *error prone* and, therefore, mutagenic.

The SOS response in prokaryotic cell induces a special set of survival enzymes that relax fidelity of replication, permitting the replication complex to synthesize DNA opposite the damaged template strand. The result is often mutations.

It may be mentioned that *DNA polymerase II* is one of the important proteins that are induced as a result of the SOS response. Unlike *DNA polymerase III*, which is stopped by such lesions, this enzyme can replicate past them. Precise function of the other proteins of this specialized system is unknown. In essence, special properties of DNAP II help the replication past the DNA lesions.

IX. Defective DNA Repair and Human Diseases

Inability to carry out certain stages of DNA repair results in human diseases discussed below.

1. *Xeroderma pigmentosum (*XP; Greek: *xeros*, dry + *derma*, skin) is a group of recessively inherited disorders due to mutations in genes encoding polypeptide(s) of the nucleotide excision repair. This system is responsible for the repair of bulky DNA lesions, including the pyrimidine dimers that are formed by UV radiation in sun-exposed skin. At least seven subtypes of XP have been identified in different patients, each caused by the deficiency of a different polypeptide in the repair system.

   Individuals with this disease are extremely sensitive to sunlight, developing freckles and ulcerative lesions on the sun-exposed skin. Moreover, they develop skin cancer at a 2000-fold greater rate than normal, usually during the first and second decades of life. In some subtypes, a variety of seemingly unrelated symptoms develop, such as progressive neurological degeneration, developmental deficits, and opacification or ulceration of the cornea. The only effective treatment for this otherwise fatal disease is complete avoidance of sunlight.
2. *Cockayne's syndrome (CS)* is a rare recessively inherited disease, also associated with defective nucleotide excision repair. In addition to the genes that are defective in XP, two additional genes are mutated in CS. Individuals with CS are hypersensitive to UV radiation, show growth retardation and neurological degeneration, but do not have an increased incidence of skin cancer.
3. *Bloom's syndrome, Fanconi's anaemia and ataxia-telangiectasis* are collectively referred to as **chromosome breakage syndromes** as they are thought to be caused by defects in the repair of double-strand DNA breaks. Fanconi's syndrome is a lethal aplastic anaemia, Bloom's syndrome is a premature aging disorder, and ataxia-telengiectasia is characterized by severe abnormalities in various organ systems and a high incidence of lymphoreticular cancer. Patients are highly sensitive to radiations, so their exposure to diagnostic X-rays should be minimized.
4. *Hereditary non-polyposis colon cancer (HNPCC)*, is a dominantly inherited cancer susceptibility syndrome in which the post-replication mismatch repair is defective. One of these mismatch repair genes (*hMSH2*) is located on the chromosome 3 and the other on the short arm of chromosome 2. HNPCC is responsible for approximately 6% of all cases of colorectal cancer; and risk of cancers of other tissues mildly increases, including the small bowel, stomach, biliary system and ovary.

Note: Repair systems are absent in mitochondria, resulting in mutations, which may accumulate with time to cause a large number of genetic diseases (discussed elsewhere). The mitochondrial mutations may also be related to the process of ageing.

X. Genetic Recombination

The exchange of polynucleotide strands between separate DNA segments is called genetic recombination. It requires cutting and covalent joining of DNA. There are two types of genetic recombination: site specific- and general-recombination.

A. Site Specific Recombination

It occurs between two short, specific DNA sequences. The process of recombination is effected by an enzyme that can recognize specific sequences either on one or on both DNA molecules (donor and target sites). For example, the *viral integrase*, which effects integration of the λ phage, recognizes both the phage DNA and the target site on the bacterial chromosome. Likewise, the *retroviral integrase* recognizes the LTRs of the retroviral cDNA. The *transposase* is likewise specific for the inverted terminal repeats of its own insertion sequence (see section on jumping genes).

B. General Recombination

It occurs between DNA segments with extensive sequence homology, hence also known as **homologous recombination.** It neither requires any specific enzyme systems nor any specific DNA sequences; it depends only on sequence similarity between the recombining molecules. Foreign DNA, for example, can be installed in a host's chromosome by this mechanism provided it has sufficient sequence homology with the chromosomal DNA.

A prototypic model for homologous recombination between homologous DNA duplexes was proposed by Robin Holliday in 1964. As shown in Figure 21.23, this type of recombination results in a reciprocal exchange between two duplex DNA molecules. The corresponding strands of two aligned homologous DNA duplexes are nicked, and the nicked strands cross over to pair with the nearly complementary strands on the homologous duplex, thereby forming a cross-shaped intermediate called the **Holliday intermediate** (Fig. 21.24). This intermediate is then resolved into two duplex DNAs by the successive action of an *endonuclease* (ruvC in *E. coli*).

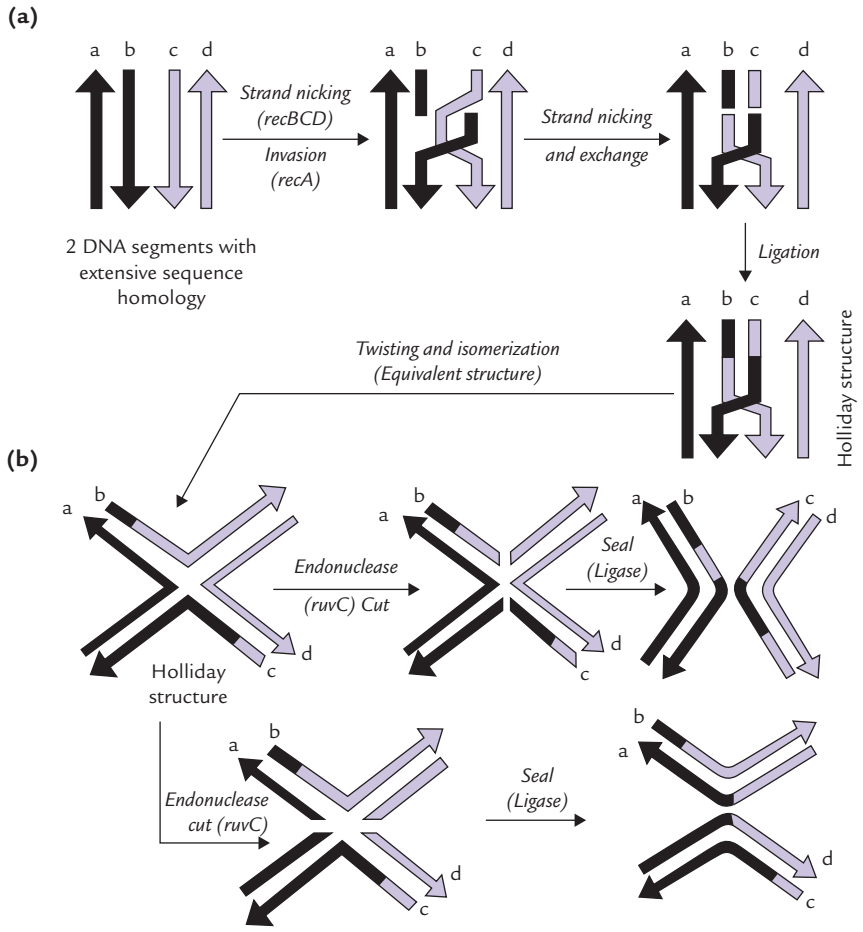


Fig. 21.24. Key steps in general recombination, (**a**) Formation of Holliday intermediate, (**b**) Resolution of Holliday intermediate.

Several enzyme proteins act successively to bring about the recombination:

1. A complex of recB, recC and recD proteins, which act as *endonuclease* and *helicase* to generate single-stranded DNA.
2. The recA protein, which catalyzes the invasion of a DNA duplex by the single-stranded DNA.
3. The ruvC protein resolves the Holliday intermediate.

General recombination involves exchange of the homologous DNA segments. A crossover structure (Holliday junction) is formed that can be subsequently resolved.

Homologous recombination is used for recombining genetic information in bacteria. In higher organisms, during *prophase* of meiosis, pairs of genes may exchange by crossing-over between homologous chromosomes, thereby promoting genetic diversity. Genes may also move from a given site to a non-homologous site by transposition; the mobile genetic elements are called transposons, discussed earlier in the chapter.

Though transposons shuttle among different locations, their movement does not seem to be of any advantage to the bacterial cell. Rather they can inflict deleterious effect by jumping into a useful gene and halting its expression. In view of these, mobile genes are, in general, thought to be selfish DNA elements.

Exercises

Essay type questions

1. Discuss the events which proved that DNA is the genetic material.
2. Outline the structural features of DNA double helix. What are the distinctive structural features of eukaryotic DNA?
3. What are the alternative higher order forms of DNA and how are they different from B- DNA? Discuss role of histones in DNA packing.
4. Outline the experiments that show that DNA replicates by semiconservative mechanism. Describe the mechanism of DNA replication.

5. Explain the basis of semi-discontinuous DNA synthesis. Discuss roles of primosomes and replisomes in replication.
6. Describe causes of DNA damage and the repair mechanism.

Write short notes on

1. DNA Supercoiling
2. Major and minor grooves
3. Unwinding proteins
4. DNA-Denaturation
5. *Topoisomerases*
6. Pseudogenes
7. DNA Sequencing
8. Jumping genes
9. Histones
10. Autoradiography
11. Primosome
12. Replisome
13. *Exonuclease* activities of *DNA polymerase*
14. Gene recombinations
15. Frameshift mutations
16. Nucleotide excision repair
17. SOS response

CLINICAL CASES

CASE 21.1 A chronic smoker with persistent cough

A 30-year-old man consulted a physician for having recurrent respiratory tract infections. He was well known to the clinic staff for his frequent visits because of problems related to chronic smoking. On one of his earlier visits to the clinic he had admitted to smoking 15 cigarettes per day, avoiding advice of the physician and the statutory warning on the cigarette packets that smoking would be injurious to health. He had been frequently treated with antibiotics, cough syrup and other usual measures for the respiratory infections. Mostly, he took these measures himself without consulting a physician. Presently he felt that the infection was more severe because even after 5 days of the above treatment, there were no signs of improvement.

X-Ray chest showed a nodule. Keeping the possibility of some cancerous transformation in mind, the physician referred him to the cancer wing of the local hospital. Sputum cytologic analysis and other investigations confirmed that the patient had cancer. Detailed analysis, including percutaneous fine needle aspiration (FNA) showed it to be small cell carcinoma (SCLC).

Q.1. How could the smoking habit of the patient result in cancer?

Q.2. The patient feels that his friend also smokes about the same number of cigarettes, but never developed such serious problem. Could genetic element be responsible for this?

CASE 21.2 A 25-year-old woman with high grade fever

A 25-year-old woman was investigated for having high grade fever. She was diagnosed as having insulin-dependent diabetes mellitus 15 years ago, and had been managed with insulin since then. Presently her blood glucose level (fasting) was 196 mg/dl, suggesting loss of control of diabetes. Urinolysis showed the following results:

Urine sugar	++
Urine albumin	++
Pus cells	+++/HPF

Gram negative bacteria were detected, which were rod shaped (e.g. Bacilli). These findings were suggestive of urinary tract infection which was further investigated by collecting a fresh urine sample and performing culture and sensitivity test. *E. coli* were grown on the culture plate and found sensitive to norfloxacin.

The patient responded well to the drug and her fever subsided on the third day after starting the treatment.

Q.1. What is the mechanism of bactericidal effect of norfloxacin?

Q.2. Are patients with diabetes mellitus more prone to get urinary tract infections? Give reason for your answer.

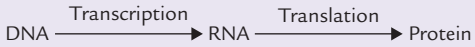
Q.3. Does genetic element play any role in development of diabetes? Explain.

CHAPTER

22

MOLECULAR BIOLOGY II: TRANSCRIPTION AND TRANSLATION

The genetic information encoded in DNA determines the kinds of proteins to be synthesized. However, protein synthesis cannot be immediately directed by DNA. Rather, it is RNA which is synthesized initially using DNA as a template (i.e. transcription). The RNA then serves as a template during protein synthesis (i.e. translation). Thus, the flow of genetic information in a normal cell is as follows:



Various types of RNA molecules are known. Some serve as information carrying intermediates during protein synthesis, whereas others are integral components of the protein synthesizing machinery. In this chapter, synthesis of RNA according to DNA template (**transcription**) in prokaryotes is considered first. This is followed by a description of **genetic code**, which defines relationship between base sequence of the RNA and amino acid sequence of the polypeptide. Finally, the process of **translation** of base sequence of the RNA into the corresponding amino acid sequence of the polypeptide chain is described.

After going through this chapter, the student should be able to understand:

- Types of RNA and distinctive features of the three types of RNAs (mRNA, rRNA and tRNA).
- *Transcription:* initiation, elongation and termination; roles of *RNA polymerase* (RNAP) and other protein factors involved in this process; post-transcriptional modifications and processing of RNAs; and inhibitors of transcription.
- *Genetic code:* general properties, concept of wobble; role of tRNA as an adaptor molecule in translation, and of ribosomes as organelles of protein synthesis.
- *Translation:* mechanism, initiation, elongation and termination in prokaryotes; post-translational modifications, protein targeting and inhibitors of protein synthesis.

I. Types of RNAs

The RNAs are polynucleotides, linked by phosphodiester bonds. They differ as a group from DNA in the following aspects:

1. They are much *smaller in size* (60–20,000 bp; DNAs often exceed 10^6 bp) and are mostly single-stranded.
2. The sugar they contain is *ribose* instead of *2′-deoxyribose* of DNA.
3. They contain a pyrimidine base, *uracil, instead of thymine* of DNA.

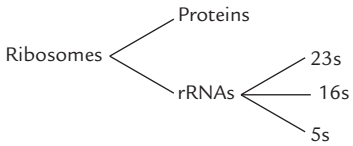
Their major role is in gene expression, in contrast to DNA which stores genetic information.

There are three major types of RNAs: **ribosomal RNA** (*rRNA*), **transfer RNA** (*tRNA*) and **messenger RNA** (*mRNA*). These types differ from one another in terms of size, structure and function.

Ribosomal RNA

rRNA plays important enzymatic and structural role, as constituent of ribosomes which are the protein synthesizing machines of the cell. The other components of ribosomes are protein molecules which are associated in

a complex way with the rRNAs. There are three distinct size species of rRNA in prokaryotic cells: 23S, 16S, and 5S (Table 22.1).



The same types are present in eukaryotic mitochondria, whereas in eukaryotic cytosol there are four size species (28S, 18S, 5.8S and 5S). Ribosomal RNA is most abundant of all types of RNAs, making up around 80% of the total RNA in the cell.

(**S** is the **sedimentation coefficient**, i.e. the sedimentation velocity per unit of centrifugal force. It is usually expressed in units of 10^{-13}S, which are known **Svedbergs**, named after the inventor of the ultracentrifuge. It is related to the molecular mass of the compound being subjected to sedimentation.)

Transfer RNA

This RNA (tRNA) is the smallest of the three major species of RNAs having 73 to 93 nucleotide residues and a sedimentation coefficient of 4S. The tRNAs contain certain **unusual bases** and have extensive intrachain base pairings. *They bind amino acids covalently and deliver them to the ribosome for protein synthesis.*

Further details about structure of tRNA and its role as an adaptor molecule in protein synthesis are discussed later in this chapter.

Messenger RNA

Of the three types of RNAs, mRNA is the most heterogeneous type in terms of size, though it comprises only 5% of the total RNA in the cell. Its sequence is complementary to the nucleotide sequence of the template DNA. Consequently, it carries "*working copies*" of genetic information, originally contained in the DNA, into the cytosol. Once in the cytosol, its nucleotide sequence is translated as a polypeptide chain; the relationship between its nucleotide sequence and the amino acid sequence of the polypeptide is defined by a genetic code.

Special structural characteristics of eukaryotic mRNA include

(i) a long sequence of adenosine monophosphates (*poly A tail*) on 3′ end of the RNA chain, plus
(ii) a methylated guanine nucleotide "*cap*" attached to the 5′ end through a unique 5′-5′ triphosphate linkage.

Note: Only about 5% of the total cellular RNA in *E. coli* is mRNA, although approximately one-third of the total RNA synthesized in this bacterium is mRNA. This is accounted by short lifespan (few minutes only) of mRNA.

During transcription, synthesis of mRNAs, tRNAs and rRNAs occurs, using DNA as a template. The tRNA is the smallest of the three major species of RNAs; rRNAs are the most abundant (80% of total) and mRNAs are the most heterogenous type.

II. Transcription

Transcription *is a cellular process during which RNA is synthesized using DNA as a template.* All three types of cellular RNAs are copied from DNA template in the process of transcription. Only mRNA contains instructions for protein synthesis. The other two types (rRNA and tRNA) are not translated, but perform other vital functions during protein synthesis.

Transcription is similar to replication in terms of chemical mechanism, polarity (direction of synthesis) and use of a template. It differs from replication in two important aspects:

1. Transcription does not require a primer.
2. It involves only short segments of DNA. Within these segments, one of the separated strands of DNA serves as a template (i.e. the template strand); the other strand of DNA is the coding strand.

Gene expression requires the synthesis of RNA by *RNA polymerase* (transcription), which is the first step in the mechanism by which genes direct protein synthesis.

Table 22.1. RNA molecules in *Escherichia coli*

Type	Sedimentation coefficient (S)	Percentage of total	Size (number of nucleotides)	Location (in eukaryotes)
rRNA	23	80	3700	Ribosomes, nucleolus
	16		1700	
	5		120	
tRNA	45	15	73–93	Cytoplasm
mRNA	Variable	5	Variable	Nucleus, cytoplasm

RNA Polymerase (RNAP)

RNA polymerase (RNAP) is the principal enzyme of transcription in prokaryotes, which **synthesizes all three types of cellular RNAs**—mRNA, tRNA and rRNA—in *E. coli*. It acts according to instructions given by a DNA template and does not need a primer. It is a complex enzyme (500 kD), consisting of various subunits. The subunit structure of the holoenzyme is **α_2 ββ′σ** (Table 22.2). The sigma (σ) subunit is loosely bound to the other subunits. It is not required for the catalytic activity, but it enables the *RNA polymerase* to initiate the transcription by recognizing certain specific sequences on the DNA. Thereafter it gets dissociated from the holoenzyme, leaving behind the core enzyme (α_2 ββ′).

Each of the subunits of the holoenzyme has a distinct role. The **β′-subunit** helps in the binding of DNA template; while the **β-subunit** binds the ribonucleoside triphosphate substrates. The **α-subunit** is necessary to reconstitute active enzyme from separated subunits. The **σ-subunit** plays an important role in initiating transcription.

RNA polymerase is a template directed enzyme, which carries out polymerization in 5′ → 3′ direction. Unlike *DNA-polymerase*, it does not need a primer. It consists of various subunits; each submit performing a specific function.

A. Three Stages of Transcription

Transcription occurs in three stages, i.e. initiation, elongation and termination.

Initiation

Before starting transcription, RNAP has to find the gene. Genes possess recognition sequence immediately upstream (i.e. on 5′ side) of the sequence that would be transcribed (Fig. 22.1). Transcription is initiated when *RNAP* interacts with the recognition sequences on the coding strand of DNA, called the **promoters**. Binding of the RNAP to promoter directs transcription of the adjacent segment of DNA.

Table 22.2. Subunits of prokaryotic *RNA polymerase* (RNAP)

Subunit	Mass (kD)
α*	40
β	155
β′	166
σ	95

* There are two α subunits in a single *RNAP* molecule.

How does the enzyme find the promoter? The α_2ββ′σ holoenzyme first positions itself on the duplex DNA, forming transient hydrogen bonds with exposed base pairs. This binding is non-selective, occurring with a moderate affinity between the holoenzyme and the bases. Then the holoenzyme slides along DNA. As it encounters the promoter sequence (which is recognized by the σ-subunit), it stops moving further. The transcription initiation complex is now formed, called the **closed complex**; it consists of the α_2ββ′ σ holoenzyme plus the duplex DNA. The duplex DNA must be unwound at this stage, so that one of its strands may serve as template. This unwinding is brought about by the σ-subunit, starting at the conserved AT-rich sequence (Fig. 22.1). This sequence, called the **Pribnow box** (after David Pribnow who described it in 1975), lies in a region about 10 base pairs upstream of the site where mRNA synthesis starts (i.e. transcription initiation site). A segment of nearly 17 base pairs of DNA is unwound in this manner. This results in formation of the **open complex**, which consists of the unwound DNA plus the α_2ββ′ σ holoenzyme.

Thus, the σ-subunit plays major role in initiation of the transcription by:

1. Permitting RNAP to recognize specific initiation site on the coding strand of DNA.
2. Causing unwinding of the duplex DNA.

Role of σ Subunit There has been considerable progress in our understanding of the mechanism by which the σ-subunit helps specific recognition. It does so by decreasing affinity of the *RNAP* for the general regions of DNA by a factor of about 10^4. Consequently, the relative affinity of the *RNAP* for the promoter region becomes much higher, and so the two bind readily.

The sigma subunit of the *RNA polymerase* holoenzyme recognizes and binds to the promoter region of DNA, and this positions the core enzyme (α_2ββ′) to initiate transcription.

Role of promoter sequences: The promoter sequences are clustered approximately 10 base pairs (called the **–10 bp**

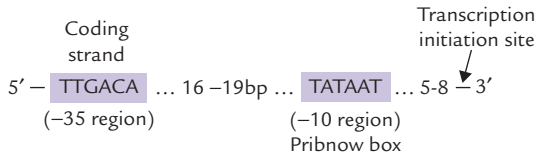


Fig. 22.1. Prokaryotic promoter showing the −10 sequence (Pribnow box) and –35 sequence. By convention, the first nucleotide of the template DNA that is transcribed into RNA is denoted +1, the transcriptional start site (bp = base pairs).

sequence) and 35 base pairs (called the **–35 bp sequence**) upstream of the transcriptional initiation site. It is noteworthy that all promoter sequences are recognized by the same σ-subunit, even though the promoters of different genes do not have identical nucleotide sequence. However, a *consensus sequence* of the most commonly encountered bases can be deduced.

- The **–10 sequence** has the consensus TATAAT. It is named **Pribnow box** after the discoverer. It is an important recognition site that interacts with the σ factor of *RNA polymerase*.
- The **–35 sequence** has the consensus TTGAGA and is important in DNA-unwinding during transcriptional initiation (Fig. 22.1).

It was puzzling as to *why different genes have different promoters*. The most plausible explanation is that different genes have to be transcribed at different rates. Some are transcribed up to 10 times per minute, whereas others are transcribed only once in 10 minutes. The factors accounting for different rates of transcription are:

(a) affinity of binding of the *RNAP* holoenzyme to the promoter site, and
(b) the rate of transition from the closed to the open complex.

These processes depend on base sequence of the promoter. Greater the resemblance between the promoter and the consensus sequences, faster are the above processes, and greater in the rate of transcription. The sequences of strong promoters correspond well with the sequences shown in Figure 22.1, whereas weaker promoters have sequences that differ from these.

The promoter sequences are located on the coding strand and not on the template strand of the DNA. Nucleotide sequences of different promoters show deviations from the consensus sequences to varying extents (consensus sequence is the "average" of sequences of many genes).

Despite its key role in initiation, the σ-subunit does not participate in transcription further. It dissociates from the core enzyme after formation of approximately 10 phosphodiester bonds of the new RNA. The dissociation of the σ-subunit marks the end of the initiation phase of transcription and sets stage for the next phase of transcription, i.e. the elongation phase. The dissociated σ-subunit then joins another core *polymerase* to initiate another round of transcription. Thus we see that the holoenzyme is involved in selection and initiation also, whereas role of the core enzyme is elongation only.

Initiation involves binding of the *RNA polymerase* enzyme to a promoter upstream of the gene. The enzyme binds to a 40–60 bp region that contains two conserved promoter elements, the −10 sequence (Pribnow box) with the consensus TATAAT, and the −35 sequence with the consensus TTGACA.

Elongation

After dissociation of the σ-subunit, affinity of the core enzyme for the promoter decreases markedly. It now moves along the adjacent base sequences of the DNA. The core enzyme plays a pivotal role in the elongation phase, catalyzing the polymerization. New nucleotide units are incorporated in the nascent RNA, one at a time, according to the base pairing rule. Thus, A in DNA is transcribed to U in mRNA, G is transcribed to C, T to A, and C to G. The polymerization occurs in the 5′→3′ direction like in case of replication. *RNAP* is processive, i.e. a single enzyme molecule can remain attached to the template and carry out transcription till the end. The region containing *RNA polymerase*, DNA and nascent RNA is called a **transcription bubble**, because it contains locally unwound (melted) bubble of DNA. The extent of unwinding area is about 17 base pairs per *polymerase* molecule (Fig. 22.2). Nascent RNA forms a hybrid helix with the template DNA strand in this region.

The precursor molecules for new RNA synthesis are ribonucleoside triphosphates. The 3′-hydroxyl of the RNA is so positioned that it attacks the innermost phosphate of the incoming precursor, forming a new phosphodiester bond. The pyrophosphate that is released drives forward the reaction. This is analogous to the mechanism of "new nucleotide incorporation", described in replication (Fig. 21.12).

Following initiation, the σ subunit dissociates from *RNA polymerase* to leave the core enzyme (α ββ′), that uses the antisense stand of DNA is used as a template, and continues RNA synthesis in a 5′→3′ direction using the four ribonucleoside 5′ triphosphates as precursors.

Length of the DNA-RNA hybrid, and the extent of unwound area in duplex remains constant as the *RNA polymerase* moves along the DNA template. This indicates that *DNA is rewound upstream at the same rate as it is unwound downstream*. Furthermore, as *RNA polymerase* causes local unwinding of the duplex DNA, its movement is associated with generation of waves of positive supercoiling ahead of it and of negative supercoiling behind it. Such transcription driven supercoilings are relieved by action of *topoisomerases*.

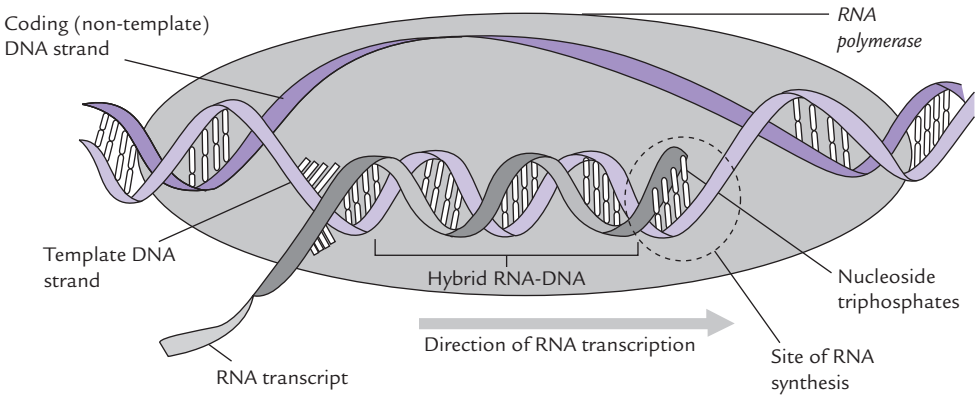


Fig. 22.2. Transcription bubble: *RNA polymerase* takes instructions from a DNA template to synthesize complementary RNA. Only one of the strands of DNA acts as template (i.e. template strand); the other one is a coding strand (or sense strand).

Table 22.3. Comparison of bacterial *DNA polymerase* and *RNA polymerase*

	DNA polymerase	*RNA polymerase*
Substrates	dATP, dGTP, dTTP, dCTP	ATP, GTP, UTP, CTP
Primer required	Yes	No
Exonuclease activities	Yes	No
Subunit structures	8 subunits in DNAP III Single in DNAP I	α_2 $\beta\beta'$ σ
Bases added per second	10–20 by DNAP I 600–1000 by DNAP III	$\approx$ 20

RNA polymerase resembles *DNA polymerase* in several ways, but it differs from the latter in two important aspects:

1. It does not require a primer; it can initiate chain synthesis by assembling the first nucleotide itself.
2. It lacks *exonuclease* activities; thus, in contrast to *DNA polymerase*, it does not edit the nascent polynucleotide chain. Consequently, the fidelity of transcription is much lower than that of replication. The error rate in transcription is 10^4 to 10^5 times more than that in DNA replication. But relatively higher error-rate in transcription is tolerated since RNA is synthesized and degraded continuously; thus, a single error causes little damage. It is of less consequence to the cell since these errors are not transmitted to the daughter cell, or to the next generation. By contrast, any error in replication produces alteration in permanently stored genetic information, which is transmitted to progeny.

Comparative features of RNAP and DNAP are given in Table 22.3.

Sense (+) and antisense (–) strands: It merits re-emphasis that only one strand of the duplex DNA is copied, directing synthesis of new RNA in a given region of genome (Fig. 22.3); it is the template or non-coding strand, also called the **antisense (–) strand**. The other strand is sense-stand, also called coding (or non-template) strand. The RNA produced has same sequence as the the **sense (+) strand** (except that T replaces U).

It is evident from what has been described so far, that nucleic acid synthesis (whether of DNA or RNA) requires a DNA template. Thus, genetic information coded in DNA is the prime factor for directing the type of protein synthesized.

The antisense strand is used as a template during transcription.

Termination

RNAP must know the defined site at which to stop RNA synthesis, so that the appropriate size of transcript is produced. This part known as transcription termination is probably the least understood part of RNA synthesis. It occurs by two well characterized mechanisms. The first mechanism, **rho-dependent termination**, requires the action of a protein factor, called rho (ρ), which recognizes certain termination signals. This halts movement of RNAP along the DNA template. The other mechanism does not require participation of the rho (ρ) factor i.e. **ρ-independent termination**).

A consistent feature of all transcriptional termination sites is the presence, i.e. the base sequence of one DNA strand, read in one direction is same as that of the other DNA strand read in the opposite direction (Fig. 22.3). A sequence of this type is called **palindrome**. When palindrome is transcribed, the base sequence of the RNA transcript is self-complementary.

Therefore, the mRNA transcript of this region forms a self-complementary **"hairpain" structure** due to internal base pairing (Fig. 22.3). The hairpin loop is often rich in GC base pairs which interact more strongly than AU base pairs. The GC rich region is often followed by a sequence of 6–8 uridine (U) residues. Formation of this secondary loop-structure dislodges the RNAP from the DNA template, resulting in termination of the RNA synthesis in the U stretch.

The ρ (rho) factor is a hexameric protein that causes termination of some transcripts.

How does the rho factor participate in termination is still not clear. Two activities of the rho *factor* are probably involved: the *ATPase* action and the DNA-RNA *helicase* action.

1. The *ATPase* activity releases energy which permits the rho factor to move towards the termination site.
2. *RNA-DNA helicase* activity helps in separation of the nascent RNA from DNA template at the termination site.

Though rho factor plays important role, specific termination can occur in its absence also. Certain termination sites are present on the template, which do not require the rho factor. Transcription stops at such sites even in absence of this factor.

The transcription bubble travels along the DNA as the RNA chain is elongated stepwise, till it is terminated in response to specific secondary structural elements in the transcript and may require the action of the rho factor.

Recently some more proteins have been reported to mediate termination. For example, the nusA protein enables *RNA polymerase* in *E. coli* to recognize a characteristic class of termination sites. Lambda phage synthesizes antitermination proteins that allow certain genes to be transcribed and expressed.

Direction of movement of the replication and the transcription machinery is same to reduce collisions (Box 22.1).

B. Post-transcriptional Modifications

The initial product of *RNA polymerase*, base sequence of which faithfully reflects that of the gene from which it was transcribed, is called the **primary transcript**. For a primary transcript to be functional, it must be processed before leaving the nucleus. Processing may involve endolytic cleavage (to cut-out unwanted sequences from the primary transcript), splicing, chemical modifications, terminal base additions, etc. Such modifications of RNA after its synthesis by *RNA polymerase* are collectively referred to as *post-transcriptional modifications*. Of the three major types, mRNAs are processed extensively in eukaryotes, but rarely in prokaryotes. rRNAs and tRNAs are subject to processing both in eukaryotes and in prokaryotes.

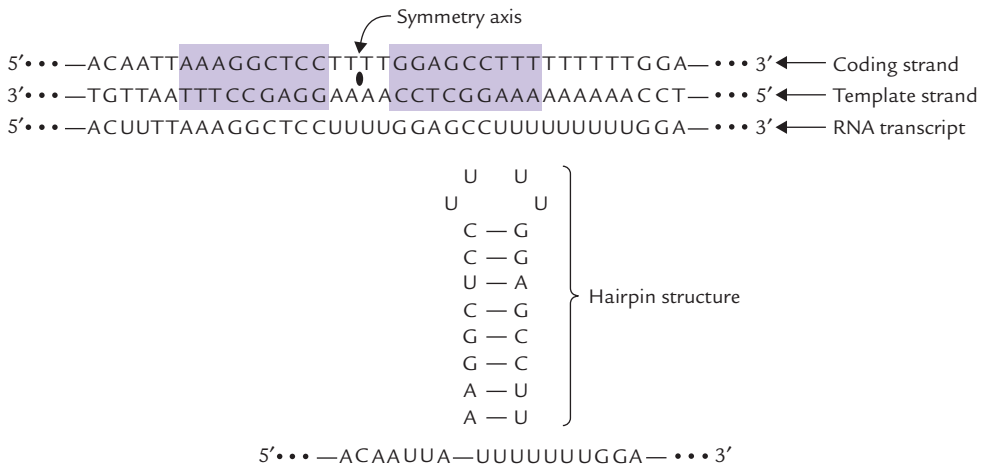


Fig. 22.3. Base sequence of mRNA transcript (transcribed by a palindrome) is self-complementary, and forms a hairpin loop.

BOX 22.1

Collision Between *DNA Polymerase* and *RNA Polymerase*

In rapidly dividing bacterial cells, replication and transcription are likely to occur at the same time: new DNA molecules are synthesized even as genes are being transcribed. However, movement of DNA replication machinery along the circular chromosome is many times faster than rate of movement of transcription machinery. Collision between the two complexes (replication complex and transcription complex) therefore is inevitable. What happens when the two complexes collide? Two possible situations may arise: the first, when both the complexes are moving in the same direction and the second, when they are moving in opposite directions. In the first situation, the replication fork passes through the *RNA polymerase* within, displacing this enzyme and leaving it fully competent to resume RNA chain elongation. However, the movement of the two complexes in opposite directions causes replication fork to collide head-on with the transcription complex. As a result, the *DNA polymerase* from pauses and the replication shows down. The collision also causes dissociation of the *RNA polymerase* its template strand. This is followed by jump of the enzyme (*RNA polymerase*) from one template strand to another which may sometimes halt the transcription altogether. Finally, the collision causes the growing RNA chain to dissociate from the original template DNA strand and hybridize with the newly synthesized daughter DNA strand of the same sequence, before RNA elongation resumes.

In view of the hazards associated with the collision, it is understandable that in many bacterial and phage genomes, many genes are oriented in such a way that replication and transcription complexes move in the same direction.

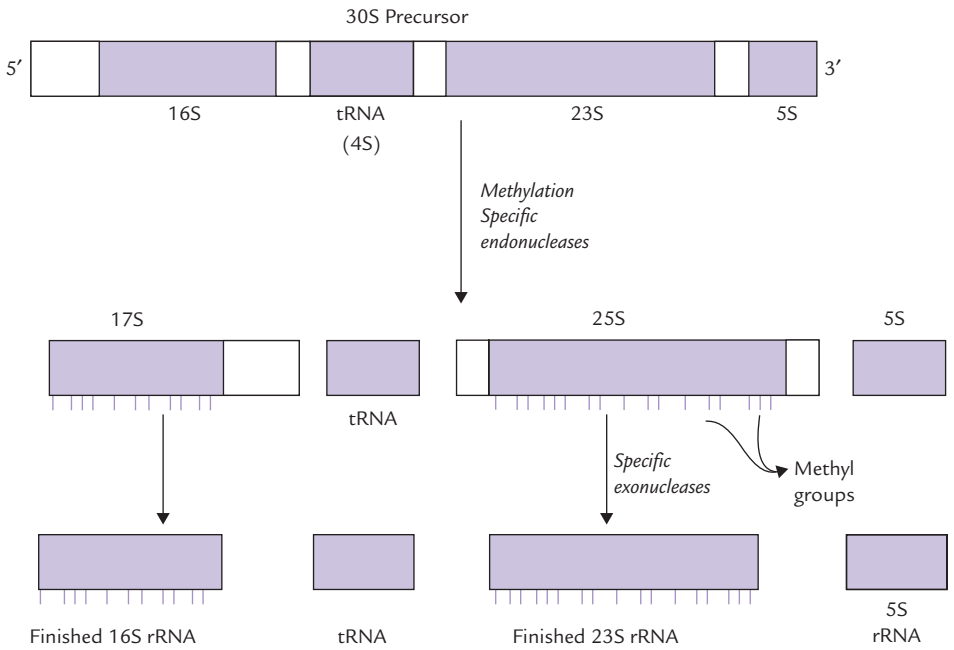


Fig. 22.4. Cleavage of the 30S primary transcript produces 5S, 16S and 23S rRNAs, and a tRNA.

mRNA transcripts of protein-coding genes in prokaryotes require little or no modification before translation. However, ribosomal RNAs and transfer RNAS are synthesized as precursor molecules that require processing by specific *ribonucleases* to release the mature RNA molecules.

rRNA

The rRNAs are initially synthesized as a large precursor RNA both in eukaryotes and prokaryotes. In *E. coli*, this precursor contains sequences of 16S, 23S and 5S rRNAs. In some cases, one or several tRNAs are also contained within the precursor. In order to obtain rRNAs (and tRNAs) of the correct size from this precursor, a concerted action of specific ***ribonucleases*** *of two major types endoribonucleases* and *exoribonucleases*, are required.

- *Endoribonucleases* cleave phosphodiester bonds within the primary transcript to release individual RNAs, and
- *Exoribonucleases* trim these RNAs by removing excess nucleotides from their 5′ and 3′ ends till a molecule of correct size is produced (Fig. 22.4).

The large precursor RNA contains one copy each of 16S, 23S and 5S RNAs so that they are generated in equal amounts. This mechanism of synthesis ensures that the three rRNAs are available in appropriate amounts for the assembly of ribosome.

Bacterial rRNAs contain some *methylated ribose residues*, which are produced post-transcriptionally by methylating enzymes, using *S-adenosyl methionine* as the methyl group donor (Fig. 22.5).

Ribosomal RNAs are derived similarly from a single large precursor in eukaryotes as well (Fig. 22.6).

tRNA

In both, prokaryotes and eukaryotes tRNAs are similarly excised from large precursors by *ribonucleases*. An end sequence containing cytidine, cytidine and adenosine (CCA) is added to the 3′end of the chains. Most tRNAs contain several unusual or chemically modified bases and sugars. Specific nucleotides are modified post-transcriptionally to yield these unusual bases (Fig. 22.5).

As noted in *E. coli*, three rRNA molecules and a tRNA molecule are obtained from a single primary transcript. Other primary transcripts may contain several different tRNAs or several copies of the same tRNA in addition to rRNAs.

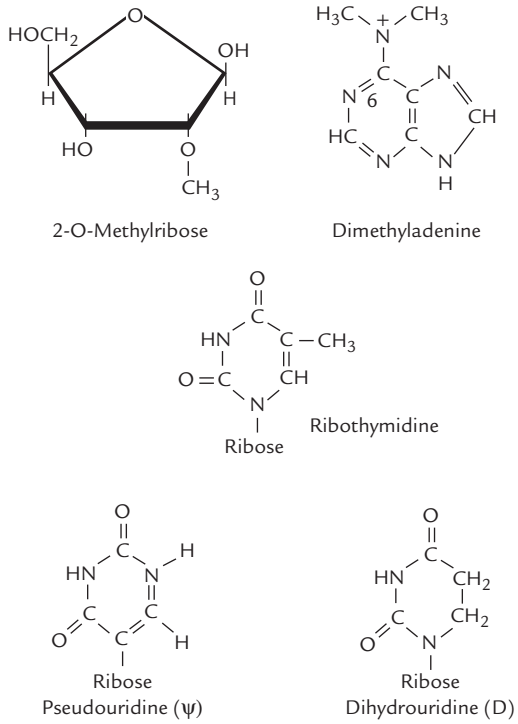


Fig. 22.5. Some modified bases.

mRNA

Most proteins in prokaryotes are encoded by a single uninterrupted DNA sequence that is copied without alteration to yield a mRNA, which is rarely subject to any post-transcriptional processing. In eukaryotes, the coding sequences (exons) are interrupted by non-coding sequences (introns) and so the primary transcript of mRNA is synthesized as a larger primary transcript, known as heterogenous nuclear (heteronuclear, hn) RNA. The latter undergoes complicated processing (described in Chapter 24).

The mRNA is processed most extensively in eukaryotes but rarely in prokaryotes. The processing of rRNA, includes nucleolytic cleavage, methylation and splicing; while processing of tRNA transcript includes addition, removal and modification of nucleotides.

C. Antibiotic Inhibitors of Transcription

Transcription can be inhibited by certain antibiotics that inhibit specific biological processes.

Agents that Bind DNA

Actinomycin D, an antibiotic isolated from *Streptomyces antibioticus*, contains a planar phenoxazone ring system that becomes intercalated between two G-C base pairs in

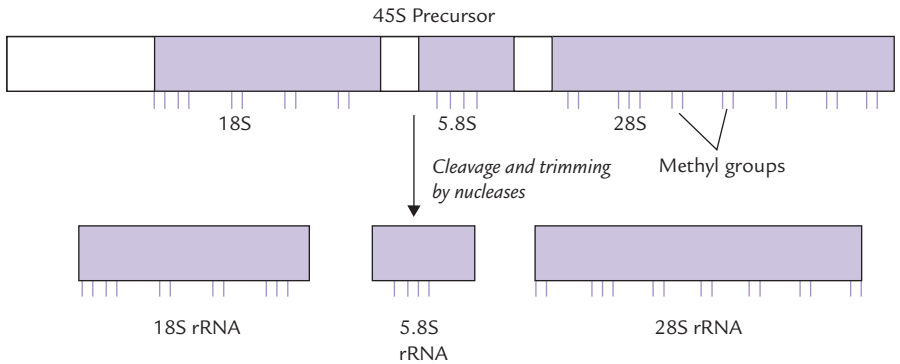


Fig. 22.6. Processing of the rRNA precursor in eukaryotes.

double-stranded DNA. In addition, this drug has two oligopeptide tails which fit into the grooves of the DNA. When actinomycin D is present RNAP cannot transcribe past the bound drug.

DNA has the same structural design in all organisms. Therefore, actinomycin D is toxic for eukaryotes as well as prokaryotes, and it cannot be used for the treatment of bacterial infections. However, it has been found to be very useful in treatment of **Wilm's tumour**, a rare childhood tumour involving the kidney (nephroblastoma). Reason for such specific effects of this drug is not clear.

Agents that Bind RNAP

Rifampicin, a synthetic derivative of a bacterial compound rifamycin B, inhibits transcription in a different way. It binds with RNAP and prevents formation of the first phosphodiester bond in the new RNA chain. The bacteria are not destroyed by this treatment but lose their ability to grow. Thus, the drug is bacteriostatic, not bactericidal. Rifampicin has been successfully used for the treatment of a variety of infections, including tuberculosis. Unfortunately, the bacteria can become resistant to it by point mutation that alters the characteristics of the rifampicin binding site on the β-subunit of *RNAP*.

Anthramycin is a relatively new drug, which can inhibit both RNA and DNA synthesis.

III. Genetic Code

The genetic information flows from the gene to protein, with mRNA serving as information carrying intermediate. *Genetic code defines the relationship between the sequence of bases in DNA and the sequence of amino acids. DNA* is the repository of all genetic information that directs protein synthesis. Information is coded in the form of a base sequence, as noted earlier. The nucleotide sequence of DNA has been described as a four letter language since it comprises four bases present in DNA (A, T, G, and C). These four code letters are ultimately translated into amino acid sequence. Translation of the four letter language of bases (A, T, C, G) into 20 letter language of amino acids (there are 20 primary amino acids) is an important aspect of molecular biology.

Protein synthesis begins with transcription of specific genes to form mRNA and ends with the assembly of amino acids into the final protein product. The base sequence of mRNA is complementary to that of the DNA which serves as a template. *A sequence of three bases in mRNA codes specifically for one amino acid*. The primary structure of mRNA and the protein synthesized is colinear (Fig. 22.7).

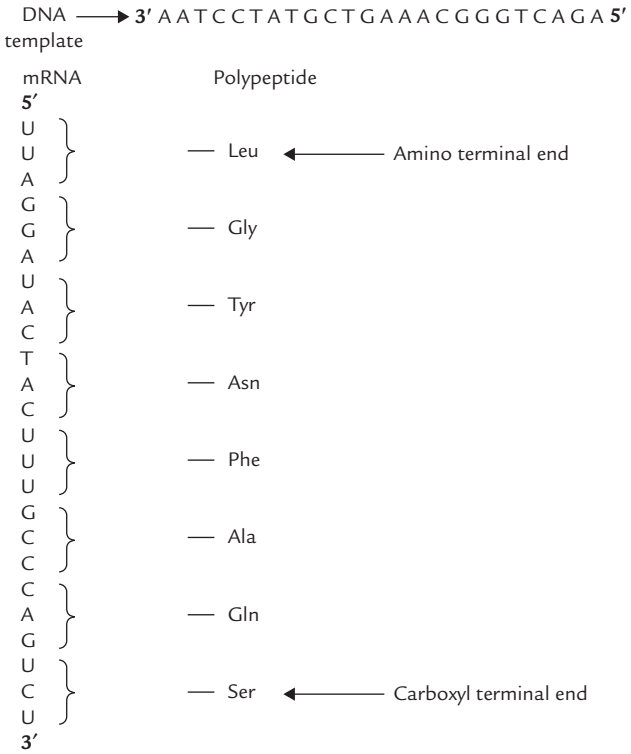


Fig. 22.7. Base sequence of the DNA, the mRNA transcript and the amino acid sequence of polypeptide are colinear.

Three bases of a given sequence in a mRNA serve as code letters or codon for one amino acid; for example, UUU codes for phenylalanine and AAA for lysine. Initially it was thought that the code letters (A, U, G, C) in a group of two are the genetic words or codons for amino acids. But by this arrangement only 16 (4^2) combinations of bases are possible, which are not sufficient to code for 20 primary amino acids. Four bases in groups of three (triplet) can yield 64 (4^3) combinations, which are sufficient for 20 amino acids. Therefore, it was established that genetic code consists of three bases and that *amino acid sequence of a protein is defined by a linear sequence of triplet codons*.

The relationship between the base sequence of the mRNA and the amino acid sequence of the polypeptide is defined by a genetic code. It is a triplet code in which a sequence of three bases on the mRNA, the codon, specifies an amino acid.

A. Deciphering the Genetic Code

What is the specific three letter code word for a given amino acid? Initial experiments conducted by Marshall Nirenberg and Heinrich Matthaei provided a major breakthrough in understanding the nature of the genetic code.

1. These workers carried out synthesis of a polyribonucleotide chain of polyuridylate (designated as poly U). The synthetic chain was incubated with a mixture of the 20 primary amino acids and an *E. coli* extract, the latter providing essential components required for protein synthesis. The synthetic poly U chain may be regarded as mRNA, containing a series of the triplet codon UUU. It was observed that a polypeptide containing phenylalanine residues was synthesized. No other amino acid was incorporated in this chain. This observation led these workers to conclude that this triplet (UUU) coded for phenylalanine.
2. When the experiment was repeated using synthetic poly C, the polypeptide that was synthesized contained only proline residue (polyproline). Therefore, the triplet CCC codes for proline.
3. Using the same approach these workers reported that the triplet AAA codes for lysine.

The synthetic polynucleotides used in these experiments were synthesized by action of the enzyme, *polynucleotide phosphorylase*. This enzyme catalyzes formation of RNA polymers from the dinucleotide precursors in vitro.

$$(RNA)_n + \text{Dinucleotide ribonucleoside} \rightleftharpoons (RNA)_{n+1} + P_i$$

The polynucleotide phosphorylase is not a template-directed enzyme and the composition of polymer synthesized by it solely depends on the nature of the substrate molecules. For example, if the substrate is UDP, polyU is synthesized, and formation of polyA occurs if the substrate is ADP. If more than one substrate is used, the nucleotide composition of the polymer reflects their relative concentrations. For instance, if the enzyme is presented with a mixture of five parts of ADP and one part of CDP, it will synthesize a random polymer containing maximum triplets of the sequence AAA; fewer of AAC, ACA, and CAA triplets; still fewer CCA, CAC and ACC; and least number of CCC triplets.

By observing relative quantities of various amino acids incorporated in the polypeptide chain, it was possible for Nirenberg and co-workers to identify the triplets coding for them. For example, lysine was incorporated in maximum amount in the above experiment, which indicated that it is encoded by AAA. Least amount of proline was incorporated, indicating that it is encoded by the triplet CCC. Nirenberg was awarded Nobel Prize in 1969 for deciphering the genetic code.

A complementary approach was provided by **H. Gobind Khorana** (Nobel Prize, 1968) who developed methods to synthesize polymers with repeat sequence of two or four bases. Study of the nature of polypeptides that were synthesized, in response to these RNAs as messengers, permitted unambiguous codon assignments. For example, polymers with defined sequence of two bases (A and C) were synthesized. These copolymers (AC)n had two alternating codons: CAC and ACA. The polypeptide chain synthesized in response to this polymer contained histidine and threonine in equal proportions. This observation, when combined with the earlier information, led to the conclusion that CAC codes for histidine and ACA for threonine. Similar studies, when applied to polymers having four bases with repeat pattern, provided further information regarding assignment of codons.

Specific amino acids coded by 61 of the 64 codons were thus identified. Further, it was found that codons, namely UAA, UAG and UGA, do not code for any amino acid. They are termed as **termination codons** or **stop codons**, since translation is terminated whenever these codons are encountered. All the 64 codons are shown in Table 22.4.

From Table 22.4, it is possible to know the amino acid that corresponds to a given codon. For example, the codon 5′ (AUG) 3′ on mRNA specifies methionine; and 5′ (ACA) 3′ or 5′ (ACG) 3′ or 5′ (ACC) 3′ specify threonine.

Table 22.4. The genetic code

First position (5′ end)	Second position				Third position (3′ end)
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

B. Major Features of Genetic Code

Degeneracy

Perhaps the most striking feature of the genetic code is that it is degenerate, meaning that *a given amino acid may be specified by more than one codon*. Since 61 codons exist

for 20 amino acids, it becomes evident that one amino acid has on an average three codons.

However, degeneracy is not uniform because some amino acids, such as arginine, have six codons. On the other hand, amino acids like tryptophan and methionine are coded by a single codon.

Codons that specify the same amino acid are called ***synonyms***. For example, CAU and CAC are synonyms for histidine. Examination of the synonyms shows that they differ from one another in the third base of the triplet (i.e. the one at 3′ direction). For example, isoleucine is coded by the triplets AUU, AUC and AUA. The first two bases (read from 5′→3′ direction) are same in all the three codons, which thereby serve as the primary determinates of specificity.

Biological significance of degeneracy is not clear. Perhaps it minimizes the deleterious effects of mutations.

Unambiguous

Though degenerate, the genetic code is not imperfect or ambiguous because no codon specifies more than one amino acid. A given codon specifies one and only one amino acid.

The unambiguous nature of genetic code suggests that replacement of a base by another may change the codon, so that it specifies a different amino acid. This may result in incorporation of wrong amino acid in the polypeptide chain that is synthesized. Several disorders are known to arise in this way (**Case 22.1**).

Universal

A given codon specifies a particular amino acid in different organisms. Thus, in course of evolution, the information specified by the genetic code has remained invariant. There is, however, an exception to it in case of the initiation codon AUG, which determines N-formyl methionine in prokaryotes and methionine in eukaryotes. In addition, some minor variations do occur in genomes of mitochondria and chloroplast (Box 22.2).

The near universality of the genetic code shows that all life on earth has descended from a common ancestor.

Non-overlapping and Commaless

The successive codons occur one after another, meaning that they do not share any nucleotides. The codons are aligned without overlap and without empty spaces in between.

Colinear

The sequence of amino acids in the polypeptide—from the amino end to the carboxy end—corresponds to the base sequence of a gene (from the 5′ end to the 3′ end).

Stop Codons

The stop codons (UAA, UAG and UGA) signal end of polypeptide chain synthesis. Unlike the coding codons, which are read by the corresponding tRNAs, the stop codons are read by certain specific proteins, called **release factors**.

One of the amino acid coding codons, AUG, is known as **initiation codon** since it signals beginning of the polypeptide chains in both prokaryotes and eukaryotes.

There are 61 amino acid codons and three stop codons.

C. Codon-Anticodon Pairing

During protein synthesis, the primary structures of mRNA and the protein synthesized are colinear. However, the codon on mRNA cannot be directly recognized by

BOX 22.2

The "Standard" Genetic Code is Not Universal

The phenomenon that one kind of organism (*E. coli*) can accurately translate the genes for quite a different organism (e.g. humans) is the basis of genetic engineering. This had also led to the assumption that the "standard" genetic code was universal. Further studies have, however, proved that the "standard" genetic code, although very widely utilized, is not universal. Alternate genetic codes have, evolved in certain mitochondria which are variants of the "standard" genetic code. For example, in mammalian mitochondria, UGA specifies tryptophan rather than "stop" and AGA and AGG are "stop" rather than arginine. Apparently mitochondria which contain their own genes and translating machinery are not subject to the same evolutionary constraints as are nuclear genomes.

Likewise, an alternate genetic code has been shown in certain ciliated protozoa as well.

the corresponding amino acids as there are no specific interactions between bases and amino acids. Rather a codon is recognized by the anticodon on the tRNA that carries the amino acid corresponding to that codon. *The three nucleotides on an mRNA codon pair with the three nucleotides of a complementary tRNA anticodon in an antiparallel fashion*. The binding is said to be antiparallel because the first base of codon (X′) pairs with the third base of anticodon (X), and the third base of codon (Z′) pairs with the first base of anticodon (Z).

Anticodon	3′ – X – Y – Z – 5′
Codon	5′ – X′– Y′ – Z′– 3′

Since there are 61 amino acid coding codons, it is expected that the same number of tRNAs must be present. Actually, however, number of tRNAs is much less. It can, therefore, be logically concluded that one tRNA must be able to base pair with more than one codon.

Crick proposed the **wobble hypothesis** to explain how a single tRNA can recognize several (degenerate) codons. He assumed that the first two codon-anticodon pairings have normal Watson-Crick geometry, but there is freedom of base pairing between the third codon base (Z′–3) and the first (5′–Z′) anticodon base. The latter can get involved in non-Watson-Crick base pairing with two or three different bases; the allowed pairings for the third codon-anticodon position are listed in Table 22.5. Thus, if U is present as the first anticodon base, it can pair not only with A but also with G. Thus, such anticodons can base pair with two different codons.

Anticodon	3′ – X – Y – U – 5′
Codon	5′ – X′ – Y′ – A – 3′
Anticodon	3′ – X – Y – U – 5′
Codon	5′ – X′ – Y′ – G – 3′

The first anticodon position (termed *Wobble position*) commonly contains inosine, so that such anticodons can recognize three different codons. By combining the above structural insights with logical deduction, it may be seen that wobble hypothesis accounts for codon degeneracy.

A cosideration of the various wobble pairings indicates that *at least 31 tRNAs are required to translate all 61 coding triplets of the genetic code*. Actually, however, there are 32 tRNAs in the minimal set because translation initiation requires a separate tRNA. Most cells have more than 32 tRNAs (usually 60), some of which have identical anticodons.

Table 22.5. Imprecision in the binding of the first anticodon base to the third codon base

First anticodon base	Third codon base
C	G
A	U
U	A or G
G	U or C
I	U, C or A

IV. Translation

Translation is far more complex than replication and transcription. In the latter two processes, a sequence of bases was simply copied into another sequence of complementary bases. Also, the bases in template interacted specifically by base pairing with the bases of newly synthesized polynucleotide chain. Simply stated, a four-letter language of four bases (A, T, G, C) was copied into four-letter complementary language of four bases. However, such simple design is not possible in translation because it involves translation of the information inherent in the nucleotide sequence (of DNA and mRNA) into a colinear amino acid sequence of polypeptide. Since there is no specific binding interaction between bases and amino acids, this process is evidently more complex.

Translation requires not only the template mRNA and the amino acid substrates for the polypeptide chain, it also needs participation of a number of other biomolecules. In fact, a coordinated interplay of more than a hundred molecules, including tRNA-activating enzymes and protein factors is required in the process of translation. Ribosomes serve as work-benches for all these components.

A. tRNAs are the Adaptor Molecules

The tRNAs are small oligonucleotides, containing 70–90 nucleotides. They are used to *convert the information of bases in mRNAs into the corresponding sequence of amino acids in proteins, thereby serving as adaptor molecules*. They are suited for this role because they can (a) bind with codon on the mRNA, and (b) link with an amino acid corresponding to that codon.

All tRNAs contain an anticodon triplet, which interacts by specific base pairing with the codon on the mRNA chain to select the correct amino acid. For example, a tRNA molecule with the anticodon AAA must attach to phenylalanine since UUU is the complementary codon for this amino acid. The tRNA specific for phenylalanine is depicted as $tRNA^{Phe}$ and so on for each of the 20 amino acids, using the first three letters of amino acid names as abbreviations.

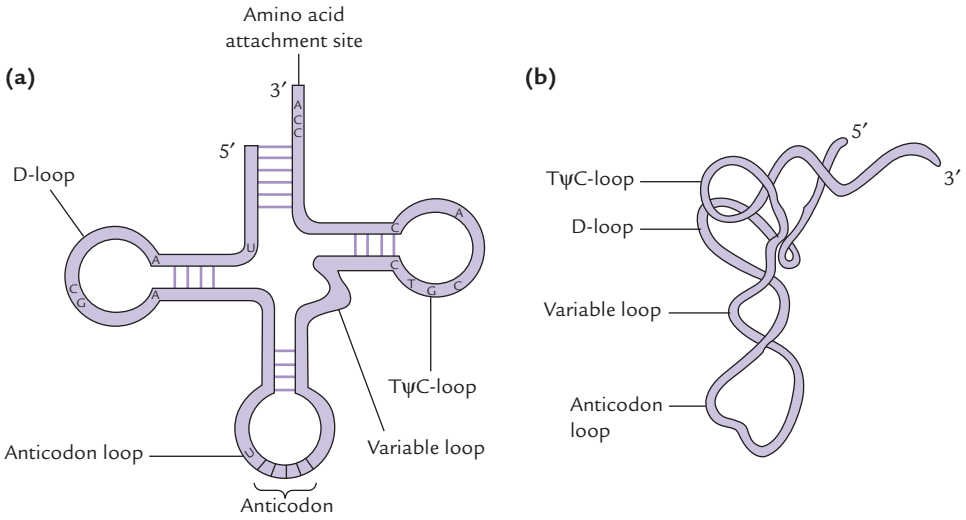


Fig. 22.8. Structure of tRNA (**a**) The clover leaf structure, (**b**) The tertiary structure.

There are nearly 60 different tRNAs in the cell, and each of them presents only one amino acid to ribosome. Though base sequences of tRNAs differ from one another, the following features are common to them all:

1. The tRNA molecule is folded in a **clover leaf structure**, with three base paired regions and three loops (Fig. 22.8). Most important of these loops is the one containing the anticodon. Three bases of anticodon pair with the codon on the mRNA during protein synthesis. The other two loops—D (*dihydrouridine*) loop and TψC (*thymine pseudouridine-cytosine*) loop—are named after the conserved bases in their sequence. The D-loop contains dihydrouridine and the TψC-loop contains both pseudouridine and ribothymidine. Recognition of the tRNA by the *aminoacyl-tRNA synthetase* rests on the D-loop; and the TψC-loop binds the tRNA-amino acid to the ribosomal surface during protein synthesis.

   Structures of the dihydrouridine, ribothymidine and pseudouridine are shown in Figure 22.5. In addition, a *variable loop* (also called *extra arm*) is also present in most tRNAs.
2. The 3′ terminal ends with the sequence, CCA which links covalently with an amino acid.
3. The 5′ terminal is covalently linked with a phosphate group.

Thus, *a tRNA has two functionally important arms: one of these links with amino acid and the other binds a codon*.

There is at least one type of tRNA molecule specific for each of the 20 primary amino acids. For some of these amino acids, more than one type of tRNA molecules are present. At least 32 tRNAs are required to recognize all amino acid codons, but most cells contain many more than 32, as discussed later.

The amino acid codons are recognized by tRNAs during protein synthesis, and each tRNA presents an amino acid to the ribosome. Many tRNAs recognize more than one codon.

B. Attachment of Amino Acids to tRNA Molecules

Amino acids are covalently linked to the corresponding tRNAs by soluble cytoplasmic enzymes, *aminoacyl-tRNA synthetases* to form the amino acid derivative of the tRNA. The latter is called aminoacyl-tRNA and the tRNA with attached amino acid is said to be aminoacylated or *charged*. The term *uncharged* tRNA refers to a tRNA molecule lacking an amino acid. The overall acylation reaction, which involves ATP breakdown to supply energy is

$$\text{Amino acid} + \text{tRNA} + \text{ATP} \rightleftharpoons \text{Aminoacyl-tRNA} + \text{AMP} + \text{PPi}$$

The process is referred to as **amino acid activation**. The reaction is reversible because the ester bond between amino acid and tRNA has essentially the same energy level as the phosphoanhydride bond in ATP. Nevertheless, the reaction is driven to completion because the inorganic pyrophosphate is hydrolyzed to two molecules of Pi.

Amino acids are activated by an ester bond with the 3′-terminal of the tRNA.

Dual specificity of aminoacyl-tRNA synthetase: An *amino acid synthetase* is specific for (a) the amino acid, which is

activated, and (b) the tRNA, which would accept the activated amino acid.

(a) *There is a synthetase specific for each of the 20 amino acids.* Usually only one *aminoacyl synthetase* exists for each amino acid, thus each bacterial cell has at least 20 different species of these enzymes. However, for a few amino acids, more than one *synthetase* does exist (such amino acids are specified by more than one codon).

(b) *A given synthetase is specific for the prospective acceptor tRNA molecule also*: the *synthetase* specific for leucine would load leucine to $tRNA^{Leu}$ only. If, however, an amino acid can be bound to more than one tRNAs the same enzyme can recognize all prospective acceptor tRNAs. This means that each enzyme recognizes one or more specific tRNAs and attaches the appropriate amino acid to it.

The *enzymology of activation* is quite similar to that for activation of fatty acids by *thiokinase*. The reaction similarly occurs in two stages as shown here.

$$\text{Amino acid} + \text{ATP} \longleftrightarrow \text{Aminoacyl-AMP} + \text{PPi}$$
$$\text{Aminoacyl-AMP} + \text{tRNA} \longleftrightarrow \text{Aminoacyl-tRNA} + \text{AMP}$$

(There are nearly 60 different tRNAs, and each can attach to only one amino acid.)

Proofreading and editing functions: These additional functions of *aminoacyl-tRNA synthetases* enhance fidelity of the amino acid attachment to tRNA. This is important because it is on the accurate selection of the correct amino acid that the accuracy of translating mRNA into protein depends. Usually the active site of a *synthetase* is highly specific for its substrate amino acid, but errors do occur, especially when it has to distinguish between very similar amino acids, such as valine and isoleucine, which differ in a single methylene ($-CH_2$) group. The *synthetase* specific for isoleucine can mistakenly activate valine and attach it to $tRNA^{Ile}$ with an error rate of more than 1%, which would result in an unacceptable rate of errors in translation, unless there was a corrective mechanism. A '**proofreading**' mechanism exists to prevent this, which depends on the presence of an additional catalytic site in the *aminoacyl-tRNA synthetase*.

Paul berg demonstrated that this catalytic site causes quantitative hydrolysis of the mistakenly activated valine (val AMP) rather than forming val-$tRNA^{Ile}$.

C. Ribosomes: The Organelles for Protein Synthesis

Ribosomes are large, catalytically active ribonucleotide-protein particles that serve as workbenches for protein synthesis. Though they do not form a separate compartment in a cell, they have been classified as organelles. In bacteria they float freely in the cytoplasm or attach with plasma membrane, while in eukaryotes they remain either free in cytoplasm or attached to the membrane of endoplasmic reticulum (ER). About 1600 ribosomes are present in an *E. coli* cell. Eukaryotic cell, on the other hand, contains more than a million.

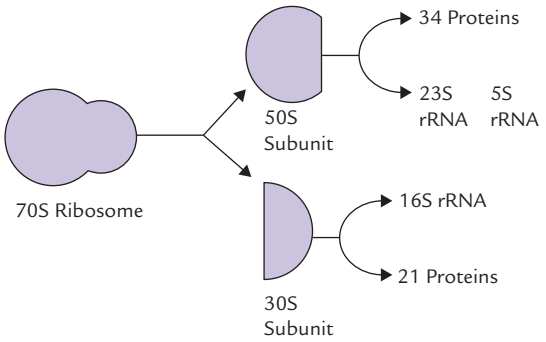


Fig. 22.9. Ribosome, the organelle for protein synthesis.

During protein synthesis, the ribosome binds to a site near the 5′ end of the mRNA and subsequently translates it in the **5′→3′ direction**. Concomitantly, the polypeptide chain is synthesized in the amino to carboxy terminal direction.

A ribosome consists of two subunits: a large subunit and a small subunit, held together by non-covalent interactions. According to their sedimentation rate in the ultracentrifuge, the bacterial ribosome subunits are termed 30S and 50S subunits, and the ribosome itself is characterized as 70S (Fig. 22.9). The 30S subunit contains 21 proteins and a 16S rRNA molecule. The 50S subunit contains 34 proteins and two rRNA molecules, a 23S species and a 5S species. About two-third of the mass of an *E. coli* ribosome is RNA, whereas the other third is protein.

Note: The 16S rRNA plays an important role in initiating protein synthesis at a specific site, as discussed later.

Eukaryotic ribosomes are much larger; 80S ribosome is formed from 40S and 60S subunits.

Ribosome is large, catalytically active ribonucleoprotein particle that serves as workbench of protein synthesis. It reads the mRNA in the 5′→3′ direction while the polypeptide grows stepwise in the amino to carboxy terminal direction.

D. How Does Ribosome Know Where to Begin Protein Synthesis

The entire mRNA is not translatable, translation begins at a precise point in mRNA. It is of vital importance that

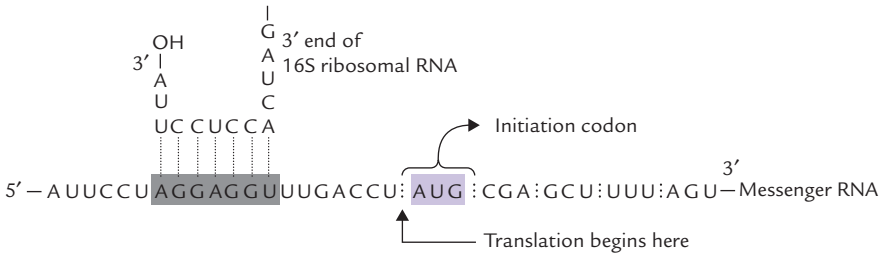


Fig. 22.10. Base pairing between the purine rich sequence in mRNA and the 3′ end of the 16S rRNA permits initiation of translation at the initiation codon.

there is an absolute precise initiation by ribosome; in other words, it must initiate correctly. An error of even a single base would be a reading frameshift, resulting in synthesis of a useless protein instead of that specified by the gene. Two kinds of signals determine the site of initiation: (a) the **AUG codon**, and (b) the **Shine–Dalgarno sequence.**

AUG Codon

The start site is **AUG**, *less frequently GUG codon*, located at least 25 nucleotides away from the 5′-terminal of the mRNA. However, all AUG codons do not necessarily signal beginning of translation: an AUG codon may be located further down the mRNA where it codes for methionine. *How is that initiation occurs only at the first AUG and not at the ones located at more internal positions?* This confusing situation is resolved by the fact that there are two different tRNAs, both specific for methionine. They have the same anticodon, but one tRNA is used exclusively for initiation and the other exclusively for adding methionine in the elongation process. What then determines the different functions of the two methionine-specific tRNA species? The answer will be evident in the following sections.

Protein Synthesis Starts with N-formylmethionine: The methionine which gets attached to the initiator tRNA ($tRNA^{fMet}$) is formylated on its amino group by a *transformylase* in prokaryotes. This enzyme uses N^{10}-formyltetrahydrofolate as formyl donor. As a result, the prokaryote proteins, for the reasons that are not clear, are synthesized with N-formylmethionine as the first unit.

$$\text{Met–tRNA}^{fMet} + N^{10}\text{ Formyl THF} \rightarrow N^{10}\text{ Formyl Met tRNA}^{fMet} + \text{THF}$$

In some proteins, this amino terminus is retained, but more commonly it is deformylated, or fMet is removed by a proteolytic cleavage.

Shine–Dalgarno Sequence

It is a purine rich sequence centred about 10 nucleotides at the 5′ end of the initiator codon (consensus sequence AGGAG). This sequence helps to position the ribosome at the beginning of each protein coding region. It is complementary to a portion of the 16S rRNA in the small ribosomal subunit (Fig. 22.10). These sequences interact through hydrogen bonding of complementary base pairs, and this interaction helps to target the ribosome to the protein-coding regions of the mRNA. It is possible that the efficacy of initiation is modulated by strength of this interaction.

E. The Process of Protein Synthesis

Translation is a dynamic process that involves the interaction of enzymes, tRNA, ribosomes and mRNA in specific ways to produce a protein molecule. The protein is synthesized from the N-terminus to C-terminus. This complex process is normally divided into three steps: **initiation, elongation** and **termination**.

A co-transcriptional initiation of translation is specific for prokaryotes.

Initiation

Initiation of protein synthesis takes place when:

(i) A ribosome (both small and large subunits) assembles on the mRNA.
(ii) The initiator tRNA ($tRNA^{fMet}$) occupies a specific site (P site) on ribosome.

The process begins by formation of a ***30S initiation complex*** between the 30S ribosomal subunit, the mRNA and formyl met-$tRNA^{fMet}$ (Fig. 22.11). GTP serves as a source of energy to drive this process. Three **initiation factors** (IF1, IF2 and IF3) are also required for the formation of this complex, as discussed below:

- **IF1** and **IF3** increase the rate of dissociation of the ribosomal subunits and maintain the dissociated state, respectively.
- **IF2**, which contains a bound GTP, is required for the binding of the initiator tRNA to the 30S initiation complex.

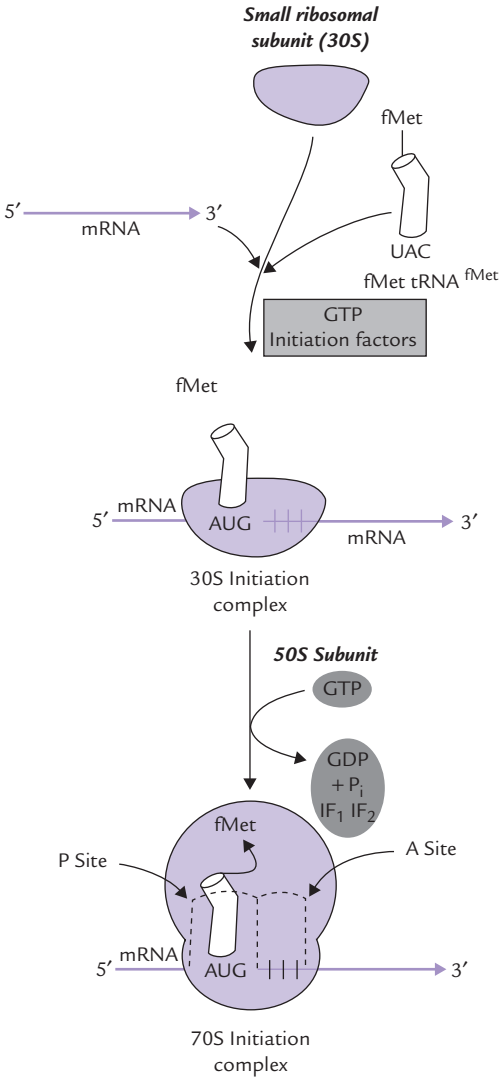


Fig. 22.11. Initiation reactions of protein synthesis require an initiator tRNA charged with formyl methionine (fMet tRNA^{fMet}). The ribosomal subunits assemble with the fMet tRNA^{fMet}, the mRNA and initiation factors along with GTP, to form 30S initiation complex first and 70S initiation complex subsequently.

At this stage, the 50S subunit joins the 30S initiation complex; simultaneously, the GTP molecule is hydrolyzed to GDP and phosphate. IF2 and IF3 also depart from the ribosome. Joining of the 50S and the 30S subunits results in the formation of ***70S initiation complex***. In this complex, the initiator tRNA occupies a specific site on the ribosomal surface, called the **P site**. The initiator tRNA is so positioned that its anticodon pairs with the initiating AUG (or GUG) codon on the mRNA. A second site on ribosomal surface, called the **A site**, is still empty (Fig. 22.11). It would be occupied subsequently by the incoming aminoacyl-tRNA molecules during the elongation phase.

Elongation

In this stage, the amino acids are added to the growing polypeptide chain in a stepwise fashion, one at a time, in the amino to carboxy terminal direction. In addition to the 70S complex, the other cellular components required in this stage are as below:

- **Aminoacyl-tRNA**, complementary to the next codon on the mRNA that is positioned in the A site.
- Three soluble proteins: EF-Tu, EF-Ts and EF-G called **elongation factors**.

The elongation cycle proceeds in three steps:

Step 1: Codon Recognition

The cycle begins with placement of an aminoacyl-tRNA corresponding to the second codon (empty) A site on the ribosome. The protein **elongation factor Tu (EF-Tu)**, a GTP-binding protein, is responsible for the specificity of placement: it delivers the complementary aminoacyl-tRNA whereas its anticodon arm binds the codon on mRNA by complementary base pairing (Fig. 22.12).

Placement of an aminoacyl-tRNA on the A site is not a passive process. It requires energy that is obtained by hydrolysis of GTP, bound to EF-Tu. Furthermore, EF-Tu is important for ensuring that correct codon-anticodon base pairing occurs. If, however, the anticodon does not base pair with the mRNA codon, the EF-Tu quickly dissociates it and sets it free. If, however, the codon-anticodon match, the bound GTD is hydrolyzed to GDP and phosphate; the EF-Tu then vacates the complex and the aminoacyl-tRNA remains in the A site, ready for peptide bond formation.

The GDP remains tightly associated with EF-Tu. The EF-Tu is made available for another round of protein synthesis by a reaction cycle of this elongation factor (Box 22.3).

Step 2: Formation of a Peptide Bond

At this stage, both A and P sites of a ribosome are occupied by an aminoacyl-tRNA and the stage is set for the formation of peptide bond (Fig. 22.12). This is accomplished by transfer of the formylmethionine unit from the initiator codon (in the P site) to the amino group of the amino acid residue on the aminoacyl-tRNA (in the A site) to from a dipeptidyl-tRNA (Fig. 22.13). This reaction does not require an external energy source, because the free energy content of the ester bond in the fMet-tRNA^{fMet} (~7.0 Kcal/mole) exceeds the energy required to form a peptide bond (1.0 Kcal/mole) that of peptide bond (≈1 Kcal/mole). The enzyme catalyzing this reaction *peptidyl transferase*, of the large ribosomal subunit is not a ribosomal protein. Rather it is an enzyme activity of the 23S RNA in the large subunit. The 23S rRNA, therefore is a **ribozyme**, or an RNA enzyme.

At the end of the second step of the elongation cycle, an uncharged tRNA occupies the P site, whereas a dipeptidyl-tRNA occupies the A site (Fig. 22.12).

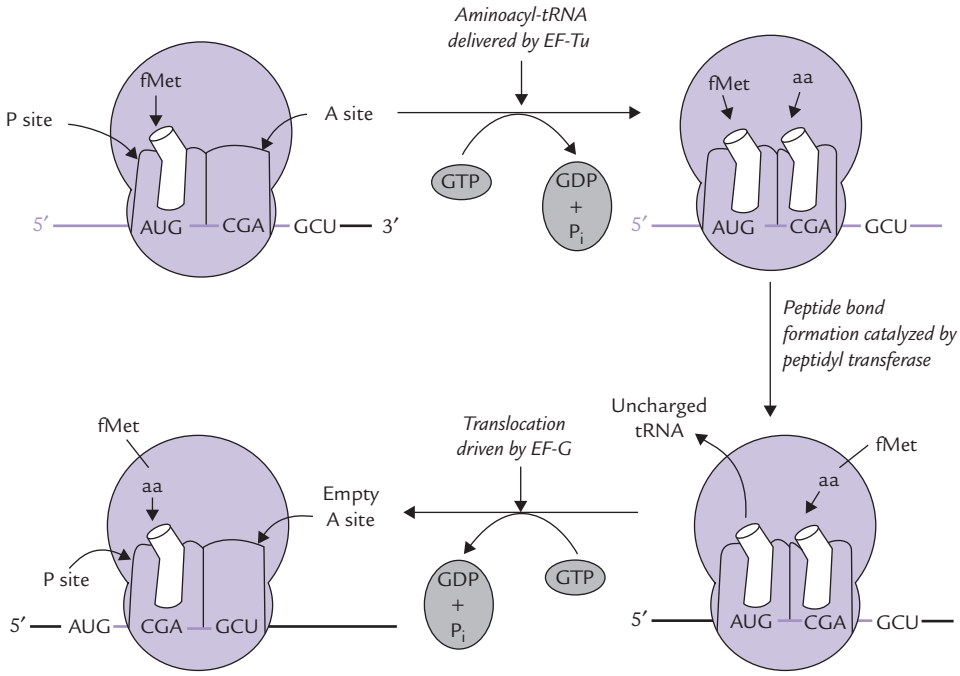
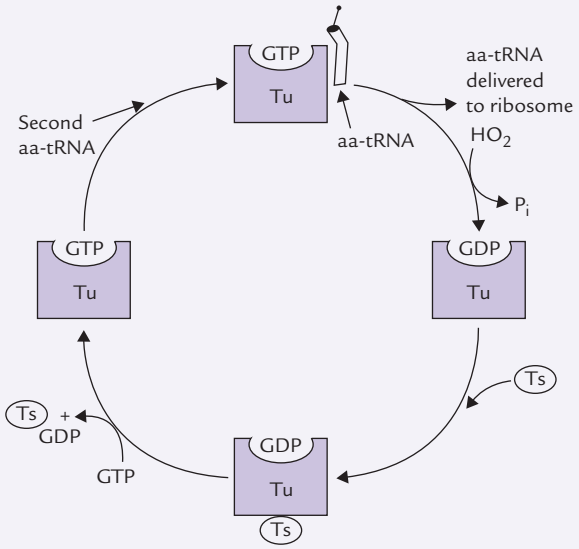


Fig. 22.12. Elongation phase of protein synthesis: codon recognition, peptide bond formation and translocation.

BOX 22.3

Reaction Cycle of Elongation Factor Tu

The elongation factor Tu is involved in delivering the correct aminoacyl-tRNA at the A site of ribosome. The energy for this process comes by GTP hydrolysis. GDP so formed remains tightly attached with the EF-Tu. The following reaction cycle ensures availability of this elongation factor for the next round.



The Tu-GDP complex is bound by another elongation factor, EF-Ts. A complex of Tu-Ts and GDP is thus formed. It is then split by binding of GTP to form Tu-GTP complex; release of Ts-GDP concomitantly occurs. The Tu-GTP complex is now ready to enter the next round of protein synthesis.

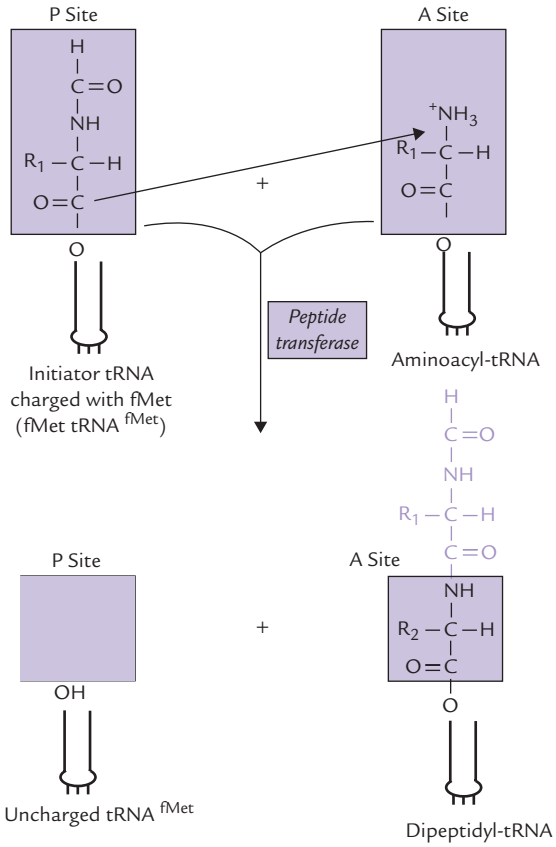


Fig. 22.13. Formation of the first peptide bond through transfer of N-formylmethionine unit to the amino group of the second aminoacyl-tRNA.

Step 3: Translocation

In the final of the elongation cycle, an energy driven movement of the dipeptidyl-tRNA from the A site to the P site occurs, and the uncharged tRNA is released from the ribosome. Codon-anticodon pairing remains uninterrupted during translocation, and therefore, ribosome moves along the mRNA by three bases. Translocation requires a GTP-binding **elongation factor**, called **EF-G**. Hydrolysis of the bound GTP provides energy for this step.

This completes a single elongation cycle. In this cycle, the dipeptidyl-tRNA is present at the P site, and the A site is ready to receive the appropriate aminoacyl-tRNA. Occupancy of the A site by aminoacyl-tRNA starts another cycle in which the same series of events are repeated. In this way, the elongation and translocation steps are repeated until a termination codon moves into the A site.

The ribosome synthesizes polypeptides at a rate of approximately *15 amino acids per second*.

This is comparable with the speed of RNA synthesis, which is about 20 nucleotides per second.

Termination

Translation termination requires **termination factors** (also called *release factors*) that recognize **stop codons** (UAA, UAG, and UGA). These codons do not have matching tRNAs. When one of these codons appears in the A site, one of the two termination factors (RF1 and RF2) binds with it. RF1 binds with UAA and UAG; RF2 binds to UAA and UGA. A third termination factor, RF3, contains bound GTP. It associates with the RF1 or RF2 and the complex causes the protein (that is attached to the last tRNA molecule in the P site) to be released. This process is an energy dependent reaction catalyzed by the hydrolysis of GTP, which transfers a water molecule to the end of the protein, thus releasing it from the tRNA. After release of the newly synthesized protein, the ribosomal subunits, tRNA and mRNA dissociate from each other.

Some stepnic in protein synthesis require help of soluble cytoplasm proteins, which are known as initiation factors, elongation factors and termination factors. Their action is driven by GTP hydrolysis.

Protein synthesis is energetically expensive: Ribosomal protein synthesis requires input of considerable energy, which is provided by both ATP and GTP:

- Two high-energy phosphate bonds are required for the formation of each aminoacyl-tRNA. Additional ATPs may be required at this step for the correction of errors.
- GTP energy is required at various steps during each elongation cycle: for placement of the aminoacyl-tRNA, for translocation and for termination.
- GTP energy is also required for the formation of initiation complex.

Thus, formation of each peptide bond requires expenditure of at least **five** *high-energy phosphate bonds.* Evidently, protein synthesis is an expensive process. In rapidly growing bacterial cell, protein synthesis may consume 30–50% of the total metabolic energy. Humans spend around 5% of the basal metabolic energy for this purpose.

Expenditure of large amount of energy during protein synthesis may appear wasteful. However, it is one of the important factors making *nearly perfect fidelity* possible, in translation of the genetic message. Moreover, it provides a large thermodynamic push in the direction of protein synthesis.

Polysomes

In prokaryotes, the efficiency of translation is greatly enhanced by polysomes, the clusters of 10 to 100 ribosomes. These ribosomes are closely spaced along a mRNA chain and translate it simultaneously (Fig. 22.14). Activity of polysomes allows highly efficient use of the

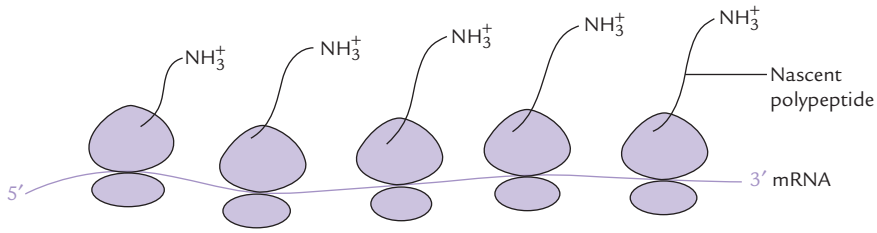


Fig. 22.14. Diagrammatic representation of a polysome.

mRNA. This is important because the bacterial mRNAs have extremely short half lives. They are degraded by the *nucleases* within a few minutes of their formation. Therefore, they must be translated with maximum efficiency; and this is made possible by polysomes.

Each ribosome can synthesize only one polypeptide at a time, but an mRNA has a number of ribosomes lined up.

In bacteria, the processes of transcription and translation are tightly coupled. The mRNAs start getting translated even before transcription is complete. Such close coupling is not possible in eukaryotes as the mRNAs have to be transferred out of the nucleus into cytosol, before they can be translated.

F. Inhibitors of Protein Synthesis

A number of antibiotics bind to various sites on ribosomes and interfere with individual steps of protein synthesis. Since growth and survival of the cell is not possible without protein synthesis, use of antibiotics eliminates the invading bacteria. Thus, antibiotics are potent tools with which infectious diseases are treated.

Some antibiotics that inhibit eukaryotic and prokaryotic translation are described below. Because of structural differences between prokaryotic and eukaryotic ribosomes, most of these antibiotics are selective either for prokaryotes or for eukaryotes.

1. *Streptomycin* binds to the 30S ribosomal subunit of prokaryotes, causes misreading of mRNA and thereby prevents formation of the initiation complex.
2. *Tetracycline* binds to the 30S ribosomal subunit of prokaryotes and inhibits binding of aminoacyl-tRNA to the A site.
3. *Chloramphenicol* competitively inhibits the peptidyl *transferase* activity in prokaryotes, thereby interfering with elongation of the peptide chain.
4. *Erythromycin* binds to the 50S subunit of prokaryotes. It prevents translocation.

These four compounds affect protein synthesis on 70S ribosomes of prokaryotes. Mitochondria also contain the 70S type ribosomes and so these compounds inhibit mitochondrial protein synthesis as well.

5. *Ricin and cyclohexamide* target eukaryotic 60S and 80S ribosomes, respectively.
6. *Puromycin* has structural resemblance with the aminoacyl-tRNA. It forms a peptide bond with the growing peptide and prematurely terminates protein synthesis. It is active in both prokaryotes and eukaryotes.
7. *Diphtheria toxin* prevents the translocation step in eukaryotes (described in Chapter 24).

Many microorganisms have become resistant to common antibiotics by undergoing mutations.

The inhibition of protein synthesis prevents cell growth, as stated above. But it does not kill the bacteria immediately because most bacteria can survive for considerable periods without protein synthesis. Therefore, *most antibiotics that act on ribosomes are not bactericidal but* ***bacteriostatic*** (streptomycin is an important exception). Moreover, bacteria can become resistant to these antibiotics by mutations that affect the target of drug action (Box 22.4). For example, resistance to streptomycin results because of mutation in the gene for a protein of 30S ribosomal subunit (e.g. S_{12}) to which this antibiotic binds.

G. Post-translational Modification of Polypeptide Chain

The nascent polypeptide chain undergoes several modifications which convert it to the biologically active form. In course of its formation, or following its release, the *polypeptide chain is* ***folded*** *into its native conformation*. The native conformation depends on the primary structure (amino acid sequence) of the polypeptide. It permits formation of maximum number of intrachain interactions, such as hydrogen bonds, or van der Waal forces, ionic,

BOX 22.4

Chemotherapeutic Agents for Treatment of Infections and Cancer

Undesirable life forms are present in the body during infections (e.g. bacteria, fungi, parasites, etc.) and cancers (e.g. the cells having undergone cancerous transformation). Chemotherapeutic agents are used for eliminating these. A good therapeutic agent should not cause damage to the "own cells" in the body. The invading prokaryotic cells (bacteria) can be killed more easily than the eukaryotic cells (fungi or parasites) because prokaryotic cells are more different from the "own cells". By the same logic cancer cells are most difficult to eradicate.

Treatment of infections involves drugs that interfere with transcription or protein synthesis. Rifampicin, an inhibitor of transcription, is used to treat tuberculosis. Streptomycin, tetracycline and erythromycin, the inhibitors of protein synthesis on prokaryotic ribosomes, are used to treat many types of infections. Chloramphenicol affects mitochondrial ribosomes and must be used with caution.

Treatment of cancer involves drugs that interfere with DNA synthesis. For example, 5-fluorouracil prevents the conversion of dUMP to dTMP, thereby reducing formation of thymine nucleotides required for DNA synthesis. Methotrexate also inhibits this reaction by limiting the supply of the coenzyme (THF) required in the above reaction. It does so by preventing formation of the THF from its more oxidized precursor, dihydrofolate (Chapter 18).

Very few drugs are currently available to treat viral infections. Viruses convert the cell's DNA, RNA and protein producing machinery to the generation of viral genes and proteins. Thus, the host's workbench is used for the propagation of the invading virus. A drug that has been used with limited success is azidothymidine (AZT), an analogue of thymidine. This drug is phosphorylated in the cell and inhibits retroviral reverse *transcriptase*, thus preventing the copying of viral RNA into complementary DNA (Chapters 24 and 25). As a result, it serves as a DNA chain terminator. It is being used for HIV infections, related to AIDS.

and hydrophobic interactions. In this way, the linear or one-dimensional genetic message in the mRNA is changed into the three-dimensional protein.

In addition to folding, several **processing reactions** also occur in the polypeptide chain. Some of these are as below:

1. ***Covalent modifications:***
   (i) Phosphorylation: Phosphate group is attached on hydroxyl group of serine, threonine or tyrosine by *protein kinases*.
   (ii) Glycosylation: Carbohydrate side chains are attached to serine or threonine hydroxyl groups (O-linked) or to asparagine (N-linked) within the endoplasmic reticulum or the Golgi apparatus.
   (iii) Hydroxylation: Proline and lysine residues of the α-chains of collagen may be extensively hydroxylated in the endoplasmic reticulum (Chapter 5).
   (iv) Addition of prosthetic groups: Prosthetic groups required for activity of some proteins are attached after the polypeptide chain leaves the ribosome.
2. ***Proteolytic processing:*** Many proteins are initially synthesized as larger, inactive precursors which are not functionally active. Active molecule is formed by proteolytic cleavage of the precursor by specific *endopeptidases* and removal of the cleaved portion. For example, trypsinogen and chymotrypsinogen are larger precursor molecules, which are converted to trypsin and chymotrypsin, respectively, by proteolysis.
3. Formation of disulphide cross-links between cysteine residues link the adjacent polypeptide chains. These linkages are formed soon after the spontaneous folding.
4. Removal of N-terminal formylmethionine occurs in prokaryotes.

V. Protein Targeting

A large number of proteins are synthesized in a cell, which get distributed to various locations. Some of these reside in cytoplasm, whereas others move to specific cellular organelles, or get inserted into membranes, or exported from cell to extracellular functional sites.

Thus newly synthesized proteins must be delivered to a specific subcellular location or exported from the cell where it can carry out its appropriate function. This phenomenon is termed ***protein targeting.***

How do the newly synthesized proteins get directed and coveyed to a specific target sites?

Some mechanism must exist so that the proteins destined for export, for insertion into membranes, or for specific cellular organelles are distinguished from proteins that would reside in the cytoplasm.

A. Targeting of Secretory Proteins

Secretory proteins are synthesized by ribosomes bound to the **rough endoplasmic reticulum (RER).** These proteins have an N-terminal signal peptide (or **signal sequence**), which enables the protein to move across the membrane of endoplasmic reticulum (ER) into the lumen where it folds into its final conformation. Vesicles then bud off from the ER membrane and carry the protein to the Golgi complex where it gets glycosylated. Other vesicles then carry it to the plasma membrane. Fusion of these transport vesicles with the plasma membrane finally releases the protein to the cell exterior.

Signal hypothesis: It refers to a special mechanism that is operative for exportation of secretory proteins (Fig. 22.15). A typical secretory protein differs from a cytosolic protein in having a **signal sequence** usually comprising first 20–30 amino acids on the amino terminal end of the protein. Signal sequences differ in different polypeptides, but all of them have the following common features:

- A highly hydrophobic sequence of 10–15 residues at the centre of the signal sequence; leucine, isoleucine and phenylalanine being common residues in this region.
- At least one positively charged residue (arginine and lysine) near the amino terminal end.
- Five residues, more polar than the hydrophobic core precede the cleavage site at the carboxy terminal.

The signal sequence directs the protein to the ER membrane and targets it into the ER lumen and be exported.

A simplified version of signal hypothesis is shown in Figure 22.15. The signal peptide is recognized by a **signal recognition particle (SRP),** which is a complex of 7S RNA

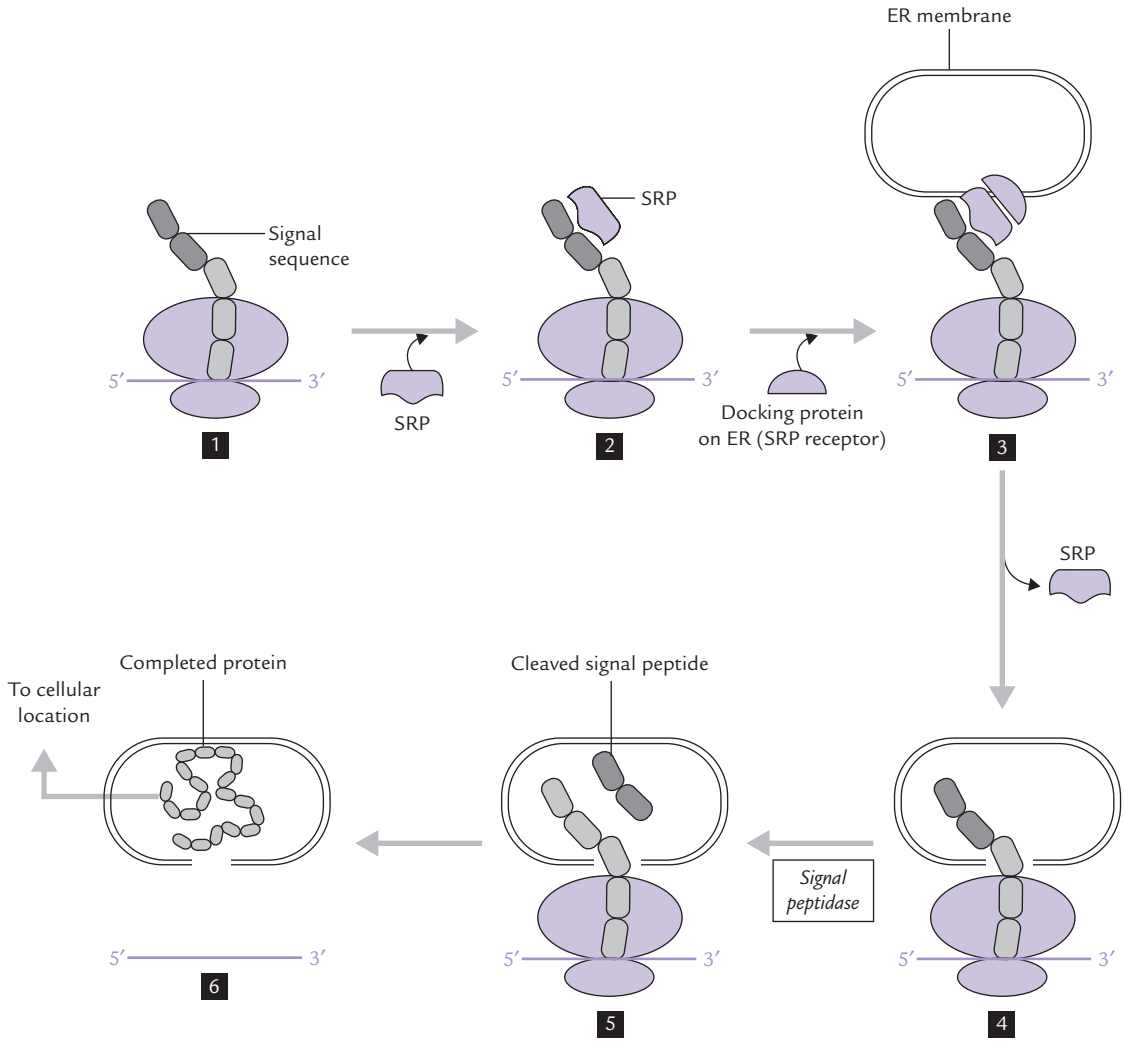


Fig. 22.15. Protein targeting to various destinations in cell occurs by recognition of signal sequence by an SRP (signal recognition particle). The complex of SRP-signal sequence ribosome is recognized by the SRP receptor on ER, and the signal sequence is inserted through a pore in the membrane. Release of signal sequence resumes translation, and the completed protein is delivered to its cellular location.

and six proteins. SRP binds to the signal peptide and stops translation of the remainder of the protein (1–2).

The ribosome-mRNA-SRP complex now binds to a docking protein (called **SRP receptor**), located on the surface of ER. (3) At this stage, the SRP is set free from the signal peptide, and translation now continues once more. By this time a pore is created in the ER membrane (by a *protein translocator*), through which the nascent polypeptide passes into lumen of ER (4).

The signal sequence is no longer required beyond this stage and is removed by a *signal peptidase*. The removal of signal sequence occurs even before the translation is completed (5). The translation continues and the rest of the polypeptide also passes through the membrane pore and enters the lumen of the ER (6). The polypeptide is transferred from ER to Golgi apparatus, and then to its final destination.

Membrane protein targeting: The membrane proteins are also synthesized on the RER but get struck into the ER membrane (and hence ultimately the plasma membrane) rather than being into the ER lumen.

Secretory proteins are transported through the pore in the ER membrane to enter the lumen but the membrane proteins get stuck in the ER membrane.

B. Nuclear and Mitochondrial Protein Targeting

In eukaryotic cells, there exists completely separate types of translocations—into nucleus and into mitochondrial compartment. In contrast to the proteins discussed, which are synthesized by ribosomes on rough ER, synthesis of the proteins destined for mitochondria and nucleus occurs in free cytoplasmic ribosomes. Delivery to their respective cellular destinations is determined by **targeting sequences**. It is not a cotranslational event as these proteins are translocated only after their synthesis is complete.

Nuclear proteins: Proteins synthesized in the cytoplasm translocate into the nucleus via special pores. Short targeting sequences, rich in basic amino acids lysine and arginine, are present throughout the length of the proteins and mark them to pass though these pores.

***Mitochondrial proteins*:** The mitochondrion, as already mentioned, has protein synthesizing machinery of its own, but a vast majority of mitochondrial proteins (95%) must be translocated from the cytoplasmic compartment. These proteins are initially targeted to the outer mitochondrial membrane (OMM), requiring an amino-terminal sequence of about 12–70 amino acids. This sequence is rich in basic amino acids and serine and threonine, and forms an amphipathic α-helix in which one side is positively charged and the other side is largely hydrophobic. One or more of these features attach the protein to its receptor on the OMM. After this initial targeting, the subsequent partitioning between the different mitochondrial compartments (the intermembranous space, the inner membrane, or the mitochondrial matrix) is a complex event, and requires several proteolytic events.

Finally, the mitochondrial proteins, but not the nuclear proteins, must be unfolded before transport. This is explained by the large size of the nuclear pores which can accommodate transport of a protein in its native state.

Exercises

Essay type questions

1. In what ways does RNA synthesis differ from DNA synthesis? Discuss roles of sigma factor and rho factor in the initiation and termination, respectively, of the *E. coli* gene transcription.
2. Compare DNA and *RNA polymerase* with respect to overall structure, substrates, mechanism of action, error rate, and template specificity. Summarize the post-transcriptional modification of prokaryotic mRNA, rRNA, and tRNA.
3. Define genetic code and discuss its characteristic features. Why do you think 61 codons are actually used to specify the 20 amino acids?
4. By means of diagram(s) explain the process of initiation of transcription of a gene in *E. coli.*
5. Discuss actions of inhibitors of transcription and translation.
6. How are newly synthesized proteins delivered to their correct destinations?
7. Describe the participation and, where known, the role of GTP in protein synthesis. Why is translation considered an expensive process?
8. Describe the post-transcriptional modifications that the primary transcript undergoes in prokaryotes.
9. Draw a diagram of tRNA indicating its binding sites. What are the functions of different loops in tRNA?

Write short notes on

1. Sigma factor
2. Ribosomes as work-benches in translation
3. rho factor
4. Codon-anticodon recognition
5. tRNAs as adaptor molecules
6. Characteristics of genetic code
7. Degeneracy of genetic code
8. Initiation codon
9. Shine–Dalgarno sequence
10. Mitochondrial DNA

CLINICAL CASE

CASE 22.1 A child with fast deteriorating vision

A 5-year-old child had difficulty in dark adaptation. Suspecting vitamin A deficiency, his family physician advised vitamin A supplementation. No improvement was, however, observed. On the contrary, deterioration in vision continued and soon there was appearance of scotoma in the visual field. Loss of peripheral vision followed soon, which progressed to such an extent that only small central field of vision was left. Central acuity was also lost subsequently, and the patient was declared legally blind at the age of 18 years.

Presently, the ophthalmoscopic examination showed extensive and widespread degeneration of rods and cones. Retinal pigment deposition around mid fundus was detected.

Defect in some photoreceptor cell protein (i.e. light sensitive retinal pigment) was suspected, and these proteins were subjected to further investigations. Gene for opsin (the protein portion of the visual pigment, rhodopsin) was identified and cloned. No change was detected in the nucleotide sequence. Gene for peripherin—a protein located in the outer segment disk membrane of rods—was also similarly investigated. A point mutation was detected in 216th codon, as below:

Patient	**Normal subject**	
RNA	5′-GUU-3′	5′-CCU-3′
Amino acid	Leucine	Proline

Q.1. How might the mutation affect structure of peripherin?

Q.2. How might the above structural alteration lead to impairment of vision?

MOLECULAR BIOLOGY III: REGULATION OF PROKARYOTIC GENE EXPRESSION AND MOLECULAR GENETICS

CHAPTER

23

Bacterial genome is capable of encoding about 2000 polypeptides, while eukaryotic genome codes for 50,000–100,000 polypeptides. Not all genes are expressed at a given instant; only a fraction finds expression according to body needs. Evidently, elaborate mechanisms exist in all organisms—prokaryotes as well as eukaryotes—that regulate gene expression. Such regulatory mechanisms in prokaryotes and their modes of action are described in this chapter.

After going through this chapter, student should be able to understand:

- Constitutive and inducible proteins, enzyme repression and enzyme induction.
- Fundamental structural design of the **operon** and its genetic components.
- Roles of regulatory gene, repressor and corepressor, catabolite activator protein (CAP), cyclic AMP and attenuator in case of ***lac* operon** and ***tryp* operon**; regulatory mechanisms in viruses; and significance of **protein-DNA interaction** in regulation of transcription.
- Aetiology of various categories of genetic diseases, and learn about the analytical tools to study them.

I. Constitutive and Inducible Proteins

Some proteins, called constitutive proteins, are present at a constant concentration during the life of a cell. Their rates of expression are not selectively modulated, and the genes encoding them appear to be 'switched on' all the time. Enzymes and other proteins that are needed at all times, and in amounts that do not vary from time to time, are constitutive, for example, the glucose metabolizing enzymes. Glucose (the commonest sugar) and other sugars are shunted on to the glucose pathways, therefore these enzymes are needed under all circumstances.

The promoters of the genes coding for the constitutive proteins have no 'on' and 'off' switch, they are always 'on'.

Inducible proteins, on the other hand, are *synthesized only when the cellular need for them arises*. A classical example of inducible proteins in prokaryotes is *β-galactosidase*, the enzyme that converts the disaccharide lactose to one molecule each of glucose and galactose. When *E. coli* grows in a medium containing glucose as the carbon source, *β-galactosidase* is not required and only about five molecules of the enzyme are present in the cell. When bacteria are transferred to a cup of milk wherein the principal sugar is lactose, 5000 or more molecules of *β-galactosidase* occur in the cell. Clearly, bacteria respond to the need to metabolize lactose by increasing the synthesis of *β-galactosidase* molecules.

$$\text{Lactose} \xrightarrow{\beta\text{-galactosidase}} \text{Glucose} + \text{Galactose}$$

If the lactose-induced bacteria are now transferred to a glass of lemonade, which contains sucrose, the synthesis of *β-galactosidase* stops as rapidly as it began. Instead, synthesis of sucrose metabolizing enzyme, *sucrase*, is accelerated now. The phenomenon of enhanced synthesis of these enzyme proteins is called **induction**.

The enzymes of biosynthetic pathways, likewise, are required only when the end product of the pathway is

not available from an external source. The enzymes of tryptophan synthesis, for example, are synthesized only when the cell has to grow in a tryptophan-free medium. Conversely, when tryptophan is available in the medium, synthe sis of these enzymes is suppressed; this phenomenon (of turning off synthesis of enzymes that are no longer needed) is called **repression.**

Regulation of gene expression in eukaryotes is much more complex. Regulation is required not only for making metabolic adjustments but also for tissue differentiation. All cells of our body have the same genes, but different cells synthesize different proteins. Haemoglobin, for example, is synthesized in the bone marrow, and insulin only in the β-pancreatic cells. These differences arise because of cell-specific gene expression. Gene regulation in eukaryotes is described in Chapter 24.

II. Operon: A Coordinated Unit of Gene Expression

As mentioned earlier, *eukaryotic genes are single units but many prokaryotic genes are grouped together,* the individual groups being under transcriptional control of a single promoter. The *RNA polymerase* transcribes through the entire group, thus creating a *polycistronic* mRNA molecule with coding instructions for several proteins. Such a group of genes, organized into a functional unit with its single promoter control, is called an **operon** (the promoter being part of the operon). Apart from these two (**promoter** and **structural genes**) the other genetic elements of an operon include a regulator ***i* gene** (*i* for inducibility), and an additional control site called **operator**.

In the following sections, the fundamental principles of regulation of transcription in prokaryotes is described with a focus on the lactose and the tryptophan operons of *E. coli*. A detailed examination of these operon systems illustrates the basic concepts in gene regulation.

An operon is transcribed as a single unit, giving rise to a polycistronic mRNA that directs more-or-less a simultaneous synthesis of each of the encoded polypeptides.

A. Lactose Operon

An approximate arrangement of various genetic elements of ***lactose (lac) operon*** is shown in Figure 23.1. It contains a **promoter**, an **operator**, the **regulatory *i* gene** and three adjacent **structural genes.** The structural genes code for the related enzymes or associated proteins that are needed to metabolize lactose.

(a) The ***structural genes*** are

- the *Z* gene for *β-galactosidase* that hydrolyzes lactose to free glucose and galactose,
- the *Y* gene for *lactose permease*, which is needed to transport lactose into the cell, and
- the *A* gene for *galactoside transacetylase* that brings about acetylation of various non-metabolizable and potentially toxic β-galactosides. The latter may be imported into the cell and the *transacetylase* is concerned with their removal.

(b) ***Promoter (P)***, *the binding site* for *RNA polymerase,* is located at the upstream end of the three structural genes. Because the genes are contiguous and are transcribed into a single polycistronic mRNA, a single promoter is sufficient for them. The ribosome can synthesize three different polypeptides from this polycistronic mRNA because the coding sequence of each gene ends with a stop codon.

(c) The ***i* gene** codes for a protein called the ***lac repressor***, which binds to short regulatory DNA sequence

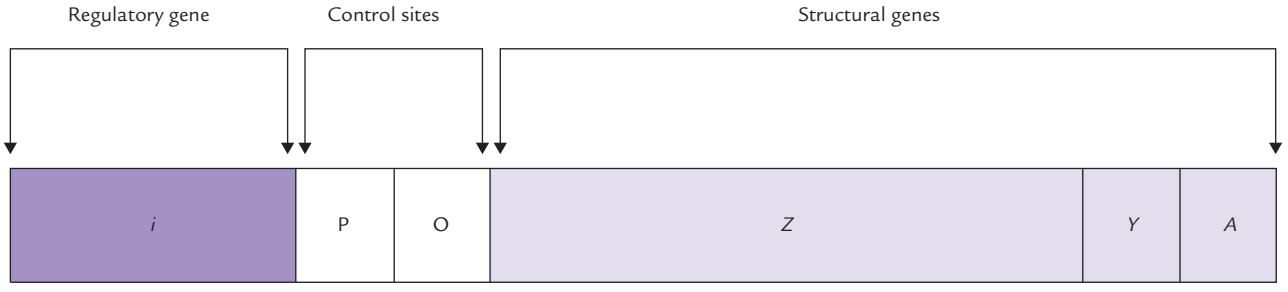


Fig. 23.1. Genetic elements in *lac* operon. Promoter (P), operator (O), and structural genes are contiguous in all bacterial operons, the regulatory gene (i) may or may not be located next to operon (Z = gene for *β-galactosidase*, Y = gene for *lactose permease*, A = gene for *β-galactoside transacetylase*.) O binds repressor protein (encoded by *i* gene) and P binds *RNA polymerase*.

lying between the three protein-coding genes and their shared promoter. This sequence is called the (iv) **operator (O).**

(d) Finally, there is a stretch of DNA adjacent to the promoter to which a regulatory protein can bind (not shown in Fig. 23.1). As described later, this protein is called the **catabolite activator protein** (CAP). It is given this general name because it is involved in the induction of several other enzyme proteins involved in catabolism of their substrates.

Before we systematically go through the control mechanism of *lac* operon, the following points merit attention:

1. The ***lac promoter*** on its own is a **weak** promoter so that the *RNA polymerase* does not readily bind to it and initiate transcription. Therefore, the three gene products are normally made in minute amounts (*basal levels*) because they are not required unless lactose is encountered.
2. When lactose is available as the sole energy source, there is an almost instant burst of synthesis of the three proteins so that the cell can use lactose as a carbon and energy source.

The *E. coli lac* operon contains genes encoding the proteins mediating lactose metabolism and the regulatory genetic sites that control their expression.

Role of *Lac* Repressor in Control Mechanism of the *Lac* Operon

The *lac* repressor is an allosteric, tetrameric protein with four identical subunits. The *i* gene encoding this protein is transcribed constitutively at a low rate. In the absence of lactose in the medium, the *lac* repressor has a strong affinity for the operator. There is some overlap between the operator and the *RNA polymerase*-binding site (promoter), thus the *RNA polymerase* is unable to bind to the promoter when the operator site is occupied by the *lac* repressor (Fig. 23.2a). This prevents the transcription of the structural genes.

But if lactose is present in the medium (Fig. 23.2b), this repressed state of the gene is altered because of the formation of a small amount of 1,6-allolactose. This minor product of *β-galactosidase* reaction binds to the repressor protein, causing an allosteric change in the

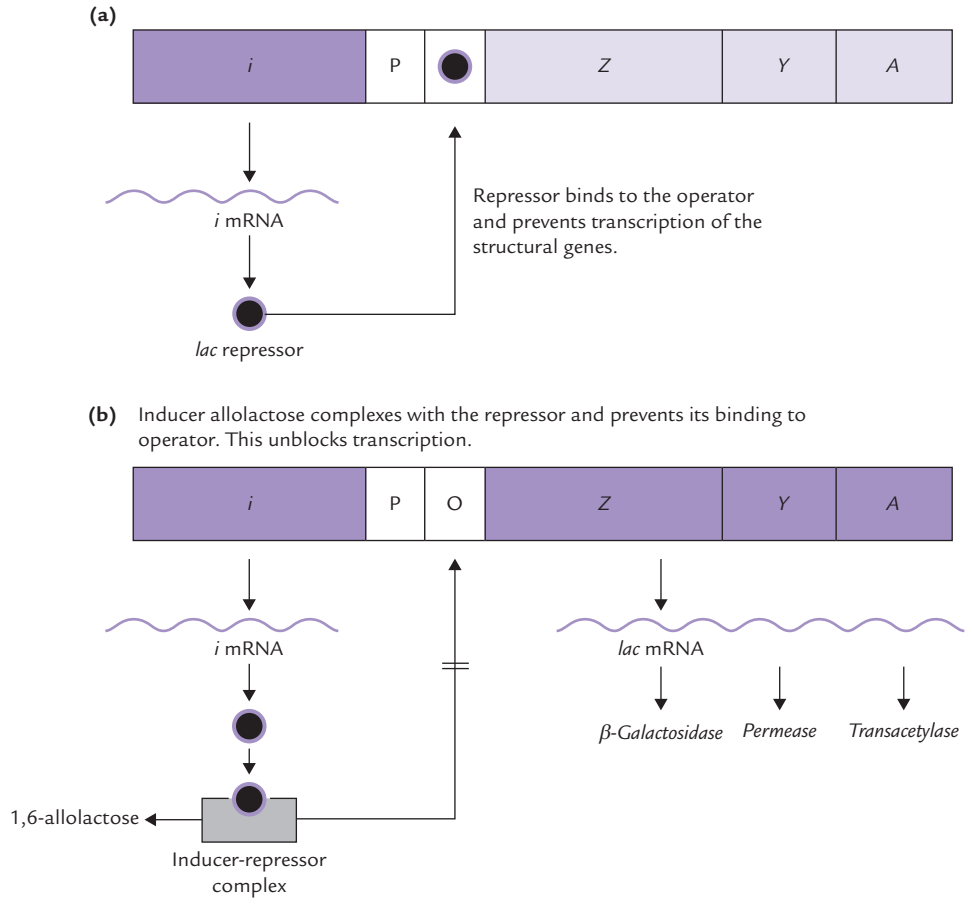


Fig. 23.2. Repressed state (**a**) and induced state (**b**) of *lac* operon. Allolactose, a minor side product of lactose is the inducer, and the protein encoded by the *i* gene is the repressor.

protein, which then dissociates from the operator and unblocks the operon. With the unblocking of the operon, the *RNA polymerase* is now free to move down the operon and transcribe the structural genes. Thus, 1,6-allolactose functions as an **inducer** of the *lac* operon.

Role of Catabolite Activator Protein

As noted earlier, promoter of the *lac* operon is intrinsically weak and permits a high rate of transcription only after binding of a catabolite activator protein (CAP). This protein binds to a stretch of DNA adjacent to promoter. This fascilitates interaction of the promoter with *RNA-polymerase*, thereby causing a high rate of transcription.

However, activity of CAP is dependent on cAMP; it can bind to DNA only after cAMP is bound to it.

When CAP binds, the promoter becomes strong. However, CAP does not bind unless cAMP is bound to it.

CAP is an allosteric protein (a dimer of two identical 22 kD subunits), whose activity is modulated by cAMP level in the cell. The amount of cAMP produced in the cell is in turn dependent on the availability of glucose. It (cAMP) is produced rapidly when glucose is not available, and is scarce when glucose is plentiful. Two situations are commonly encountered:

1. ***When glucose is scarce:*** The cAMP is produced, the cAMP-CAP complex readily forms and binds the CAP site; *RNA polymerase* can now bind to the promoter to enhance transcription of the *lac* operon.
2. ***When glucose is plentiful:*** cAMP is scarce in the cell, CAP does not bind, the *RNA polymerase* therefore does not bind effectively, and *lac* operon transcription is minimal.

These situations are illustrated in Figure 23.3.

The above discussion shows that the *enzymes of lactose metabolism are induced only when glucose is depleted* (Fig. 23.4). This is advantageous because glucose is metabolized more easily than lactose and the bacterium saves expense for the synthesis of *lactase*. Not only the *lac* operon but several other catabolite operons are repressed in the presence of glucose. This phenomenon is called **catabolite repression**.

Glucose regulates the transcription of many catabolic operons.

How does CAP facilitate interaction of *RNA polymerase* with the promoter site? As shown in Figure 23.3 the binding sites for *RNA polymerase* and the CAP-cAMP complex are juxtaposed, and the CAP-cAMP complex provides additional site of interaction for *RNA polymerase*. Consequently, binding of the *RNA polymerase* is enhanced and transcription is accelerated by factor of 50.

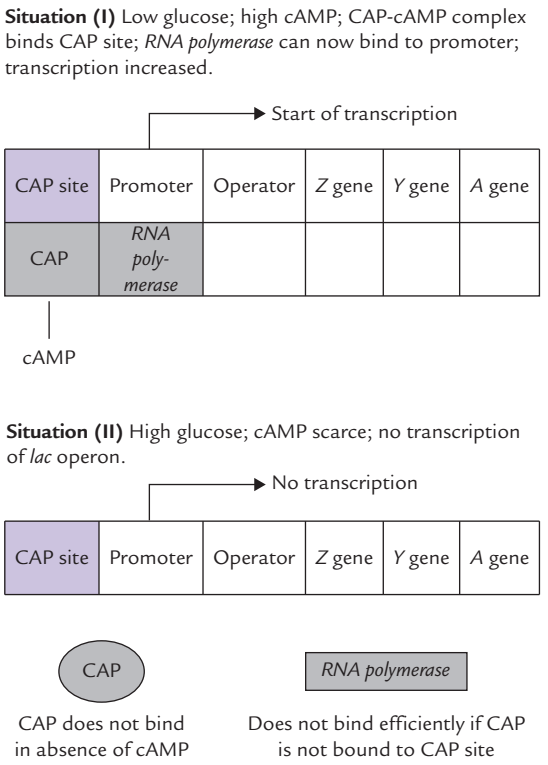


Fig. 23.3 Expression of *lac* operon in presence of high or low levels of glucose.

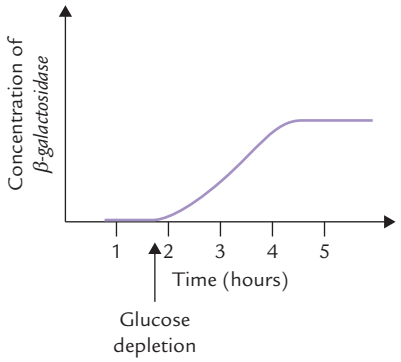


Fig. 23.4. Induction of enzymes of lactose metabolism by glucose depletion.

B. Tryptophan Operon

The *lac* operon was the first understood example of prokaryote operon control, but it applies to many other pathways, such as *tryptophan* operon (*trp* operon). Transcription of *trp* operon is controlled by two mechanisms: **operator control** and **attenuation**.

Operator Control

Trp operon is an anabolic operon; it contains the five structural genes (*E, D, C, B, A*) needed to produce the enzymes involved in tryptophan synthesis (Fig. 23.5), and is similarly controlled by a tryptophan repressor protein encoded by a separate operon. In this case, tryptophan bound to the repressor protein causes it to block the operator site. This prevents most of the mRNA transcription so that the expression of structural genes is blocked. Conversely, when tryptophan is not present in the medium, the repressor is not bound to the operator site, and thus the gene transcription is permitted.

Many anabolic operons contain genes encoding proteins required for biosynthetic pathways. The product of the pathway acts as (co)repressor of the pathway.

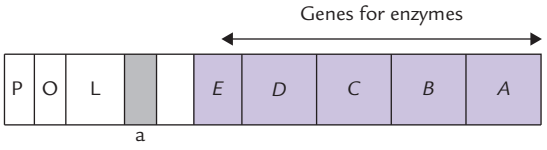


Fig. 23.5. Genetic elements in *tryptophan* operon showing control sites; **promoter** (P), **operator** (O) and **attenuator** (a). Genes for the leader sequence (L) and the enzymes of the tryptophan pathway (*E, D,* C, *B* and *A*) are also shown.

Control by Attenuation

Operons responsible for the biosynthesis of amino acids, including the *trp* operon are controlled by an attenuation mechanism also. This mechanism can *sense change in the intracellular level of the amino acid and adjusts the rate of transcription accordingly*. Its operation depends on the fact that in prokaryotes translation of mRNA starts as soon as a small stretch of the mRNA molecule has been synthesized, and therefore *polymerase* is closely followed by the ribosome doing the translation.

The operon initially codes for a peptide only 14 amino acids in length; it is called the *leader peptide*. It contains tryptophan residues at the amino acid positions 10 and 11. This has significant effect on the regulation, as discussed below:

Met-Lys-Ala-Ile-Phe-Val-Leu-Lys-Gly-**Trp-Trp**-Arg-Thr-Ser

When tryptophan is plentiful, the ribosome promptly adds these residues in the attenuation leader sequence and moves along to the next segment (the segment 2) of mRNA. As it interacts with the segment 2, the next two segments (3 and 4) internally base pair to form a termination hairpin structure, the transcription is thereby aborted (Fig. 23.6).

When tryptophan is scarce, the ribosome is stalled at the tryptophan codon of the segment 1. The segment 2 of the mRNA now interacts with segment 3 and therefore segments 3 and 4 cannot base pair. Thus, the termination

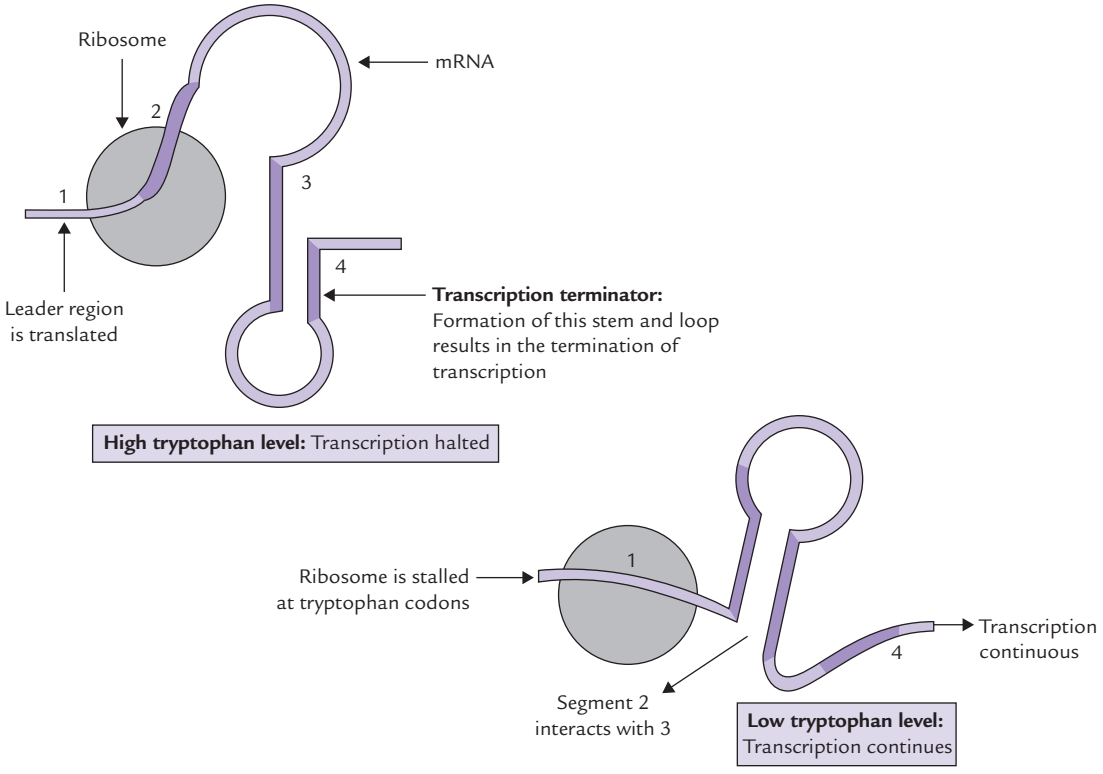


Fig. 23.6. Control of termination of transcription in tryptophan operon depending on concentration of this amino acid.

structure is not formed and therefore, the transcription continues.

Tryptophan operon encodes for six-gene products that are required for the synthesis of tryptophan.

To summarize:

- *When tryptophan is abundant*, a rapid translation of the primary transcript occurs. This causes termination of transcription by generating a secondary structure in mRNA, which acts as a termination signal for *RNA polymerase*.
- *When tryptophan is scarce*, translation becomes slow and so the termination structure is not formed. Therefore, transcription continues under these circumstances.

Note: The classical operon models described do not apply to eukaryotes. The functionally related protein-coding genes of eukaryotes are widely separated, not transcribed as polycistronic mRNA.

III. Protein-DNA Interaction in Transcriptional Regulation

Role of protein-DNA interaction is of vital significance in regulating gene transcription. The proteins are broadly divided into two types: **specific factors** and **regulatory proteins**.

A. Specific Factors

These proteins help *RNA polymerase* (RNAP) to recognize specific sites where transcription begins (i.e. the promoter sequences; Box 23.1). **Sigma (σ)** subunit of *RNA polymerase*, discussed earlier, is a typical example of such a specific factor. It is involved in specific promoter recognition and binding. It alters specificity of *RNA polymerase* for a given promoter (or set of promoters).

Any alteration in this protein is naturally expected to alter specificity of initiation. For example, exposure to heat produces a change in the σ-subunit (sigma 70 to sigma 32 conversion occurs). Sigma 32 now recognizes a different set of promoters. The latter controls the expression of a set of genes that make up the heat-shock response. Thus, the holoenzyme containing a different sigma subunit called σ^{32} recognizes heat shock promoters, which leads to synthesis of a series of protective **heat-shock proteins (HSPs)**. The HSP enables the bacteria to adapt to the challenge posed by heat.

Note: σ^{32} is called so because it has a mass of 32 kD.

B. Regulatory Proteins

Regulatory proteins bind with the DNA sequence present in the vicinity of the transcriptional initiation site (i.e. the operator region) and thereby regulate the rate of transcription. Examples include *lac* repressor, CAP-cAMP complex, tryptophan holo-repressor, discussed earlier. The regulatory proteins are actually *allosteric proteins with binding and non-binding conformations*. They can recognize the regulator DNA sequences because the edge of the DNA bases are exposed in the major and minor grooves of the double helix. Some general features of regulatory proteins are described here.

1. Activities of the regulatory proteins are controlled through interaction with various modulators such as cAMP, 1,6-allolactose and tryptophan. However, other control mechanisms are also known, such as covalent modulation or interaction with other regulatory proteins (i.e. protein-protein interaction).
2. Binding of a regulatory protein to the regulatory DNA sequence can have dual effect on the rate of transcription; the rate can be increased or decreased. For example, binding of CAP-cAMP (to the operator) enhances the rate of transcription, and therefore, this effect is called ***positive control***. Conversely, binding of *lac* repressor and tryptophan holo-repressor can

BOX 23.1

Promoter Sequences

Sequence of nucleotides in the promoter region determines affinity of *RNA polymerase* for this region. In *E. coli*, certain "*consensus sequences*" have been worked out, which have extremely high affinity for *RNA polymerase*. Resemblance with the consensus sequence determines affinity of a given promoter sequence for *RNA polymerase*. For example, the promoter sequence for genes coding for the central catabolic enzymes show remarkable resemblance with the consensus sequences. As a result, they have high affinity for the corresponding *RNA polymerase*, which in turn results in high rate and frequency of transcription.

Regulation at the level of promoter sequences is the only way in which expression of the constitutive genes is regulated.

diminish the transcriptional rate: this effect is called ***negative control***.

3. The regulator proteins are **oligomeric** in nature. For example, CAP is a dimer and *lac* repressor is a larger oligomeric protein consisting of multiple subunits. The subunits exhibit positive cooperativity which enhances the net response of these proteins to allosteric modulators.
4. Structures of the repressor proteins exhibit **two-fold symmetry** that matches the near palindromic base sequence of the operator. *Symmetry matching is an important general feature of protein-DNA interactions.*

Regulation during initiation of transcription is the best documented regulatory point. As in all other metabolic processes, the first step is the efficient stage for regulation. Unnecessary investment of energy is prevented by an early regulatory point. Moreover, initiation of transcription is an excellent point for coordinated regulation of several genes whose products have interdependent roles. For example, following DNA damage, several enzymes of DNA repair have to be synthesized, coordination of which best occurs at the level of transcription.

Transcriptional regulation is based on protein–DNA interactions. DNA sequences near the transcriptional start site bind regulatory proteins resulting in enhancement of transcription (**positive control**) or inhibition (**negative control**).

It is worthwhile to mention that regulation of gene expression can occur at various other levels also, apart from regulation of the synthesis of primary RNA transcript, as described in this chapter. Other processes that serve as potential points for regulation are:

1. Post-transcriptional processing of mRNA.
2. Degradation of mRNA.
3. Regulation of translation.
4. Post-translational processing of the polypeptide chains.
5. Protein degradation.

Role of some of these processes in regulating gene expression in eukaryotes is described in Chapter 24.

IV. Regulatory Mechanism in Viruses

The regulatory mechanism is quiet complex in viruses the infectious particles that consist of nucleic acid and protein. Although the viruses have no cellular structure and metabolism, they can replicate within a host cell by exploiting the host's cell enzymes, metabolic energy and ribosomes. DNA viruses substitute their own genomic DNA for the host-cell DNA, directing synthesis of viral mRNA and viral proteins. Thus, DNA viruses (and indeed all viruses) abuse the host-cell ribosome for the synthesis of their own proteins.

When a virus, for example, lambda phage, infects a bacterial cell, it can follow a lysogenic pathway or a lytic pathway. In **lysogenic pathway** the virus integrates its DNA into the host-cell chromosome where it replicates with the latter; and in the **lytic pathway** the viral DNA replicates independently till the host cell is destroyed. For details refer to Box 23.2.

BOX 23.2

The Viral Lifestyles within the Host: Lysogenic and Lytic

After the virus enters a cell it can immediately replicate and produce new virus particles that destroy its host cell by brute force (lytic); or it chooses an alternative lifestyle (lysogenic), wherein it integrates its DNA into bacterial chromosome. The integration requires a virally encoded *integrase*. The viral DNA now becomes part of the host-cell chromosome, where it can remain inert *indefinitely* being replicated with the host DNA. The integrated viral DNA is called a prophage, the bacterium is characterized as lysogenic and this mode of virus replication is called the **lysogenic pathway**. It produces a large population of otherwise normal bacterial cells, each carrying a prophage.

The **lytic lifestyle** is so called because it is accompanied by the lysis (destruction) of the host cell. A single lytic cycle takes some 20 minutes and approximately 200 progeny viruses are released from the lysed host cell. In each infection cycle, synthesis of viral mRNA and coat proteins occurs. New virus particles are assembled from the replicated viral DNA and the newly synthesized coat proteins. The phage coded enzymes: a *phospholipase* and a lysozyme are synthesized close to the end of the infection cycle; these enzymes destroy the bacterial cell wall resulting in cell lysis.

Exposure to ultraviolet light, certain chemicals or ionizing radiations may change lysogenic lifestyle into lytic lifestyle. Following the exposure there is activation of the SOS functions designed to institute repair, but in lysogenic cells these functions cause excision of the prophage from DNA by a mechanism too complex to be described here. The excised prophage enters the lytic cycle of replication.

What determines the pathway—lysogenic or lytic—the virus would take? Lambda phage virus illustrates this well. The viral DNA contains an operator region known as the *right operator* (O_R), which comprises three distinct segments: O_R1, O_R2 and O_R3 (Fig. 23.7). Upstream of the operator region is a **repressor gene**, and downstream there is a ***cro* gene**. Alternative expression of either the repressor gene or the *cro* gene determines whether the phage proceeds on the lysogenic or the lytic pathway.

In ***lysogenic pathway*** the repressor gene of the lambda phage becomes activated and codes for a repressor protein. The latter binds to O_R1, and this interaction represses the expression of *cro* gene, at the same time promoting further expression of the repressor gene. *The repressor protein is believed to maintain the lysogenic state, probably by preventing the transcription of all other phage genes* (except its own). It even makes the lysogenic bacterium resistant to re-infection by lambda phage because the genes of any incoming lambda phage become repressed as well.

In ***lytic pathway*** the bacterial *protease*, *rec* A, hydrolyzes the repressor protein, thus removing the inhibition on the *cro* gene. The latter is expressed to form *cro* protein, which binds to O_R3, the site adjacent to the repressor gene, to inhibit the latter. This in turn switches off the synthesis of the repressor protein, and thereby removing inhibition on transcription of phage genes, including the genes of the lytic pathway. (One of such de-repressed genes, the *xis* gene, codes for an *excisionase* that dissects the prophage out of the bacterial chromosome). In this way there is tilting of the balance entirely towards the lytic pathway, and interestingly, like rats leaving the sinking ship, the virus ultimately leaves its troubled host.

V. Genetic Diseases

After studying the organization of the prokaryotic gene, its expression and regulation of expression, the stage is now set for understanding the eukaryotic gene which is much more complex. The haploid genome of a eukaryotic cell contains more than 100,000 functional genes, which are scattered among billions of base pairs (3.5×10^9). While the details about eukaryotic gene expression will be presented in the next chapter, a brief introduction to human genetic diseases is given here.

More than one-third of the genes present in each human being are different from the ones present in the majority of the population. This remarkable degree of genetic variation among normal people accounts for much of the naturally occurring variation in body traits, such as height, intelligence, colour of skin, etc. Moreover, these genetic differences produce marked variations among individuals to handle environmental challenges, some of which can cause diseases. Thus, *every disease can be visualized as net outcome of interplay of genetic make up of a given individual and the surrounding environment*. Significance of the role played by each of these two factors, varies. In certain diseases, the genetic component is so overpowering that it always expresses itself in a predictable manner. Extraordinary environmental challenges are not required for their expression. Such diseases are termed **genetic diseases**.

A. Analytical Tools

Analysis of genetic material is tedious and time consuming due to its extreme complexity. Consequently, the study of genetic diseases has been difficult in the past. For example, in a given genetic disorder, the genetic alteration may involve a single or a few genes only, which is negligible compared to the total size of the human genome. It is an extraordinarily tedious task to pinpoint a defect from such vast expanse of genetic material.

With remarkable advances made in molecular biology and development of computers in recent years, the nucleotide sequences of long stretches of DNA have been found out. The *Human Genome Project* has completed the task of deciphering the entire sequence of the human genome. Because of such advances, it is now possible to get a clearer insight into genetic disorders.

Some enzymes acting on DNA and certain techniques have contributed towards a better understanding of genetic disorders. Some of these are discussed below:

1. *Restriction endonucleases* cause cleavage of huge DNA with remarkable precision to produce smaller defined

Fig. 23.7. Expression of repressor gene promotes lysogenic pathway and expression of cro gene favours lytic pathway.

fragments which are easier to handle and analyze (Chapter 25).

2. *Cloning techniques* amplify a given DNA fragment. The amplification implies production of several identical copies of DNA from a single parent DNA.
3. *Synthesis of specific probes* permits identification of a specific nucleotide fragment and its subsequent manipulation.

B. Categories of Genetic Disorders

The genetic diseases have been divided into five major groups, as shown in Table 23.1.

Chromosomal Disorders

The human diploid genome comprises 46 chromosomes. There are two copies of each chromosome, therefore a total of 23 pairs of chromosomes. Loss, gain, or abnormal arrangement of one or more of these chromosomes results in chromosomal disorders.

Of the 23 pairs of chromosomes, 22 pairs are **autosomes**, and the remaining two chromosomes are the **sex chromosomes**, XX or XY. Females have two X chromosomes and males have one X chromosome and one Y chromosome. The most important of the autosomal syndromes involve a *numerical abnormality*; sometimes there may be presence of three rather than two copies of a particular chromosome. Such defects are referred to as ***autosomal trisomies***, most important of which is **trisomy 21** (three copies of chromosome 21). It is also the commonest autosomal defect and is responsible for about 95% cases of *Down's syndrome* or *mongolism* which is characterized by mental retardation. **Trisomy 13** is characterized by *cleft lip*. In addition, certain chromosomal disorders involve sex chromosomes. These include the disorders with 47 chromosomes where an extra sex chromosome is present (XXY, XYY and XXX), or those with 45 chromosomes where a sex chromosome is lacking (XO).

Chromosomal abnormalities are common; about 20% of the pre-implantation embryos have chromosomal defects. Spontaneous abortions eliminate most of these, reducing the live-birth frequency to about 0.6%.

Table 23.1. Classification of genetic disorders

I	Chromosomal	Excess loss translocation
II	Single gene defect	Autosomal dominant Autosomal recessive X-linked: dominant or recessive
III	Multifactorial	Involves a number of genes
IV	Somatic cell gene defect	
V	Mitochondrial mutations	

Single Gene Defect

These disorders are because of defect of a single gene, i.e. there is a primary error in the DNA code. They display simple (Mendelian) inheritance pattern.

The single gene defects can be subdivided according to the chromosome on which the mutant gene is situated, and also based on the nature of the trait itself. Thus a trait that is determined by a gene situated on autosome is inherited as an **autosomal trait**; and this may be either *dominant* or *recessive*. A trait determined by a gene situated on one of the sex chromosomes is known as sex-linked, which may also be either dominant or recessive. About 1200 diseases of this type are known, some of which are listed in Table 23.2. With a few exceptions, most of these diseases have rare occurrence. However, their importance cannot be ignored because, as a group, they constitute an important cause of morbidity and death.

The term **dominant** implies that a mutation will clinically manifest when an individual has a single dose of this mutation (i.e. a case of heterozygosity). **Recessive** implies that a *double dose* (or a case of homozygosity) is required for clinical manifestation. It is noteworthy that the genes themselves are not dominant or recessive; the clinical pattern they produce may be classified as dominant or recessive.

Since a single mutant gene is responsible for a particular disease, it can be concluded that pathogenesis of the disease is due to an **abnormality in a single protein molecule**.

Table 23.2. Some single gene defects

Autosomal dominant	Autosomal recessive	X-linked recessive
Achondroplasia	Albinism	Glucose 6-phosphate
Gilbert's syndrome	Cystic fibrosis	*Dehydrogenase* deficiency
Hyperlipoproteinaemia (Type II)	Fanconi's syndrome Glycogen storage diseases	Hunter's syndrome
Osteogenesis imperfecta	Phenylketonuria Niemann–Pick disease	Nephrogenic diabetes insipidus
Acute intermittent porphyria	Tay–Sachs disease Wilson's disease	Lesch–Nyhan syndrome

For example, in sickle cell anaemia, having thymine instead of adenine at a specific site in the gene that codes for the β-chain of haemoglobin, produces a substitution of valine for glutamic acid in the protein sequence. The abnormal protein that is produced due to a single gene defect may be a structural protein (as in sickle cell anaemia); or an enzyme protein (as in phenylketonuria). In several other disorders also, a single gene alteration may result in impaired activity of the protein it encodes (**Case 23.1**).

Understanding of the genetic defects at a molecular level have given new significance to the old genetic terms "recessive" and "dominant". A **recessive pattern** of inheritance is usually seen in disorders in which the affected protein is an enzyme. In a heterozygote for such disorder, the gene that is normal provides sufficient enzyme to prevent clinical symptoms. The **dominant** inheritance, on the other hand, is commonly seen with the disorders in which some structural protein is affected. This is understandable in view of the fact that presence of a single defective structural protein, even in combination with other normal components, can affect function of the overall structure. Thus, the heterozygote with one normal and one defective gene (coding for structural protein) manifests clinically.

The genes responsible for the X-linked disorders are located on the X chromosome. Therefore, the clinical risk and severity of the disease are different for the males and the females. Since a female has two X chromosomes, she may be either heterozygous or homozygous for a mutant gene. Males, on the other hand, have only one X chromosome, so they manifest clinically whenever they inherit the gene.

Multifactorial Disorders

Some diseases, such as diabetes mellitus, hypertension, and manic depression do not show simple Mendelian mode of inheritance because they are caused by an *interaction of multiple genes and multiple exogenous or environmental factors.* The disease may develop due to a complex interplay of a variety of factors and, therefore, have a multifactorial aetiology. Diabetes mellitus, for example, is a genetically programmed auto-immune disease in which genetic factors, auto-immunity and possibly some viruses play an important role.

The multifactorial disorders develop more often in the immediate family of the affected person compared with the general population. Further, if the disease develops in one of the identical twins, the other twin who is genetically identical (monozygotic twins) has greater chance of developing the disease.

Somatic Cell Gene Defects

Certain mutations in the somatic cells lead to **tumourigenesis**. They are not inherited except by the cells in the lineage of the mutated cell.

Mitochondrial Mutations

Mutations in the mitochondrial genes also may give rise to abnormalities. Mitochondrial myopathies and Leber's H optic neuropathy (LHON) are disorders arising due to abnormalities in the mitochondrial DNA.

Treatment of genetic disorders will be described in Chapter 25.

Exercises

Essay type questions

1. Draw a neat sketch to show various genetic elements in a lac operon and describe how they function. Outline the advantages and disadvantages of arranging genes in an operon.
2. Describe structure and function of tryptophan operon.
3. Describe regulatory mechanisms for gene transcription in viruses.

Write short notes on

1. Catabolite repression
2. Lysogenic pathway
3. Lytic pathway
4. Protein-DNA interactions
5. Tryptophan operon
6. Polycistronic mRNA
7. Consensus sequences in promoters

CLINICAL CASE

CASE 23.1 A 1-year-old child with fatigability and breathlessness

A 1-year-old child was suspected of having congenital methaemoglobinaemia based on his clinical history and low concentration of *cytochrome b_5 reductase* in erythrocytes.

Gene for *cytochrome b_5 reductase* was isolated and sequenced. A mutation was detected in the codon 57: instead of 5′-CGC-3′ found in normal individuals, 5′-CAG-3′ was present. *Cytochrome b_5 reductase* of this patient was heat labile and more susceptible to trypsin treatment than that of a normal individual.

Q.1. How does the above mutation decrease intracellular level of *cytochrome b_5 reductase*?

Q.2. Provide an explanation for the oxygen deficiency symptoms of this patient?

CHAPTER 24

MOLECULAR BIOLOGY IV: EUKARYOTIC CHROMOSOME AND GENE EXPRESSION

Size of human genome is much larger and its organization is much more complex than that of prokaryotes. It contains about 1000 times more DNA than *E. coli* (3.1×10^9 base pairs vs 4.2×10^6 base pairs). The genomic material of a human cell is packaged in 23 pairs of chromosomes. Each chromosome contains only one duplex DNA, which may be several centimeters long. In contrast to the circular shape of the bacterial DNA, the **eukaryotic DNA is linear.** A unique feature of the latter is the presence of vast expanse of non-coding DNA sequences. These sequences separate some 1000–6000 genes present in a single DNA molecule. According to the findings by Human Genome Project, human beings contain **33,000–44,000 genes** in their genome.

The total length of all the DNA in a single human cell is about 2 m (in *E. coli* it is 1.4 mm only). The DNA is tightly packaged in association with proteins called **histones.** Structural features of the complex between the two, referred to as nucleohistone, have been described earlier (Chapter 21).

The processes of DNA replication, transcription and translation in eukaryotes are fundamentally similar to those of prokaryotes. However, they are more complex, and have certain distinctive features. Specialized features of eukaryotic gene and its expression will be discussed in this chapter.

After going through this chapter, the student should be able to understand:

- Differences in the structural organization of genome of eukaryotes (from prokaryotes) and know their implications for gene control.
- Role of eukaryotic *DNA polymerases;* steps of eukaryotic cell cycle and biochemical events in each step; details of steps of transcription (i.e. binding of *polymerase,* initiation, elongation and termination); and post-transcriptional modification and processing of RNAs (mRNA, rRNA and tRNA).
- Fundamental differences between the eukaryotic and the prokaryotic protein biosynthesis and various levels at which control of gene expression can be exercised in eukaryotes.

I. Replication has Multiple Sites of Origin

DNA replication has thousands of origins in humans. Approximately one origin exists for every 10,000 to 100,000 base pairs. Bidirectional replication from these origins, termed **replication bubbles,** proceeds semiconservatively until the daughter strands are complete. The new duplexes then separate; each contains a parent and a daughter strand (Fig. 24.1).

Multiple origins make the replication process sufficiently fast so as to complete replication of a DNA molecule in about 9 hours. It is noteworthy that eukaryotic (*DNAPs*) are much less efficient when compared to their bacterial counterparts. The α-*DNAP* (a eukaryotic enzyme) can add only 100 base pairs per second, 10 times slower than the prokaryotic *DNAP III*. This may be because of more complicated arrangements of DNA in eukaryotic chromosomes. Considering the large size of eukaryotic chromosome, it would take long time (up to 3 weeks) to replicate it fully if there were only one pair of replication

fork per chromosome. Multiple sites of origin ensures that replication of total genome takes less time than replication of a bacterial chromosome.

The eukaryotic DNA replication starts from several origins—approximately one in every ten thousand to one lakh base pairs, and therefore replication does not take much longer in humans compared to bacteria.

All regions of a chromosome are not replicated simultaneously. Rather, many replication bubbles (consisting of two replication forks moving in opposite directions) will be found in one region of the chromosome, but none in the other. The ends of chromosomes, called ***telomeres*** play a vital role in eukaryotic replication.

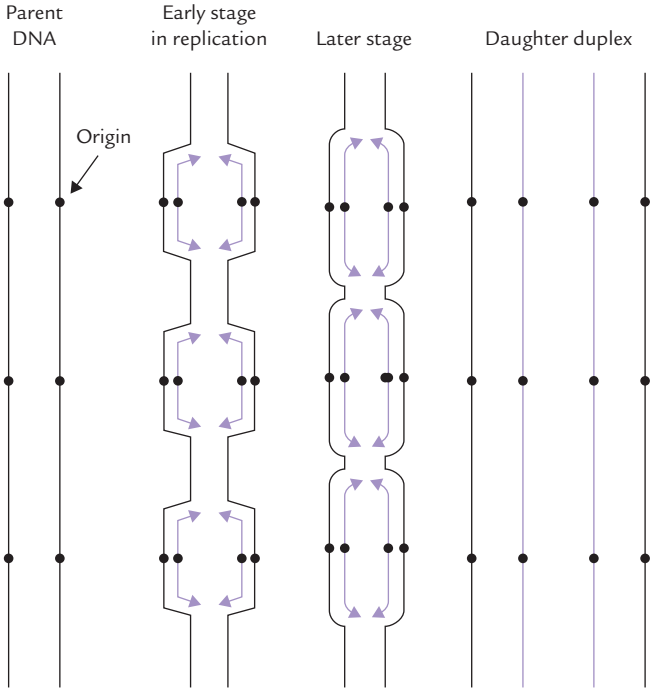


Fig. 24.1. Replication of the eukaryotic chromosomal DNA starts from multiple origins, one every ten thousand to one lakh base pairs (••• indicates sites of origin).

A. Eukaryotic *DNA Polymerases*

Five distinct eukaryotic *DNA polymerases* are known, which are designated in order of their discovery. These are: *DNA polymerases* **α, β, γ, δ** and **ε** (Table 24.1). The eukaryotic *DNAPs* are not as well characterized as their bacterial counterparts.

Polymerase α is composed of four distinct subunits with different catalytic activities. One of its subunits constitutes a *primase* enzyme, capable of synthesizing short RNA transcript that can serve as primer for DNA chain elongation. *Polymerase* α is a *moderately processive* enzyme: it can add about 100 nucleotides per binding event. Its activity correlates with cellular proliferation, being high when there is rapid cell division and dropping to a low level in quiescent cells. Evidently, this enzyme has a direct role in chromosomal DNA synthesis.

It was believed that *polymerase* α is the *polymerase* for the synthesis of the leading strand. However, recent evidence suggests that *polymerase δ may be the key enzyme for the leading strand, whereas polymerase α is primarily for the lagging strand*. This appears beneficial as the enzyme with moderate processivity is more suited for the enzyme synthesizing the lagging strand. The *polymerase δ* possesses proofreading ability *(3′–5′ exonuclease* activity), which contributes to the fidelity of replication.

Polymerase β and ε are **for repairing** the damaged DNA like δ-DNAP; ε-DNAP also has high processivity and possesses 3′ to 5′ *exonuclease* activity. *Polymerase γ* most likely copies the mitochondrial DNA. It possesses 3′ to 5′ *exonuclease* activity though it is not significant. Therefore, error rate during mitochondrial replication is much higher than that during chromosomal replication. The *polymerase* γ has some *primase* activity.

All eukaryotic *DNAPs* except γ are located in the nucleus; the *γ-DNAP* is a mitochondrial enzyme. Repair and polymerization are performed by different enzymes, and likewise different enzymes appear to be involved in synthesizing the leading and the lagging strands.

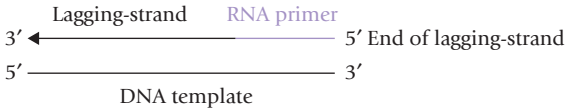
Table 24.1. Properties of eukaryotic *DNA polymerases*

	α	β	γ	δ	ε
Location	Nucleus	Nucleus	Mitochondrion	Nucleus	Nucleus
Mass (kD)	>250	36–38	160–300	170	256
Function	Lagging strand synthesis	Repairs leading strand	Mitochondrial replication	Leading strand synthesis	Repairs lagging strand
Primase activity	Yes	No	Yes	No	No
Inhibitors	Aphidicolin	Dideoxy NTPs	Dideoxy NTPs	Aphidicolin	Aphidicolin

All DNAPs except α and β possess 3′ to 5′ *exonuclease* activity.

B. Role of Telomeres in Eukaryotic Replication

The ends of mammalian chromosomes are formed by the telomeres (Greek: *telos* = end) which consist of the repeat sequence (TTAGGG)n. This sequence is repeated in tandem (one after another) between 500 and 5000 times in each telomere. Telomeres play a crucial role in eukaryotic DNA replication. Without them a linear chromosome would become progressively shorter with each cell division and the cell's descendants would ultimately die from loss of essential genes. This is because of the inability of *DNA polymerase* to synthesize the extreme 5′ end of the lagging strand.



As diagrammed above, the *DNA polymerase* can only extend an RNA primer that is paired with the 3′ end of a template strand.

Removal of this RNA primer would leave a gap that cannot be filled by *DNA polymerase* (recall that *DNA polymerase* operates only in 5′ to 3′ direction; it can only extend an existing primer, and the primer must be bound to its complementary strand; Fig. 21.13). Consequently, in the absence of a mechanism for completing the lagging strand, DNA molecule would be shortened by the length of an RNA primer with each round of replication.

Mechanism must exist to prevent such shortening, referred to as **telomere erosion.** This is accomplished by *telomerase*, the enzyme which synthesizes and maintains the telomeric DNA. It adds tandem repeats of the **telomeric sequence** TTGGGG at the extreme 5′ end of the lagging strand. This telomeric sequence now acts as primer: the *DNA polymerase adds* an okazaki piece at its 3′ end to initiate a new round of replication.

Telomerase acts without benefit of template. Therefore, it provides the template itself. The template is 150-nucleotide RNA which is an integral part of the enzyme. Thus *telomerase* is a ribonucleoprotein (complex of protein and RNA).

Telomerase, Aging and Cancer

Telomerase is required for mortality. This is evident from the following observations:

1. Germline cells which express *telomerase* are immortal.
2. Somatic cells that are not capable of expressing *telomerase* have only a finite life span. Lack of *telomerase* activity may contribute to cell aging and ultimate death.
3. Besides germline cells, cancer cells express *telomerase* and are immortal.

Though lack of *telomerase* activity in somatic cell appears to be a disadvantage, does it offer some selective advantage to multicellular organisms? A possibility is that mechanism of cellular aging and death protects multicellular organisms from cancer.

Because *telomerase* apparently stabilizes even short telomeres, inhibition of *telomerase* should eventually lead to complete loss of telomeres and hence cessation of cell division. This hypothesis makes *telomerase* inhibitors an attractive target for anti-tumour drug development.

C. Cell Cycle

In eukaryotic cells, replication of DNA and cell divisions occur at different times, separated by gap phases. These processes ultimately lead to the formation of a pair of daughter cells from a parental eukaryotic cell, and are together referred to as the cell cycle.

Total duration of the cell cycle ranges from 12 to 24 hours. This period is mainly occupied by two active phases: the **synthetic phase** (S-phase) and the **mitotic phase** (M-phase), as shown in Figure 24.2. Intervening between these two active phases are **gap phases** termed G_1 and G_2.

- During the **M-phase,** growth and division of the cell (**mitosis**) occurs. Duration of this phase is about one hour.
- During the **S-phase**, *duplication of DNA* occurs. New DNA strand is synthesized based on the formation in the pre-existing DNA template to form two daughter DNA molecules. A copy of the daughter DNA would subsequently, pass into each of the daughter cells. The S-phase occupies about 7 hours.
- The **G_1-phase**, which precedes the S-phase, is the *preparatory phase for replication*. In this phase, the enzymes

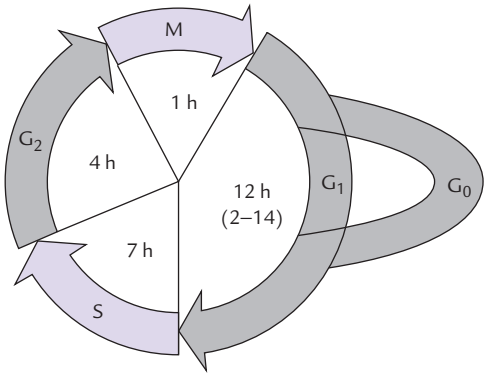


Fig. 24.2. Diagrammatic representation of the cell cycle. Mitotic (M) and synthetic (S) are the active phases, while G_1 and G_2 are the intervening gap phases. G_0 is the silent phase wherein cells are not involved in the cell cycle, but can enter the cycle as and when required.

necessary for replicating the genome are produced and the genomic DNA is checked for integrity. Its duration is variable (12 hours: range, 2–14 hours).

- In the **G_2-phase**, which follows the S-phase, the *completion of replication* of genome is confirmed. It occupies 4 hours.

Finally, at the end of a cycle the duplicate daughter chromosomes are segregated, the nuclear membrane is disrupted and the daughter chromosomes are distributed equally on either side of the parental cell.

The S-phase refers to the time when DNA synthesis occurs, and the M-phase refers to the time of mitosis. G_1 and G_2 are gap phases, before and after DNA synthesis respectively.

G_0 phase: The events of the cell cycle are ordered, as shown in Figure 24.2. However, the cells leaving the M-phase are not absolutely committed to entering the G_1 phase; they may enter a silent phase called G_0 phase. Cells in the G_0 phase are said to be viable but not in the cycle; they are capable of entering the cell cycle under appropriate circumstances. Duration of this phase shows great variation: few hours in most cells to several years in neurons. The duration may be subject to availability of growth factors.

Regulation: Various phases of the cell cycle are meticulously controlled. A number of regulatory enzymes are involved to ensure correct operation of the cell cycle. The regulatory process is very complex, but a simplified account is presented below. The regulatory enzymes are *protein kinases* which are commonly referred to as the *cell cycle kinases*. A cyclic variation in catalytic activities of these enzymes is observed as the cell cycle proceeds from one phase to another. Activities of *protein kinases* in turn depend on regulatory subunits, called **cyclins.** The cyclins make periodic appearance during the cell cycle. Although they are synthesized throughout the cell cycle, they are abruptly destroyed during mitosis.

Accumulation of the cyclins during cell cycle promotes phosphorylation of the *protein kinases.* This results in activation of the latter (these enzymes being active in the phosphorylated state), and the activated enzyme causes phosphorylation of chromosomal scaffold proteins, histones and nuclear lamins. These changes signal onset of mitosis. This is followed by appearance of *proteases* which remove the cyclins. The result is M-phase to G_1-phase transition.

II. Untranscribed Human DNA

The organization of eukaryotic DNA is strikingly different from prokaryotic DNA. *E. coli* has 4.2 million base pairs, most of which encode for proteins. A total of approximately 2000 polypeptides are encoded by the bacterial genome. In contrast, *most of the human DNA is not involved in protein synthesis.* Only about 3–4% is active in this regard and the rest of the 96–97% DNA is non-coding and remains untranscribed. This can be estimated from the fact that the size of human genome (3.1×10^9 base pairs) is nearly a thousand times as large as that of the *E. coli* genome, which is sufficient to code for more than 15 lakh polypeptides. But the number of polypeptides it actually encodes is much smaller (about 100,000); it is about 50-fold, rather than a 1000-fold, greater than the number of polypeptides encoded by the *E. coli* genome. Thus, it is far less compact (Box 24.1). *Most of the non-coding sequences are present between the adjacent genes, while a few are present within the gene.* Since no obvious function is performed by the transcriptionally inactive DNA, it is referred to as the **junk DNA**.

Most of the human DNA remains untranscribed.

What is the function of the non-coding DNA which exceeds several billion base pairs? A convincing answer is yet to come. No doubt, this has been a mysterious scenario, it is challenging as well to elucidate the function of this seemingly useless DNA.

BOX 24.1

Human DNA is Less Compact than Bacterial DNA

Human DNA is strikingly different from bacterial DNA in being less compact as far as its expression is concerned. For instance, the functional globin genes of humans in the α- and β-clusters are far apart from one another. These clusters together occupy 35–60 kb, which means that average of about 10–12 kb DNA encodes a single globin chain. But the actual amount of DNA needed for encoding a globin chain (made up of an average of 150 amino acid residues) is only 0.45 kb. This is less than about 4% of the DNA clusters. This observation can be accounted only by presence of large segments of non-coding, junk DNA.

III. Gene Distribution Along DNA

Genes are transcribed as continuous sequences, but only some segments of the resulting mRNA molecule contain information that codes for the protein product of the gene. These sequences are called **exons.** The regions between exons are called **introns**, which are the untranslated, intervening sequences lying within the gene. Thus the exons are interspersed with long stretches of introns. *The introns are transcribed along with exons, but the intron sequences are subsequently removed by specific enzyme systems.* After removal of these sequences, a continuous coding sequence is produced (Fig. 24.3). Most human genes have few to several dozens of introns. The total length of introns may exceed that of exons by several times.

Exons are those parts of the gene that are represented in mature RNA, whereas introns do not appear in the functional RNA product.

A. Repetitive Sequences in Human Genome

A vast majority of DNA in human genome consists of unique or nearly unique sequences. Other sequences are repeated hundreds or thousands of times and are scattered throughout the genome. They constitute approximately 30–40% of the human DNA. Two major classes of the repetitive sequences have been recognized: **highly repetitive sequences** (also called simple sequence DNA) and **moderately repetitive sequences.**

Highly Repetitive Sequences

These are present at $> 10^6$ **copies** per genome. They are clusters of nearly identical oligonucleotide sequences, each between two and a dozen base pairs, that are repeated over and over in tandem arrangement (next to each other) in a DNA molecule. Highest concentrations of such sequences are present (i) in centromere, (ii) in telomere regions of eukaryotic chromosomes, and (iii) smaller agglomerations are located outside these, in junk DNA.

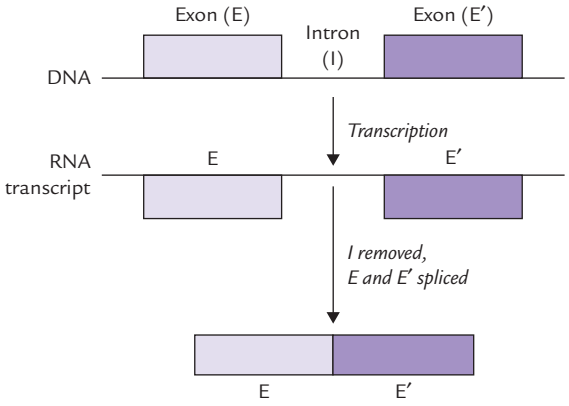


Fig. 24.3. Production of a functional RNA molecule in eukaryotes.

(i) **Centromeres**, the region attached to the microtubular spindle during mitosis, help align chromosomes and facilitate recombination. They contain short oligonucleotide sequences that are repeated more than a million times in the genome.

(ii) **Telomeres**, as already stated, form ends of mammalian chromosomes and consist of a repeat sequence (TTAGGG)n. This sequence is repeated in tandem between 500 and 5000 times in each telomere.

(iii) Outside telomere and centromere, the **repetitive simple sequence DNA** is present in junk DNA of introns or untranscribed spacers. Each simple repeat unit consists of 2–5 base pairs. A repeat unit occurs from few to 50 times in a given location and the same repeats may be present in a number of different locations. The most abundant repeat is a dinucleotide repeat (AC)n that occurs at about 50,000 to 100,000 different sites. Repeats of 3, 4, or 5 bases are also common, although they are not as numerous as the dinucleotide repeats.

At least a dozen human neurological diseases result from excessively repeated *trinucleotide sequences* (Table 24.2, Box 24.2).

Satellites, Minisatellites and Microsatellites

The DNA in centromeric and telomeric regions is called **satellite DNA**. After partial *endonuclease* digestion, when the mixture of pieces is analyzed by a density gradient method, most of the DNA runs in one band, but the repetitive DNA runs as a 'satellite' (separate from the main band). Likewise, the repetitive DNA fragments outside centromeric and telomeric regions present as *microsatellites* (total length below 1000 nucleotides) or as *minisatellites* (total repeat length more than 1000 nucleotides).

Table 24.2. Some diseases associated with trinucleotide repeats

Disease	Repeat	Site of repeat
Fragile X syndrome	GGG or CCG	5′ UTR
Myotonic dystrophy	CTG	Upstream region, 3′ UTR
Friedrich's ataxia	GAA	Intron
Spingobulbar muscular atrophy	CAG	Exon
Huntington's disease	CAG	Exon

BOX 24.2

Trinucleotide Repeat Diseases

Several human neurological diseases called trinucleotide repeat diseases are caused due to excessively repeated trinucleotide sequences. In some types of these diseases, massive expansion (usually hundreds of copies) of a trinucleotide occurs in the non-coding region of a gene, such as in a region upstream of the transcription start site, in the 5′ or 3′ untranslated region (UTR), or in an intron (Table 24.2). Once the expansion reaches the critical stage, it begins to interfere with gene function and then result in a clinical syndrome. For example, in myotonic dystrophy (that results from repeated CTG sequences in the upstream region and 3′ UTR), aberrant expression of a *protein kinase* occurs. The severity of the symptoms, such as progressive muscle weakness and wasting, correlates with the number of CTG repeats ($>$2000 in some cases).

Other types of trinucleotide repeat diseases are caused by moderate expansion of a CAG triplet in the protein coding region of a gene. This triplet codes for glutamine. Expansion of this triplet yields non-functional proteins that kill cells, particularly in the nervous system. Such as loss of neurons occurs in *Huntington's disease*, a devastating condition characterized by progressively disordered movements (chorea), cognitive decline, and emotional disturbances. The disease is genetically dominant and invariably fatal, with an age of onset of ~40 years and follows a 10 to 20 years course. The repeated CAG sequences are present in a gene which codes for a 3145 residue polypeptide called **Huntington**. Normally they are present in 11–34 copies but increase to 37–876 copies in affected individuals.

The trinucleotide repeat diseases exhibit *genetic anticipation*, which refers to worsening of clinical phenotype with successive meioses. It is due to progressive enlargement of the trinucleotide expansions during meiosis. Because overall length of the repeat sequence(s) correlates with the age of onset of the disease, descendants of an affected individual tend to be more severely affected and at an earlier age.

Moderately Repeated Sequences

Such sequences ($< 10^6$ copies per haploid genome) occur in segments of 100 to several thousand base pairs that are interspersed with larger blocks of unique DNA. The best characterized one of these sequences belongs to **Alu family**: human genome contains 300,000–500,000 widely distributed *Alu sequences* that are on average 80–90% homologous with their consensus sequence (Alu is named so because most of its ~300 base pair long segments contain a cleavage site for the restriction *endonuclease Alu I*).

The Alu sequences and similar sequences measure up to few hundreds. Bases are referred to as **short interspersed repeat elements** (*SINEs*). In addition, there are **long interspersed repeat elements** (*LINEs*), which are relatively longer sequences of about 5000 base pairs. They may occur up to 10,000 times per cell. Like SINEs, the LINEs are also transcriptionally active and are usually mobile segments of DNA.

No definite function has been assigned to the moderately repetitive sequences, which are therefore termed as selfish DNA. This otherwise harmless DNA appears to be a molecular parasite that, over many generations, has disseminated itself throughout the genome through transposition. Since it puts additional metabolic burden on the replication machinery of the cell, theory of natural selection would predict its elimination. Yet this has been avoided probably because relative disadvantage of replication of additional base pairs of selfish DNA in a billion base pairs genome is so slight that its rate of elimination would be balanced by its rate of propagation.

The repetitive sequences are produced by the duplication of sequences, as discussed under gene families and pseudogenes in Chapter 21. This causes formation of additional copies of the genes which have no effect on the phenotypes.

B. Exon Shuffling

Exons from different locations are believed to have combined to yield new functional genes. This is important in view of evolution of genes. *Exons appear to be the building blocks from which various new genes have been constructed in course of the evolution.* For example, fibronectin, a connective tissue protein, consists of repetitive modules; some of these modules have close relatives in the genes for some proteins of the clotting system.

Relation of Introns to Evolution

Though no specific function has been unravelled for introns, they are believed to be important for the evolution of eukaryotic genes. During evolution, a new trait is produced by development of new genes. Mutations in the existing gene produces new gene. However, mutation in the functional gene may have deleterious consequences for the organism. Hence the non-functional genes are

mutated to produce new traits. Simply stated, the introns may be prototype of the future functional genes.

IV. Transcription

Transcription is more complex in eukaryotes. It occurs mainly in the nucleus, where only one of the two DNA strands, i.e. the antisense (–) strand is copied by *RNA polymerases*. Transcription also occurs in mitochondria to a limited extent where both strands are copied, but one is subsequently degraded. Not more than 3–4% of the human DNA codes for proteins, and the remaining 96–97% consists of the non-coding sequences. *Most of the non-coding sequences are transcriptionally inactive and lie between the adjacent genes.*

Other non-coding sequences are present within the genes. These non-coding sequences (called introns) are transcribed, but they do not contain information that codes for the gene's protein product. They are removed by nuclear enzymes before the mRNA reaches ribosomes.

Most of the human genes are patchwork of exons (whose mRNA transcripts are processed to mature RNA) and the untranslated intervening sequences, the introns. After removal of introns, the exons form a single, continuous coding sequence (Fig. 24.3).

Some genes have only 1–2 introns, others may have several dozens, whereas some are intronless (e.g. histone genes).

A. Types of *RNA Polymerases*

In contrast to a single *RNA polymerase* (RNAP) in prokaryotes, in eukaryotes there are **four *RNA polymerases***—three nuclear and one mitochondrial (Table 24.3). This was demonstrated by studying the action of α-amanitin (a toxin from mushroom, *Amanita phalloides*) which inhibits different RNAPs to different extents. Molecular weight of these enzymes range from 500,000 to 600,000.

Table. 24.3. Properties of eukaryotic *RNA polymerases*

Type	Location	RNA product synthesized	Sensitivity to α-amanitin
I	Nucleolus	5.8S, 18S and 28S rRNAs	Insensitive
II	Nucleoplasm	mRNA precursors (hnRNA)	Extremely sensitive
III	Nucleoplasm	Pre-tRNAs, 5S rRNA	Moderately sensitive
Mitochondrial	Mitochondria	Mitochondrial RNAs	Insensitive

RNAP I is found in the nucleolus. It synthesizes a large 45S RNA precursor which is subsequently cleaved to three smaller molecules (5.8S, 18S and 28S rRNAs). **RNAP II** is the principal enzyme which synthesizes almost all mRNA precursor, also known as *heteronuclear (hn) RNA*. It is present in nucleoplasm and is inhibited by low concentration of α-amanitin. **RNAP III** is responsible for the formation of small species of RNAs, such as 5S rRNA and tRNAs. It is moderately sensitive to inhibitory action of α-amanitin, being inhibited only by high concentration of the latter. It also transcribes the gene for the 7S RNA associated with signal recognition particle (SRP). The latter is involved in the translocation of proteins across the endoplasmic reticulum membrane.

Eukaryotes have separate *RNA polymerases* for the synthesis of mRNA, rRNA and tRNA. *RNA polymerases* transcribe defined segments of DNA into RNA with a high degree of selectivity and specificity.

B. Chemistry of mRNA Synthesis

The basic processes involved in eukaryotic mRNA production are the same as in prokaryotes. The process occurs in three stages: **initiation**, **elongation** and **termination**.

Initiation

About 10^5 initiation sites are present in the entire DNA. RNAP recognizes these and moves along the DNA rapidly, incorporating about 1000 base pairs per second in the new RNA.

How does RNAP recognize the correct DNA strand and initiate RNA synthesis at the beginning of a gene? RNAP binds to its initiation site through the base sequence known as the **promoter**—the basic recognition unit signalling that there is a gene nearby and can be transcribed. A promoter lies on the same strand as the gene being transcribed and is commonly referred to as ***cis*-acting element**. Other *cis*-acting elements include *response elements* and *enhancers*, which (together with promoters) influence specificity and efficiency of transcription. Although the promoters are essential for the initiation of transcription, they are not alone in influencing the efficiency of transcription.

Promoters: These are perhaps the most important and fundamental elements required for initiation of transcription. Promoters are nucleotide sequences in front (upstream) of the start site, located up to –200 base pairs

Table 24.4. Some promoter elements in eukaryotes

Element	Location	Sequence
TATA box	Centred at –25 bp	TATAA
GC box	Between –40 bp and –110 bp	GGGCGG
CAAT box	Between –40 bp and –110 bp	GGCCAATCT
Ig octamer	Up to –200 bp	ATTTCGAT

bp = base pairs.

(Table 24.4). They ensure correct positioning of the *RNA polymerase II* at the transcription start site (i.e. the first base copied into RNA), which is designated as number 1. Promoters of different genes look quite different, but some basic key elements occur with great regularity in all of them. These elements may be present in varying combinations, some elements being present in one gene but absent in another. However, some form of promoter element is virtually always present.

Some of the common elements found within the promoter region are: **initiator region, TATA box,** and the **upstream elements** such as GC and CAAT boxes.

- ***Initiator region:*** The start site itself has a recognizable short sequence between nucleotides −3 and +5, called initiation (Inr). It is the simplest form of promoter known to be recognizable by RNAP II, and its structure varies from gene to gene. However, in general, it has the nucleotide sequence of Py2CAPy5 (P = pyrimidine base, C = cytosine, A = adenine).
- ***TATA Box*** (*Hogness Box*): About 80% genes have a TATA box centred at about −25 bp whose sequence is reminiscent of the Pribnow box in the prokaryote promoter. It has an 8 bp consensus sequence that consists entirely of adenine-thymine base pairs, although very rarely a guanine-cytosine pair may be present.
- ***Upstream elements:*** Further upstream elements, within about 40–200 or so base pairs of the start site are elements or boxes. These are short DNA sequences that are recognized by specific proteins called *transcription factors*. Three common upstream elements are the **CAAT box,** the **GC box,** and an eight base pair **octamer box;** these are mostly located at sites within the −100 to −200 region in different genes. They occur less regularly than TATA box; for example, CAAT and GC consensus sequences occur only in 10–15% of eukaryotic promoters.

Table 24.4 summarizes location and consensus sequence *cis*-acting elements seen within promoters.

Note that boxes and elements are accepted terms, but these short stretches of bases are not in any sense boxes nor do they have anything to do with atomic elements.

Gene transcription requires several regulatory elements, such as promoters, enhancers, silencers and response elements to be present in the region of the gene.

Enhancers and silencers: These are *regulatory DNA sequences*, which respectively increase or decrease the rate of transcription by binding to regulatory proteins. It is the bound proteins that increase or decrease the rate of transcription, thereby acting as specific transcription factors.

Positive control of transcription by enhancer elements is far more common than negative control by silencer sequences.

Some remarkable features of the enhancers are:

- They can be *located far away from the gene*—even thousands of base pairs away.
- Their orientation in the DNA can be reversed without impairing their function.
- Enhancers often contain a number of elements that are recognized by *transcription factors*.

Enhancers may lie upstream or even downstream of the start-point of transcription. For example, in the immunoglobulin gene the enhancers may be present downstream, located within the intron of the gene being actively transcribed.

Response elements: These elements are nucleotide sequences that allow specific stimuli, such as steroid hormones, cyclic AMP, or insulin-like growth factors to *control gene expression*. They are often found within 1 kilobase of the start point, and a single gene may possess, any number of different response elements.

An interesting and somewhat disturbing aspect of eukaryote gene control is its apparent haphazard nature. No one element is essential for transcription and different genes have different combinations of elements. Twenty per cent of genes lack TATA boxes, some lack other elements and some have multiple copies of an element. It is quite different from the comparatively orderly situation in prokaryotes.

Mechanism of Initiation

Most of the promoter elements promote transcription only if they are located on the correct (non-template) strand of the double helix and in the correct 5′ to 3′ orientation. The sequence-specific DNA-binding proteins that recognize and bind to promoter elements are called ***general transcription factors***. They assemble on promoter to form a complex, termed *basal transcription complex*, which enables RNAP to initiate transcription. In addition to these general transcription factors, the eukaryotic cells contain other regulatory proteins, which act as ***specific transcription factors***, regulating transcription of some, but not all genes.

The basal initiation complex: In prokaryotes, the sigma subunit of *RNAP* recognizes the correct binding site on a promoter and binds directly to the DNA. In eukaryotes, there is no functional equivalent of the sigma subunit; instead a whole retinue of proteins, called **general transcription factors** (GTFs), assemble into a complex on the promoter. The *RNA polymerase* attaches to this complex, termed the basal initiation complex, and this is needed for the transcription of all type-II genes.

Figure 24.4 shows formation of the complex. First a general transcription factor, universally present in eukaryote cells, called *TATA-binding protein* (TBP), attaches to TATA box (Fig. 24.4). Eight or more other proteins called TAFs (TBP associated factors) then get associated with this, forming a complex known as TFII-D. This is followed by attachment of other general transcription factors to DNA in the region of the promoter; to from the basal initiation complex. The GTFs serve three important roles: (1) Facilitates attachment of *RNAP II* to the promoter at the correct nucleotide for initiation, (2) destabilizes (*helicase* activity) the DNA at the promoter and (3) initiates transcription.

A description of functions of all GTFs is beyond the scope of this book; however, some have been outlined in Table 24.5.

Histones inhibit initiation of transcription: When tightly bound by histones, the DNA of native chromatin is unavailable for transcription. A histone may be bound over the promoter or enhancer of a gene, thereby inhibiting initiation of transcription. During transcription, the regular packaging of DNA into nucleosomes is altered so that the decondensed chromosome is accessible to transcriptional proteins. Apparently, histones have to be displaced from promoters and enhancers by competitive binding of certain protein factors, which thereby disassemble chromosomes. A distinct class of protein factors has been identified, for example, **RSF (remodelling and spacing factor)**, which facilitates transcription initiation on chromatin templates in vitro. Another one is **FACT (facilitates chromatin transcription)** that promotes elongation of RNA chains through nucleosomes. Together, RSF and FACT disassemble nucleosomes and permit transcription initiation to occur.

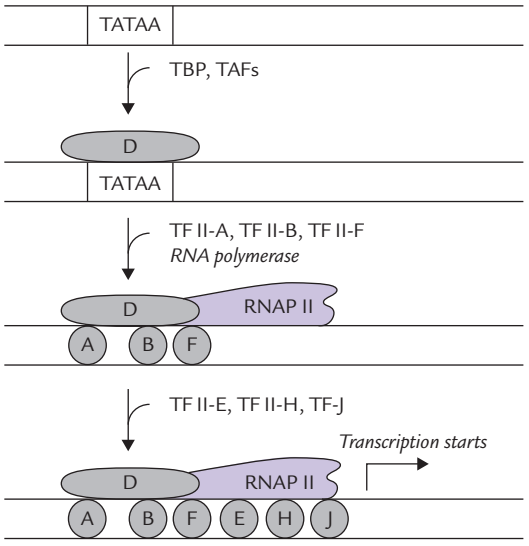


Fig. 24.4. Transcription factors assemble on promoter to form a complex which enables *RNAP* to initiate transcription. Placement of various factors in this diagram is arbitrary; their exact position is not shown (TBP = TATA-binding protein, TAFs = TBP associated factors; TFII-D = Transcription factor D for *polymerase II* [likewise TFII-B, TFII-F, TFII-E, TFII-H and TFII-J refer to the respective transcription factors; D, A, B...J are abbreviations for these transcription factors]).

GTFs that control the initiation of transcription at the TATA box and the protein factors that disassemble nucleosomes, together with specific DNA sequences are important for initiation of transcription.

The given description makes it clear that the initiation stage is more complex in eukaryotes and is better controlled as per cellular requirements. The next two stages—elongation and termination—are remarkably similar to those in prokaryotes.

Elongation

RNA chain elongation proceeds in the same way as in prokaryotes. As the RNAP moves along the DNA template, sequential addition of nucleotide units occurs, directed by the base pairing rule.

Termination

Signals of termination are not fully characterized in eukaryotes. In fact, existence of these signals is doubted in eukaryotes.

Table 24.5. The General transcription factors

Factor	Subunits	Function
TFII-D	13	Recognizes TATA box; recruits TFII-B (composed of factors TBP and TAFs)
TFII-A	3	Stabilizes TFII-D binding
TFII-B	1	Orients RNAPII to start site
TFII-E	2	Recruits TFII-H (*helicase*)
TFII-F	2	Destabilizes non-specific RNAPII-DNA interactions
TFII-H	9	Promotes promoter melting by *helicase* activity

* RNAPII contains nine subunits.

C. Transcription Factors

General characteristics: Transcription of a human gene is initiated and regulated by a number of transcription factors. *These factors must bind to double-stranded DNA for transcription to occur.* The site-specific binding results from the ability of a relatively small area of the protein to come into close contact with the double helix of the DNA. The regions of these proteins that contact the DNA are called *DNA-binding regions* or *motifs*, and are highly conserved between species. These proteins are active in a dimeric state, either as homodimers of two identical subunits or as heterodimers of slightly different subunits. The response elements of these proteins show a **dyad symmetry** (dyad means two units treated as one, or a group of two) that matches the symmetry of the dimers.

The binding of proteins with DNA involves weak *hydrogen bond* formation between amino acid side chains of the protein and the DNA bases, though additional stabilizing bonding to the sugar-phosphate-sugar backbone may occur. Individually a hydrogen bond is weak, but the average DNA-binding protein may have 20 or more sites of contact, which considerably increases the strength and specificity of the contact.

Structural elements: The following structural elements have been recognized in the DNA-binding proteins:

1. *DNA-binding domain*: It is the protein recognition motif that makes contact with the DNA bases, predominantly in the major groove of the double helix, where the edges of the bases are exposed. Contact is made by an α-helix, 20–40 amino acid long, having a high content of basic amino acid residues. Positive charges on these amino acids permit interaction with the negative DNA bases. The DNA-binding domain is highly conserved among different species.
2. *Dimerization domain*: It allows the protein to assume the active dimeric state. The dimerization is effected by an amphipathic α-helix that forms a two-stranded coil.
3. *Transcription activating (or inhibiting) domain*: It is required for increasing (or decreasing) the strength of transcription.

Gene expression is regulated by transcription factors, which bind their response elements (on DNA) in a dimeric form to stimulate (or sometimes inhibit) transcription.

Major classes: Four major classes of DNA-binding domains in transcription factors have been recognized (Fig. 24.5). Salient features together with examples of each are as follows:

1. ***Helix-turn-helix proteins:*** This was the first type to be identified.
   - *Structural features*: It has two α-helices linked by a β-turn (Fig. 24.5). One of these two (called the recognition helix) grips DNA by fitting into the major groove of the double helix. This allows precise alignment of the transcription factor in relation to the DNA sequence recognized.
   - *Examples*: Proteins regulating embryonic development, e.g. homeodomain proteins and lambda repressor.
2. ***Leucine zipper proteins***
   - *Structural features*: The whole molecule is an α-helix, having a (i) DNA-binding basic region, and (ii) a dimerization helix which has hydrophobic leucine residues consistently on one face of the helix. The regularly spaced leucine residues form a hydrophobic phase, which allows two such subunits to interact by hydrophobic forces between the leucine side chains. This brings about dimerization of the two subunits. The term "zipper" is a misnomer resulting from the initial belief that the leucine residues were interdigitated.
   - *Examples*: Growth regulators, proto-oncogenes, proto-oncogene products, e.g. c-Myc, c-Fos, c-Jun.
3. ***Zinc-finger proteins***
   - *Structural features*: In this motif, there are loops or fingers of amino acids with a zinc ion at their core. Each finger consists of approximately 12 amino acids. The zinc atom attached to four amino acid residues (either 4 cysteines, or two cysteines and two histidines) stabilizes the finger-like structure.
   - *Examples*: Intracellular receptors to which steroid hormones bind have variants of the zinc-finger theme.
4. ***Helix-loop-helix proteins***
   - *Structural features*: It contains two DNA binding α-helices, linked by a non-helical loop.

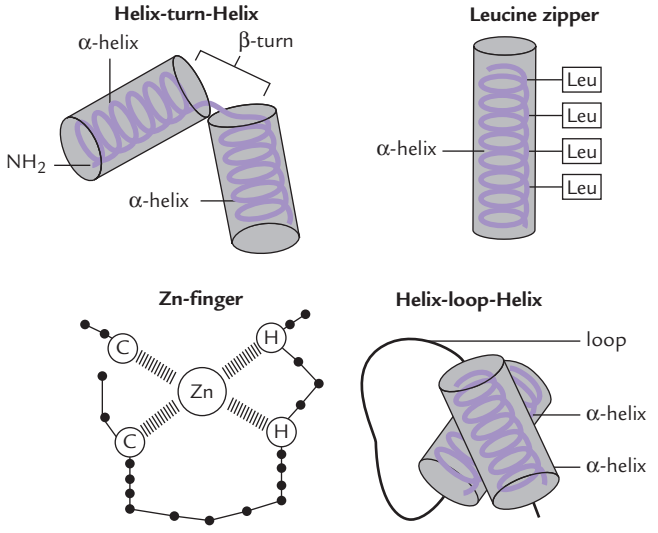


Fig. 24.5. The four main classes of DNA-binding domains in transcription factors.

- *Examples*: Muscle cell specific transporter factor, e.g. MyoD. It is believed that helix-loop-helix motifs mediate negative influences on gene expression.

Functions: Specific interaction of the DNA-binding proteins with response elements is required not only during development, but also within tissues of the mature organism for a number of important functions. These include:

1. *Differential expression of genes in tissues*: Though DNA compositions from all cells of the same organism are same, some genes are optimally expressed in certain tissues and remain silent in others. Genes for transthyretin (a plasma protein), for example, are expressed in liver only. About 10 upstream binding sites have been recognized for at least five different DNA-binding proteins. These proteins are present in much higher concentration in hepatocytes than in any other cell type, and so transthyretin synthesis occurs only in liver.
2. *Growth and development:* Cell growth, differentiation of stem cells into fully differentiated cells, and embryogenesis are controlled by transcriptional regulator proteins. Sequential appearance or disappearance of these proteins in course of growth and differentiation effects regulation of transcription.
3. *Activation of hormonal response elements*: The receptors for steroid hormones, thyroid hormones, retinoic acid and calcitriol are transcription factors (zinc-finger proteins). Formation of hormone-receptor complex activates the latter, allowing it to bind response elements and stimulate transcription.
4. *Expression of viral genes:* Several viral genes can be expressed only after the cellular transcription factors bind to the promoters or enhancers of the viral genes.

Regulation of activity: Transcription factors are regulated at various levels by following means:

1. *Covalent modification*: Some transcription factors are phosphoproteins, and are regulated by phosphodephosphorylation, effected respectively by *protein kinases* and *phosphatases.* Second messengers of the water-soluble hormones regulate transcription by inducing phosphorylation of transcription factors.
2. *Allosteric effects*: Some small molecules have been observed to bind transcription factors reversibly, and act as positive or negative modulators of their activities.
3. Several transcription factors are regulated at the level of their own synthesis.

D. Post-transcriptional Processing

All three classes of RNA are synthesized as larger primary transcripts, also known as **heterogeneous nuclear** (*heteronuclear, hn*) **RNAs**. hnRNAs are usually much larger than the RNAs found in the cytoplasm and require processing to a small size before they can carry out their functions.

The extent and type of processing are different for each type of RNA. Processing of tRNA and rRNA is similar to that in prokaryotes, as described in Chapter 22. However, *eukaryotic mRNA undergoes extensive modifications, in contrast to prokaryotes where it is rarely subject to any post-transcriptional processing.*

Processing of mRNA

Three processing steps occur almost uniformly with eukaryotic mRNA, as summarized in Figure 24.6.

The 5′ end of the mRNA is capped: Shortly after initiation of mRNA synthesis, a guanosine residue is attached at the 5′ terminus in an unusual 5′-5′ triphosphate linkage. The N-7 of the terminal guanine is then methylated by S-adenosylmethionine. This unit **7MeG-5′PPP-5′G is called a cap** (Fig. 24.7). The cap is thought to protect the 5′mRNA from the action of *5′-exonucleases*, to facilitate its binding to the ribosome and is important for subsequent splicing reactions.

The 3′ end of the mRNA receives a polyadenylate tail: Most eukaryotic mRNAs contain a long (up to 250 residues) poly(A) tail at the 3′ end, which is added to the mRNA before it can leave the nucleus. The polyadenylation signal resides in a highly conserved **AAUAAA** consensus sequence, lying near 3′ end of the primary transcript. This sequence is recognized by a specific *endonuclease* that cleaves the RNA approximately 20 nucleotides downstream. The newly created 3′ terminus, however, serves as primer for the enzymatic addition of up to 250 adenine residues.

The polyA tail is believed to associate with proteins that retard action of *3′-exonucleases.*

The removal of introns and rejoining of exons: The eukaryotic genes are patchwork of exons, which are represented in the mature mRNA and introns, which do not appear in the mature mRNA as they are spliced out of the primary transcript. The removal of introns requires nuclear enzyme complexes, called **spliceosomes**.

The basic structures of rRNAs and tRNAs in eukaryotic and bacterial cells are similar, but the eukaryotic mRNAs have a 5′(m^7Gppp) cap and a 3′[A]n tail, and can encode only a single protein (prokaryotic mRNAs lack such modifications on their 5′ and 3′ ends and are mostly polycistronic).

Mechanism of Action of Spliceosomes

The spliceosome contains five small RNAs (U1, U2, U4, U5 and U6) each between 106 and 185 nucleotide long. These are associated with proteins to form small nuclear ribonucleoprotein particles (mRNPs or "snurps"). The

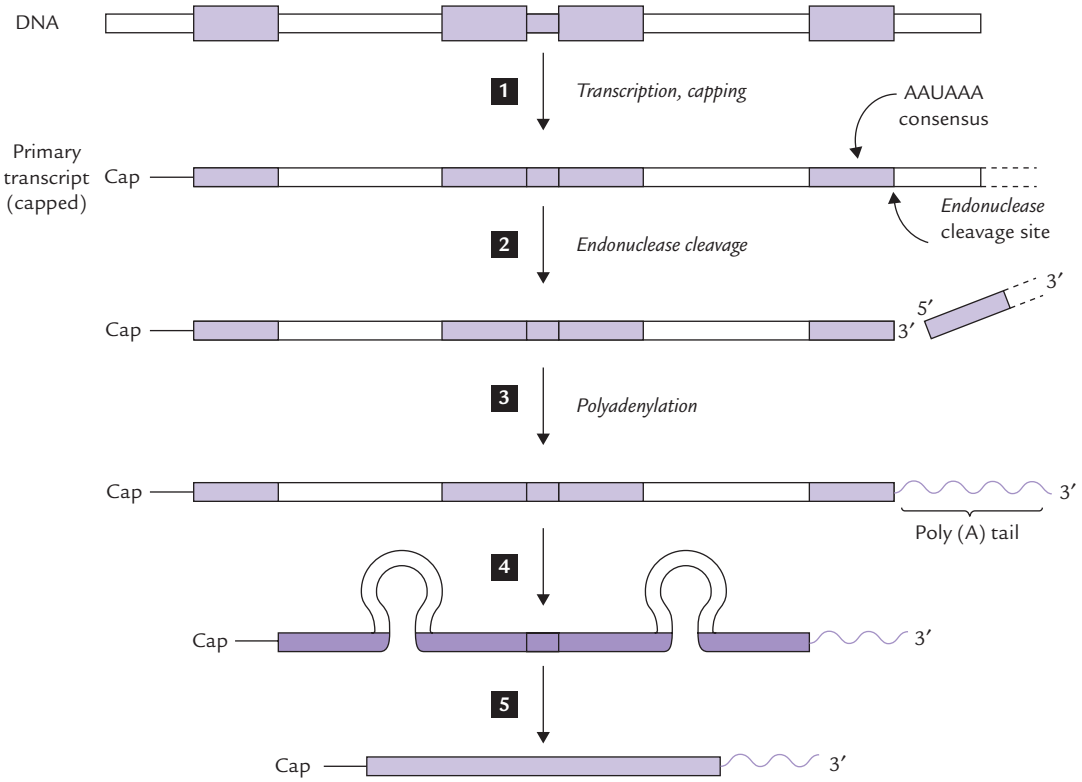


Fig. 24.6. Post-transcriptional processing of mRNA in eukaryotes. All processing steps occur in the nucleus. The mature mRNA is transported into the cytoplasm, where it is translated by the ribosomes. Steps **4** and **5** take place on the spliceosomes. ▬▬ = exon.

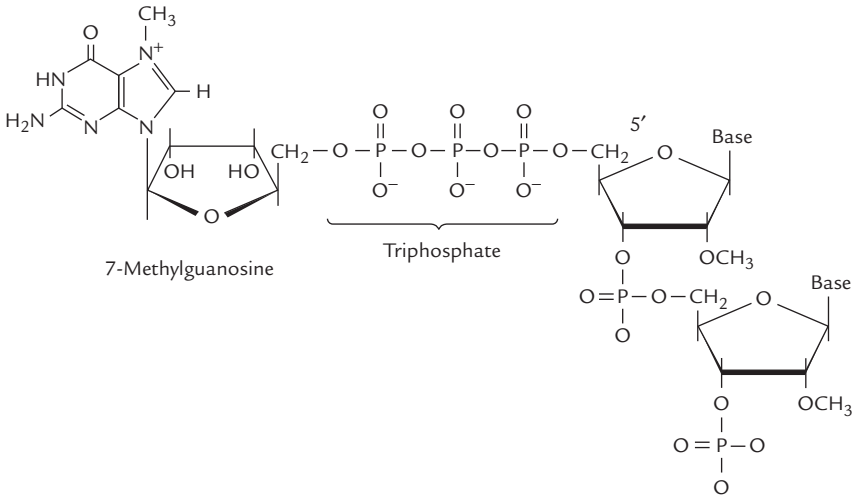


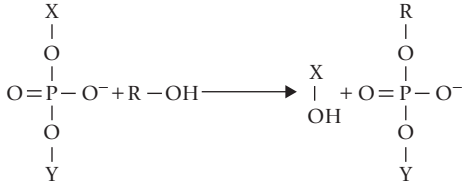
Fig. 24.7. The cap structure.

individual small RNAs apparently recognize specific conserved nucleotide sequences in the mRNA precursor by RNA-RNA base pairing. As shown in Figure 24.8, these conserved sequences are present at:

- The 5′ exon-intron boundary (5′ splice site).
- The 3′ intron-exon boundary (3′ splice site).
- A branch site within the intron, approximately 30 bases from the 3′ end.

Roles of snRNAs are summarized in Table 24.6.

The removal of an intron and rejoining of two exons is based on **two transesterification reactions.** In this, a phosphodiester bond is transferred to a different –OH group, as diagrammed below.



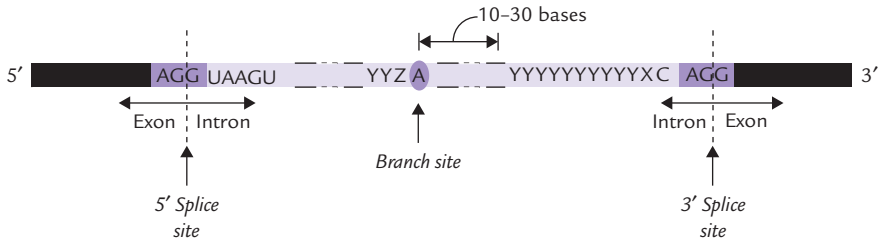


Fig. 24.8. The conserved nucleotide sequences in the mRNA.

Table 24.6. The small nuclear RNAs

SnRNA	Size (no of nucleotides)	Function
U1	165	Binds 5′ splice site
U2	185	Binds the branch site within an intron
U4	145	Helps assemble the spliceosome*
U5	116	Binds 3′ splice site
U6	106	Helps assemble the spliceosome*

* Spliceosome = several kinds of snRNPs + an mRNA precursor.

If X-Y were an RNA chain, it would be broken. Note that energy is not required in the process and there is no hydrolysis.

The removal and rejoining can be considered to occur in the following steps (Fig. 24.9):

1. *Positioning:* The transcript is positioned by base pairing with the U1- and U2-snRNA, which respectively bind to the 5′ splice site and the branch site within the intron.
2. *The first transesterification:* The 2′-OH of an adenine residue at the branch site within the intron plays a crucial role in the first transesterification. This 2′-OH group attacks the phosphate bond of a guanosine residue at the 5′ splice site, forming a lariat structure (named so because of its resemblance to cowboy's lariat). This breaks the chain at the 3′ end of the exon 1.
3. *The second transesterification:* Breaking of chain at 3′ end of exon 1 produces a free 3′-OH which attacks and cleaves the 5′ end of exon 2, thus joining the two exons. The intron is released in the lariat form.

It is evident that RNA itself is able to catalyze cutting and splicing reactions, hence it is referred to as *self-splicing catalytic RNA* (Box 24.3).

Removal of introns takes place in the nucleus and requires ribonucleoprotein particles called spliceosomes.

Note: Mutations at a conserved site may interfere with the correct removal of introns from the precursor mRNA. Abnormal splicing may result in diseases, for example β-thalassaemia, where production of β-haemoglobin chains is greatly reduced.

Systemic lupus erythematosus, an autoimmune disorder, results due to production of antibodies against the snRNPs. It is a serious condition, often fatal.

The most remarkable feature of eukaryotic gene expression is the extensive post-transcriptional processing of mRNA by nuclear enzyme systems. The processing may involve specific exo-and endonucleolytic cleavages to cut out unwanted sequences from the primary transcript, or removal and rejoining of segments of the transcript. A number of other modifications also occur, which include 3′- and 5′-additions, base and sugar modifications, changes in tertiary conformations, etc.

Processing of rRNAs

The genes for 18S, 5.8S and 28S rRNAs are clustered in a unit that is tandemly repeated many times. This cluster of three genes is transcribed by RNAP in the nucleolus to yield a 45S RNA. This large precursor (13 kb) is enzymatically modified and cleaved to yield the mature 18S, 5.8S and 28S rRNAs (Fig. 22.6). The modification occurs through methylation of over 100 of its 14,000 nucleotides, mostly on the 2′-hydroxyl groups of their ribose units. The methylated 45S RNA then undergoes a series of enzymatic cleavages and trimming brought about by *nucleases.* These changes ultimately yield the 18S, 5.8S and 28S rRNAs.

Processing of tRNA

Transcription by RNAP III produces tRNA. A distinctive feature of its transcriptional activity is that it recognizes promoters, located within the coding region of the gene. Initial modifications in the primary transcript include cleavages at the 5′ and 3′ ends, and removal of introns. Modifications of some nucleotide units occur to produce the unusual bases of the tRNA such as pseudouridine (ψ), ribothymidine (T) and dihydrouridine (D). Addition of the sequence CCA to the 3′ end of tRNA also occurs. The reaction is catalyzed by the enzyme *nucleotidyl transferase;* CTP and ATP serve as substrates for the enzyme.

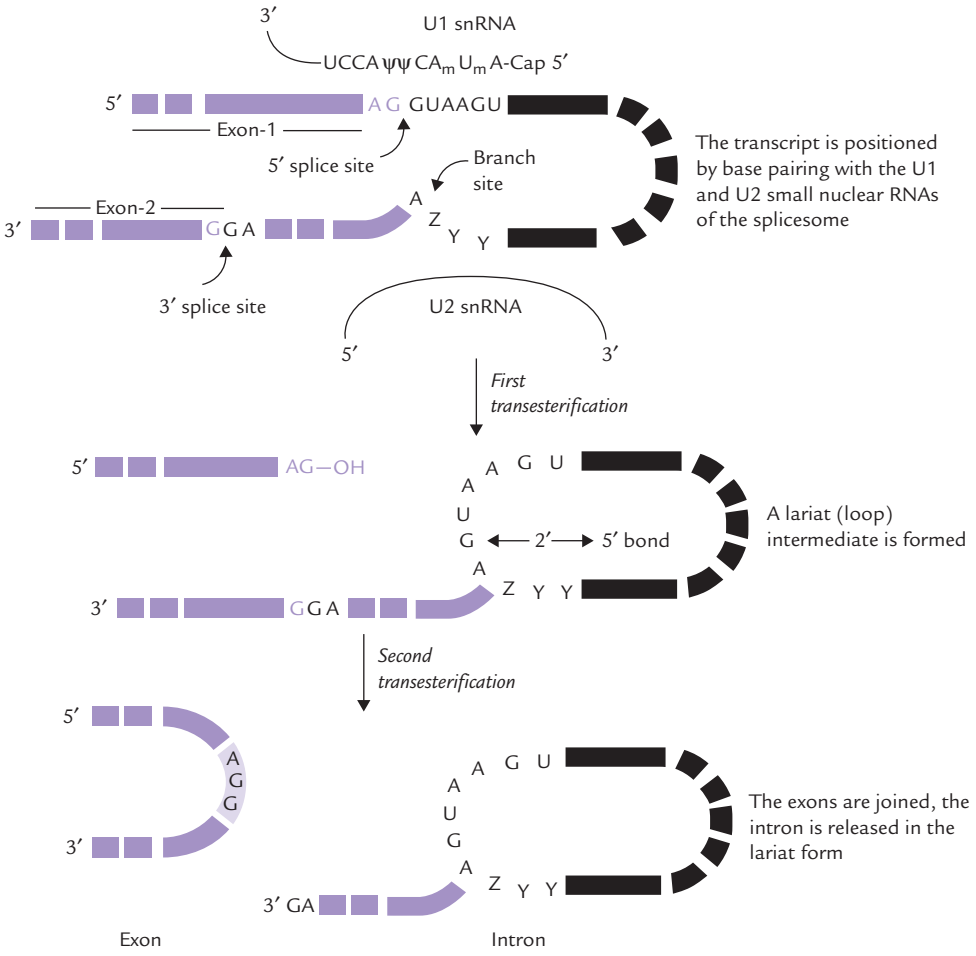


Fig. 24.9. Splicing of introns from the transcripts of protein-coding genes in eukaryotes. The lariat intermediate has a 2′ → 5′ (A→G) phosphodiester bond (Z = purine, Y = pyrimidine, ψ = pseudouridine, A_m and U_m = 2-O-methylated adenosine and uridine).

BOX 24.3

Ribozymes

The enzyme-like action of snRNA in removal of introns challenges the earlier notion that ascribes catalytic activity only to proteins. It shows that some RNAs possess catalytic activity similar to those of proteinaceous enzymes. Such catalytic RNAs called ribozymes were discovered by Thomas Cech and Sidney Altman, for which they were awarded Nobel Prize in 1989. Other examples of ribozymes are: *peptidyl transferase, RNase* P (which generates ends of tRNA) and spliceosomes.

Ribozymes possess a substrate specificity, which is determined via nucleotide base pairing between complementary sequences contained within the enzyme and the RNA substrate. Just like protein enzymes, the ribozyme cleaves its RNA substrate at a specific site and then releases it, without itself being consumed in the reaction.

Therapeutic potentials: The ribozymes are being investigated for the treatment of diseases that are caused by the inappropriate expression of a mutated RNA. In these cases, development of a ribozyme specific for such RNA could result in selective degradation of the substrate, eliminating it from the cell and inhibiting the disease process. This type of treatment is conceptually attractive, but would require several more years of research to be practically available.

V. Translation

Translation in eukaryotes is remarkably similar to that in prokaryotes. However, there are four notable differences:

1. The first amino acid that is incorporated in the eukaryotic polypeptide is *not N-formylmethionine,* but methionine. A special initiator tRNA is required for initiation, which is distinct from that used in elongation.

2. *No Shine–Dalgarno sequence* (which defines start site in prokaryotes) has been detected in the eukaryotic mRNA. Instead, a group of protein factors attach to the cap (the methylated guanine at 5′ end), which enable specific initiation of transcription at the AUG codon nearest to the 5′ end.
3. *Eukaryotic ribosomes are larger* (80S with 60S and 40S subunits) and have extra rRNAs and proteins.
4. Instead of binding with the rRNA, the eukaryotic mRNA binds with proteins in the small ribosomal subunit.

Knowledge of the last difference (eukaryotic mRNAs do not bind rRNA) has important implications in genetic engineering. Because of their inability to bind rRNA, the eukaryotic mRNAs cannot be translated by bacterial ribosomes. The only way to make it possible is to artificially fuse the coding sequence of the eukaryotic mRNA with such a sequence, that is capable of binding with the bacterial ribosome (the sequence is fused in the 5′-untranslated region of the eukaryotic mRNA).

Though the process of protein synthesis in prokaryotes and eukaryotes is similar, it is far more elaborate and complicated in the latter and needs participation of a number of protein factors (Fig. 24.10).

A. Process of Translation

Initiation

Initiation of protein synthesis takes place when a ribosome (both large and small subunits) has assembled on

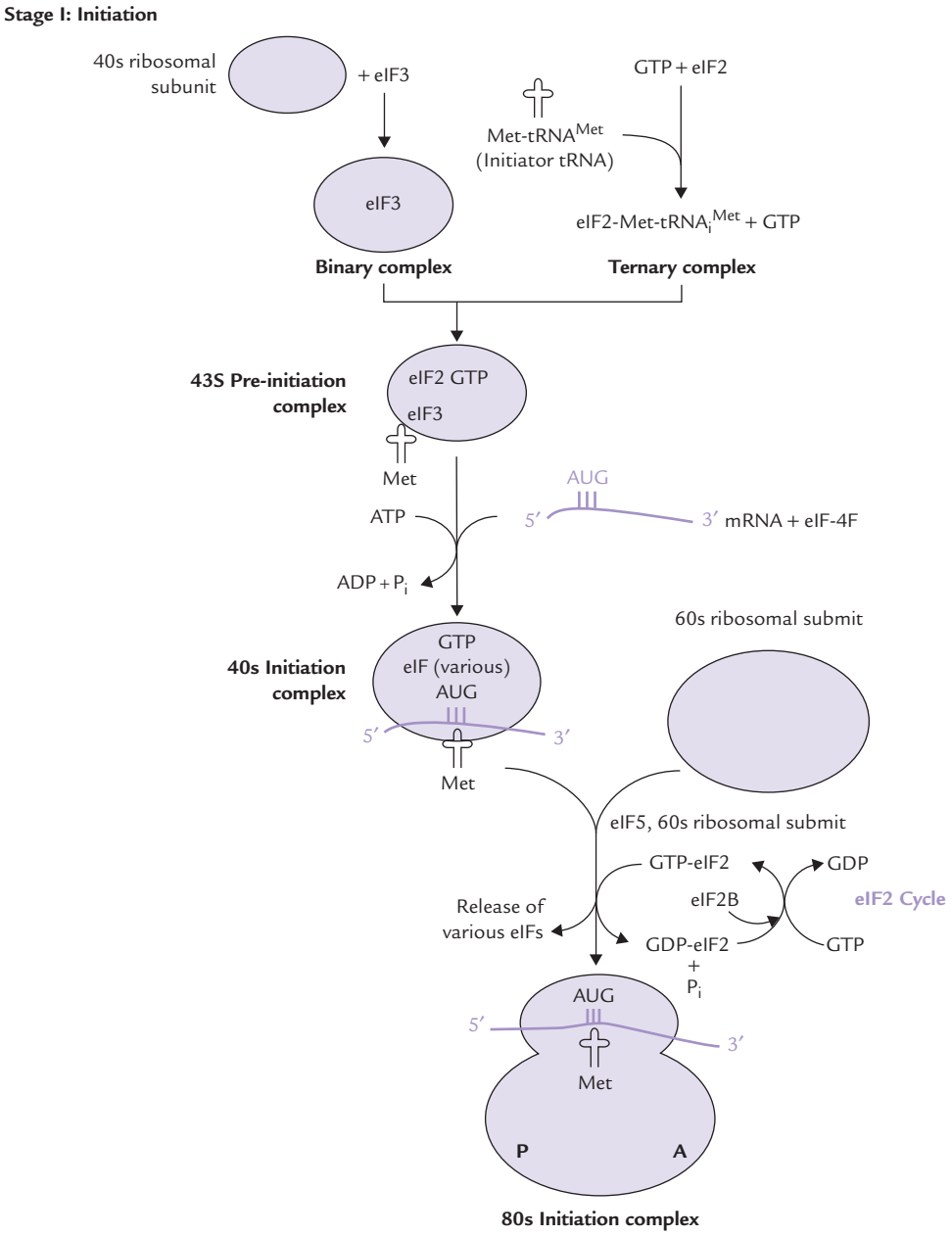


Fig. 24.10. The eukaryotic protein synthesis. *(Continued on page 517)*

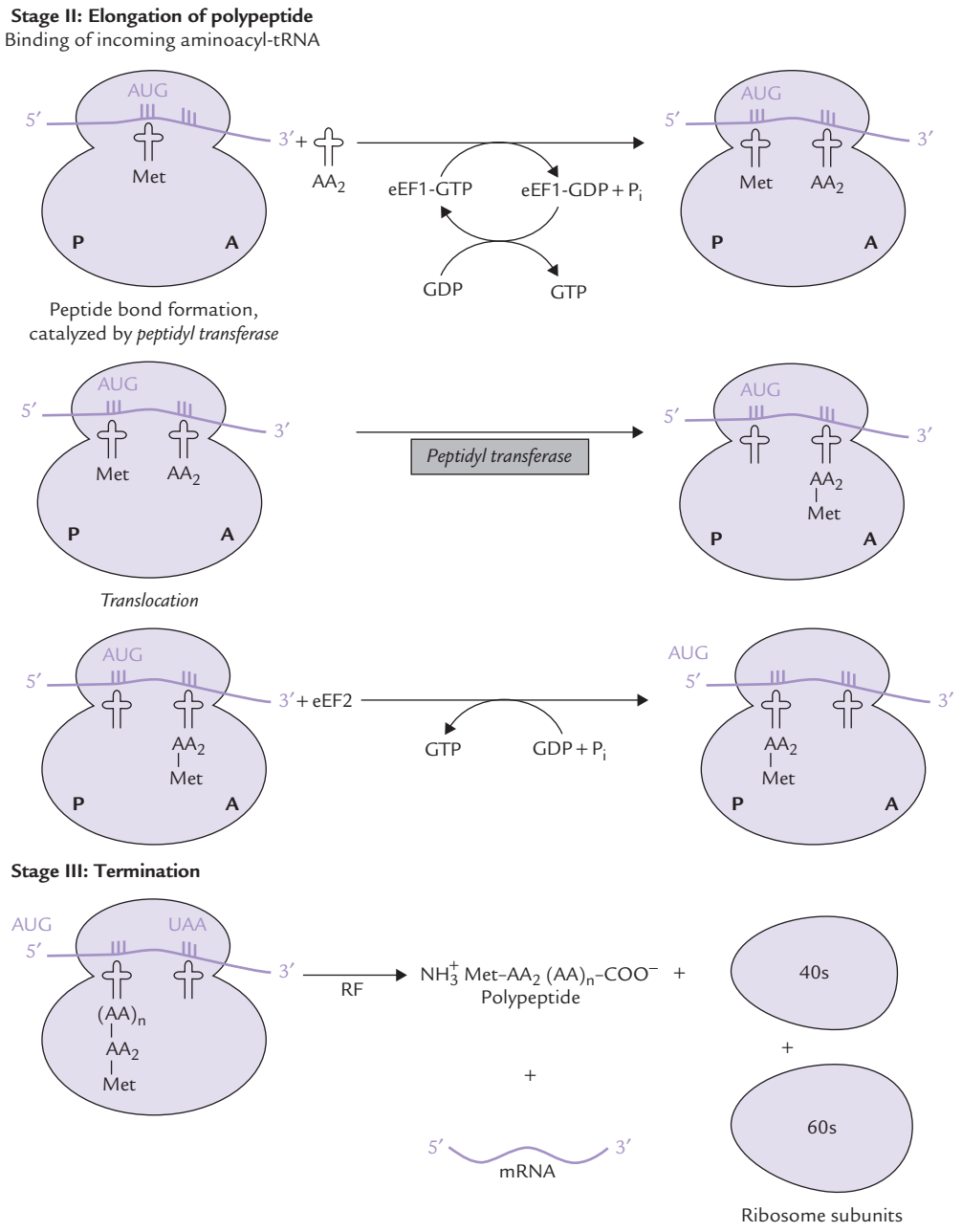


Fig. 24.10 (*Continued*). The eukaryotic protein synthesis.

the mRNA and the P site is occupied by the initiator codon. This complex is formed by the action of proteins known as **initiation factors**. In eukaryotic cells there are at least 12 different initiation factors, which help promote the association of the small ribosomal subunit with the mRNA and a charged initiator tRNA (Table 24.7).

Specific initiation of translation requires four steps:

- Dissociation of ribosome into its 40s and 60s subunits.
- Formation of a 43s pre-initiation complex, consisting of the initiator tRNA, GTP, eIF-2 and the 40s subunit.
- The mRNA is bound to the pre-initiation complex, forming the 40s initiation complex.
- Joining of the 60s ribosomal subunit to form the 80s initiation complex.

Details of these steps are as follows:

Step 1: Dissociation of Ribosome

The initiation factors eIF1 and eIF3 bind to the 40S ribosomal subunit, which favours its dissociation from the 60S subunit.

Step 2: Formation of 43s Pre-initiation Complex

Binding of GTP to eIF2 occurs. It gets attached to the activated initiator tRNA charged with methionine (Met-tRNA^{Met}), resulting in formation of a ternary complex

Table 24.7. Eukaryotic initiation factors

Name	No. of subunits	Functions
eIF1	1	mRNA binding to 40s, prevents association of 40s and 60s ribosomal subunits
eIF2	3	met-tRNA binding to 40s
eIF3	9	Prevents reassociation of ribosomal subunits
eIF4A	1	Stimulates, *helicase* binds simultaneously with eIF4F
eIF4G	1	mRNA binding. Acts as a scaffold for assembly of eIF4E and eIF4A in the eIF4F complex
eIF4F	4	mRNA binding
eIF4E	1	Recognition of mRNA 5′ cap
eIF5	1	Association of 40s and 60s subunits; release of eIF2 and eIF3, *ribosome-dependent GTPase*

eIF = Eukaryotic initiation factors.

(Fig. 24.10). Simultaneously, the 40s ribosomal, subunit is complexed with eIF3 (and probably eIF1 also) to form binary complexes.

The binary and the ternary complexes then join together, forming the 43s pre-initiation complex. This complex is stabilized by earlier association of eIF1 and eIF3 to the 40s subunit.

Step 3: Formation of 40s initiation complex

Binding of mRNA to the pre-initiation complex occurs next. This is effected by the eIF4F, which is a complex of three proteins: eIF4E, 4A and 4G.

- eIF4E physically recognizes and binds to the 5′-cap structure of the mRNA.
- eIf4A binds and hydrolyzes ATP and exhibits RNA *helicase* activity. (Unwinding of the mRNA secondary structure is necessary to allow access to the ribosomal subunits.)
- eIF4G aids in delivering the mRNA to the 43s pre-initiation complex. It also ensures proper alignment of the mRNA on the 43s complex, resulting in formation of the 40s initiation complex.

The initiation complex so formed is at some distance away from the initiation codon, AUG. By an ATP-driven mechanism, it moves along the mRNA till it encounters the first AUG triplet. This scanning mechanism has two interesting repercussions:

1. There can be only one initiation site per mRNA molecule, so eukaryotic mRNAs are *monocistronic* (code for single polypeptide) as contrasted to prokaryotic mRNAs, which are often polycistronic with Shine–Dalgarno sequence provided for the initiation of translation of each cistron.
2. Efficiency of translation is decreased in improperly capped mRNAs.

Step 4: Formation of 80s Initiation Complex

The 40s initiation complex, having migrated to AUG, now joins 60s subunit to form 80s complex. The joining is facilitated by eIF5. The energy needed to stimulate formation of the 80s complex is provided by hydrolysis of GTP bound to eIF2. The GDP bound form of eIF2 then binds to eIF2B, which stimulates the exchange of GTP for GDP on eIF2. This is termed as the eIF2 cycle (Fig. 24.10). It ensures availability of the GTP bound to eIF2 (GTP-eIF2) for additional rounds of translation to occur.

With formation of the 80s complex, all initiation factors are released. This completes the initiation phase and the protein synthesis is now ready to begin. At this stage, the initiator tRNA (charged with methionine) is bound to the mRNA within the **P site** (peptide site) of ribosome. The other site within the ribosome to which incoming charged tRNAs bind is termed the **A site** (for amino acid site).

It is noteworthy that assembly of initiation complex was driven by hydrolysis of GTP and the movement of the complex down the mRNA was driven by hydrolysis of ATP.

Elongation

The elongation process is similar to that of prokaryotes, involving sequential addition of amino acids to the carboxyl end of the growing polypeptide chain by formation of peptide bonds (Fig. 24.10). A notable difference is that the elongation factors involved in the process are different. These are termed eEF1 and eEF2, the former consists of two subunits, which function in a manner analogous to prokaryotic EF-Tu and EF-Ts, bringing the charged tRNA molecule to ribosome. The eEF2 functions like the prokaryotic EF-G, moving the ribosome one codon down the mRNA (i.e. translocation). Though they are functional equivalents of each other, the prokaryotic and the eukaryotic elongation factors are not interchangeable. eEF2 is a specific target for inactivation by diphtheria toxin.

Once the correct charged tRNA molecule has been delivered to the A site of the ribosome, *peptidyl transferase* catalyzes the formation of a peptide bond between the amino acids in the A site. **Translocation** occurs next because the A site needs to be freed in order to accept the next aminoacyl-tRNA. As described in Chapter 22, during translocation the ribosome is moved along the mRNA such that the next codon of mRNA resides under the A site. Following translocation, the eEF2 is released and the whole process is begun again for addition of next amino acid. (For additional information on *peptidyl transferase*

BOX 24.4

***Peptidyl Transferase* is Not a Typical Enzyme**

The enzyme activity of *peptidyl transferase* is located in the *ribosome,* but none of the ribosomal proteins has the capacity to catalyze the elongation reaction (i.e. formation of peptide bond from an ester bond).

Even when all the proteins of the large ribosomal subunit (that contains the *peptidyl transferase* activity) are removed, the rRNA appears to remain capable of catalyzing the elongation reaction. These observations show that the enzymatic activity of *peptidyl transferase* is contained in the rRNA rather than in the proteins that associate with the ribosome. The rRNA behaves therefore, as an RNA enzyme or ribozyme.

refer to Box 24.4). The elongation process is remarkably similar in prokaryotic cells, but the factors are different, which explains the utility of antibiotics that specifically target the prokaryotic factors (Chapter 22).

Termination

Termination of protein synthesis in both prokaryotes and eukaryotes is accomplished when the A site of the ribosome reaches one of the stop codons of the mRNA. A eukaryotic **releasing factor (RF)** recognizes these codons and causes the protein that is attached to the last tRNA molecule in the P site to be released. This process resembles that in prokaryotes, but there are three RFs in prokaryotes (Chapter 22).

After termination, the ribosomal subunits, tRNA and mRNA dissociate from each other. The eIF3 associates with the small ribosomal subunit, and this prevents its reassociation with the 60S subunit in the absence of mRNA. Finally, eIF2 binds to the small ribosomal subunit, setting the stage for translation of another mRNA.

Post-translational modification and antibiotic inhibitors of translation have been described in Chapter 22.

Diphtheria Toxin Inactivates eEF2

Diphtheria is a bacterial infection of the upper respiratory tract, caused by a toxic protein secreted by *Corynebacterium diphtheriae*. The toxin causes death (necrosis) of the mucosal cells, leading to severe pathologic changes and airway obstruction. Initially, it interacts with a receptor on the surface of a sensitive cell, and is proteolytically cleaved to yield two fragments, an A fragment and a B fragment. The B fragment facilitates the penetration of the A fragment through the cell membrane, where it inactivates elongation eEF2. It does so by catalyzing the ADP ribosylation (transfer of an ADP-ribose moiety from NAD^+) of a single unusual amino residue present in eEF2 (diphthamide), which is formed by post-translational modification of histidine. This ADP ribosylation irreversibly blocks the capacity of eEF2 to carry out the translocation step of protein chain elongation, there by blocking the protein synthesis. A single molecule of the fragment A is sufficient to kill the cell because of its capacity to cause enzymatic inactivation of thousands of eEF2 molecules. Function of diphthamide in eEF2 is not known, but it is clearly critical.

Interestingly, the gene for diphtheria toxin does not naturally reside in the bacterial genome. It is carried as a prophage (viral DNA that integrates into the bacterial chromosome). Only lysogenic strains (the ones harbouring prophage) of bacteria are pathogenic; the non-lysogenic strains are peaceful members of the normal bacterial flora on our skin and mucous membrane.

B. Protein Folding and Prion Diseases

As discussed in Chapter 4, protein folding may start and make considerable headway while translation is still in progress, suggesting that the amino acid sequence of a polypeptide is an important determinant of its final native conformation. However, it is not the only determinant. The constituent amino acids can associate with one another in a number of ways and accordingly the polypeptide chain can fold in various shapes. It is theoretically possible that the polypeptide tries out every conceivable internal association pattern until it arrives at the correct folding association and thereby assume the native 3-dimensional structure—the one having biological activity. However, such process would require millions of years to find the folded form, while in the cell it all has to happen over in a time-scale of a minute or so. Evidently, protein folding is not entirely an automatic unaided process, but requires participation of other factors. Two classes of proteins have been described: one class is of molecular chaperones and the other is of conventional enzymes or proteins.

Molecular Chaperones

These proteins aid production of correct spatial arrangement. They are concerned with kinetics of folding and not with the nature of the final folded form, which is determined by the amino acid sequence (Chapter 4). They can recognize and bind to partially folded proteins (or more appropriately, partially unfolded proteins), and prevent improper associations that may occur; for example,

between hydrophobic patches. Such associations could prevent the formation of correct folding. Binding of chaperones to certain reactions of polypeptides is believed to stabilize the partially folded molecule until a stage is reached when correct associations can occur.

Chaperones ensure proper associations; of course, the word chaperone literally means elderly lady entrusted with responsibility of looking after unmarried girls during social occasions probably to ensure correct associations. Chaperones have enormous requirement for ATP energy.

Heat shock proteins: These are examples of chaperones that are produced in response to heat, and other forms of stress, such as toxins, heavy metals, free radicals, radiations, bacteria, etc. (Chapter 22). Some of the major families of heat shock proteins (HSP) are HSP-10, HSP-60 (called chaperonin), HSP-70 and HSP-90; the numbers represent the molecular weight in kD.

Enzymes/Proteins Involved in Protein Folding

These include *protein disulphide isomerase, proline isomerase,* and immunoglobulin binding protein.

- *Protein disulphide isomerase* (*PDI*): This enzyme shuffles the disulphide bonds in the polypeptide chains to ensure its correct positioning. If an incorrect disulphide bond were to be formed, PDI cleaves it and reforms a correct one, so that the folding can correct itself.
- *Peptidyl proline isomerase* (*PPI*): Wherever a proline residue occurs in a peptide linkage, the configuration can be *cis* or *trans*. PPI plays the role of shuffling proline residues between the configurations so as to permit the whole protein to assume the correct configuration.
- *Immunoglobulin-binding protein*: It is more punishing in its action, being able to recognize the incorrectly folded immunoglobulins and cause their destruction.

Prion Diseases

The prion diseases (also known as spongiform encephalopathies) are a group of transmissible, fatal neurological degenerative diseases that affect humans and animals. These include *Creutzfeldt-Jacob disease* and *Kuru* in humans. The disease in sheep and goats is designated as scrapie and in cows as spongiform encephalopathy (mad cow disease). It is not clear whether the diseases are genetic or infectious.

The diseases are characterized by the accumulation of altered forms of normal proteins. These proteins, commonly termed *"prion proteins"* have certain changes in 3-dimensional structures. The major change is replacement of α-helices by β-sheets. This forms an abnormal isoform, called *prion protein-cellular form* (PrPC) in affected brains. The *scrapie form of the prion protein* (PrPSc) has a high α-helical content and is devoid of β-pleated sheets.

It is believed that the abnormal conformation of the protein makes it resistant to action of proteolytic enzymes. These proteins are highly infectious in nature. Moreover, the progression of infectious prion disease appears to involve an interaction between PrPC and PrPSc, which induces a conformational change of the α-helix rich PrPC to the β-pleated sheet-rich conformer of PrPSc.

C. Translation in Mitochondria

Most of the cellular DNA resides in the nucleus, but a small fraction (less than 1%) resides in mitochondria. The mitochondrial DNA is a circular duplex with a length of less than 20,000 base pairs in mammals (16,569 base pairs in humans), and there can be five to ten copies in each mitochondrion. The mtDNA is genetically active. The protein coding genes are transcribed by mitochondrial *RNA polymerase* and translated by mitochondrial ribosomes that are structurally different from those in cytoplasm.

Though the mitochondrion is a self-replicating organelle with its own DNA and protein synthesizing machinery, it is far from self-contained because its DNA codes for only a small fraction of mitochondrial proteins (the same is true of chloroplasts). The nuclear DNA of the cell codes for the rest of the mitochondrial proteins, which are synthesized in the cytoplasm and transported into the mitochondria. mtDNA codes for a total of 13 polypeptides, most of them being components of electron transport chain, using a code slightly different from the virtually universal code.

It is not known with certainty why mitochondria have their own DNA, though it is believed that mitochondria evolved from engulfed prokaryotic cells that became symbiotic (a similar theory holds for the chloroplasts). Most of the original prokaryotic genes were transferred to the nucleus, leaving the mitochondria with a rudimentary but still functional protein synthesizing system. This **endosymbiont hypothesis** thereby suggests that our genome is partly descended from a primordial eukaryote and partly from a symbiotic prokaryote, and is therefore of hybrid nature (Box 24.5).

VI. Regulation of Eukaryotic Gene Expression

The regulatory mechanisms for controlling expression of eukaryotic genes operate at various levels. Some of these involve certain changes in genes such as gene loss, gene

BOX 24.5

The Endosymbiont Hypothesis

The hybrid nature of mitochondrial DNA, as proposed by endosymbiont hypothesis, is supported by the following facts:

- The mitochondrial translation has prokaryotic characteristics as compared with the eukaryotic features of cytoplasmic translation.
- The sequences of rRNA in mitochondria and bacteria have considerable structural resemblance.
- The mtDNA and the prokaryotic DNA are closely related.

The mitochondrial DNA is for the most part maternally inherited; this is probably the result of the egg cell contributing much more cytoplasm to the zygote than the sperm. Therefore, it has little chance for genetic recombination. Mitochondrial DNA is exposed to oxidative stress and therefore subject to a high rate of mutation (about 10 times more than nuclear DNA). These mutations result in inherited disorders with defects in oxidative phosphorylation, e.g. mitochondrial myopathies.

rearrangement, gene amplification, etc. Others operate at the level of gene transcription, which has overriding importance in control of gene expression, both in eukaryotes and prokaryotes.

Several additional mechanisms are also present in eukaryotes (not in prokaryotes), which operate at the post-transcriptional level at any or all of the steps from RNA to protein. These steps include RNA processing (using different promoters and polyadenylation sites, and alternative splicing), selection of mRNAs to be transported from the nucleus to the cytoplasm, selective stabilization or degradation of specific mRNA molecules in the nucleus or cytoplasm. By control over these mechanisms, gene expression may be altered more than 1000-folds.

A. Changes in Genes

Gene Amplification

By this mechanism, repeated initiation of DNA synthesis occurs. Additional copies of the gene may be produced in this manner. The drug **methotrexate** causes production of hundreds of copies of the gene responsible for the enzyme *dihydrofolate reductase.* In electron microscopy, it is seen as a big replication bubble (Fig. 24.11).

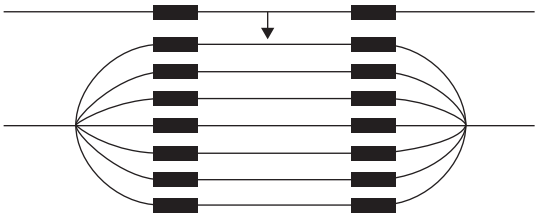


Fig. 24.11. Amplification of gene.

Gene Rearrangement

A segment from the DNA moves from one location to another on the genome. This forms new combination of genes so that several different proteins are encoded. For example, various portions of the antibody-producing genes lie at distant location of the DNA of the germ line cell. By gene rearrangement, DNA segments are brought together and combined to produce the **active gene**.

Production of the active heavy chain gene of immunoglobulin by gene rearrangement is shown in Figure 24.12.

- The heavy chain consists of a variable region (V_H) and a constant region (C_H). The genes coding for these two regions that lie far apart in genome of an immunocyte, are brought close and joined to yield the active heavy chain gene.
- Furthermore, the complete V_H gene (the functional form) is formed by assembly of three different types of genes: the variable gene, V; the diversity gene, D; and the joining gene, J. These genes are scattered throughout the genome of the immunocyte. They can be transposed next to one another and joined to yield the complete V_H gene.

Various components of the V_H originate from different DNA segments that lie far apart in the genome.

The V_H gene is then joined with the constant region gene (C_H) to yield the heavy chain antibody gene (Fig. 24.12).

Gene Loss

Certain genes may be lost from a cell so that the functional protein encoded by them are not produced. The gene loss may be partial or even complete. For example, during differentiation of red blood cells, the nuclei are extruded resulting in complete loss of all genes. Sometimes the gene may not be lost but its expression may stop (Box 24.6).

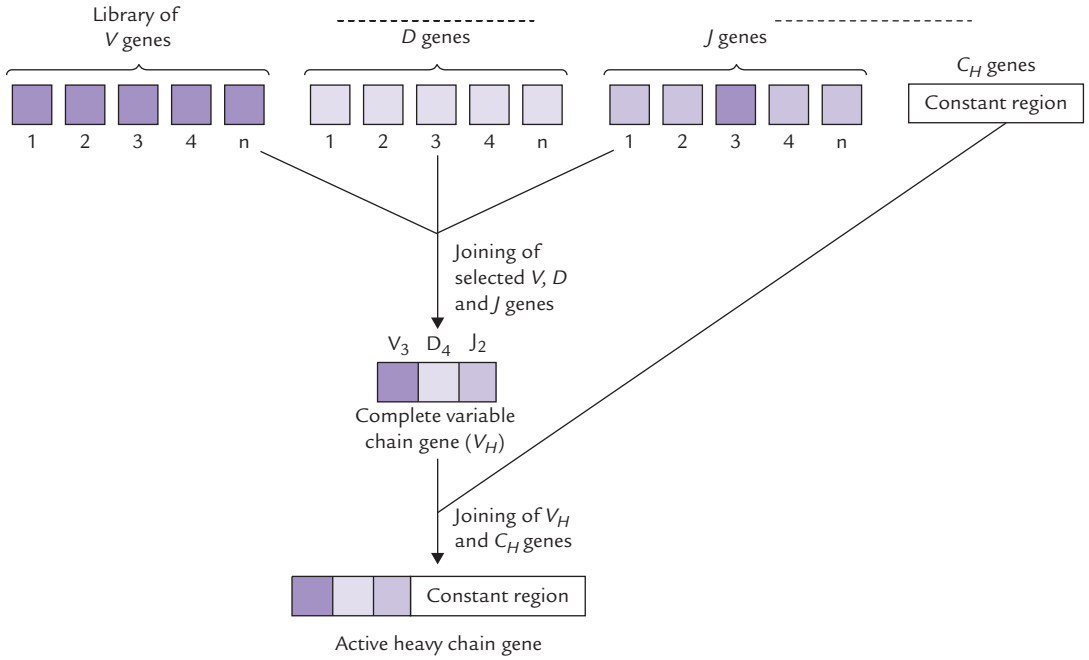


Fig. 24.12. Assembly of an antibody heavy chain gene by gene rearrangement.

BOX 24.6

Gene Switching

It is a phenomenon in which transcription and translation of a given gene stops, and instead expression of a related gene is turned on. Thus, gene expression is shifted from one gene to another. An example is seen during elicitation of immune response. During primary response, IgMs are synthesized, whereas in secondary immune response, synthesis of IgG occurs. Thus, switching of genes from IgM to IgG occurs.

Another example of gene switching is seen in haemoglobin synthesis where genes for different types of polypeptide chains are expressed at different stages of life:

- In embryo: Two α- and two ε-chains are synthesized (embryonic haemoglobin).
- Intrauterine life: Two α- and two γ-chains are synthesized (fetal haemoglobin: HbF).
- After birth: Two α- and two β-chains (HbA1); or two α and two δ chains (HbA2) are synthesized. The gene switching permits replacement of embryonic haemoglobin by fetal haemoglobin by sixth month of intrauterine life, and then replacement of the fetal haemoglobin by adult types of haemoglobins after birth.

Modification of DNA

Modification of certain bases affects transcriptional activity of the gene. When cytosine is methylated at its fifth position to methylcytosine, reduction of the transcriptional activity of the concerned gene results. *More the **methylation**, less is the frequency of transcription of the affected gene.*

Most of the methylcytosine is present in the sequences, 5′-CG-3′ throughout the genome. In fact, more than 70% of the susceptible sites are methylated in this way. In a double-stranded DNA, the complementary cytosine is also methylated (because the methylating system uses the hemimethylated sequence as a preferred substrate), giving rise to a palindromic sequence:

5′–C^{m}G–3′
3′–GC^{m}–5′

So much is control of methylation on the rate of transcription, that the highly methylated areas of DNA are transcriptionally inactive.

Methylation of DNA regulates gene expression.

B. Regulation at the Level of Transcription

Regulation of eukaryotic gene expression occurs principally at the level of transcription. Various genetic elements, e.g. promoters,

enhancers, etc. are involved in regulating the transcriptional activity, as already described in this chapter.

C. Regulation at Post-transcriptional Level

Use of Alternative Promoters and Different Polyadenylation Sites

Some genes have evolved a series of promoters that confer tissue-specific expression. Use of these alternative promoters in different tissues result in transcripts differing in the 5′ portion and the translated polypeptides differing in the N-terminal sequence. The best example of the use of alternative promoters in humans is the gene for dystrophin, the muscle protein that is deficient in Duchenne muscular dystrophy. This gene uses different promoters, each associated with its own first exon, so that each mRNA and its protein has a tissue specific N-terminal sequence. Thus, brain, muscle, and retinal specific proteins differ in N-terminal amino acid structures.

In other genes, alternative polyadenylation sites can be used, and this produces transcripts with different 3′ ends. These complexities suggest that genes be redefined (Box 24.7).

Tissue-specific Splicing

In some primary transcripts, sequences that are treated as introns by some cells are treated as exons by others. This leads to formation of different mRNAs and different proteins from a single gene.

mRNA Editing

RNA editing involves the enzyme-mediated alteration of RNA in the cell nucleus before translation. The process may involve insertion, deletion or substitution of nucleotides in the RNA molecule. The substitution of one nucleotide for another, for example, results in alteration of a codon to yield tissue-specific transcripts. Apolipo-protein B (apoB) illustrates this. The gene for this protein encodes a 14.1 kb mRNA transcript in the liver and a 4536 amino acid protein product, apoB100 (Fig. 24.13). In the small intestine, the editing of nucleotide 6666 of apoB100 mRNA, by changing cytidine to uracil, generates a stop codon in the intestinal mRNA. The translation product of this edited mRNA (7 kbp) is therefore shorter (2152 amino acids) than the unedited mRNA.

Regulation by *Nuclear RNase*

Several *nuclear RNase* extensively degrade the primary transcriptor mRNAs before the latter can emerge into the cytoplasm as mature, translatable RNAs. This may prevent inappropriate expression of genes, and also protects the cell from viruses whose mRNAs are degraded in the nucleus.

The mRNA Stability

The survival time of mature mRNA in the cytoplasm is regulated by relative activities of *nucleases* that degrade

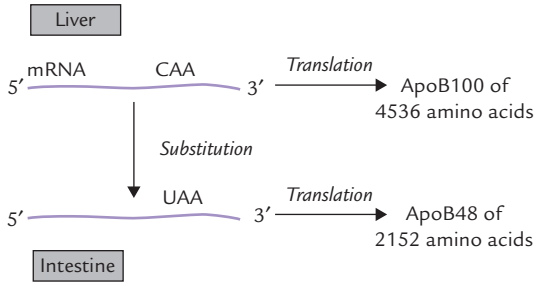


Fig. 24.13. RNA editing of the apoB gene gives rise to tissue specific transcripts.

BOX 24.7

Gene Redefined

The earlier notion that gene is a segment of DNA, with exons and introns which code for a single protein product has been challenged because of the following observations:

1. Two or more mRNA species may arise from a single region of DNA because of differences at the level of transcription.
2. Different protein products may arise from a given segment of transcribed DNA because of differences at the post-transcriptional level.

Therefore, there is now tendency to refer to such DNA segment as a transcriptional unit. A transcriptional unit comprises not only the exons, introns and regulatory elements, but also various molecular events that modify the transcription process or the events modifying the primary RNA transcript (i.e. the post-transcriptional modifications). This is a shift from the earlier definition of a gene (a single-strand of DNA with exons, introns, response elements and promoters) to one of a complex gene that encapsulates not only these structural elements but also the dynamic processes giving rise to the final gene product or products.

the mRNA and by mRNA-binding proteins that prevent *nuclease* attack. *An average eukaryotic mRNA molecule has a lifetime of 3 hours*. In certain highly differentiated cells, it is prolonged so that production of a large amount of a single type of protein by its translation is made possible. For example, cells of the chicken oviduct which makes ovalbumin, contain only a single copy of the ovalbumin gene per haploid set of chromosomes, but the cellular mRNA is long lived.

D. Translational Regulation

This refers to the number of times a finished mRNA molecule is translated. Regulation of gene of ferritin, an iron-binding protein, in iron deficiency and iron excess illustrates this mode of regulation.

In conditions of iron excess (haemochromatosis) there is an increase in the synthesis of ferritin; and in iron deficiency there is a decrease in the synthesis of ferritin and increase in the synthesis of the transferrin receptor protein (Chapter 19). The mRNAs for both ferritin and transferrin receptor mRNA contain a specific sequence known as the **iron-response element** (IRE), to which a specific IRE-binding protein can bind. When iron is deficient the IRE-binding protein binds the ferritin mRNA and prevents translation of ferritin. It also binds the transferrin receptor mRNA and prevents its degradation. *Thus in iron deficiency, ferritin concentrations falls low and transferrin receptor concentration rises*. A reverse series of events occurs in iron excess and translation of ferritin mRNA increases, whereas transferrin receptor mRNA undergoes degradation.

Various regulatory mechanisms act cooperatively to give rise to tissue-specific gene products (Box 24.8).

E. Differential Expression of a Parental Allele of a Gene

There are 23 pairs of chromosomes, each of which has genes that are present on both chromosomes: they are biallelic. One of the alleles is inherited from father and the other from mother, and normally both are identical. Under normal circumstances, either of the two alleles has equal chance for expression without preference being given to either. However, in humans certain genes have been identified, in which only one allele—either maternal or paternal—is preferentially expressed. This occurs despite the fact that both are perfectly normal or identical. As a result of such restriction of the expression of biallelic genes, only 50% of the product is produced. The mechanisms for such a restriction are: **allelic exclusion, X chromosome inactivation** and **genomic imprinting**.

Allelic exclusion implies tissue-specific expression of a single allele product (for example, synthesis of a single immunoglobulin chain in a B cell from one allele only); **X chromosome inactivation** implies switching off an entire X chromosome, thereby inactivating all X-linked genes located on it. **Genomic imprinting** refers to effect of methylation of RNA (refer to Box 24.9 for details).

F. Need for Elaborate System of Gene Regulation in Eukaryotes

Finally, it is worthwhile to consider as to why such an elaborate system of gene regulation is required in eukaryotes and not in prokaryotes. Compared to prokaryotes where most, if not all, genes are expressed at the same time in eukaryotes there is need to express different genes in different tissues and at different times. Though all cells

BOX 24.8

Tissue-specific Expression Involves More than One Processing Patterns

Gene expression is regulated at various levels and each of these is crucial to the proper functioning of the cell and the-organism. Various regulatory mechanisms act cooperatively and give rise to tissue-specific gene products. The calcitonin gene illustrates this. It consists of five exons (and intervening introns) and uses two alternative polyadenylation signalling molecules. It can yield two products upon translation: calcitonin and calcitonin-related gene peptide.

In medullary C cells of thyroid, calcitonin is produced through use of one polyadenylation signalling molecule associated with exon 4 to transcribe a primary transcript comprising exons 1–4. The associated intronic sequences are removed to produce an mRNA molecule that yields calcitonin upon translation. However, in neural tissue, a second polyadenylation signalling molecule next to exon 5 is used. This results in pre-mRNA having the same organization as the gene; there are all five exons and the intervening introns. A different form of mature mRNA is produced from it after all introns as also the exon 4 is spliced out. This leaves an mRNA comprising exons 1–3 and 5, which is then translated into the so-called calcitonin-related gene peptide (GRGP), an important growth factor.

Thus, a single gene can be expressed differently in different tissues of the same organism.

BOX 24.9

Genomic Imprinting

Inheritance of some genes in a silent state from one parent and in a fully active state from the other parent suggests that there must be some mechanism that marks out one allele as maternal and one as paternal. One postulated mechanism is through methylation, which impairs gene expression. The genes are said to be imprinted (by methylation) and this phenomenon is known as genomic imprinting.

Though in Mendelian inheritance, the effect on any allele on the offspring is expected to be the same regardless of the parent from which it came, imprinting (because of methylation of base) changes this. It causes the phenotype to vary depending on the parent who donated the allele. For instance, if a gene is expressed in a developing embryo only if derived from the paternal allele, the most likely explanation for this is that the female allele is methylated in egg, but the paternal allele remains unmethylated. Such is the case for many genes, although the sex of the active allele may be the reverse.

contain identical sets of genes, one cell type (for example muscle) expresses only certain genes and never expresses the liver-specific genes. Moreover, during embryonic development certain genes are required at an early stage of the differentiation process and others later in the development of the specific tissue. Thus, gene control needs are much more complex in differentiated eukaryotes.

Regulation of eukaryotic gene expression occurs principally at the level of transcription, but it does encompass several other processes, including gene amplification, gene rearrangement, processing, transport and stability of mRNA and translation.

Exercises

Essay type questions

1. Describe the transcription and post-transcriptional modifications in eukaryotes.
2. At which points in protein synthesis do fidelity mechanisms operate? Explain why the mechanism of initiation of translation in eukaryotes is not compatible with polycistronic mRNA.
3. Describe the process of protein synthesis in eukaryotes. Explain the participation and, where known, the role of GTP in protein synthesis.
4. Describe the regulation of gene expression in eukaryotes.
5. Explain the role of chaperones and other proteins in folding of polypeptide chains. What diseases may be associated with improper protein folding?

Write short notes on

1. *Cis*-acting elements
2. Gene amplification
3. Diphtheria toxin
4. Spliceosomes
5. Telomers
6. Transcription factors
7. Mitochondrial DNA
8. *Reverse transcriptase*
9. TATA Box
10. Genomic imprinting

CHAPTER

25

RECOMBINANT DNA TECHNOLOGY AND GENETIC ENGINEERING

Development of recombinant DNA technology is a major breakthrough in molecular biology. New combination of unrelated genes can be constructed in laboratory by means of the DNA manipulating techniques, which also make it possible to isolate and determine base sequences of individual genes, to manipulate them in any desired way, and to transfer genes from one species to another. The transferred genes can be placed in a functional role in many different types of cells to produce the protein it codes for in whatever amount is required. The associated technologies can be used in medical diagnosis of genetic diseases. The general biological applications are almost unlimited.

About 33,000–44,000 genes are scattered among billions of base pairs (3.1×10^9) that make up the human genome. To identify a given gene from this vast expanse of base pairs, separate it, and produce multiple identical copies of it, has been a challenging task for the researchers, it resembles not so much as finding a needle in a haystack, as finding a particular piece of hay in a haystack. The molecular biology changed it all, making it possible to isolate and amplify a specific gene. The student is introduced to this fascinating and ever expanding field in this chapter.

After going through this chapter, the student should be able to understand.

- Uses and mechanism of action of various enzymes that are employed to manipulate nucleic acids (restriction enzymes, *terminal transferases, ribonuclease H,* etc.); role of cloning vectors (plasmids, phages and cosmids) in transferring the gene of interest to host cell; and application of techniques of genetic engineering in isolating and amplifying a gene of interest.
- Difference between two types of DNA libraries: genomic and complementary.
- Polymerase chain reaction (PCR based DNA cloning).
- Specific methods used in analysis of DNA.
- Special applications of recombinant DNA technology in medical diagnosis and treatment (gene therapy).
- Current status and future goals of the Human Genome Project.

The term **recombinant** refers to recombining different segments of DNA, and **cloning** refers to production of multiple identical copies of the recombinant DNA, which are therefore of the same origin. The essence of cloning involves separating a specific gene or a segment of DNA from a large chromosome, attaching it to a carrier (**vector**), and introducing it into a suitable cell where it is propagated. In addition to such cell-based cloning, another method is enzyme-based (represented by **polymerase chain reaction**), which is a simpler, quicker and a more robust technique for mass DNA amplification.

It is worthwhile to first examine the properties of some of the enzymes and be familiar with some techniques used in DNA cloning.

I. Techniques and Enzymes Used in Manipulation of DNA

It has been possible to manipulate DNA because of three key factors: the annealing properties of nucleic acids and availability of certain enzymes that act on nucleic acids in a specific manner. In addition, cloning vectors (vehicles) play the vital role of carrying DNA into the suitable host cell.

A. Annealing Properties

Any two segments of a single-stranded DNA or RNA, having continuous complementary sequences of 20 or more bases in common, form complementary base pairs to form **duplex** structure. Duplex may be formed between DNA and DNA, RNA and RNA, or between DNA and RNA to generate a hybrid.

An application of annealing properties is construction of nucleic acid probes which provide useful means for detecting genes and mRNAs.

B. Enzymatic Reactions

The commonly used enzymes in manipulation of DNA are enumerated here.

Restriction Endonucleases

These are bacterial enzymes that cleave specific **palindromic sequences** in double stranded DNA (Fig. 25.1). Werner Arber (Nobel Prize 1978) showed that entry of phages into host bacteria is restricted by such bacterial enzymes. Therefore, they are named as *restriction endonucleases* (RE), or simply restriction enzymes.

More than 800 types of REs are known; and more than 400 of them are available commercially. The restriction enzymes serve as the most "exquisite scalpels" cleaving the double-stranded DNA very selectively at specific sites, called the ***restriction sites***. Each restriction site is a short sequence of 4–6 base pairs, and is palindromic, showing a two-fold rotational symmetry. As shown in Table 25.1, the top strand of each palindromic sequence

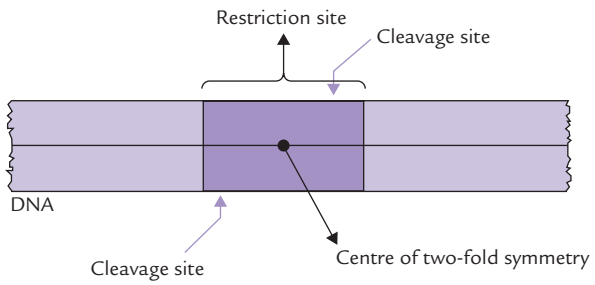


Fig. 25.1. *Restriction endonuclease* cleaves DNA at specific restriction site.

Table 25.1. Specificity of some *restriction endonucleases*

Restriction endonuclease	Cleavage site
1. *Produce flush ends*	
(a) *Hind II*	↓ 5′ – G – T – Py – Pu – A – C – 3′ • 3′ – C – A – Pu – Py – T – G – 5′ ↑
(b) *Hae III*	↓ G – G – C – C • C – C – G – G ↑
(c) *Hpa I*	↓ 5′ – G – T – T – A – A – C – 3′ • 3′ – C – A – A – T – T – G – 5′ ↑
2. *Produce sticky ends*	
(a) *Eco RI*	↓ 5′ – G – A – A – T – T – C – 3′ • 3′ – C – T – T – A – A – G – 5′ ↑
(b) *Taq I*	↓ T – C – G – A • A – G – C – T ↑
(c) *Eco RII*	↓ * 5′ – N – C – C – N – G – G – 3′ • 3′ – N – G – G – N – C – C – 5′ * ↑
(d) *Hind III*	* ↓ 5′ – A – A – G – C – T – T – 3′ • 3′ – T – T – C – G – A – A – 5′ ↑

Arrows (→) indicate cleavage sites, • indicates the axis of two-fold rotational symmetry, * indicates methylation sites in the parent organism.

has the same nucleotide sequence (from 5′ to 3′ end) as the bottom strand, when read in the 5′ → 3′ direction. Therefore, if the structure is rotated 180 degrees, it remains the same; for example, turning the page upside down will not affect the above structures.

Some restriction enzymes can recognize the sites that are relatively small, e.g. four nucleotides such as *Hae III*, or larger sites, e.g. six nucleotides in the case of *Eco RI*. Enzymatic digestion of a DNA molecule by different REs produces several different fragments of varying sizes depending upon the cutting frequency of the RE. Each RE cuts DNA into different fragment sizes, which is not necessarily the same size as those cut by another enzyme (Fig. 25.2). If a piece of DNA from a species is exposed to a specific RE, a characteristic array of DNA fragments is produced; this is called **restriction map**.

Restriction endonucleases cleave double-stranded DNA very selectively at certain specific palindromic sequences. Bacterial cells have these enzymes to destroy the invading foreign DNA.

Nomenclature

The nomenclature of the restriction enzymes is based on the bacteria from which they are isolated. It consists of a three-letter abbreviation: the first letter comes from the genus of the bacterial source, the second and the third letters are derived from the name of the bacterial species. For example, the restriction enzyme *Hae* is isolated from the bacterium, *Haemophilus aegypticus.* An additional Roman numeral indicates the order in which the enzyme was discovered in a particular organism. *Hae III* indicates that this is the third enzyme isolated from *H. aegypticus*, whereas *Hae I* is the first enzyme so isolated from this organism. Likewise, *Taq I* refers to the first RE isolated from *Thermus aquaticus* and *Hind II* refers to the second enzyme from *H. influenzae* (Table 25.1).

Most REs do not cleave in the centre of their recognition sequence but rather one or two base pairs away from the symmetry axis on both strands of DNA. As a result, a double-stranded DNA with short, single-stranded ends is produced. Such single-stranded ends are called **sticky ends** (or cohesive ends). Examples of such REs include Eco RI and Taq I (Table 25.1). Other REs, such as Hae III or Hpa I, cleave vertically the palindromic sequence, leaving two ends that do not have any overhanging nucleotides; these are **blunt ends** or **flush ends**.

Restriction Enzyme Cutting Frequency

Each restriction enzyme cleaves its own cutting site; variation in just one nucleotide adjacent to the cutting site may be sufficient to eliminate *endonuclease* activity for that particular enzyme. Depending on how many sites for a given enzyme are present in DNA, fragments of varying sizes are produced when that DNA is digested by a restriction enzyme. The frequency of cutting sites for various different enzymes varies considerably from enzyme to enzyme, and accordingly the REs are divided into two groups: (a) the ***frequent cutters***—enzymes whose recognition site is very common throughout DNA, e.g. *Pvu II* and *Hae III*; and (b) the ***rare cutters***—enzymes whose recognition sites are uncommon, often involving extended DNA sequence, e.g. *Not I*.

Size of the cut fragments (termed restriction fragments) differs in the two groups: the *frequent cutters generating many small fragments and rare cutters generating fewer but larger restriction fragments*. These differences serve as indispensable tools when it is intended to map the location of certain genes to chromosomal locations.

Biological Functions of REs

Bacterial cells have *restriction enzymes to destroy invading foreign DNA*, for example of the lambda phage which injects its DNA into an *E. coli* cell. The bacterial enzymes recognize the restriction sites and promptly cut the DNA

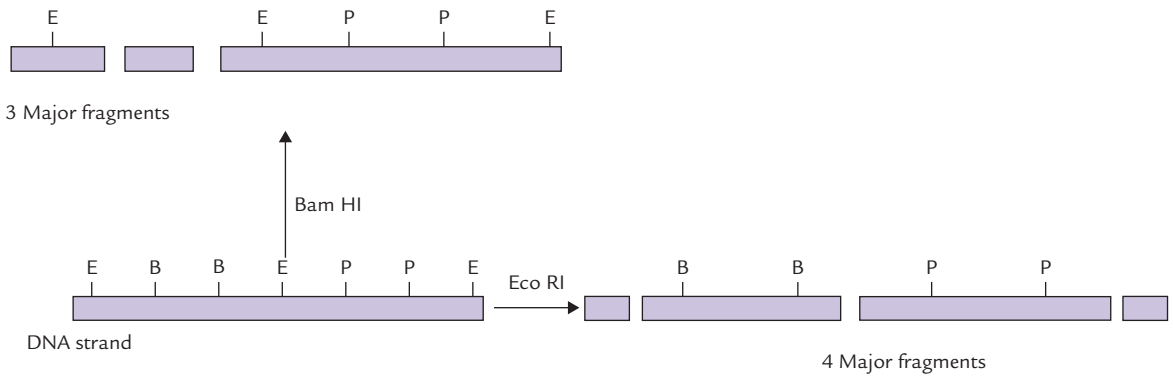


Fig. 25.2. Restriction enzyme digestion of DNA by different restriction enzymes may result in different restriction fragments (E = *Eco RI* site, B = *Bam H1* site, P = *Pst I* site).

to prevent phage infections: it restricts the ability of foreign DNA to infect the cell. A question arises as to why the restriction enzyme does not recognize the same restriction site on the bacterial DNA (a tetramer sequence recognized by Hae III for example must be occurring many times in DNA—statistically every 4^4 base pairs) and cleave at that site. *The cell guards against cutting its own DNA by* ***methylation***: it adds a methyl group to one of the bases in all of the recognition sequences on each new DNA strand soon after it is synthesized. This does not interfere with the base pairing or gene function, but the restriction enzyme no longer recognizes the methylated sequence and the cell's own DNA is, therefore, protected from attack by that enzyme.

This system, however, is not absolutely reliable because occasionally the intruding virus encounters the methylating enzyme first; and once methylated, it is resistant to the RE.

Terminal Transferases

They catalyze the addition of homo-oligonucleotide sequences (poly dN) to the 3′ end of double-stranded DNA molecule (Fig. 25.3a). No template is required for these reactions. The 3′ ends of DNA serve as primers. A given enzyme can catalyze addition of many copies of a particular nucleotide only, and there are separate enzymes for each of the four deoxyribonucleotides. For example, the enzyme catalyzing addition of the poly-dA sequence is different from that catalyzing addition of the poly-dT sequence.

Reverse Transcriptase (RT or RNA-Dependent DNA Polymerase)

RT is an enzyme encoded in the genome of RNA-containing retroviruses. It catalyzes *formation of DNA on an RNA template* using deoxyribonucleoside triphosphates as substrates (Fig. 25.3b). Primer is required by *RT*. The DNA thus synthesized has a base sequence complementary to that of the RNA template and is called the **complementary DNA** or **cDNA.** For details regarding action of *RT* refer to Boxes 25.1 and 25.2.

Ribonuclease H

It is an *endonuclease* that catalyzes random hydrolysis of phosphodiester bonds of RNA, but only when the latter forms a hybrid duplex structure with DNA (Fig. 25.3c).

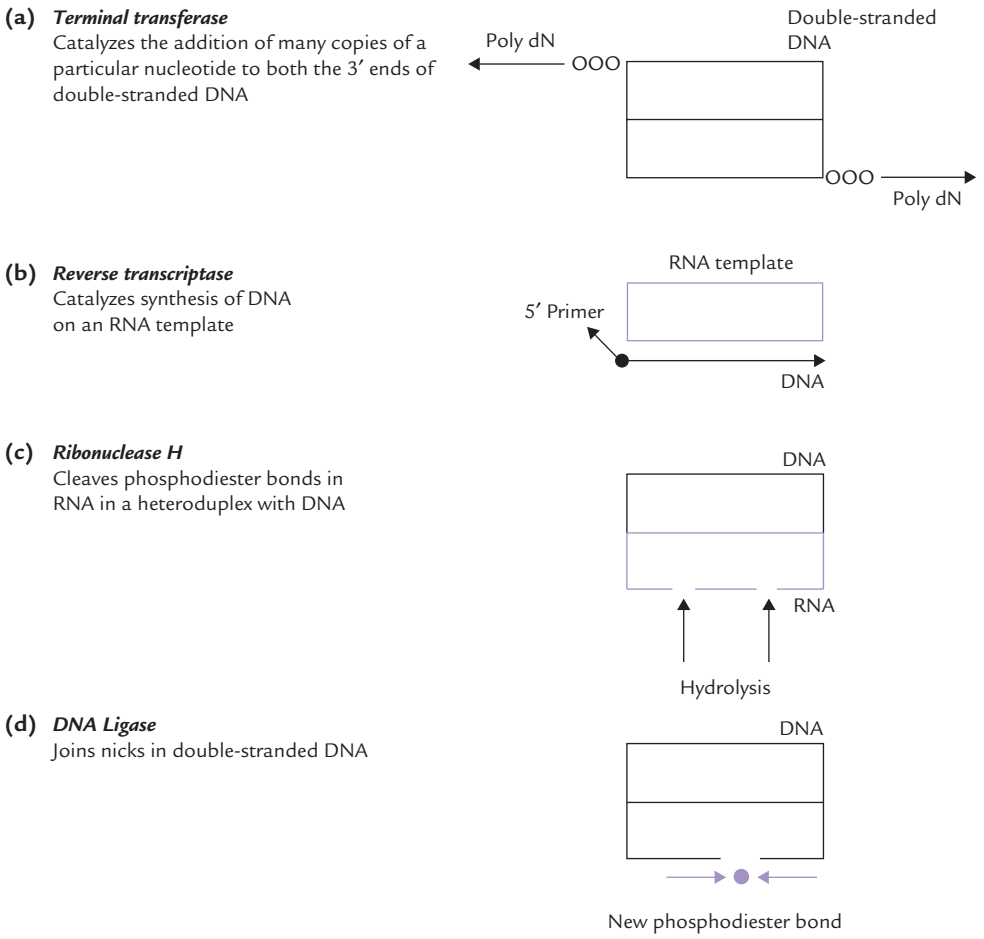


Fig. 25.3. Some enzymatic reactions that involve DNA.

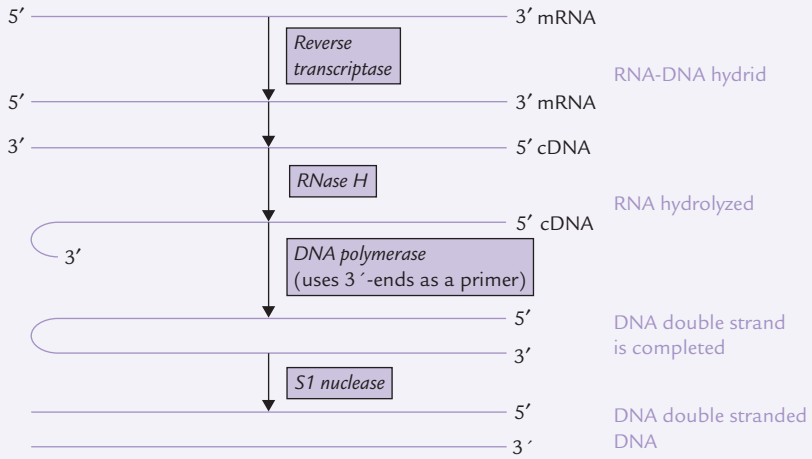
BOX 25.1

Reverse Transcriptase

Copying of RNA template to synthesize DNA by *reverse transcriptase* was initially considered antithetical to the central dogma of molecular genetics. However, there is no thermodynamic prohibition to the *RT* action and, in fact under certain conditions, *DNA polymerase* can also copy RNA templates. *RT* is an essential enzyme of **retroviruses** which are the RNA-containing viruses, such as human immunodeficiency virus (HIV; the causative agent of AIDS). The enzyme was independently discovered in 1970 by **Howard Temin** and **David Baltimore**.

RT catalyzes the first step in the conversion of the single-stranded RNA genome of virus to a double-stranded DNA. After the virus enters the cell, its *RT* synthesizes complementary DNA from the viral RNA template to yield an RNA-DNA hybrid helix. (The DNA synthesis is primed by a host cell tRNA whose 3′-end unfolds to base pair with a complementary segment of viral RNA.) The viral RNA strand is then nucleolytically degraded by an *RNase H* (an *RNase* activity that hydrolyzes the RNA of an RNA-DNA heteroduplex). The single-stranded DNA is now left. This DNA acts as a template for the synthesis of its complementary DNA by *DNA polymerase*.

Finally, action of *S1 nuclease,* a single-strand *endonuclease*, yields a double-stranded DNA that is then integrated into the host cell chromosome.



RT is a useful tool in genetic engineering because of its ability to transcribe mRNAs to complementary DNA strand. The latter can be used, for example, to express eukaryotic structural genes in *E. coli*. Since the bacterial cell cannot splice out the intron segments, the use of genomic DNA to express a eukaryotic structural gene in *E. coli* would require prior excision of the introns. This is a difficult feat and highly unfavourable thermodynamically.

DNA Ligases

These enzymes catalyze ATP or NAD^+-dependent joining of nicks in double-stranded DNA (Fig. 25.3d).

C. Cloning Vectors

The cloning vector is a self-replicating entity which is capable of carrying DNA into the host cell. Essential properties of a cloning vector are as here.

1. It must contain at least one specific *palindromic region* for a RE.
2. It must be capable of *autonomous replication*.

A cloning vector may be a ***plasmid***, ***bacteriophage***, or a ***cosmid***. An essential technique in genetic engineering is to introduce the human gene into a cloning vector. Covalent binding of the two results in production of ***chimeric DNA*** or ***recombinant DNA***, which is then carried into a suitable host cell (Fig. 25.4). The inserted gene is replicated, transcribed and translated in the new setting, thus undergoing mass amplification. The name 'chimeric' derives from *chimera*, a mythological creature with the head of a lion, the body of a goat and the tail of a serpent.

Cloning vector, a self-replicating entity, helps in the *in vivo* propagation of DNA by effectively introducing it into a bacterium or some other suitable host cell.

Plasmids

Plasmids are circular, extrachromosomal DNA molecules of 2000–200,000 base pairs, present in some prokaryotic

BOX 25.2

Treatment of HIV Viral Infection by AZT

A treatment methodology developed for AIDS takes advantage of the fact that *reverse transcriptase* is an error prone enzyme. It not only lacks the *exonuclease* activities of *DNA polymerase,* but also makes mistakes during identification of its substrate. A potent drug for AIDS treatment is azido-2′ 3′-dideoxythymidine (AZT), which is metabolized in the body into azide. TPP, which is a thymidine triphosphate (TPP) analogue. The *HIV reverse transcriptase* misincorporates azido-TPP into the RT-transcribed viral genome instead of its natural substrate, TPP. Because the 3′ group of azido-TPP cannot form a phosphodiester bond with subsequent nucleoside triphosphate, it results in blocking of further chain elongation. As a result the virus cannot be replicated.

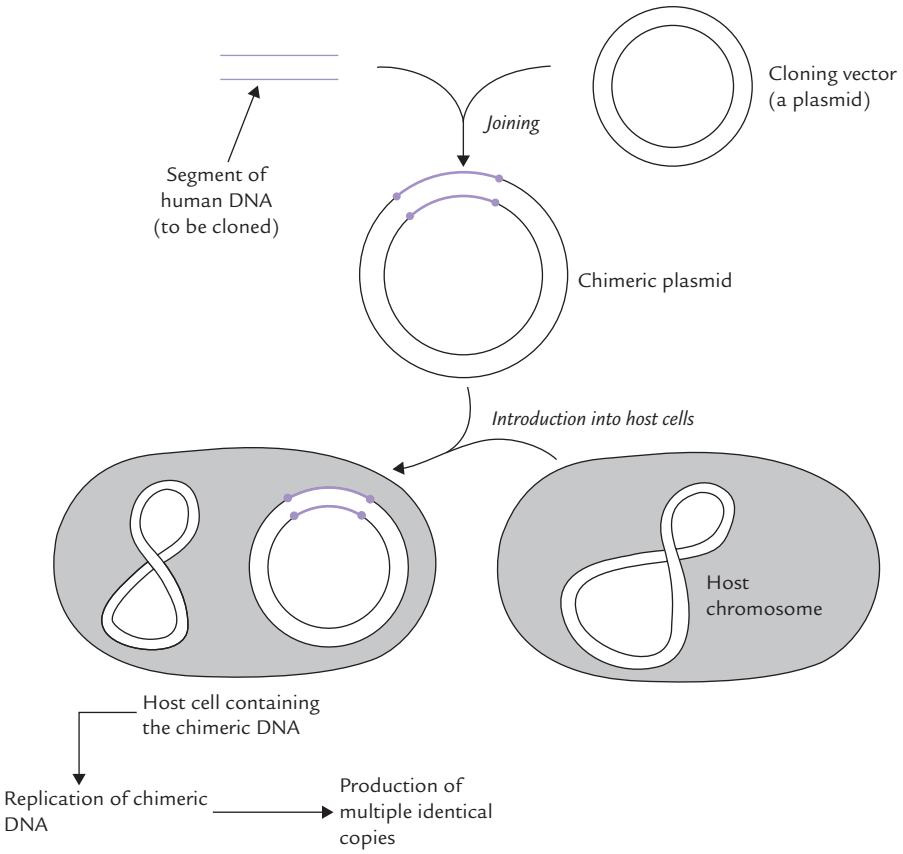


Fig. 25.4. Role of cloning vector plasmid in carrying chimeric DNA into the suitable host cell where it is amplified.

cells. Most plasmids are very small and contain only a few genes, compared with the prokaryotic chromosomal DNA which contain thousands of genes. In some cells, however, plasmids may be quite large. A single cell may contain 10–20 copies of plasmids. They undergo replication with each cell cycle, which may or may not be synchronized to the chromosomal DNA.

Plasmids usually carry genes that impart resistance to antibiotics. Thus, a bacterium having a plasmid, which has genes for ampicillin resistance, will be resistant to this antibiotic. Plasmids serve as efficient means of delivering a desired DNA segment into a host cell; the process is called **transfection**. Two remarkable properties of plasmids enable them to perform the stated function:

1. They can pass from one cell to another and also from one species of bacteria to another; for instance, some plasmids readily move from the *E. coli* cell to the *Salmonella typhimurium* cell. This explains the observation that mixing of the *S. typhimurium* cells with an ampicillin resistant strain of *E. coli* makes the former cells also resistant to ampicillin. The genes for ampicillin resistance, present in a plasmid of *E. coli* are transmitted to the *S. typhimurium* cells in this case.

2. Second important property is that foreign genes can be inserted into plasmids quite easily and may then be carried into a bacterium where they become part of the host-cell genome. Technical details about the role of plasmid in gene manipulation will be described later in this chapter.

Plasmid vectors suffer from a serious disadvantage: they can accommodate only small pieces of DNA, that is, up to about 5000 base pairs (5 kb). Larger inserts tend to be deleted randomly during replication of the plasmid.

Bacteriophages

Lambda (λ) phage is another choice vector. In view of the disadvantage in case of plasmids (the plasmids cannot carry large fragments of DNA) there is a need of other cloning vectors capable of carrying larger fragments. Lambda (λ) phages were found to be capable of carrying much larger DNA inserts, as large as 40 kb.

The genomic material of λ phage comprises a linear, double-stranded DNA of 48,513 base pairs (Fig. 25.5). It is packaged into the capsid protein within the host cell. The genome has two distinctive properties which permit easy insertion of foreign DNA into it:

1. About one-third of the λ phage genome (lying in the middle section) is not essential for integration into the host-cell chromosome. This nonessential DNA can be replaced by an insert DNA, without affecting properties of the phage genome.
2. The DNA fragments present at the ends (referred to as ***cos sites***) form packaging signals. They are essential for packaging the DNA into the phage particles.

Insertion of Foreign DNA into Bacteriophage

It involves the following steps:

- The middle section is removed first by certain *restriction endonucleases* (such as *Eco RI*). Nearly one-third of the phage DNA is thus removed, leaving behind two pieces having a combined length equal to approximately 70% of the unit genome (Fig. 25.5).
- Incorporation of a suitably long DNA insert is done next, to create a ***recombinant DNA***. The DNA insert should be of nearly the same size as the middle-section (~15 kb).
- The insert DNA is then linked covalently at its ends by *DNA ligase* to create a recombinant DNA.
- **Packaging reactions** follow, wherein the recombinant DNA is mixed with capsid proteins and the infective phage particles are assembled by *in vitro* packaging. This results in the formation of λ virions, harbouring the foreign DNA.

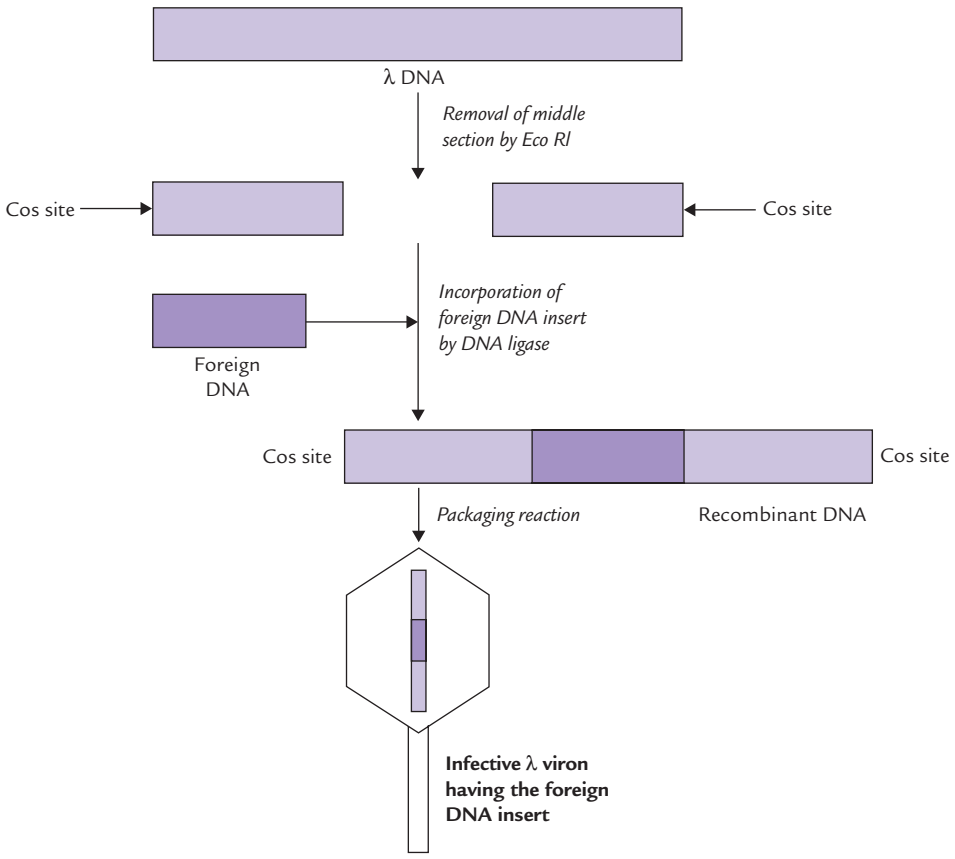


Fig. 25.5. Lambda phage as cloning vector.

An advantage of using the λ virions is that it "**injects**" its DNA into the bacterium with an efficiency much higher than that of plasmids. The injected DNA is either (i) incorporated into the bacterial chromosome to replicate during mitotic division, or (ii) it replicates within the cell independent of the bacterial chromosome. Such independent replication leads to synthesis of new viral particles, which can then lyze the host cell and then infect adjacent cells.

Cosmids

A cosmid is an artificially constructed circular DNA molecule of 5000–7000 base pairs. It *combines useful features of both, the plasmid and the λ phage*, because of the presence of the following elements:

1. An origin of replication.
2. Packaging signals of λ phage, called *cos site*.
3. In addition, the cosmid contains an antibiotic resistant gene and several restriction sites where foreign DNA can be inserted. The cosmid permits insertion of large fragments of DNA, up to 35–45 kb long.

Others

Larger inserts (up to 3000 kb can be cloned by using modified chromosome from either bacteria (***bacterial artificial chromosome, BACs***), or yeast (***yeast artificial chromosome, YACs***).

II. DNA Amplification and Cloning

Production of multiple copies of a gene from the same origin (**cloning**) is central to the study of molecular biology and genetics. The cloned DNA can be used for many purposes; the human genome project, for example, used cloned DNA to sequence the human genome. Moreover, several techniques in molecular biology require *visualization* or *quantification* of the DNA and this requires mass amplification.

Two methods for the mass amplification of DNA are introduced in the following section.

- ***Cell-based cloning:*** DNA is amplified *in vivo* by a cellular host, such as a replicating bacteria (commonest), baker's yeast, frog oocytes or cultured hamster's ovary cells. The desired DNA template is introduced into a cellular host. As the number of replicating host cells increases (by cell divisions), the number of copies of the desired DNA template also increases correspondingly.
- ***Enzyme-based cloning:*** This method, represented by polymerase chain reaction (PCR), involves *in vitro* DNA amplification. It concentrates on the targeted increase in the number of copies of a specific DNA sequence.

A. Cell-based Cloning

This method has been introduced earlier (Fig. 25.4). It uses recombinant DNA in the replicating bacteria, and is based on the ability of the latter to sustain presence of the recombinant DNA. The term **recombinant** refers to any DNA molecule that is artificially constructed from two pieces of DNA not normally found together. One of these pieces will be the ***target DNA*** that is to be amplified and the other will be the *replicating vector* or ***replicon*** (a sequence capable of initiating DNA replication).

This method requires separation of target DNA, its insertion into vector to produce chimeric (recombinant) DNA, and introduction into host cell where it is amplified. It is a complex process which can be divided in the following major steps:

- Isolation of the target DNA.
- Introduction of target DNA into replicon.
- Transformation of the host cell.
- Isolation of DNA insert or its protein product.

Step 1: Isolation of the Target Gene

Isolation of a specific gene of interest is a difficult task because a given gene represents a very small part of the total genomic material. There are three general approaches for obtaining the target gene:

- *Reverse transcription of mRNA:* This procedure involves isolation of mRNA complementary to the target gene and obtaining cDNA by reverse transcription of the mRNA. Though it is an indirect approach, this method is considered simpler because of the relative ease with which mRNA and then cDNA is obtained.

  Note that the cDNA contains all coding regions of the target gene (without any introns).
- *Shotgun approach:* It involves (i) fragmentation of the entire genome, (ii) joining of all the DNA fragments with plasmids opened by the same *restriction enzyme*, and (iii) selection of the plasmids containing the desired gene. These aspects are described later in this chapter (see genomic library and Southern blotting).
- *Chemical synthesis of the gene:* The DNA that codes for a small protein may be chemically synthesized. However, since genetic code is degenerate, it is not possible to know the precise DNA sequence corresponding to the protein. More DNAs than one have to be synthesized.

Step 2: Introduction of Target DNA into Replicon

The target DNA is introduced into a replication vector (mostly plasmid vector) using *restriction enzyme* to cut the target and the replicon DNA. The enzyme used is the one

that produces sticky ends. The plasmid, when digested with such an enzyme, for example *Eco RI*, becomes linear DNA strand with sticky ends complementary to the target (Fig. 25.6). Annealing of the two yields a new circular ***chimeric (recombinant) plasmid*** with the target gene inserted.

Creation of sticky ends by **homopolymer tailing** can also be used to construct recombinant plasmid with the enzyme *terminal transferase* (Fig. 25.3a). A poly(A) tail may be attached to one DNA fragment and poly(T) to another, so that they can base pair to yield recombinant DNA.

Step 3: Transformation of the Host Cell

This refers to cellular uptake of plasmid DNA. This step (also called *transfection*) is critical in cell-based cloning because each of the successfully transformed cells would act as individual replication unit, capable of amplifying the target DNA in huge amount (Fig. 25.7). However, the cell membrane of bacteria is selectively permeable, restricting passage of large molecules such as DNA into the cell. Efficiency of transformation is very low: only 1–2% of cells may take up plasmid DNA, and often only a single plasmid is introduced during transformation. However, the permeability of the cell membrane can be increased by factors such as electric currents (***electroporation***) and high-solute concentration (***osmotic stress***), so that DNA can move into the cell more freely. Such processes are said to render the cells more competent, i.e. they can take up foreign DNA from extracellular fluid.

The recombinant DNA has to be brought into the host bacterial cell by transformation. Transformation is rather inefficient but it can proceed with reasonable yield if the recombinant plasmid is mixed with the bacteria in presence of high solute concentration of if electric current is passed.

Selection of transformed cells: Following transformation, the next step is to identify/select the cells that have been successfully transformed. Such cells contain the recombinant plasmids, with target DNA inserted. Their selection

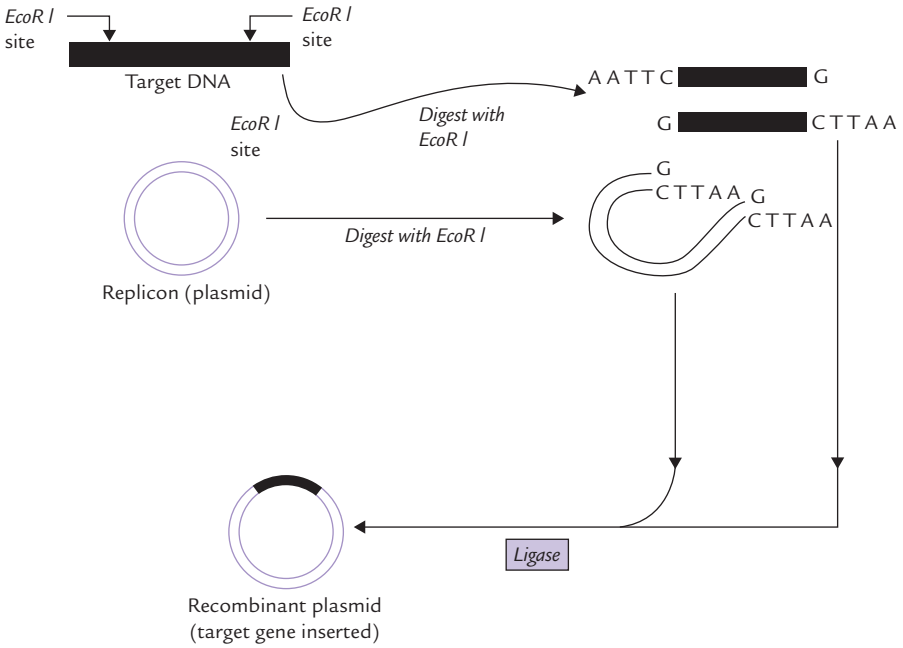


Fig. 25.6. Formation of a recombinant plasmid containing a target gene for cloning. Both the plasmid and the target DNA are cut with the same restriction enzyme and mixed together. The cohesive ends of each DNA reanneal and are ligated together to form the recombinant plasmid molecule which would be introduced into bacterial host cells.

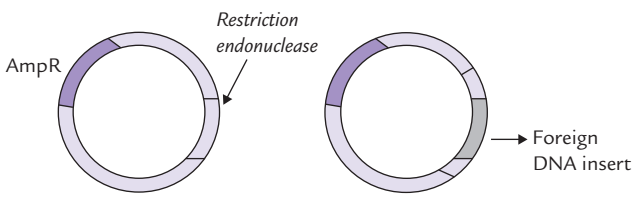


Fig. 25.7. Structure of a plasmid containing **(a)** genes conferring resistance to antibiotics ampicillin (**AmpR**), **(b)** the inserted DNA within the plasmid. The antibiotic resistance genes are useful for selection of the transformed cells.

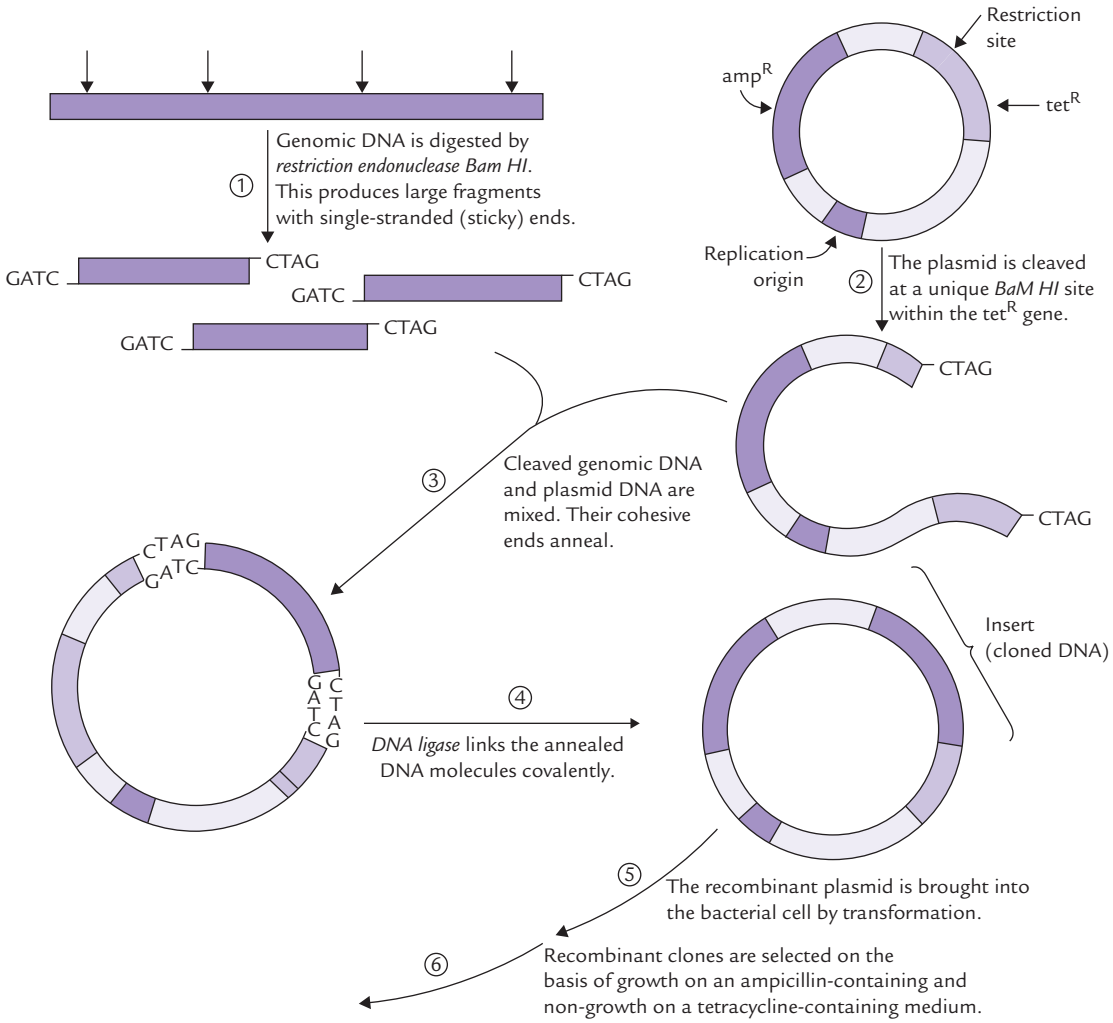


Fig. 25.8. The use of plasmid pBR322. This plasmid has been constructed specifically as a cloning vector.

makes use of the antibiotic resistance gene(s) present in the plasmid vector. *If the vector contains an antibiotic resistance gene the transformed host cells will become resistant to this antibiotic*. Plating on medium containing the relevant antibiotic will allow only the transformed host cells to grow. The non-transformed cells will be destroyed by the antibiotic. The following example will clear this point.

- The plasmid shown in Figure 25.7 contains genes that confer resistance to ampicillin.
- After transformation, the bacteria are cultured in a medium that contains ampicillin.
- The transformed cells will not be destroyed by this antibiotic because the gene for ampicillin resistance are present in them.
- The non-transformed bacteria (not containing the plasmid DNA), will however be destroyed by ampicillin.*

*To learn further about these aspects, the student may refer Figure 25.8 gives a detailed account (of use of the plasmid pBR_{322}), which the undergraduates may skip.

Cell-based cloning uses recombinant DNA in replicating bacteria or other host cells. Selectable markers can identify the transformed host-cells.

As shown in the Figure 25.8, this plasmid contains genes for ampicillin resistance and tetracycline resistance. Several *restriction endonucleases* (*BamHI* in this example) cleave the plasmid at unique sites within either the ampicillin resistance (amp-R) gene or the tetracycline-resistance (tet-R) gene. In the application shown in Figure 25.8, the restriction enzyme cleaves the plasmid at the gene for tetracycline resistance. Thus, the tet-R gene is destroyed while the amp-R gene remains intact. Bacteria transformed by this recombinant plasmid can grow in the presence of ampicillin but not of tetracycline. They are grown in suitable medium to form colonies.

Properly transformed bacteria can grow in presence of ampicillin but not in presence of tetracycline.

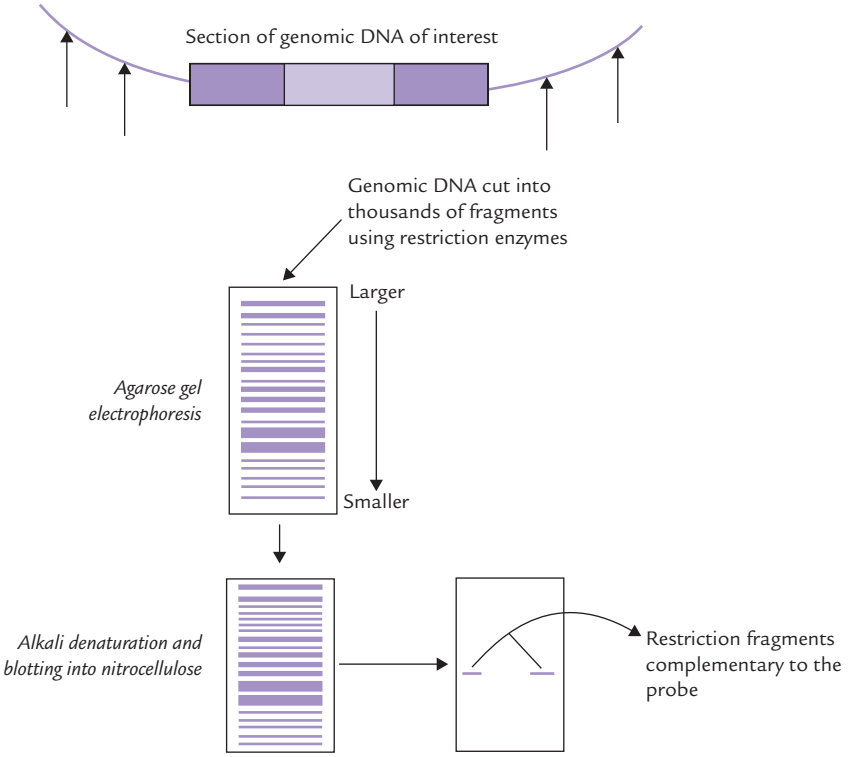


Fig. 25.9. Identification of a target DNA sequence by southern blot technique.

BOX 25.3

Production of Human Recombinant Insulin by Cell-based Cloning

Insulin was the first biologically important human peptide to be produced by means of recombinant DNA technology. One crucial step in the process is obtaining insulin mRNA. How to separate insulin mRNA from several other mRNAs in the β-pancreatic cell? A general way is to lyze the cells, collect the polyribosomes by centrifugation and treat them with an antibody directed against insulin.

The mRNA is used as a template for the synthesis of human insulin cDNA (by reverse transcription). The cDNA contains all the coding regions of the insulin gene. It is used as template for transcription and translation to obtain insulin. To do so, the cDNA is cloned into a plasmid vector for *in vivo* amplification. By this means, large amount of human insulin is produced and it has virtually totally replaced animal insulin in the treatment of diabetes.

Further selection of colony having the desired DNA: The cells containing the chimeric-plasmids are grown to form colonies. However, there will be many more colonies also. This is because the bacteria containing other plasmids (not the desired chimeric plasmid) also grows to form distinct colonies. In this way we obtain a large collection of bacterial clones, each containing a random piece of DNA insert. Identification of the cells that contain the desired DNA insert is the only problem left now. This requires the process of ***screening*** done by labelled DNA probe complementary to the desired DNA insert. The probe may be an isolated DNA fragment or a synthetic oligonucleotide (discussed later).

Step 4: Isolation of DNA Insert or its Protien Product

Once a clone with the desired DNA insert has been identified, the vector can be recovered from the bacterial clone and the **insert can be excised** with the same *restriction endonuclease* that had been used for the construction of the recombinant plasmid. In order to isolate large quantities of protein product of the inserted DNA, the cells are grown under conditions that are favourable for production of proteins. The protein is then isolated from these cells by usual methods. Several medically important proteins are thus obtained (Box 25.3).

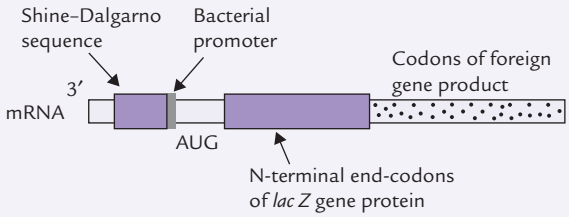
Usually the recombinant DNA technology successfully produces human proteins using enzymatic machinery of

BOX 25.4

Expression Vector

The recombinant DNA methodology suffers from a major drawback, which is—the cloned genes can be rapidly propagated, but they cannot be effectively expressed in bacterial cells. Hence this methodology is not effective for generating the protein products of the cloned genes. This is a major pitfall because an important goal of recombinant DNA studies is to propagate a foreign gene in bacteria and obtain the product in a biologically active form. To achieve this goal, expression vectors, which facilitate expression of the DNA inserts, have been developed. The following conditions must be met during their construction:

1. Since bacteria can express the recombinant eukaryotic gene only when properly positioned transcriptional and translational control sequences are present, a strong bacterial promoter region is added to the upstream of the eukaryotic gene. It is then inserted into the vector.
2. The mRNA transcript of a recombinant eukaryotic gene must contain the bacterial ribosome binding sequence (the Shine–Dalgarno sequence) required to properly orient it with functional bacterial ribosome.
3. Only complementary DNA inserts (not genomic DNA) must be used. This is because the complementary DNA can be effectively expressed in bacterial cell (bacteria cannot remove intron sequences from primary transcript).



The whole of the cloned complementary DNA is inserted within a bacterial genome, commonly in the *lac Z* gene. The mRNA transcript of the recombinant DNA contains the following elements in that order; *lac* Z-Shine–Dalgarno sequence, the bacterial promoter, the start codon, and the codons for first few amino acids of the bacterial protein (encoded by a portion of the 3′-end of the *lac Z* gene). This is followed by codons of the foreign gene product.

The engineered protein product is a fusion protein that contains few N-terminal amino acids of the *lac Z* gene protein followed by the complete amino acid sequence of the cloned complementary DNA.

bacteria. However, *bacterial RNA polymerase may not recognize the human promoter regions, and so expression of human genes may be minimal.* Use of **expression vectors**—these include bacterial plasmids into which appropriate transcriptional and translational signals have been added—permit effective expression of human genes in bacterial cells (Box 25.4).

The aforementioned discussion centres around plasmid vectors, which appear to represent *ideal replicon* for DNA amplification. They undergo intracellular replication, are passed vertically from the parent cell to each daughter cell, and (unlike chromosomal DNA) are copied many times during each cell division. Comparative features of other vectors, such as lambda phage, cosmid, BACs and YACs, have been discussed earlier.

B. Enzyme-based Cloning

It involves use of PCR (described later) for *in vitro* amplification of DNA.

III. DNA Library

A. Genomic Library

Genomic library *refers to a large collection of bacterial cells, each containing a random piece of human genomic DNA.* For constructing a gene library, the entire DNA of cell is cleaved into small pieces by using different REs. All these cut fragments are then introduced into appropriate vectors. This forms a large collection of different recombinant clones, which are then introduced into the host bacterial

cells to form the gene library. All the genes of an organism are represented in the gene library. In order to produce a complete gene library for *E. coli*, about 1500 fragments are required, whereas about 10 lakhs fragments are required for human gene library.

B. cDNA Library

In a genomic library, all the genomic material is represented but only about 3% of the cloned DNA codes for proteins. If a collection of only expressed DNAs were to be made, then RNA and not DNA would be the starting point.

***cDNA library** is a collection of all the expressed DNA of a particular cell type or tissue*. For example, a cDNA from pancreatic β-cell contains clones with cDNA for proinsulin. On the other hand, a cDNA library from bone marrow cell contains many clones with cDNA for α- and β-chains of haemoglobin. Thus, for a cDNA library the tissue of origin is important. For a genomic library, the tissue of origin is unimportant because the genomic material is same in all cell types of an organism.

Building of cDNA library: The mRNA is extracted from a specific tissue. It is used as a template for synthesis of complementary DNA strand; the enzyme catalyzing this synthesis is *reverse transcriptase*, which yields a single-stranded cDNA (Fig. 25.3). Double-stranded cDNA is then obtained from it by adding the *DNA polymerase*. The latter is incorporated in plasmid, λ phage or cosmid and introduced into host bacterial cell.

Genomic libraries contain all the DNA of an organism, and cDNA libraries contain only expressed DNA.

Screening of genomic library: It is possible to screen the genomic library and obtain a gene of interest from it. Although a daunting task, use of *nucleotide probes* has made it possible.

IV. Nucleotide Probes

To search a desired DNA sequence of interest from a vast array of DNA fragments present in a genomic library, a reagent is needed which would react only with the correct fragment and ignore the rest. This type of reagent is called a nucleotide-based probe. It is a single-stranded piece of DNA (sometimes RNA), which can range in size from as little as 15 bp to several hundred kilobases (Table 25.2). It can identify, through base pairing, a specific DNA fragment of the library, which contains complementary sequence. A probe DNA, for example, will form complementary base pairing with another DNA strand (termed template) if the two strands are complementary and a sufficient number of hydrogen bonds are formed.

Table 25.2. Characteristics of nucleic acid probes

Probe type	Size	Origin
DNA	0.1–100 kb	Cell-based DNA cloning, PCR
RNA (or riboprobe)	1–2 kb	RNA transcription from plasmid (or phage vectors)
Oligonucleotide	15–50 nucleotides	Chemical synthesis

An effective molecular hybridization between the probe and the template requires that both must be single stranded. Likewise, a mRNA can be used as a probe: it will bind to the DNA fragment that contains exon sequences of its gene. RNA probe, termed **riboprobe**, can be produced by *in vitro* transcription of cloned DNA inserted into a plasmid vector. **Synthetic oligonucleotide probes**, constructed by chemical methods, are most commonly used. The probe has to be at least 15–18 nucleotides long because shorter sequences may be present, by chance, at multiple sites in the genome (e.g. a trinucleotide every 4^3 base pairs). Oligonucleotides of this size can be easily synthesized.

A. Probes must have a Label to be Identified

To render them recognizable, probes are labelled with the **radioisotopes**, such as ^{32}P or tritium. These probes can be detected by autoradiography, which involves placing the sample in direct contact with the photographic material, usually an X-ray film. Alternatively, end-labelling probes with **fluorescent tags** can be used. The latter are visible under the UV lamp.

B. Techniques for Labelling Probes

There are two general ways in which a labelled nucleotide can be incorporated into the structure of the probe:

1. ***End-labelling:*** Addition of a labelled group to one terminal of the probe is done, for example, by exchanging a labelled γ-phosphate from ATP with a phosphate from the 5′-terminal on (single or double-stranded) DNA.
2. ***Polymerase-based labelling:*** Using a *DNA polymerase*, multiple-labelled-nucleotides are incorporated into the probe during DNA synthesis. Such a reaction requires dNTPs, and it is customary to have one of them to be labelled, e.g. dGTP. Because on an average 25% of the nucleotides incorporated are labelled, *this*

type has a higher specific activity than the end-labelling where only terminal nucleotide is labelled.

C. Uses of Nucleotide Probes

1. To search specific DNA sequences of DNA library, as discussed.
2. In Southern and Northern blot techniques, probes are used to identify DNA or RNA fragments respectively, as described later.
3. In diagnosis of genetic disorders, such as sickle cell anaemia, thalassaemia, cystic fibrosis, etc.

Probe gets involved in the formation of heteroduplex with template DNA, and this is the key to usefulness of molecular hybridization.

V. Blotting Techniques

These are standard techniques for the identification of a specific DNA, an RNA or a protein from a vast expanse of others. The technique for DNA identification is termed Southern blot, whereas Northern blot is for RNA and Western blot for protein identification.

A. Southern Blot Technique

This technique, used for the identification of a specific DNA sequence, is named after its originator, ***E.M. Southern***, who developed the method in 1970s. Supposing it is proposed to identify a restriction fragment that carries a particular DNA sequence, for example, exon of a protein coding gene (*11-γ-hydroxylase;* Fig. 25.9), the method would involve the following steps:

1. ***Extraction*** of genomic DNA from the cells, and cutting into fragments by *restriction endonuclease* to prepare a ***restriction digest***. The latter is likely to contain thousands of double-stranded restriction fragments.

2. ***Electrophoretic separation*** of different fragments on the basis of their size in a cross-linked agarose or polyacrylamide gel; the smallest fragments move furthest, whereas the larger ones are retarded by smaller 'gaps' in the gel. This method can separate fragments ranging in size from 100 bases to approximately 20 kb in length (above 40 kb resolution is minimal). Furthermore, the method is sufficiently powerful to separate even those restriction fragments that may differ in length by only one or a few nucleotides (refer to Box 25.5 for more information on electrophoretic separation).

3. ***Denaturation*** of the DNA fragments to render them single-stranded by soaking the gel slab in a strong alkali solution. This step is needed because the nucleotide probe (single-stranded oligonucleotide that is complementary to the desired DNA sequence on a restriction fragment) can identify only a single-stranded DNA by hybridization.

4. ***Transfer*** of the single-stranded DNA to a sheet of *nitrocellulose filter* (or nylon membrane); the process of transfer involves the passage of solute through the gel and into the filter, passively, carrying the DNA. Transfer of DNA from the semisolid phase of a gel to a solid phase in this manner is referred to as **blotting** (hence the name blot).

   This step is important because the restriction fragments are not tightly bound to the semisolid gel and therefore at risk of being easily leached out of the gel slab. *Nitrocellulose or nylon immobilizes the single-stranded DNA to form a replica of the DNA from the gel, which is still accessible to hybridization reaction with probes.*

5. ***Addition of nucleotide-based probe*** (RNA or single-stranded DNA) by dipping the nitrocellulose filter into a neutral solution of the probe and holding at a suitable temperature and time period to allow hybridization of the probe with the matching DNA fragment. The filter is then washed to remove the excess unbound probes. The radioactively labelled probes are generally used.

BOX 25.5

Pulsed Field Gel Electrophoresis

The molecules of DNA can migrate through the small gaps between agarose particles during electrophoretic separation. But larger DNA molecules, which are obtained by rare cutting restriction enzymes, encounter hindrance because of their larger size (often several hundred kilobases) and their coiled tertiary structures. They are separated by using a special technique; in which polarity of the electric field is changed periodically. It means that the current is applied in different directions at alternating times, which results in uncoiling of the large DNA fragments. The uncoiled fragments can then slowly migrate between ('**snake through**') the agarose particles. This technique enables us to size fractionate DNA of fragments over 100 kb.

6. ***Autoradiography or fluorescent scanning*** determines the position of the labelled probe, base paired with the sequence of interest.

Southern blotting is prototype of permanent DNA storage, wherein a replica of the gel slab with its separated restriction fragments is made on the nitrocellulose filter. The replica can act as template for any number of different probes.

The DNA transferred from the semisolid to the solid phase, if preserved properly, forms a permanent record of the digested DNA and then acts as a template for any number of different probes. Either DNA or RNA probes can be used. Subsequently, analogous techniques were developed for the transfer of RNA and proteins, and adopting the theme of direction, they were termed Northern and Western blots, respectively (Table 25.3).

B. Northern Blot Technique

This technique (which is similar to the Southern blot technique) is used for **identification of RNAs** by *electrophoresis, blot-transfer* and *probing* with radioactive RNA or single-stranded DNA probes (Table 25.3).

Table 25.3. Blots used in molecular biology

Blot	Probe	Template
Southern	Nucleotide	DNA
Northern	Nucleotide	RNA
Western	Antibody	Protein

C. Western Blot Technique

It is a method for (electrophoretic) separation and **identification** of a **specific protein** from a mixture. In contrast with the Southern and Northern blots, which use nucleotide-based probes, this technique relies upon the ability of a **monoclonal antibody** to bind to a protein that it has been raised against.

A combination of some of the techniques described above permit analysis of long stretches of DNA from genomic library. The method used is **chromosome walking** (refer to Box 25.6 for details).

In situ Hybridization

It is a **modified version of hybridization** in which *labelled probes are used to detect complementary DNA sequences in tissue sections*. The term *in situ* is applied because the DNA is not extracted in this process. The following steps are involved:

- Histology slides are prepared from the tissue under study and examined by traditional staining techniques.
- DNA is then denatured with NaOH and specific probe is layered over it.
- The probe would hybridize, on the microscope slide, with the complementary sequences of the denatured DNA in tissue sections (or whole cell).

BOX 25.6

Chromosome Walking

Analysis of long stretches of DNA gives important information regarding arrangement of genes and their regulatory elements in a chromosome. To make such analysis from genomic library is not a simple and straightforward procedure because, currently it is not possible to clone long DNA fragments. Size of an insert in a genomic library is very small (few thousand base pairs) when compared to eukaryotic genes, which may be several hundred thousands base-pair long. Therefore, analysis of long stretches of DNA from genomic library is a difficult feat. It is made possible by a combination of techniques. The method is called chromosome walking or *overlap hybridization* which permits analysis of long DNA stretches of 300 or more kilobases and defining gene arrangement therein. It is a multistep procedure as discussed below:

1. First, *overlapping restriction fragments* are generated and cloned in a DNA library. The cloned fragments will contain overlapping fragments with other cloned sequences.
2. The *DNA library is screened* for a sequence of interest with a DNA or RNA probe. A reactive clone is thereby identified with this (original) probe.
3. The reactive cloned fragment (insert) is now *sequenced*. A new probe is then synthesized which recognizes a sequence near the end of the insert. This new probe recognizes a new clone with an overlapping insert.
4. Once the new clone is identified, its insert is sequenced as well and another *new probe is synthesized* for an end sequence of this fragment.
5. This procedure can be *repeated till whole gene is sequenced* from the cloned fragments in a genomic library.

- After the probe is properly fixed, it is autoradiographed; hybridization complexes are seen as silver grains on photographic film.

The **advantage** of this technique is that it allows correlation of findings of cytopathological examination with the presence of an intracellular viral pathogen, as shown by use of probe.

This technique has emerged as an important diagnostic tool and is particularly useful for detection of intracellular viral pathogens. Some modified versions of *in situ* hybridization are as below:

1. ***Fluorescent in situ hybridization (FISH):*** It is a non-isotopic *in situ* hybridization (NISH) technique where **fluorescent labelled probes** are used. It is faster, safer, more sensitive, and shows greater resolution compared to the ordinary *in situ* work discussed above. Probes used in FISH are short, of about 20–25 nucleotides long. The slide is visualized using ultraviolet light. Under the ultraviolet light, the probes emit fluorescence, intensity of which depends on the extent of hybridization.
2. ***Use of riboprobes:*** Riboprobes are RNA probes produced by *in vitro* transcription of cloned DNA inserted into plasmid (or phage) vectors (Table 25.2). *The labelled radioprobes can be used to detect complementary RNA in tissue sections*. To maximize the efficiency of the hybridization, **cRNA probe** is preferred. cRNA is single-stranded and is complementary to the transcribed mRNA of the gene. Therefore, it hybridizes with any mRNA present in the tissue section with high efficiency, and serves as a useful diagnostic tool (**Case 25.1**).

VI. Applications of Recombinant DNA Technology

The knowledge of genes and techniques for manipulating them is revolutionizing medicine. The genes for several genetic diseases have been mapped and sequenced already, and molecular methods for diagnosis of these diseases have been developed. Treatment based on manipulation of gene expression was no more than an intellectual pastime, but is a reality now. A variety of proteins are being produced by genetically engineered micro-organisms, and widely used as therapeutic agents. However, treatment based on transfer of intact gene into the cells of patients with genetic diseases is still in experimental stage.

The major applications of recombinant DNA technology are in:

- Medical diagnosis
- Gene therapy
- Commercial preparation of proteins and hormones
- Construction of useful organisms
- Basic applications, and several others described in this section.

A. Medical Diagnosis

The DNA-based diagnostic tests can *identify structural variants (point mutations) as well as dysfunctional and absent genes*. They are also used to identify a recessive-disease-gene in unaffected carriers, or even in healthy people to estimate disease risk. Prospective parents may be screened, and if identified as carriers of a severe, debilitating disease, may be advised not to have a child or to go for prenatal or preimplantation diagnosis.

Cloned genes are currently being used as hybridization probes for the detection of cancer and genetic diseases. Even infectious diseases can be diagnosed by specially designed probes which identify DNA of pathogen. Some representative procedures and applications of DNA-based diagnosis to investigate a possible mutation include the following (a far from exhaustive list):

Sickle Cell Mutation Analysis

Sickle cell disease is caused by a single base substitution (*A to T transversion*) in the β-chain gene of haemoglobin. This mutation can be easily recognized by a procedure based on *restriction endonuclease* cleavage and Southern blotting as it changes length of a restriction fragment. This is because the mutation occurs within a palindromic sequence recognized by a *restriction endonuclease, Mst-II*. The mutation results in obliterating this cleavage site so that it is not cleaved by ***Mst-II***; Example:

Coding strand Template strand	↓ 5′-CCTG**A**GGAG-3′ 3′-GGACTCCTC-5′ ↑	Normal gene **cleaved** by *Mst-II*
Coding strand Template strand	5′-CCTG**T**GGAG-3′ 3′-GGACACCTC-5′	Mutated gene **not cleaved** by *Mst-II*

In the Mst-II restriction digest, the size of the restriction-fragment produced from the mutated gene will be different from that produced from the normal gene (Fig. 25.10). This can be exploited for the molecular diagnosis of the disease by Southern blotting, using a probe for a sequence 5′ of the mutation.

- In the absence of mutation a 1150 base pair (1.15 kb) fragment is produced and identified by the probe.
- If the mutation is present, Mst-II is unable to cleave the mutated site and so a larger restriction fragment

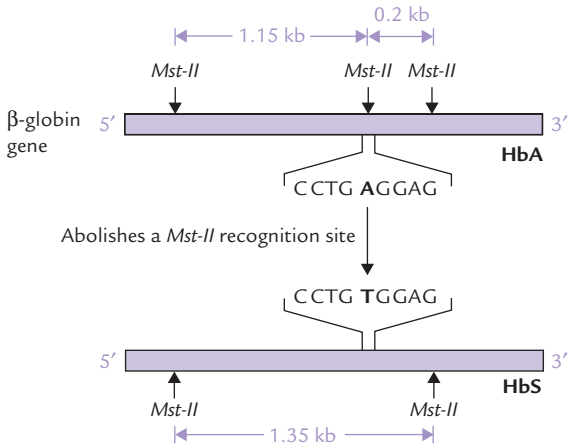


Fig. 25.10. Molecular diagnosis of sickle cell mutation by electrophoresis of an *Mst-II* restriction digest, transfer to nitrocellulose filter and analysis with a probe.

(1350 base pair) is generated. The probe sequence will therefore be present on the 1350 bp (1.35 kb) fragment.

The above method can be used for diagnosis of various other gene abnormalities where a mutation may, rather than obliterating or deleting pre-existing cleavage site, creates a new one.

Sickle cell mutation, which abolishes the cleavage site of the *restriction endonuclease*, Mst-II, can be diagnosed by Southern blotting using a probe for a sequence 5′ of the mutation.

Analysis with Allele-specific Probes

Some mutations do not obliterate (or create) a restriction site, therefore the method shown in Figure 25.10 cannot be used in their diagnosis. A different diagnostic approach, which aims to directly identify a sequence variation, is thus needed for discriminating between the normal and the mutated DNA. It involves Southern blotting and analysis with **two different probes**: one specific for the normal sequence and the other specific for the mutated sequence.

Thus, two probes to identify the two sequence variants (***alleles***) of the gene are used. Note that the probes identify sequence variants in this case, rather than identifying restriction fragments of different lengths. The probes may have to distinguish between sequences that differ in only a single base, for example, in sickle cell anaemia, and therefore, require rigorous adjustment of the conditions during annealing (base-pairing). Since annealing is favoured by elevated temperature and low ionic strength, these conditions must be present during the assay.

This method is applicable only when molecular nature of the mutation is known.

Allele-specific probes are small oligonucleotides that are used for rapid diagnosis of point mutations by probing the blotted DNA template under highly stringent conditions, e.g. low ionic concentration and high temperature.

Allelic heterogeneity: A number of different mutations in the same gene may disrupt the normal function of the protein product, and so lead to the same disease. This phenomenon, called allelic heterogeneity, is *observed in most single gene disorders*, for example β-thalassaemia. In this disorder, approximately 100 different mutations in the β-chain gene, all leading to the same disease, are known. Each individual mutation requires its own pair of allele-specific probes, which is not an easy feat. Thus, allelic heterogeneity is an important limitation to the use of allele-specific probes.

Dot-blot Analysis

It is a rapid and relatively inexpensive procedure for detection of a mutation, where crude DNA is examined directly without any prior amplification or electrophoresis. The extracted DNA is denatured, applied directly to nitrocellulose filter and then probed with allele-specific oligonucleotide probes. Evidently, the procedure is fast, relatively inexpensive, and is suitable for testing large number of people, such as during population screening.

Steps

1. Sample DNA is extracted from different individuals and denatured.
2. Every sample of the denatured DNA is applied onto two different nitrocellulose filters, each in a single dot (dotting).
3. One filter is dipped into a solution with a fluorescent-labelled probe for the normal sequence, and the other into a solution with a probe for the mutant sequence.
4. Excess probe is washed off and the bound probe is visualized under the UV lamp.
5. In normal individuals, probe for the normal sequence will bind, whereas in the individuals carrying the mutation, the probe for the mutant sequence will bind.

Uses

1. *Population screening for single mutation:* Dot blotting is less costly and nearly instantaneous as compared to more expensive Southern blotting based procedures (takes few days), and PCR based procedures (require few hours). It is more suitable for screening populations for a single mutation or polymorphism.

2. *Multiplex genetic testing*: DNA of a single individual is dotted on different filters and each of these filters is analyzed with a different probe. By this approach it is possible to detect different mutations, and so identify individuals who carry susceptibility genes for such multifactorial diseases as osteoporosis or Alzheimer's disease.

A number of newer techniques for the clinical analysis of DNA are being developed, and there is a shift from experimental DNA analysis to diagnostic services in clinical settings.

Recombinant DNA technology is likely to replace complex biochemical assays as the major biological-scientific technology in most laboratories.

B. Gene Therapy

Advances in the field of DNA recombinant technology have made it possible to ameliorate or cure a number of inherited disorders as well as diseases caused by somatic mutations. The *goal of gene therapy is to replace (or repair) the defective, disease-causing genes by normal functioning copies.* It involves the following approaches:

1. ***Gene replacement***, which involves removal of a defective sequence and its replacement with a normal one.
2. ***Gene correction***, wherein a pathological change in the nucleotide sequence is repaired.
3. ***Gene augmentation*** (most commonly used), defined as introduction of genetic material into cells without any attempt to delete or modify the endogenous genetic material.

In essence, gene therapy uses the capacity of a patient's cells to synthesize the therapeutic agent from the introduced gene.

Gene therapy may be either **somatic cell gene therapy** which targets the patient's somatic cells; or the **germline gene therapy** which targets the germline cells. However, at present only somatic gene therapy is permitted in humans. It does not attempt to change all somatic cells, but the gene is delivered only to the affected tissue. For example, in *Duchenne muscular dystrophy*, an intact dystrophin gene is targeted specifically to muscle tissue. Likewise, in *cystic fibrosis* the intact gene for the cystic fibrosis (CF) chloride channel is targeted to the lung epithelium.

Germline gene therapy is considered unethical as it carries the risk of transmitting genetic modifications to the offspring, and so not permitted.

The procedure of *gene therapy involves isolation of the gene, its incorporation into a suitable carrier, followed by its delivery into the target cell.*

Delivery Methods of Therapeutic Gene into Target Cell

Several techniques are currently available for delivery of therapeutic genes into host cell. However, none of these techniques is efficient. There are two major obstacles to effective gene delivery: first, the large therapeutic genes do not cross the biological membrane easily; and second, the foreign DNA after entering the cell is rarely integrated in chromosome. Normally, they remain extrachromosomal, do not replicate and are degraded by *nucleases.* As a result, they are diluted with successive mitotic divisions.

Some of the commonly practiced methods for gene delivery are:

Physical Methods

The gene is encapsulated in *liposome*, an artificial vesicle in which aqueous core is surrounded by lipid bilayer. The bilayer fuses with the plasma membrane of the target cell and the enclosed gene is released into the cytoplasm. Exposure to high voltage enhances cellular uptake of exogenous DNA into the cell (*electroporation*).

Receptor-mediated Endocytosis

The foreign gene is covalently linked to a ligand that is taken up by receptor mediated endocytosis. Only the cells possessing the receptor (for the ligand) are transformed by this method. An undesirable outcome is that most of the foreign DNA is directed to lysosomes for degradation. To prevent this, endosomes are disrupted with adenovirus.

Retroviruses

The most advanced technique for targeting therapeutic genes to mammalian cells involves retroviral vectors (Fig. 25.11). Retroviral vectors are artificial and disabled constructs obtained by modification of certain retroviruses:

- The key viral genes (*pol*, *gag* and *env*) are replaced by the foreign (therapeutic) gene.
- The long terminal repeat (LTR) are retained (LTR are required for integration of the viral DNA in the host chromosome). The ψ sequences, required to pack viral RNA in viral protein coat, are also retained.

To assemble virus with the above recombinant genetic information, however, poses a major problem. This is because the retroviral vector DNA cannot synthesize viral proteins (recall, it lacks *gag*, *pol* and *env* genes). To solve this problem the retroviral vector DNA is introduced into

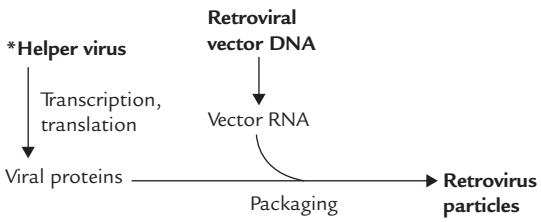
Fig. 25.11. Preparation of retroviral vector carrying recombinant genetic information.

tissue culture, infected with a **helper virus**, which contains the key viral genes (*pol*, *gag* and *env*). Therefore, within the culture cell, a cooperative action of the two types of viruses (recombinant virus and helper virus) ensures that the recombinant DNA is transcribed and the RNA packaged into the virus particle. (Note that the retrovirus particle so produced also contains *reverse transcriptase* and *integrase.*)

After the retrovirus enters the host cell, the following events follow:

- A cDNA copy of the recombinant RNA genome is produced (by *reverse transcriptase*).
- The cDNA is integrated into host chromosome (by *integrase*).
- Because of its inability to synthesize viral proteins and produce infectious viral particles, the engineered RNA genome becomes a permanent part of the host chromosome. Instead of producing viral particles, it is expressed to generate the desired protein product.

Because of the reason that viral proteins are not coded by them, the retroviruses have little or no cytotoxicity and are, therefore, called "**clean vectors**".



*The helper virus is also called packaging-defective virus: it lacks the ψ sequence required for packaging.

Use of retroviral vector, though considered a highly advanced technique, suffers from **two disadvantages**. First, the retroviral gene transfer known technically as transfection is not a particularly efficient process. In most experiments, less than 10% of the cells are transfected. Second, the retroviruses are not able to infect non-dividing cells (with sole exception of AIDS virus) and this is true for the retroviral vectors as well, which are therefore, poorly suited for gene therapy of neurons and muscle fibres.

Gene therapy mostly involves delivery of the therapeutic genes to somatic cells of the patient with the help of retroviral vectors.

DNA Viruses

DNA viruses have also been explored as potential gene transfer vectors for gene therapy. Strategies have been developed for incorporating normal genes into the adenoviral genome. *Adenoviruses*, the large (35 kb) double-stranded DNA viruses which are minor respiratory pathogens in humans, have been used to transfer genes of *cystic fibrosis transmembrane conductance regulator* (CFTR) protein to the respiratory epithelium in patients with cystic fibrosis (respiratory epithelium is the usual host cell for adenoviruses). In the adenoviral vectors, one of the viral genes is replaced by the therapeutic CFTR gene. However, the viral proteins can still be synthesized by the vector. This gives rise to cytotoxicity and immunological responses. Also, the therapeutic benefits of the adenoviral vector are transient because it rarely, if ever, incorporates its DNA into the host cell genome. However, these vectors can be applied in higher titers than the retroviral vectors, and they are also capable of infecting the non-dividing cells. Therefore, they can be effectively used for gene therapy of neurons and also muscle fibres.

Achievements and Limitations of Gene Therapy

In addition to the conditions discussed above, gene therapy has been attempted in various other disorders.

Haemophilia B (Deficiency of Factor IX)

A clotting disorder, is treated by retroviral transfer of factor IX gene into hepatocyte. Although the intracellular level of factor IX reaches only 0.1% of normal after this treatment, the clinical manifestations of the defective gene are mitigated. Haemophilia A, caused by deficiency of factor VII is also treated by retroviral transfer of this factor in patient's fibroblast, followed by implantation of the latter under the patient's skin.

Severe Combined Immunodeficiency (SCID)

The first success for gene therapy was reported in the year 2000 by a French medical team, who succeeded in curing SCID, a disease caused by deficiency of *adenosine deaminase* (ADA). The molecular defect is a mutation located on chromosome 20. The ADA deficiency results in accumulation of adenosine, deoxyadenosine and dATP. All the three compounds are toxic to lymphocytes. The immune responses, both cell-mediated and humoral, are

suppressed and so the patient is susceptible to repeated infections. Treatment involves isolation of the ADA genes from normal healthy cells and their introduction into retroviruses. These retroviruses are then used to introduce the gene into lymphocytes of the patient *in vitro*. The cells so transformed are introduced back into the patient. These cells produce ADA normally. Furthermore, no side effects have been observed. Therefore, this line of treatment holds promise for the future. However, one of the problems is that the introduced gene may not express at the desired level in the new environment.

Non-genetic Conditions

Some non-genetic conditions such as *rheumatoid arthritis* and *cancers* are also treated by gene therapy. Rheumatoid arthritis is an inflammatory process mediated by the cytokine interleukin-I (IL-1). The transfected gene encodes a receptor-blocking protein that prevents the interaction of IL-1 with its receptor. The vector carrying the gene is injected directly into the inflammed joint. For the treatment of cancers, the tumour cells are transfected with cytokines such as IL-2 and tumour necrosis factor (TNF). The aim is to stimulate immune response against tumour cells.

Other diseases under study for somatic cell gene therapy are factor *VIII deficiency, HGPRT deficiency, pyrimidine nucleotide phosphorylase deficiency*, etc.

Germline Gene Therapy

Studies on germline gene therapy have been conducted on *transgenic mice*. The fertilized mouse eggs are harvested and injected with desired DNA molecules. These are then reimplanted in the pseudo-pregnant female mouse. The technique has been used in murine gene disorders, such as growth hormone deficiency and γ-globin deficiency. Though beneficial effects have been observed in animals presently, this therapy appears to have no application in human genetic disorders (see transgenesis).

Current Status of Gene Therapy

Gene therapy is still in the experimental stage and at present only a few diseases are treated by it. There are several reasons for its limited use, such as lack of consistent results and lack of ideal vector. *Most non-viral vectors lack precise targeting ability, whereas use of viral vectors carries risk of carcinogenesis.*

However, in spite of several technical difficulties, recent advances in molecular and cellular biology raise hope that gene therapy will soon start playing an increasingly important role in clinical practice and have applications in several fields of medicine.

A new approach for treatment of hereditary disorders aims at increasing expression of an endogenous gene. For example, hydroxyurea enhances expression of the γ-globin gene. In patients with sickle cell anaemia, this approach increases synthesis of nonsickling haemoglobin. *Antisense technology* is also emerging as an effective mode of treatment (refer to Box 25.7 for details).

C. Commercial Preparation of Proteins and Hormones

Many human proteins can be used as therapeutic agents. These include hormones such as insulin and growth hormones; blood-clotting and fibrinolytic proteins such as clotting factors VIII and IX and tissue type plasminogen activator; and humoral mediators of immune response such as the interferon and interleukins. Traditionally, these proteins have been obtained from human tissue. But this method suffers from a serious drawback in that transmission of serious diseases may occur. Many recipients have been reported to have contracted AIDS virus after treatment with contaminated proteins. Moreover, the yield from the human tissue is very poor and the price of production is very high.

The cloning strategies described above are sometimes not useful: they may be able to propagate DNA, but effective transcription and translation of the latter do not occur. Consequently, the genes are not expressed. The use of expression vectors, is therefore, favoured.

D. Construction of Useful Organisms

Some bacteria that produce increased amounts of industrial chemicals such as methanol, ether, and butanol from biological wastes have been genetically engineered. Marine bacteria capable of metabolizing petroleum have been constructed. These are useful in clearing oil spills. The future possibilities of these methods are enormous and highly promising. For example, attempts are being made to introduce the nitrogen fixation system into non-nitrogen fixing bacteria and plants. If successful, these measures will reduce the dependence on nitrogen fertilizers.

E. Basic Applications

Site-directed Mutagenesis

This is a technique that can introduce controlled alterations of selected regions of a DNA molecule. It may involve the insertion or deletion of selected DNA sequences, or replacement of a specific nucleotide with a different base. It can also be accomplished by replacing

BOX 25.7

Antisense Technology

Gene therapy primarily aims to treat the diseases caused by lack of a normal gene by introducing the therapeutic gene into cells. A different approach is used in some other diseases that are caused by overexpression of an undesirable gene. In many types of cancers for examples, the malignant cell grows out of control because of overexpression of growth promoting genes, the oncogenes. If expression of oncogenes is brought under control, it may serve as an effective mode of treatment. Also, viral diseases can be treated by agents that check the expression of essential viral genes. Such approaches form the basis of antisense technology which aims to block expression of undesirable genes. It is based on the use of oligonucleotides that are complementary to the base sequence of the undesirable mRNA. The oligonucleotides hybridize with the mRNA and thereby block translation. Both DNA and RNA nucleotides have been used as antisense agents. To prevent non-specific hybridization with non-target sequences, the antisense must have a length of at least 16–18 nucleotides.

A major problem in the use of the antisense oligonucleotides is that they show poor cellular uptake. Moreover, they are rapidly degraded by the *cellular nucleases*. To overcome the latter problem, synthetic oligonucleotides have been constructed in which the phosphodiester bonds are replaced by *nuclease* resistant linkages. But even these resistant analogues are not easily taken up by their target cells, and therefore, must be used at high concentration—an economically unfavourable situation considering, high cost of the antisense agents.

Currently, antisense technology is at an experimental stage but it is expected to play important role in future. Severe viral diseases, such as AIDS will probably be the first one for which antisense therapies will be used.

Antisense genes: Use of antisense genes combines the principle of gene therapy and antisense technology. A synthetic gene is introduced into the cell whose RNA transcript is complementary to the mRNA of the undesirable gene. The two hybridize so that the RNA transcript of the undesirable gene is not translated.

the appropriate section of the isolated gene with a synthetic oligonucleotide whose base sequence is such that it codes for the altered amino acid. The altered activity of the mutant protein helps us to know the function of the particular amino acid in a protein.

Knockout Mice

These are useful animal models developed by inducing gene disruptions. They are proving to be very useful animal models *for the study of human genetic diseases.* They also help to ascertain the biological significance of proteins by observing the abnormalities (if any) when the functional protein is missing.

F. Analytic Techniques

The recombinant DNA technology is used in a number of analytic techniques, such as polymerase chain reaction, restriction fragment length polymorphism (RFLP), DNA typing, Northern blotting, Southern blotting and Western blotting.

G. Transgenesis

It refers to the process of introducing exogenous genes into germline of an organism, which changes its phenotypic characteristics and which would be passed on to the successive generations. Some success has been achieved in the transgenesis experiments in mice. The rat growth hormone gene was introduced into the germline of mice, and the resulting transgenic mice grew to more than twice the normal size. Because the gene of interest must be inherited in the germ line, it is essential to use vectors (containing the gene of interest) with appropriate regulatory elements for introducing the gene into the nucleus of the fertilized eggs. The eggs are then transplanted into the uterus of receptive females for the development of the potential transgenic off-springs.

H. Agricultural and Livestock Industries

The process of transgenesis is being utilized in both the plant and livestock industries, with an aim to develop robust plants and animals that are more resistant to diseases and infections. Some transgenic animals are developed to obtain high levels of expression of therapeutically important proteins. For example, a desired human protein can be produced in sheep by attaching the isolated human gene to the signaling part of the milk-protein-gene. This results in secretion of that (desired) protein into milk. This allows us to obtain large amounts of human protein from the milk of transgenic animals.

BOX 25.8

DNA Vaccine

This refers to plasmid DNA carrying a cloned gene. The gene is expressed after the plasmid DNA is injected into tissues. The gene encoding a viral protein may be thus be used to prevent viral disease. In experiments with mice, the gene coding for the core protein of the influenza virus was cloned into a plasmid vector and injected into mice. The mice developed immunity not only to the strain of influenza virus from which the gene was delivered but also from other strains of influenza virus. DNA vaccines have a definite advantage over the conventional vaccines that use inactivated or attenuated viruses, in that the viral proteins expressed intracellularly stimulate both humoral and cell-mediated immunity. Fragments of the synthesized viral protein are carried to the cell's surface where they stimulate CD8+ cytotoxic T cells and, thereby, cell-mediated immunity.

More applications of recombinant DNA technology are described later in this chapter along with uses of *polymerase* chain reaction and RFLP.

I. Enhancement Engineering

Genes targeted to germline can permanently change genetic makeup of a species. For example, humans lack the enzyme *gulonolactone oxidase* required for synthesis of vitamin C. The transfer of a gene for this enzyme would restore our ability to synthesize vitamin C and push scurvy into oblivion. This is an example of enhancement engineering. In contrast to gene therapy, which is aimed at repair/replacement of the defective genes, the goal of enhancement engineering is to improve the genetic constitution in general.

J. Other Uses

Several vaccines are prepared using recombinant DNA technology (refer to Box 25.8 to know the process of DNA vaccine preparation). The first vaccine prepared by this technology was hepatitis B vaccine. The engineered proteins are used as food additives to enhance the nutritional value of food.

VII. Restriction Fragment Length Polymorphism

Restriction endonuclease cleaves the DNA only at some specific recognition sequence, the **restriction site**. In case a restriction site is disrupted, either by a pathologic change in the DNA sequence resulting in a disease (a mutation), or a naturally occurring variation in the DNA sequence (polymorphism), it may create new sites or obliterate the existing sites. This may result in fragments of different lengths following digestion by restriction enzyme—*restriction fragment length polymorphism* (RFLP). Such RFLP can be used either to identify disease-causing mutations or to study variations in non-coding DNA, which can be used in the study of genetic linkage.

A. Origin of RFLP

The following sequence variations—mutations or polymorphism—can change the pattern of restriction-sites to generate RFLP:

1. ***Base substitution:*** A single base substitution may create a new restriction site or delete a pre-existing one.
2. ***Insertion or deletion mutations*:** Bases are added to or removed from a DNA segment between two restriction sites. This changes length of DNA between the two cleavage sites, resulting in the polymorphism.
3. ***Variable number of tandem repeats*:** Tandem repeats are short sequences of DNA—between two to a dozen base pairs—repeated over and over in tandem arrangement (next to each other) in the DNA molecule. There is much variation between individuals in number of repeats of these short sequences. This hypervariability is referred to as *variable number of tandem repeats* (*VNTR*); for example, a given sequence may be repeated less than 5 times in some people and more than 20 times in others.

Mutation or polymorphism may cause different patterns of restriction sites, generating DNA fragments of different lengths, a phenomenon termed restriction fragment length polymorphism.

B. Clinical Applications

1. ***Diagnosis of genetic defects:*** RFLP can be used to identify a disease-causing mutation because of a single

point mutation, creating or abolishing a restriction site. This has been discussed earlier with reference to sickle cell anaemia (Fig. 25.10) A more recent use is to map and isolate genes for hereditary diseases in which the gene responsible (for the disease) is unknown.

2. ***Linkage studies:*** If the polymorphism is located close to a defective gene on the same chromosome, the two are transmitted together from one generation to the next (simply for being on the same DNA molecule). Therefore, inheritance of the defective gene can be traced by tracing the inheritance of the associated restriction fragments. The advantage of RFLPs is that they can be used even if the sequence of the defective gene is unknown. Moreover, allelic heterogeneity, probably the most serious obstacle to DNA-based diagnosis, does not affect the use of RFLPs.
3. ***Study of genetic individuality:*** RFLPs, which are inherited and conserved in a Mendelian manner, are highly conserved in different individuals and therefore, allow study of genetic individuality at molecular level. The fact that probability of obtaining identical pattern (of restriction fragments) in two randomly selected individuals is only 1 in 10^{19} makes this technique applicable to problems of genetic individuality.
4. ***Detection of changes in DNA sequence:*** RFLPs can also be used to detect larger pathological changes in the DNA sequence, either deletions or duplications. In case of gene deletion, a restriction site may be abolished, leading to disappearance of a fragment on Southern blot. In case of gene duplication a new gene may be formed having a different pattern of restriction sites, which allows its detection.
5. ***Others:*** Value of RFLP in detecting the presence of the disease allele in heterozygotes (carriers) is obvious. It is also useful as a diagnostic test for detecting the individuals in which the disease has not yet manifested.

VIII. Polymerase Chain Reaction (PCR)

It is a technique for adjective amplifying a single template DNA. It permits generation of up to 10^9 copies of the target nucleotide sequence in a matter of few hours. PCR can be used to amplify DNA sequences from any type of source: bacterial, viral, plant or animal.

A. Requirements for PCR

A Standard PCR requires the following:

1. ***DNA template***—the piece of **target DNA** that is to be amplified.
2. ***Amplimers***—small oligonucleotide **primers** that would hybridize with the complementary DNA sequences upstream of the template and act as a starting point for the synthesis of the newly amplified DNA strands.
3. ***Polymerase enzyme***—an enzyme that would catalyze the formation of DNA during the amplification reaction. A thermostable ***DNA polymerase*** of bacterial origin is most commonly used.
4. ***dNTPs*** (deoxynucleoside triphosphates)—these are the substrate molecules for the new DNA synthesis by the *polymerase.*

B. Steps

The average PCR involves about 30–35 cycles of reactions that provide sufficient DNA, about 10^9 copies of the original DNA. Each cycle comprises three steps: **denaturation, annealing** and **elongation** (Fig. 25.12).

1. ***Denaturation:*** The isolated duplex DNA containing the target sequence is heated to separate the two strands. Generally, heating at 94°C for one minute is enough to ensure that the template (as also the primers) are single stranded.
2. ***Annealing of primers:*** The mixture is cooled to allow formation of heteroduplexes of primer and template. The temperature for this (45°C) is based on the Tm of the expected duplex (usually 3°–5°C below Tm). *The primer sequence is complementary to a short nucleotide sequence in the regions flanking the target sequence.* A large excess of 10^8 molar excess primers must be added to ensure that they promptly anneal with the flanking sequences.
3. ***Elongation:*** Deoxyribonucleoside triphosphate precursors (dATP, dGTP, dCTP and dTTP) are then incorporated into the growing DNA chain by the *polymerase* enzyme. The process of PCR has been greatly facilitated by the discovery of **thermostable enzymes**, which can withstand the heat required during the denaturation step and still function as *polymerase*. The most commonly used enzyme, *Taq polymerase* is isolated from *Thermus aquaticus*, a thermophile bacterium that was isolated originally from a hot spring in Yellowstone National Park. The other such enzyme, *Pfu I*, also functions effectively at higher temperatures and can survive repeated heating to 94°C.

The polymerase chain reaction is a novel in vitro method of copying a single template DNA in vitro. PCR is used widely in all aspects of biomedical research, in particular the human genetics.

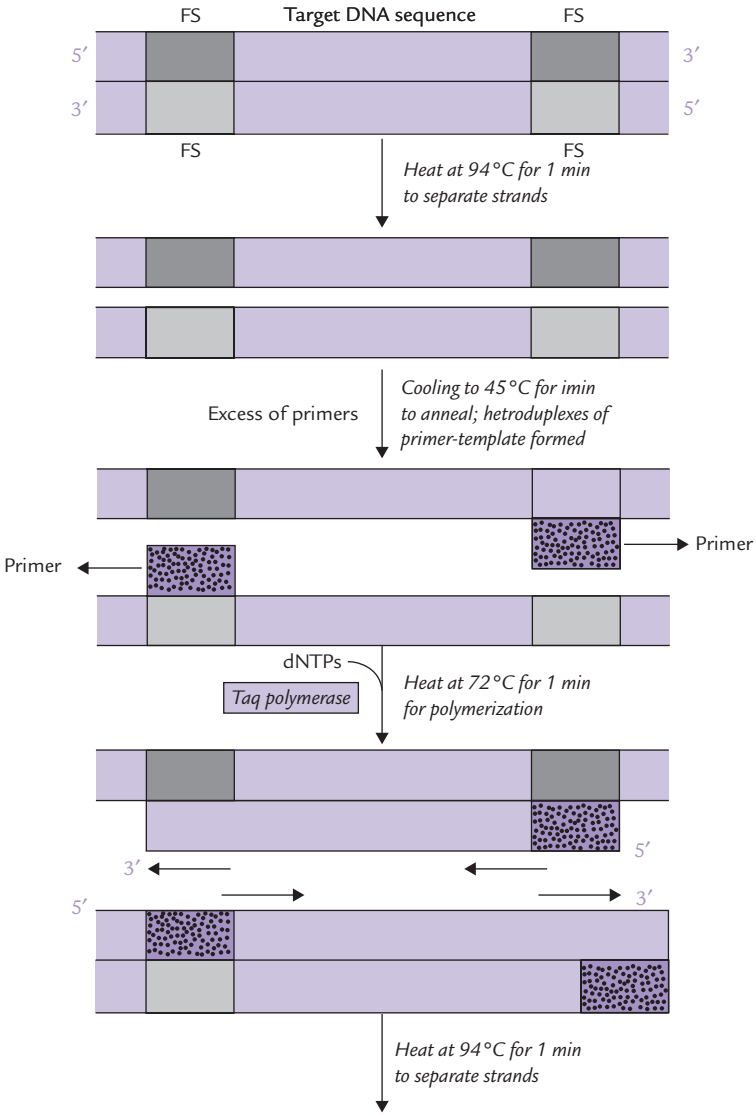


Fig. 25.12. The polymerase chain reaction for amplification of DNA sequences of interest. dNTPs (deoxynucleoside triphosphates) = Substrate molecules for new DNA synthesis. ▭ and ▭ = Flanking sequences (FS), ▩ = Primer.

These three steps—denaturation, primer annealing and primer extension—can be carried out repetitively just by changing temperature of the reaction mixture in a suitable **thermocycling device**.

- When temperature is raised to 94°–100°C, *denaturation* occurs.
- Lowering the temperature to about 45°C optimizes *annealing*.
- A temperature of 72°C permits *polymerization*. The thermostable nature of the *Taq I polymerase* makes it feasible to carry out PCR in a closed container. Further addition of reagent is not required after the first cycle.

PCR is a simple and quick technique to amplify target DNA, but it requires prior knowledge of the nucleotide sequences in the regions flanking the target sequences. If these flanking sequences are known, then it is possible to construct the well-designed, specific primers, which will only amplify the DNA of interest. Primers that are non-specific, will amplify sequences other than the target DNA to produce haphazard results. If, for example, the 3′-primer contains sequences that contain tandem repeats of a base, the specificity of the amplification may be reduced. This is because such repetitive sequences are widely distributed throughout the genome, and so unwanted sequences may be amplified.

Reverse transcriptase PCR: cDNA is generated from mRNA template by reverse transcription. The cDNA is then amplified by PCR to produce large amount of cDNA. This overall process of producing cDNA from mRNA is called *reverse transcriptase PCR* (***RT-PCR***). By this technique, a huge amount of cDNA may be generated, which can be used to synthesize useful proteins.

C. PCR-based DNA Cloning versus Cell-based Cloning

Advantages

PCR has three principal advantages over cell-based cloning:

1. *Cost and time*: Compared to cell-based cloning, PCR is less expensive and quicker. The denaturation step takes 1 minute, the annealing step takes 3–5 minutes and elongation step 1–5 minutes. Thus, *30 cycles would take 3–6 hours to cause 1 lakh-fold amplification*. The cell-based cloning can be expected to yield such results in several days or weeks.
2. *Sensitivity*: PCR can amplify a DNA template to usable amounts even if a few nanograms of target DNA is available. The primers are added to the DNA in about 10^8 molar excess, together with a generous amount of the enzyme and substrates. After that, the only additional requirement is a thermal cycling device (heat to denature and cool for annealing), which causes exponential increase in DNA.
3. *Robustness*: PCR is able to amplify DNA that is often badly degraded, or is present in inaccessible sites, e.g. formalin-fixed tissue.

Disadvantages

Following are the disadvantages of PCR-based over cell-based cloning:

1. ***Inaccuracy:*** The *Taq polymerase* has no proofreading activity, so it has an error rate on the order of one base mismatch every few thousand base pairs.
2. ***Small target DNA amplification:*** PCR effectively amplifies small target sequences of 200–1000 bases, but cannot reliably amplify larger sequences. However, use of ***long PCR methods*** allows sequences up to 20 kb or more to be amplified.
3. ***Prior information of flanking sequences:*** This is essential for synthesizing flanking primers, and may require prior cell-based cloning to derive the flanking nucleotide information.

D. Applications of PCR

The major use of PCR is to provide massive quantity of *DNA for prenatal* ***diagnosis, pre-implantation diagnosis, diagnosis of infectious diseases, forensic medicine studies*** among others.

PCR in Prenatal Diagnosis

PCR can be used for the prenatal diagnosis of genetic defects early in pregnancy. Sample DNA is extracted from the fetal cells, obtained by either chronic villus sampling at 9th week of gestation or by aminocentesis at approximately 15th week, and is amplified (by PCR). The PCR products are electrophoresced, denatured, transferred on nitro-cellulose filter and then probed with complementary nucleic acid sequence for specific detection of a gene abnormality.

For example, mutation in about 70% of the cases of **cystic fibrosis** consists of deletion of three successive base pairs that code for the amino acid phenylalanine in position 508 of the polypeptide chain (Phe^{508} mutation). A short section of DNA around the site of mutation is amplified. Since the mutated sequence yields a PCR product three nucleotides shorter than normal, the two products can be easily separated by polyacrylamide gel electrophoresis. This method can be employed for all small insertion/deletion mutations.

In case a crippling mutation is detected by this technique, selected abortion of the affected fetus may be carried out. However, an ethical dilemma is associated with this approach.

PCR in Pre-implantation Diagnosis

In view of the above ethical considerations, the DNA sample is obtained from an alternate source instead of the growing fetus. *In vitro fertilization* is carried out and the embryo is grown in a test tube to the 8- or 16-cell stage, when a single cell is removed for extracting its DNA. Removal of a single cell does not impair further growth of the embryo.

Further analysis is carried out by electrophoretic separation of PCR products, detection of genetic abnormality, if any, by radioactive probes. Typically, 5–8 ova are fertilized in vitro at the same time, and all developing embryos are subjected to the test. Only disease-free embryos are re-implanted.

If only small amount of DNA is available, for example, in prenatal and pre-implantation diagnosis, the DNA of interest has to be amplified by PCR, before it can be analyzed.

A major limitation of the above method is that the quantity of DNA from a single cell is too small to be amplified reliably even by "standard PCR". A more sensitive method—**PCR with nested primers**, is required to accomplish this feat. A relatively large section of the target DNA is initially amplified through a 20- to 30-cycle PCR to generate approximately 10^6 molecules. These molecules are then subjected to a second round of PCR (20–30 cycles) with a new pair of primers, which yields more than 10^{10} molecules of the PCR products.

BOX 25.9

DNA Fingerprinting

Traditional method for the identification of an individual has been fingerprinting. A more advanced identification technique, based on DNA typing of microsatellite repeat is DNA fingerprinting. It makes use of the fact that a large *number of repeat sequence* have been identified in the human genome, which are variable from individual to individual in the number of repeats in the sequence. The method involves choosing primer sites on either side of the selected repeat sequences and using PCR to amplify the latter. Running the product on an electrophoretic gel yields a band of a size depending on the number of repeats in the amplified segment. By selecting a number of such DNA segments to amplify, a pattern of bands highly characteristic of an individual is obtained.

Application to Forensic Medicine

PCR amplification serves as a potent tool for investigating cases of assault or rape, and answering simple questions such as is this man father of this child? Because PCR requires only small quantities of DNA, *a blood stain* or *drop of semen* provides sufficient DNA for **DNA fingerprinting**, a technique that relies on a high degree of polymorphism of microsatellite repeats, discussed earlier (Box 25.9).

Tandem repeats are used for DNA fingerprinting.

Linkage Studies of Disease Trait

The high degree of polymorphism makes microsatellites very useful for the study of genetic linkage, particularly of complex disorders such as diabetes mellitus. Study of the PCR-amplified microsatellites has successfully established the linkage of disease traits, for example type 1 diabetes mellitus, to certain areas of the human genome.

Diagnosis of Infectious Diseases

Bacterial, viral or protozoal DNA is amplified for diagnosis of diseases before they manifest clinically. As few as 10 tubercle bacilli (the bacteria causing tuberculosis) per million human cells, for example, can be detected using PCR amplification.

Archaeology and Paleontology

PCR is opening new fields of molecular archaeology and molecular paleontology because of its potential in cloning DNA fragments from mummies and the remains of extinct animals. Amplifications of the rare, surviving fragments of the ancient DNA samples is providing useful insight into evolution. Comparisons of DNA sequences of the extinct and living species can reveal missing links in evolution.

Others

These include chromosomal walking, introduction of mutations *in vitro* to test their effect in biological systems and amplification for DNA sequencing.

E. PCR-based Methods in Advanced DNA Research

Several methods used in DNA research are based on use of PCR. Some of the more important ones are given here.

Restriction Site Polymorphism (RSP)

The restriction fragments obtained after digestion with appropriate restriction enzymes are PCR amplified prior to their detection. Thus, RSP is considered a *small scale equivalent of RFLP*.

Mutation Detection by Allele-Specific PCR

Standard PCR uses two primers to amplify both strands of the DNA, and relies on the fact that each primer will amplify the target strand. However, PCR can be performed using allele-specific primers, which will amplify one allele of the target gene. Their use is based on the principle that *amplification of target DNA relies upon the presence of complete hybridization of primer and template:* polymerization will be most effective when the primer and target are 100% complementary. If, in particular, the base at the 3′ end of the primer does not match the base in its target strand, the new DNA synthesis is severely impaired.

Application of allele-specific primers to detect single base substitution in sickle cell anaemia, is discussed here.

Allele-specific primers are constructed whose 3′-terminal base corresponds to the site of mutation. One of these primers (the s-primer) contains a sequence, the extreme 3′ base of which recognizes the (HbS) mutation. Therefore, this primer will hybridize only with the mutant (sickle cell) DNA, and amplify it. The extreme 3′ base of the other primer (the a-primer) detects normal allele, and therefore it amplifies only the normal allele.

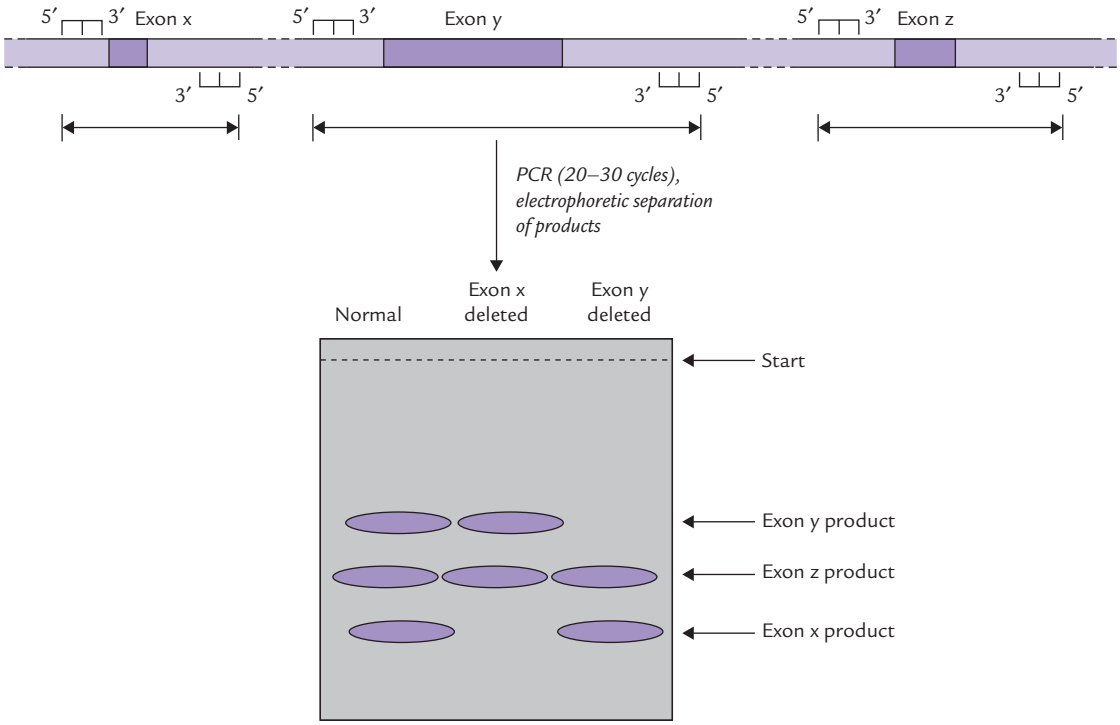


Fig. 25.13. The principle of deletion scanning with the PCR. Note that the primer pairs are designed to generate PCR products of different lengths that can be separated from each other by polyacrylamide gel electrophoresis (▬ = exons; ⌊⌋ = primer).

5′ CACCTGACTCCTGT 3′

← **s-primer** (recognizes the sickle cell allele; amplifies mutated gene).

5′ CACCTGACTCCTGA 3′

← **a-primer** (recognizes the normal allele; amplifies normal gene).

PCR is performed separately with the two allele-specific primers, the PCR products electrophoresed and then identified.

Detection of Deletions in Genes Causing Diseases

Some genetic diseases are caused by large deletions of DNA segments that remove one or several exons, or occasionally whole of the gene. PCR is used to identify such deletions by a technique termed **deletion scanning**, which is performed using primers that generate PCR products of different sizes (Fig. 25.13). Its application in the diagnosis of Duchenne muscular dystrophy (DMD) is discussed here.

DMD is a severe X-linked disorder in which the deletions affect the gene for the muscle protein, dystrophin. It is the largest of all human genes, with 79 exons scattered over a total length of more than 2 million base pairs. One or several exons are deleted in DMD; but some exons are more likely to be deleted than others and are termed *deletion prone exons*. Specific amplification of some of these exons using primers, which generate PCR products of different sizes for each exon, can be used to screen quickly for the presence of a disease-causing deletion in an affected individual (or determine if the fetus of a carrier possesses the deletion). Such PCR is carried out in a single reaction using a single template and 5–10 different primer pairs, each generating a separate amplification product. This form of PCR is called **multiplex PCR.**

Detection of Microsatellite Repeats

Microsatellite repeats are the short nucleotide sequences that are present in non-coding regions, either intergenic or within introns. A typical microsatellite contains tandem repeats of a simple dinucleotide, trinucleotide or tetranucleotide (see Chapter 24). Amplification of the DNA including the microsatellite repeat is performed by PCR using a radio-labelled primer and the products are then size-fractionated onto a polyacrylamide gel to aid resolution.

One important feature of these microsatellite repeats is that they are highly pleomorphic, i.e. the number of tandem repeats in a particular microsatellite may vary among different individuals. For example, a sequence may be repeated only 5 or 6 times in some people and more than 20 times in others. Indeed, we can expect different repeat numbers not only in different people but also on the two homologous chromosomes that carry

the repeat in the diploid cells. As a result of this high degree of polymorphism, *microsatellite repeats are of great value* in the study of genetic linkage and the forensic application of DNA fingerprinting.

IX. Human Genome Project: Current Status and Future Goals

Human Genome Project (HGP) is a large multicentric, international collaborative venture, the main aim of which is to obtain the DNA sequence of the entire human genome. It is a gigantic effort considering the fact that human genome (with 33,000–44,000 genes; 3.1×10^9 base pairs) is 25 times as large as any previously sequenced genome and eight times as large as sum of all such genomes. Ever since its initiation a decade earlier, it has been marked by relentless coordinated drive involving at least 20 groups from USA, UK, France, Japan, Germany and China, thus reflecting the scientific community at its best. The main **objectives** set out early in history of the project include:

1. To obtain complete sequence of pooled DNA extracted from cells donated by several anonymous donors, so as to determine the sequence of DNA in each chromosome.
2. To construct genetic map for facilitating genetic linkage studies.
3. To discover all human genes to allow further study of human genetic diseases.
4. To develop simplified and automated technology for DNA sequencing process.

HGP was initially a public research effort, but subsequently a private company, **Celera Genomics** (Rockville, MD) undertook the sequencing task. In a joint announcement by the US Government and Celera Genomics, a successful completion of the project was declared in June, 2000. About 90% of the gene rich portion of the genome (euchromatin) was declared to be sequenced and assembled; the latter term used to describe the process of using computers to join up bits of sequences into a larger whole. There were, however, gaps and errors in the sequence and the two groups (Cellera Genomics & US Government) agreed to continue cooperation to complete the sequencing task as soon as possible.

By independent methods the two groups declared the following **results**:

- The total number of genes in the human genome is in the range of 33,000 to 44,000.
- Approximately 75% of these genes have the same DNA sequence in all individuals, except for those with rare mutations.
- Any two human genomes are approximately 99.9% identical in sequence. The apparently insignificant difference of 0.1% has a highly significant effect on personality, behaviour, intelligence, disease susceptibility and other traits.

Building upon these achievements, the **goals for the HGP have been expanded**, as discussed below (Source: Collins et al., New goals for the US HGP: *Science*, 1998; 282: 683):

1. Establish the complete human genome sequence and to make it freely accessible.
2. Improve the sequencing technology by developing new and more effective methods.
3. Analyze sequence variations in the human genome, such as **single nucleotide polymorphisms** (SNPs) and other DNA sequence variations.
4. Develop technology for **functional genomics**. It includes: development of additional cDNA resources and technology for detailed analysis of gene expression; comprehensive study of functions of non-protein coding sequences; and encourage development of technology for global protein analysis.
5. Study **comparative genomics** by completing the genome sequence of some model organisms (e.g. *Drosophilia*, mouse, etc.) which would enhance our understanding of the human genome.
6. Consider **ethical, legal** and **social implications** of the vastly expanding knowledge base. It is anticipated that clash of this new and advanced knowledge with the pre-existing ethical and philosophical perspectives may result in undesirable consequences, which have to be taken care of.
7. **Develop bioinformatics** and **computational biology**, to impart advanced training to young scientists and encourage establishment of academic careers in genomic research.

The **medical implications** of the huge amount of genetic information obtained from the HGP are tremendous. It would serve as resource for identification of the human disease genes. For example, the oncogenic sequence changes in cancer cells can be directly identified by comparing cancer genome sequences against draft genome. Advancement in biotechnology with regard to development of useful genes would expand the scope of gene therapy and open new ways of combating disease (Collins FS. Positional cloning moves from peridiitional to traditional. *Nature Genetics* 1998; 9: 347).

It may even initiate new fields such as **pharmacogenomics** which would individualize therapies depending on genetic make up of the patient.

Exercises

Essay type questions

1. What are the steps involved in isolating a particular clone from a genomic library?
2. How does a restriction enzyme differ from a pancreatic DNase? What is meant by blunt ends and sticky ends as applied to DNA molecules?
3. What is meant by genomic clone and cDNA clone, and how do they differ from each other? Give one example each to give their specific applications in genetic engineering.
4. Enumerate the importance and principle of the polymerase chain reaction, specifying the requirements for it to be used.
5. What is Human Genome Project? State its implications in the study of health and disease.
6. Write briefly about *reverse transcriptase*, its biological significance and usefulness in recombinant DNA technology.

Write short notes on

1. Expression vectors
2. Cosmids
3. Restriction analysis
4. DNA-fingerprinting
5. Complementary DNA
6. Antisense therapy
7. *Reverse transcriptase* PCR
8. DNA vaccines
9. RFLP
10. Western blotting
11. Dot-blot analysis
12. Allelic heterogeneity
13. Knockout mice
14. Transgenesis
15. Enhancement engineering
16. Site-directed mutagenesis

CLINICAL CASE

CASE 25.1 Tissue *in situ* hybridization in diagnosis of a metastasis

A 47-year-old man has hypertension. His serum concentration of aldosterone, cortisol and 18-hydroxycortisol were found to be elevated. Liver imaging revealed the presence of a circular lesion of 2 cm diameter, which probably was a metastasis. Further tests suggested presence of adrenal carcinoma. To confirm whether the hepatic lesion was outcome of a metastatic spread from the adrenal cortex, or was it a primary growth of liver, a biopsy obtained from the lesion was studied by in situ hybridization with riboprobes. Indeed, the lesion was found to be a metastasis.

Q.1. How does in situ hybridization help us to arrive at the above diagnosis?

Q.2. Enumerate some other uses of this technique.

CHAPTER

GASTROINTESTINAL PATHOBIOCHEMISTRY

26

Digestion, absorption and subsequent assimilation of dietary components are responsible for the transformation of *food to flesh*. This process is complex and depends on integrated activities of various organs, mainly those of the alimentary tract.

A battery of enzymes and cofactors present in various gastrointestinal secretions are essential for digestion. Coordinated action of these enzymes causes disintegration of the large polymers that constitute bulk of the ingested nutrients. The process occurs in a stepwise and controlled fashion which ensures degradation of macronutrients like proteins, lipids and carbohydrates to their building block precursors. This is essential because only small molecules can be absorbed from the gastrointestinal wall.

In this chapter, mode of action of various proteolytic, lipolytic and other enzymes of digestion and mechanism of absorption of various products of digestion are described. This is followed by description of biochemical tests for assessing gastrointestinal and hepatic functions.

After going through this chapter, the student should be able to understand:

- **Digestion and absorption** of major nutrients: carbohydrates, proteins and lipids.
- **Organ function tests:** pancreatic function tests, intestinal function tests and liver function tests.

I. Digestion and Absorption

Most nutrients are large polymers. Their disintegration requires coordinated action of a number of enzymes and associated cofactors. Approximately 30 g of digestive enzymes are secreted per day. The whole process of digestion consists of hydrolytic cleavage reactions catalyzed by these enzymes, in which macromolecular nutrients are hydrolyzed to their monomeric building blocks.

The extent to which various nutrients are hydrolyzed, and then utilized varies. Utilization of starch and glycogen is nearly complete: these molecules are completely degraded to their monomeric unit, glucose, which is then readily absorbed. On the other extreme are the indigestible compounds like dietary fibres, which remain unutilized and are excreted as such. The *major processes involved in digestion and absorption* are:

1. **Mechanical homogenization** of food and mixing of the ingested solids with gastrointestinal secretions.
2. **Secretion of**
   - digestive *enzymes* that hydrolyze dietary polymers to oligomers and dimers.
   - *electrolytes, acids or bases* to provide an environment for optimum digestion.
   - *bile acids* to emulsify lipids.
3. **Hydrolysis** of the oligomers and dimers by intestinal surface enzymes.

Along the length of the gut, various fluids, electrolytes, acids or bases, and enzymes are added to aid in mixing, hydration and digestion of food.

A. Carbohydrates

Carbohydrates provide a major share of the daily caloric requirement. Dietary carbohydrates consist of digestible compounds such as starch, glycogen, lactose and sucrose.

In addition, certain indigestible **fibres** of plant origin, such as *cellulose, hemi-cellulose, pentosans*, and *inulin* are present in normal diet, which cannot be degraded by digestive enzymes of non-ruminants.

Since starch and glycogen provide bulk of the dietary carbohydrates, they will be considered first in some detail. **Starch** is a plant polysaccharide, consisting of linear chains of glucose molecules linked by α(1→4) glycosidic linkages, and branch points linked by α(1→6) glycosidic linkages (Chapter 2). Glycogen is a polysaccharide of animal origin, having a similar structure as starch, but is more extensively branched. Digestion of these polysaccharides begins in mouth in humans by action of the *salivary* enzyme, ***α-amylase***. This enzyme hydrolyzes the α(1→4) bonds to release smaller oligosaccharide fragments. However, it gets little time to act because as soon as the food reaches the stomach, it is acidified and the acidic pH stops the action of this enzyme (optimum pH for salivary *amylase* is 6.9). No further digestion of carbohydrates occurs in the stomach.

When the acidified gastric contents reach duodenum, the low pH stimulates release of *secretin,* an intestinal hormone, which helps elevation of pH to neutral range. It does so by stimulating bicarbonate release from the exocrine pancreas. Neutral pH is optimum for the action of the ***pancreatic amylase***, the principal enzyme for digestion of starch (and glycogen). Like *salivary amylase*, it is also a *dextrinogenic endosaccharidase*, specific for hydrolyzing the α(1→4) linkages. The term endo implies that it is capable of hydrolyzing those bonds that lie towards the core of the starch molecule. This distinguishes *amylases* (salivary and pancreatic) from the *exosaccharidases* of plant origin, which can act only on the terminal α(1→4) linkages. *Amylases cannot hydrolyze the branch linkages, i.e. α(1→6) linkages, being specific only for the α(1→4) linkages*.

Extensive action of *α-amylase* in intestine cleaves the starch molecule into smaller fragments such as maltose, maltriose, and short oligosaccharides. The oligosaccharides may be linear or α-limit dextrins (5–9 glucose units with a branch point). Further hydrolysis of these products is carried out by surface enzymes of the small intestinal epithelium cells. These enzymes are also referred to as *brush border enzymes*. They are firmly attached to the cell surfaces with their catalytic domains protruding into the intestinal lumen.

The final product of digestion of starch (or its animal equivalent, glycogen) is glucose through a complex series of reactions. The initial digestion involves *amylase*, which occurs free in the lumen, whereas the final processes involve brush border enzymes, which are attached to the enterocyte mucosal membrane.

Some **brush border enzymes** and their actions are as below:

1. *Maltase:* This enzyme possesses *α(1→4) glucosidase* activity which enables it to cleave maltose into two glucose residues. It can also hydrolyze short linear oligosaccharides of up to 9-carbon unit length.

2. *Lactase:* It degrades lactose into glucose and galactose. Its action is slower than the other brush-broder enzymes which have excess capacity to hydrolyze their substrates. Moreover, quantity of the *lactase* is just about sufficient to degrade the lactose that is presented to it. This is in contrast to the other brush border enzymes, synthesis of which can be induced, if required.

3. *Isomaltase or sucrase:* This enzyme is initially synthesized as a single polypeptide chain, which is later cleaved into two subunits, each having a distinct enzymatic activity. Both these subunits get embedded in the brush border, where:

(i) The *isomaltase activity* accounts for cleaving the α(1→6) linkages in isomaltose and the α-limit dextrins.

(ii) The *sucrase activity* accounts for the hydrolysis of sucrose into glucose and fructose.

In addition, some other *disaccharidases* and *oligosaccharidases* are present in the intestinal brush border, as summarized in Table 26.1.

Thus, concerted action of various enzymes results in breakdown of dietary carbohydrates to produce monosaccharides such as D-glucose, D-galactose, and D-fructose.

Table 26.1. *Disaccharidases* and *oligosaccharidases* of the intestinal brush border

Enzyme	Cleavage specificity
Maltase	Maltose, maltotriose; also acts as *exoglycosidase* on α(1 → 4) bonds at the non-reducing end of starch and starch-derived oligosaccharides
Lactase*	Lactose; also cellobiose[#]
Cerebrosidase*	Gluco- and galactocerebroside
Sucrase	Sucrose; also maltose and maltotriose
Isomaltase	α(1 → 6) bonds in isomaltose and α-limit dextrins
Trehalase	Trehalose•

* The *lactase* and *cerebrosidase* activities reside in two different globular domains of the same polypeptide.

[#] Cellobiose is a disaccharide of two glucose residues in β(1 → 4) glycosidic linkage.

• Trehalose is a disaccharide with a structure of α-D-glucopyranosyl-α-D-glucopyranoside (in 1,1 glycosidic linkage); common only in mushrooms.

These compounds are absorbed by **facilitated diffusion**, i.e. through mediation of the carrier proteins present in the plasma membrane (Chapter 7). Based on studies, carried out in rat-intestine, it has been estimated that the relative rates of absorption of the monosaccharides are as follows: glucose, 100%; galactose, 11%; fructose, 43%; mannose, 10% and xylose, 15%.

Transport of glucose and galactose into the cell can be coupled with passive diffusion of sodium ions. As the sodium diffuses across the cell membrane of enterocyte along the concentration gradient, glucose and galactose are also transported along with it. This mode of transport is referred to as the ***secondary active transport*** (Chapter 7). After absorption, the monosaccharides pass into portal circulation.

Absorption of monosaccharides resulting from digestion of dietary carbohydrates occurs via specific carrier-mediated mechanisms, which demonstrate substrate specificity and stereospecificity.

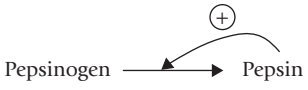
B. Proteins

Digestion of Dietary Proteins Begins in the Stomach

The gastric secretion contains ***hydrochloric acid*** and ***pepsinogen***, a zymogen. Both play important role in protein digestion. The most potent stimuli for gastric secretion are dietary proteins: partially digested proteins and amino acids are more effective stimuli than the intact food proteins. These compounds first stimulate release of a hormone, **gastrin**, from the gastric antrum (Fig. 26.1). Gastrin in turn stimulates the gastric parietal cells (to release hydrochloric acid) and the chief cells (to release pepsinogen). Pepsinogen, an inactive precursor form, is subsequently converted to *pepsin* which is primarily responsible for the digestive activity of gastric juice.

Hydrochloric acid performs several important functions.

(a) It acts as an antiseptic and lowers the pH of the food mixture to about 1.5–2.5. This pH range is optimum for the action of *pepsin*.
(b) Hydrochloric acid brings about denaturation (unfolding) of the polypeptide chains, which exposes the peptide bonds for enzymatic action.
(c) Low pH (below 5) induces the conversion of *pepsinogen* to its activated form, *pepsin*, by removal of 42 amino acids from the N-terminal of the polypeptide chain.

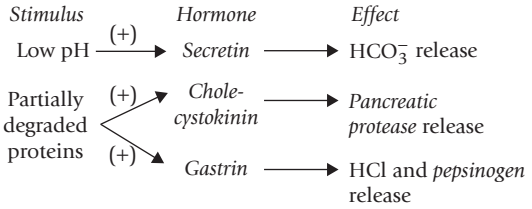


Further conversion proceeds autocatalytically, i.e. pepsin itself activates more of pepsinogen molecules (Fig. 26.1).

Pepsin is an ***endopeptidase***, i.e. it cleaves the peptide linkages lying towards the core of the polypeptide chain. Moreover, it is **site-specific** in action, cleaving only those peptide bonds whose carboxy terminal amino acid is an aromatic amino acid (phenylalanine, tyrosine, or tryptophan). Its action yields partially disintegrated proteins.

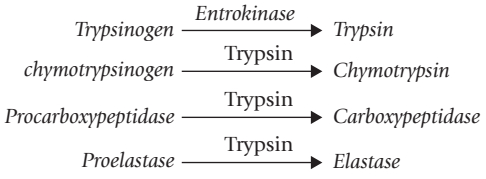
Bulk of Protein Digestion Takes Place in Duodenum

As the acidic gastric contents reach duodenum, their low pH acts as stimulus for duodenal mucosa, inducing it to release ***secretin*** (Fig. 26.1) (*Secretin was the first hormone ever to be identified*). Its predominant action is to stimulate release of water and bicarbonate ions from the exocrine pancreas. The bicarbonate ions elevate pH of the duodenal contents towards the neutral range, which is optimal for the action of *pancreatic enzymes*, including those of protein digestion (Table 26.2). These enzymes are secreted in response to another hormone released from the duodenal mucosa, i.e. *cholecystokinin*, earlier known as *pancreozymin*. The release of this hormone is stimulated by the products of protein digestion.



The pancreatic proteases are released in the form of inactive precursors, the form in which they are stored in the exocrine cells. Subsequently, the inactive precursors (**zymogens**) are activated in the intestinal lumen. This arrangement acts as safeguard against autodigestion of pancreatic cells since the activated pancreatic enzymes that are capable of digesting the cellular components are generated away from the cell.

Trypsinogen, *chymotrypsinogen*, *procarboxypeptidase* and *proelastase* are some examples of inactive precursors of the pancreatic *proteases*. Trypsinogen is initially activated by a duodenal enzyme, *enterokinase* which is located in the brush border. Once trypsin has been generated in this way, it acts as a major stimulus for the activation of other zymogens: chymotrypsinogen, *procarboxypeptidase* and *proelastase*.



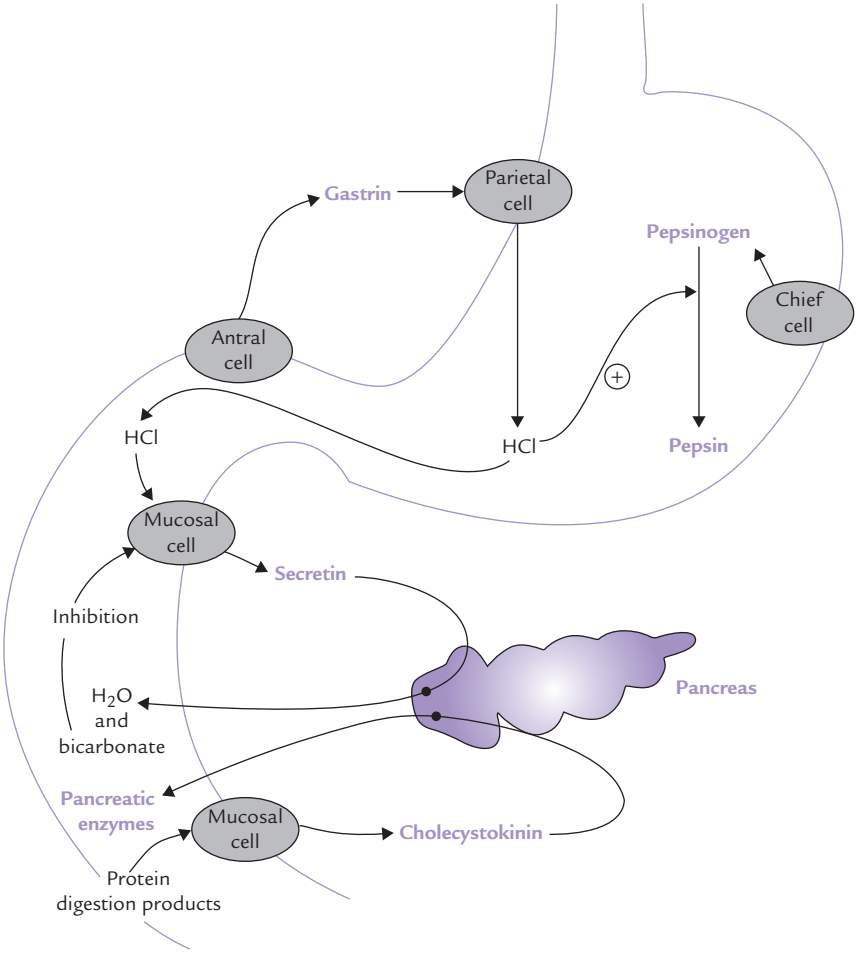


Fig. 26.1. Release and action of intestinal hormones.

Table 26.2. Enzymes of protein digestion

Enzyme	Source	Type	Catalytic mechanism
Pepsin	Stomach	*Endopeptidase*	*Carboxyl protease*
Trypsin	Pancreas	*Endopeptidase*	*Serine protease*
Chymotrypsin	Pancreas	*Endopeptidase*	*Serine protease*
Elastase	Pancreas	*Endopeptidase*	*Serine protease*
Carboxypeptidase A	Pancreas	*Exopeptidase*	*Metalloprotease* (Zn^{2+})
Carboxypeptidase B	Pancreas	*Exopeptidase*	*Metalloprotease (Zn21)*

Like pepsin, the pancreatic *proteases—trypsin, chymotrypsin* and *elastase*—are also **site-specific *endopeptidases***.

- *Trypsin* cleaves those peptide bonds the carboxy terminal of which is contributed by the basic amino acid (lysine and arginine).
- *Chymotrypsin* hydrolyzes the ones in which an aromatic amino acid forms the carboxy terminal.
- *Elastase* acts on the peptide linkages at carboxy terminal of small amino acids such as valine, leucine and alanine.

Combined effect of *pancreatic peptidases*, which have different substrate specificity with respect to peptide bond cleavage, is to produce abundance of free amino acids and oligopeptide fragments.

In contrast with these enzymes, the *carboxypeptidase* is referred to as an ***exopeptidase*** (Table 26.2). This is because *carboxypeptidases* (both A and B type) act only on the terminal peptide bond of the polypeptide chain: their action catalyzes successive removal of a single amino acid from the carboxy terminal of the polypeptide chain. Cleavage specificity of **type A** is hydrophobic amino acid at carboxy terminal, and that of **type B** is basic amino acid at carboxy terminal.

Final Stages of Digestion Occur in Small Intestine

Various *peptidases* are released in small intestine which act at multiple sites. This greatly enhances the overall efficiency and speed of proteolysis. Their concerted action cleaves the polypeptide chain into multiple oligopeptide fragments.

In final stages of protein digestion, hydrolysis of these oligopeptides into constituent amino acids occurs. Enzymes operating at this stage are called *peptidases* (*dipeptidase* and *aminopeptidase*). They are located in the brush border membrane. *Aminopeptidase* catalyzes removal of a single amino acid from the N-terminal end of the oligopeptide fragment. The *dipeptidase*, as is evident from its name, cleaves dipeptides to yield two amino acids.

Protein digestion is completed by *endopeptidases: dipeptidases,* and *aminopeptidases.*

Absorption

Amino acids produced by the action of these enzymes are actively absorbed, being coupled with passive diffusion of sodium (i.e. **secondary active transport**). There appears to be at least six different carrier proteins for amino acid transport across the luminal mucosa. The dipeptides and tripeptides can also be absorbed, although at a much slower rate. About 5% of the ingested proteins remain unabsorbed and are excreted in faeces.

C. Lipids

An average diet provides 90–100 g lipids, **90%** of which are **triacylglycerols**. The dietary triacylglycerols are termed saturated or unsaturated fats depending on whether the glycerol backbone is esterified with the saturated fatty acids or the unsaturated fatty acids. The other dietary lipids are **phospholipids** and **cholesterol esters**. In addition to dietary lipids, the gastrointestinal system has to handle about 1 g of cholesterol and 5 g of lecithin that enter intestine through bile each day.

Digestion of lipids is more complex than that of proteins or carbohydrates because the hydrophobic nature of lipids limits the digestive process to the lipid–water interphase. This is in contrast with the proteins and the carbohydrates, which being hydrophilic are thoroughly exposed to action of hydrolytic enzymes. For instance, the enzymes catalyzing digestion of carbohydrates and proteins, such as *endosaccharidase* and *endopeptidases* are capable of hydrolyzing the internal bonds of these molecules, including those lying near the core.

Digestion of lipids starts in the mouth by action of the enzyme ***lingual lipase***, which is secreted by the sublingual glands. As the food mixes with saliva, this enzyme begins hydrolytic removal of the fatty acids esterified to the C-1 and the C-3 of the triacylglycerol molecule. As a result, the dietary triacylglycerol is converted to the corresponding diacylglycerol and then to monoacylglycerol.

The hydrolytic process continues in the stomach also due to action of another lipolytic enzyme, the ***gastric lipase***.

The partially hydrolyzed food then enters the duodenum in the form of an unstable emulsion. Further digestion occurs in the duodenum, where the following processes occur:

Emulsification

The emulsion is stabilized and subsequently dispersed by the action of **bile-salts**, namely *sodium taurocholate* and *sodium glycocholate* (Chapter 12). These compounds are synthesized from cholesterol in hepatocytes, and secreted into the duodenum through the hepatobiliary route. The bile-salts are major constituents of bile.

Bile salts are amphipathic in nature, i.e. they have a polar portion (termed head) and a non-polar portion (called tail). The polar heads form a hydrophilic coating on surface of the lipid particles which face the aqueous exterior. The non-polar tails, on the other hand, extend into the interior of these particles (see Fig. 3.3). The above-stated action of the bile salts is supplemented by phospholipids and other emulsifiers already present in food. This results in *dispersion of the lipid particles into the finer emulsion droplets*. As a result, the surface area of the lipid: water interphase, on which the digestive enzymes can act, is increased several folds.

Bile salts are essential for solubilizing lipids during the digestive process. Lipids with lowest water solubility are most dependent on bile-salts for their digestion.

Enzymatic Hydrolysis

The ***pancreatic lipase*** is the principal enzyme of fat digestion, carrying out bulk of the digestion in the duodenum. This enzyme is secreted in an inactive precursor form, *prolipase*, along with its cofactor, called *colipase*. The latter is also initially secreted in a precursor form called *procolipase*.

*The **colipase** positions itself on the surface of a fat particle and helps in anchoring the pancreatic lipase to the fat particle* (Fig. 26.2). The enzyme then carries out fat degradation: the triacylglycerol particles are hydrolyzed to fatty acids, diacylglycerol and monoacylglycerol.

As in case of other pancreatic enzymes, the neutral pH required for the optimum activity of *lipase* is attained by action of bicarbonate ions.

Hydrolysis of other dietary lipids, cholesterol esters and phospholipids also occurs in the duodenum as below:

$$\text{Cholesterol esters} \xrightarrow{\textit{Esterase}} \text{Cholesterol} + \text{Fatty acid}$$

$$\text{Lecithin} \xrightarrow{\textit{Phospholipase}} \text{Lysolecithin} + \text{Fatty acid}$$

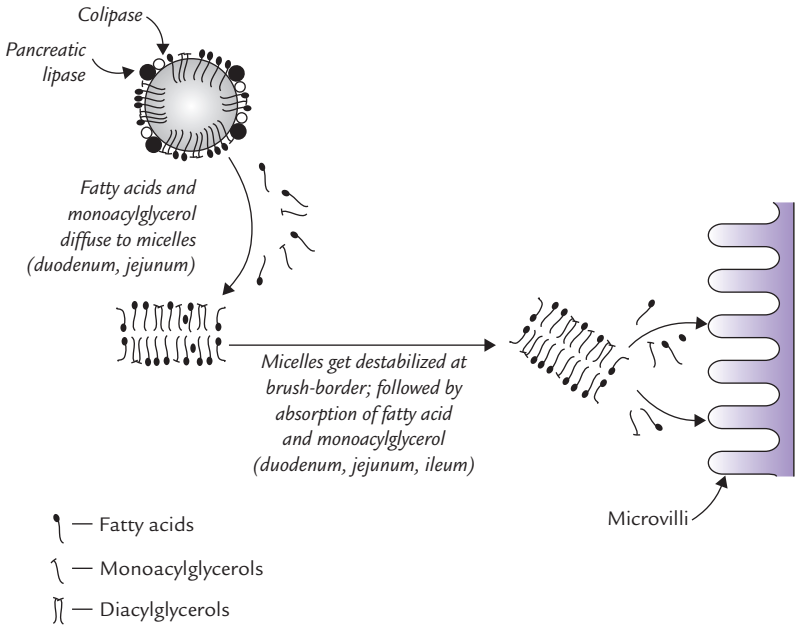


Fig. 26.2. Enzymatic hydrolysis of fat droplet, micelle formation and absorption of products of digestion in brush border.

Micelle Formation

The fatty acids and other simple molecules formed by enzymatic hydrolysis of the dietary lipids are aggregated with bile salts to form mixed micelles. In this type of micelle, aggregation of different types of molecules occurs. The polar heads of the bile salts from the hydrophilic exterior that faces the aqueous medium; non-polar tails of the bile salts form the hydrophobic core, in which fatty acids and monoacylglycerols are held. A single emulsion particle can from 10^6 micelles, each of about 20 nm diameter. Micelles can accommodate small quantities of cholesterol, carotene and other such non-polar substances as well.

Absorption

Absorption of the products of lipid digestion occurs by *passive diffusion*, mainly in the proximal jejunum. The micellar complex gets destabilized when it faces the relatively acidic pH prevailing at the brush border of enterocytes. Liberation of the monoacylglycerols, fatty acids, cholesterol and other non-polar substances, held in the hydrophobic interior, occurs as a result of the destabilization. This is followed by passive diffusion of these substances across the luminal cell membrane. Some fatty acids such as oleic acid and linoleic acid cross the luminal cell membrane by *facilitated diffusion* to enter the enterocyte.

Within the enterocyte, resynthesis of triacylglycerols and cholesterol esters occurs in endoplasmic reticulum. These lipids, together with small quantities of phospholipids and apolipoprotein B-48 are incorporated into **chylomicrons** (Chapter 12). The chylomicrons cross the cell membrane (serosal aspect) of enterocytes and pass into the lymphatic vessels. The triacylglycerols having short chain fatty acids directly enter the portal circulation. Some short chain fatty acids of chain length of 6–10 carbon atoms do not even need esterification in the cells and can directly enter the portal circulation.

Digestion and absorption of **nucleic acids** are discussed in **Chapter 20**.

II. Organ Function Tests

A. Gastric Function Tests

These tests evaluate acid secretion by the stomach. A sample of gastric juice is obtained by aspiration via a nasogastric tube. The parameters generally measured in the sample are **titrable acidity** and **acid output**. However, titrable acidity is affected spuriously by two factors, dilution and neutralization. The aspirated juice may be diluted by food, saliva, pancreatic secretion or bile. Further, these secretions may be alkaline in nature and hence tend to neutralize gastric acid.

Acid output of the stomach, which is the product of titrable acidity and volume of secretion, is the more commonly used test. It corrects for dilution but not for acidity.

However, in spite of these limitations, useful information has been obtained by these simple biochemical tests. Acid output may be measured under basal conditions, or after administration of some gastric acid-stimulating factor, as discussed below:

Acid output measurement
- **Basal acid output (BAO)** – after overnight fast
- **Maximum acid output (MAO)** – after test-meal or administration of histamine/pentagastrin

The sample of gastric juice is collected over a defined time period (usually one hour) by nasogastric aspiration. Following an overnight fast, the patient is directed to swallow a nasogastric tube, passed through one of his nostrils. The tube is directed radiologically and positioned with its tip in the gastric antrum. Samples of the gastric juice are collected by suction every 15 minutes, for a total period of one hour, i.e. **basal secretion.** Volume of the secretion collected in one hour is recorded and its acid content, i.e. **basal acid output** or BAO is determined by titration with alkali (Normal BAO: 0–17 mmol/h). The patient is kept fasting till end, which ensures that the gastric cells remain unstimulated. The acid output may also be measured following stimulation of the gastric secretion, i.e. **stimulated secretion.** The stimulus is provided by *test meal* or by exogenous agents such as *histamine* and *pentagastrin*. The test results are tabulated as below:

	Volume	Time	Acid output
Basal secretion	—	—	—
Stimulated secretion	—	—	—

The physiological stimulation by test meal may lead to inaccurate results because the food components may dilute the gastric secretion and neutralize its acid content. Measurement of acid secretion following **histamine secretion** is, therefore, the more widely used test. Histamine interacts with specific receptors on gastric cells called H_2 *receptors*, which in turn leads to stimulation of acid secretion. However, use of histamine suffers from a serious drawback. Histamine also interacts with another type of receptor called H_1 *receptor*. The latter interaction elicits a widespread vasodilatation which evokes several undesirable side effects, including headache and flushing. Fortunately, prior administration of benadryl, an H_1 blocker, prevents these effects without affecting the acid secretion.

Pentagastrin, a synthetic analogue of gastrin, is a highly potent and predictable stimulus for acid secretion. Subcutaneous administration of pentagastrin stimulates the parietal cells to secrete maximum acids, i.e. **maximum acid output** (MAO).

Gastric analysis involves collecting gastric secretions for a baseline period to determine basal (or unstimulated) gastric secretion. Next, a parietal cell stimulant is administered and gastric juice is collected to evaluate maximum secretory ability.

The aforementioned tests are principally used to investigate the conditions in which gastric acid secretion is either excessive or inadequate. Excessive gastric acid secretion (*hyperchlorhydria*) is an important factor in pathogenesis of duodenal ulcers. Decreased acid secretion (*hypochlorhydria*) is seen in conditions where gastric mucosa is damaged, such as chronic gastritis and gastric carcinoma. When the damage is so extensive that no acid output can be detected by the aforementioned tests the condition is called **achlorhydria.** This condition is most frequently seen in patients with atrophic gastritis, but also occurs in pernicious anaemia and gastric carcinoma.

Newer tests like endoscopy and contrast radiography make direct visualization of gastric mucosa possible, and therefore the mentioned tests (involving measurement of acid output) are becoming obsolete. However, they are still of much use in the following conditions:

1. *Zollinger Ellison syndrome:* The underlying defect in this condition is carcinoma of the gastrin-producing cells. Excess gastrin causes a persistent stimulation of the parietal cells, resulting in marked increase in the acid output even in the unstimulated state. Consequently, the *ratio of the basal acid output and the maximum acid output rises up to 0.6%* (Normal: BAO, 0–17 mmol/h; MAO, 4.7–58.4 mmol/h).
2. *Pernicious anaemia:* Extensive damage of the gastric mucosa occurs in this disorder, resulting in achlorhydria.
3. *Following gastrectomy:* Gastrectomy entails removal of a portion of the acid-secreting stomach wall. It is performed for treating acid hypersecretion.

Insulin secretion test is performed to check whether vagus nerve has been properly sectioned in the patients treated by vagotomy. Vagotomy is performed for treating gastric hypersecretion. The test is based on the principle that insulin-induced hypoglycaemia stimulates vagal center, which in turn serves as a potent stimulus for acid secretion. Therefore, in normal subjects, insulin injection is followed by increased acid secretion. However, following vagotomy, this is not expected, provided the vagus nerve has been properly sectioned.

In a typical test (***Hollander test***), 0.1–0.2 units/kg body weight of regular insulin is injected, which is followed by sampling of the gastric juice. The test, however, is not commonly used due to associated risk of hypoglycaemia.

Measurement of plasma gastrin concentration is now the first-line test for the investigation of atypical peptic ulceration such as duodenal ulcer resistant to medical treatment, recurrent ulcer after surgery, multiple duodenal ulcers and jejunal ulcers.

B. Pancreatic Function Tests

Changes in both internal and external secretions of the pancreas may be observed in pancreatic diseases. As regards the internal secretions (insulin, glucagon, etc.) conditions in which the α- and the β-cells are affected, have already been considered (Chapter 15). The tests described below are mostly concerned with exocrine secretions of the organ, i.e. sodium bicarbonate, and the enzymes, *amylase, lipase*, and the group of *proteinases*, still referred to as trypsin. The tests can be divided into two categories: direct tests and indirect tests.

Direct Tests

These involve quantitative estimation of the output of fluid, bicarbonate, and enzymes in the sample obtained by duodenal intubation, following pancreatic stimulation.

Collection of duodenal contents is always preceded by an overnight fast. On the morning of the test, the pancreatic secretions are stimulated by exogenous hormones or by test meal. This is followed by collection and analysis of the duodenal contents.

Secretin-cholecystokinin Test

The stimulus for the pancreatic secretion is provided by administration of hormones, secretin and cholecystokinin. During collection of the duodenal contents, special care is taken to prevent contamination by gastric secretion. A *double lumen tube* is used for this purpose, which consists of a longer and a shorter tube placed together. The length of longer tube is about 25 cm more than the shorter one. The tube is positioned, under radiological control, in such a way that the shorter tube ends at pyloric antrum and the longer one in the duodenum near the opening of the pancreatic duct. Continuous aspiration in both these tubes permits recovery of uncontaminated duodenal contents from the longer tube. The samples are collected in containers chilled in ice.

At the end of the collection period, the total volume of the aspirate is measured and the total bicarbonate output and the enzyme activities are determined. Different investigators favour different enzymes, but *amylase* and a proteolytic enzyme (trypsin or chymotrypsin) are probably best. The range for normal response is very wide; Table 26.3 gives typical lower limits of the normal response.

The **sensitivity** of this test for the diagnosis of exocrine pancreatic-insufficiency is approximately 85% and the **specificity**, approximately 90%. The test has high diagnostic significance in a number of pancreatic disorders, discussed below.

Table 26.3. Secretin-cholecystokinin test

Test	Lower limit of normal response
Volume of aspirate	150 mL
Peak bicarbonate concentration	90 mmol/L
Peak tryptic activity	30 IU/mL
Peak amylase activity	270 IU/mL

Diagnostic significance: Following an episode of acute pancreatitis, the bicarbonate and the enzyme outputs are depressed, but return to normal as the pancreas recovers (**Case 26.1**). In **chronic pancreatic insufficiency**, values of the bicarbonate and enzyme outputs are persistently decreased, indicating extensive acinar damage. The fall in bicarbonate secretion is more marked in the early stages than the (reduction in) output of enzymes, which is observed late in the disease process. In fact, such a dissociation is highly suggestive of chronic pancreatitis. In **carcinoma pancreas**, particularly when the growth is located in the head region, secretion of enzymes tends to be affected to an extent greater than that of the bicarbonate ions. However, progression of any pancreatic disease eventually results in depression of both. Such changes may be observed in extra-pancreatic disorders, such as duodenal ulcer, gall bladder disease and haemochromatosis.

Other Direct Tests

Several simpler tests have been developed in which a *physiological stimulation* is employed for stimulating the exocrine pancreatic secretion. Most common is the **Lundh test** in which the stimulus is provided by a *test meal containing corn oil, skimmed milk powder and dextrose*. The samples are collected by a single lumen tube for a period of 2 hours and analyzed for tryptic activity. Reduced activity (positive test) is seen in about 90% of the cases of chronic pancreatitis but in only 79% of cases of pancreatic carcinoma. The test is most definite for tumours of head of the pancreas.

Note: The uncontaminated pancreatic juice can also be collected by **endoscopic cannulation** of the pancreatic duct. However, this method does not appear to offer any definite advantage.

Indirect Tests

Determination of Enzymes in Serum and Urine

In pancreatic disorders involving exocrine cell damage, the pancreatic enzymes may be absorbed into the portal blood stream and be found in increased quantities in the

peripheral blood (**Case 26.1**). Increased urinary excretion may also follow.

Serum amylase: It is a digestive enzyme secreted by pancreas and salivary glands. Most of the *amylase* present in blood circulation is of pancreatic origin, therefore a pancreatic pathology is commonest cause of increased serum *amylase* activity. In acute pancreatitis, greatest elevation of the enzyme activity is observed (often exceeds 1000 Somogyi units per 100 mL), whereas in chronic pancreatitis and carcinoma of the pancreas the serum enzymes are raised to a lower extent. Less significant increase occurs in acute parotitis, mumps or calculous disease of the salivary glands, and in some abdominal conditions such as perforated peptic ulcer, acute peritonitis and intestinal obstruction. Urinary *amylase* in most conditions is increased when the serum *amylase* is increased.

Serum lipase: It is an enzyme for lipid digestion secreted by exocrine pancreas. Its activity is likewise increased in acute pancreatitis.

Study of faeces: Since pancreatic enzymes are primarily involved in the digestive process, in severe pancreatic deficiency evidence of impaired digestion of foodstuff may be obtained from the stools. Analysis of fecal matter shows increased fat (*steatorrhoea*) and *excess nitrogen*. Quantitative estimation of the average daily excretion of fat and nitrogen is made when the patient is given a standard diet containing known amounts of fat and protein.

Increased serum *amylase* activity most commonly suggests a pancreatic pathology. Urinary excretion of the enzyme follows. Faecal excretion of fats and nitrogen is also seen.

Determination of sweat chloride: This test is important in the diagnosis of cystic fibrosis (fibrocystic disease of the pancreas or mucoviscidosis). There may be an abnormal increase in the chloride content of sweat: a concentration above 60 mEq/L is diagnostic of this condition. The genes which undergo mutation in this autosomal recessive disease have been recently identified and sequenced (Case 7.1).

Tubeless Pancreatic Function Tests

The patient is administered certain labelled compound that may serve as substrate for a pancreatic enzyme. The digestion product of this compound is then measured, usually in urine sample. Decreased excretion indicates impaired digestion and hence pancreatic insufficiency. The following two tests are commonly employed.

Fluorescein dilaurate test: The substrate used in this test is fluorescein dilaurate. When administered orally, it is hydrolyzed by *pancreatic esterase* to release fluorescein. The latter is absorbed in the small intestine, conjugated in liver to form fluorescein glucuronide and excreted in urine. Since bile salts are also needed for the activity of esterase, the test effectively assesses combined pancreaticobiliary function.

False positive results are obtained in case of any defect in hepatic conjugation, intestinal absorption or renal excretion. To avoid any such error of interpretation, an equivalent amount of free fluorescein is given on the following day and urinary excretion of this compound is measured. The results are then compared.

^{14}C PABA test: A synthetic peptide, BT-PABA is the labelled compound administered in this test. It is hydrolyzed to PABA by chymotrypsin and excreted in urine. The PABA excretion in the urine is measured and the results are expressed as ratio of the amount excreted to the dose administered. This test is also affected by the extrapancreatic factors mentioned earlier, but the sensitivity and specificity are high (up to 90%).

Demonstration of an Associated Abnormality

Certain pancreatic disorders are associated with *hyperlipidaemia* or *hypercalcaemia*. Therefore, detection of these biochemical aberrations indirectly indicates a pancreatic disorder.

Some familial disorders involving pancreas are associated with *aminoacidurias*. Measurement of these parameters serves as a useful pointer towards the above stated pancreatic disorders.

Visual Procedures

These include radiographic methods such as ultrasonic scanning and computed tomography scanning.

Pancreatic function tests are mostly used to measure the ability of the pancreas to produce enzymes, proteins and bicarbonate, and to secrete an adequate volume of fluid into the duodenum.

C. Intestinal Function Test

Test for Carbohydrate Digestion and Absorption

A number of tests are available to study disorders affecting the carbohydrate digestion and absorption. Some of the most commonly used tests are given here.

Tests of Disaccharidase Deficiency

In these tests, the blood glucose response is observed following oral administration of a disaccharide. The principle

of these tests is that the administered disaccharide is hydrolyzed to the constituent monosaccharide units by the intestinal brush border *disaccharidases*; for example, lactose and maltose, the commonly used disaccharides, are hydrolyzed as below:

$$\text{Lactose} \xrightarrow{\textit{Lactase}} \text{Glucose} + \text{Galactose}$$

$$\text{Maltose} \xrightarrow{\textit{Maltase}} \text{Glucose} + \text{Glucose}$$

Glucose, produced by the action of *disaccharidases*, is then absorbed to cause elevation of blood glucose level. Thus, the blood glucose response serves as an indicator of the *disaccharidase* activity. In case of impairment of the activity of *lactase* (or *maltase*), the blood glucose elevation following administration of lactose (or maltose) is impaired.

The commonest of the *disaccharidase* deficiency disorders is *lactose intolerance,* in which *lactase* is deficient. In the affected children, the blood glucose response, following oral administration of lactose (50 g), is considerably depressed as compared to the normal subjects.

Tests for Monosaccharide Absorption

These tests involve oral administration of a monosaccharide followed by measurement of its plasma concentration and urinary excretion. In some disorders affecting intestinal absorption, the blood concentration as well as the urinary output of the ingested monosaccharide are depressed. The test results are, however, not affected by deficiency of the digestive enzymes.

A number of monosaccharides were initially tried, but xylose has been found most suitable for this test. It is a plant monosaccharide, that is rapidly absorbed (without undergoing any change in intestine) and excreted as such. Xylose is rapidly eliminated in urine, there being no renal threshold for it.

Different investigators have recommended different doses of xylose, although 5 g is the most commonly used dose. The compound is administered orally. Blood xylose concentration is determined in a sample drawn one hour after the administration. Urinary excretion of this compound is measured in samples collected at regular intervals, for a total period of 5 hours. In a normal subject, the blood xylose concentration reaches a maximum of 250 mg/dL, and the urinary excretion exceeds 4 g in the 5-hour period.

Apart from the rare congenital abnormalities of the glucose transport system, decreased absorption mostly occurs as a consequence of an impaired absorptive surface area. Abnormal results are always found in severe celiac disease and disorders of the proximal small intestine (**Case 26.2**). However, in milder diseases, the test results may not be affected. Therefore, the xylose excretion test cannot be used as a screening test for malabsorption, but may be useful in differential diagnosis of steatorhea as an alternative method.

Other monosaccharides (especially glucose) have been tried instead of xylose. But the tests involving glucose lack specificity because the results are affected in several extra-intestinal disorders as well (for example, in diabetes).

Hydrogen Breath Test

This test provides convenient alternative for more cumbersome tests than just described. It is especially useful for the diagnosis of lactose intolerance. As mentioned earlier, in this disorder the lactose cannot be hydrolyzed and therefore, starts accumulating in the intestine. Bacterial fermentation of the accumulated lactose produces a number of by products including hydrogen, which is exhaled. Since hydrogen is not normally present in the expired air, its detection in breath serves as an indicator of the *disaccharidase* deficiency.

Others

Less common *disaccharidase* deficiencies include *maltase* deficiency and *sucrase-isomaltase* deficiency. The definitive tests for the specific diagnosis of these conditions are **histopathological examinations** of the biopsy sample. **Measurement of the appropriate enzymes** in biopsy sample also has high diagnostic value (**Case 26.2**).

III. Liver Function Tests

A number of biochemical tests, referred to as liver function tests, are available to make a quantitative assessment of the hepatic cell activity. These tests involve measurement of certain parameters in blood, urine, and stool samples, and in some cases, in exhaled air. Since these parameters are directly related to function of the hepatocytes, alternation of their levels may reflect hepatic disorders. Thus, they serve as the first line investigations for the diagnosis of these disorders.

These tests are also useful for follow up studies in chronic hepatic disorders; periodic estimations of these parameters provide important information regarding response to therapy.

Liver function tests are based on various functions of this organ and can be accordingly divided into five categories:

A. Tests based on excretory function of liver.
B. Tests based on role of liver in intermediary metabolism.
C. Tests based on synthetic function of liver.
D. Tests based on detoxification function of liver.
E. Tests based on diagnostic enzymes.

A. Tests Based on Excretory Function of Liver

Serum Bilirubin Estimation

Bilirubin, a catabolic end product of haem, is generated in reticulo-endothelia system, metabolized in liver and excreted by biliary system (Chapter 16). The normal plasma concentration of bilirubin is less than 1.0 mg/dL; however, increased concentration (> 3.0 mg/dL) is readily recognized clinically because of the yellow colour that the bilirubin imparts to the skin (jaundice).

Estimation of serum bilirubin is based on its reaction with ***diazo reagent*** (*van den Bergh reaction*) to form purple coloured azobilirubin. Because of its solubility in the aqueous medium, the conjugated bilirubin (bilirubin diglucuronide) is capable of directly reacting with diazo reagent. It is, therefore, referred to as the **direct bilirubin**. The unconjugated bilirubin, on the other hand, being non-polar, reacts only after it has been solubilized by methanol. It is referred to as the **indirect bilirubin.**

Estimation of direct and indirect bilirubin is useful for delineating the predominant hepatic pathology. In obstructive pathology, predominant rise occurs in indirect bilirubin. On the other hand, in case of hepatocellular damage, rise in serum concentrations of both direct and indirect bilirubin occurs.

Further details of bilirubin metabolism and the biochemical tests based thereupon are discussed in Chapter 16.

BSP Elimination Test

BSP is a drug that is taken up from the blood circulation by the liver where it is metabolized. The metabolized drug is subsequently excreted through bile. When a measured dose of BSP is administered, rate of its disappearance from the blood circulation depends on the functioning liver cell mass. Normally, only about 5% of the administered dose remains in blood circulation 45 minutes after the administration. In cases of hepatic impairment, the ability of the damaged cells to take up the circulating drug falls. Consequently, a larger percentage of the administered drug is detectable in blood after 45 minutes.

B. Tests Based on Role of Liver in Intermediary Metabolism

Liver is the major organ for metabolism of carbohydrates and proteins; in lipid metabolism its importance stands next only to adipose tissue. The under-mentioned tests are based on the role of liver in intermediary metabolism.

Galactose Tolerance Test

Liver is the major organ for metabolism of galactose. In case of hepatic impairment, metabolism of galactose proceeds at a rate much slower than normal. Half-life of the circulating galactose, which is 12 ± 2.6 minutes, increases in this instance.

The test is carried out by giving a measured dose of galactose intravenously and estimating its blood level at specific intervals. In case of hepatic impairment, the blood galactose levels remain elevated for a longer time and to a greater extent. Galactose tolerance test is one of the most sensitive of all liver function tests.

Amino Acid Profile

Serum amino acids are taken up from blood circulation by hepatocytes, where *amino acid oxidases* metabolize them. For example, aromatic *amino acid oxidase* brings about catabolism of the aromatic amino acids: tyrosine, tryptophan and phenylalanine. The enzyme activity is depressed in hepatic disorders, resulting in elevated blood level of these (aromatic) amino acids. Thus, circulating levels of the aromatic amino acids relate inversely with liver cell function.

Conversely, *levels of the branched chain amino acids*, namely leucine, isoleucine and valine **fall** in hepatic disorders. Therefore, amino acid profile is a useful indictor for the diagnosis of hepatic disorders.

C. Tests Based on Synthetic Functions of Liver

Serum Protein Estimation

Liver is an important site for the synthesis of a variety of biomolecules, including plasma proteins. In fact, *all subfractions of the plasma proteins, with sole exception of γ-globulins, are synthesized in liver*. In most hepatic disorders, impaired synthesis of these molecules occurs, resulting in hypoproteinaemia and hypoalbuminaemia. Albumin: globulin **(A:G) ratio** is a more reliable indicator of hepatic dysfunction than total protein concentration.

Prothrombin Time

Hepatocytes synthesize most coagulation factors, including factor II (prothrombin). Prothrombin levels (assessed by measuring prothrombin time) may be reduced in hepatocellular damage because of impaired synthesis. Vitamin K is necessary for the synthesis of prothrombin, and therefore, deficiency of vitamin K is accompanied by a bleeding tendency and a **prolonged prothrombin time**. This can be corrected by parenteral administration of vitamin K (provided hepatocellular function is normal).

Hepatic protein synthesis is important, as the majority of plasma proteins are synthesized in the liver; and, hepatocellular disease may alter protein synthesis both qualitatively and quantitatively.

D. Tests Based on Detoxification Function of Liver

Several endogenously produced metabolites and exogenously administered chemicals (xenobiotics) that are potentially hazardous, are converted to relatively harmless substances in hepatocytes (Chapter 15). This process, known as detoxification, involves oxidation and attachment of polar groups, which render these substances hydrophilic so that they can be excreted through the renal or the hepatobiliary routes.

Arterial Ammonia Levels

Ammonia is a highly toxic substance that is produced in the peripheral tissues from amino-group catabolism. It is also produced in colon by bacterial fermentation, and in muscles by adenine breakdown. It is carried to liver (as amino-group of glutamine) where it is channeled into urea cycle. Since urea cycle operates only in liver, in *hepatic impairment*, decreased ammonia detoxification occurs, which leads to *elevated blood ammonia levels*.

The arterial ammonia level serves as a more reliable indicator of hepatic impairment than the venous-blood ammonia level. This is because the latter keeps fluctuating due to continuous addition of ammonia in veins from various sources. These levels are further increased due to formation of venous shunts that bypass the liver.

^{14}C Amino-Antipyrine Excretion Test

The labelled ^{14}C amino-antipyrine is catabolized in liver; the end product is labelled (^{14}C) carbon dioxide, which is subsequently exhaled. In hepatic disorders, decreased catabolism of this compound occurs, resulting in decreased ^{14}C carbon dioxide in breath.

Thus, measurement of the labelled $^{14}CO_2$ in the exhaled air is a useful indicator of hepatocellular damage.

E. Tests Based on Diagnostic Enzymes

Diagnostic utility of enzymes like *transaminases*, *alkaline phosphatase*, *γ-glutamyl transpeptidase* and *5′-nucleotidase* in hepatic disorders was described earlier in Chapter 5. Some of these enzymes are finding increasing use in the diagnosis, being early indicators and sensitive markers of the hepatic disorders. Estimation of some enzymes which were earlier used as research procedures have been successfully introduced in clinical laboratories as routine tests; for example, serum *arginase* activity is a highly specific and sensitive indicator of hepatocellular damage.

The battery of tests described finds use not only for *diagnosis* of liver disorders, but also for their treatment and follow up (**Case 26.3**). These tests reflect a pathological process and not any disease specifically. Nowadays, highly sophisticated diagnostic tools based on imaging techniques are available. But their routine use is not possible due to the high costs involved. Therefore, the biochemical tests retain their utility and still remain the first-line investigations. A decision to go for the imaging techniques is usually taken based on the reports of these tests.

Exercises

Essay type question

1. Discuss various tests used for follow up studies in a patient suffering from a chronic liver disorder. Explain the role of diagnostic enzymes in the patient.

Write short notes on

1. Ven den Bergh reaction
2. Galactose tolerance test
3. Zollinger–Ellison syndrome
4. Basal and maximum acid output
5. Intestinal brush border enzymes
6. Secretin-cholecystokinin test
7. Xylose excretion test

CLINICAL CASES

CASE 26.1 Excruciating abdominal pain in an alcoholic

A 47-year-old man was brought to the hospital emergency in a disoriented state, having pain in the epigastrium, which radiated to the back. The pain started suddenly and had become very severe in a few minutes. This was followed by repeated vomitings. The patient was seen on an earlier occasion also because of a problem related to ethanol abuse. At that time, his brother had admitted that he took about half a bottle of whisky each day since his wife's death four months ago.

On examination, the epigastric region was extremely tender. Radiographic examination was unremarkable, without any evidence of intestinal obstruction or perforation.

Biochemical tests		
Test	**Patient's report**	**Reference range**
Serum urea	62 mg/dL	15–45 mg/dL
Serum creatinine	1.4 mg/dL	0.6–1.4 mg/dL
Serum calcium	8.1 mg/dL	8.7–10.7 mg/dL
Serum protein	5.4 g/dL	6.0–7.8 g/dL
Serum albumin	3.0 g/dL	3.6–7.4 g/dL
Serum *amylase*	500 Somogyi Units/dL	60–150 Somogyi Units/dL
Aspartate transamine (AST)	34 U/L	10–35 U/L
Alanine transaminase (ALT)	36 U/L	10–40 U/L
Alkaline-phosphatase	76 U/L	40–100 U/L

Q.1. What is the most probable diagnosis?

Q.2. State the biochemical abonormality in the above disorder.

Q.3. Which other enzymes have diagnostic significance in the above disorder?

Q.4. How does this condition affect the parameters of renal function: urea and creatinine?

Q.5. Mention the pathophysiological alterations that result in hypoalbuminaemia and hypocalcaemia.

Q.6. One day during this painful episode, the patient had mild hyperglycaemia (fasting blood glucose, 116 mg/dL). Provide a biochemical explanation for this observation.

CASE 26.2 A 3-year-old malnourished child

A 3-year-old boy was referred for investigation of weight loss, failure to thrive and frequent diarrhoea. Stools were semi-formed and greasy in appearance.

The child looked pallor, emaciated and was (a) below the third centile for height, and (b) below the tenth centile for weight.

He was irritable, listless and had markedly reduced appetite. He had severe muscle wasting, especially of the limbs, buttocks and shoulder girdle. The abdomen was slightly distended and tympanitic.

The history and examination findings being suggestive of malabsorption, the following intestinal function tests were performed:

1. Fecal fat excretion: Measurement of the 24-hour fecal fat excretion showed markedly increased fat excretion.
2. Xylose excretion test was performed with 5 g oral load of this monosaccharide. Serum xylose was 80 mg/dL one hour after the oral administration, and urinary xylose excretion was 1.4 g in the given 5-hour period.

Blood sample was analyzed with the following results:

Test	**Patient's report**	**Reference range**
Haemoglobin	9.7 g/dL	12–16 g/dL
Serum protein	5.7 g/dL	6.0–7.8 g/dL
Serum albumin	02.9 g/dL	3.6–5.4 g/dL

Examination of peripheral smear showed hypochromic, microcytic red blood cells.

A biopsy specimen was obtained from the jejunum. It showed total villous atrophy: the surface epithelium was altered with a sparse brush border, and the cells were cuboidal rather than normal columnar. There was infiltration of inflammatory cells in the subepithelial *layer*.

Q.1. Interpret result of the intestinal function tests.
Q.2. What is the most probable diagnosis? State the underlying lesion in this condition.
Q.3. Comment on the other biochemical test results.
Q.4. The patients suffering from the above disorder are known to develop secondary lactose intolerance. Explain why?
Q.5. What treatment do you suggest for this patient?

CASE 26.3 Bloody vomiting in a 30-year-old man

A 30-year-old man was admitted in the hospital emergency following a bout of bloody vomiting. He was mildly disoriented, had incoherent speech and was smelling of alcohol. He admitted being a heavy drinker for the past 12 years. On enquiry he revealed having suffered from a similar attack of haematemesis 4 months earlier, and despite instructions to keep away from alcohol, he continued to drink. He had been complaining of nausea, loss of appetite, tiredness, irregular bowels and pain in right hypochondrium for the past 4–5 weeks. Four days earlier he noticed dark urine and pale stools.

On examination, yellow sclera, suggestive of jaundice, was observed. Mild ascitis, and hepatomegaly were detected; the liver being palpable just below the costal margin and mildly tender. Blood sample was analyzed.

Biochemical tests		
Test	**Patient's reports**	**Reference range**
Serum bilirubin	7.8 mg/dL	0.1–1.1 mg/dL
Aspartate aminotransferase	380 U/L	10–35 U/L
Alanine aminotransferase	336 U/L	10–40 U/L
Alkaline phosphatase	42 U/L	40–100 U/L
Gamma glutamyl		
Transpeptidase	360 U/L	10–55 U/L (in males)
Urine examination:		
Bilirubin	++	−ve
Urobilinogen	+	−ve

Q.1. What is the most probable diagnosis?
Q.2. Comment on the biochemical test results.
Q.3. What is Zieve's syndrome associated with alcoholism?
Q.4. Besides the abnormal liver function tests discussed above, what other biochemical abnormalities may be present in this patient?
Q.5. Suggest treatment for this patient.

CHAPTER

FREE RADICALS IN HEALTH AND DISEASE

27

Free radical is a molecule or molecular fragment that contains one or more unpaired electrons in its outer orbit and has an independent existence. A free radical is designated by a superscript dot ($R^{\bullet}$). In body, the oxidative reactions normally ensure that molecular oxygen is completely reduced to water. Incomplete reduction of oxygen which generates *oxygen-free radicals* is an infrequent occurrence but plays vital role in health and disease, as discussed in this chapter. Some oxygen free-radicals of biological importance include *superoxide radical* ($O_2^{\bullet}$), *hydroperoxyl radical* ($HO_2^{\bullet}$), and *hydroxy radical* ($OH^{\bullet}$).

After going through this chapter, the student should be able to understand:

- Incomplete reduction of oxygen and other sources of free radicals; generation of free radicals; the chain reaction; the damaging and the beneficial effects of reactive oxygen metabolites.
- Scavenging of oxygen free radicals, the antioxidant defences and their role in aetiology of diseases.

I. Incomplete Reduction of Oxygen

Oxygen is the final acceptor of electrons in the electron transport chain (ETC). Normally, four electrons are transferred to molecular oxygen to form a water molecule; this is the final event of the electron transport chain.

$$4e^- + O_2 \xrightarrow{4H^+} 2H_2O$$

Thus, a complete reduction of the molecular oxygen requires **four electrons**. Sometimes oxygen settles for a fewer electrons at a time, thereby undergoing incomplete reduction. The products of incomplete reduction are referred to as **reactive oxygen intermediates**. They are highly reactive, being capable of reacting with a number of biomolecules. Because of their strong oxidant and cytotoxic properties, they are implicated in a number of pathological processes.

The incomplete reduction of oxygen during aerobic metabolism generates reactive oxygen intermediates such as the superoxide and hydroxy radicals.

A. Reactive Oxygen Intermediates and Free Radicals

The reactive oxygen intermediates are referred to as *oxygen free radicals* (OFR), *reactive oxygen metabolites* (ROM), or simply *active oxygen* (AO). These include, among others, **superoxide**, **hydrogen peroxide**, and **hydroxy radicals**.

1. ***Superoxide radical*** is produced when a single electron is transferred to oxygen. It is, therefore, referred to as the single electron reduction product of oxygen. It is both an anion and a free radical.

$$O_2 + e^- \longrightarrow O_2^- \text{ (Superoxide)}$$

2. ***Hydrogen peroxide:*** The *two-electron reduction* product of oxygen is hydrogen peroxide (H_2O_2).

$$O_2^- + e^- \xrightarrow{2H^+} H_2O_2$$

Since this reaction generates a non-radical product (H_2O_2) from the free radical reactants, it is called **dismutation**.

The dismutation reaction may occur spontaneously or it is catalyzed by the enzyme *superoxide dismutase (SOD)*. When spontaneous, its rate is very slow.

$$O_2^- + O_2^- + 2H^+ \xrightarrow{SOD} H_2O_2 + O_2$$

3. ***Hydroxy radical:*** The three-electron reduction product of oxygen is hydroxy radical *(•OH)* which is the *most powerful free radical*. It is produced when an electron is transferred from ferrous ion to hydrogen peroxide. This results in lysis of O–O bond; one of the fragments appears as the hydroxy radical, as shown below:

$$H_2O_2 + Fe^{+2} \longrightarrow OH_2^- + {}^{\bullet}OH + Fe^{+3}$$

The above reaction is referred to as the **Fenton reaction**.

A summary of all these reduction steps of oxygen is presented in Figure 27.1.

Superoxide radical is the precursor of other reactive oxygen intermediates; hydroxy radical is a much stronger oxidant, H_2O_2 is relatively harmless.

Another iron-catalyzed reaction for generation of hydroxy radicals is **Haber-Weiss reaction**.

$$O_2^- + H_2O_2 \xrightarrow{Iron} O_2 + {}^{\bullet}OH + OH^-$$

This reaction is more complicated than shown here.

B. Transition Metals

Like ferrous iron, copper in the cuprous state can also react with hydrogen peroxide to yield the hydroxy radical. The *ferrous iron* and *cuprous copper* are more reactive than their oxidized counterparts, ferric and cupric, respectively. They are referred to as the ***transition metals***. Evidently they are important in oxygen-free-radical biochemistry. In their absence, the superoxide and hydrogen peroxide are harmless and rapidly removed.

Because ferrous iron plays an important role in generating the deleterious free radicals, it is never found in free or loosely chelated form in healthy state. When transported in plasma, it is bound with apotransferrin (to form transferrin) and, while being stored, it is bound with apoferritin (to form ferritin). However, superoxide radical is capable of releasing the ferritin bound iron by causing ferric to ferrous conversion.

$$O_2^- + \text{Apoferritin–ferric} \rightarrow O_2 + \text{Apoferritin–ferrous}$$

Since ferrous has low affinity for apoferritin, it is liberated and, in turn, seriously exacerbates the tissue damage.

Transition metals, e.g. ferrous or cuprous, are important in generating the most powerful free radicals, e.g. hydroxy radical (•OH). Therefore, they never occur in free, unbound form in the body.

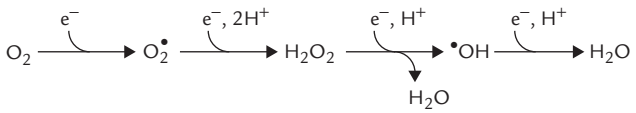


Fig. 27.1. Partial reduction of molecular oxygen by fewer than four electrons generates highly reactive products **$O_2^{\bullet}$** is one electron reduction product, **H_2O_2** is 2 electron reduction product •OH is three electron reduction product. $O_2^{\bullet}$ = superoxide radical, •OH = hydroxy radical, **H_2O_2** = hydrogen peroxide.

II. Generation of Oxygen Free Radicals

A. Electron Leakage

Major source of free radicals in cells is *"electron leakage"* from the electron transport chain. Normally, efficiency of the electron transfer reactions is so high that most of the oxygen is completely reduced to form water and less than 2% forms oxygen free radicals (OFR).

Simply stated, the electron leakage is minimum under normal circumstances. *Transition metals which can amplify the free radical response, are kept sequestered.* However, these measures are not completely effective and some "leakage" is inevitable due to improper operation of electron transport chain. This results in generation of OFRs, even under normal circumstances.

B. Normal Oxidation-reduction Reactions

Generation of free radicals is common during normal metabolism. Examples include:

1. Autoxidation of certain compounds including adrenaline, thiols, ascorbic acid, etc. Such reactions are greatly amplified through participation of transition metals.

2. Flavin coenzymes, present in the peroxisomes, are especially active in generating H_2O_2 (**Box 27.1**).
3. Enzymes, such as *xanthine oxidase*, *aldehyde oxidase*, and *dihydro-orotate dehydrogenase* can also generate OFRs.

C. Exogenous Agents

Toxic compounds, such as carbon tetrachloride, are capable of generating oxidative stress (Chapter 12). Exposure to *ionizing radiations* damages tissues by producing free radicals. Lights of certain wavelengths can cause photolysis of covalent bonds to produce free radicals. *Cigarette smoke* also contains dangerously high concentration of free radicals.

D. Respiratory Burst

In all the examples discussed so far, free radicals are accidently generated and are potentially hazardous for the body. However, free radical production is sometimes required in biological systems. For example, *phagocytes generate free radicals* with a specific purpose of destroying the engulfed bacteria. Certain enzymes (such as, *ribonucleotide reductase*) generate free radicals at their active site, which helps the catalytic process (Chapter 20). Certainly, *free radicals have beneficial effects as well.*

Phagocytosis by white blood cells: Phagocytosis is an important defence mechanism, especially against the invading bacteria. Neutrophils and monocytes are important components of this mechanism. Both can engulf the microorganisms, foreign particles and cellular debris, and subsequently kill the invading microorganisms through *generation (and participation) of the free radicals*.

Steps:

1. The enzyme ***NADPH oxidase*** catalyzes formation of superoxide radical from oxygen and NADPH (Fig. 27.2).
2. By dismutation, H_2O_2 is generated, as discussed earlier.
3. The hydrogen peroxide is involved in generation of another lethal oxygen species: **hypochlorous acid** (HOCl) and hydroxy radicals. Hypochlorous acid is formed by combination of chloride ions with the hydrogen peroxide. The enzyme catalyzing this conversion, *myeloperoxidase* (MPO) is present in neutrophil granules.

$$H_2O_2 + 2Cl^- \xrightarrow{MPO} 2HOCl$$

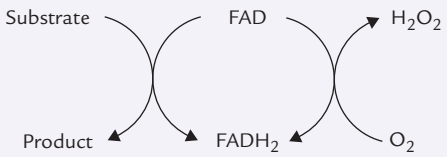
Neutrophils and monocytes are activated upon exposure to bacteria and other stimuli; *NADPH oxidase* plays a key role in the process of activation (the respiratory burst).

Both hypochlorous acid and hydroxy radicals have potent microbicidal action. Thus, they strengthen body

BOX 27.1

Generation of Hydrogen Peroxide by Flavin Coenzymes

The flavin coenzymes, FAD and FMN, are tightly bound prosthetic groups. This is in contrast to the NAD^+ and $NADP^+$ which are loosely bound to the apoenzymes. Thus, after being reduced in a dehydrogenation reaction (to $FADH_2$ and $FMNH_2$ respectively), the flavin co-enzymes cannot diffuse to a different enzyme, where they may be reoxidized. Rather they get oxidized when they transfer their hydrogen directly to the molecular oxygen resulting in the formation of hydrogen peroxide.



Amino acid oxidase (coenzyme: FMN) and *D-amino acid oxidase* (coenzyme: FAD) generate hydrogen peroxide in this manner (Chapter 13).

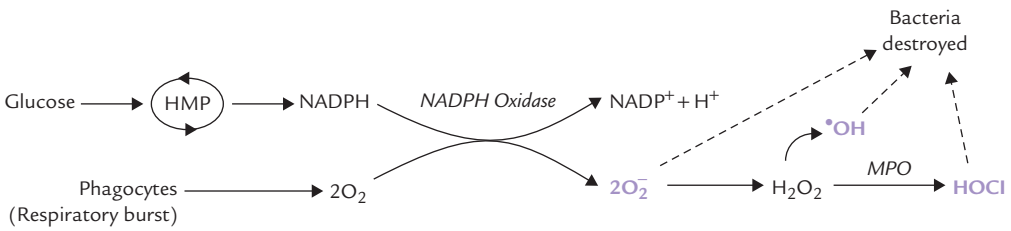


Fig. 27.2. Generation of superoxide radicals in neutrophils by *NADPH oxidase* (MPO = *myeloperoxidase*).

defences by destroying the invading microbes. If the above pathway is impaired due to some reason, the neutrophils are unable to generate free radicals. Such patients are highly susceptible to bacterial infections because although their phagocytes ingest the bacteria normally, they are not able to destroy the ingested pathogen. The condition is fatal in early life (**Case 27.1**).

Hydrogen peroxide is not a free radical: All reactive oxygen metobolites are erroneously believed to be free radicals, though technically speaking, the latter term is reserved for only those substances which contain a single, unpaired electron. Thus, superoxide is a free radical but hydrogen peroxide is not.

The respiratory burst: It has been observed that cellular consumption of oxygen is greatly increased following ingestion of bacteria by the phagocytic cells (or following interaction of these cells with some external stimuli such as membrane binding of immunoglobulins). In neutrophils, the oxygen consumption may rise up to 50-fold, i.e. the **respiratory burst**.

III. Damage Produced by Free Radicals

All types of biomolecules can be damaged by free radicals because of their extremely high reactivity; for example:

Lipids

Lipids are most susceptible to damaging effects of free radicals. Cell membranes, being rich in *polyunsaturated fatty acids (PUFAs), are especially prone to the damage*. This is because the oxidative damage to PUFA proceeds in a self-amplifying manner, which is also referred to as the chain reaction. Initial reaction of the free radical with PUFA generates more free radicals. Each free radical, in turn, generates more of free radicals.

$$RH \xrightarrow[metal]{{}^{\bullet}OH,\ H_2O} R^{\bullet} \quad (Reaction\ 1)$$

$$R^{\bullet} + O_2 \longrightarrow ROO^{\bullet} \quad (Reaction\ 2)$$

$$ROO^{\bullet} + RH \longrightarrow ROOH + R^{\bullet} \quad (Reaction\ 3)$$

$$ROOH \xrightarrow{Cu,Fe} ROO^{\bullet} + H^{+} \quad (Reaction\ 4)$$

RH = target PUFA; **R•** = initiating radical (or carbon centred radical; **ROO•** = fatty acid peroxyl radical, **ROOH** = Lipid hydroperoxide.

The chain reaction is initiated with interaction of PUFA with a free radical (O_2^{-}, $HO_2^{\bullet}$, or $^{\bullet}OH$). Removal of hydrogen atom follows to generate R•—the carbon-centred radical (***Reaction 1***). It promptly adds oxygen to form the fatty acid peroxy radical (ROO•, ***Reaction 2***). The latter is termed *carrier of the chain-reaction* since it can attack another polyunsaturated lipid molecule (***Reaction 3***). This reaction generates a lipid hydroperoxide (ROOH). Thus, net result of Reactions 2 and 3 is the conversion of R• to ROOH. The latter can break to form more of oxidizing species, e.g. peroxyl radical (***Reaction 4***) and aldehydes. Damage by these compounds is not confined to their site of production since they can diffuse to other parts of the cell.

Reaction 1 falls into the **initiation phase**; and Reaction 2–4 make up the **propagation phase** of this chain reaction.

Termination of this reaction chain occurs when a peroxyl radical reacts with another such radical to form inactive products.

$$ROO^{\bullet} + ROO^{\bullet} \longrightarrow RO{-}OR + O_2$$

$$R^{\bullet} + R^{\bullet} \longrightarrow R{-}R$$

$$ROO^{\bullet} + R^{\bullet} \longrightarrow RO{-}OR$$

Proteins

Proteins are far less susceptible than PUFA to free radical damage. Mechanism of protein damage is still not very clear. It is possible that protein conformation may be altered because of oxidation of sulfhydryl group, or modification of certain amino acid residues such as proline or methionine.

Nucleic Acids

They are also damaged by free radicals, though to a lesser extent than PUFA. DNA is undoubtedly a vulnerable target. Chain break or alterations in nitrogenous bases may lead to cell death, mutations or carcinogenesis.

Haemoglobin

Superoxide can oxidize haemoglobin to form **methaemoglobin**; about 3% of haemoglobin is converted to methaemoglobin each day. This is hazardous for the body since methaemoglobin is incapable of transporting oxygen.

> Free radicals are extremely short-lived (approximately 1×10^{-9}s), but they readily extract electrons from other molecules, converting them to free radicals and thereby initiating a chain reaction.

IV. Free Radical Scavenger Systems

Several protective mechanisms have evolved in cells that serve to minimize the toxic effects of free radicals. These free radical scavenger systems or **antioxidant defence mechanisms** can be divided into two categories:

- **Preventive mechanisms** which prevent the generation of free radicals.

- **Interceptive mechanisms** which destroy the free radicals that are accidentally generated.

A. Preventive Mechanisms

The most important preventive mechanisms against free radical generation are (a) the efficiency of electron transport, and (b) sequestration of transition metals, as discussed earlier. In addition, the peroxide decomposing enzymes, such as *glutathione peroxidase* and *catalase*, play important roles:

1. ***Glutathione peroxidase*** *(GP)*, a cytosolic enzyme, reductively eliminates hydrogen peroxide and organic hydroperoxides, thus preventing unchecked buildup of peroxides in cells. Peroxides are eliminated by reaction with reduced glutathione (GSH), catalyzed by GP.

$$2GSH + H_2O_2 \xrightarrow{GP} GSSG + 2H_2O$$

The oxidized glutathione (GSSG) is reconverted to reduced glutathione by the enzyme *glutathione reductase* (GR), using the reducing power of NADPH.

$$GSSG + NADPH + H^+ \xrightarrow{GR} 2GSH + NADP^+$$

Detoxification of H_2O_2 in this way prevents generation of free radicals from it (remembering that, by definition, H_2O_2 is not a free radical). Some types of *glutathione peroxidase* require selenium (Se) this is the reason why Se appears to have antioxidant activity.

2. ***Catalase*** is another enzyme causing decomposition of peroxide to yield water and oxygen. It is a haem-containing enzyme, present in blood and tissues (which is why hydrogen peroxide bubbles when applied to wounds). Highest concentration of *catalase* is present in *peroxisomes* (up to 40% of the total proteins). This enzyme is used when large quantities of H_2O_2 are generated.

B. Interceptive Mechanisms

These exist in both aqueous and membrane compartments of the cell and comprise an enzyme (*superoxide dismutase*) and several non-enzyme substances.

1. ***Superoxide dismutase*** *(SOD)* is the only known enzyme that takes a free radical (superoxide) as its substrate and hence acts as a scavenger. Different isoenzymes of *SOD* are described. The isoenzyme of *SOD* present in mitochondria is Mn^{2+}-dependent, and the other, present in cytosol requires copper. An extracellular copper-zinc-dependent isoenzyme has also been reported.
2. ***Non-enzyme substances*** include vitamin E, retinoids, ascorbate, ceruloplasmin, transferrin, ferritin, ubiquinone, uric acid and bilirubin.
   (a) Vitamin E and retinoids being **lipophilic**, act in biological membranes.
   (b) The others are **aqueous phase antioxidants**.
   (c) Some of these agents are also called **chain-breaking antioxidants** since they terminate the lipid peroxidation chain reactions. For instance, *α-tocopherol* (Toc OH, a vitamin E compound) does so by intercepting lipid peroxyl radical ($ROO^\bullet$).

$$ROO^\bullet + Toc\ OH \rightarrow ROOH + Toc\ O^\bullet$$
$$Toc\ O^\bullet + ROO^\bullet \rightarrow \text{Interactive products}$$

Tocopherol is the most potent chain-breaking antioxidant in tissues. The other example is β-*carotene*. It is, however, less effective.

The antioxidant action of vitamin C and β-carotene have been explained in Chapter 18. Consumption of foods rich in the above vitamin nutrients has been recommended for reducing risk of certain chronic health problems. For example, the populations that consume foods rich in vitamin C are known to have reduced risk of stomach cancer; foods containing β-carotene may offer protection against many cancers, including those of lungs, stomach, oesophagus and oral cavity.

A battery of antioxidants disposes off the free radicals; *superoxide dismutase* is considered the first-line defence. *Glutathione peroxidase*, *catalase*, and some vitamins also offer protection.

C. Others

Caffeine and *ceruloplasmin* are good antioxidants. *Vitamin A* and *cysteine* are mild, non-specific antioxidants.

Action of antioxidants is supplemented by processes that repair the damage caused by free radicals, e.g. DNA repair enzymes (Chapter 21).

V. Free Radicals in Aetiology of Disease

Free radicals have been implicated in degeneration of aging and a number of human diseases such as cancer, inflammatory disease, and reperfusion injury, and in the aging process (Table 27.1).

The list is growing and there is tendency to implicate free radicals in pathogenesis of several diseases where no other pathogenic mechanism can be elucidated. *To establish or*

Table 27.1. Disorders caused by free radicals

1. Age-related diseases	Parkinsonism, Alzheimer's disease
2. Diseases of the eye*	Retrolental fibroplasia, retinopathy of prematurity, cataractogenesis
3. Respiratory diseases	Adult respiratory distress syndrome, bronchopulmonary dysplasia in premature infants
4. Chronic inflammatory disease	Rheumatoid arthritis
5. Cancers	
6. Atherosclerosis and myocardial infarction	
7. Diabetes mellitus	
8. Reperfusion injury and shock-related injury	

* Eye tissue has highest concentration of free radical scavenging enzymes.

rule out involvement of free radicals in a disease is difficult because of their extremely short lifetime.

Specific role of oxygen derivatives in pathogenesis of several diseases is, however, established beyond doubt:

Parkinsonism

Excessive production of hydrogen peroxide by *monoamino oxidase* (MAO) has been implicated as a major factor for the neuronal degeneration in patients with Parkinson's disease. The **dopaminergic nigrostriatal neurons** that are destroyed in this disease have shown *high MAO activity*. Thus, treatment with MAO inhibitors has been found to be an effective mode of therapy in this disorder.

Atherosclerosis

Oxidized low density lipoproteins (LDL), formed by action of free radicals, are readily taken up by the macrophages, producing **foam cells**. These cells accumulate beneath the endothelial layer of the arterial wall, and this triggers onset of atherogenesis (Chapter 12).

Diabetes Mellitus

Free radicals induce destruction of the pancreatic β-cells, and this has been implicated in etiopathogenesis of the **type-1 diabetes**.

Cancer

It is firmly established that when mutations occur within DNA, cancer may arise. Free radicals cause mutations by inflicting chromosomal damage and inhibiting the DNA repair processes.

Alzheimer's Disease

It is a chronic and progressive neuro-degenerative disorder, characterized by loss of neurons in motor, sensory and cognitive systems in elderly people (onset: 7th to 9th decade of life). The neuropathological hallmarks (senile plaque and neurofibrillary tangles) were first described by *Alois Alzheimer* a German psychiatrist, in 1907. The disease has genetic predisposition, and free radicals play important role in causing the neuronal damage.

Male Infertility

Males are at a greater risk of incurring free radical damage because they tend to accumulate large body stores of iron (iron plays important role in free radical generation, as discussed earlier). These free radicals are known to reduce sperm viability and motility, hence contributing to male sterility. Women are protected till menopause.

Rheumatoid Arthritis

Excessive generation of free radicals by macrophages (respiratory burst) is one of the major causes in the pathogenesis of several inflammatory diseases, including rheumatoid arthritis.

Thus, medical importance of free radicals is widely acknowledged, and yet not entirely clear. Some consider oxygen as the most important and least avoidable environmental mutagen. Laboratory animals are known to die in an atmosphere of oxygen, and premature infants have developed retrolental fibroplasia. Unfortunately, oxygen therapy of this condition results in blindness.

Exercises

Essay type questions

1. Name some free radicals and explain how are they generated. Discuss various defensive mechanisms that tend to minimize toxic effects of free radicals.
2. Explain how free radicals damage various biomolecules. Discuss their role in aetiology of human diseases.

Write short notes on

1. Antioxidant vitamins
2. *Superoxide dismutase*
3. Transition metals
4. Respiratory burst
5. *Glutathione peroxidase*

CLINICAL CASE

CASE 27.1 A 1-year-old child with recurrent skin infections

A 1-year-old boy was brought to the hospital OPD with complaints of recurrent skin infections. Since this suggested a generalized impairment of body defences, the patient's leukocytes were isolated for investigation. They exhibited normal motility in response to the chemotactic peptides. Moreover, intracellular production of inositol triphosphate (IP_3), and the subsequent protein phosphorylation in response to these peptides was normal. However, intracellular production of hydrogen peroxide in response to these peptides was impaired. Intracellular $NADP^+$ level was also low.

Q.1. Identify the defective protein in this case.

Q.2. What is the relevance of determining intracellular IP_3 production in this patient?

CHAPTER

28

ENERGY METABOLISM AND NUTRITION

Nutrition is the science that deals with food and nutrients and the way the body uses them. It also deals with interaction of nutrients and their balance in relation to health and disease. The word nutrient refers to certain parts of food that the body uses for energy, growth and replacement of worn out structures. Thus, nutrition is the process of utilization of food by living organisms.

Human nutrition can be divided into three categories: undernutrition, overnutrition and ideal nutrition. In a developing country like India, undernutrition is the major concern, whereas in some affluent societies overnutrition is a serious problem. The concept of ideal nutrition is attracting attention presently. It explores long-term effects of nutrition on health.

A balanced intake of nutrients in quantities required by the body is essential to maintain good health. This ensures adequate growth and development as well. Carbohydrates, fats, proteins, vitamins, and minerals are some nutrients present in diet. Carbohydrates and fats are required primarily for providing energy. The dietary proteins, on the other hand, are mainly required for growth. Moreover, they provide amino acids which serve as precursors for a number of specialized products. Carbohydrates, fats and proteins are collectively referred to as **macronutrients** since they are required in relatively larger quantities, that is, to the extent of several grams per day. In contrast, vitamins and minerals are required in much smaller quantities: few micrograms to few milligrams per day. Accordingly, they are referred to as **micronutrients.**

After going through this chapter, the student should be able to understand:

- Calorific values of food materials, components of energy requirements of the body; basal metabolic rate, specific dynamic action and physical activity; recommended dietary intake of nutrients: and proximate principles of food.
- Nutritional importance of carbohydrates, dietary fibres, fats and proteins; parameters defining protein quality such as biological value, chemical score, protein efficiency ratio and net protein utilization; nitrogen balance; and protein sparing action of carbohydrates and fats.
- Disorders related to undernutrition and overnutrition.

I. Calorific Values

The chemical energy in food is released in the body by oxidation to provide for the human energy requirements. The calorific value refers to the energy content of a food material. Its estimation involves measurement, of the energy produced by the combustion of foodstuffs in calorimeter. This method of estimation is based on *Hess' Law,* which states that the energy given off in a chemical reaction is the same, no mater by which intermediate steps it is carried out. Thus, the energy liberated (as heat) in the combustion of glucose ($C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O$) is exactly the same whether it is carried out in the laboratory or in the body, so long as the end products in each case are carbon dioxide and water. Therefore, by measuring the heat released in a calorimeter on burning a given quantity of fat, carbohydrate or protein, it is possible to say what must be the energy liberated by the body consuming known quantities of the three foodstuffs. A correction, however, has to be made especially in case of

proteins, as discussed later. The calorimeter used is usually the **bomb calorimeter,** which can be filled with oxygen and ignited electrically.

In nutritional studies, the unit of energy used traditionally, has been the calorie. One calorie represents the thermal energy required to raise the temperature of 1 gm of water by 1°C (15°–16°C). In human physiology, energy is generally expressed in terms of kilocalories (kcal); one kcal is equals 1000 cal. In the International system of units, the unit of energy is kilo-joule (kJ); one 1 kJ is the energy required to move a mass of 1 kg by 1 meter distance by a force of one Newton. It is related to kcal as below:

$$1\,\text{kcal} = 4.128\,\text{kJ; or}$$
$$1\,\text{kJ} = 0.24\,\text{kcal}$$

Table 28.1 gives energy yield from various macronutrients during body metabolism. It may be observed that the energy content of fat is more than twice as that of carbohydrate and protein. The calorific values of the latter two macronutrients are nearly the same. This is because both are completely oxidized to the same end products, carbon dioxide and water, both within the body and in the bomb calorimeter (see Hess' Law). Proteins, however, are not completely burnt; one of the biological end products, urea, is different from that obtained in a calorimeter. Since urea molecule contains some amount of chemical energy, the calorific value of protein in the body (4 kcal/mole) is significantly less than that obtained in a bomb-calorimeter (5 kcal/mole 4 kcal/mole).

One gram of carbohydrates and proteins provides 4 kcal each, whereas lipids are energy dense, providing 9 kcal per gram. Energy is used to do work and to maintain integrity of the body.

Unlike these three principal foodstuffs, the other nutrients, namely vitamins and minerals have zero calorific value. However, they play a supplementary role by promoting several reactions that generate metabolic energy from the principal foodstuffs.

A. Respiratory Quotient (RQ)

This is defined as the molar ratio of carbon dioxide produced per unit of oxygen consumed during combustion of a foodstuff. For the combustion of glucose,

$$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + H_2O$$

the RQ is 6/6 = 1, which is assumed to be the case for all carbohydrates. The RQ for fats is lower because of their lower oxygen content. For the combustion of palmitic acid,

Table 28.1. Energy yield from various macronutrients during body metabolism

Macronutrient	Calorific value in kcal/g	
	In body	In bomb calorimeter
Carbohydrates	4	4.1
Fats	9	9.4
Proteins	4	5.4

$$C_{16}H_{32}O_2 + 23O_2 \rightarrow 16CO_2 + 16H_2O$$

the RQ is 16/23 = 0.71, which is assumed to be the case for all fats. For proteins, which are sums of their component amino acids that may show considerable variants, the RQ also varies accordingly; its value is generally taken as 0.8. On a mixed diet, the RQ value varies with relative proportions of different foodstuffs. For a normal Indian diet, its value is around 0.84.

RQ may be used to measure energy expenditure of an organism. This method may be referred to as *indirect calorimetry* (because it does not involve direct measurement of heat evolved by burning of the foodstuff). It is important to remember that average kcal yield per liter oxygen consumed is 4.65 when fats are metabolized (RQ = 0.71), and 5.11 per liter oxygen when carbohydrates are burnt (RQ = 1). Since the difference is small, it may be neglected, or an average value of 4.825 kcal is taken for a person on a mixed diet. An apparatus for measuring gas exchange gives oxygen consumption over a given period, and total energy spent by the individual is calculated based on the formula that each liter of oxygen consumed represents energy production of 4.825 kcal.

Energy expenditure can be determined by direct calorimetry (heat generated), or indirect calorimetry (from measurement of oxygen consumption and carbon dioxide production).

II. Components of Energy Requirements

Total energy required by an average normal person is the sum of three energy-requiring body processes. These are:

1. The basal requirement (the basal metabolic rate).
2. The requirement for diet-induced thermo-genesis (specific dynamic action of food).
3. The requirement for physical activity.

Besides the above three, extra provision of energy has to be made for growth, pregnancy and lactation.

A. Basal Metabolic Rate (BMR)

BMR is defined as the energy expended by the body when the voluntary activity is at a minimum. The subject should be awake, at complete physical and mental rest, 12-hour after the last meal, and at an equable temperature (about 25°C). The metabolic rate during sleep is less than BMR.

Why should there be any expenditure of energy in absence of physical activity? To find the answer one must envision the body as an unlikely agglomeration of essentially unstable compounds, dissolved in or surrounded by a very precise but unusual salt solution. This combination has to be maintained for survival, as a rule, above the temperature of its surroundings. This is not possible without expenditure of a good deal of energy, which is tentatively identified with BMR. Thus, *BMR reflects the energy required for maintaining the integrity of the organism and for sustaining vital functions under basal conditions*. Some such functions include pumping of blood by heart, conduction of nerve impulse, gastrointestinal motility and work of breathing. BMR also accounts for such processes as protein turnover, glycoside bond biosynthesis and maintenance of cation differences across membrane, especially the sodium-potassium gradient. A large proportion of the BMR is ascribable to the need to maintain this gradient (Chapter 7).

Normal Values

BMR is in the order of *1 kcal/kg body weight/h* for human beings. It is higher in children than in adults, a little higher in younger adults than in the elderly and a little higher in males than in females. BMR takes into account the body surface area as well, and therefore, expressed as kcal expended per square meter body surface area per hour (kcal/sqm/h respectively). Variations to the extent of ±15% of these are considered normal.

Measurement of BMR

The BMR is measured after some 12 hours of fasting in the post-absorptive state to avoid the so-called specific dynamic action: a surge of heat production after eating. The subject should be at complete rest, but awake, and in comfortable surroundings (at about 25°C). Any deviations from these conditions may alter the measured BMR to a significant extent.

Most commonly, the **Benedict-Roth apparatus** is used for the measurement of BMR. The subject is asked to breathe into the mouth piece of the apparatus and the volume of oxygen consumed over a 6-minute period is determined. This is multiplied by 10 to obtain the oxygen consumption in one hour. An approximation of how much calories of energy the subject is generating in an hour is provided by multiplying the oxygen consumption by the factor of 4.825, which represents the average kilocalorie yield per liter oxygen consumed. A person consuming 15 L oxygen per hour is generating 15 L/h × 4.825 kcal/L = 72.3 kcal/h. This multiples to **1737 kcal per day.**

BMR is generally calculated on the basis of the person's surface area, which is estimated using the ***Du Bois equation*:**

$$\log A = 0.425 \log W + 0.725 \log H + 1.8564$$

where **A** is the subject's surface area in m^2, **W** is the weight in kilograms and **H** is the height in centimeter. A 75 kg person who is 1.8 m (5.91 feet) tall has a surface area of about 1.94 m^2. In all mammals, BMR expressed in this fashion is approximately the same, **900 kcal/m^2/day**.

Factors Affecting BMR

A number of physiological and pathological conditions affect BMR.

Body size and weight: The BMR shows wide variations, taking into account individual size and weight variations. For instance, BMR in *humans* is about **32 kcal/kg/day**, whereas in a *mouse* it is **212 kcal/kg/day**.

Among humans, BMR varies with body weight, but the rise is not proportional. For instance, in males with body weight 60 kg, 70 kg, and 80 kg, the BMR is 2590, 3010, and 3710 kcal/day, respectively. This is because adipose tissue is not as metabolically active as lean body mass. In fact, *BMR is more appropriately expressed per kilogram of lean body mass or fat-free mass.*

Surface area, which is related to body weight and height, also affects BMR, as mentioned earlier. Person of lean built of a given body weight has a greater surface area compared to an obese subject of the same body weight, and therefore, BMR is higher in the lean individual.

Age: The BMR is significantly higher in infant and growing child than in adult, and a little higher in young adult than in the elderly. After 40 years of age, the BMR decreases at the rate of about 2% per decade of life (Table 28.2). Age affects BMR, largely through a decrease in lean body mass after adulthood. The decrease in BMR with advancing years is usually not accompanied by changes in eating habits, which is partly the reason for overweight and hypothermia in old age.

Table 28.2. Basal metabolic rate (in kilocalories/hour/square meter body surface area) in different age groups

	Age in years						
	1	5	15	25	40	55	80
Male	53.0	49.3	41.6	37.3	35.4	35.4	33.0
Female	53.0	48.4	37.8	35.4	34.9	33.2	31.1

Sex: The gender differences in BMR (higher in males by about 5%) usually reflect the lower proportion of lean body mass in women, when compared to men of the same body weight.

Genetic differences: The BMR may vary by up to 10% between subjects of the same sex, age, body weight and fat content.

Racial differences: The BMR in Eskimos is significantly higher compared to that in other ethnic groups. However, studies to find differences between various ethnic groups, such as Asians, Africans and Caucasians have so far proved *conflicting and inconclusive*.

Nutritional status: In starvation, there is an adaptive decrease in the BMR, over and above that which results from the decrease in lean body mass. The converse is true when energy intake is increased.

Endocrinal state: BMR is increased in hyperthyroidism because thyroid hormones have a stimulatory effect on metabolism. In hypothyroidism, BMR may decrease by up to 40%, leading to weight gain. Other hormones, such as epinephrine and growth hormones, tend to raise BMR.

Climate: BMR increases in cold climate. In the persons living in tropical climate, BMR is about 10% less than those living in temperate zones.

Drugs: Smoking (nicotine) and coffee (caffeine and theophylline) increase the BMR, whereas β-blockers tend to decrease it.

Lifestyle: In physically active individuals, notably the athletes, BMR shows substantial increase. Stress and anxiety also tend to elevate BMR, which is mainly due to increased secretion of epinephrine.

Disease states: The BMR increases in infections, burns and cancer. In fever, the BMR rises by about **11%** for every **1°C** rise in body temperature.

BMR is the minimum amount of energy required when voluntary activity is minimal, in the post-absorptive state. It averages to 36–38 kcal/sqm/h for males, 32–35 kcal/sqm/h for females; and is influenced by a number of factors including age, lifestyle, climate and thyroid functions.

B. Specific Dynamic Action (SDA)

Another component of energy expenditure is specific dynamic action of food, which refers to increased heat production after eating. It is also called **diet-induced thermogenesis** or post-prandial thermogenesis. SDA is believed be due to the *energy expended in digestion, absorption, transport and subsequent processing of food.* SDA was originally attributed solely to the metabolic handling of proteins, but it is now recognized as an effect produced by the consumption of all dietary fuels. The consumption of protein does, however, produce the greatest loss of energy (20–30% of intake), compared to fats (11–13%) or carbohydrates (5–6%). Relatively higher SDA of proteins is primarily to meet the energy requirements for oxidative-deamination, synthesis of urea and biosynthesis of proteins.

SDA is a wasteful expenditure from viewpoint of cellular energetics. Out of every 100 g of proteins consumed, the energy available for doing useful work is 20–30% less than the calculated value. For instance, assume a person taking 250 g of protein, which should produce $250 \times 4 = 1000$ kcal. But the body must spend some energy (about 20% = 200 kcal) to extract this 1000 kcal. So the net value of 25 g protein is 800 kcal (1000 minus 200).

On mixed diet, value of SDA is around 10%. The percentage of energy lost being 10% a person must take food worth 110 kcal to obtain 100 kcal. Thus, additional calories must be added during diet planning to account for SDA. Finally, SDA shows considerable individual variation and this has been implicated as an important factor, which allows some persons to maintain their normal body weight after overeating.

SDA of foodstuffs is the increase in energy (heat) output after eating. It is highest for proteins (30% of the calorific value), and lower for carbohydrates (5%) and mixed diet (10%).

C. Physical Activity

It is the third important factor that must be taken into account while calculating the energy requirements. Its value varies from one individual to another, and even in the same individual from time to time. It depends on the *occupation, physical activity* and *lifestyle* of the individual. For convenience, the activity level may be divided into three groups: **sedentary, moderate** and **heavy**.

To calculate total amount of energy for carrying out the muscular activities in 24 hours, the type and duration of all activities are carefully recorded. Table 28.3 shows energy expenditure for some of the common body activities. In general, a sedentary person requires about 30–50% more energy than the BMR requirement, whereas a highly active person may require 100% or more calories above the BMR.

Table 28.3. Energy expenditure in some common physical activities (over and above BMR)

Activity	kcal/kg body wt/hour
Fast running or dancing	9.7
Bicycling (13 miles/hour)	9.4
Swimming	8.7
Chopping wood	6.1
Walking	1.98
Brisk walk (4.5 miles/hour)	5.8
Mopping floors	3.9
Gardening, digging	3.1
Dressing	0.8
Eating	0.5
Standing	0.5
Reading/Writing	0.5
Sitting	0.4
Lying quietly	0.1

D. Total Energy Requirement of an Average Normal Person

Total energy requirement of a person is calculated by adding the above three components of energy requirement:

Physical activity + BMR + SDA

- **Physical activity:** The energy required for different grades of physical activities carried out in last 24 hours, is calculated. Consider energy requirement of an 80 kg man with the following daily activities: sitting, 8 h; lying down, 4 h; walking, 1 h, standing, 0.5 h driving, 1 h; reading/writing, 1 h; gardening, 0.5 h; cycling, 0.2 h. These activities account for 828 kcal in 24 hours (the student is advised to refer to Table 28.3 and calculate).
- **BMR:** The energy requirement for physical activity is added to BMR. The BMR may be calculated from the approximation:

  BMR = 1 kcal/kg/h

  This equals $1 \times 80 \times 24 = 1920$ kcal. Adding the two numbers, 828 and 1920, we get a total of 2748.
- **SDA:** The above value of 2748 equals calorie requirement excluding the SDA. A 10% surcharge is added to take into account the SDA. This amounts to about 274 kcal to give a **grand total** calorie requirement of $274 + 2748 = \mathbf{3022\ kcal/day}$.

Daily energy demands of most adult men vary between 2400 kcal (in sedentary workers) to 4000 kcal (in heavy workers). The corresponding figures in women are 1900–2600 kcal per day. Table 28.4 shows recommended kcal allowances for individuals of different age groups and body weights.

Table 28.4. Recommended kcal allowances per day

	ICMR recommendation
Man (55 kg)	
Sedentary work	2400
Moderate work	2800
Heavy work	4000
Women (45 kg)	
Sedentary work	1900
Moderate work	2200
Heavy work	2600
Later half of pregnancy	+300
Lactation	+700
Infants	
0–6 months	120/kg body wt
7–12 months	100/kg body wt
Children	
1–3 years	1200
4–6 years	1500
7–9 years	1800
10–12 years	2100
13–15 years	
Boys	2500
Girls	2300

Energy requirement of an individual is calculated taking into account the basal metabolic rate, the specific dynamic action and physical activities.

As noted earlier, an extra provision of energy must be made during **infancy, childhood** and pre-pubertal growth spurts (during 10–15 years of age). This is because biosynthesis of new tissues occurs during these stages of life, which requires input of considerable chemical energy. It has been estimated that laying down of 1 kg of new tissue in a year requires about 81 kcal/day extra energy. Likewise, an extra provision must be made during **pregnancy** to meet the requirements of the growing fetus. During **lactation**, an extra provision for synthesizing milk is needed; elaboration of one liter of milk (containing 642.8 kcal) requires at least 1285 kcal of energy.

III. Recommendations for Dietary Nutrients

Amount of nutrients required to meet the physiological needs of an individual is termed ***recommended dietary allowance (RDA).***

Table 28.5. Recommended daily allowance (RDA) of proteins in different age groups

Nutrient	Requirement per day per kg body weight
Proteins	
Males	1 g
Females	1 g
Children	
Infants	2.4 g
Up to 10 years	1.75 g
Boys (> 10 years)	1.6 g
Girls (> 10 years)	1.4 g
Pregnancy and lactation	
Pregnancy	2 g
Lactation	2.5 g

The Indian Council of Medical Research has prescribed RDAs for all essential nutrients, after suitably modifying the recommendations of WHO and FAO for Indian conditions.

RDA of a given nutrient must be differentiated from the amount that must be consumed merely to avoid deficiency disease. RDA refers to the amount that is required for optimizing health, and not the amount merely to avoid the disease. For example, RDA for vitamin C is about 60 mg, but the amount required for preventing scurvy is only about 10 mg/ day. Intake of a nutrient at or near RDA provides considerable safety margin against the development of deficiency disorders.

RDA for proteins is 70 g for a 70 kg adult man. The same individual would require about 80 g fats and 365 g carbohydrates each day. RDA for the micronutrients is much smaller, for example, RDA for vitamin E is 10 mg/ day of α-tocopherol equivalents (see Chapter 18).

Factors Affecting RDA

Age, sex and several other factors determine RDA (Table 28.5). For example, infants require about 2.4 times as much proteins per day as required by adults to support growth; males require greater amount of a given nutrient since their body mass is more; and RDA is enhanced in physiologically stressful states such as *pregnancy* and *lactation* and in patients with *injury* or *illness*. However, certain exceptions do occur; for example, RDA for iron in women is more than in men since women must replace the amount lost during menstruation.

Recent studies have indicated that for optimal health, it is sometimes necessary to consume certain nutrients in amounts larger than RDA. For example, daily intake of vitamin E in amounts three times the RDA significantly reduces risk of coronary artery disease. Conversely, excessive intake of certain nutrients is hazardous, for example, risk of coronary artery disease is enhanced if fats are taken in excessive amounts.

IV. Proximate Principles of Food

Food plays a much wider role than merely providing energy and biochemical needs of life. In fact, culture, availability, economics, religion, fads and several other factors play an important role in determining the type of food consumed in a given society. Agricultural societies, including those in India, subsist mainly on cereals. However, in non-agricultural societies like Eskimos of Greenland and hunters of Africa, meat is the major food. In such foods, nearly one-third of the total energy is provided by proteins and a major contribution for the rest comes from animal oils and fats. This is in sharp contrast with the average **Indian diet** where carbohydrates, proteins and fats account for about 75–80%, 10–12% and 10–15%, respectively, of the total energy provided. In between these two extremes lies the continental diet of the affluent Western societies. In a typical **continental diet**, contribution by carbohydrates is much less (*40–45%*), whereas fats and proteins account for more energy as discussed later.

V. Nutritional Importance of Carbohydrates

Dietary carbohydrates are major components of an Indian diet, contributing up to 60–70% of the total caloric requirements of the body. However, from nutritional viewpoint they are not essential for human beings as all carbohydrates can be synthesized in the body. They are regarded as a relatively cheap source of energy: the carbohydrate-rich foods cost less. There are two groups of carbohydrates in the diet: **available carbohydrates** that can be assimilated and utilized for energy production: and dietary fibres that are indigestible, and therefore constitute **unavailable carbohydrates** in the diet.

A. Available Carbohydrates

Some important available carbohydrates present in diet are *starch, sucrose, lactose, fructose* and *glucose*. Of these, **starch** is the most abundant and cheapest source of energy, being present in most commonly available foods such as cereals, pulses, tubers, etc. On cooking it becomes more soluble and accessible to digestive enzymes. *Amylase* hydrolyzes the α(1→4) linkages of starch to form maltose, and

oligosaccharide units, which are further hydrolyzed to glucose by the brush-border enzymes (Chapter 26). Glucose is the major source of energy for most organs and tissues; erythrocytes, brain and other parts of central nervous system depend almost exclusively on glucose as a fuel.

Lactose is present in milk, and is therefore the major carbohydrate for the breast-fed infants. **Sucrose** imparts sweetness to the diet because of its excessive sweetening property. However, the same property restricts its consumption; the daily intake may be as low as 4 g/day. It has been observed in experimental animals that feeding of excess sucrose results in hypercholesterolaemia and hypertriglyceridaemia. In humans, it predisposes to the development of dental caries. In view of these observations, only a limited intake of sucrose has been suggested; and avoided altogether in diabetic individuals and in those attempting weight reduction. Jaggery, a sweetening agent, is better since besides sucrose it provides iron as well.

Rates of assimilation of different saccharides differ to a significant extent. This is indicated by the observation that following consumption of these saccharides different responses are obtained in the

(a) post-prandial blood glucose elevation, and
(b) insulin release.

The above responses are lower with the complex carbohydrates (for example starch) than with the simple carbohydrates (glucose or fructose). The elevation of blood glucose level following fructose intake is less than the elevation seen after intake of an isocaloric amounts of glucose. Lactose is least effective in this regard, causing the least elevation of the glucose response curve.

Functions

Carbohydrates play a major role in body's energy balance and are involved in a number of other vital functions, discussed earlier chapter. To summarize a few important ones, carbohydrates are required for the synthesis of fats and non-essential amino acids, for oxidation of fats (acetyl CoA, a product of fatty acid oxidation combines with oxaloacetate, a product of carbohydrate metabolism), and have protein sparing action (discussed later).

Carbohydrates are the most abundant dietary constituents, and yet they are not essential nutrients.

B. Dietary Fibres

Fibres, an important component of the diet, comprise plant cell components that cannot be broken down by human digestive enzymes. It is, however, incorrect to assume that fibres are completely indigestible, since some of them are at least partially broken down by intestinal bacteria. Fibres are complex carbohydrates having varying degrees of solubility. The more **insoluble fibres** include *cellulose* and *lignin*; **soluble** ones include pectins and gums; and, **partly soluble** ones are *arabinoglycans* (mucilage).

Physiological Significance

Being largely indigestible, the dietary fibres do not provide any energy. Yet, certain invaluable health benefits are provided by dietary fibres, as discussed here:

1. ***Increased bowel motility:*** Fibres can absorb significant amount of water because of their predominantly hydrophilic nature. Thus, the ingested fibres attract large quantity of water into intestinal lumen, which results in increased bowel motility. This is especially useful in constipation and in patients with haemorrhoids, diverticulosis and colon cancer.
2. ***Elimination of toxic compounds:*** Fibres can bind various toxic compounds, including certain carcinogens and bacterial toxins, and eliminate them through fecal route.
3. ***Cholesterol lowering effect:*** The binding properties of fibres enable them to absorb organic substances such as cholesterol and eliminate them in faeces, so as to lower plasma cholesterol concentration.

   Further, fibres bind bile salts and reduce their enterohepatic circulation. This in turn enhances cholesterol to bile salts conversion and promotes its disposal from the body.

   Water-soluble fibres act by additional mechanisms: they lower serum cholesterol levels either due to their effect on insulin levels (insulin stimulates cholesterol synthesis and export), or other metabolic effects (perhaps caused by end products of partial bacterial digestion).

   Thus, high fibre diet is recommended in patients with hypercholesterolaemia.

Cellulose and lignin are the more insoluble among fibres (the plant cell components that cannot be digested by gut enzymes), and are beneficial with regard to colonic functions. The more soluble, gums and pectins, have been associated with the lowering of blood cholesterol.

4. ***Anti-hyperglycaemic effect:*** Fibres form a vicious gel in stomach and intestine to slow the rate at which various nutrients, most importantly carbohydrates, are digested and absorbed from intestine. Thus, the rise in blood glucose as also the subsequent rise in insulin levels are significantly impaired if fibres are ingested along with carbohydrate-containing foods. This accounts for their utility in diabetes mellitus.

Table 28.6. Dietary fibres

Fibre	Chemical nature	Physiological effect
Cellulose	Polymer of glucose	Increases weight and bulk of faeces, promotes colonic peristalsis, decreases fecal transit time
Hemi-cellulose	Polymer of pentoses, hexoses and uronic acid	Retains water to increase bulk of faeces, increases bile acid excretion
Lignin	Aromatic alcohol	Antioxidant, hypocholesterolaemic
Pectin	Polymers of galactose, galacturonic acid, rhamnose and arabinose	Improves glucose tolerance in diabetes, increases bile acid excretion
Mucilage	Branched arabinoglycan	Increases bile acid excretion, hypocholesterolaemic

5. ***Satiety effect:*** Fibres significantly increase bulk of the diet, so that one gets a feeling of fullness after ingestion. This bulk-enhancing property gives a feeling of satiety even without consumption of excess calories.

High-fibre diet has several health benefits, including reduced incidence of cancer of colon, cardiovascular disease and diabetes mellitus.

Chemical nature and physiological effects of some dietary fibres are given in Table 28.6.

Adverse Effects

Excess fibre ingestion is not unproblematic as it may lead to several complications:

1. Fibres provide binding sites for **divalent metals**, calcium, iron and zinc, making these metals less bio-available.
2. They decrease intestinal protein absorption.
3. Some fibres are degraded by intestinal bacteria, causing flatulence and abdominal discomfort.

VI. Nutritional Importance of Fats

Fats are considered as richer source of energy compared to carbohydrates and protein since they provide more energy per unit mass (Table 28.1). From energy perspective, the most important dietary fat is triacylglycerol (TAG) since it constitutes more than 90% of the total dietary lipids. The TAG may contain saturated, monounsaturated and polyunsaturated fatty acids (PUFA). Smaller amounts of other lipids, e.g. phospholipids, glycolipids, cholesterol, etc. may also be present in foods.

In Western countries, the percentage of calories derived from fats is high (about 40°C) compared to developing countries (about 15%). Regarding Indian conditions, the fat intake is highest in Punjab and lowest in Orissa and Kerala. Current recommendations are—a maximum of 35% of the energy intake as fats. These recommendations, which are aimed primarily at influencing cholesterol metabolism, further suggest intake of less than 300 mg/day of cholesterol; 30 gm/day of PUFAs, and only 10% of the total energy intake as saturated fatty acids.

Dietary fat content of some common food items are shown in Table 28.7. The dietary fats may be divided into two types: visible fats and invisible fats. The visible fats are oils, butter, ghee, etc. which are pure lipid forms. The invisible fats are part of other food items, e.g. egg, fish, meat, cereals and oil seeds. Recommended daily intake of visible fats is 10% of total calorie intake or about 20 g/day.

Functions

In addition to providing energy, fats increase the palatability of food and are mainly responsible for the feeling of satiety after meals. They constitute the favoured, cooking medium all over the world. Dietary fats are also required for the intestinal absorption of fat-soluble vitamins (A, D, E and K). Moreover, the dietary fats provide the polyunsaturated fatty acids (PUFAs) that are required for various body functions.

Role of PUFAs

1. They are components of phospholipids and form biomembranes.
2. They are required for esterification of cholesterol.
3. They serve as precursors of prostaglandins and leukotrienes.
4. They have hypocholesterolaemic effect and, therefore, offer protection against atherogenesis.

The ω-6 bring down the plasma LDL level. However, they lower plasma levels of HDL as well, thus reducing the cardioprotective effect. The ω-3 PUFAs reduce serum triglyceride levels. Moreover, they increase production of TXA_3, which is less thrombogenic than TXA_2. Thus, consumption of the ω-3 PUFAs decreases platelet aggregation, which accounts for their anti-thrombogenic properties. The monounsaturated fatty acids also lower the serum cholesterol levels when substituted for the saturated fatty acids. An additional advantage is that they do not lower HDL levels.

Table 28.7. Macronutrient content of some common food items and their calorific values

Food item	Carbohydrate (g/100 g)	Fat (g/100 g)	Protein (g/100 g)	Energy (kcal/100 g)
Rice	78.25	0.50	6.9	345
Wheat	71.2	1.70	12.10	351
Maize	66.2	3.60	11.10	341
Bengal gram	60.9	5.30	17.10	350
Soya bean	20.9	19.50	43.20	430
Dry peas	56.5	1.10	19.70	315
Groundnut	20.3	40.1	26.70	548
Egg (hen)	2.0	13.3	13.3	177
Fish	0.2	0.6	22.6	91
Mutton (muscle)	Traces	13	18.5	196
Chicken (roasted)	1.25	8.7	25.8	182
Milk (cow)	4.40	4.10	3.20	69
Milk (buffalo)	5.10	8.80	4.30	117
Milk (human)	7.50	3.40	1.10	66
Cheese	6.1	25.1	20.50	332
Potato	22.6	0.1	1.60	98
Sweet potato	28.2	0.3	1.20	123
Tapioca	38.7	0.20	0.70	161
Banana	37.0	0.50	1.10	156
Mango	11.8	0.50	0.70	52
Papaya	9.2	0.10	0.60	42
Apple	13.3	0.10	0.30	55
Cabbage	4.6	0.10	1.80	22
Carrot	10.6	0.20	0.96	49
Cucumber	2.5	0.10	0.40	13
Mushroom	5.1	0.70	3.80	41

However, in recent times various hazardous effects following over-consumption of PUFAs have been reported. Excess PUFA may lead to production of free radicals that may be injurious to the cell unless antioxidants are available. It is, therefore, advisable to consume a proper mixture of fats, containing a balanced proportion of saturated and unsaturated fatty acids. The present recommendation is that not more than 30% of the total fats should be in the form of PUFA.

There has been considerable debate regarding utility of fats from plant sources vis-à-vis those from animal sources. The animal fats generally contain a higher proportion of saturated fatty acids (with the exception of fish which contains mostly unsaturated fatty acids), whereas the fatty acids from plants are mostly unsaturated (with the exception of coconut oil which contains mostly saturated fatty acids). Finally, though vegetable oils are rich in PUFA, they are mostly unpalatable because of unpleasant odour and unpleasant taste. Free fatty acids and substances responsible for bad odour, taste and rancidity are, therefore, removed to yield refined oils. Vegetable oils are refined by treating them with steam, alkali, etc.

VII. Nutritional Importance of Proteins

Recommended daily intake of proteins is 1.0 g/kg body weight. Dietary protein intake is considered adequate if it supplies up to 12% of body's energy needs. In children, the protein requirement is higher because of a higher growth rate. The protein requirement is higher in both, pregnancy and lactation (Table 28.5).

Dietary proteins are needed for:

1. Body growth, and repair and maintenance of tissues.

2. Replenishing the amino acids lost in urine, faeces, saliva, skin, hair and nails.
3. Replenishing amino acids in the (amino acids) pool, after these amino acids are used up for the synthesis of non-protein nitrogenous compounds and for other bodily activities.

There are minimum daily requirements for each of the major constituents of the diet: 150 g for carbohydrates (to avoid ketosis), 100 g for fats (to provide sufficient linoleic acid and fat-soluble vitamins and to limit the overall bulk of diet) and about 50 g for proteins (corresponding to a daily nitrogen loss on a protein-free diet).

Excess dietary intake of proteins, or low intake of carbohydrates and fats results in protein degradation for obtaining energy.

A. Protein Quality

Proteins from different sources differ in their ability to support growth and to maintain and repair the body tissues. This ability is an indicator of quality of protein.

Protein quality depends on amino acid composition, digestibility and efficiency with which the amino acids are absorbed and subsequently incorporated in the body proteins. Some of the parameters used to define protein quality are **biological value**, **protein efficiency ratio** and **chemical score**.

Biological Value (BV)

It is the percentage of absorbed nitrogen retained in the body

$$BV = \frac{\text{Nitrogen retained}}{\text{Nitrogen absorbed}} \times 100$$

Table 28.8. Protein quality of some foods items

Food item	Chemical score	Protein efficiency ratio	Biological value
Egg	100	4.5	96
Cow milk	65	3.0	84
Meat	70	2.8	80
Fish	60	3.0	85
Rice	60	2.0	64
Wheat	42	1.7	58
Bengal gram	44	1.7	58
Ground nut	44	1.7	54
Soya bean	57	1.7	64

Biological value is most commonly used parameter for assessing protein quality.

BV is an empirical measure of the efficiency of protein as a supply of essential amino acids. For example, biological value of egg protein is very high (96) because it contains an adequate quantity of all essential amino acids and more importantly, it supplies these amino acids in the needed proportions.

Generally speaking, *the animal proteins have higher biological value* (Table 28.8). However, gelatin, prepared from animal collagen, is an exception since it lacks several essential amino acids. Consequently its biological value is low.

The proteins obtained from plant sources have lower biological value since they lack one or other essential amino acid. The missing amino acids are called the **limiting amino acids.** This limitation of plant foods is overcome when these foods are consumed in combinations. For example, wheat is deficient in lysine but contains other essential amino acids in adequate amounts. Kidney beans contain lysine in adequate amounts, though deficient in methionine. When these two food items are combined, the body gets the required supply of both methionine and lysine.

Thus, dietary combinations of higher biological value are designed by appropriate combination of foodstuffs. This is referred to as the **mutual supplementation of proteins.** The following combinations are commonly seen in Indian diets: rice + pulses; wheat + pulses; soya bean + wheat; and cereals + legumes + curd. Another way to improve biological value is addition of the limiting amino acid to a foodstuff, e.g. lysine can be added to wheat flour. This process is called **fortification** with (limiting) amino acid. Likewise, vegetable proteins can be fortified with their limiting amino acids: lysine, methionine, threonine and tryptophan. The amino acids for fortification are commercially manufactured at economic prices and then mixed in appropriate proportions to foods.

A mixture of different incomplete proteins (deficient in different amino acids) can be as satisfactory as a diet containing abundant complete proteins.

Protein Efficiency Ratio (PER)

It is a less commonly used parameter for assessing the protein quality. It is defined as gain in body weight for each unit weight of protein ingested. Apparently, the amount of dietary protein used for maintenance is ignored.

$$PER = \frac{\text{Gain in body weight (g)}}{\text{Protein fed (g)}}$$

Eggs, milk and meat have high PER (2.5) and legumes, lentils, cereals and nuts have PER between 0.5 and 2.5. PER of gelatin is less than 0.5.

Chemical Score

The essential amino acid content can also be expressed in terms of chemical score. It is calculated as here:

$$\frac{\text{mg of limiting amino acid in one gram test protein}}{\text{mg of same amino acid in one gram of reference protein}} \times 100$$

The egg protein is taken as the reference protein; its` chemical score is 100.

Digestibility Coefficient (DC)

It refers to the percentage of ingested protein absorbed into the bloodstream (after digestion) in the gastrointestinal tract.

Net Protein Utilization (NPU)

This is a measure of utilization of dietary protein for synthesizing body proteins, calculated by the formula:

$$\text{NPU} = \frac{\text{Nitrogen retained}}{\text{Intake of nitrogen}} \times 100$$

NPU is a better index than biological value since it takes into account the digestibility factor also. In fact, it depends on both digestibility coefficient and biological value.

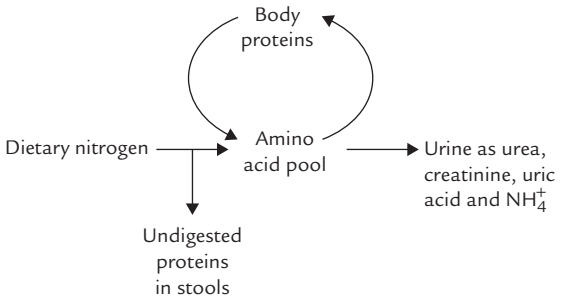
$$\text{PU} = \text{DC} \times \text{BV}/100$$

Nutritive value of proteins is assessed by various parameters such as biological value, protein efficiency ratio, net protein utilization, etc.

B. Nitrogen Balance

Nitrogen balance refers to comparative amounts of the nitrogen entering the body and that leaving it. A normal adult with adequate protein intake should be in **nitrogen equilibrium**: the amounts of incoming and outgoing nitrogen are equal. Since the exclusive source of nitrogen in the body are dietary proteins (16% of which is nitrogen), the nitrogen balance studies are used for assessing utilization of dietary amino acids for synthesis and for studying the balance between anabolic and catabolic processes.

A 100 g sample of dietary protein contains approximately 16 g nitrogen, of which, approximately 83% leaves the body as urea, 7% as ammonium ion, and 10% as organic waste products such as uric acid and creatinine. Most of this nitrogen is eliminated via urine but 1–2 g of nitrogen from undigested protein are excreted in stools also. These observation are diagrammed here.



In addition to the state of nitrogen balance in a normal healthy body, there are two abnormal states, **positive** and **negative-nitrogen balance.**

A **positive nitrogen balance** is observed when nitrogen intake exceeds nitrogen excretion, it implies that amount of body protein increases. It is seen in situations where tissue growth occurs; for example, in *growing children, convalescing adults and in pregnant women*.

Negative nitrogen balance, on the other hand, reflects nitrogen excretion exceeding the nitrogen intake. It occurs when the dietary protein intake is inadequate, or when some physiological stress is encountered, such as *trauma, burns,* or *emaciating illness*.

C. Protein Sparing Action

Dietary intake of carbohydrates and fats indirectly influences protein requirement of an adult. If dietary consumption of carbohydrates and fats is inadequate, some dietary protein must be used for energy generation. Thus, protein becomes unavailable for building and replacing tissues; consequently, dietary requirement of protein increases to make up this loss. Conversely, when sufficient calories are obtained from carbohydrates and fats, protein is no longer required for providing energy, and therefore it can be used for building up body proteins. The proteins are thus spared by carbohydrates and fats; this action is referred to as protein sparing action.

Carbohydrates are more efficient in sparing proteins than fats, probably because carbohydrates can be used by almost all tissues.

D. Reasons for Increase in Protein Requirement

- **In active growth**, more dietary proteins are required to supply amino acids to the growing tissue. Pregnancy, infancy, childhood and adolescence are some examples.
- Prolonged illness, trauma and surgery cause a major **catabolic response** to meet the increased energy needs of the body. This response is mediated by epinephrine, glucocorticoids, glucagon and certain cytokines which

Table 28.9. Balanced vegetarian diet for an adult male

	Sedentary	Moderate	Heavy
Cereals	290	340	450
Pulses	100	120	120
Leafy and other vegetables	200	200	200
Roots and tubers	200	225	250
Fruits	40	4040	
Milk	200	200	200
Fats and oils	35	3540	
Sugar and jaggery	35	3545	
Ground nut	10	1040	
Sprouts	50	5030	
All values in grams.			

are secreted in increased amounts. Breakdown of body proteins occurs, which must be replaced by dietary intake so as to prevent negative nitrogen balance.

- In **old age** also, increased dietary protein intake is required. This appears paradoxical as the requirement should fall since there is little growth. However, older people need and generally consume fewer calories. Therefore, proteins should provide a larger percentage of total calories. Further, absorption of dietary proteins may be less due to digestion and absorption problems, common in old age.

VIII. Balanced Diet

A balance diet is the one that has all the macronutrients and micronutrients recommended for individual's age, sex, occupation and lifestyle. It supplies all the nutrients in amounts slightly more than the minimum requirement to withstand any stressful state. Composition of balanced diet differs depending on socio-cultural background, economic status, availability of foods, etc. Keeping in view the commonly available foods, in India (cereals, pulses, vegetables, roots & tubers, milk, etc), the composition of balanced diet for a vegetarian adult male, with sedentary lifestyle and moderate and heavy lifestyle, are formulated as shown in Table 28.9.

Special allowances should be made in the formulation of balanced diet during physiologically stressful states, such as *pregnancy* and *lactation*.

Pregnancy: Fetal growth, tissue differentiation, and organogenesis requires energy, which is met by mother's nutritional resources. Hence, marked increase in requirement of energy and protein, in addition to calcium, iron, folate and iodine is recommended.

Lactation: Like pregnancy, requirement of various nutrients increases during lactation. However, vitamins A and C are required additionally, but requirement of iron reverts to normal.

Table 28.10. WHO classification of protein-energy-malnutrition

	Deficiency in weight for height	% of body weight compared to standard weight	Oedema
Kwashiorkor	+	60–80	+
Marasmus	+	< 60	Nil
Marasmic kwashiorkor	+	< 60	+

IX. Nutritional Implications of Diseases

Nutritional disorders can be classified into two broad categories: those arising due to undernutrition and those resulting due to overnutrition. The affluent urban elite of India are becoming more prone to the overnutrition-related health problems.

A. Undernutrition: Protein Energy Malnutrition (PEM)

The dietary resources in the developing countries are limited, therefore, protein energy malnutrition is a common problem. Infants and children are the worst sufferers. The signs and symptoms of PEM vary from case to case depending on the degree of inadequacy. However, it is a common practice to divide all cases of PEM into two major categories: **marasmus** and **kwashiorkor**.

- *Marasmus* results from an inadequate intake of both proteins and energy.
- *Kwashiorkor*, on the other hand, results due to inadequate intake of proteins with relatively adequate energy intake. It is encountered in children between 1 and 4 years of age (marasmus usually occurs under one year of age).

In India, marasmus is the most frequent presentation of severe PEM, pure kwashiorkor is seldom seen. Mostly one comes across *marasmic kwashiorkor* in which there are clinical features of both the disorders (Table 28.10).

Marasmus

It is likely to occur among poor families, when there is a rapid succession of pregnancies, and early and often abrupt weaning, followed by inadequate artificial feeding. Thus, the diet is low in both calories and protein.

Table 28.11. Comparative features of marasmus and kwashiorkor

	Marasmus	Kwashiorkor
1.	Diet with inadequate calories	Diet **insufficient in proteins and calories**
2.	Usually seen in early weaning, or infants never breast-fed (under an year of age)	Usually seen in late weaning (1–4 years of age)
3.	Oedema not present	Oedema characteristically present
4.	Severe cachexia is seen	Body weight decreased or normal
5.	Serum albumin normal or slightly decreased	Hypoalbuminaemia present
6.	Somatic protein compartment depleted	Visceral protein compartment depleted
7.	**Face shrivelled** and monkey-like	**Face puffy,** protuberant abdomen
8.	Voracious appetite	Impaired appetite (anorexia)

A child suffering from marasmus *looks like a little old person with skin and bone*. He has a thin, wasted appearance and weight is less than the desirable weight for his age. If the dietary deficiency continues for long, the developmental milestones get delayed and the condition ultimately leads to permanent stunting of growth and development. Not only the physical development, but the mental development is also affected (Table 28.10).

Kwashiorkor

In this type of malnutrition, the child has a *plump appearance due to oedema*. Oedema results from lack of adequate plasma proteins that normally maintain the distribution of fluids between the intra- and the extra-vascular compartments. Protuberant abdomen (due to oedema) is often seen, which is a deceptive presenting feature in view of muscle wasting and stunted growth of the child. Other symptoms associated with kwashiorkor are brittle hair, diarrhoea, dermatitis of various forms, and retarded growth. *In contrast to early weaning in marasums, late weaning, often extending over 2 years, is characteristic of kwashiorkor.*

Comparative features of marasmus and kwashiorkor are given in Table 28.11.

PEM exists either as the oedematous protein deficiency (from **kwashiorkor**), or in the overall nutritional deficiency from **marasmus,** among infants and young children particularly belonging to under-developed countries. They result from inadequate diets (after weaning) often compounded by the effects of recurrent infections.

Both marasmus and kwashiorkor have similar devastating results. There is reduced ability of the afflicted individuals, to ht off infection in both conditions. Mortality is very high, the underlying cause of death is mostly some infection rather than starvation itself. Increased susceptibility to infection is due to decreased immunological response; reflected by: Various components of immune system affected are as follows:

- The number of T lymphocytes is decreased, resulting in decreased cell-mediated immune response.
- Generation of phagocytic cells, immunoglobulins, interferons and other components of immune response is diminished.

Biochemical Changes in PEM

A number of biochemical parameters are altered in the following conditions:

- Decreased plasma levels of **serum albumin** and **other serum proteins** are most commonly seen. In kwashiorkor, the serum albumin may fall to a value as low as 2 g/dl, whereas in marasmus it need not be so low.

  The level of **retinol-binding protein** (RBP) is also characteristically lowered.
- **Glucose tolerance** is often normal in both the conditions, but hypoglycaemia is often seen in a marasmic child.
- In kwashiorkor, blood urea may decrease and excretion of creatinine in urine is reduced due to loss of muscle mass.
- The cellular immunity tests are also affected.
- Features of associated deficiency of vitamins and minerals (mostly iron) are commonly seen.
- Hypokalaemia and dehydration occur when there is diarrhoea.

Treatment

Treatment of PEM comprises three components:

1. Administration of adequate calories (and proteins in kwashiorkor).
2. Control of infections.
3. Treatment of associated disorders mentioned above.

B. Overnutrition

Diseases due to overnutrition are becoming more prevalent all over the world, especially among city dwellers. These diseases are influenced by **dietary excess** and **dietary imbalances**. For example:

- High calorific and high fat diet is associated with ischaemic heart disease—the biggest killer in urban settings. High fibre diet is beneficial for such individuals.
- Saturated fatty acids consumption is associated with increased risk of certain cancers, especially cancer of colon, breast and prostate.

Obesity

Obesity is a state in which excess fats (triacylglycerols) accumulate in the body. Increased number and/or size of adipocytes occurs in obesity.

Common Causes

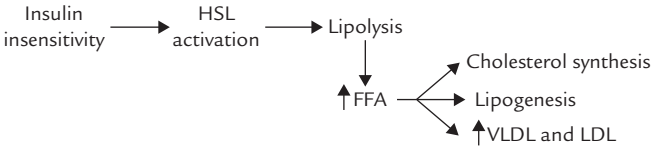
One of the most serious problems among the urban affluent populations is consumption of **excessive calories** with **inadequate exercise,** of which obesity is the natural consequence. But obesity is not a simple problem of **excessive calorie consumption**; several other factors contribute to its development.

- **Genetic predisposition** plays an important role. Genetic predisposition has been suggested based on the familial incidence of obesity. If one parent is obese, 40–50% chances are there that the children would also suffer from obesity. However, no single gene is responsible for obesity.
- **Leptin** (Greek: *Leptos* = thin) is a 16-kD polypeptide, encoded by the *obese* genes in adipocytes, that influences appetite control system in the brain. These proteins cause decreased food intake, thus representing "satiety signal" in the body. Defect in the leptin-coding gene is seen in ob/ob mice, who tend to overeat and develop obesity. When injection of leptins is given to these animals, they eat less and lose weight.
- Certain **metabolic aberrations** are commonly associated with obesity. *Hypothyroidism, hypogonadism, Cushing's syndrome* and *hypopituitarism* may lead to obesity. Women are more prone to become obese during puberty, pregnancy, and after menopause.

Biochemical Profile

The most commonly affected parameters are the **lipid parameters**. Elevation of serum concentration of free fatty acids, cholesterol and triacylglycerol commonly occur. These parameters poorly respond to dietary restrictions. The underlying cause of the altered lipid profile can be traced to decreased sensitivity of peripheral tissues (including adipocytes) to insulin. The number of insulin receptors on cell surface is decreased (e.g. down-regulation) so that these tissues respond poorly to insulin, even though plasma insulin level is elevated. Since insulin inhibits activity of the *hormone-sensitive lipase* within adipocytes, this inhibition is impaired in obesity. Consequently, activity of the *hormone-sensitive lipase* increases, resulting in mobilization of fatty acids from adipocytes. This leads to increase in circulating free fatty acid levels.

The excess fatty acids are carried to the liver and degraded to acetyl CoA through β-oxidation. Since acetyl CoA is a precursor for cholesterol and fatty acids (and hence triacylglycerols), increased generation of these molecules results. The excess cholesterol and triacylglycerols are released (after incorporation in very low density lipoprotein particles) leading to high circulating levels of these lipid components.



The major ill-effects of obesity are reduced lifespan and *coronary artery disease* due to the prevailing alterations of lipid profile. Obese individuals are at higher risk of developing ***diabetes***; *nearly 80% of the adult onset diabetic individuals are obese*. The underlying cause of diabetes in obese people is decrease in number of insulin receptors and decreased sensitivity to insulin. Moreover, decreased sensitivity to insulin results in hyperglycaemia and hence a persistent stimulus for the pancreatic β-cells. **Hyperinsulinaemia** is the natural consequence. Hyperinsulinaemia in turn causes stimulation of sympathetic nervous system, which results in vasoconstriction and retention of sodium and water. Both these changes lead to ***hypertension***.

The obese individuals are at increased risk of developing hypertension, diabetes and coronary heart disease.

The best ***treatment*** for the obese individuals—in fact the only effective mode of treatment—is **reduction of body weight**. All the aforementioned metabolic changes get significantly reversed if the ideal body weight is attained.

This goal can be accomplished by reducing the intake of calories and performing controlled exercise. Frequent small meals with lots of vegetables have been found especially effective. Fat restricted diet may retard the ageing process also and thus prolong the lifespan.

Diet and Coronary Heart Disease (CHD)

Coronary heart disease is a potentially lethal condition; the mortality is very high if timely diagnosis is not made and proper treatment is not initiated. The underlying defect in CHD is deposition of **atherosclerotic plaques** on the inner surface of the small- and medium-sized arteries. A plaque consists of *deposits of cholesterol, cholesterol esters, and cellular debris*. As the disease progresses, these deposits reduce or even block the blood flow. The blood supply to the peripheral tissues is impaired resulting in inadequate supply of oxygen and nutrients to these tissues. The consequences depend on the blood vessel that is affected:

- When the blockage occurs in the coronary vessels, damage to myocardium results. The cardiac tissue may become non-functional, and the condition is termed *myocardial infarction* (MI), commonly referred to as heart attack.
- Blockage in cerebral blood vessels may result in stroke.
- Risk of developing *thrombosis* is also greatly enhanced in atherosclerosis.

Several dietary factors are thought to influence the incidence of coronary artery disease, most notably cholesterol. As discussed in Chapter 12 **LDL cholesterol is bad cholesterol, whereas HDL cholesterol is cardio-protective**. Relatively higher HDL levels are seen in women before **menopause**, which may account, in part, for lower incidence of CHD in them.

Bearing in mind the above correlations, the following measures are advised in the susceptible individuals.

1. *Reduction of dietary intake of cholesterol*. Cholesterol, being a product of animal metabolism, is found only in foods of animal origin; egg yolk and organ meats are extremely rich sources. In contrast, the plant products—even vegetable oils—contain no cholesterol. In view of these facts, necessary dietary modifications must be made in diet so that intake of cholesterol rich foods is reduced.
2. *Change in lifestyle* in such a way that adequate *physical exercise* becomes part of daily routine.
3. *Increased dietary intake of fibres*, although exact role of high-fibre diet on coronary artery disease is still controversial. Mode of action of fibres and their beneficial effects have already been discussed in this chapter.
4. *Increased intake of dietary antioxidants*, such as vitamins C, E, and β-carotene. These compounds chemically inactivate the oxidative radicals derived from molecular oxygen and hence protect LDL against oxidation. This offers protection against heart diseases as confirmed by certain studies which brought out that daily intake of 100 International Units of vitamin E (RDA 15 IU) may reduce death from coronary artery disease by approximately 40%.
5. *Hypocholesterolaemic drugs* are usually recommended in middle-aged men with very high plasma cholesterol levels.

Type of fatty acids consumed have a significant bearing on serum cholesterol levels. For example, the **saturated fatty acids are harmful** because tend to elevate serum cholesterol levels, especially the LDL fraction. Therefore, the foods rich in these fatty acids, such as meat products and some vegetable oils like coconut and palm oil, must be avoided. It is interesting to note that all saturated fatty acids do not have cholesterol raising effects: stearic acid (18-C) and short-chain saturated fatty acids (< 10-C) tend to raise serum cholesterol to a lesser extent than the 12-C to 16-C saturated fatty acids.

Exercises

Essay type questions

1. What is meant by balanced diet? Explain the nutritional principles involved in its formulation. Formulate the balanced diet requirement for a 72 kg adult of 40 years.
2. What are the components of energy expenditure? Discuss various factors that effect BMR.
3. Explain importance of proteins in human nutrition. Why is SDA value of proteins higher than that of other major nutrients?
4. Discuss the protein-energy malnutrition, comparing features of marasmus and kwashiorkor. Mention biochemical findings in kwashiorkor.

Write short notes on

1. Basal metabolic rate
2. Specific dynamic action
3. Dietary fibres
4. Nitrogen balance
5. Biological value of proteins
6. Mutual supplementation of proteins
7. Protein efficiency ratio

CHAPTER

BIOCHEMICAL ENDOCRINOLOGY

29

Hormones are chemical messengers, secreted by cells of specialized tissues called endocrine glands, and transported by blood to stimulate specific functions of distant tissues or organs. They play key role in the intercellular communication and coordination of responses. Together with nervous and immune systems, the hormones are important constituents of the *signalling system* of the body that mediates interactions between various tissues. It is essential to have such interactions between different tissues and organs, each of which is especially designed for performing its specialized function. The discovery of existence of hormones, or chemical messengers (as they were originally called), stems from experiments of *Bayliss* and *Starling* (1902). These workers demonstrated that injection of the duodenal mucosa extract into the blood stream stimulated flow of pancreatic juice. (The factors present in the duodenal mucosa must have reached the pancreas through blood).

The word hormone is of Greek origin, which means "*to arouse to activity*". By the classic definition, a hormone is synthesized in one tissue, secreted in the circulatory system and transported as *mobile messenger* to control metabolic and biological activities in the target cells. However, this definition is too restrictive and it is now well established that some hormones can act on the adjacent cells in a given tissue (*paracrine function*) as well as on the cells in which they are synthesized (*autocrine function*). Examples: interleukin-2 is autocrine hormone for it is being produced by T cells and stimulates proliferation of T cells; and prostaglandins have paracrine function since they act on nearby cells. In this chapter, general characteristics and mode of action of various classes of hormones are described.

After going through this chapter, the student should be able to understand:

- Chemical diversity, biosynthesis, and transport of hormones to the target tissues; their interaction with *receptors* on target cells, concepts of *signal transduction* (receptor-effector coupling); and biochemical basis for development of disease in case of defect in receptors or in events following hormone-receptor interaction.
- Working of hormone at cellular and molecular levels, concept of membrane and intracellular receptors, and various types of *second messengers* and their mode of action.
- The above concepts will be illustrated with the help of clinical cases.

I. General Characteristics of Hormone Systems

Hormones are chemically diverse. They transmit messages or chemical signals (i.e. **endocrine signalling**), which can control almost any aspect of cellular functions and is exquisitely suited for orchestration of physiological responses throughout the body. Molecular basis of action of hormones is different for various groups of hormones. It is interesting to note that all *hormones act by* ***regulating pre-existing processes****; none of the known hormones is an enzyme or a coenzyme.*

A. Chemical Diversity of Hormones

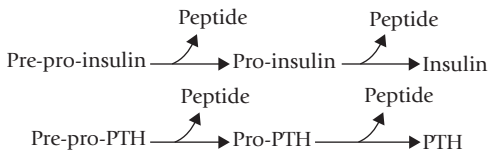
Substances of diverse origin and chemical nature may serve as hormones. Based on their chemical nature, the

hormones have been categorized into the following four groups:

1. ***Lipid hormones:*** Most lipid hormones are derived from cholesterol, such as adrenocortical hormones, sex hormones and calcitriol. Others are derived from arachidonic acid, such as prostaglandins.
   The cholesterol-derived hormones contain a steroid nucleus and are *lipophilic* in nature; they readily traverse the cell membrane of their target cells and interact with the *cytoplasmic receptors*.
2. ***Amino acid hormones:*** These hormones are produced by enzymatic modification of an amino acid molecule. For example, both *epinephrine* and *thyroxine* are derived from tyrosine molecule.
3. ***Peptide and protein hormones:*** These hormones are made up of amino acids, joined by peptide bonds. Smallest of them is thyrotropin releasing hormone (TRH), a hypothalamic-releasing factor, which consists of only three amino acids. Other examples include antidiuretic hormone (9 amino acids), glucagon (29 amino acids), parathormone (84 amino acids), and growth hormone (191 amino acids).
4. ***Glycoprotein hormones:*** A glycoprotein hormone consists of peptide chain to which carbohydrate moieties are covalently attached. The latter are necessary for the biological activity of these hormones. Examples include pituitary hormones (TSH, LH and FSH), and chorionic gonadotropin (hCG) of placental origin.

B. Biosynthesis

Biosynthetic mechanisms for hormones are diverse. Some hormones are initially synthesized as large precursor proteins which are converted to the biologically active forms by removal of specific peptide sequences. *Insulin* (MW 5500), for example, is initially synthesized as an inactive precursor, **pre-pro-insulin** (MW 11,500). A sequential removal of two peptide sequences results in production of an insulin molecule. Likewise, *parathormone* (PTH), an 84 amino acid peptide, is formed from the 115 amino acid precursor, **pre-pro-parathormone** by successive removal of two peptide segments.



Perhaps the most exaggerated example is that of thyroxine, a single amino acid hormone, which is processed from a 115 amino acid glycoprotein precursor, **thyroglobulin.**

Conversion of each of the precursor proteins mentioned above to its activated form takes place in the endocrine gland of its origin only. For example, insulin is synthesized (from pre-proinsulin) in β-cells of the pancreas, and parathormone (PTH) is produced from the precursor pre-pro-PTH in the parathyroid glands. Other precursor molecules are activated in distant peripheral tissues. ***Peripheral activation*** of androgen precursors (dehydroepiandrosterone and androstenedione) of adrenal cortex to potent androgens (testosterone and dihydrotestosterone) is an example. Similarly, 25-hydroxycholecalciferol is modified to the active form, calcitriol by 1-hydroxylation in the renal tubules (Chapter 18). Other hormones like glucocorticoids and mineralocorticoids, secreted from the adrenal cortex, do not need any modification; they are synthesized and secreted in the final form only.

Biosynthesis of Insulin

Like several other peptide hormones that are processed from larger precursor molecules, insulin is also synthesized from a precursor. The synthesis begins in the rough endoplasmic reticulum (RER). The precursor, called **pre-pro-insulin**, is a single chain peptide with 103 amino acids (Fig. 29.1). Production of mature insulin of 51 amino acid from this larger precursor involves the following sequence of events:

- Removal of a hydrophobic amino acid sequence of 24 amino acids, called the signal peptide or the *leader sequence*, from the N-terminus of pre-pro-insulin. Occurs first it is removed by *signal peptidase* in the ER immediately after translation.
- The remaining structure, known as **proinsulin**, consists of the insulin sequence interspersed by a connecting *C-peptide* (C for connecting). C-peptide is required for proper folding of the prohormone and formation of correct disulphide bonds. Its structure is far less conserved than that of insulin. Its length varies from 27–35 amino acids (therefore, proinsulin may contain 78 to 86 amino acids).
- At the final stage of insulin synthesis, proinsulin is cleaved at paired-basic amino acid residues to yield the hormonally inactive C-peptide fragment, and active fragment, the insulin.
- Both are packaged into secretory vesicle and are released from the cell.

C-peptide is released in an amount equimolar to insulin. This is exploited in the clinical laboratories to assess the β-cell function in diabetic patients under insulin treatment. Although the circulating insulin of these patients comes in part from their own pancreas and in part from therapeutically injected insulin, the circulating *C-peptide level is directly proportion to the insulin release from the patient's β-cells.*

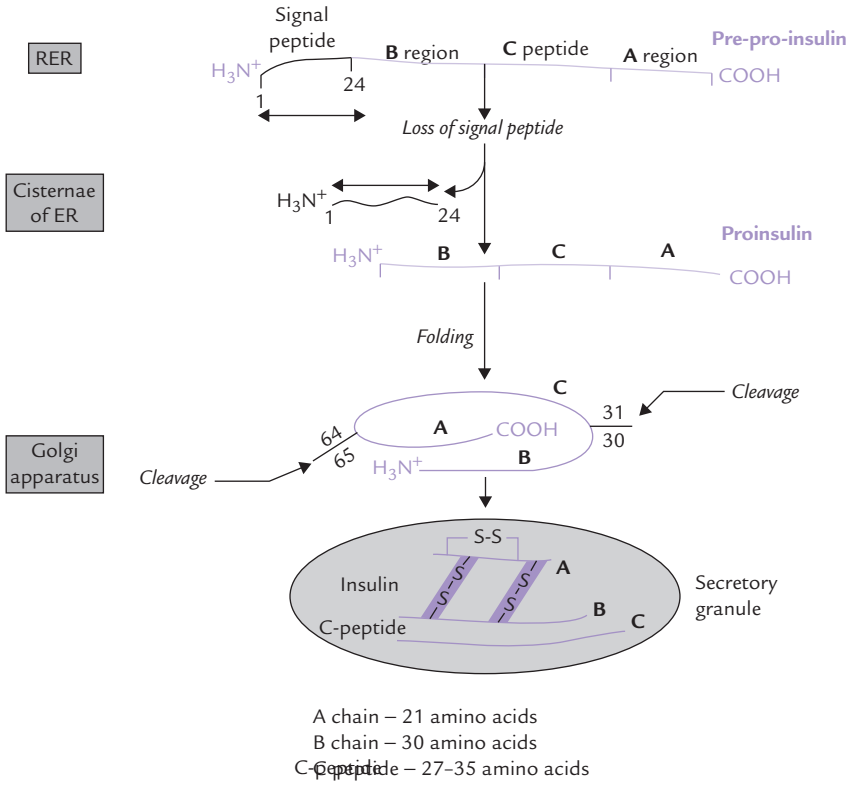


Fig. 29.1. Synthesis of insulin from pre-pro-insulin precursor by removal of 2 peptide fragments: the signal peptide sequence and the C-peptide. Signal peptide consisting of the first 24 amino acids at the amino terminus is cleaved by a *peptidase*, yielding pro-insulin. The C-peptide is removed by an enzyme in secretory granule, yielding insulin (ER = endoplasmic reticulum).

Table 29.1. Carrier proteins regulating delivery and biological half-life of the hormone

Hormone	Carrier protein	Per cent free hormone	Half-life $\left(t_{\frac{1}{2}}\right)$ in blood
Tri-iodothyronine (T_3)	TBG, TBPA	0.3	1.5 days
Thyroxine (T_4)	TBG, TBPA	0.03	6.5 days
Aldosterone	Albumin	50	Few minutes
Oestrogen	SHBG	Uncertain	Uncertain
Testosterone	SHBG/Albumin	3.1	Uncertain
Cortisol	CBG	8.7	1.5–2 hours

TBG = thyroxine-binding globulin, TBPA = thyroxine-binding prealbumin, SHBG = sex hormone-binding globulin, CBG = cortisol binding globulin (half-life of peptide hormones is less than a minute).

C. Transport

The peptide hormones circulate in the blood in free form, unbound to any transport protein. This is due to their polar (hydrophilic) nature. In contrast, the steroid hormones and the thyroid hormones are predominantly hydrophobic in nature and therefore, cannot circulate in blood entirely in an unbound form. They are mostly transported to their site of action by carrier proteins, where they exert their action and are inactivated by further metabolism. Relatively smaller percentage of the total circulating hormone is left in the free form (Table 29.1). The thyroid hormones are bound with two specific carrier proteins: thyroxine-binding globulin (TBG) and thyroxine-binding pre-albumin (TBPA), which together bind nearly all the T_3 (tri-iodothyronine) and T_4 (thyroxine) in plasma.

In addition to playing a key role in hormone transport, the carrier protein also prevents rapid clearance of the hormone and thereby, prolongs its half-life. Because of this, the half-life of the steroids and thyroid hormones, which mostly exist in bound states, is much longer than

that of the peptide hormones which is less than 30 minutes (Table 29.1). The avidity of binding also determines the half-life. T_4, which binds the carrier proteins more tightly than T_3, has a longer half-life compared to that of T_3.

Only the free form of a hormone, and not the bound form, is able to interact with the cellular receptors. Therefore, it is the free form that is responsible for inducing metabolic and biological effects in target cells. The bound form merely acts as a *circulating reservoir*, releasing the free hormone in the immediate vicinity of the target tissue, as per cellular requirements. Therefore, the **ratio of free : bound forms**, and not the total circulating hormone concentration, is true reflection of the observed biological activity. For example, in case of aldosterone, which is the most potent mineralocorticoid, high free : bound ratio is observed. This high ratio among other factors, accounts for its high biological activity.

The peptide hormones circulate free (unbound), whereas the steroid and thyroid hormones are mostly bound with carrier proteins. The latter act as circulating reservoir of hormone, and also prolong its half-life.

D. Target Tissue Concept

The physiological and biochemical effects of a given hormone are elicited only in a specific tissue, known as its target tissue. A target tissue has **specific receptors** with which the hormone interacts. The hormone-receptor interaction triggers a series of events that subsequently lead to elicitation of the biological effects of the hormone. For example:

- Target tissue for thyroid-stimulating hormone (TSH) is the thyroid gland, where this hormone stimulates synthesis and secretion of the iodothyronines (T_3 and T_4).
- Similarly, adrenal cortex, the target tissue for ACTH, responds to ACTH by increasing steroidogenesis.
- Other hormones (e.g. insulin, growth hormone, and cortisol) have more than one target tissues, including liver, muscle and adipose tissue, where they influence a variety of metabolic processes.

More recently, as the knowledge about the hormone-receptor interaction has increased, the definition of target tissue has been expanded. Presently it includes all such tissues which have cellular receptors for the given hormone, irrespective of whether the biological response is elicited or not. For instance, endothelium is the target tissue for insulin because of the presence of insulin receptors on endothelial cells, even though the latter are unresponsive to this hormone.

Hormones, like other extracellular messengers and neurotransmitters, initiate their cellular effects by binding to specific receptors on their target tissue(s).

E. Feedback Concept

Blood levels of the target tissue hormones (e.g. thyroid gland) have an effect (mostly inhibitory) on the secretary activities of either the hypothalamus or pituitary. An elaborate system of feedback effects, termed the **short feedback loop** and **long feedback loop** exist, which relies on a precise signalling, and it helps to maintain the circulating plasma concentration of the hormones to the required levels. Such feedback mechanism is best illustrated by the **hypothalamo-pituitary-target gland axis** (Fig. 29.2).

- The hypothalamic-releasing factor stimulates the release of the anterior pituitary hormone (called *trophic hormone*).
- The trophic hormone enhances the synthesis and/or secretion of the hormone from the target gland.
- The blood level of the target gland hormone thereby rises and exerts inhibitory effect (i.e. *negative feedback effect*) on the anterior pituitary and the hypothalamic secretions.
- The net result of these inhibitions is that secretion of the trophic pituitary hormone decreases. Diminished level of the pituitary hormone decreases stimulus for the target gland, so that secretion of the target gland hormone also falls.

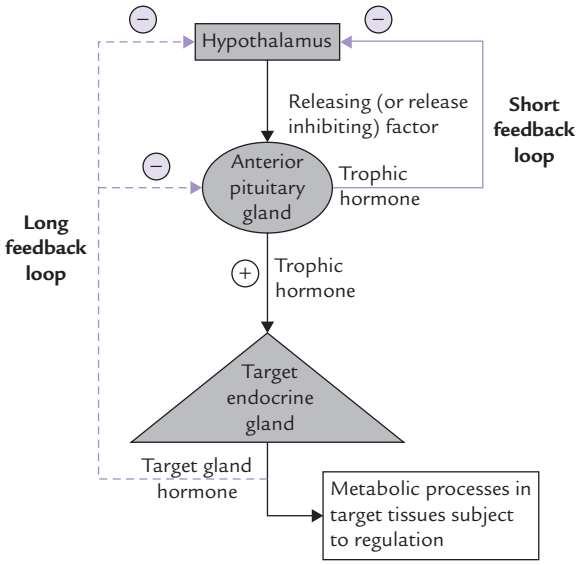
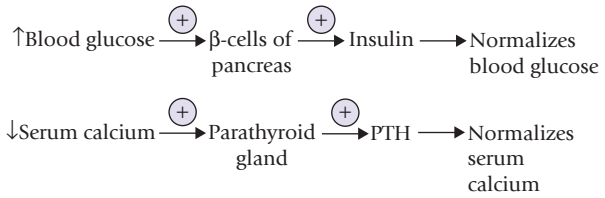


Fig. 29.2. Hypothalamic-pituitary-target gland axis in negative feedback; ⊖ indicates negative feedback effect.

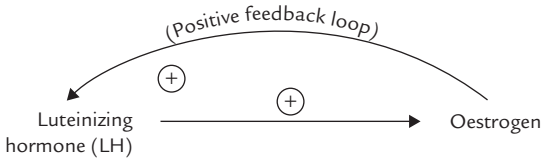
In this way, a self-regulatory loop is formed referred to as **long feedback loop** (Fig. 29.2). The pituitary hormone may also directly cause feedback inhibition of the hypothalamic secretion (i.e. **short feedback loop**). Defect in this axis can disturb the normal regulation and cause hormonal disorders (**Case 29.1**).

In some cases, the feedback modulator of a given hormone may be a metabolite, the circulating level of which is controlled by the hormone itself. For example, raised blood glucose concentration is the most potent (positive) modulator of insulin synthesis and secretion from the β-cells of pancreas. Similarly, decreased serum calcium is a potent stimulus for parathormone (PTH), which in turn mobilizes calcium from the bony reservoir to restore the serum calcium levels.



Not all Feedbacks are Negative

In addition to the negative feedback mechanisms, those for *positive feedback modulation* are also known. A classic example is stimulation for release of luteinizing hormone (form anterior pituitary) by increased plasma levels of oestrogen. It is noteworthy that oestrogen is itself released from ovarian follicles in response to the luteinizing hormone (LH).



The oestrogen-mediated release of LH secretion is called **LH surge**, which induces ovulation during menstrual cycle (Chapter 30).

Failure of feedback mechanisms may disturb the precisely controlled systems described above. These may result in a disease, as exemplified in **Case 29.2.**

Central nervous system also exerts important influence on these mechanisms and may even override the effects of the feedback regulations. In fact, cascades of signals that start from external or internal environment are first transmitted to the CNS. These may involve components of the limbic system such as hippocampus and amygdala. These structures innervate the hypothalamus in a specific region, which responds by liberating hypothalamic-releasing factors.

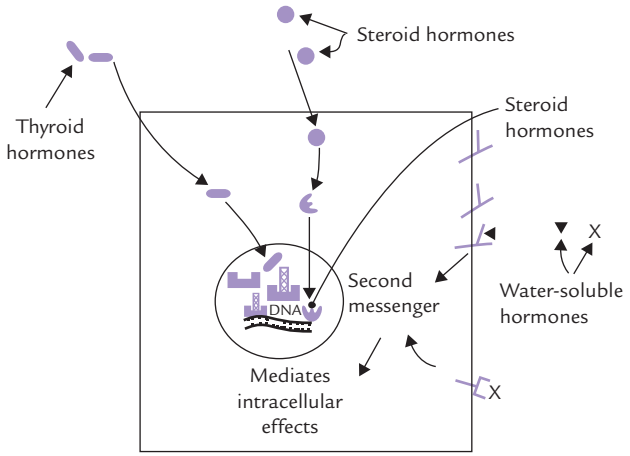


Fig. 29.3. Cellular locations of receptors for hormones. Receptors for thyroid hormone are intra-nuclear; steroid hormone receptors are cytosolic; water soluble hormone receptors are on cell surface. ▬ = thyroid hormone receptor, ᴗ = steroid hormone receptor, ⅄ and ⅂ = cell surface receptors.

II. Hormone Receptors

A. General Characteristics

Receptors are cell-associated recognition molecules that play a crucial role in the hormone action.

- *The receptors for water-soluble hormones* (peptides, proteins, or glycoproteins) are present on the **cell surface**. Interaction of hormones with receptors stimulates certain molecules namely second messengers, which mediate biochemical functions intracellularly.
- *The receptors for lipophilic hormones* such as the steroids and the thyroid hormones are **located intracellularly** (Fig. 29.3).

The lipophilic hormones readily traverse the cell membrane because of their hydrophobic nature and enter the cell. Intracellularly, they interact with specific receptors located either in the cytosol (e.g. steroid hormones) or within the nucleus (e.g. thyroid hormones) to form hormone-receptor complexes. The latter serves as the intracellular messengers, through which biochemical functions are mediated.

Thousands of receptor molecules may be present in a single cell. They bind the hormone with a high affinity. The interaction is highly specific, involving electrostatic and hydrophobic interactions, and is reversible in most cases.

B. Receptor-effector Coupling

The hormone-receptor interaction has a unique feature, i.e. Receptor-effector coupling, implying that occupancy

of a receptor with its specific ligand molecule influences several intracellular events. This is because of the two distinct functional domains present in a receptor molecule: the **binding domain** and the **coupling domain**. The binding domain recognizes and binds the hormone; the signal is transmitted (*signal transduction*) through the coupling domain to the intracellular proteins, usually enzymes. The activity and/or intracellular concentration of an enzyme changes in response to the signal; and it is through such change that the effect of the hormone is mediated.

A classic example is the hormone, epinephrine, which binds with the extracellular binding domain of the receptor, located on the cell surface of the hepatocyte. Following signal transduction, *adenylate cyclase*, an intracellular enzyme, is activated. This enzyme influences several metabolic pathways by triggering a series of events discussed later in this Chapter. Thus, the hormone may influence several intracellular events even without entering the cell.

C. Regulation of Receptors

Sensitivity of the target tissue to the hormone is determined by number of receptors on the target cell, and the affinity of these receptors for the respective ligand is important in this regard. Both these factors are subject to regulation. Prolonged exposure to an elevated concentration of the hormone results in decreased sensitivity of the target tissue to the hormone (**desensitization**). Desensitization involves two mechanisms: down regulation and covalent modulation of the receptors.

Down Regulation

This implies internal sequestration of the receptors, so that fewer receptors are available on the cell surface for interaction with the agonist. Moreover, the internalized receptors get segregated from other components of the response system, resulting in decreased response of the target tissue. Down regulation is a reversible process; removal of the agonist results in return of the receptors onto the cell surface (**Case 15.1**).

Covalent Modulation

This involves phosphorylation of the receptors which causes uncoupling of the binding and the signal transmitting functions of the receptors. This impairs the response of the target tissue.

A few hormones, such as prolactin and angiotensin, *"up regulate"* their own receptors, thus amplifying the target tissue sensitivity. Such increase in the number of receptors occurs within minutes and serves as an efficient way of regulating the biological response elicited by the hormone.

Receptors are allosteric proteins that bind their ligand with high affinity and selectivity and then transduce the signal to elicit responses in the interior of the cells. Many receptors lose their responsiveness after over-stimulation, either because they get phosphorylated or because they undergo endocytosis (down regulation).

D. Structure of Receptors

Purification and characterization of receptors have been difficult. This is because it is difficult to obtain the requisite amount of receptors for such analysis. With development of recombinant DNA technology, this difficulty has been overcome. Samples of adequate size can be obtained by recombinant techniques and subjected to analysis. Structures of various hormone receptors are being elucidated in this way. Structure of the acetylcholine receptor, for instance, has been determined by molecular cloning techniques and is known to consist of four peptide subunits. The steroid hormone receptors have also been shown to be proteins; some of these are single chain proteins, while others are heterodimers. Details about insulin receptors—an outcome of extensive research in recent years—are known.

Insulin Receptor

Insulin binds with a specific receptor type in target tissues. The receptor is a membrane protein with an unusual enzymatic activity. It is a *tetramer*, consisting of **two alpha subunits** and **two beta subunits** linked by disulphide bonds. The beta subunits possess *tyrosine kinase* activity while the alpha subunits provide binding site for insulin.

1. **Binding of insulin to the alpha subunits** occurs in the extracellular domain, as shown in Figure 29.4. This switches on the ***tyrosine kinase*** activity of the beta subunits. The above stimulation causes phosphorylation of protein in intracellular domain on tyrosine side chains. (Phosphorylation of tyrosine is in contrast to that of hormone stimulated *protein kinases A* and *C*; the latter phosphorylates serine and threonine residues on their substrates.)
2. The substrate specificity of the insulin receptors is rather unusual: after ligand binding, the receptors tend to aggregate in the lipid bilayer of membrane and phosphorylate one another. This activity is called **autophosphorylation.**

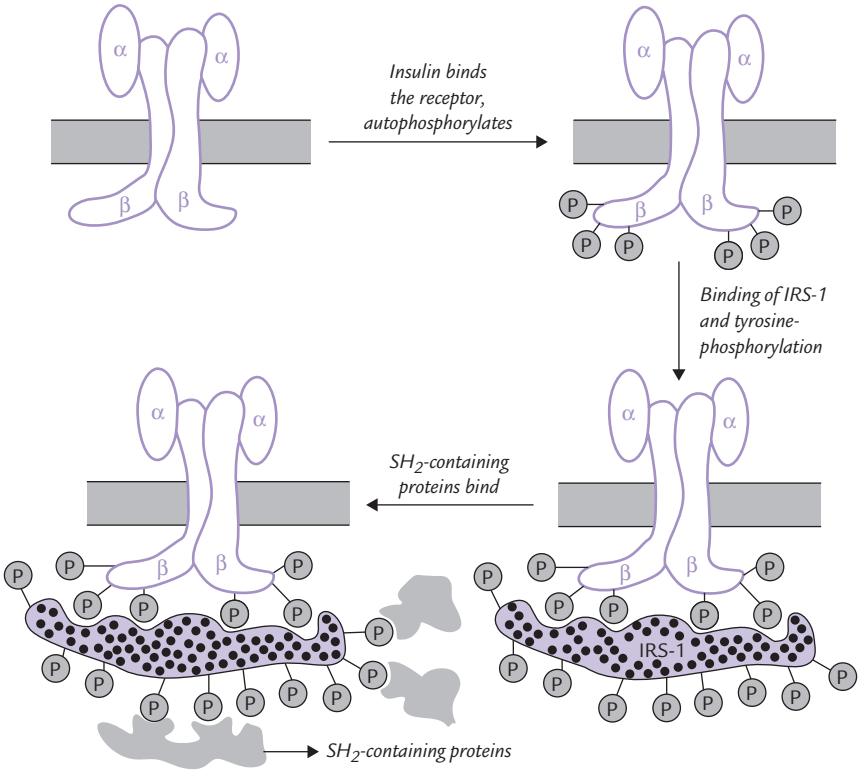


Fig. 29.4. The insulin receptor (P indicates phosphoryl group, IRS-1 = insulin receptor substrate-1, SH = src homology).

3. The next event after autophosphorylation is **binding of the receptor to 131-kD phosphoprotein, called IRS-1** (insulin receptor substrate 1), which becomes extensively phosphorylated on approximately 20 tyrosine residues (Fig. 29.4).
4. These phosphotyrosine residues on IRS-1 serve as the **docking sites** for the SH_2-containing proteins, meaning that a number of proteins having a specialized domain, the SH_2 domain (SH for *src* homology), bind with the tyrosine-phosphorylated sites on IRS-1.
5. Though the signalling pathways are still poorly known beyond this stage the following changes in SH_2-containing proteins have been defined:
   - Some of the SH_2-containing proteins get tyrosine phosphorylated by IRS-1, and then participate in **multistep phosphorylation cascades** (one protein phosphorylating another, and so on). These cascades will eventually alter the activities of enzymes of various pathways, e.g. glycogen and fatty acid metabolism.
   - Some other effector proteins are recruited to the plasma membrane, where they facilitate glucose transport. Few such transport proteins are normally located on the cell membrane. Binding of insulin to the receptor initiates a rapid mobilization of intracellular stores of such transporter to the plasma membrane.

Receptors for insulin (and many growth factors) are membrane proteins with an intrinsic *tyrosine kinase* activity. They autophosphorylate, in response to ligand binding, and transmit signals via protein binding to the tyrosine phosphorylated sites.

E. Hormone Receptors and Diseases

Abnormalities of hormone receptors may impair hormone action, which may cause various receptor related diseases (Table 29.2). Based on the nature of abnormality of the receptor, these diseases are divided into three categories:

1. The first category includes the diseases in which ***binding of the hormone and the receptor is impaired***; in some cases, no binding has been detected. Diseases developing due to defective receptors for calcitriol (vitamin D resistant rickets, type II), ADH (congenital nephrogenic diabetes insipidus), PTH (pseudohypoparathyroidism) and fall into this category.
2. In the second category, ***antibodies of IgG class, directed against a specific hormone receptor,*** have been detected. Interaction of these antibodies with the receptor may have varying effects on the hormone action. The antibodies may

Table 29.2. Diseases related to abnormalities of hormone receptors

Disease	Receptor	Defect
Vitamin D resistant rickets; type II	Calcitriol	Receptor deficiency
Congenital nephrogenic diabetes insipidus	ADH	Receptor deficiency
Pseudo-hypoparathyroidism	PTH	Receptor deficiency
Acanthosis nigricans	Insulin	Antibody (IgG) blocks hormone-receptor interaction
Graves' disease	TSH	Antibody stimulates TSH receptor
Myasthenia gravis	Acetylcholine	Antibody enhances turnover of acetylcholine receptors
Asthma	β-Adrenergic	Antibody blocks β-adrenergic binding
Non-insulin dependent diabetes mellitus	Insulin	Decreased receptors on target cells
Obesity	Insulin	Decreased receptors on target cells

ADH = antidiuretic hormone; PTH = parathormone; TSH = thyroid stimulating hormone.

(a) block the action of the hormone (acanthosis nigricans with insulin resistance),
(b) mimic hormone binding (Graves' disease), or
(c) enhance receptor turnover (myasthenia gravis).

3. In the third category, the ***receptor regulation is defective.*** For example, in patients with non-insulin dependent diabetes mellitus and obesity, **down regulation** of the insulin receptors occurs. The target cell insensitivity and glucose intolerance develop in these patients, in spite of elevated plasma insulin levels. With weight reduction, the receptors return to the cell surface resulting in improvement of the target tissue response **(Case 15.1).**

   In some cases, more than one-receptor defect may be present. For example, in acanthosis nigricans, the patient is mostly obese. Therefore, in addition to the primary receptor defect (insulin-receptor binding decreased), down regulation of receptors also occurs. Some clinical cases discussed in this chapter illustrate as to how a variety of hormonal disorders can develop because of defective receptors **(see Cases 29.3 to 29.5).**

III. Mechanism of Action of Hormones

Hormones have been classified in **two major groups** *based on their mechanisms of action* (Table 29.3). The mechanism of action depends upon the location of the hormone receptor and the nature of the signal transmitted following the hormone-receptor interaction.

Group I

This group comprises the *lipophilic hormones* that are derived from cholesterol and the thyroid hormones (exceptions—T_3 and T_4). These hormones interact with cytoplasmic or nuclear receptors respectively. The hormone-receptor complex itself acts as an intracellular messenger and directly influences the gene expression (Fig. 29.3).

Group II

This group comprises the *peptide, protein*, and *glycoprotein hormones*. These hormones bind with the surface receptors, located on the plasma membrane of the target cells. The hormone-receptor interaction transmits a signal across the membrane, that results in elevation of the intracellular level of an intermediary molecule, the so called ***second messenger*** (the hormone itself is the first messenger). The second messenger acts as a signal-conducting molecule, through which the biological effects of a hormone are mediated.

Depending on the chemical nature of the second messenger generated, group II hormones are further divided into three subgroups (IIa, IIb and IIc):

- ***Group IIa*** hormones employ ***cAMP*** as the second messenger (Table 29.3). The concept of second messenger was introduced with the discovery of cAMP by *Sutherland,* who reported raised intracellular cAMP levels following binding of epinephrine with plasma membrane of pigeon erythrocytes.
- The second messenger for the ***group IIb*** hormones is *cGMP* and those for the **group IIc** hormones is ***calcium or phosphatidylinositides*** (or both).
- Hormones of ***group IId*** employ some multistep phosphorylation cascade that has not been fully identified/settled.

Most hormones induce formation of a second messenger, which carries out long-distance signalling throughout the cell. Cyclic AMP, cyclic GMP, diacylglycerol and inositol triphosphate (IP_3) are some common second messengers.

Table 29.3. Classification of hormones based on mechanism of action

Group I: Hormones that bind to intracellular receptors	
Ia. Cytosolic receptors	
Glucocorticoids	
Mineralocorticoids	Oestrogens
Progestins	Calcitriol
Ib. Nuclear receptors	
Thyroxine (T_4)	
Triiodothyronine (T_3)	
Group II: Hormones that bind to receptors located on cell surface	
IIa. Second messenger is ***cAMP***	
Human chorionic gonadotropin (hCG)	Calcitonin
Luteinizing hormone (LH)	Glucagon
Follicular-stimulating hormone (FSH)	Lipotropin (LPH)
Antidiuretic hormone (ADH)	Opioids
β-Adrenergic catecholamines	Parathormone (PTH)
α_2-Adrenergic catecholamines	Somatostatin
Thyroid-stimulating hormone (TSH)	Angiotensin II
Melanocyte-stimulating hormone (MSH)	
Adrenocorticotrophic hormone (ACTH)	
Corticotropin-releasing hormone (CRH)	
IIb. Second messenger is ***cGMP***	
ANF (atrial natriuretic factor)	
NO (nitric oxide)	
IIc. Second messenger is ***calcium*** or ***phosphatidylinositides*** or both	
Gonadotropin-releasing hormone (GnRH)	Angiotensin II
Thyrotropin-releasing hormone (TRH)	Vasopressin
Acetylcholine (muscarinic)	Gastrin
α-Adrenergic catecholamines	Cholecystokinin
IId. The intracellular messenger is a ***protein kinase*** or ***phosphatase cascade***	
Insulin	Growth hormone (GH)
Epidermal growth factor (EGF)	Prolactin (PRL)
Nerve growth factor (NGF)	Oxytocin
Fibroblast growth factor (FGF)	Erythropoietin
Chorionic somatomammotropin (CS)	
Insulin-like growth factors (IgF-I, IGF-II)	

Some hormones use different secondary messengers in different tissues. Action of PTH, for example, is mediated via cAMP in renal cells and phosphoinositide/Ca^{2+} in bone cells. Likewise, vasopressin also employs two different messengers in renal cells and muscle cells.

The intracellular events that follow the hormone-receptor interaction play a role of vital significance in eliciting the biological effects of the hormone.

A. Mechanism of Action of Group I Hormones

Sterol-derived Hormones

The hormones of this subgroup readily diffuse through plasma membrane of the target cells and encounter specific, high-affinity **cytosolic receptors** (Fig. 29.3).

Formation of hormone-receptor complex is followed by change in its conformation and surface charge, which results in activation of the complex. The activated hormone-receptor complex is able to selectively bind specific regions of DNA (called **hormone responsive elements**) and enhance transcription of specific genes.

Why is the receptor able to bind with specific DNA sequences only after its interaction with the hormone? The reason is that the unstimulated receptor normally resides in the cytoplasm complexed to cytoplasmic proteins that mask its DNA-binding domain. After the hormone binding, the receptor releases its cytoplasmic-binding proteins, and the DNA-binding domain is thereby exposed. Together with the bound hormone, the activated receptor is thus able to bind the hormone responsive elements (HRE).

The HREs are located next to the hormone regulated genes, usually within a few hundred base pairs upstream of the start site for transcription.

Direct action of the hormone-receptor complex upon various cell organelles and membranes have also been reported. In addition, some hormones (glucocorticoids) effect post-translational processing of some proteins. However, of all these actions, the most predominant one is on gene transcription. Further details about molecular mechanisms involved in the process are given in Chapter 24.

Iodothyronine Hormones

These hormones also change expression of specific genes, like the sterol-derived hormones. However, they do not associate with the cytoplasmic receptors. Rather, they associate with certain high affinity receptors present within the nucleus. Thus, receptors for these hormones are also ligand-binding transcription factors (Fig. 29.3).

The receptors for steroid hormone and thyroid hormones are transcription factors that encounter their hormonal ligand either in the cytoplasm or in the nucleus. The receptor-hormone complex binds to specific response elements in the regulatory sites of genes and affects the rate of transcription.

B. Mechanism of Action of Group II Hormones

Hormones of this group are considered *first messengers*, and their intracellular effects are elicited through mediator molecules termed *second messengers*. The hormones bind with ***cell surface receptors***, which are integral membrane glycoproteins, having three functional domains:

1. The *extracellular domain* that binds the hormone.
2. One or more *transmembrane α-helices* that penetrate the lipid bilayer.
3. *Intracellular domain* that is coupled with an effector mechanism.

Most hormone receptors belong to a family of membrane glycoproteins with seven membrane spanning α-helices. They transmit the signal through cascade of protein-protein interactions.

The hormone receptor interaction transmits signal across the cell membrane to target proteins by the following sequence of events (Fig. 29.5):

- Binding of hormone with the extracellular domain of the receptor induces conformational changes in the receptor.
- These changes are transmitted to the transmembrane α-helices, from where they reach the intracellular domain, which is associated with a guanine nucleotide-binding

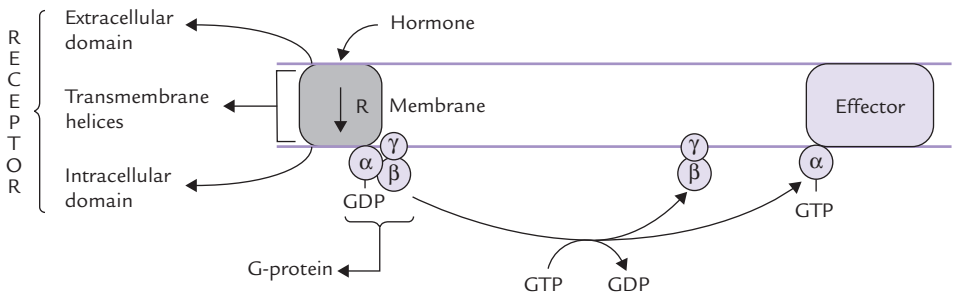


Fig. 29.5. The G-protein signalling. In the resting state the complete G-protein (αβγ) is associated with the receptor (R) and GDP. Binding of hormone to the receptor causes cleavage of the G-proteins and the products dissociate from the receptor to reach their target proteins (effectors) whose properties they affect by allosteric mechanism.

regulatory protein or **G-protein**. It is a trimeric complex of three subunits (α β and γ) associated with GDP in its inactivated state.

- The conformational change in the intracellular domain of the receptor activates G protein allosterically. The activation involves GDP-GTP exchange, and dissociation of G-proteins into βγ and α-GTP subunits.
- Both these products bind to target proteins in plasma membrane, called **effectors**. This results in allosteric modulation of the effector.

Most hormones act on plasma membrane receptors that operate in conjunction with heterotrimeric G-proteins: a group of regulatory molecules that mediate intracellular responses. The hormone-receptor interaction activates G-protein, and the components of the activated G-protein act on effector proteins in the plasma membrane.

The best known **effectors** are second messenger synthesizing enzymes, such as ***adenylate cyclase,*** and ***phospholipase C, guanylate cyclase,*** etc. These enzymes, when activated, generate second messengers which serve as mediators of the hormone action. The commonest second messengers are:

- Cyclic AMP (cAMP)
- Phosphatidylinositides/calcium system
- cGMP and nitric-oxide

cAMP as Second Messenger (Group IIa Hormones)

The series of events that lead to activation of *adenylate cyclase* and consequent generation of cAMP are discussed in brief below (Fig. 29.6):

1. The hormone (H) binds to its receptor (R) and the signal is transmitted to the G-protein: a trimer of α-subunit (47 D), β-subunit (37 kD) and γ-subunit (7 kD). The β- and γ-subunits are tightly associated with each other; the α-subunit binds only loosely with βγ.
2. The G-protein responds to the signal from the H-R complex by undergoing a conformational change, which greatly reduces affinity of the α-subunit for GDP. The **GDP,** therefore, **dissociates** and is replaced quickly by GTP.
3. Once GTP is bound, it induces **dissociation of the trimeric complex** into α-GTP subunit and βγ complex.

$$\frac{\alpha\beta\gamma}{GDP} \xrightarrow[GDP]{GTP} \alpha\text{-GTP} + \beta\gamma$$

α-GTP, called active-α, is a potent stimulator of *adenylate cyclase*.

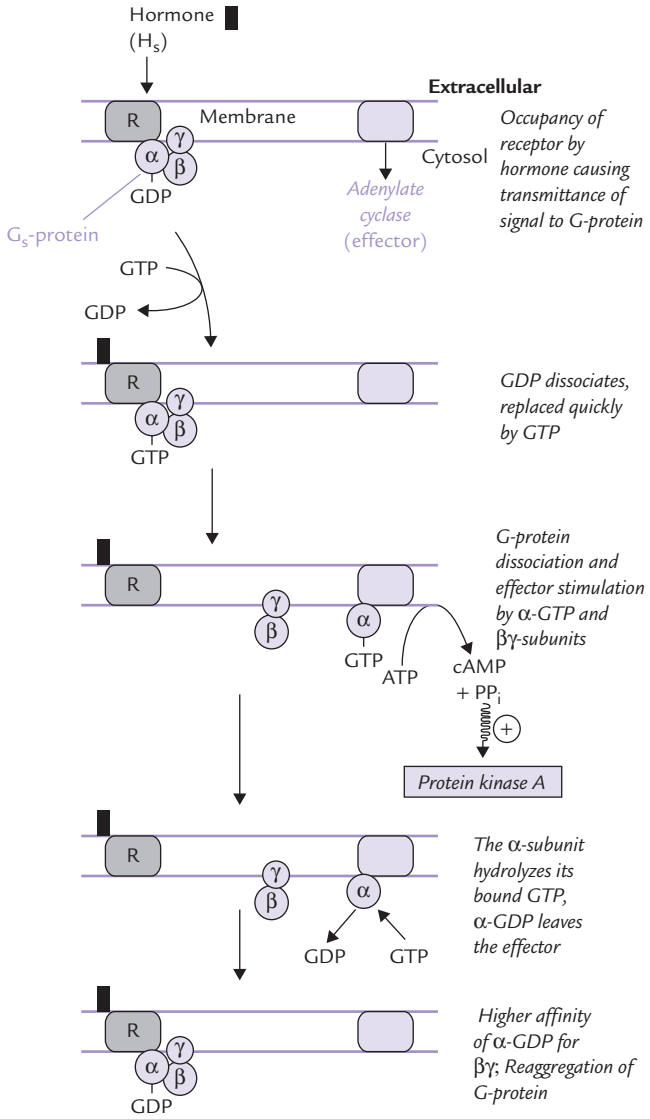


Fig. 29.6. Signal transduction through a G-protein-linked receptor.

4. Action of the active-α is *self-limiting*. This is because of the *intrinsic GTPase* activity of the active-α, that hydrolyzes the bound GTP to GDP and inorganic phosphate.

$$\alpha - GTP \longrightarrow GDP + P_i$$

5. The α-GDP is biologically inactive. It dissociates from *adenylate cyclase*, and thereby ceases to activate it. Further, it reassociates with the βγ subunits to form the trimeric ($\alpha_s\beta\gamma$) complex again.

The **α-GTP** (also called active α_s) **stimulates *adenylate cyclase.*** *Some isoforms of adenylate cyclase are stimulated by βγ-subunits.* The stimulated enzyme generates cAMP, as discussed in the following section.

The components of G-protein are membrane bound messengers that transmit signal from the receptor to the effector.

Action of cAMP: An Overview

The *adenylate cyclase*, stimulated by the hormone stimulated G-protein, acts on ATP:

$$ATP \longrightarrow cAMP + PP_i$$

The inorganic pyrophosphate is immediately hydrolyzed by *pyrophosphatases*, making the reaction irreversible.

The cAMP activates an intracellular enzyme called cAMP-dependent *protein kinase* A (*PKA;* **A** is the designation for c**A**MP).

How does cAMP cause activation of protein kinase A? The mechanism is discussed below.

Protein kinase consists of four subunits: two catalytic and two regulatory (Fig. 29.7). Its subunit structure is R_2C_2. In absence of cAMP, the C and R are bound together and the enzyme is inactive in this form. Four cAMP molecular bind to the inactive *protein kinase* and cause dissociation of R and C subunits.

The catalytic C subunits phosphorylate a variety of protein substrates in the cytoplasm and the nucleus. These include several metabolic and regulatory enzymes. (Table 29.4), which ultimately cause biochemical responses.

The cAMP cascade amplifies hormonal stimulus: *Protein kinase A* is fully activated at submicromolar concentration of cAMP, so that it is very sensitive to small changes in *adenylate cyclase* activity. The hormonal response is **amplified** following receptor binding because a single receptor can generate several active-α molecules. Each active-α acts on *adenylate cyclase* long enough to cause synthesis of hundreds of cAMP molecules, and so on.

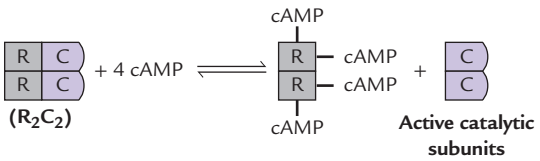


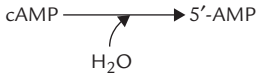
Fig. 29.7. Binding of 2 molecules of cAMP to each R-subunit results in conformational changes that lead to dissociation of ($cAMP_2$-R_2) dimer from the C-subunit. The monomeric, active C-subunits, then proceed to phosphorylate serine and threonine residues in target enzymes. **C** = catalytic subunits, **R** = regulatory (inhibitory) subunits.

Table 29.4. Intracellular substrates for *protein kinase A*

Intracellular protein	Tissue	Effect of phosphorylation
Glycogen synthase	L, M	Inhibition
Phosphorylase kinase	L, M	Stimulation
Pyruvate kinase	L	Inhibition
Hormone sensitive lipase	A	Stimulation
Acetyl CoA carboxylase	L	Inhibition
Phenylalanine hydroxylase	L	Stimulation
Inhibitor-1	L	Stimulation

L = liver; M = muscle; A = adipose tissue.

Dissociation of cAMP: cAMP is degraded by a group of enzymes, called *phosphodiesterases*.



Phosphodiesterases are inhibited by methylxanthines such as caffeine, theophylline and amniophylline, which thereby potentiate effect of cAMP.

Note The G-protein discussed above is actually stimulatory G-protein (G_s), having subunit composition of $\alpha_s\beta\gamma$. Other types of G-proteins such as inhibitory G-protein, G_q and G_{12} are also known, as discussed later.

The Inhibitory G-Protein (G_i)

Some hormones (H_i) inhibit the *adenylate cyclase activity* (Fig. 29.8). The effect is mediated through a different receptor type (R_i) and an ***inhibitory G-protein*** complex (G_i). This complex consists of three subunits (α_i $\beta\gamma$); the structure of $\beta\gamma$-subunits of G_i is similar to those of the stimulatory G-protein, but α-subunit of G_i (α_i) is different with a molecular weight of 45,000 (Recall that the *α-subunit of the stimulatory G-protein is* α_s, having molecular weight of 47,000). The hormone-receptor (H_i-R_i) interaction results in dissociation of the complex in two parts: α_i-GTP and $\beta\gamma$. The former has a negative effect on the activity of *adenylate cyclase* and therefore on cAMP production. This explains fall in intracellular cAMP level by certain hormones (H_i), such as α_2*-adrenergics, opioids, somatostatin* and *angiotensin II* (Table 29.5).

cAMP is synthesized by a membrane enzyme, *adenylate cyclase* that can be both stimulated and inhibited by receptor-controlled G-proteins. cAMP activates the enzyme *protein kinase A*, a multifunctional *kinase*, which alters the activity of many target proteins by phosphorylating them.

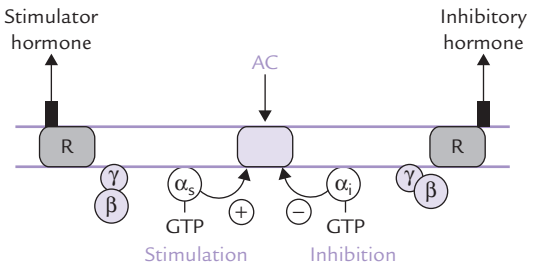


Fig. 29.8. Regulation of *adenylate cyclase* by the stimulatory and inhibitory hormones acting via stimulatory G-protein (G_s) and inhibitory G-protein (G_i), respectively. Most cells contain both G_s-linked receptor (R_s) and G_i-linked receptors (R_i). The actual activity of *adenylate cyclase* depends on the balance between the stimulatory and the inhibitory hormones (AC = *adenylate cyclase*).

Abnormalities of the hormone-G protein-*adenylate cyclase* axis may result in impaired action of hormones, as exemplified in **Case 29.4.**

Other Types of G-proteins

Humans have approximately 20 different G-protein α-subunits, five different β-subunits and 10 β-subunits. These subunits are expressed in different combinations in different cell types. According to structure and function of their α-subunit, four types of G-proteins have been recognized.

The first two—G_s-proteins and G_i-proteins—which stimulate and inhibit *adenylate cyclase* have been already discussed. The other two groups are as follows:

- ***Gq proteins*** stimulate a phosphatidylinositol-specific *phospholipase C* (Fig. 29.9). The latter forms the second messengers inositol triphosphate and diacylglycerol (discussed later).
- ***G-12 proteins*** regulate ion channels.

Table 29.5. Hormones that use cAMP as the intracellular messenger may be stimulatory (Hs) or inhibitory (Hi) for *adenylate cyclase* system. The stimulatory hormones increase the intracellular cAMP concentration, and the inhibitory ones decrease it

Hs	Hi
Human chorionic gonadotropin (hCG)	Acetylcholins
Luteinizing hormone (LH)	α_2-adrenergies
Follicular-stimulating hormone (FSH)	Somatostatin
Antidiuretic hormone (ADH)	Angiotensin II
β-Adrenergic catecholamines	
Thyroid-stimulating hormone (TSH)	
Melanocyte-stimulating hormone (MSH)	
Adrenocorticotrophic hormone (ACTH)	
Corticotropin-releasing hormone (CRH)	
Calcitonin Glucagon Lipotropin (LPH)	
Parathormone (PTH)	

Most important effectors for hormone regulated G-proteins are the second messenger synthesizing enzymes, such as *adenylate cyclase* and *phospholipase C*, but some other cellular responses are also regulated by this mechanism. These include regulation of activities of a number of ion channels, such as K^+, $C1^-$ and Ca^{2+} channels.

Modification of G-protein by Some Bacterial Toxins

The stimulatory (G_s) as well as the inhibitory (G_i) G-proteins are targets of some bacterial toxins: the G_s for the cholera toxin, and the G_i for the pertussis toxin.

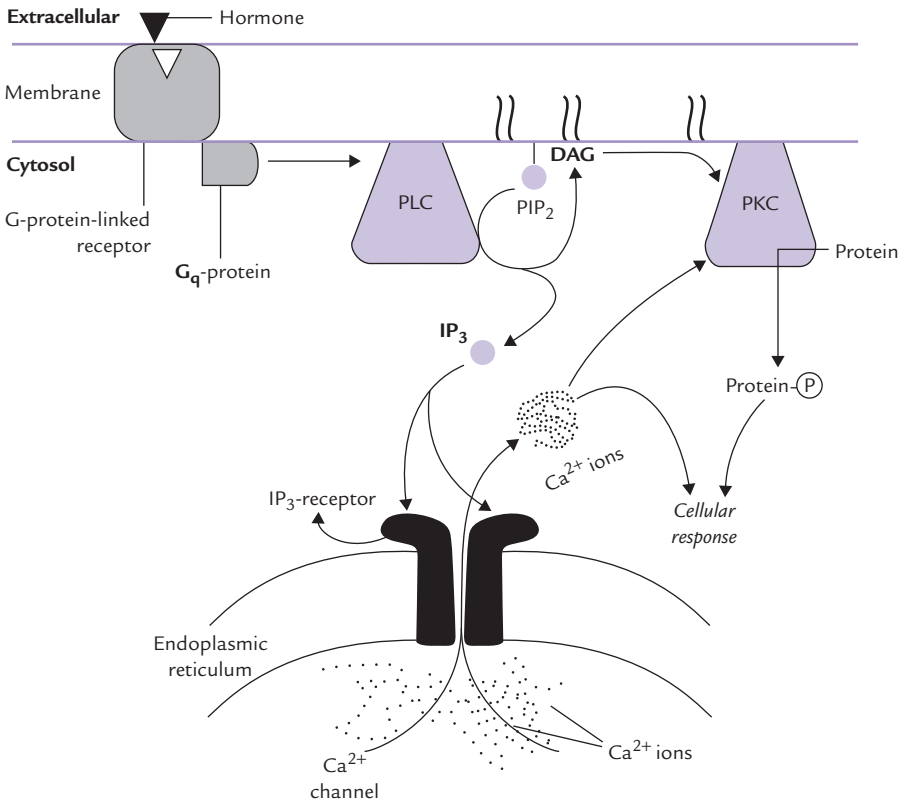
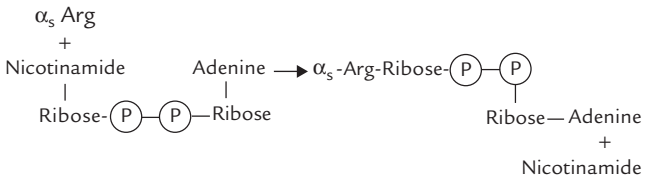


Fig. 29.9. The phosphatidylinositol 4,5-bisphosphate (PIP_2), a membrane phospholipid, is hydrolyzed by *phospholipase C* (PLC) to generate two second messengers: inositol triphosphate (IP_3) and diacylglycerol (DAG). IP_3 is released into cytosol where it mobilizes intracellular stores of calcium; and DAG, a membrane anchored second messenger, activates a family of signalling enzymes known as *protein kinase C* (PKC).

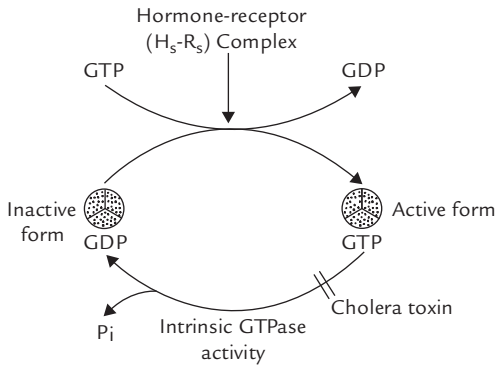
Cholera is an intestinal infection caused by *Vibrio cholerae*. The offending bacterium produces a lethal toxin that produces severe diarrhoea, which is often fatal. This enterotoxin binds with a ganglioside Gm_1 on the surface of intestinal mucosal cells initially, and then one of its subunits enters the intestinal mucosal cell, where it acts as an enzyme. Intracellularly, this subunit brings about covalent modification of the α_s-subunit of the G_s-protein. The covalent modulation it causes is referred to as ***ADP-ribosylation***:



Activity of the ADP-ribosylated α-subunit is selectively modified as below:

1. It loses its GTPase activity.
2. It retains capability of binding GTP, thereby getting activated itself and stimulating *adenylate cyclase*.

Thus, the α-subunit gets activated normally, but remains trapped in this (activated) form only. The protein is "*turned on*" but is "*locked*" in the activated form, having lost its "*turn off*" switch. Persistent stimulation of *adenylate cyclase* follows so that excessive cAMP is generated. Flooding of the cell with cAMP is the natural consequence. This causes profuse secretion of water and electrolytes, resulting in the dreaded diarrhoea.



The ***pertussis*** toxin modifies the α_i-subunit of the G_i by covalent modulation. It causes ADP-ribosylation of this subunit, which makes G_i unable to inhibit the *adenylate cyclase*. Consequently, the latter is "locked" in an activated form, and so the cell is flooded with cAMP. *Bordetella pertussis*, the bacterium which produces pertussis toxin, is the cause of **whooping cough**.

Phosphatidylinositides/Calcium System as Second Messenger (Group IIc Hormones)

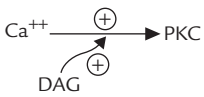
This system generates two intracellular second messengers: *inositol 1,4,5-triphosphate* **(IP_3)** and *diacylglycerol* **(DAG)**. The hormones of group IIc (and some neurotransmitters) that act through this system are shown in Table 29.3. They bind to a cell-surface receptor, and the receptor-hormone complex activates a G-protein of the **Gq** family (Fig. 29.9). The latter activates a membrane bound enzyme *phospholipase C* (PLC), which acts on a variety of substrates, present on the inner leaflet of the plasma membrane. Phosphatidyl inositol 4,5-bisphosphate (PIP_2) is the most important of such substrates because its cleavage forms two intracellular messengers, **IP_3** and **DAG**.

Inositol 1,4,5-Triphosphate (IP_3)

This water soluble second messenger is generated at the plasma membrane initially. It diffuses through the cytoplasm and binds to the IP_3 receptor on the endoplasmic reticulum. This binding *releases intracellular calcium stores*. The Ca^{2+} in turn influences a variety of biochemical processes, particularly through the mediation of calmodulin (Table 29.6).

Diacylglycerol (DAG)

DAG is non-polar and so remains membrane associated, where it functions as a second messenger, activating *protein-kinase-C* (PKC). The latter translocates from the cytosol to the membrane and gets activated by DAG. This activation requires ionic **c**alcium (hence the **C** in designation), and DAG greatly increases affinity of PKC for calcium.



The activated PKC in turn phosphorylates a wide range of target proteins (e.g. *HMG CoA reductase*, *tyrosine hydroxylase*, and several signal transduction molecules), resulting in altering their functional status. Interestingly, *PKC* can phosphorylate the same protein targets as *PKA* but, whereas *PKC* generally phosphorylates the protein at serine residues, *PKA* usually phosphorylates threonine residues.

Table 29.6. Enzymes regulated by calcium/calmodulin

• Adenylate *cyclase*	• Guanylate *cyclase*
• *Phosphorylase kinase*	• Glycogen *synthase*
• Calcium-magnesium ATPase	• Phospholipase A_2
• Pyruvate *carboxylase*	• Pyruvate *dehydrogenase*
• Calcium-phospholipid-dependent protein *kinase*	• *Myosin kinase* • Cyclic nucleotide *phosphodiesterase*

BOX 29.1

cAMP and Calcium Regulate Gene Transcription

Though calcium and cAMP are primarily involved in modulating activities of intracellular proteins and enzymes, they can also enhance gene transcription. Several cAMP regulated genes that have palindromic sequence (TGACGTCA) in the promoter and enhancer regions have been identified. After their release from the tetrameric complex in presence of cAMP, the active catalytic subunits of the *protein kinase A* translocate to the nucleus, where they phosphorylate several transcription factors, including the cell response element binding (CREB) protein. CREB is phosphorylated at a single serine residue (Ser-133). The phosphorylated CREB forms dimers which interact with the above palindromic sequences. The interaction causes stimulation of transcription. Although, the dephosphorylated CREB can also bind with the response element, it can stimulate transcription only after phosphorylation of a single serine residue (Ser-133).

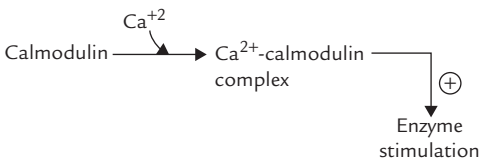
Calcium also acts on the cAMP-regulated genes, but in a different way. A calcium-calmodulin activated *protein kinase* (CaMK-II) phosphorylates a different serine residue (Ser-142) in CREB. This phosphorylation inhibits transcription. In this way, calcium acts antagonistically to cAMP in the regulation of gene expression. In some other cell types, a different type of calcium-calmodulin activated *protein kinase* (CaMK-IV) is present, which can phosphorylate the Ser-133, and thereby bring about transcriptional activation. Thus, in these cells calcium acts synergistically with cAMP in regulation of gene expression.

It is evident from the given discussion that **both second messengers** (*DAG and IP_3*) **act synergistically.** DAG activates PKC by increasing its affinity for calcium, and calcium is made available by IP_3.

Hydrolysis of phosphatidyl inositol bisphosphate by phospholipase C generates two second messengers: **inositol triphosphate,** which opens calcium channels in ER, and **DAG,** which activates *protein kinase* C.

Activation of the *phospholipase C* system also causes influx of calcium (form ECF), to cause further elevation of intracellular calcium concentration. It does so by increasing cell permeability so that calcium moves from high concentration in extracellular space (1–2 mmol/L) into cell (0.1–10 μmol/L).

Calmodulin is a calcium-binding and calcium sensing protein, which is an important mediator of intracellular effect of Ca^{2+}. It has four binding sites for calcium. Occupancy of these sites with calcium activates the calmodulin by inducing alterations in its conformation. The activated calcium-calmodulin complex has ability to interact with a number of enzymes; the interaction changes activities of these enzymes (Table 29.6).



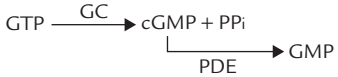
Detailed studies on the enzyme **glycogen phosphorylase** have shown that the enzyme consists of four subunits (α, β, γ, δ), and calmodulin itself constitutes one of the four subunits (δ). Binding of calcium to the calmodulin subunit alters its conformation, which in turn alters conformation of the enzyme protein and hence its activity.

Cytoplasmic calcium is a ubiquitous intracellular signal. Several downstream signalling events mediated by calcium are modulated by a calcium binding protein, calmodulin. Binding of calcium induces a conformational change, allowing calmodulin to bind to and modify the activity of target enzymes.

Though calcium (and cAMP) primarily acts by modifying activities of target enzymes, it may also regulate their synthesis (Box 29.1).

cGMP as Second Messenger (Group IIb Hormones)

cGMP is an important intracellular messenger in retinal cells and some non-retinal cells. Steps involved in the synthesis and degradation resemble the corresponding steps of cAMP. It is generated from GTP by the enzyme ***guanylate cyclase*** (GC), and degraded by the enzyme ***phosphodiesterase*** (PDE), which terminates its action.



The chief action of cGMP in the *non-retinal cells* is to activate ***protein kinase G*** (action of cGMP in retinal cells has been explained in Chapter 18).

BOX 29.2

Atrial Natriuretic Factor (ANF)

ANF is a peptide of 28 amino acids, which relaxes some vascular beds. It acts by stimulating the membrane bound form of GC, which is actually the intracellular domain of the receptor for ANF. The ANF is released from the atrium of heart in response to increased blood volume, increased blood pressure, and increased salt intake. It acts on kidneys to increase the glomerular filtration rate, urine volume and sodium excretion. This effect is supplemented by decrease in release of two hormones which normally increase blood pressure and salt retention in the body: the renin (from juxtaglomerular apparatus) and aldosterone (from zona glomerulosa of the adrenal cortex).

BOX 29.3

Unusual Aspects about NO

There are some unusual aspects about NO as a first messenger. First, it is a gas, unlike most other compounds that act as first messengers. Second, it is a membrane *permeant* paracrine messenger, which diffuses freely not needing any secretory vesicle. NO is highly unstable; it decomposes spontaneously within a few seconds of its formation. Therefore, its action is terminated without needing any enzymes. Finally, it contains an unpaired electron, and therefore, a free radical.

$$\underset{\text{(Inactive)}}{PKG} + 4\ cGMP \rightleftharpoons \underset{\text{(Active)}}{PKG.\ cGMP_4}$$

Activation of PKG occurs by allosteric mechanism. PKG is a serine/threonine *protein kinase*. It mediates most of the cGMP effects by phosphorylating a number of intracellular proteins.

cGMP can also act directly by binding to ion-channels, or by altering activities of some *phosphodiesterases*. However, physiological substrates for cGMP are largely unknown.

In contrast to *adenylate cyclase*, an integral membrane enzyme, the GC exists in two forms: the **membrane bound** form and the **cytoplasmic** form.

1. ***Membrane bound GC:*** This form of GC is actually the intracellular domain of a receptor for the peptide, *atrial natriuretic factor* (ANF). The receptor contains an extracellular ligand-binding domain, a single transmembrane helix, and an intracellular domain that acts as *guanylate cyclase*, all on the same polypeptide. The intracellular GC domain requires binding of ANF to the extracellular domain for its catalytic activity (Box 29.2). Once activated, it generates cGMP that activates the *protein kinase* to mediate most of the cGMP effects, as discussed earlier.

2. ***Cytoplasmic GC:*** The principal GC of the vascular smooth muscles is the soluble cytoplasmic form. It is *not activated by the ANF, but by nitric oxide*; and generates cGMP, which mediates vasodilatation.

Cross-talk: Cyclic nucleotides are not specific for their respective *protein kinases*. Thus, cAMP can activate *PKG* in vascular smooth muscles to cause vasodilatation; and cGMP can excessively stimulate *protein kinase A* to cause fluid secretion and diarrhoea.

Signalling by Nitric Oxide (NO)

Nitric oxide causes vasodilatation and is an important regulator of blood pressure. It acts by stimulating the cytoplasmic GC, thereby increasing cGMP. NO is unique intracellular messenger (Box 29.3) for being membrane soluble. It readily diffuses to nearby cells and increases cGMP level in them. Such a phenomenon occurs in vascular endothelial cells and the nearby smooth muscles: NO synthesized in the endothelium diffuses into the smooth muscle cells where it increases cGMP. The latter relaxes smooth muscles to cause vasodilatation.

NO is synthesized by the enzyme, *NO synthase* in the endothelial cells from one of the nitrogen molecules in the side chain of arginine.

$$\text{Arginine} \xrightarrow[O_2\ \ H_2O]{NADPH} \text{Hydroxy-arginine} \xrightarrow[O_2\ \ H_2O]{\frac{1}{2}\ NADPH} \text{Citrullin} + NO$$

The reaction sequence involves NADPH and the products include NO, citrulline and $NADP^+$. Activity of *NO synthase* is stimulated by calcium-calmodulin. Therefore, the agents that elevate cytoplasmic calcium concentration increase the *NO synthase* activity.

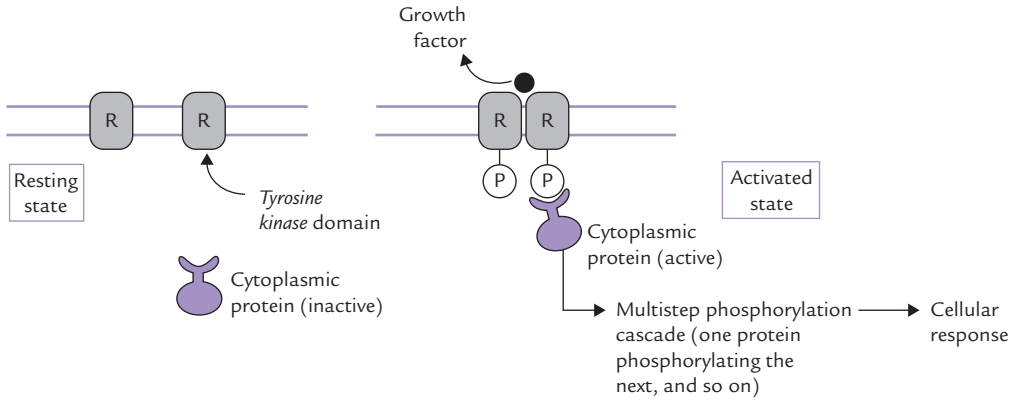


Fig. 29.10. Receptor *tyrosine kinase* activation. Growth factor binding turns on the *tyrosine kinase* activity in the intracellular domain, resulting in its (auto-) phosphorylation. The autophosphorylation enables the intracellular domain to bind cytoplasmic proteins, which get tyrosine phosphorylated, and initiate a phosphorylation cascade, that ultimately leads to cellular response (R = receptor).

cGMP is synthesized either by a membrane-bound *guanylate cyclase* that functions as a receptor for atrial natriuretic factor, or by a cytoplasmic *guanylate cyclase* that is stimulated by nitric oxide. Nitric oxide and cGMP mediate vasodilatation.

The *chemical agents* that generate NO during their metabolism relax smooth muscles. This forms the basis for their use in treatment of angina pectoris and other vascular disorders by some NO generating agents, such as nitroprusside and nitroglycerine. *Physiological agents* that increase intracellular calcium concentration also relax smooth muscle by the same mechanism (i.e. by stimulating NO synthesis). Histamine, acetylcholine, bradykinin, thrombin and ATP are some examples of such agents. On the other hand, the agents that decrease NO generation by inhibiting *NO synthase* cause elevation of blood pressure. These observations amply illustrate that NO is an important regulator of blood pressure under ordinary circumstances.

Protein Kinase or Phosphatase Cascade as Intracellular Messenger

Hormones of this group, such as insulin, and a variety of growth factors, interact with receptors whose C-terminal intracellular domains have *tyrosine kinase* activity. After binding of ligand to the extracellular domain, the intracellular domain of these receptors cause auto-phosphorylation of proteins on tyrosine side chains (Fig 29.10). This is termed as *intrinsic ligand activated tyrosine kinase activity*. Phosphorylation at tyrosine side chains initiates a multistep phosphorylation cascade that may involve several *protein kinases*, *phosphatases* and other regulatory proteins. This leads to generation of several gene products, which mediate the hormone action. Like other intracellular signals, activities of these *protein kinases* must be turned off after they have delivered their message. The off switch is provided by *tyrosine phosphatases*, enzymes that dephosphorylate phosphotyrosyl residues. Defect in these events may lead to diseases (**Case 30.5**).

Several receptors—generally those involved in binding of ligands involved in growth control differentiation and the inflammatory response either have *intrinsic tyrosine kinase* activity or are associated with cytoplasmic proteins that are *tyrosine kinase*. The *tyrosine-kinase* activity is different from those of *protein kinases* discussed so far (e.g. the hormone-stimulated *protein kinase A, C* and *G*) which phosphorylate serine and threonine residues in their substrates.

IV. Neurotransmitters

Like hormones, neurotransmitters are also important **extracellular messengers**. They bring about *transmittance of a targeted message from a neuron to the responding cells*. A neurotransmitter is released by a neuron at synapse to effect another post-synaptic neuron or organ (e.g. heart, lungs). Its effect is confined to the synapse and, unlike hormones, it is into released into the bloodstream. In addition to amino acids and their derivatives (described in Chapter 4), other compounds that may serve as neurotransmitters are peptides, purines, gases (e.g. NO), etc.

- *Peptides*: A number of physiologically active peptides, including the enkephalins, cholecystokinin, octapeptide, vasoactive intestinal peptide (VIP) somatostatin, substance P (an undecapeptide in primary afferents and the basal forebrain) are well-established neurotransmitters.

- *Purines*, such as ATP and adenosine in CNS and sympathetic nerves.
- *Nitric oxide*, which is derived from arginine in CNS and genitourinary system.
- *Miscellaneous*, such as acetylcholine that acts in parasympathetic nerves and CNS.

Exercises

Essay type questions

1. What do you understand by second messenger concept in hormone action?
2. Discuss the second messenger mechanism of G-proteins in signal transduction.
3. Give schematic diagram of insulin receptor, mention its important properties and discuss its functional significance.
4. Describe the mechanism of action of epinephrine at molecular level.
5. Describe the activity of each component of the signalling pathways based on *adenylate cyclase, receptor tyrosine kinases* and phosphoinositides.

Write short notes on

1. Second messenger
2. Down regulation
3. Nitric oxide
4. Insulin receptor
5. Cholera toxin
6. Pertussis toxin
7. Target tissue concept
8. Feedback concept
9. Pseudohypoparathyroidism
10. Cyclic GMP

CLINICAL CASES

CASE 29.1 Listlessness and cold intolerance in a 37-year-old man

A 37-year-old civil servant reported in the medical OPD with complaints of weakness, constipation, weight gain and easy fatigue. He had slowed up mentally, become indecisive and had stopped his daily morning walk. He was becoming intolerant to cold and insisted to switch off the fans and to keep the windows closed. Often he felt that he did not have enough energy even to perform the routine activities. Physical examination revealed sallow skin and coarse, lack-luster hair. He was mildly hypertensive: blood pressure was 122/95, and pulse rate was 66 per minute. The peripheral reflexes were tested and a slow quadriceps relaxation was observed. There was no goitre.

Investigations:

Test	Patient's report	Reference range
Serum triiodothyronine (T_3)	40 ng/100 mL	70–180 ng/100 mL
Serum thyroxine (T_4)	3.8 μg/100 mL	5.5–12.0 μg/100 mL
Serum TSH	0.7 U/L	< 5.7 U/L

TSH injection restored the circulating T_3 and thyroxine levels to normal. Injection of pyroglutamylhistidylprolinamide did not produce any change in serum levels of T_3, T_4 or TSH. The patient died in a motor accident; the pituitary gland was recovered at autopsy.

Q.1. Is this a case of primary hypothyroidism?
Q.2. Pinpoint the pathological lesion in the patient?
Q.3. What further test would you perform to confirm the above diagnosis?

CASE 29.2 A 12-year-old boy with bulging eyeballs

A 12-year-old boy presented in the medical OPD with swelling in neck, which was first noticed about three months back by his mother. She was concerned that his eyes appeared more prominent, although there was no deterioration in vision; and that he was losing weight in spite of a ravishing hunger. He had excessive moistness of skin, about which he felt embarrassed while he was among his friends. Family physician's report showed that he had always been nervous, emotionally labile, very thin despite extraordinary appetite; he also had tremors, frequent bowel movements, excessive sweating, and his cardiac rhythm showed disturbances.

On examination, the boy appeared anxious, restless, and fidgety. The skin was warm and sweaty. A fine tremor of fingers and tongue together with hyper-reflexia were present. The eyeballs were bulging with widened palpebral fissures; blinking was infrequent with a lid lag. The heart rate was rapid at 140 per minute, pulse was irregular and the pulse pressure was increased. Enlargement of the thyroid gland to about three times its normal size was detected.

Investigations:

Test	Patient's report	Reference range
Serum thyroxine	18.1 μg/100 mL	5.5–12 μg/100 mL
Serum triiodothyronine (T_3)	260 ng/100 mL	70–180 ng/100 mL
Serum TSH	15.1 U/L	< 5.7 U/L
Serum cholesterol	161 mg/dL	150–250 mg/dl
Plasma glucose (fasting)	78 mg/dL	< 100 mg/dl

Magnetic resonance image of the pituitary was normal.

Q.1. What is the most likely diagnosis?
Q.2. What further tests can be performed to substantiate the above diagnosis?

CASE 29.3 A 4-year-old child with rachitic skeletal changes

A 4-year-old boy was brought to OPD with skeletal deformities suggestive of rickets. He appeared extremely weak and was not able to walk without support. On examination, severe muscular weakness was detected. Radiological examinations showed widening of distal ends of the shafts of long bones. Bone mineralization was also found to be decreased.

Investigations:

Test	Patient's report	Reference range
Serum calcium	7.2 mg/dL	8.5–10.5 mg/dL
Serum parathormone	82 ng/dL	10–65 ng/dL

These results were in line with the above tentative diagnosis. However, serum calcitriol level was normal. This was unexpected since in rickets serum calcitriol falls.

Treatment with calcitriol resulted in improvement in symptoms only when this steroid was administered in very high doses. Fibroblasts were obtained by skin biopsy and cultured. Addition of radioactive calcitriol to cultured fibroblasts was followed by little binding of the two. Addition of larger amounts of calcitriol, however, resulted in increased binding.

Q.1. Identify the biochemical defect.
Q.2. Explain biochemical basis of the clinical features and the test results of this child.
Q.3. What exactly is wrong with the defective protein of this child?

CASE 29.4 A 5-year-old child with skeletal deformities

A 5-year-old boy had protuberant abdomen and was unable to walk with a steady gait. On examination, profound muscular weakness was detected. Skeletal deformities were also observed, suggesting a possible diagnosis of rickets. However, results of biochemical investigations were remarkably different.

Investigations:

Test	Patient's reports	Reference range
Serum phosphate	5.2 mg/dL	2.5–4.5 mg/dL
Serum calcium	8.1 mg/dL	8.5–10.5 mg/dL
Serum parathormone	56 ng/L	10–65 ng/L

Renal biopsy was performed. Production of cAMP in the isolated cells, following addition of parathormone, was measured. It was found to be less compared to the cAMP production in the cells obtained from a normal subject.

Q.1. Point out the biochemical defect in this patient.
Q.2. Could there be some defect in G-protein in this patient? How would you investigate?

CASE 29.5 Glucose intolerance and hyperpigmented skin patches in an obese child

A 7-year-old girl presented with complaints of generalized weakness, lethargy, and tiredness. She was unusually obese and had brown to black, velvety hyperpigmented, thickened skin, most prominently on the posterior and lateral folds of neck. Other areas involved, though to a lesser extent, were elbows, knuckles, dorsum of the neck, and the skin fold areas in the axilla and the groin. Blood sample was sent to the biochemistry laboratory for analysis. Random blood glucose was elevated (210 mg/dL). Other test results are as below:

Investigations:

Test	Patient's report	Reference range
Blood glucose		
Fasting	168 mg/dL	< 100 mg%
2 h after glucose load of 75 g	228 mg/dL	< 140 mg%
Serum insulin	67 μU/mL	6–24 μU/mL
Fasting		

The child was admitted in the medical ward and insulin treatment was started. Blood glucose was monitored daily for stabilizing the daily dose. A high dose (200 units a day) was required for controlling the hyperglycaemic state.

Q.1. Identify the biochemical defect in the child.
Q.2. What further test would you perform for evaluation of the above biochemical abnormality?
Q.3. Suggest a test for identifying the precise cellular defect.

CHAPTER

30

ACTION OF HORMONES

Overall regulatory role of the endocrine system was discussed earlier (Chapter 29), emphasizing the working of hormones at cellular and molecular levels, the concept of membrane and intracellular receptors, the nuclear receptors, the second messengers, and their mode of action. The present chapter will focus on arrangement of various endocrine systems, and explain the biosynthesis, chemical nature, and functions of individual hormones. The clinical syndromes arising because of overactivity or underactivity of individual hormones will also be discussed.

By the end of this chapter the student should know:

- Arrangement of hypothalamus-pituitary hormone system, chemical nature and effects of the hypothalamic factors, the trophic and non-trophic hormones elaborated by anterior pituitary and the posterior pituitary hormones.
- Biosynthesis, chemical nature, and functions of the hormones elaborated by thyroid, adrenals, gastrointestinal tract and gonads; and explain biochemical basis and signs and symptoms of diseases associated with the hormone over- or under-production, e.g. Addison's disease, Graves' disease, Cushing's syndrome, etc.

Several other endocrine systems are also present in the human body, which are not considered in this chapter. Some of these are discussed in previous chapters as part of the physiologic function they control. Thus, the student is referred to Chapter 10 for glucose homeostasis, Chapter 19 for water and electrolyte balance, and Chapter 31 for calcium and phosphorus metabolism.

I. Hypothalamus-pituitary System

A significant group of hormones is secreted by peripheral endocrine glands in response to signals from the brain. The brain structures responsible for initiating such signals are the **hypothalamus** and the **pituitary gland** (Fig. 30.1). Pituitary gland or hypophysis is an oval organ weighing about 0.6 g and is encased in the body cavity of the skull (sella turcica) below the brain. It is demarcated into two parts: the anterior pituitary (***adenohypophysis***), which accounts for 80% of the gland, and the posterior pituitary (***neurohypophysis***). The two lobes are connected by **pars intermedia**, which is rudimentary in humans. Pituitary gland lies below the hypothalamus and is connected with it via the **pituitary stalk**, which contains a complex array of axons and blood vessels.

Two connecting pathways are present in the pituitary stalk: (a) a bundle of nerve fibres, leading into the posterior pituitary, and (b) a vascular network, the *hypophyseal portal system,* joining the anterior pituitary with the hypothalamus. Various hypothalamic hormones, and some pituitary hormones are initially synthesized in hypothalamus and transported to pituitary via these two connecting pathways.

The hormones of posterior pituitary are initially synthesized and packaged in the supraoptic and paraventricular nuclei of the hypothalamus, and transported along the nerve fibres to posterior pituitary for storage prior to release in circulation. The hypothalamic release or release-inhibiting hormones are also synthesized in various nuclei of the hypothalamus, but they travel via the portal system to the anterior pituitary, where they elicit the secretion of various trophic or non-trophic hormones. The anterior pituitary hormones act on target

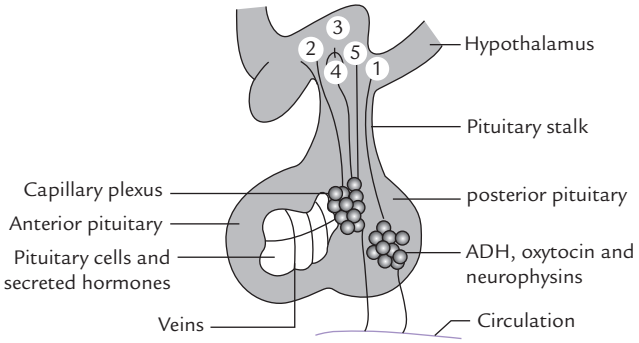


Fig 30.1. The hypothalamo-pituitary regulatory system. The posterior pituitary hormones are synthesized in supraoptic and paraventricular nuclei (1); and the anterior pituitary releasing/or release-inhibiting hormones in various hypothalamic nuclei (2–5). The posterior pituitary hormones are transported (from hypothalamus) along nerve fibres to posterior pituitary, whereas anterior pituitary hormones travel with portal system to anterior pituitary.

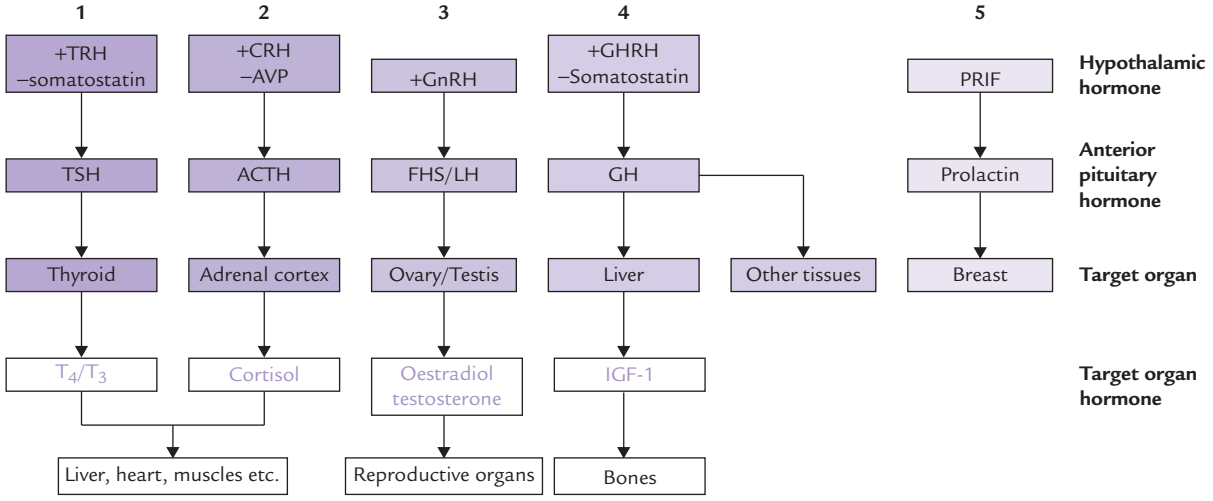


Fig. 30.2. The hypothalamo-pituitary-target organ axes (1–5). See text for the expanded forms of abbreviations.

endocrine and non-endocrine organs to regulate a broad range of biological functions. Figure 30.2 depicts a bird's eye view of the hypothalamo-pituitary regulatory system.

The release, and in some cases production, of each of the pituitary hormones is under tonic control of at least one hypothalamic release or release-inhibiting hormone.

II. Hypothalamic Hormones

At least **seven hormones** are secreted by the hypothalamus. Since they are involved with homeostatic mechanisms, their secretion is not constant, rather they are released in bursts (i.e. pulsatile and episodic release). Their **secretion is under tight feedback control of the pituitary hormones** or the hormones secreted by peripheral endocrine glands. Stimuli from central nervous system also controls secretion of hypothalamic hormones. There are five separate endocrine axes, as shown in Figure 30.2.

Actions of the hypothalamic hormones on their target cells in the pituitary are mediated by (a) **calcium phosphoinositide** system or (b) via **cAMP second messenger system**.

Thyrotropin-releasing Hormone (TRH)

It is a modified tripeptide (pyroglutamate-histidine-proline) synthesized as a 26 kDa prohormone. TRH stimulates synthesis and secretion of thyroid-stimulating hormone (**TSH**) from the thyrotroph cell. It binds to the cell surface receptors that are linked to *phospholipase-C*. The resulting phosphoinositides stimulate the release of calcium from intracellular storage sites and so lead to exocytosis of TSH. The latter stimulates synthesis and release of the thyroid hormones (T_3 and T_4).

Corticotropin-releasing Hormone (CRH)

It is a 41 amino acid peptide secreted by the paraventricular nucleus. It acts via the cAMP second messenger system to stimulate release of the pituitary adrenocorticotropic hormone (**ACTH**). The latter in turn acts on

adrenal cortex to stimulate release of adrenocorticosteroids, mainly cortisol.

Arginine Vasopressin (AVP)

It is a nine amino acid peptide, synthesized by both supraoptic and paraventricular nuclei. It elicits an **inhibitory effect** on pituitary, opposite to that of CRH, by altering the intracellular calcium-ion channels.

Gonadotropin-releasing Hormone (GnRH)

It is a 20 amino acid peptide synthesized as a 92 amino acid precursor by various hypothalamic nuclei. It has a half-life of about 3 minutes and stimulates the synthesis and secretion of both follicle-stimulating hormone (**FSH**) and luteinizing hormone (**LH**) from the same gonadotroph cell-type in the anterior pituitary. Its action is mediated via calcium-phosphoinositide system.

Growth Hormone-releasing Hormone (GHRH)

It is a 44 amino acid peptide, synthesized as part of a 108 amino acid prohormone. It binds to its receptor on the pituitary somatotroph cell and triggers both the adenylate *cyclase* and intracellular calcium-calmodulin system to stimulate **GH** transcription and release. GH in turns promotes soft tissue and bone growth.

Growth Hormone Release-inhibiting Hormone (GHRIH)

In view of the wide range of action of GH, it is not surprising that an inhibitory hormone also exists, in addition to a stimulatory one to influence the GH secretion. GHRIH, also called **somatostatin**, is found in two forms—with 14 and 28 amino acids; both are produced from a 116 amino acid gene product. They inhibit GH secretion by lowering intracellular cAMP concentration. *Inhibition of TSH secretion is also brought about by somatostatin.*

Prolactin Release-inhibiting Factor (PRIF)

Prolactin (PRL) is unique among the pituitary hormones in that it is under predominant inhibitory control from the hypothalamus. Furthermore, the controlling agent (PRIH) is a very simple molecule, **dopamine**, which works by inhibiting *adenylate cyclase. Dopamine inhibits both prolactin synthesis and secretion.*

Some of these hypothalamic hormones are synthesized in other tissues also. **Somatostatin**, for example, is found **α-cells of pancreas** in higher concentrations than in hypothalamus, and it plays an important role in regulating secretion of insulin and glucagon from islet of Langerhans.

III. Anterior Pituitary Hormones

Anterior pituitary or adenohypophysis is an endocrine target organ for hypothalamic hormones. The hypothalamic hormones reach pituitary via **portal system** and induce cause it to release its hormones into the general circulation. By conventional staining three cell types were identified in anterior pituitary: chromophobe, eosinophil and basophil. The cells are now classified by immunological means, and at least **five cell types** have been identified—somatotrophs secrete GH, lactototrophs secrete PRL, thyrotrophs secrete TSH, gonadotrophs secrete LH and FSH, and corticotrophs secrete ACTH.

The hormones are secreted as either peptides (ACTH, GH, PRL, chorionic somatomammotropin) or glycoproteins (TSH, LH, FSH).

A. The Master Gland

Four of the seven pituitary hormones act on target endocrine organs, and are termed as trophic hormones. These are: *TSH*, *ACTH*, *LH* and *FSH*. The remaining are non-trophic hormones, which act primarily on non-endocrine target tissues or organs. Because of widespread actions of these hormones, the anterior pituitary affects growth, thyroid activity, sexual function, lactation, metabolism of carbohydrate, protein and fat, as well as skin pigmentation (Fig 30.2).

In view of its influence on a vast variety of biochemical processes, pituitary is known as the foremost or master endocrine gland. Accordingly, excess or deficiency of pituitary hormones influence a vast range of functions. Excess is usually due to a pituitary tumour and almost always involves only one hormone. Isolated deficiencies of pituitary hormones occur but are rare, with GH being most common. More commonly, however, the pituitary hormone deficiencies involve more than one, and eventually all the anterior pituitary hormones (**panhypopituitarism**). There is generalized endocrine dysfunction, and resultant symptoms relate to thyroid or adrenal insufficiency, or there is loss of sexual function.

B. Classification

Based on structural characteristics, the anterior pituitary hormones have been traditionally divided in three broad groups:

1. ***The growth hormone-prolactin-chorionic somatomammotropin group***: *Growth hormone* (*GH*; somatotropin), *prolactin* (*PRL*) and *chorionic somatomammotropin* (*CS; placental lactogen*) are protein hormones, exhibiting

considerable structural similarities with one another. They are believed to have originated by duplication of a single ancestral gene. The last one (CS) has no definite function in humans.

2. ***The glycoprotein hormones***: Three anterior pituitary hormones, *thyroid-stimulating hormone* (*TSH*), *follicle-stimulating hormone* (*FSH*), and *luteinizing hormone* (*LH*) are glycoprotein in nature. The last two are known as gonadotropins. Human chorionic gonadotropin (hCG), a placental glycoprotein hormone, resembles LH in nature.
3. ***The pro-opiomelanocortin peptide family*** **(POMC):** It is a 241 amino acid precursor protein (MW 31,000), which is cleaved to yield a number of hormonally active peptides, including the *adrenocorticotropic hormone* (*ACTH*), *endorphins* and *melanocyte-stimulating hormone* (*MSH*). ACTH regulates synthesis and secretion of adrenal cortical hormones, endorphins are endogenous analgesics and MSH regulates pigment production by melanocytes.

Growth Hormone (GH)

GH elicits a wide range of actions in regulating growth and intermediary metabolism. Gene for GH is located as a part of the GH-CS gene family in region q 22–24 on the long arm of chromosome 17.

Chemistry

Human growth hormone is a peptide comprising 191 amino acids (MW 21,000) that is synthesized initially as a 29 kDa prohormone. Nearly two-third of GH in circulation is associated with a binding protein that is identical to the extra-cellular domain of the GH receptor. This binding protein prolongs the half-life of GH in plasma (20 minutes). The normal human pituitary contains approximately 10 mg of GH, less than 5% of this is released each day.

Biochemical Functions

GH is essential for postnatal growth and for normal carbohydrate, lipid, protein and mineral metabolism (Fig. 30.3).

1. ***Effect on growth:*** GH interacts with its receptor to induce complex series of intracellular events that lead to transcription of many enzymes and peptides, called **growth factors** or **somatomedins**, which mediate the growth promoting effects of GH. These compounds are synthesized primarily by the liver under the influence of GH and have strong metabolic and mitogenic effects especially on *cartilage, adipose tissue* and *striated muscle*. Two types of growth factors are known: insulin-like growth factor-I (IGF-I) and insulin-like growth factor-II (IGF-II). ***IGF-I***, also known as ***somatomedin C*** (formerly sulfation factor), is a 70 amino acid straight-chain basic peptide produced by the liver in response to GH. It is more active than IGF-II, a 67 amino acid peptide, and seems to be directly related to growth. It acts through receptors similar to insulin receptors and is linked to intracellular *tyrosine kinase* activity.

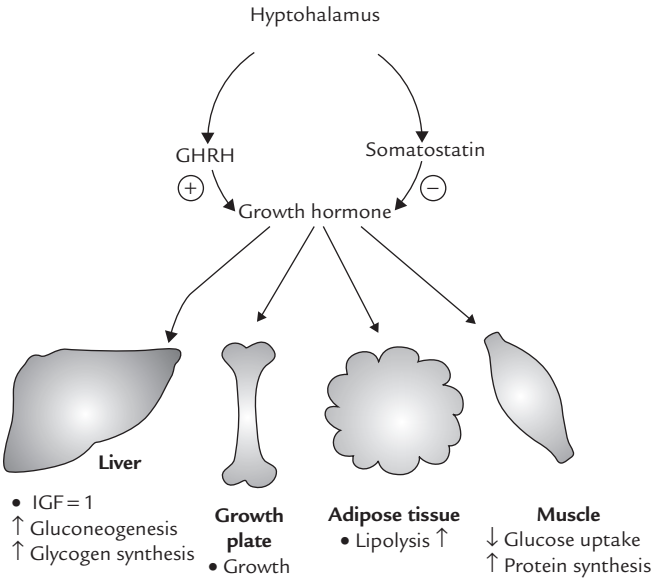


Fig. 30.3 Anabolic effects of growth hormone on various tissues.

2. ***Effect on protein metabolism***: GH promotes protein synthesis and cell proliferation in both skeletal and non-skeletal tissues. It increases cellular up-take of amino acids and stimulates transcription. Overall, its growth promoting effects are insulin-like, and tend to turn the nitrogen-equilibrium towards *positive nitrogen balance*.
3. ***Effect on lipid and carbohydrate metabolism***: Effects of GH on lipid and carbohydrate metabolism are *opposite to those of insulin*, and not mediated via IGF-I. GH promotes glycogenesis and gluconeogenesis in liver and decreases utilization of glucose in muscles and other tissues. These actions result in increase of blood glucose level, hence GH is a diabetogenic hormone.

   GH *promotes lipolysis in adipose* tissue to elevate the circulating free fatty acid levels. Oxidation of fatty acids in liver is increased and ketogenesis is enhanced, particularly in diabetes.
4. ***Effect on mineral metabolism***: GH increases incorporation of calcium and phosphate in growing bones. This effect is probably mediated via IGF-I and results in increase in bone growth and its mineralization. Incorporation of sodium, magnesium and chloride in bones and of sulphates in cartilage is also increased.
5. ***Effect on lactation***: GH has structural similarity with prolactin, which accounts for its prolactin-like effect of increasing milk production in lactating animals.
6. ***Others***: GH stimulates cellular differentiation and augments actions of other trophic hormones.

Regulation

GH is controlled by a dual system, namely growth hormone-releasing hormone (GHRH) and growth hormone release-inhibiting hormone (GHRIH), which respectively increases and decreases the release of GH. GHRH, also called **somatocrinin**, is a peptide with 44 amino acids, and GHRIH or **somatostatin** is a peptide with 14 amino acids and an S-S bond.

GH secretion is greatly enhanced during sleep when the plasma GH concentration may reach up to 100-fold greater than the baseline. This lends a scientific touch to the proverb *"Sleep well and grow well"*. Stress, exercise, food intake and obesity also influence GH secretion.

GH inhibits its own secretion via a *short feedback loop*. IGF-I also causes inhibition of GH by a dual mechanism: (a) it inhibits GHRH secretion, and (b) stimulates somatostatin secretion (Fig. 30.4).

GH is maximally secreted in children and young adults, chiefly during sleep. IGF-I has a long loop effect on the elaboration of growth hormone from anterior pituitary.

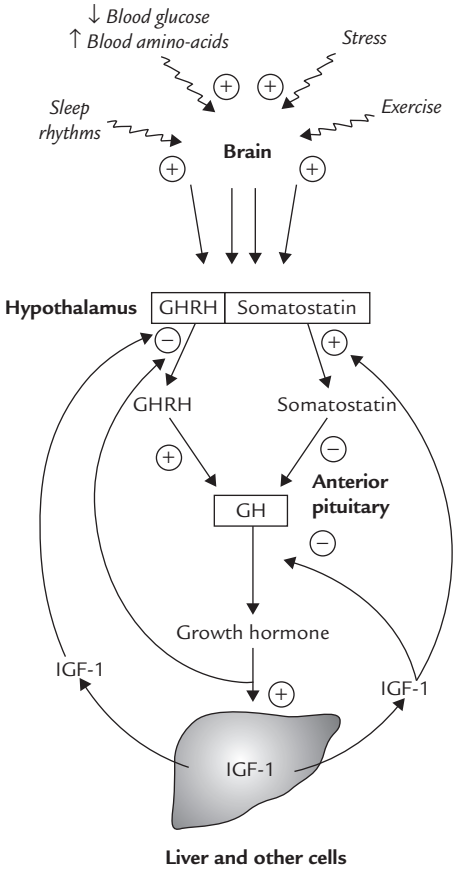


Fig. 30.4. Various factors stimulate release of GHRH and somatostatin (GHRIH) from hypothalamus. GHRH is stimulatory and somatostatin is inhibitory to release of growth hormone from anterior pituitary. GH and IGF, participate in feedback loops (GHRH = somatocrinin, GHRIH = somatostatin).

Clinical Disorders of GH Secretion

GH deficiency in children results in short stature, while GH excess in children leads to **gigantism**. In adults, since the epiphyses of the long bones have closed, linear growth is not possible. Therefore, the adult form of GH excess, called **acromegaly**, manifests by thickening of tissues and a change of facial appearance. GH excess is almost always due to GH-secreting pituitary adenoma, although ectopic GHRH from a pancreatic tumour has been described. GH deficiency may occur as an isolated problem or as part of the general picture of *panhypopituitarism*.

Prolactin

Chemistry

It is a polypeptide of 199 amino acids (MW 20,000), whose amino acid composition is similar to that of GH (16% sequence homology). It is synthesized in pituitary **lactotrophs** as a larger prohormone by transcription of a gene located on chromosome 6.

Regulation of PRL Secretion

Secretion of prolactin from anterior pituitary is increased by a hypothalamic hormone, prolactin release factor (PRF). Prolactin release-inhibiting factor (PRIF), most likely dopamine, inhibits prolactin secretion. Dopamine interacts with receptors on pituitary lactotroph (called D2 receptor) to inhibit *adenylate cyclase* and consequently inhibits both prolactin synthesis and secretion.

Biochemical Functions

In association with other pregnancy-related hormones, prolactin assists initiation and maintenance of lactation, and promotes full development of breast tissue. It binds to its receptor in mammary tissue and stimulates the synthesis of several milk proteins, including lactalbumin.

Thyroid-stimulating Hormone (TSH)

TSH is synthesized in the pituitary **thyrotrophs** and acts on thyroid gland to increase synthesis and secretion of thyroid hormones. It is also called thyrotrophic hormone or **thyrotropin**.

Chemistry

TSH is a 28 kDa glycoprotein (MW 30,000). The carbohydrate portion is a complex mixture of acetylated sugars, sialic acid and sulphate. The protein portion is made of non-covalently linked **α** and **β** chains. The former comprises 92 amino acids and is identical to the α-chain present in other pituitary glycoprotein hormones (FSH and LH). The β-chain differs in the three hormones and determines biological specificity of each.

Biochemical Functions

TSH acts on thyroid gland and influences virtually every aspect of thyroid hormone biosynthesis and secretion, as described later. It acts via a specific membrane receptor on thyroid cell and stimulates *adenylate cyclase* with a consequent increase in cAMP level.

Regulation

The release factor for TSH is the hypothalamic thyrotropin releasing hormone (TRH). Circulating thyroid hormones inhibit secretion of TSH (also of TRH) by feedback mechanism.

Thyrotropin-releasing hormone (TRH) is synthesized in the hypothalamus and transported via portal circulation to the pituitary where it causes release of TSH. The latter acts on thyroid and enhances synthesis and release of thyroid hormones (T_3 and T_4).

Gonadotropins

FSH and LH are called gonadotropins as they stimulate the development and secretory activity of gonads-testis and ovary. Both are secreted in the male and female from the same cell type, the gonadotrophs.

Chemistry

FSH and LH are both glycoproteins with molecular weight of approximately 28 kDa. Each comprises an identical α-subunit (shared with TSH) and a specific β-subunit. The β-subunit of both FSH and LH is composed of 115 amino acids and each has two carbohydrate chains.

The gene for FSH β-chain is on chromosome 11, while that for LH β-chain is on chromosome 19. The latter is located close to the gene for the β-subunit of human chorionic gonadotropin (hCG), with which it has considerable homology.

Biochemical Functions

Both LH and FSH activate *adenylate cyclase* and increase cAMP production, through which their effects are mediated.

In males: LH and FSH act concertedly to promote spermatogenesis in testes.

- LH induces testosterone production by Leydig cells (interstitial cells) of the testes. Testosterone brings about the meiotic divisions necessary for spermatogenesis.
- FSH acts on Sertoli cell of the testes and induces synthesis of several proteins, including androgen-binding protein (ABP) and inhibin. ABP is secreted into the seminiferous tubule lumen where it binds testosterone (or its active form dihydrotestosterone). This ensures high local testosterone concentration which promotes spermatogenesis (Fig. 30.5).

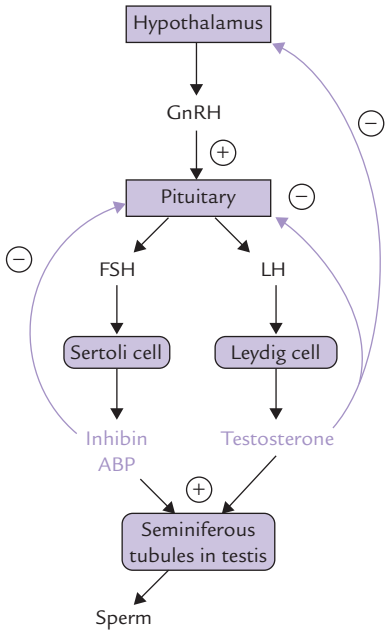


Fig. 30.5. Control of the Hypothalamic-pituitary-gonadal axis (ABP = androgen-binding protein, GnRH = gonadotropin-releasing hormone).

- Inhibin participates in negative feedback loop to inhibit secretion of FSH from anterior pituitary.

In females: FSH and LH have different functions in females, on the basis of which they have been given names. FSH promotes oestradiol synthesis leading to follicular maturation, while LH leads to follicle rupture and oocyte release. Further details are given later (see menstrual cycle).

Follicle-stimulating hormone and luteinizing hormone regulate testicular functions in males, though they are given their names on the basis of their function in females.

Regulation

Gonadotropin-releasing hormone (GnRH) regulates the secretion of FSH and LH (Fig. 30.5). The pituitary secretion of these hormone is subject to feedback regulation by the target organ hormones.

- LH is inhibited by the target gland sex hormones, testosterone and oestradiol. The latter inhibits pituitary LH, whereas testosterone inhibits both hypothalamic GnRH and pituitary LH.
- FSH is subject to feedback inhibition by inhibin.

Regulation is more complex in females and is linked with menstrual cycle, discussed later.

Human chorionic gonadotropin (hCG) is a placental glycoprotein, structurally similar to LH. It is secreted by

BOX 30.1

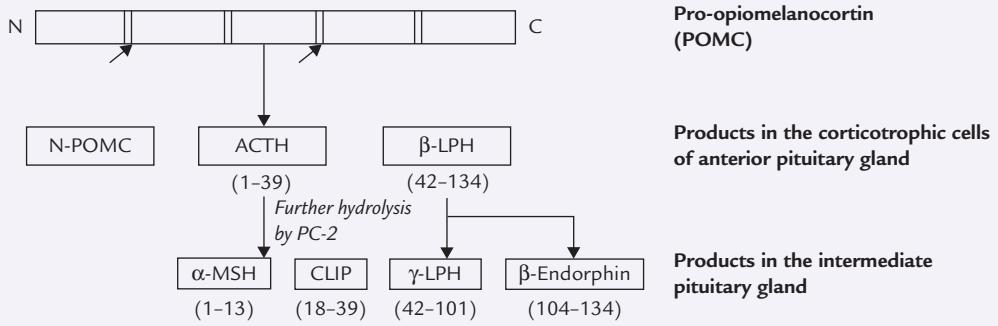
Pro-opiomelanocortin (POMC)

POMC comprises 285 amino acid residues (MW 31000) and serves as a precursor of β-endorphin, ACTH and various forms of melanocyte-stimulating hormones in the pituitary gland. As shown in the diagram below, the main products in anterior pituitary are the N-terminal portion of the molecule (N-POMC), adrenocorticotropin (ACTH) and β-lipotropin (β-LPH). No further processing of these peptides occurs in human pituitary gland; the intermediate lobe of pituitary, which is capable of further processing these peptides being rudimentary in humans.

In the intermediate lobe of pituitary of those animals that have such a structure (including human fetus), and in hypothalamus of many species, both ACTH and β-LPH undergo further hydrolysis to yield other active peptides. These include α-melanocyte-stimulating hormone (α-MSH) and corticotropin-like intermediate lobe peptide (CLIP) from ACTH; and β-endorphin and γ-LPH from β-LPH.

Adrenocorticotropin (ACTH): This has been discussed

β-Endorphin: This is 31-amino acid peptide which includes behavioural changes when injected in the third ventricle in animals. It probably acts, along with enkephalins, to produce stress analgesia. In neurons of central nervous system, it acts as a neurotransmitter and a neuromodulator. When present in blood circulation, β-endorphin has no effect.



||Paired basic amino acid residues, sites of hydrolysis shown by arrows

Melanocyte-stimulating Hormone (MSH) They stimulate the melanin producing melanocytes. In humans, where the intermediate lobe of pituitary is not well developed, the MSH activity is contained in the larger fragments ACTH, β-LPH and γ-LPH.

Corticotropin-like Intermediate Lobe Polypeptide (CLIP) It is a biologically inactive peptide.

Lipotrophin (β- and γ-LPH) These peptides have stimulatory effect on *hormone sensitive lipase*. It is now understood that the differential processing of POMC is because of the presence of different combinations of processing enzymes in different tissues. In anterior pituitary gland, the proteolytic enzyme, PC-1 alone is present, which cleaves the POMC into three fragments. In intermediate lobe (of pituitary) and hypothalamus, the enzyme PC-2 is also present in addition to the closely related PC-1. The PC-2 causes further cleaving of the ACTH and the β-LPH.

fertilized ovum; the secretion begins immediately after implantation. It maintains corpus luteum during pregnancy. Its detection in urine forms the basis for **diagnosis of pregnancy**.

Adrenocorticotropin (ACTH)

Chemistry

ACTH is formed in pituitary corticotrophs from first 39 amino acids of POMC. POMC is cleaved to release several hormonally active peptides including ACTH (Box 30.1).

Adrenocorticotropin stimulates synthesis and release of glucocorticoids and to some extent mineralocorticoids from the adrenal cortex. The biological activity of ACTH resides in the N-terminal 24 moieties.

Biochemical Functions

ACTH acts via the cAMP second messenger system. The major effects are:

1. **Synthesis and release of steroid hormones from the adrenal cortex** is induced by ACTH. The conversion

of cholesterol to pregnenolone, the first step in steroidogenesis, is enhanced by ACTH. This leads to synthesis of glucocorticoids, and to a lesser extent other steroid hormones as well.

2. Prolonged exposure to high ACTH levels causes **lipolysis** in adipocytes and increased insulin secretion from the pancreas.

Regulation

The ACTH secretion from pituitary mainly occurs under the influence of

(a) two hypothalamic peptides, corticotropin-releasing hormone (**CRH**) and arginine vasopressin (**AVP**), both hypothalamic peptides are under the influence of cerebral factors, and
(b) negative feedback effect of adrenal cortex hormones.

- **CRH** is a 41 amino acid peptide secreted by the paraventricular nucleus, which has a stimulatory effect on ACTH secretion. It acts via the cAMP second messenger system.

  AVP, is a nine amino acid peptide, which **inhibits** the ACTH secretion. It is synthesized by both the supraoptic and paraventricular nuclei, and it acts by altering intracellular calcium ion channels (Fig. 30.2).
- The **negative feedback** occurs by high concentrations of glucocorticoids, particularly cortisol, which inhibits the secretion of ACTH as well as CRH. Further details about this hypothalamo-pituitary-adrenal axis are given in a subsequent section of this chapter (see Fig. 30.11).

Secretion of each of the anterior pituitary hormones is influenced by a stimulus provided by the hypothalamus; in all, there are five separate endocrine axes within the hypothalamo-anterior pituitary regulatory system.

IV. Hormones of Posterior Pituitary

Posterior pituitary is a source of the circulating vasopressin and oxytocin. The principal action of vasopressin is to increase reabsorption of water by the renal tubules, and because of this action it is also known as **antidiuretic hormone** (ADH). Oxytocin is important in parturition and lactation in females; its role in males is unknown.

Both the posterior pituitary hormones are non-apeptides with disulfide bonds, differing from each other by two amino acids only. They are actually synthesized (and packaged) in hypothalamus: oxytocin in supraoptic nucleus and ADH in paraventricular nucleus. *From hypothalamus they travel as granules down the axons of hypothalamohypophyseal tract into the posterior pituitary, where they are stored.* In hypothalamus, both the hormones are produced as larger precursors: vasopressin as pre-pro-vasopressin and oxytocin as pre-pro-oxytocin (Fig. 30.6). The former contains a carrier protein neurophysin II, and likewise pre-pro-oxytocin contains neurophysin I. Upon reaching neurohypophysis they generate ADH and oxytocin by proteolysis.

Hormones of posterior pituitary (oxytocin and vasopressin) are actually synthesized in specific nerve cells in the hypothalamus (supraoptic and paraventricular nuclei) and travel along axons into posterior pituitary, where they are stored, prior to release.

Oxytocin

Biological activities of oxytocin are linked to parturition (child-birth) and expression of milk from breast. The hormonal release from posterior pituitary occurs in response to nipple stimulation (suckling). As with prolactin, *suckling*

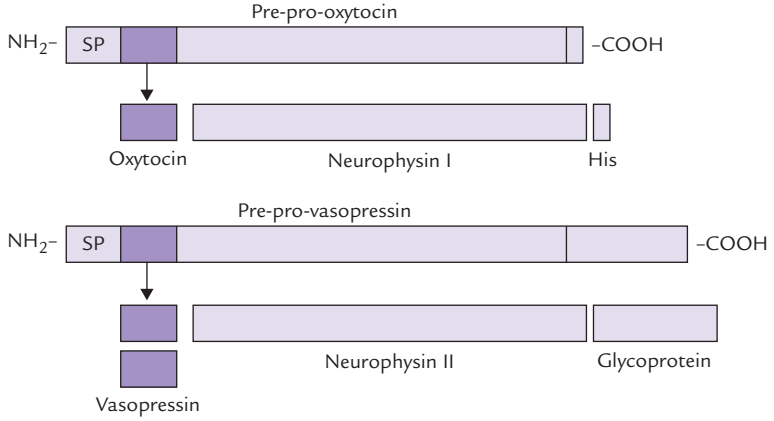


Fig. 30.6. Proteolytic maturation of pre-pro-vasopressin and pre-pro-oxytocin (SP = signal peptide).

induces oxytocin secretion via the neural pathways. The other stimuli for oxytocin release are vaginal and uterine distention.

Functions

1. *Uterine contraction:* Oxytocin stimulates contraction of uterine smooth muscles in pregnant women at term. Blood oxytocin levels have been observed to increase just before parturition. Some synthetic analogues of oxytocin such as syntocinon and syntometrine are used to induce labour at term.
2. *Milk ejection:* Oxytocin stimulates contraction of the myoepithelial cells of breast, leading to milk ejection in response to suckling.

Oxytocin exerts its effects by binding to receptors on myometrial or myoepithelial cell membranes and initiating formation of cAMP or cGMP. The hormone appears to be metabolized by the liver and kidneys or the lactating mammary gland.

Vasopressin

Vasopressin is primarily concerned with the regulation of water balance in the human body. It stimulates water retention to increase blood volume and blood pressure.

Functions

Vasopressin stimulates water absorption from distal tubules and collecting ducts of kidneys. It interacts with specific receptors on epithelial cells, resulting in stimulation of *adenylate cyclase* and cAMP production. The latter promotes water reabsorption by making the cell membrane permeable to water. This is called **facultative reabsorption** of water.

Vasopressin helps maintain fluid balance of the body by adjusting the facultative reabsorption of water according to requirements of the body. For instance, when the osmolarity of the blood is higher than normal, the osmoreceptors (of hypothalamus) sense the rise in osmolarity and induce vasopressin secretion. In addition, the patient experiences a thirst sensation, although a different set of osmoreceptors is involved in this response. The overall effect is to dilute the blood by increasing blood volume with water.

Another less significant effect of vasopressin is to cause peripheral *vasoconstriction* resulting in raised blood pressure.

***Hypersecretion of ADH*:** In addition to tonicity and volume, ADH release may also be stimulated by nausea, pain, stress, various infections, and vascular and neoplastic disorders. Such conditions may result in syndrome of inappropriate ADH (**SIADH**), in which ADH is secreted continuously, resulting in excessive water reabsorption and decreased plasma osmolarity.

***Hyposecretion of ADH*:** Deficiency of ADH, termed **diabetes insipidus** (DI), results from destruction of the posterior pituitary or hypothalamus secondary to neurosurgical procedures, tumour, trauma or degenerative processes. Limitation of ADH secretion in DI results in increased water excretion to as high as 10 L/day, the normal being about 1.5 L/day.

Rarely, diabetes insipidus results from hereditary defect in receptor in renal tubules, when it is referred to as **hereditary nephrogenic diabetes insipidus**. Acquired form of nephrogenic diabetes insipidus is known to result from long term administration of lithium in treatment of depressive illness.

V. Hormones of Thyroid Gland

Thyroid gland is the largest endocrine gland in humans, weighting about 20 g. It synthesizes two main hormones, triiodothyronine (T_3) and tetraiodothyronine or thyroxine (T_4). Calcitonin, a hormone involved in calcium homeostasis, is secreted by special cell types called *parafollicular C-cells*, or *simply C-cells*.

T_3 and T_4 are the only hormones in humans and higher animals that contain organically bound iodine (Fig. 30.7). They are necessary both in the long term regulation of the metabolic rate and in the regulation of growth, development and tissue differentiation.

A. Biosynthesis of Thyroid Hormones

Site of Synthesis

Thyroxine (T_4) is produced exclusively by the thyroid gland. T_3 is produced by deiodination of thyroxine (Fig. 30.7). This process may occur in the thyroid gland, in target tissues or in other peripheral tissues (e.g. liver and kidneys). Deiodination of thyroxine may also occur at the inner ring to form biologically inactive reverse T_3 (r T_3).

Substrates for Thyroid Hormones

There are two important substrates for thyroid hormone synthesis: (i) thyroglobulin, the intrinsic substrate, and (ii) elemental iodide, the extrinsic substrate.

1. ***Thyroglobulin* (Tgb):** It is a homodimeric glycoprotein (MW 6,669,000) that is synthesized in the rough endoplasmic reticulum of the thyroid follicular cells and secreted (by exocytosis) into the follicular lumen, where it is stored in the extracellular **colloid**. It contains 134 tyrosyl residues, up to 40 of which become iodinated, but no more than 8 to 10 of these are processed to active hormones.

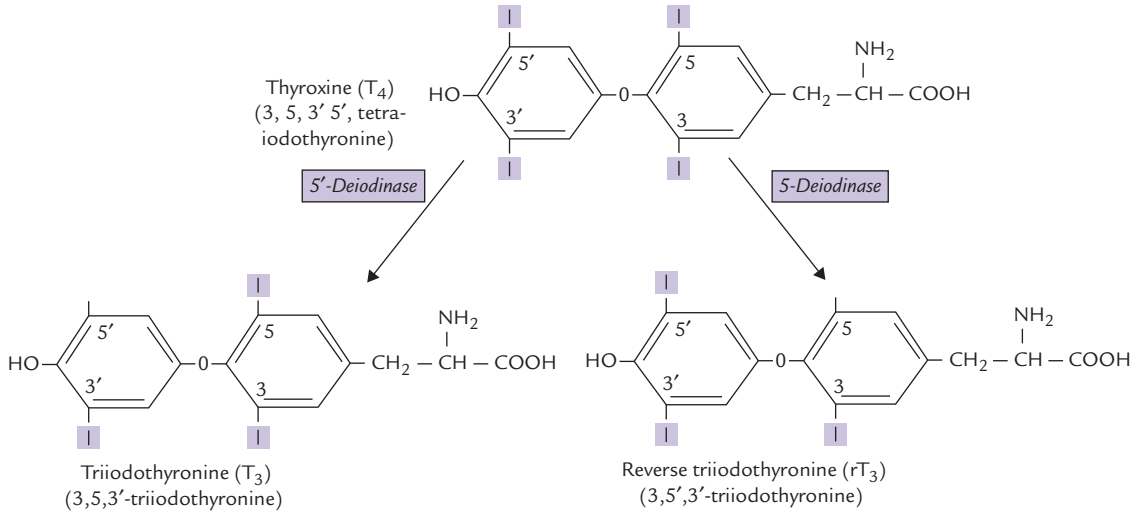


Fig. 30.7. Structures of thyroid hormones. T_4, T_3, and rT_3. Approximately 80% of T_4 is converted into equal amounts of T_3 and rT_3. Thyroxin is synthesized in largest amounts but T_3 is biologically more active. rT_3 is an inactive form.

2. ***Elemental iodine***: The extrinsic substrate iodine, in the form of iodide, is derived from dietary sources, such as drinking water, fish, cereals and iodinated salt. Daily requirement of iodine is 150–200 µg and plasma concentration of iodide is less than 0.2 µg/dL. But thyroid gland readily removes the ion from circulation and concentrates it intracellularly.

Steps in Synthesis

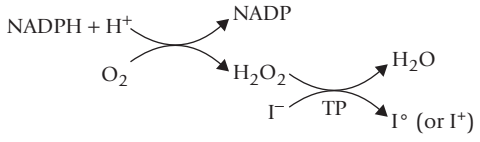
Step 1: Iodide Uptake

The follicular cells in the thyroid gland take up and concentrate iodide against a steep concentration gradient (about 20 : 1). It is an energy requiring step and is ***rate-limiting*** for the pathway of thyroid hormone synthesis. The uptake is mediated by a *sodium-iodide symporter* (an intrinsic membrane protein with approximately 14 transmembrane segments), and is controlled by TSH. From the follicular cell, iodide can enter the lumen of the thyroid follicule by facilitated diffusion.

Antithyroid drugs, such as thiocyanate and perchlorate inhibit iodide uptake (Fig. 30.8).

Step 2: Iodide to Iodine Oxidation

The iodide is oxidized by the enzyme *thyroperoxidase* (TP) to a more reactive form, iodine. The reaction requires hydrogen peroxide, which is supplied by an *NADPH-dependent oxidase* system, similar to that of leukocytes (Chapter 27).

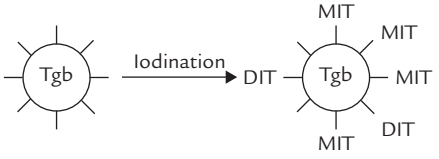


Thyroid is the only tissue capable of oxidizing iodide (I^-) to higher valence states (I^+ or I°). The enzyme catalyzing this step, *thyroperoxidase* is a haem-containing enzyme that is bound to the apical (luminal) surface of the plasma membrane. It is a large tetrameric protein (MW 60,000), and its catalytic domain faces follicular lumen.

Antithyroid drugs such as thiourea, thiouracil and methimazole inhibit oxidation of iodide.

Step 3: Iodination of Tyrosyl Residues in Tgb

Thyroglobulin acts as a precursor for T_3 and T_4. It contains several tyrosyl groups to which the reactive iodine attaches; the process is referred to as *organification of iodine*. It requires hydrogen peroxide and catalytic action of *thyroperoxidase*. The tyrosyl residues are iodinated first at position 3 to form mono-iodo-tyrosine (MIT) and then at position 5 to form di-iodotyrosine (DIT).



Step 4: Coupling Reactions

Coupling of two iodotyrosyl residues results in the formation of a thyroid hormone (iodothyronine). When two DIT residues are thus coupled, formation of a tetra-iodothyronine (thyroxine or T_4) residue results. One MIT residue may be coupled with a DIT to form a tri-iodothyronine (T_3) residue.

$$MIT + DIT = T_3;\ DIT + DIT = T_4$$

Normally, about **99%** of the hormone produced by the thyroid gland is **T_4**.

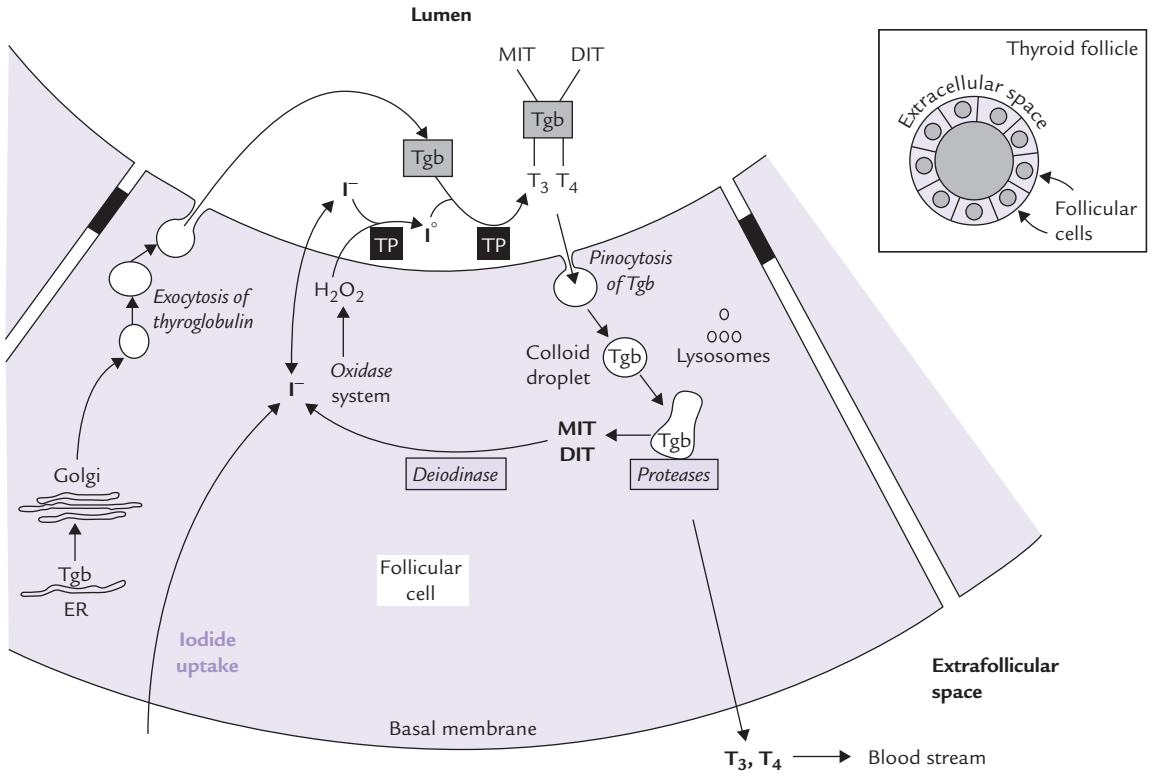


Fig. 30.8. Cellular mechanism for T_3 and T_4 synthesis and release into blood circulation. The follicular cells produce thyroid hormones, which are then stored in the lumen of the spherical follicle in a material called thyroid colloid. Tgb from these stores is degraded, forming thyroid hormones, which are finally released in blood circulation (Tgb = thyroglobulin, TP = *thyroperoxidase*).

The coupling reactions are catalyzed by *thyroperoxidase*.

Thyroid hormones, derived from protein-bound tyrosine (also require iodine) may be considered as pedal of metabolism. T_4 is produced exclusively in the thyroid gland and is more abundant than T_3, which is the biologically active form.

B. Storage and Release of Thyroid Hormones

The thyroid hormones synthesized in the follicular space are still attached to the Tgb molecule. Thus thyroid follicles store an appreciable amount of thyroid hormones, covalently bound to the amino acid sequence of Tgb. The thyroglobulin therefore is considered a storage form of T_3 and T_4 in the colloid, containing several weeks' supply of these hormones.

When need for thyroid hormone arises, thyroglobulin enters follicular cells by pinocytosis, and undergoes proteolytic degradation by *lysosomal proteases*. This releases T_3 and T_4, which are secreted into blood circulation. The MIT and DIT residues that are also present in the Tgb molecule are deiodinated in the cell, and the released iodide ions are recycled (reutilized for hormone synthesis).

It is important to note that ***virtually all steps of thyroid hormone biosynthesis are promoted by TSH***, including iodide uptake, thyroglobulin synthesis, iodination, coupling, and pinocytosis of thyroglobulin.

C. Transport of T_3 and T_4

Thyroid hormones are transported bound to two specific binding proteins. More than 99% of T_3 and 99.9% of T_4 thyroxine-binding globulin (**TBG**; 54 kD) and thyroxine-binding prealbumin (**TBPA**) or transthyretin. The hormones first bind to TBG, and when it is saturated, to TBPA. Very small amounts are bound with **albumin** or remain in free form; the latter accounts for all biological actions of thyroid hormones.

T_3 binds with proteins with less avidity than T_4. About 99% of T_3 and 99.9% of T_4 are bound to a protein. Therefore, relative concentrations of free, unbound hormones are more balanced, even though the total plasma level of T_4 is approximately 50 times higher than that of T_3 (80 ng/mL versus 1.5 ng/mL).

The protein binding prevents rapid renal clearance of the hormones and also protects them from enzymatic attack. As a

result, their biological half-lives are remarkably long: 6.5 days for T_4 and 1.5 days for T_3.

Only the free, unbound form of a hormone is biologically active. T_3 is more potent than T_4 (about 4 times).

D. Metabolic Fate of T_3 and T_4

Approximately 90% of the hormone released from the thyroid gland is T_4, but part of it is converted to T_3 by deiodination in target tissues by 5′-*deiodinase*. About two-third of the circulating T_3 is derived not directly from the thyroid gland but rather by deiodination of T_4 in peripheral tissues, especially the liver and the kidney. T_3 is approximately four times more potent than T_4, which supports the current theory that *T_4 is a pro-hormone for T_3*.

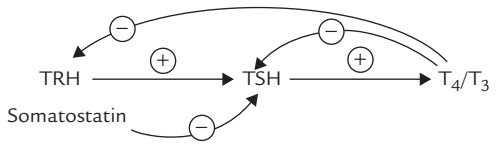
Alternatively, T_4 may undergo deiodination in the inner ring to form biologically inactive reverse T_3 (Fig. 30.7). Approximately 33% of the thyroxine that is secreted each day produces T_3, whereas another 40% produces reverse T_3.

The deiodination reactions release iodine that enters plasma and is recycled for new hormone synthesis.

Approximately 75% of T_4 is metabolized by deiodination, producing active T_3 and inactive rT_3. The remaining T_4 is deactivated by deamidation or decarboxylation and by conjugation with sulphate or glucuronide. It is subsequently excreted via bile and, to a lesser extent, in urine.

E. Regulation of T_3 and T_4 Synthesis

Synthesis and secretion of thyroid hormones are regulated by *hypothalamoanterior pituitary—endocrine axis* (Fig. 30.2). The regulation begins with the hypothalamus. TRH, a modified tripeptide is synthesized in the hypothalamus and transported via the portal circulation to the pituitary, where it promotes exocytosis of TSH. The secretion of TSH appears to be regulated by an *interplay of negative feedback from circulating free T_3 and T_4, TRH, and an inhibitory neurotransmitter somatostatin*. The negative feedback by thyroid hormones, in fact, occurs at both hypothalamic and pituitary levels. At the pituitary level, free T_4 and T_3 inhibit TSH secretion by decreasing both the biosynthesis and release of TSH through regulation of gene transcription and TSH glycosylation.



F. Biochemical Functions of Thyroid Hormones

The thyroid hormones increase the metabolic activities in nearly all cells of the body. They diffuse through the plasma membrane and interact with the nuclear receptors. The hormone-receptor complex binds to regulator sequences in DNA (hormone responsive elements; HRE) to increase expression of specific genes. The protein thus synthesized mediates effects of the hormone.

The major effects are as below:

1. ***Calorigenic effect or thermogenesis:*** Thyroid hormones cause uncoupling of oxidative phosphorylation, resulting in heat generation (i.e. calorigenic effect). There is swelling of mitochondria and basal metabolic rate is increased.
2. ***Metabolic effects:*** Thyroid hormones increase cellular metabolism. Initially they stimulate transcription and protein synthesis in various tissues, resulting in positive nitrogen balance. Glucose absorption from intestine and its utilization in various tissues is enhanced. Gluconeogenesis in liver and kidney, and lipolysis in adipose tissue are also increased, resulting in increased concentration of glucose and free fatty acids in plasma.

   There is induction of the synthesis of ***7α-hydroxylase***, the regulated step of synthesis of bile acids from cholesterol. This effect may be related to the observed increase of the plasma cholesterol level in patients with hypothyroidism.
3. ***Physiological effects:*** Thyroid hormones are required for normal physical growth because they increase synthesis of structural proteins (may stimulate GH synthesis). Mental development also requires normal levels of iodothyronines. Other effects include increase in heart rate and cardiac output, rate and depth of respiration, gastrointestinal motility and secretion of digestive juices.

G. Clinical Disorders of Thyroid Functions

Insufficient formation of thyroid hormones is known as **hypothyroidism** and overproduction is known as **hyperthyroidism** or **thyrotoxicosis**. Both are very common in clinical practice affecting almost 3% of the population, and nine times as many women as men are affected. More than 95% of thyroid diseases originate in the thyroid gland, (mostly due to autoimmunity) and the rest are accounted by **hypothalamic** or **pituitary causes**.

Hypothyroidism

The commonest cause is failure of the thyroid gland known as **primary hypothyroidism**. In adults, the cause

of primary hypothyroidism is often spontaneous autoimmune disease (**Hashimoto's thyroiditis**) or destructive therapy for hyperthyroid states, but some cases are of unknown aetiology. *Hypothalamic* and *pituitary causes* of hypothyroidism occur regularly, often as a result of impaired TSH secretion secondary to pressure from an adjacent tumour.

If hypothyroidism occurs in adults, it leads to **Myxoedema**, with widespread subcutaneous oedema, decreased basal metabolic rate, sluggishness of mental activity and a broad range of non-specific symptoms, outlined in **Case 30.1**. The biochemical parameters used for specific diagnosis include measurement of serum T_3, T_4 and TSH. In addition, measurement of serum autoantibodies for thyroglobulin and *thyroperoxidase* are carried out.

In children, hypothyroidism has serious consequences with severe and irreversible mental deficiency, stunted growth and multiple physical deformities. This condition is called **cretinism**. Hypothyroidism is treated by oral administration of thyroxine.

Hyperthyroidism

The most common cause of hyperthyroidism is **Graves' disease**, which is caused by an IgG **auto-antibody** known as LATS (long acting thyroid stimulator). It binds with the TSH receptor in the thyroid gland and stimulates them. Toxic multinodular goitre and toxic adenoma are other important causes of hyperthyroidism. Pituitary TSH secreting adenomas are known, but are extremely rare causes of hyperthyroidism.

The clinical features include weight loss, fatigue, weakness, sweating, palpitations, nervousness, etc. The typical biochemical laboratory parameters utilized for diagnosis and monitoring are T_3, T_4 and TSH estimations **(Case 30.2)**.

Simple or Diffuse Goitre

This results from deficiency of iodine, and the affected patient may suffer from hypothyroidism or may remain euthyroid. Synthesis of T_3 and T_4 decreases and the TSH release from the pituitary is disinhibited, leading to overstimulation of the thyroid gland by TSH. Hypertrophy and hyperplasia of the follicular cells and accumulation of excess thyroglobulin in the follicular lumen results; this phenomenon is known as **simple- or euthyroid-goitre**. It is seen in many mountainous areas of the world, where the soil and the plants grown on it are often deficient in iodide. Use of *iodized salt* is, therefore, recommended in these regions.

Thyroid hormones, TSH and TRH, interact by way of a hypothalamic-pituitary-thyroid axis, therefore a thyroid disorder might arise due to diseases of the hypothalamus, pituitary or thyroid gland.

VI. Hormones of Adrenal Cortex

Adrenal glands lie in relation to the upper poles of the kidneys. Each gland consists of outer cortex and inner medulla, which are *embryologically and histologically distinct* from each other. The cortex comprises three layers—Zona glomerulosa, Zona fasciculata and Zona reticularis which secrete a number of steroid hormones (>50). The adrenal medulla is part of the sympathetic nervous system. The medullary cells synthesize and store adrenaline, along with noradrenaline and dopamine.

Chemistry

The three zones of adrenal cortex secrete different classes of steroid hormones. **Zona glomerulosa**, the outermost and the smallest, secretes *mineralocorticoids*, the C-21 steroids that principally affect water and electrolyte balance. **Zona fasciculata**, the middle zone produces *glucocorticoids* (C-21) which mainly affect carbohydrate metabolism, and also lipid and protein metabolism to a smaller extent. Androgens and oestrogens are also produced by this zone, but are quantitatively insignificant. **Zona reticularis**, the innermost zone, produces *androgens* (C-19) and *oestrogens* (C-18), and to a small extent, glucocorticoids also. Structures of some adrenal steroids are illustrated in Figure 30.9.

The adrenal steroids participate in regulation of energy metabolism (glucocorticoids) and mineral metabolism (mineralocorticoids); the gonad-steroids are concerned primarily with sexual development and reproduction.

Synthesis of Steroid Hormones

All steroid hormones originate from cholesterol (Fig. 30.10). Cholesterol is obtained either from endogenous synthesis or from LDL. On the basis of weight, adrenal cortex (and sex glands), has the highest rates of cholesterol synthesis in the body. Cholesterol esters are also stored abundantly in these cells. Hormonal stimulations by such trophic hormones as ACTH activate a *cholesterol esterase* that provides a quick supply of cholesterol (from cholesterol esters).

Mineralocorticoids are recognized by the aldehyde group at C-18, and oestrogens by aromatic nature of ring A. Hydroxy or keto group is present in positions 3 (all), 11 (corticosteroids), and 17 (glucocorticoids).

Steps involved in the biosynthesis of hormones or adrenal cortex are as follows:

Step 1: Cholesterol is first acted upon by *desmolase*, a mitochondrial side chain cleavage enzyme. It cleaves

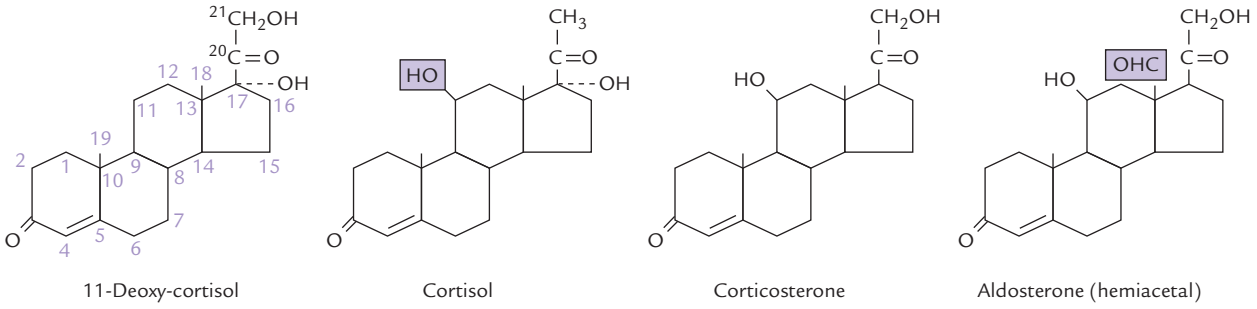


Fig. 30.9. Structures of common corticosteroids. Cortisol and ll-deoxycortisol are glucocorticoids; corticosterone and aldosterone are mineralocorticoids.

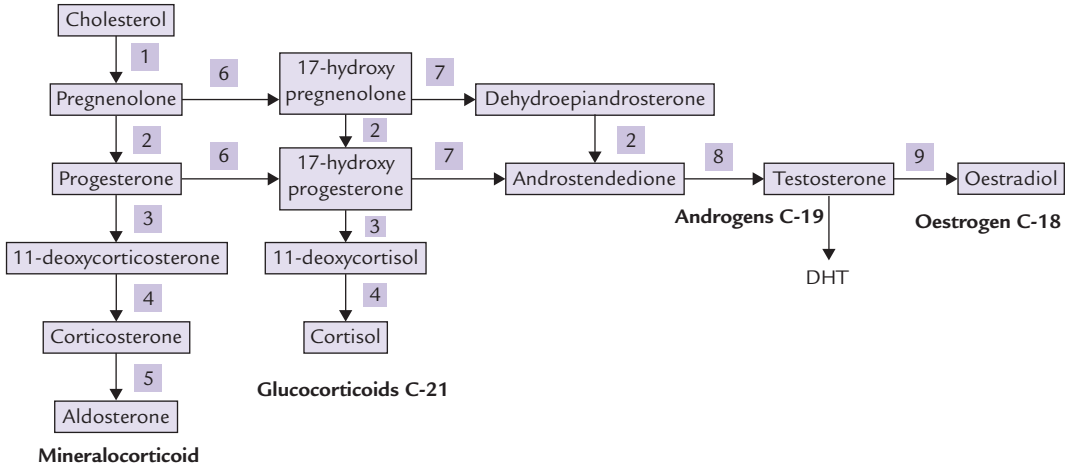


Fig. 30.10. Summary of steroid hormone biosynthesis. **(1)** *Desmolase*, **(2)** *3β-Hydroxysteroid dehydrogenase: Oxosteroid isomerase.* **(3)** *21-Hydroxylase*, **(4)** *11-Hydroxylase*, **(5)** *18-Hydroxylase* and *18-Hydroxydehydrogenase* **(6)** *17-Hydroxylase*, **(7)** *17–20 Lyase*, **(8)** *17-Keto-reductase*, **(9)** *Aromatase.*

off a 6-carbon unit forming a 21-product, pregnenolone, which is the biosynthetic precursor of all other steroid hormones.

Step 2: Pregnenolone is converted to *progesterone* by microsomal and cytoplasmic enzymes, e.g. a *dehydrogenase* and an *isomerase*.

Step 3: Progesterone is further converted into *glucocorticoids and mineralocorticoids* by hydroxylations. These reactions are effected by *monooxygenases* which require cytochrome P-450 as an intermediate electron carrier.

$$\text{Steroid} + O_2 + \text{NADPH} + H^+ \longrightarrow \text{Steroid-OH} + \text{NADP}^+ + H_2O$$

The *17-hydroxylase* and *21-hydroxylase* are microsomal and *11-hydroxylase* is mitochondrial. A summary of all hydroxylation reactions is given in Figure 30.10.

Pregnenolone is generated from cholesterol by the mitochondrial side chain cleavage enzyme, *desmolase*. It serves as an intermediate in all steroid hormone biosynthetic processes.

Synthesis of all steroid hormones is *stimulated by* ACTH, which stimulates *desmolase* so that availability of pregnenolone is increased.

Transport, Metabolism and Excretion

Cortisol is transported by a specific α_1-globulin called **cortisol-binding globulin (CBG)** or **transcortin**. It binds approximately 70% of the circulating cortisol, another 20% is bound to albumin and the rest is transported in free form. The free form only is the biologically active fraction. Aldosterone is bound mainly to albumin.

Metabolism of steroid hormones occurs mainly in liver. The double bond in the ring is reduced, the keto group is also reduced, and the reduced compound is conjugated with glucuronic acid and to a smaller extent with sulphate. The conjugated compounds, termed 17-ketosteroids and 17-hydroxy steroids, are excreted in urine. A smaller amount is excreted in faeces.

Biochemical Functions of Adrenal Steroids

The C-21 corticosteroids, glucocorticoids and mineralocorticoids are potent metabolic regulators and

immunosuppressants. Many of their biological effects can be understood best as adaptations to sustained stressful situations.

1. ***Effect on carbohydrate metabolism:*** As the name glucocorticoids suggests, these hormones have a major role on increasing glucose production by enhancing virtually every step in the gluconeogenesis. In the liver, synthesis of the enzymes of gluconeogenesis (*pyruvate carboxylase*, *PEP carboxykinase*, *fructose bisphosphatase* and *glucose 6-phosphatase*) is increased. In addition, protein breakdown in skeletal muscle and other extrahepatic tissues is increased, which supplies amino acid precursors for gluconeogenesis. Part of glucose 6-phosphate formed by gluconeogenesis is diverted into glycogen synthesis, and liver glycogen stores are increased.

Tissue uptake and metabolism of glucose is decreased by glucocorticoids. The net effect of increased production and decreased utilization is elevated blood glucose concentration.

Cortisol is the major glucocorticoid synthesized in the inner two zones of the adrenal cortex. It has a major influence on gluconeogenesis. It is under the direct control of pituitary ACTH.

2. ***Effect on lipid metabolism*:** Glucocorticoids enhance mobilization of adipose tissue triacylglycerols by inducing synthesis of the hormone-sensitive *lipase*. They also decrease lipogenesis in adipocytes; the net result is hyperlipidaemia.
3. ***Effect on protein metabolism*:** Protein and RNA synthesis are stimulated in the liver but inhibited in peripheral tissues (muscle), with protein breakdown products acting as gluconeogenic substrates. The glucocorticoid-induced protein and amino acid catabolism explains negative nitrogen balance in a seriously ill patient (chronic stress), and muscle wasting in any protracted disease.
4. ***Other effects:*** Several effects of glucocorticoids are elicited only when these hormones are present in high concentrations. They have a wide range of suppressant effect on humoral and cell-mediated immunity, and thus provide useful form of therapy as *anti-inflammatory* and *anti-allergic agents*. It is likely that some of the *immunosuppressant* effects are important in the control of immune response in normal physiology.

Glucocorticoids also influence heart vasculature, blood pressure, water excretion and electrolyte balance (*sodium retention and potassium elimination*). They also influence bone turnover through a variety of mechanisms, and the presence of high concentrations of glucocorticoids for prolonged periods results in osteoporosis. Glucocorticoids can inhibit linear growth and cell division in several tissues. Finally, they decrease synthesis of eicosanoids by inhibiting *phospholipase* A_2, and increase secretion of gastric juice.

Functions of the other C-21 corticosteroids (the mineralocorticoids) are described in a subsequent section.

Mineralocorticoids: The more important effects of mineralocorticoids is to promote tubular reabsorption of sodium ions and to increase secretion of potassium from renal tubular cells into the tubular fluid. The two effects are coupled, i.e. *sodium is antiported in exchange for potassium.* Mineralocorticoids also promote tubular reabsorption of chloride ions and tubular secretion of hydrogen ions.

Because sodium retention is accompanied by reabsorption of water, it results in expansion of the extracellular fluid volume. Mineralocorticoids also stimulate sodium conservation by the sweat glands and the mucosal cells of the colon, but in normal circumstances these effects are trivial.

In contrast to glucocorticoids (required for energy metabolism), mineralocorticoids are concerned mainly with mineral balance.

Regulation of Synthesis and Secretion of Adrenal Steroids

Adrenocorticotropic hormone (ACTH) alters the rate of biosynthesis and secretion of the glucocorticoid cortisol from the Zona fasciculata and reticularis. ACTH also increases secretion of the adrenal androgens, and aldosterone to a lesser extent.

Four major mechanisms appear to control ACTH release and hence cortisol secretion, namely hypothalamic factors, negative feedback, diurnal rhythm and stress. These are illustrated in Figure 30.11.

- ***Hypothalamic factors*,** namely CRH and AVP have been discussed earlier.
- ***Negative feedback*** operates via inhibition of CRH and ACTH synthesis and secretion by high concentrations of plasma glucocorticoids (long loop). Conversely, reduction in level of plasma cortisol leads to an increased secretion of ACTH. The feedback effect occurs within two time frames: the fast and the slow-feedback.
- Cortisol secretion is controlled by an inherent ***diurnal rhythmicity*** of CRH secretion and consequently of ACTH release. The plasma cortisol level is about 10 times higher at 08.00 h than at 24.00 h.

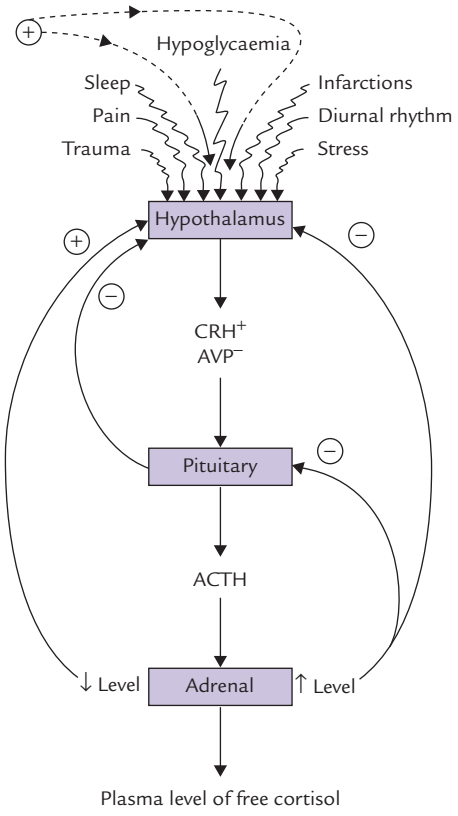
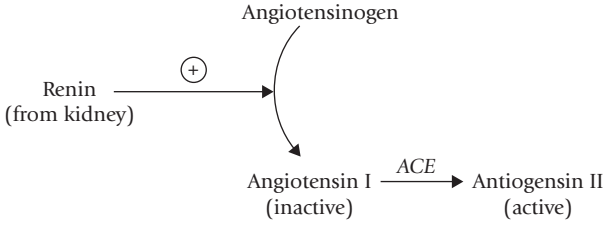


Fig. 30.11. Feedback regulation of cortisol biosynthesis. ACTH = adrenocorticotropic hormone, AVP = arginine vasopressin, CRH = corticotropin-releasing hormone.

- **Stress**, such as trauma, pain, apprehension, fever and hypoglycaemia act directly on hypothalamus. Thus, they can override the negative feedback mechanism and the diurnal rhythm to enhance cortisol secretion.

Renin-angiotensin system: Aldosterone, the most important mineralocorticoid, is primarily controlled by renin-angiotensin system (though other factors, including ACTH also regulate aldosterone synthesis). The major stimuli for this system are hypovolaemia, hyponatraemia and renal ischaemia. Specialized cells in the juxtaglomerular apparatus of the nephron sense these changes and respond by secretion of renin, a *protease*. This enzyme cleaves off a decapeptide from the amino terminal of a plasma protein of liver, termed angiotensinogen. This decapeptide termed angiotensin I is biologically inactive. It is processed to biologically active angiotensin II by angiotensin converting enzyme (ACE), an endothelial enzyme in lung capillaries.



Angiotensin II stimulates aldosterone secretion from the Zona glomerulosa cells. It also has powerful hypertensive effects: it contracts vascular smooth muscle directly and indirectly by enhancing the release of norepinephrine from sympathetic nerve endings. ACE inhibitors (enalapril, lisinopril) are used in treatment of *hypertension*.

Finally, angiotensin II is degraded into inactive fragments by *angiotensinases* released from vascular epithelium.

Dysfunctions of Adrenal Cortex

Decreased hormonal secretions by adrenal gland may result in hormonal deficiency (hypoadrenocorticism) or excess (hyperadrenocorticism).

Hypoadrenocorticism: Hypoadrenocorticism is a rare but life-threatening condition. It may arise in hypothalamic-, pituitary-, or adrenal-failure.

Addison's disease or primary adrenal insufficiency is caused by the destruction of adrenal cortex. The most common causes of Addison's disease are autoimmune destruction of the gland and tuberculosis in that order. Production of both glucocorticoids and mineralocorticoids is decreased.

- Deficient glucocorticoids production leads to hypoglycaemia between meals. Mobilization of fats is also decreased, resulting in decreased availability of fuels and energy. This leads to weakness and inability to withstand stress.
- Decreased mineralocorticoid production results in renal loss of sodium (so hyponatraemia) and retention of potassium (so hyperkalaemia). Acidosis develops due to decreased excretion of hydrogen ions. Pathological sodium loss shrinks the extracellular fluid volume, causing hypotension.

Hyperpigmentation is a characteristic feature of Addison's disease and distinguishes it from adrenal insufficiency secondary to pituitary or hypothalamic disease. It occurs because of loss of negative feedback on pituitary and consequent increased production of ACTH. The structure of this hormone contains part of the amino acid sequence of melanocyte-stimulating hormone. The latter is responsible for skin pigmentation (POMC), and so hyper-pigmentation occurs; common sites are face, neck and dorsum of hands.

Acute adrenal insufficiency is a life-threatening medical emergency and needs prompt treatment with cortisol replacement and fluids.

Hyperadrenocorticism: Hyperfunction of adrenal cortex can result in three distinct syndromes: Cushing's syndrome, primary aldosteronism and adrenogenital syndrome, which respectively result from excess of cortisol, aldosterone and adrenal androgens.

Cushing's syndrome: It results from excessive endogenous secretion of cortisol, the sources of which may be

classified into ACTH-dependent (pituitary overactivity) and ACTH-independent (primary adrenal disease). Benign adenoma and carcinoma comprise majority of the primary adrenal causes. The clinical features associated with Cushing's syndrome are summarized in Table 30.1 and laboratory diagnosis is discussed in Chapter 34 (see **Case 34.1**).

Table 30.1. Clinical features of Cushing's syndrome

Feature	Finding in Cushing's syndrome
Fat metabolism	Central obesity, buffalo hump, supraclavicular fat pad. Atherosclerosis, hyperlipidaemia
Protein catabolism	Proximal myopathy and muscle wasting. Thin skin, easy bruising, wide purple striae. Poor wound healing
Carbohydrate metabolism	Hyperglycaemia
Electrolyte balance	Hypertension, polyuria, lower limb-oedema
Androgen excess	Hirsutism, acne, amenorrhoea in females
Effect on bone	Osteoporosis. Growth failure in children
Neurological	Psychiatric and personality disorders

Hypersecretion of cortisol results in Cushing's syndrome and hyposecretion in Addison's disease. Both hyper- and hyposecretion may occur as a result of hypothalamic, pituitary, or adrenal failure.

The other syndromes of hyperfunction of adrenal cortex are described in Box 30.2.

VII. Hormones of Adrenal Medulla

The hormones of adrenal medulla are synthesized from tyrosine. Chemically, they are categorized as **catecholamines** because their aromatic residue is catechol (Fig. 30.12). Several catecholamines are synthesized in the chromaffin cells of the adrenal medulla, of which ***epinephrine*** is most predominant. The others, ***norepinephrine*** (the term "*nor*" denotes that the molecule does not contain "R" or methyl group) and ***dopamine*** are intermediates in epinephrine biosynthesis. Dopamine acts largely as a brain neurotransmitter, and noradrenaline and adrenaline are

BOX 30.2

Hyperfunction of Adrenal Cortex

In addition to cushing's syndrome and Addison's disease, a deranged function of adrenal cortex causes the following:

Primary aldosteronism (Conn syndrome) is a rare condition. In most cases it results from a benign adrenal adenoma in Zona glomerulosa. Other causes are hyperplasia or carcinomas of the adrenal glands and ectopic aldosterone-producing tumours.

The condition is characterized by expansion of ECF and increased blood volume. The patient may present with hypertension, polydipsia and polyuria. Increased potassium excretion results in hypokalaemia, resulting in symptoms of neuromuscular abnormalities such as weakness, paraesthesiae and tetany.

Secondary hyperaldosteronism is common and is associated with renal, cardiac, or hepatic diseases. Increased secretion of renin from kidneys in these conditions stimulates angiotensin production, which induces aldosterone secretion from adrenal cortex.

Adrenogenital syndrome (*congenital adrenal hyperplasia*; CAH): CAH is an inherited defect in corticosteroid biosynthesis. It is caused by deficiency of the enzymes responsible for the synthesis of corticosteroids (the *21-* and *11-hydroxylase*) but not of androgens. The *21-hydoxylase* is the deficient enzyme in 95% of cases of CAH, and deficiency of *11-β-hydroxylase* is a less important cause.

The block in biosynthesis and the resulting lack of glucocorticoids leads to a disinhibition of ACTH-release from the anterior pituitary gland. The excess of ACTH secretion continues to drive steroid biosynthesis, but with the pathway of corticosteroids synthesis blocked, the initially formed precursors (e.g. progesterone) are diverted into androgen synthesis (Fig. 30.10).

The enzyme deficiency in CAH may be partial or complete. A partial deficiency results in virilization in baby girls and precocious puberty in boys. But with complete deficiency, secretion of cortisol and aldosterone is severely affected, resulting in dangerous hyponatraemia and hyperkalaemia. Treatment involves administration of corticosteroids. If CAH is not diagnosed promptly, the afflicted infant may die.

Pseudo-Cushing's syndrome: Excessive alcohol intake results in hypertension, truncal obesity and acne. Preliminary investigations may show biochemical features of hypercorticism. The condition, however, resolves after few weeks of abstinence.

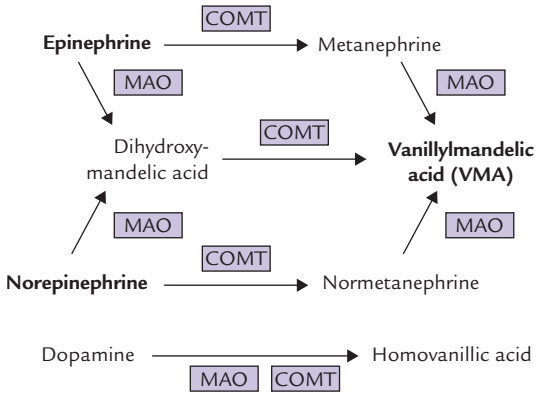


Fig. 30.12. Alternative pathways for enzymatic inactivation of catecholamines (MAO = *monoamine oxidase*, COMT = *catechol-O-methyltransferase*).

released both from chromaffin cells and from peripheral and central neurons.

Synthesis, Storage and Secretion

The synthesis of catecholamines from the common precursor tyrosine **tyrosine** has been discussed earlier in Chapter 13. Synthesis of adrenaline requires four enzymatic steps, while two and three steps suffice for dopamine and noradrenaline respectively (Fig. 13.23). Apart from adrenals, the catecholamines are synthesized in some extra-adrenal tissues, e.g. brain and sympathetic ganglia. Blood brain barrier is impermeable to norepinephrine and epinephrine, but permeable to dopamine. The dopamine synthesized in adrenals crosses the blood-brain barrier to enter the brain cells where it acts as a neurotransmitter.

Catecholamines are stored in adrenal medulla in the **chromaffin granules**. Their release is triggered by (a) splanchnic nerve impulses and (b) by other stress hormones, namely corticosteroids.

Metabolism

The enzymatic inactivation of catecholamines occurs both in liver and in target cells. Two enzymes are involved in the process: *monoamine oxidase* (MAO) and *catechol O-methyltransferase* (COMT).

- *MAO*, a copper-containing flavoprotein of the outer mitochondrial membrane, inactivates its substrates by oxidative deamination.
- *COMT* methylates one of the phenolic hydroxyl groups in its substrates, thereby greatly reducing their biological activities. Its action depends on the presence of S-adenosylmethionine (SAM), a carrier of methyl residues.

Both *MAO* and *COMT* have broad specificity and the order in which they act is immaterial (Fig. 30.12) because the end product is same: ***vanillylmandelic acid*** (from epinephrine and norepinephrine). Likewise, homovanillic acid is the end product of dopamine catabolism.

Biochemical Functions

Epinephrine is a stress hormone: stress means an environmental insult. Lowering of blood glucose, oxygen deprivation, confinement in a small or crowded quarter or participation in a marathon is to name a few causes of stress. Stress causes release of epinephrine, and to a lesser extent of ACTH and glucocorticoids. It has major effect on

- metabolic pathways (stimulation of glycogenolysis, glyconeogenesis, lipolysis). These effects are mediated via cAMP.
- smooth muscle function (causes relaxation), and on blood pressure (increases), and on blood coagulation (decreases the time of clotting by aggregating platelets).

In response to epinephrine, a sudden extra supply of glucose is delivered to muscle; the heart and lungs work harder to pump oxygen round the circulation; and the body is thus prepared to exert or to defend itself. These are typical features of the "***flight or fight***" response.

Abnormalities of Catecholamine Production

Pheochromocytomas are the tumours of the adrenal medulla or sympathetic ganglia that produce and release large quantities of catecholamines. Majority of these tumours are located in the adrenal medulla and produce both epinephrine and norepinephrine. Tumours in extra-adrenal locations produce only norepinephrine.

Excessive production of catecholamines causes increased cardiac output and peripheral vasoconstriction, leading to severe hypertension. Diagnosis is made by demonstrating enhanced excretion of vanillylmandelic acid (VMA) and the metanephrines in 24-hour urine.

In **neuroblastoma**, a pediatric tumour, free dopamine and homovanillic acid are highly elevated in urine. The latter is also increased in the urine of patients with *ganglioblastoma*.

The Receptors for Epinephrine and Norepinephrine

They are called **adrenoreceptors**. They are divided into α- or β-receptor classes and subclasses on the basis of their pharmacology. Epinephrine acts on all classes of the receptors, but norepinephrine is more specific for β-receptors. The β-blockers, such as atenolol, are used to treat hypertension and chest pain (angina) in ischaemic heart disease because they antagonize the stimulatory effects of catecholamines on the heart. Non-specific α-blockers have limited use although the more specific α_1-blockers such as prazosin and α_2-blockers such as clonidine can be used to treat hypertension. Certain subclasses of β-receptors are found in particular tissues; for example, the β_2-receptor is present in lung, and

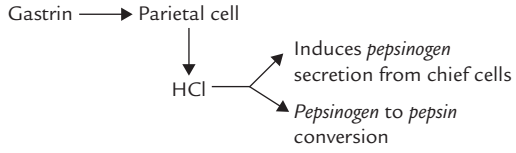
β_2-receptor agonists, such as salbutamol, are therefore used to produce bronchial dilatation in asthma without stimulating the β_1-receptor in the heart.

VIII. Gastrointestinal Hormones

A number of peptide hormones are released by specialized cells lining the stomach and small intestine (Fig. 26.1), which play important role in food digestion and absorption. The cells secreting these hormones are not organized into any discrete anatomical structure, but are dispersed throughout the GIT. A large number of hormones have been identified, but the most important ones are gastrin, secretin, cholecystokinin and gastric inhibitory peptide (GIP).

Gastrin

It is a peptide of 17 amino acids (MW 17,000) secreted by the **G cells of stomach antrum** in response to the ingestion of food. Acetylcholine and vagus nerve also stimulate its secretion. Gastrin migrates via the bloodstream to the fundic region of the stomach, which contains parietal cells and chief cells. The former are induced (by gastrin) to secrete HCl, which induces the chief cells to secrete pepsinogen (proenzyme of pepsin). The same HCl then catalyzes pepsinogen to pepsin conversion by removing of a 42 amino acid peptide from it. Pepsin initiates protein digestion in stomach. Its optimum pH generated by HCl is less than 2.



The effects of gastrin are inhibited by various agents, such as the gastric inhibitory peptide (GIP) and vasoactive intestinal peptide (VIP), produced by small intestine.

Secretin

A small peptide of 27 amino acids (MW 3000), secretin was the ***first hormone to be discovered***. It is secreted by the S cells of the duodenum in response to the acidic gastric contents entering the small intestine. It migrates to the pancreas, where it *induces release of a bicarbonate-rich solution* into the small intestine via pancreatic duct. Function of bicarbonate is to elevate pH of the acidic gastric contents to 7.0–7.5, which is the optimum pH for action of intestinal enzymes.

Other important functions of secretin are to potentiate action of cholecystokinin on pancreas, and to delay gastric emptying.

Cholecystokinin

Cholecystokinin (CCK) or cholecystokinin-pancreozymin (CCK-PZ) contains 33 amino acids (MW 4000), and is secreted by the *mucosa of small intestine* in response to entry of the food in the intestinal lumen. Products of protein and lipid digestion, namely peptides, amino acids, mono- or diacylglycerol, fatty acids and glycerol stimulate its secretion.

CCK causes *contraction of gall bladder to release bile into the small intestine* via the bile duct and *causes the release of zymogens by the pancreas*. The zymogens are later activated in the small intestine to active enzymes, and participate in food digestion (Chapter 26).

Secretin apparently potentiates the action of CCK on pancreas.

Gastric Inhibitory Peptide (GIP)

GIP is made up of 43 amino acids (MW 5100), and it is secreted by *duodenal mucosa* in response to carbohydrates and fats entering the duodenum. It inhibits gastric motility and secretion, and also stimulates release of insulin from pancreas. The latter effect explains the observation that insulin level is elevated in plasma before any substantial increase in blood glucose.

Other gastrointestinal hormones are **VIP**, **motilin**, **somatostatin** and **entero glucagon**. Their important characteristics are shown in Table 30.2. They are broadly divided in two homology groups based on their common structural characteristics (Box 30.3).

Pancreatic Hormones

The hormone-secreting, endocrine portion of pancreas is the *islet of Langerhans*, which contains **α, β, δ,** and **F cells**. Insulin and glucagon are perhaps the most influential intermediary metabolism regulating hormones; insulin is released from the β-cells in response to high blood glucose levels (e.g. after a meal) and glucagons from α-cells in response to low blood glucose levels. *The δ-cells produce somatostatin and F cells secrete pancreatic polypeptide.*

All these pancreatic hormones are peptides and not under any major pituitary, hypothalamic, or neural control.

Insulin

This hormone has profound effect on several aspects of intermediary metabolism, as discussed in appropriate places of Chapter 9 through 15. Its structure, biosynthesis, regulation, and interaction with insulin-receptor were described earlier (Chapter 29) and a detailed account of its role in integration of metabolism and metabolic abnormalities in insulin deficient state (diabetes mellitus) was presented in Chapter 15.

Table 30.2. Some gastrointestinal hormones and their important characteristics

Hormone	Chemical nature	Source	Functions
Vasoactive intestinal peptide	28 amino acids	Whole length of small intestine	Inhibits gastric motility but stimulates secretion from small intestine and pancreas
Motilin	28 amino acids	Secreted locally as paracrine agent from δ-cells of pancreas	Inhibits secretion of gastrin, CCK and pancreatic polypeptide
Somatostatin	14 amino acids	Hypothalamus adrenal medulla, pancreas, GIT	Same effect on GIT as gastrin
(Entero-) Glucagon	28 amino acids	Mucosa of small intestine	Delays passage of food through intestine releases pancreatic insulin

BOX 30.3

Two Homology Groups of Digestive Tract Hormones

GIT hormones show certain structural similarities and may be divided into two homologous groups: the gastrin family and the secretin family.

- *Gastrin family* includes gastrin and CCK; both are structurally related and share most, if not all, of the activities/characteristics. Several types of gastrins are also known, all sharing the N-terminal 14 amino acid peptide, and biologically active portion of all gastrins appear to be the N-terminal Phe-Asp-Met-Trp portion.
- *Secretin family* includes secretin, GIP, vasoactive intestinal peptide (VIP) and glucagon. All these exhibit similar structural characteristics.

Glucagon

It is a peptide of 29 amino acids (MW 3500), synthesized in α-cells of pancreas, and a small amount originating from the gastrointestinal tract. Major function of glucagon is to maintain blood glucose. It tends to ***elevate blood glucose*** by moving glucose from the cells into the bloodstream by stimulating those metabolic processes that give rise to free glucose, e.g. glycogen degradation and gluconeogenesis (see Table 15.2). Glucagon is also associated with enhanced protein degradation. In addition, it stimulates the breakdown of storage fat, whose function is to provide cellular fuel and thus spare glucose. Thus, glucagon is considered as a catabolic hormone.

Regulation of glucagon secretion is affected by blood glucose concentration, being enhanced by hypoglycaemia and inhibited by hyperglycaemia.

Pancreatic Polypeptide

It is a peptide of 36 amino acids, secreted from F-cells of islets of Langerhans in response to protein-rich meal, hypoglycaemia and exercise. It decreases the bicarbonate and protein contents of pancreatic juice.

Motilin

It is a peptide structurally similar to hypothalamic somatostatin, and is secreted locally as a paracrine agent by the δ-cells of *pancrease* (Table 30.2).

IX. Hormones of Gonads

Sex hormones are secreted by gonads (testes in males, ovaries in females) in response to pituitary FSH and LH (Fig. 30.2). In addition, the gonads are also involved in production of germ cells, and this function is influenced by sex hormones.

- The *ovarian follicular cells* secrete C-18 oestrogens, C-19 androgens and C-21 progesterone (Table 30.3).
- The *Leydig cells of testes* secrete testosterone, a C-19 androgen. A small amount of oestrogens are also formed in testes: oestradiol by aromatization of testosterone and oestriol by aromatization of androstenedione, a weak androgen produced in adrenal cortex.
- *Zona fasciculata* and *Zona reticularis* of adrenal cortex produce weak androgens, which are converted to more potent sex-hormones. The conversion may occur in gonads or in some non-endocrine tissues, such as liver and adipose tissue.

Sex hormones can be divided in three groups (Table 30.3): (a) androgens, the male sex hormones (C-19); (b) Oestrogens, the female sex hormones (C-18); and (c) progesterone (C-21), produced during menstruation and pregnancy.

Table 30.3. Sex steroid hormones and their tissues of origin

Type	Number of C atoms	Principal sources	Subsidiary sources
Androgens			
Testosterone	19	Testis	Adrenal cortex, ovary
Androstenedione	19	Adrenal cortex	Ovary, testis
Dehydroepiandrosterone	19	Testis	Adrenal cortex, ovary
Oestrogens	18	Ovary, placenta	Adrenal cortex, testis
Progesterone	21	Ovary, placenta	Adrenal cortex, testis

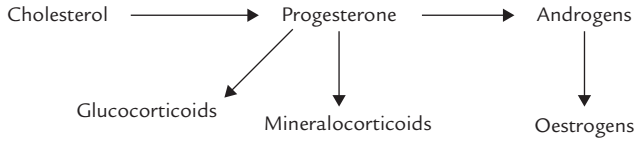
X. Androgens

These are mostly produced by the *Leydig cells* of testes, though a smaller amount originates from the adrenal glands (Table 30.3). The theca cells at periphery of the ovarian follicles also produce a small amount of androgens.

Testosterone is the major androgen synthesized in testes, its plasma level is about 0.7 μg/dL in males and < 0.1 μg/dL in females. Structures of some important androgens are given in Figure 30.13.

Synthesis

Cholesterol is the biosynthetic precursor of all steroid hormones, including androgens.

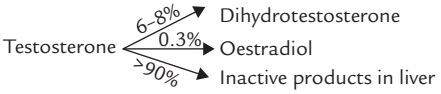


Pregnenolone, initially formed from cholesterol, is converted to progesterone by microsomal and cytoplasmic enzymes, which is converted to androstenedione by *17–20 lyase* or *desmolase* (Fig. 30.10). Androstenedione is finally converted to testosterone by a *ketoreductase*. Alternately, pregnenolone is converted to 17-hydroxypregnenolone, which forms testosterone by a similar set of reactions.

Testosterone is a pro-hormone: In the target tissues and, to a lesser extent in the testis itself, testosterone is converted to ***dihydrotestosterone*** by the enzyme *5-α-reductase*. The plasma level of DHT is only about 10% of the testosterone concentration, but DHT is considerably a more potent androgen than testosterone. Therefore, it is a widely acknowledged that testosterone acts as prohormone of the active hormone DHT.

Transport, Metabolism and Excretion

Testosterone circulates in blood in association with two proteins: **sex hormone-binding globulin** (SHBG) and **testosterone-oestrogen-binding globulin** (TEBG). Both these proteins are formed in the liver. Only small percentage of testosterone is in the free state, but it is responsible for the biologic activity of the hormone. The liver mainly metabolizes androgens to **17-ketosteroids**, which are excreted in urine.



Biochemical Functions

Secretion of androgens is minimal prior to onset of puberty. At puberty, the testes start secreting about 7000 μg of androgens each day into circulation and a further 500 μg originates from the adrenal gland.

Androgens induce development of primary and secondary sexual characteristics and also affects many specific physiological and metabolic activities:

1. ***Effect on sexual characteristics*:** Androgens cause growth and development of male internal genitalia. They stimulate differentiation of the wolffian duct system into the epididymus, vas deference and seminal vesicles. These are the *primary sexual characteristics*. During puberty, testosterone promotes appearance of *secondary sexual characteristics*, such as deepening of voice and masculine distribution of body hair, libido and

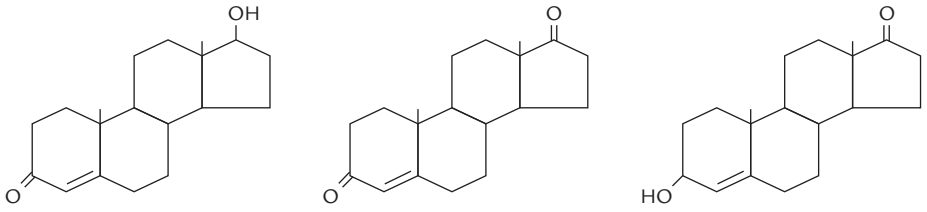


Fig. 30.13. Structures of androgens.

potency (erectile functions). It increases secretory activity of sebaceous glands, which may lead to acne.

2. ***Spermatogenesis*:** Testosterone, in association with FSH is required for normal spermatogenesis. FSH promotes secretion of ABP. This ensures high local concentration of testosterone, which is necessary for spermatogenesis in seminiferous tubules (Fig. 30.5).
3. ***Effect on nitrogen balance*:** Androgens promote RNA synthesis (transcription) and protein synthesis (translation), thereby causing positive nitrogen balance. Stimulation of protein synthesis increases muscle mass (*myotrophic effect*) in a wide range of tissues. In this capacity *testosterone is the natural anabolic steroid*.
4. ***Effect on carbohydrate and fat metabolism*:** Production of D-fructose from D-glucose in seminal vesicles is increased by androgens.

   The androgens induce synthesis of *aldolase*, thereby enhancing glycolysis.
5. ***Effect on electrolytes*:** Androgens increase renal retention of sodium, chloride and water, though to a much smaller extent than mineralocorticoids.
6. ***Effect on bones*:** Androgens promote formation of bone matrix proteins. Calcium deposition is increased because of increased availability of bone matrix. Epiphyseal fusion of long bones is enhanced.
7. ***Others*:** Basal metabolic rate is increased by androgens. Erythropoiesis is stimulated through stimulation of erythropoietin secretion.

Mechanism of Action and Regulation

Testosterone binds with the **nuclear receptors** found in muscle, bone and other tissues where it exerts its influence. The testosterone receptor complex activates specific genes, the protein products of which mediate some (if not all) of the effects of the hormone.

The affinity of DHT for these receptors is much higher than that of testosterone. Thus, testosterone is unique among steroid hormones in that further metabolism, 5 α reduction, more than doubles its affinity for androgen receptors (**Case 30.2**).

Synthesis and secretion of androgens is *regulated by hypothalamo-pituitary-gonadal axis*, described earlier.

Clinical Disorders of Androgen Secretion

Hypogonadism: Androgen deficiency may originate from a wide range of disorders of the hypothalamus, pituitary or testes. In hypothalamic disorders, deficient production of GnRH results in delayed puberty, and subnormal FSH, LH, and testosterone. Similar picture is seen in pituitary disorders, and together with hypothalamic disorder they are classified as **secondary hypogonadism**. In primary hypogonadism, there is failure of testes to produce testosterone.

Hypergonadism: It is seen in precocious puberty, a rare condition resulting either from early maturation of the normal hypothalamo-pituitary-gonadal axis or as a result of a tumour that secretes androgen (or hCG).

Testosterone regulates spermatogenesis and is a natural anabolic steroid. Its deficiency may originate from a wide range of disorders of the hypothalamus, pituitary or testes.

XI. Oestrogens

The oestrogens are a family of hormones synthesized in ovarian and extra-ovarian tissues. The three main oestrogens are **oestrone**, **oestradiol** and **oestriol** (Fig. 30.14). The first two are predominantly ovarian hormones; and all three are synthesized by placenta. Oestrogens are produced in males also: small quantity comes from testis, but most is produced in adipose tissue, liver, skin, brain and other non-endocrine tissue.

Biosynthesis

Like other steroid hormones, cholesterol serves as the precursor for synthesis of oestrogens (Fig. 30.10). Oestrogens contain an aromatic ring, which is produced by aromatization (formation of aromatic ring) of androgens-androstenedione and testosterone.

The aromatization reaction is catalyzed by the microsomal enzyme system, *aromatase* in a complex process that involves three hydroxylation steps, each of which requires O_2 and NADPH. Oestradiol is formed if the substrate is testosterone, whereas oestrone production

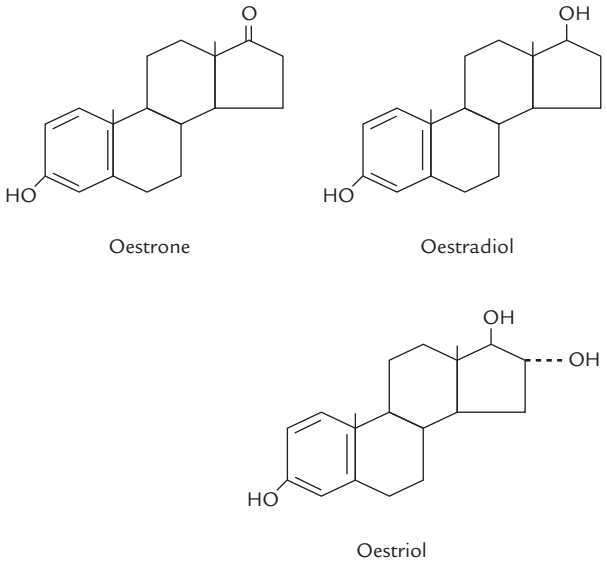
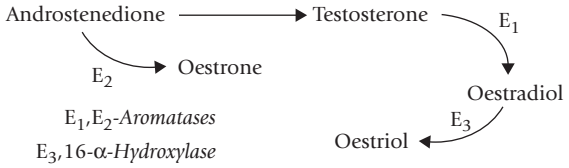


Fig. 30.14. Structures of oestrogens.

results from aromatization of androstenedione (this occurs in extraovarian tissues).



Oestradiol (E_2) is the main oestrogen secreted by the Graafian follicles of ovary, and is biologically the most active form. About 90% of the secreted oestradiol comes from ovaries. Oestrone, a weak oestrogen, is the predominant plasma oestrogen in post-menopausal women. Oestriol is the predominant oestrogen during pregnancy, secreted by the placenta. In non-pregnant women, oestriol is formed in small amounts in the liver from oestradiol and oestrone.

In males significant amounts of oestrogens are produced by peripheral aromatization of androgens. Peripheral aromatization of testosterone to oestradiol accounts for 80% production rate of the latter, and testes account for < 20%.

Transport and Metabolism

Oestrogens are transported in blood bound to SHBG. Affinity of oestradiol for SHBG is less than that of testosterone or DHT. Oestrogens can undergo conjugation with glucuronide or sulphate; the conjugation renders them water soluble so that they can be excreted in urine.

Functions

1. ***Effect on sexual organs*:** Oestrogens mediate the growth and development of maternal reproductive organs, especially the gravid uterus. They play important role in maintenance of menstrual cycle. Development of secondary sex characters is also induced by oestrogens: breasts are enlarged and hair develop in pubic and axillary regions.
2. ***Effect on nitrogen balance*:** Oestrogens promote transcription and translation, and so nitrogen balance becomes positive under their influence.
3. ***Effect on bones*:** Oestrogens enhance osteoblastic activity to promote skeletal growth and calcification, which accounts for growth spurt during puberty. The decalcification and osteoporosis developing in post-menopausal women is accounted for by lack of oestrogens.
4. ***Effect on electrolytes*:** Renal retention of sodium, chloride and water is enhanced by oestrogens. This effect is much weaker compared to that of mineralocorticoids.
5. ***Cardioprotective effects*:** Oestrogens lower total cholesterol and LDL, and also increase HDL level, thereby improving lipid profile. This explains low incidence of coronary artery disease and atherosclerosis in women of reproductive age group.
6. ***Effect on fat deposition*:** Oestrogens increase fat deposition in adipose tissue, and this may partly explain increased fat deposition (5–10%) in women compared to males of the corresponding age group.
7. Oestrogens activate the enzyme *transhydrogenase*, which catalyzes transfer of reducing equivalents from NADPH to NAD^+. In post-menopausal women this reaction being impaired, the NADPH tends to accumulate and is diverted towards lipogenesis causing weight gain and obesity.

Clinical Disorders of Oestrogen Secretion

1. ***Primary hypogonadism*:** This may be due to gonadal agenesis or polycystic ovary, both leading to deficiency in ovarian function.
2. ***Secondary hypogonadism*:** This is due to hypothalamic or pituitary disorders. For example, a genetic hypothalamic disorder in children (*Kallmann's syndrome*) results in deficient GnRH production, and the affected individuals present with delayed puberty and subnormal FSH, LH and testosterone.
3. ***Hypergonadism*:** Early maturation of the normal hypothalamo-pituitary-gonadal axis, an oestradiol- or androgen-secreting cyst or tumour of the ovary or the adrenal gland, or congenital adrenal hyperplasia are the common causes of hypergonadism. They cause **precocious puberty** in which sexual characteristics and reproductive potential develop prematurely.

Defects of synthesis of gonadal steroids (concerned with sexual development and reproduction) can cause abnormalities of sexual development.

XII. Progesterone

Progesterone is a 21-carbon compound, with a steroid nucleus (Fig. 30.15). It is synthesized from cholesterol by side chain cleavage, and acts as precursor for steroid hormones (Fig. 30.10). The progesterone released into blood circulation originates from corpus luteum in non-pregnant women and from placenta in pregnancy.

Transport and metabolism: Progesterone is transported in blood bound with a protein, transcortin. The major metabolite of progesterone is a pregnanediol, which is excreted in urine as a glucuronide conjugate.

Biochemical Functions

1. Progesterone promotes secretory changes in the endometrium in the second half of the menstrual cycle (after

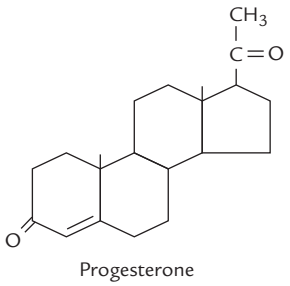


Fig. 30.15. Structure of progesterone.

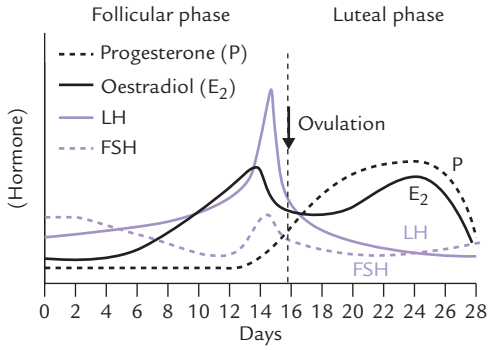


Fig. 30.16. Hormone secretion during menstrual cycle.

ovulation), and induces secretion of mucous. This is essential for implantation of the fertilized ovum.

2. Together with oestrogen, progesterone stimulates growth of glandular tissue in breasts.
3. Progesterone has a mild thermogenic effect: it elevates body temperature by 0.5–1°F. This explains mild rise in body temperature with ovulation. The cause of thermogenesis by progesterone is not clearly known.
4. In high doses, progesterone elicits mild mineralocorticoid effect, causing renal retention of sodium, chloride and water.

Regulation of Female Sex Hormones

The hypothalamic GnRH causes release of FSH and LH from anterior pituitary, which induce secretion of oestradiol (Fig. 30.2). The hormones in females are secreted in a cyclic manner during each menstrual cycle.

In females, FSH promotes oestradiol synthesis leading to follicular maturation, while LH leads to follicle rupture and oocyte release.

Menstrual Cycle

The hormonal changes occurring during a normal menstrual cycle well illustrates how various hormone functions are coordinated to achieve a common cause (Fig. 30.16). It is now clear that the pulsatile release of GnRH is ultimately responsible for the control of all hormonal changes in the normal menstrual cycle. A typical cycle has a length of 28 days (varies between 25 and 31 days). It is divided into two phases: the follicular phase (preovulatory) and the luteal phase (post-ovulatory). The length of the post-ovulatory period is remarkably constant at approximately 14 days.

1. Follicular phase: FSH is the dominant hormone of this phase, its level showing progressive rise under the influence of GnRH. FSH acts through its receptors in the ovarian granulose cells and causes the development of ovarian follicles. FSH may also enhance the production of oestradiol (from testosterone) by inducing *aromatase*. As oestradiol is secreted, FSH secretion falls (by feedback inhibition), and this combination leads to selection of a dominant follicle for further development. Initially several follicles undergo development but only a single (dominant) follicle is selected for further maturation and becomes mature or Graafian follicle.

Development of the dominant follicle continues to progress under the influence of rising oestradiol concentration, and this is turn results in more secretion of oestradiol. When oestradiol level reaches its peak (around midcycle), it causes a positive feedback at the pituitary to initiate LH surge. The LH binds to its receptor on the Graafian follicle and, in tandem with steroid hormones and other factors such as prostaglandins, causes the mature follicle to rupture and release the oocyte. This process is called **ovulation**. Approximately 14 days are required to produce this mature follicle and to cause its rupture; and these 14 days correspond to the follicular phase of the menstrual cycle.

To summarize, in the follicular phase FSH promotes oestradiol synthesis leading to follicular maturation, while LH causes the final follicular growth and ovulation.

2. Luteal phase: After the ovulation, the Graafian follicle remains in the ovary and undergoes a change and becomes known as the *corpus luteum*. The latter is the site of steroid secretion from ovary during the luteal phase. It secretes progesterone and lesser amounts of oestradiol; both these hormones sustain the oocyte. Progesterone also serves to prepare the uterus for pregnancy and the lobules of the breast for lactation. Increased level of progesterone may serve in a negative feedback mechanism to decrease the amount of LH released from the anterior pituitary.

If fertilization of the ovum does not occur after ovulation, corpus luteum function declines, resulting in a decrease in progesterone and oestradiol secretion. This brings about vascular changes in the endometrium leading

to tissue death and shedding of endometrium, causing menstruation. Fourteen days are required for the process to be completed after ovulation; this period is the luteal phase of the cycle.

The fall in steroid secretion also stimulates FSH secretion to start a new cycle.

Note: If fertilization of the ovum occurs, it is implanted in the uterus and starts producing hCG. This hormone sustains corpus luteum, so that it continues secreting hormones. The endometrial lining participates in the formation of placenta.

Menopause

The menstrual cycle begins at the onset of puberty and continues throughout the active reproductive life. Women normally reach the menopause at the age of 45–50 years, when the cycles cease to occur and this coincides with loss of ovarian function. In post-menopausal women, oestrogen and progesterone levels fall, but because of loss of their negative feedback effect, LH and FSH levels rise. Following with cessation of regular menstrual cycles:

1. Vasomotor symptoms: These are the symptoms related to contraction and dilation of small blood vessels (e.g. flushing), that occur due to oestrogen deficiency.
2. Osteoporosis: It may occur in long-term post-menopausal oestrogen deficiency because of bone loss.
3. Others: Atorphy of the sex tissues and increased risk of cardiovascular disease are the other complications.

Exercises

Essay type questions

1. Explain the arrangement of hypothalamus-pituitary hormone system and discuss function of hypothalamic release and release-inhibiting factors.
2. Discuss the chemical nature, biosynthesis and functions of steroid hormones and explain biochemical basis, and signs and symptoms of diseases associated with their over and underproduction.
3. Discuss chemical and physiologic properties, and actions of the hormones involved in digestion and metabolism of food substances. What controls blood levels of these hormones.

Write short notes on

1. Hypothalamus pituitary system
2. TRH
3. Somatostatin
4. Vasopressin
5. Prolactin
6. Reverse T_3
7. Cretinism
8. Sex steroids
9. Menstrual cycle
10. Anabolic steroids
11. GIT hormones
12. Pseudo-Cushing's syndrome
13. Abnormalities of catecholamine production
14. Conn syndrome

CLINICAL CASES

CASE 30.1 An apparent hermaphrodite with enlarged adrenals

Patient was referred to the Endocrinology department of a teaching hospital after the adrenal glands were detected to be enlarged. Circulating levels of all steroids were very low, though serum cholesterol level was normal. Detailed clinical examination revealed that the child was a genetic male, but his external appearance was predominantly that of a female, meaning that he had female external genitalia, but the internal wolffian duct structures (including epididymus, vas deferens and seminal vesicles) were male.

Cell-free extract was prepared from his adrenals, radioactive cholesterol added and production of pregnenolone was monitored. It was found to be low. Addition of 20-hydroxycholesterol to the reaction mixture was, however, followed by normal production of pregnenolone. The clinical diagnosis was lipoid adrenal hyperplasia.

Q.1. Name the defective enzyme and the reaction that it carries out.

Q.2. Why would you expect cyanide to inhibit the enzyme catalyzing this reaction?

Q.3. Why does addition of 20-hydroxycholesterol followed by normal production of pregnenolone?

CASE 30.2 Abnormalities of sexual development

External manifestations of this patient are similar to those described in Case 30.1: an apparent hermaphrodite, who is a genetic male (internal wolffian duct structures are male) but has female external genitalia. Circulating levels of testosterone were found to be normal for a male. A cell-free extract was prepared from the genital-skin fibroblasts. ^{3}H-testosterone was added and production of ^{3}H-dihydrotestosterone was monitored. It was found to be drastically decreased (by 90%). Addition of substrate ^{3}H made only a slight difference, but could not speed up the reaction to any significant extent.

The defective enzyme was purified, and its gene was cloned and sequenced. At codon 34, a mutation was detected:

Patient: AGG – codes for arginine
Normal: GGG – codes for glycine

Q.1. Identify the biochemical defect.
Q.2. How does this defect account for the phenotype—different internal and external genitalia?
Q.3. What exactly is the defect in the patient's enzyme?

CALCIUM AND PHOSPHATE: METABOLISM AND REGULATION

CHAPTER 31

Calcium and phosphate are the most abundant minerals in bones where they exist in form of crystal lattice. Nearly 99% of the total body calcium and 85% of the total body phosphate are present in these crystals of hydroxyapatite; the remainder are distributed in soft tissues, teeth and extracellular fluid (ECF). The plasma pool of these elements is relatively smaller (Ca = 9–11.5 mg/dL and phosphorous, largely as HPO_4^{2-} amounts to 2.7–4.5 mg/dL), but a multitude of cell and organ functions are dependent on the tight control of their concentrations in plasma. Under normal circumstances, small amounts of calcium, and presumably phosphate, are exchanged daily between bone and plasma as a result of constant bone remodeling, i.e. coupled processes of bone resorption by osteoclasts and bone formation by osteoblasts. The exchange process is part of a complex homeostatic mechanism which utilizes bone as a reservoir of calcium when deficiency exists, and as a store of calcium when the body is replete. It requires interplay of a number of regulatory factors, including hormones such as parathormone, calcitonin and calcitriol.

Calcium and phosphate homeostasis and various other aspects of metabolism of these minerals are described in this chapter. After going through the chapter, the student should be able to understand:

- Sources, dietary requirements, intestinal absorption, excretion and functions of calcium and phosphate.
- Role of various factors in regulation of plasma levels of calcium and phosphate; and causes, deficiency–symptoms and toxicity of these elements.

I. Calcium

Of all nutritionally important minerals, calcium occurs in largest amount in the human body: a 70 kg adult male contains 1.0–1.4 kg (25–33 g/kg of fat-free tissue) of calcium. Over 99% of the total body calcium is present in bones and teeth, and about 1% in various body fluids. *Extracellular calcium represents physiologically active fraction* being involved in a number of critical functions.

A. Nutritional Requirement and Sources

Daily dietary requirement of calcium is about 400–500 mg. This replaces the daily loss of 300–400 mg calcium in urine and an additional loss in faeces and sweat. In the growing age group (12–20 years), about *1200* mg per day is required.

It has been recommended that the daily intake of calcium in the **post-menopausal women** must be higher (around 1500 mg) since they are at risk of developing **osteoporosis**, a condition characterized by loss of bone organic matrix as well as progressive demineralization.

Several other factors influence the dietary requirement. For example, **vitamin D** is required for the optimal utilization of calcium, and therefore, adequate supply of this vitamin decreases the dietary requirement of calcium. Excess **dietary proteins**, on the other hand, may upset calcium balance by causing rapid excretion of this element. Exercise increases efficiency of calcium utilization. The calcium balance studies carried out in Peruvian Indians, who have extensive exposure to sunlight, perform adequate

exercise and subsist on low protein, vegetarian diet, indicate a need of only 300–400 mg of calcium per day.

Sources: Milk and milk products are the richest sources of calcium. Beans, cabbage, egg, fish and leafy vegetables are some other sources.

B. Functions

In Bone

Calcium contributes enormously to the physical strength of bones (and teeth). Within the bone matrix, type I collagen is the major protein (90%) and calcium-rich **crystals of hydroxyapatite** [$Ca_{10}(PO_4)_6(OH)_2$] are found on the collagen fibres. The crystals are very small (less than 0.1 nm long) and are surrounded by a hydration shell of water. Smaller amount of the bone calcium is present in form of calcium phosphate and calcium carbonate and some other calcium salts. Together, all these minerals constitute about 50% of the total skeletal mass; the rest is made up of organic elements.

The hydroxyapatite crystals contribute significantly to the strength of bones. The bone matrix consists of fibrils of collagen to which these crystals are attached in a regular fashion. Between the fibrils is the ground substance; a mixture of proteins, glycoproteins and proteoglycans.

The calcium stores in bones are in a dynamic equilibrium with the surrounding extracellular fluid. As much as 700 mg of calcium may leave or enter the bones each day. ***Thus, bones serve as prime reservoir of body calcium.*** Deposition of calcium in bones depends on serum concentration of calcium (and phosphate). When $Ca^{2+} \times PO_4^{-3}$ ion product rises beyond a limit value, calcification is promoted. A value above 70 mg/dL reflects tendency of soft tissue calcification, and below 20 mg/dL, it reflects defect of bone mineralization.

In Ionic Form

The calcium present in the body fluids, though extremely small in amount compared to that in bones, mediates a large number of physiological functions, as discussed here:

1. ***Muscle contraction*:** Calcium mediates excitation and contraction of muscle fibres. Contraction of striated muscles requires binding of calcium to troponin on thin filaments. The calcium-troponin interaction enhances reaction between actin and myosin, which triggers muscle contraction. In addition, calcium stimulates *ATPase*.

Smooth muscles do not have troponin. Their contraction is also calcium-dependent, but the calcium-sensing protein in smooth muscle is calmodulin, and not troponin.

2. ***Blood coagulation*:** Coagulation of blood occurs through a cascade of reactions, most of which require calcium.

3. ***Nerve excitability*:** Excitability and conductivity of nerves depends upon a number of cations, including calcium. A raised plasma calcium level decreases and a low plasma calcium level increases the excitability of nerves.

4. ***Activation of enzymes*:** Calcium causes direct activation of a number of enzymes, such as *succinate dehydrogenase, ATPase* and *pancreatic lipase.* Intracellularly, it interacts with **calmodulin**, a calcium-binding regulatory protein, and the calcium-calmodulin complex activates certain enzymes (Table 29.6).

As Intracellular Messenger

Cytoplasmic calcium is an important intracellular signal. It is referred to as a "second messenger" because it mediates cellular response to a wide range of stimuli in a manner analogous to cAMP. For example, it acts as a second messenger for epinephrine or glucagon in hepatic glycogenolysis. Some authors consider calcium as the second messenger in phosphoinositide system, and as a third messenger for some hormones, such as ADH (Chapter 29).

Other Functions

1. ***Neuromuscular transmission:*** Neuromuscular transmission occurs through release of acetylcholine from the motor end plate, which requires calcium.

2. ***Membrane integrity and permeability:*** Transport of a number of substances across the membranous barrier is influenced by calcium.

3. ***Action on heart:*** Cardiac muscles are dependent on calcium for the generation of rhythmic impulses. Increased calcium concentration increases myocardial contractility and vice versa.

4. ***Secretory processes:*** Secretion of water-soluble products by *exocytosis* is triggered by calcium. Examples include the release of zymogens by pancreas, insulin from β-cells, histamine from mast cells and neurotransmitters from nerve terminals. Endocytosis–exocytosis, cell motility and other such processes mediated via microtubules-microfilaments are also regulated by calcium.

Bone metabolism is closely interrelated with the metabolism of calcium, which also involves the instestine and the kidneys. Unbound calcium is physiologically active, being required for nerve function, muscle contraction, membrane permeability, blood coagulation and glandular secretion.

C. Metabolism

Absorption

Dietary calcium is absorbed in the duodenum and proximal jejunum through mediation of an intestinal **calcium-binding protein (CBP)**. This protein transfers the luminal calcium across the intestinal mucosal cell by an energy dependent active process.

Factors promoting absorption: Synthesis of CBP is enhanced by 1,25-dihydroxycholecalciferol, i.e. calcitriol, the activated form of vitamin D. This accounts for stimulation of intestinal calcium absorption by *calcitriol.*

Calcitriol → Calcium-binding protein → Calcium absorption

Table 31.1 lists various factors that increase calcium absorption.

Factors inhibiting absorption: Calcium absorption is inhibited by certain compounds such as *oxalates*, *phytates*, and *phosphates*. These compounds form insoluble calcium salts. The undigested *dietary fats* also impair calcium absorption by forming insoluble calcium soaps. Dietary *fibres* also interfere with the absorption. Thus, a large part of dietary calcium is not absorbed and is eliminated in faeces.

Plasma Calcium

Following absorption, calcium pours into blood plasma. Total plasma calcium is 9–11 mg/dL, and it exists in three forms:

1. ***Ionized calcium:*** It is the biologically active fraction of the calcium in plasma. Maintenance of its concentration within tight limits (4.5–5.0 mg/dL) is required for nerve function, membrane permeability, muscle contraction and glandular secretion.

Table 31.1. Factors affecting intestinal calcium absorption

Factors increasing absorption
Calcitriol
PTH (acts by enhancing calcitriol production)
Acidity (low pH)
Growth hormone
Pregnancy, lactation
Lactose, arginine, lysine
Factors decreasing absorption
Oxalates and phytates
High dietary fats and fibres
Phosphates
Alkalinity
Chronic renal failure (leads to impaired activation of vitamin D)

2. ***Protein bound calcium:*** The majority of the remaining calcium is mainly bound to negatively charged albumin.
3. ***Calcium complexed to substances such as citrate and phosphate*:** It constitutes a small fraction (10% of total).

Serum calcium (9–11 mg/dL)
- 50% – Ionized Calcium
- 40% – Protein-bound Calcium
- 10% – Calcium complexed with citrate, phosphate, bicarbonate

Excretion

Calcium is ultimately excreted through the following routes:

1. A large amount (~200 mg/day) is secreted into the intestinal lumen and lost in faeces.
2. About 300–400 mg/day is lost in urine; the kidneys start filtering calcium when the plasma levels exceed 7.0 mg%.
3. A smaller amount is lost in sweat.

Fecal calcium excretion varies widely in response to diet, whereas elimination through other routes remains relatively constant. Various aspects of calcium metabolism and the factors regulating them are shown in Figure 31.1.

Intestinal absorption of calcium is enhanced by calcitriol and parathormone, and decreased by phytate, oxalate, free fatty acids and fibres.

D. Regulation of Serum Calcium Levels

Calcium homeostasis is modulated by hormones. ***Parathyroid hormone***, ***calcitriol*** and ***calcitonin*** are the most important calcium regulators; others, such as *thyroxine*, *growth hormone*, and *oestrogen* play a relatively minor role.

PTH and calcitriol have hypercalcaemic effect, whereas calcitonin decreases the serum calcium levels. Their actions are interlinked: PTH enhances production of calcitriol

25-(OH)-cholecalciferol —(PTH, ⊕)→ 1,25-$(OH)_2$-cholecalciferol (Calcitriol)

Calcium balance is hormonally regulated by parathormone, vitamin D metabolites and calcitonin. The complex homeostatic mechanism utilizes bone as the reservoir of calcium when deficiency exists, and as a source of calcium when the body is replete.

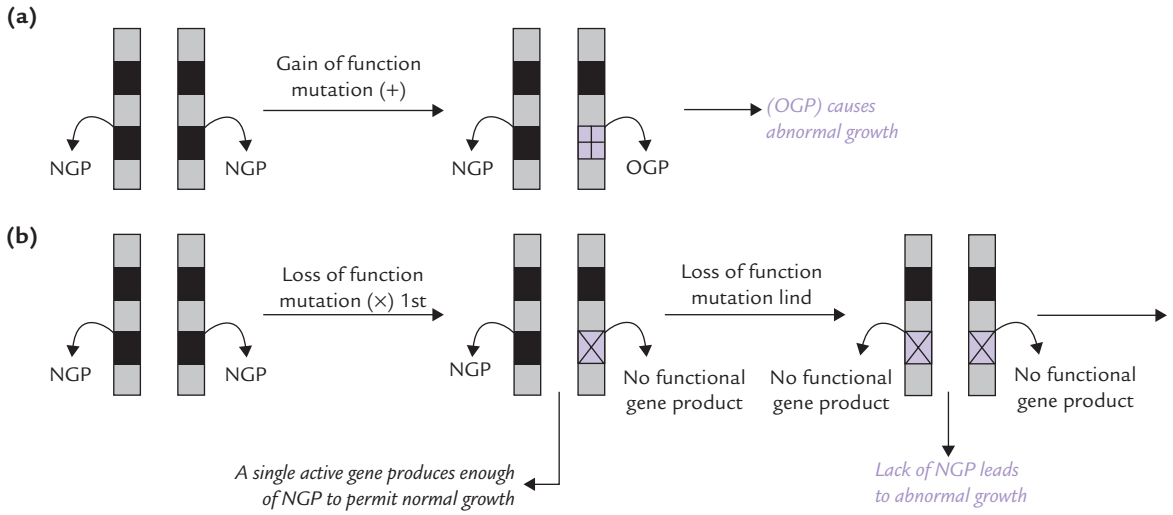


Fig. 31.1. Metabolism of calcium. Amount of calcium intake and its deposition in bone are exactly matched by its excretion in urine (EFC = extracellular fluid, PTH = parathormone, D_3 = calcitriol).

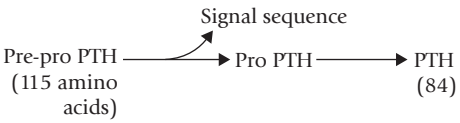
Parathyroid Hormone (Parathormone; PTH)

Source

PTH is produced by four small glands having a total weight of 0.05–0.30 g. Parathyroid glands can be found anywhere in the neck or upper mediastinum but are usually located just behind the posterior thyroid capsule.

Structure of PTH

Parathormone (PTH) is an 84-amino acid single chain peptide (MW 9500). Biological activity of the hormone resides in the N-terminal portion of the molecule: PTH 1–34 possesses full biological activity.



Synthesis and Secretion of PTH

Parathormone is secreted as a 115-amino acid precursor molecule, the *pre-pro-parathormone*. It is converted to *pro-parathormone* by removal of a 25-amino acid sequence, the so called hydrophobic signal sequence from the N-terminal. Pro-parathormone is transported to the Golgi apparatus where enzymatic removal of another 6-amino acid sequence from the N-terminal occurs, to generate the parathormone.

Synthesis of the PTH, as well as its secretion, is markedly enhanced by low ambient concentration of Ca^{2+} (and Mg^{2+}). Conversely, hypercalcaemia suppresses the parathyroid glands. The PTH levels fall in a rectilinear fashion in relation to the serum calcium levels between 4 mg/dL and 10.5 mg/dL.

Transport and Metabolism

Like other peptide hormones, PTH has a short half-life (< 20 min), because it does not bind with any transport protein in the blood circulation. It is quickly removed by the liver and the kidneys and is subjected to proteolytic degradation in these tissues. Kupffer cells are especially active in this process. Specific fragments of the PTH (PTH 34–84, and PTH 37–84) are generated in the liver; the role of kidneys may be to remove and excrete these fragments.

(A number of proteolytic enzymes, including *cathepsin B* and *D* have been identified in the parathyroid tissue itself.)

Mechanism of Action

PTH interacts with specific receptors (MW 70,000) located on the target cell surface. The receptors appear to be identical in bones and kidneys, but they are linked with different intracellular messenger systems.

Renal cells: The hormone-receptor interaction activates *adenylate cyclase* to generate cAMP.

This in turn initiates a signaling cascade, which promotes phosphorylation of proteins (by *kinases*) and finally brings about biological actions.

Osteoblasts: The hormone-receptor interaction acts through the second messangers (IP_3 and DAG) to cause release of calcium from the intracellular stores. Interaction of calcium with certain intracellular enzymes and proteins results in alteration in properties of the latter, and this brings about biological effects of PTH (Box 31.1).

In renal cells, the *adenylate cyclase*-cAMP axis is important, whereas the receptors on osteoblasts are linked with PLC, and act through phosphoinositides (IP_3) and diacylglycerol (DAG) as second messengers.

BOX 31.1

Parathormone Mobilizes Bone Calcium, Using Calcium as Second Messenger

PTH acts on the *Phospholipase* C (PLC) linked receptors on osteoblasts to elevate intracellular calcium concentration. The calcium which allosterically alters conformation of various intracellular proteins and membrane-proteins. The altered conformation of membrane-proteins results in making the membrane more permeable.

Inward movement of calcium into the cells from the surrounding calcium-rich bone matrix (along a concentration gradient) follows. Intracellular concentration of calcium thereby rises further. The initial rise in intracellular calcium concentration is followed by mobilization of calcium into blood circulation. This is because at this stage, **calcium pump**, an enzyme located on the cell membrane, is activated (its main activator is calcitriol). The activated pump promptly moves calcium out of the cells into the serum, resulting in hypercalcaemia.

Biological Effects

Target tissues of the PTH are bone, kidney and small intestine. PTH exerts a direct effect on bone and kidneys, whereas its effect upon the intestine is mediated via calcitriol.

Effect on bones: PTH acts on both osteoblasts, which build bones, and osteoclasts, which destroy bones (i.e. bone resorption). Initially, PTH interacts with the cell surface receptors on the osteoblasts and triggers a series of events that cause mobilization of calcium from bones (by activating a ***calcium pump*** on cell membrane; Box 31.1). Calcium release into the blood circulation is accompanied by release of phosphate as well, because the two are present together in the hydroxyapatite crystals. Uptake of calcium and phosphate by bone is also decreased by PTH, resulting in *hypercalcaemia* and *hyperphosphataemia*. The above effects take place almost instantaneously, and are termed short-term effects.

These are followed by the ***long-term effects*** that begin about 12 hours later. These involve differentiation of the precursor cells into the osteoclasts. Osteoclasts cause dissolution of the bone matrix, resulting in release of various constituents including the organic matrix elements, calcium and phosphate. This process is called bone resorption and it results in **decalcification** and **demineralization** of bones. Recent evidence suggests that intracellular synthesis of the enzymes of bone resorption is stimulated by the PTH.

Effect on kidneys: PTH increases calcium reabsorption from the distal convoluted tubules (Table 31.2). However, due to its hypercalcaemic effect, more calcium load is presented to the kidneys than can be reabsorbed. The excess is excreted in the urine (i.e. hypercalciuric effect; **Case 31.1**).

PTH increases renal clearance of phosphate by inhibiting its reabsorption in the proximal convoluted tubules. This results in increased urinary excretion of phosphate, i.e. **phosphaturia**. The phosphaturic effect of PTH overrides its hyperphosphataemic effect, so that *there is net fall in serum phosphate levels by PTH*.

Table 31.2. Control of calcium and phosphate metabolism

Agent absorption	Uptake by bone	Intestinal absorption	Kidney tubule reabsorption	Fecal excretion
PTH	$P_i\downarrow$ Ca↓*	$P_i\uparrow$ Ca↑	$P_i\downarrow$ Ca↑	
Calcitriol	$P_i\downarrow$ Ca↓	$P_i\uparrow$ Ca↑	$P_i\uparrow$ Ca↑	
CT	$P_i\uparrow$ Ca↑	No effect	$P_i\downarrow$ Ca↓	$P_i\uparrow$

* PTH elicits both short-term and long-term effects on bones (see text).
PTH = parathormone, CT = calcitonin.

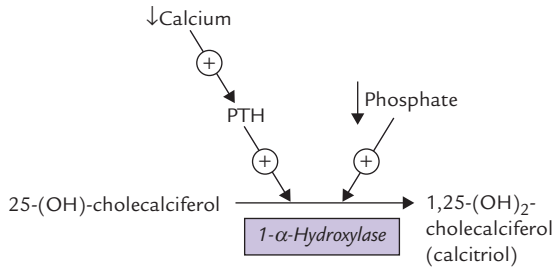


Fig. 31.2. PTH stimulates formation of calcitriol by stimulating *1-α-hydroxylase*, the renal tubular enzyme. Hypophosphataemia has same effect.

Effect on intestine: Effect of PTH on intestine is mediated via calcitriol. PTH increases calcitriol generation from 25-hydroxycholecalciferol by activating the renal tubular enzyme *1-α-hydroxylase* (Fig. 31.2). Calcitriol then enhances calcium absorption from intestine.

PTH promotes calcium reabsorption from the kidney, resorption from bone and absorption from small intestine via increased production of calcitriol.

Calcitriol

Calcitriol increases calcium absorption from the small intestine by intestinal transcription and translation of a specific gene that encodes for a **calcium-binding protein** (Chapter 18). The latter binds the available calcium in the intestinal mucosal cells and transfers it across the intestinal mucosa.

Together with PTH, calcitriol decreases calcium uptake by bone and enhances bone resorption by osteoclasts. These effects increase serum calcium (and phosphate) levels.

Effects of calcitriol on other target tissues are summarized in Table 31.2. Low calcitriol causes abnormal mineralization of newly formed osteoid as a result of low calcium and phosphate availability, which results in rickets among children, and osteomalacia in adults.

Calcitonin

Source and Structure

Calcitonin (CT) is a 32-amino acid peptide, with one disulfide bond and a characteristic N-terminal loop. It is synthesized and secreted mainly by the parafollicular C-cells of the thyroid, and to a lesser extent, by parathyroid gland and thymus.

Secretion

Secretion of calcitonin is inversely related to that of PTH. As in case of PTH, secretion of calcitonin is controlled by the ECF ionized calcium (and probably magnesium) concentration, but the stimulus for the hormone release is hypercalcaemia. This is in contrast with PTH where hypocalcaemia is the stimulus for the release. CT secretion increases linearly when calcium concentration is between 9.5 mg/dL and 15 mg/dL.

Glucagon and *pentagastrin* are potent CT secretagogues. The pentagastrin is used in the provocative test for medullary thyroid cancer, a malignant tumour of the calcitonin-secreting parafollicular c-cells.

Biological Effects

The precise role played by the CT is still unclear. It appears to prevent hypercalcaemia by decreasing the serum calcium.

Action of calcitonin on the bones is opposite to that of the PTH; it inhibits osteoclast-mediated bone resorption and, therefore, decreases the release of calcium and phosphate. It promotes the phosphate influx into the bone cells from blood circulation with a concomitant fall in extrusion of calcium. All these actions result in *hypocalcaemia* and *hypophosphataemia*, and account for effectiveness of CT in treating the hypercalcaemia associated with cancer. These actions are independent of PTH, probably because the target cells for CT (wherein it elevates the cAMP levels) are not same as the ones on which the PTH acts.

Steroid and Peptide Hormones

Several hormones for which the primary action is not related to calcium regulation, directly or indirectly affect ***calcium homeostasis***.

1. **Thyroid hormones** stimulate osteoclast-mediated resorption of bone to cause hypercalcaemia.
2. **Adrenal and gonadal steroids**, particularly oestrogen sin women and testosterone in men, have important regulatory effects, increasing osteoblast and decreasing osteoclast function. They also decrease renal calcium (and phosphate) excretion and PTH function.
3. **Growth hormone** has anabolic effects on bone, mediated by insulin-like growth factors (IGF-I and IGF-II). Growth hormone increases the urinary excretion of calcium (while decreasing that of phosphate).

Self-regulatory Loop in Calcium Homeostasis

The circulating (unbound) calcium plays a role in calcium homeostasis via self-regulatory loop. For instance, hypocalcaemia induces PTH secretion, which elicits its various actions on bone, intestine and kidneys. As a result serum calcium level rises to normal. The self-regulatory loop is illustrated in Figure 31.3.

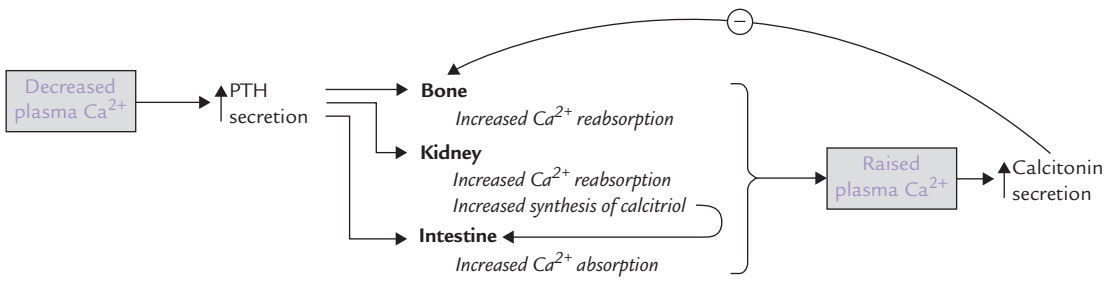


Fig. 31.3. Self-regulatory loop in regulation of serum calcium level.

E. Disorders of Calcium Metabolism

The factors discussed so far in this section ensure meticulous regulation of serum calcium level within a narrow range. However, alteration of serum calcium levels—hypercalcaemia or hypocalcaemia—may occur in a number of conditions.

Hypercalcaemia

The commonest causes of hypercalcaemia are primary hyperparathyroidism and hypercalcaemia of malignancy. Inappropriate dosage of vitamin D or metabolites, e.g. in the treatment of hypoparathyroidism or renal disease, also causes hypercalcaemia (Table 31.3).

Clinical features: There is wide variation among individuals in the development of numerous signs and symptoms. Common ones are:

1. Neurological and psychiatric features such as lethargy, confusion or depression.
2. Cardiac arrhythmias.
3. Renal features, e.g. thirst, polyuria and calculi.
4. Gastrointestinal problems, e.g. nausea, abdominal pain, vomiting and constipation.

Hypocalcaemia

Many acute and chronic illnesses lead to a decrease in serum albumin, which in turn decreases serum total calcium concentration (Chapter 5). Other important causes are:

1. *Hypoparathyroidism*: Idiopathic, after neck surgery, or occasionally due to anticonvulsant therapy.
2. *Vitamin D deficiency:* It may occur due to dietary deficiency, malabsorption, or inadequate exposure to sunlight.
3. *Renal disease*: The diseased kidney fails to synthesize calcitriol.
4. *PTH resistance*: In *pseudohypoparathyroidism,* the hormone is secreted but there is failure of target tissue receptors to respond to the hormone. Hypomagnesaemia also causes PTH resistance.
5. *Rarer causes*: Malignancy, acute pancreatitis, sarcoidosis, multiple myeloma and thyrotoxicosis.

Table 31.3. Causes of hypercalcaemia

Common causes
• Primary hyperparathyroidism
• Malignant diseases
• Iatrogenic: overdose of vitamin D or metabolites
Uncommon causes
• Thyrotoxicosis
• Calcium therapy
• Prolonged immobilization
• Drug induced: thiazide diuretic, lithium
• Multiple myeloma sarcoidosis
• Renal failure: acute and chronic

Clinical features of hypocalcaemia: These include neurological features, such as tingling, tetany; cardiovascular signs, such as abnormal ECG; and cataract.

The measurement of calcium in serum is an important test, because both hypercalcaemia and hypocalcaemia lead to clinical symptoms.

Osteoporosis

A chronic dietary deficiency of calcium accelerates the net loss of bone mass after the age of about 35 years. This process, called bone resorption, can ultimately lead to *osteoporosis*. Most common sufferers are the postmenopausal women. Osteoporosis is characterized by frequent bone fractures, which are a major cause of disability among the elderly people.

The risk of fracture in postmenopausal osteoporosis is treated by increasing the bone mass through oestrogen replacement and by calcium supplementation. Most often this treatment is given in combination with vitamin D. It is recommended that the current average calcium intake (approximately 500 mg/day) be supplemented with an additional 1000 mg/day calcium in these women. Milk and milk products are recommended since they are rich sources of the vitamin. Regular exercise also tends to increase the bone mass.

Osteoporosis, a disease associated with aging, is characterized by a significant reduction of bone mineral density with an increased susceptibility to fractures.

II. Phosphorus

Total amount of phosphorus in a 70 kg adult is about **1000 g**, of which, about 85% is present as hydroxyapatite in bones. About 10% is found in muscles and bones in association with proteins, carbohydrates and lipids. The rest is distributed in various compounds in the extracellular fluid (ECF) and the intracellular fluid (ICF).

Serum phosphate concentration of adults ranges from **2.4 mg/dL** to **4.5 mg/dL**. In children, it is higher because of higher bone turnover.

A. Nutritional Requirement

An average daily intake of **1 g** is sufficient to meet the body needs. Being a component of a number of biomolecules, phosphorus is present in a variety of foods of both plant and animal origin. *Leafy vegetables*, *milk*, *meat*, *egg* and *cereals* are especially rich sources.

B. Functions

Phosphorus plays important role in intracellular as well as extracellular fluids.

Role in ICF

1. *Component of biomolecules:* Phosphorus (as phosphate) is present in a vast array of biomolecules such as nucleic acids, nucleotides, phospholipids, and some proteins.
2. *The high-energy compounds* (e.g. nucleoside phosphates like ATP, GTP, CTP, etc.) and several coenzymes (NAD^+, $NADP^+$, FAD, FMN) also contain phosphate.
3. *Phosphate buffer* is an important intracellular buffer.
4. *Phosphorylated intermediates* play important role in energy metabolism (Chapter 9).
5. Activation of several proteins and enzymes occurs by phosphorylation.

Role in ECF

The extracellular phosphate is important for calcification. In bones, fate of phosphate is linked with that of calcium, as discussed earlier.

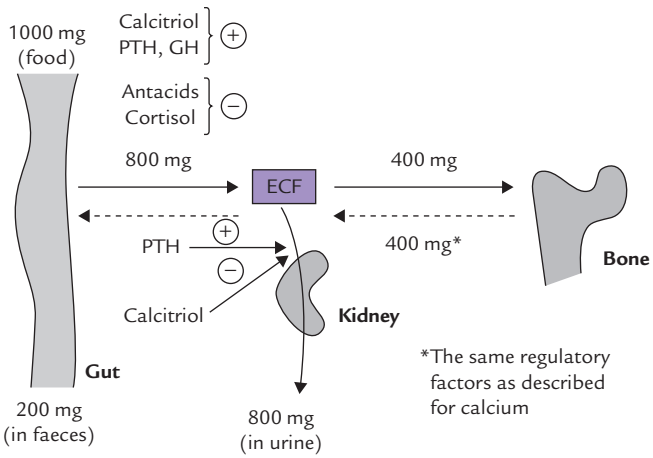


Fig. 31.4. Metabolism of phosphorus.

C. Metabolism

Most of the dietary phosphate (70–80%) is absorbed in gut; the absorption is maximum when the dietary Ca : P ratio lies between 1 : 2 and 2 : 1. The absorption requires mediation of calcitriol. PTH also increases the phosphate absorption mainly by enhancing calcitriol production. Excessive use of antacids decreases absorption of dietary phosphates. This is because heavy metal ions present in antacids (Mg^{2+}, Al^{3+} and Ca^{2+}) bind phosphate ions to form insoluble salts. Other factors that influence intestinal absorption are shown in Figure 31.4.

In serum, phosphate circulates in free ionic form (40%) or in a complex form (50%) with cations, such as, Ca^{2+}, Mg^{2+}, K^+ or Na^+. It is deposited in bones as hydroxyapatite; the process is mainly regulated by calcitriol. Excretion of phosphate occurs mainly through the renal route. About 90% of the circulating phosphate is filtered in glomeruli, and its further fate depends on action of calcitriol and parathormone on renal tubules. Calcitriol enhances the renal tubular reabsorption, but parathormone diminishes it. *Thus, calcitriol has phosphate retaining effect, which contrasts with the phosphaturic effect of PTH* (Table 31.2). Effect of PTH is far more potent than calcitriol. In all, 85–95% of the filtered phosphate is reabsorbed in renal tubules and the rest is excreted, which amounts to about 800 mg per day. About 200 mg phosphate is eliminated in faeces each day (Fig. 31.4).

D. Regulation of Serum Phosphate Levels

A complex interplay of various hormones, such as PTH, calcitriol and calcitonin keeps plasma phosphate levels within the narrow range of 4.5–5.0 mg/dL (Table 31.2). It is increased in hyperparathyroidism and decreased in hypoparathyroidism, rickets and renal distress.

E. Deficiency

Excessive excretion through kidneys or decreased absorption from intestine is the major cause of phosphate depletion. Dietary deficiency is extremely rare due to wide distribution of phosphate in a variety of foodstuffs.

Effect of phosphate depletion is most pronounced on skeletal system. Decreased mineralization of bones leading to rickets-like symptoms may occur (**Case 31.2**). In addition, abnormalities of erythrocytes, leukocytes and platelets may occur.

F. Toxicity

Renal failure, both acute and chronic, may result in phosphate retention, thereby leading to phosphate toxicity. There is a concomitant fall of circulating calcitriol level due to inhibitory effect of phosphate on the calcitriol production. Because calcitriol has hypercalcaemic effect, serum calcium levels tend to fall in such cases.

Treatment of phosphate toxicity involves oral administration of antacids, which bind phosphate in intestinal lumen and thereby inhibit its absorption. As the serum phosphate levels are lowered, a concomitant rise in calcitriol also occurs. The latter then promotes calcium absorption from the intestine to normalize the circulating calcium levels.

Exercises

Essay type questions

1. What are the intracellular functions of calcium ion in human body? Explain the manner in which calcium homeostasis is maintained.
2. Discuss hormonal regulation of serum calcium level and write a note on hypercalcaemia.

Write short notes on

1. Calmodulin
2. Hypocalcaemia
3. Calcium homeostasis

CLINICAL CASES

CASE 31.1 Visual impairment following thyroid surgery

A 52-year-old woman reported in the ophthalmology OPD with complaints of dimness of vision for the past six months. She was apparently in good health apart from her failing vision. Further enquiry revealed that she had numbness and paraesthesia in the extremities, and occasionally suffered from muscle cramps and spasms. She had undergone complete thyroidectomy for a multinodular goitre two years earlier.

Physical examination showed that the Chvostek's sign and the Trousseau's sign were positive. Ophthalmic examination revealed presence of bilateral mature cataract. The patient was admitted for lens extraction. Blood sample was sent to the biochemistry laboratory for the routine pre-operative and other relevant investigations. The results are as below:

Test	Patient's reports	Reference range
Blood glucose (F)	80.4 mg/dL	< 100 mg/dL
Serum urea	32 mg/dL	15–40 mg/dL
Serum calcium	6.4 mg/dL	9.5–10.5 mg/dL
Serum phosphate	7.8 mg/dL	2.5–4.5 mg/dL
Serum albumin	4.4 g/dL	3.6–5.4 g/dL
Serum A : G ratio	1.3	1.1–1.5
Serum alkaline phosphatase	70 U/L	401–100 U/L

Q.1. What is the most probable diagnosis? State the most likely cause of cataract.
Q.2. What further tests would be appropriate for the diagnostic confirmation?
Q.3. Why was it important to estimate serum albumin concentration in this patient?
Q.4. Explain the biochemical basis of signs and symptoms of this patient.
Q.5. What is the rationale behind the use of vitamin D in this condition?
Q.6. What are the long-standing complications of this condition?

CASE 31.2 A 2-year-old child with slow development and skeletal deformities

A 2-year-old child was brought to OPD with skeletal deformities such as bow legs, kyphosis, and bossing of frontal and parietal bones. His early development had been delayed; he could not learn to crawl and walk at appropriate age. These features were suggestive of rickets. X-Ray examination revealed demineralization, thinning of bone cortex and thickening of zones of epiphyseal cartilage; these features were also in line with the above tentative diagnosis. However, serum calcitriol estimation showed that concentration of this hormone was elevated. This was contrary to the expected low value seen in rickets. Further biochemical investigations were, therefore, performed to reach at a conclusive diagnosis.

Investigations test	Patient's report	Reference range
Serum calcium	10.8 mg/dL	8.7–10.7 mg/dL
Serum phosphate	1.9 mg/24 h	2.4–4.5 mg/dL
Urine calcium	298 mg/24 h	50–150 mg/dL
Urine phosphate	306 mg/24 h	45–150 mg/dL

The circulating parathormone level was decreased. Orally administered calcium and phosphate were found to be absorbed normally from the intestine.

In view of the clinical diagnosis of rickets, initial treatment with oral calcium and calcitriol supplementation was tried. It did not produce any improvement either in the clinical picture or in the biochemical profile. Treatment with parathormone injections also failed to produce any beneficial effect. However, a dramatic improvement was observed with dietary phosphate supplements, to the extent that the patient's symptoms virtually disappeared.

Q.1. Identify the biochemical defect in this patient.
Q.2. Propose a mechanism by which this biochemical defect leads to poor mineralization of bones.
Q.3. Explain the biochemical basis of the following findings: (a) Hypercalcinuria, (b) Elevated serum calcitriol levels.
Q.4. What treatment would you recommend for this patient?

CHAPTER

CANCER

32

Cellular proliferation is a regulated process, which ensures that cell division is coordinated with physiological demands of the tissue or the organ. Under normal conditions, cells of adult body divide only to replace those cells that die (***apoptosis***) after fulfilling their task. Derangements of the regulatory mechanisms result in uncontrolled and excessive cellular proliferation—the cells are said to be malignantly transformed, and the condition is known as **cancer**. Thus, cancer refers to a state of abnormal, excessive and uncontrolled proliferation of the cells of a tissue; the cellular proliferation far exceeds physiological requirements of the tissue. The proliferating cells may migrate beyond the tissue of origin and infiltrate the adjacent tissues. The word cancer is derived from the Greek word, *karikoma* meaning "crab": the cells from a mass of cancer tissue move away like feet of a crab.

After going through this chapter, the student should be able to understand:

- Terminology relating to cancer and distinctive features of cancer cells.
- Factors implicated in aetiology of cancer: the somatic mutations causing activation of **protooncogene** or inactivation of **antioncogenes**; the oncogenic retroviruses and DNA viruses; chemical carcinogens; and ionizing radiations.
- Mechanism of activation of protooncogenes by somatic mutations, gene rearrangements, gene amplification, promoter insertion and enhance insertion.
- Role of **growth factors** in causing malignant transformation of cell, and **cancer markers** in diagnosis and follow up cancer patients.

I. Tumours: Benign and Malignant

Excessive cellular proliferation leads to accumulation of an abnormal cell mass called *tumour*. If the growth of the tumour is self-limiting, the tumour is said to be **benign**. However, if the cells have high propensity for invasion and distant spread, the tumour is **malignant** (from the Latin, meaning "engendering harm"). Generally benign tumours are not life-threatening (e.g. moles, warts), and not even considered as cancers. The term cancer is reserved as for the diseases resulting from malignant tumours. The diseases resulting from the benign and the malignant tumours are collectively called **neoplasia** (Greek word for "new growth").

Cancer is characterized by derangements of cellular growth control, resulting in excessive proliferation and accumulation of an abnormal cell mass called tumour.

The malignant tumours are classified according to the type of somatic cells from which they are derived. For example, tumours of epithelial origin are known as **carcinomas**, and the tumours originating from connective tissue are called **sarcomas**. Although the transformed cells may retain some of the important characteristics of the cells of their origin (carcinomas produce keratin, a product of epithelial cells; and sarcomas produce certain constituents of extracellular matrix), these cells acquire some distinctive features not seen in the normal cells.

Incidence of cancers: More than 100 different types of human cancers are known, which are *second only to heart disease as the major health issue* in the developed world, accounting for about 23% deaths annually. Among children of 7–13 year age, cancer is the leading cause of death: leukaemia accounting for about 50% of cancer-related deaths. Among adults, the incidence of cancers increases with advancing age; about two-third of newly diagnosed cancers have been reported in persons over 60 years.

II. Distinctive Features of Malignant Cells

Some of the distinctive features of cancer cells are:

De-differentiation: The cancer cells assume features of the **immature stem cells**. For example, carcinoma cells lose the normal columnar or cuboidal shape of the progenitor cells and assume an undifferentiated appearance.

Moreover, the malignantly transformed cells often lose capability to synthesize specialized proteins which are synthesized by the normal cell. For example, the cancerous liver cells may stop producing the normal liver enzymes and instead start secreting certain proteins that are normally synthesized by the fetal liver; for instance, the α-fetoproteins (AFP). Such abnormal proteins are called **tumour markers**.

High mitotic rate and disordered growth: The rapidly proliferating cancer cells have high mitotic rate. These cells are capable of growing in the absence of normal mitogenic stimuli.

The growth of cancer cells is disorganized and disordered. The cells spread in all directions, often crossing the anatomical boundaries of the affected tissue. Consequently, they infiltrate into the surrounding areas.

Dedifferentiation and disordered growth are together referred to as **anaplasia**. The greater the degree of anaplasia, higher the malignant behaviour of the cells and poorer is the prognosis for the patient.

Genetic abnormalities: The cancer cells commonly show certain genetic abnormalities such as abnormal chromosomal number, deletion or translation of chromosomes, gene amplification or extrachromosomal genetic elements. These aspects of cancer cells are described later in this chapter.

Hereditary: Cancer is a manifestation of some *heritable alteration in the genome*. This enables them to divide excessively without losing the distinctive neoplastic characteristics.

Monoclonal origin: All malignant cells in a patient descend from a single aberrant somatic cell. In the other words, cancers are monoclonal in origin. For example, in patients with chronic myelogenous leukaemia (CML) the characteristic chromosomal translocation between chromosome 9 and chromosome 22 (called the *Philadelphia chromosome*) are present in all tumour cells, but not in the normal somatic cells. This observation confirms monoclonal origin of the cancer cells.

Loss of apoptosis: Apoptosis refers to death of a cell after its function has been fulfilled (Box 32.1). Cancer cells lack this property. Even in extremely adverse circumstances such as lack of growth factor, DNA damage, or other environmentally induced injuries, the cancer cells retain ability to survive and proliferate.

Biochemical alterations: The cancer cells show a number of biochemical changes such as:

(i) The **glycolytic sequence** is greatly enhanced, indicating higher energy demands of the multiplying cells.
(ii) The **membrane permeability** and cell-surface charge are altered.
(iii) Activities of certain enzymes, such as proteases, are decreased. The isoenzymes typical of fetal cells, make appearance in cancer cells.
(iv) The **antigens** present on cell membrane are **altered**; certain new antigens appear, while some antigens present on the normal cell surface are lost.
(v) Rates of synthesis of DNA and RNA are much higher than that in the normal cell.
(vi) Tumour cells can grow in the absence of normal mitogenic stimuli. In cell cultures they exhibit normal growth under conditions of growth-factor deprivation. However, the quantity of growth factors secreted by tumour cells is greater than normal cells.

Metastasis

The cancer cells can spread to distant tissues through blood or lymph. Such spread of the cancer cells and colonization of distinct tissues is called **metastasis**. On reaching the far tissues in this manner, the cancer cells form *secondary tumour*.

Metastasis is the major cause of cancer-related morbidity and mortality. Biochemical basis of metastatic spread are not clearly understood, but it is believed to require *poor adherence* to neighbouring cells in the tumour and *loss of contact inhibition*.

- ***Loss of contact inhibition:*** Normal cells growing in tissue culture show contact inhibition. When these cells are allowed to grow in tissue culture medium on a surface, such as that of a Petri-dish, they grow and divide until the surface is covered with a monolayer

BOX 32.1

Apoptosis

Apoptosis, a form of **programmed cell death,** is required to maintain integrity and homeostasis of multicellular organisms. It is an important biological process, impairment of which may lead to cancer. On the other hand, activation of the apoptosis machinery can result in neurodegenerative disorders. Therefore, apoptotic cell death is a meticulously controlled and highly ordered process. Recent research has elucidated two biochemical signals that appear to be central to regulation of apoptosis: the *cysteine proteases* or ***caspases***, and the **B-cell lymphoma protein** (bcl-2) related family members.

- The *caspases* are executor of cell death: they cause systematic disassembly of the apoptotic cell by cleaving some key target proteins.
- The bcl-2 proteins modulate cell survival by suppressing the *caspase* activity. Over-expression of bcl-2 genes decrease apoptosis, thereby leading to cancers.

Dramatic morphological changes occur in the cells that undergo apoptotic cell death, such as shrinkage of cell and condensation of chromatin. Systematic disassembly of these cells into membrane enclosed vesicles, called ***apoptotic bodies*** occurs, which is followed by:

- Disruption of cell cycle.
- Dismantling of structural components such as the cytoskeleton.
- Detachment of the cell from the surrounding extracellular matrix.
- Flagging the apoptotic cells for elimination via phagocytosis by macrophages or by neighbouring cells.

Two markers have been identified recently that serve to flag the apoptotic cells for phagocytosis. These are accumulation of sterols at the plasma membrane and translocation of phosphatidylserine to the outer leaflet of the plasma membrane.

and further growth is inhibited. Cancer cells, however, show loss of contact inhibition and continue to grow beyond the monolayer to form multilayered masses of cells.

- ***Poor adherence to surface:*** Normal cells are firmly adhered to the surface, but cancer cells show loss of such anchorage properties. Alteration in structure of the protein, *vinculin*, is possibly responsible for this.

There are typical differences between malignant neoplastic cells and the normal progenitor cells: cancer cells are immortal, grow in absence of mitogenic stimuli; have abnormally high mitotic rate, gross genetic abnormalities, disordered growth patterns, loss of contact inhibition; and exhibit features of immature stem cells.

III. Aetiology

Cancer is a multifactorial disease. A number of factors are implicated in its aetiology, viz. genetic, hormonal, metabolic, physical, chemical and environmental. However, most cancers are spontaneous and no specific cause could be held responsible for their development. Each day, a few cells may undergo spontaneous mutation and become cancerous, but such cells are promptly removed by the immune system. Cancer develops only when (a) the number of mutant cells is too large to be removed by immune system, and (b) immune system is not efficient enough to remove the mutant cells.

The factors that are implicated in cancer aetiology appear to be vastly diverse, but all of them converge at a single point. All produce certain permanent changes in DNA, which are transmissible at cell divisions. Thus, **cancer is truly a disease of genome**.

Most human tumours are non-viral in origin and arise from spontaneous or induced mutations. Certain *viruses*, *chemicals* and *ionizing* radiations are also capable of causing similar changes in DNA, and hence fostering cancerous transformation of the affected cell.

A. Somatic Mutations

Somatic mutations can cause malignant transformation of the cell in two ways:

1. By activation of the growth promoting genes called **protooncogenes**; the activated protooncogene is called **oncogene**.
2. By inactivation of the growth inhibiting genes (i.e. **antioncogenes**; also called *tumour suppressor genes*).

Certain growth promoting protein-products are encoded by the (activated) protooncogenes; examples include growth factors, the receptors for these growth factors and the components of the mitogen activated protein (MAP) *kinase* pathway. These protein-products can drive the cell into malignancy. Such mutations, causing activation of the protooncogenes, are called "**gain of function mutations**".

Growth inhibiting protein products are encoded by the anti-oncogenes. Cyclin-CDK inhibitors, the retinoblastoma protein pRb, and the p^{53} protein are some examples of inhibitory proteins. Mutations in antioncogenes, decrease production of these growth inhibiting substances, leading to malignant transformation. Therefore, these mutations are referred to as the "**loss of function mutations**".

Malignant transformation can be caused by activation of the genes that code for growth-promoting proteins; or by inactivation of genes that code for growth-inhibiting products.

Malignant transformation occurs even if only single copy of a growth-promoting protooncogene becomes activated (Fig. 32.1). The inactivation of an antioncogene, on the other hand, is effective in causing cancerous transformation only if both copies of the gene are inactivated: a single active gene can still produce enough of the tumour-suppressing protein products to permit normal growth. Most of the cancers result due to a combination of several mutations that include both *activated protooncogenes* and the *inactivated antioncogenes*. A combined effect of these two types of mutations mostly drives the cell into malignancy.

How does the Protooncogene Activation Cause Malignancy

Protooncogenes are important regulatory genes present in normal cells that play a key role in cell growth and division. Their expression is meticulously controlled. It is only when the protein product encoded by the proto-oncogene is required by the host cell, that these genes are expressed. Excessive production of such protein products provides excessive stimulation to the host cells, leading to their malignant transformation.

The mutationally activated form of protooncogene (called **oncogene**) is also known to cause malignancy. It is generated by somatic mutations, of which the commonest are point mutations. Others are gene amplification, gene translocation, truncation, etc. These changes in cellular genome, cause activation of a normal (proto-onco-) gene to abnormal and overactive (onco-) gene, which stimulates cellular growth and proliferation.

Thus, protooncogene-activation may cause malignancy in two ways:

1. By causing excessive and uncontrolled synthesis of a *structurally normal gene product.* (This usually follows a mutation in regulatory site of the gene).
2. By creating an *abnormal protein product* (**oncoprotein**) which is "superactive", i.e. it causes rapid stimulation of the cell growth and proliferation.

Excessive amount of the normal-gene-product as also the overactive, oncogenic-gene-product provide excessive stimulation for cell growth, thus driving the cell into malignancy.

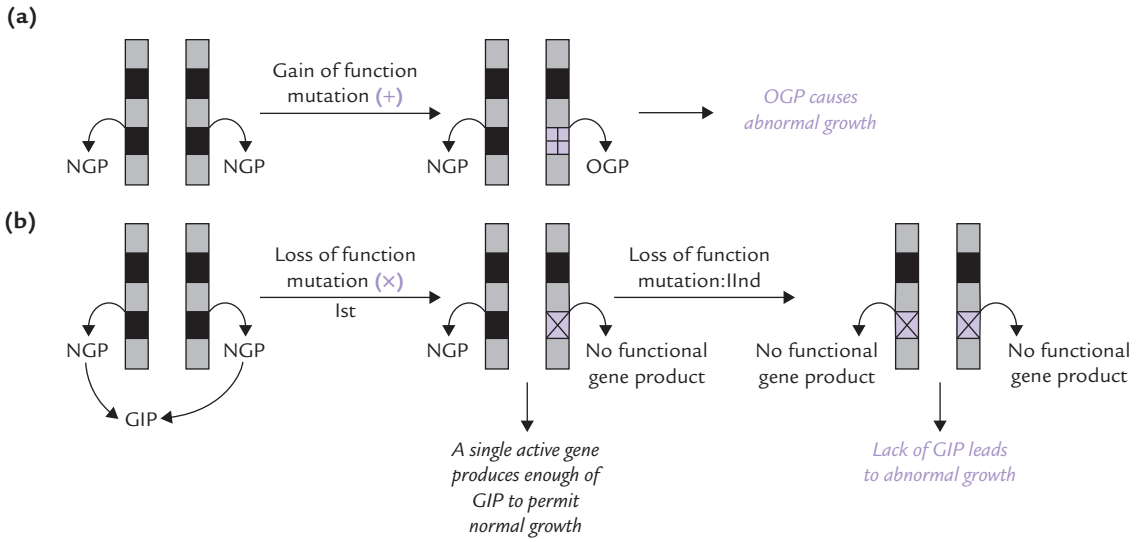


Fig. 32.1. Activation of growth promoting genes (proto-oncogenes) or inactivation of growth inhibiting genes (anti-oncogenes) cause malignant transformations, **(a)** A single mutation in protooncogene can cause malignant transformation, **(b)** In antioncogenes, a single mutation is insufficient to drive the cell into malignancy (NGP = normal gene product, OGP = oncogenic gene product, GIP = growth inhibiting product).

More than 80 protooncogenes are presently known, each one located on a specific chromosome. Some of these are listed in Table 32.1.

Antioncogene Inactivation – More Common Cause of Malignant Transformation

Antioncogenes offer protection against cancer since they code for proteins having inhibitory effect on cell division (Table 32.2). Mutation of antioncogenes, therefore, results in the loss of inhibition on cell division–a hallmark of cancerous growth (Fig. 32.1). In fact, *inactivation of antioncogenes is more common than the formation of active oncogenes in spontaneous cancers.*

The inactivating mutation can be a missense mutation, nonsense mutation, splice-site mutation, or a large deletion.

Two tumour **suppressor genes**, discussed as follows, have key roles in determining cell cycle progression and apoptosis, and are the *genes most frequently mutated in cancer cells.*

RB1 Gene

This antioncogene, also called **retinoblastoma gene** encodes a nuclear phosphoprotein with a molecular weight of 105,000. The protein, termed $p110^{RB1}$, acts specifically in retinoblasts. It binds to DNA and inhibits the expression of some growth-promoting genes. This inhibits cell division. Deletion or loss-function mutations of the *Rb* gene removes this inhibitory effect in the neural progenitor cells in the retina, resulting in their cancerous transformation.

p53 Gene

It encodes a cell cycle regulator protein of 375 amino acids, with a molecular weight of 53,000. *The p^{53} gene is expressed in almost all cells, and mutations in it are the most common genetic changes in spontaneous cancers.* It plays a vital role during the cell cycle just before replication. It monitors for any signs of DNA damage, and if damage is detected, it causes arrest of the cell cycle in the G_1-phase. This gives adequate time for repair of the damaged DNA, before the S-phase of the cell cycle begins. Alternatively, if the damage cannot be repaired in this time, it causes programmed cell death (apoptosis) of the cell. Thus, this gene plays a key role, not only in regulating G_1/S-phase cell cycle progression, but also in monitoring DNA damage.

Individuals with inherited mutations in the p^{53} gene are predisposed to develop a wide range of tumours, including sarcomas, carcinomas of the lung, breast, larynx and colon; brain tumours; and leukaemias. This syndrome, called **Li Fraumeni syndrome**, is associated with a high incidence of various cancers at an early age in the affected families.

The vast majority of cancers arise due to somatic mutations. The two tumour suppressor proteins, p^{53} and Rb, which have key roles in determining cell-cycle progression and apoptosis, are the genes most frequently disrupted in cancer cells.

B. Oncogenic Viruses

Some viruses, called *oncogenic viruses*, are capable of causing cancer by introducing a cancer-producing gene in the host cell genome (Fig. 32.2). Rous (Nobel Prize, 1966) was the first scientist to prove the role of viruses in aetiology of cancer by demonstrating that the cell-free filtrates from certain chicken sarcomas induced new sarcomas in chickens.

Viruses are nucleoproteins; their genomic material comprises either DNA (i.e. *DNA viruses*) or RNA (i.e. *retroviruses*). Certain viruses of either type are capable of producing cancer in the host cell. A selected list of cancer-causing viruses is given in Table 32.3.

Oncogenic Retroviruses

Genomic material of these viruses is RNA. It directs synthesis of complementary DNA, known as **retroviral**

Table 32.1. Some cellular protooncogenes

Oncogene	Location	Oncogene product	Subcellular
erb-B	Chromosome 7	Receptor for epidermal growth factor	M
src	Chromosome 20	*Tyrosine kinase*	M
sis	Chromosome 22	Platelet derived growth factor (PDGF)	M
alb	Chromosome 9	*Tyrosine kinase*	M
ras	Chromosome 12	P21; *GTPase*	C
myc	Chromosome 8	DNA-binding protein	N

M = membrane, N = nucleus, C = cytoplasm.

Table 32.2. Some important antioncogenes

Inherited disease	Location	Antioncogene	Encoded protein
Retinoblastoma	Chromosome 13q	*Rb*	Nuclear phosphoprotein
Wilm's tumour	Chromosome 11p	*WT-1*	Nuclear protein
Li-Fraumeni syndrome	Chromosome 17p	P^{53}	Phosphoprotein

p = short arm: q = long arm.

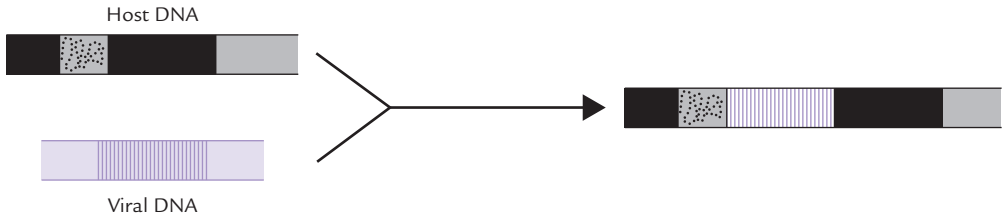


Fig. 32.2. Integration of viral gene into the host cell DNA.

Table 32.3. A selected list of oncogenic viruses

Class	Examples	Associated cancer
Retro viruses	HTLV and-II	Leukaemia, lymphoma
	Hepatitis C virus	Hepatocellular carcinoma
DNA-viruses	Human papilloma virus	Genital warts
	Epstein–Barr virus (EBV)	Cervical carcinomas, Burkitt's lymphoma, nasopharyngeal cancer
	Hepatitis B virus	Hepatocellular carcinoma
	Human Herpes virus-I	Oral cancer
	Human Herpes virus-II	Cervical carcinomas

DNA, by the enzyme *reverse transcriptase.* The retroviral DNA gets incorporated into the host cell genome, (and is now known as **provirus**). It fosters cancerous transformation of the host cell by either of the two mechanisms discussed here.

1. The retroviral DNA is expressed at a high rate under the directions of viral promoter and enhancer in the long terminal repeats (LTRs). This results in overexpression of the protein encoded by the viral oncogene (Fig. 32.3a). The excessive production of the viral oncoprotein fosters malignant transformation of the cell.
2. A second route to retroviral transformation arises when a virion with LTR is inserted into the host genome next to a cellular protooncogene. Transcription of the protooncogene is boosted up by the LTR (Fig. 32.3b). Thus, viral oncogene-insertion increases production of the protein product encoded by the neighbouring protooncogene. Excessive production of such protein, having growth enhancing action, results in uncontrolled cell division, and hence a cancerous growth follows.

Some retroviruses cause cancer by introducing an oncogene into the host cell genome or by inserting themselves next to a cellular protooncogene.

Viral Oncogenes Resemble Cellular Protooncogenes (Table 32.4)

In case of Rous sarcoma virus, the viral *src* oncogene (*v-src*) resembles a cellular oncogene (*c-src*) present in the normal cell (i.e. protooncogene). Apparently, the virus picked up this gene accidentally during a previous infection cycle. The *v-src* oncogene codes for a protein that possesses *tyrosine kinase* activity. This protein phosphorylates several intracellular proteins, which in turn are involved in mitogenic signalling. Thus, production of the *kinase* protein by the *v-src* provides a potent stimulus for cell proliferation, which ultimately results in cancer.

Several retroviruses carry viral equivalents of cellular genes. Their oncogene product(s) foster cancerous transformation of the infected cell (Table 32.4).

DNA Oncogenic Viruses

Several groups of DNA viruses have been recognized which induce development of human cancers. In contrast to the retroviruses, however, the DNA viruses neither carry viral equivalent of cellular genes (*v-onc*) nor the LTRs. Instead their genomes encode proteins that are able to interact with, and inactivate, the regulatory proteins that control the cell cycle. The genomic material of these viruses gets incorporated in the host cell genome (Fig. 32.2). *The viral genome is capable of overriding the regulatory checks and balances of the host cell and undergoes rapid replication.* Along with it the rate of replication of the host cell genome is also greatly enhanced, being boosted up by the protein product encoded by the viral genome.

Some examples of DNA oncogenic viruses: These include human papilloma virus, Epstein-Barr virus (EBV) and hepatitis B virus (Table 32.3). Human papilloma virus causes genital and skin warts, and has a possible role in development of cancer cervix. Hepatitis B virus (HBV) is involved in development of hepatocellular carcinoma. Epstein–Barr virus is a herpes virus, which has been implicated in aetiology of Burkitt's lymphoma. This cancer develops in children suffering from chronic malaria or in a transplant recipients.

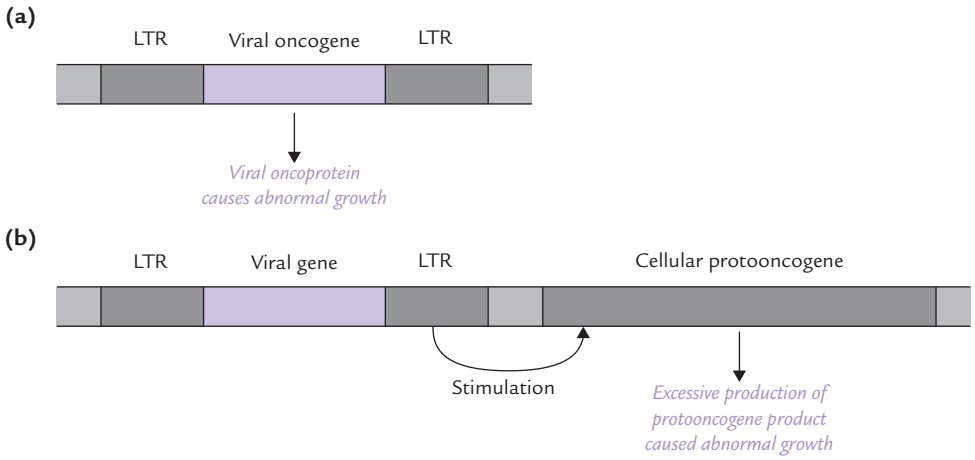


Fig. 32.3. Cancerous transformation of host cells following insertion of virion with long terminal repeats (LTR). **(a)** Excessive production of viral oncogene product causes abnormal growth, **(b)** The integrated retrovirus induces overproduction of the cellular protooncogene product which then leads to abnormal growth.

Table 32.4. Resemblance of viral oncogenes with certain cellular oncogenes

Cellular oncogenes	Virus carrying the gene	Oncogene product	Tumour (species)
src	Rous sarcoma virus	Non-receptor *tyrosine kinase*	Sarcoma (chicken)
erb-A	Erythroblastosis virus (chicken)	Transforming growth factor receptor *(TGF-α-R) kinase*	Erythroblastosis (chicken)
erb-B	Erythroblastosis virus (chicken)	Epidermal growth factor receptor	Erythroblastosis (chicken)
ras	Rat sarcoma virus	*GTPase*	Sarcoma, leukaemia (rat)

C. Chemical Carcinogenesis

Certain chemical agents, called chemical carcinogens, react with host cell DNA to induce mutations. Some of the mutant cells may die, some become normal through repair of mutation, but a few may be induced to undergo rapid multiplication. The latter action (i.e. proliferation-enhancing action) of these chemicals may result in rapid, uncontrolled cell division—a hallmark of malignant transformation.

Some of the carcinogenic chemical agents are enlisted in Figure 21.18. These may be organic or inorganic, and account for about 80% of all human cancers. They may enter the human body through diet, drugs or may be related to occupation (e.g. asbestos), or lifestyle (tobacco).

D. Radiation Energy

X-Rays, gamma rays and electromagnetic radiations are mutagenic in nature, causing cancers. Similarly, particulate radiations (α-particles, β-particles, protons and neutrons) are also capable of inducing cancer. These radiations can directly damage the DNA structure, or may produce free radicals, which in turn cause damage to DNA.

Mutations of normal cellular genes that control growth underlie development of all cancers. Various physical, chemical and environmental agents that are implicated in aetiology of cancer cause such mutations.

IV. Molecular Basis of Protooncogene Activation

A central event in carcinogenesis is the protooncogene activation. This section deals with various ways by which the activation may occur.

A. Point Mutation of Protooncogene

Certain protooncogenes, for instance the *ras* oncogene, are activated by point mutation (Fig. 32.4a). The mutated *ras* oncogene produces a protein (*GTPase*) in which amino acid is substituted at the 12th position. This alteration greatly diminishes the *GTPase* activity. Because this activity plays a key role in control of all growth, its loss manifests as excessive proliferation of cell.

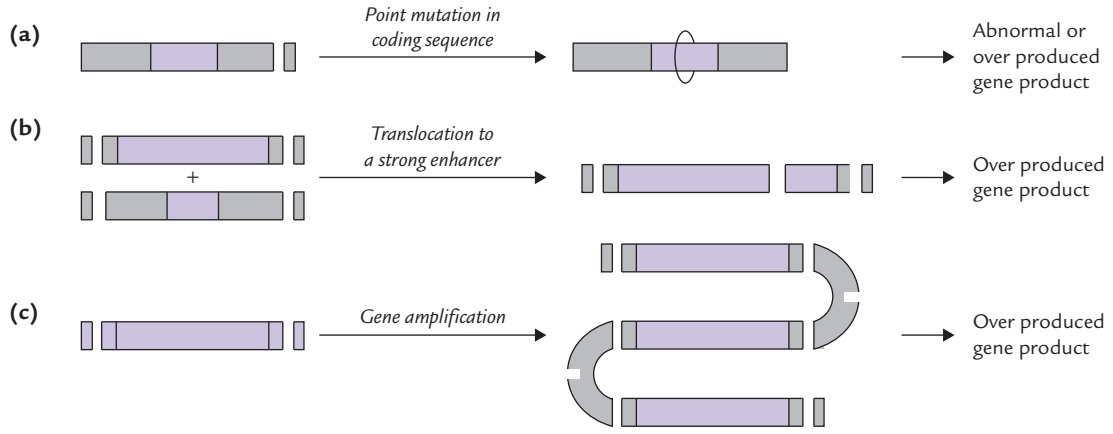


Fig. 32.4. Common mechanisms for oncogenic activation of cellular protooncogene.

B. Gene Rearrangements

A common type of gene rearrangement is chromosomal translocation which involves splitting of a piece of chromosome and joining it to some other chromosome. In this way, a gene is translocated to a transcriptionally active site where it is under the influence of some strong enhancer or promoter sequence (Fig. 32.4b). The translocated gene is overexpressed at the new site.

Chromosomal translocation has been described in the following conditions:

Burkitt's Lymphoma

The *c-myc* gene locus on chromosome 8 is split off and joined to chromosome 14. At this site the *c-myc* gene is rapidly expressed under the influence of the local enhancer sequence. Consequently, the protein product of the *c-myc* gene is produced in large amounts, which makes the cell malignant.

Leukaemias

Gene rearrangements have been described in chronic myelogenous leukaemia and acute lymphoblastic leukaemia. In these diseases, recombination between the chromosome 9 and the chromosome 22 occurs with the following consequences. Shortening (or deletion) of the short arms of the chromosome 22 occurs (due to translocation of the genetic material to the chromosome 9). The abnormal chromosome 22 so produced is called **philadelphia (Ph') chromosome**, which is seen in 80–85% patients of chronic myelogenous leukaemia (CML) and in 5–20% patients of acute lymphoblastic leukaemia (ALL). The abnormality persists throughout the course of the disease (in remission as well as in relapse), and is generally unaffected by therapy. In about 20% cases of CML, there is translocation of chromosomes 9–22, which results in activation of *c-alb* present on the chromosome 9.

Non-Hodgkins Lymphoma

Gene rearrangements have been implicated in development of non-Hodgkin's lymphoma also, where translocation of Chromosomes 14 to 18 inactivates *bcl-2* oncogene. Since *bcl-2* product suppresses apoptosis (programmed cell death), inactivation of *bcl-2* leads to tumour formation.

C. Gene Amplification

Gene amplification may cause overproduction of certain proteins because of several fold amplification of certain DNA sequences (Fig. 32.4c). Gene amplification is frequently seen in cancer cells, when the patient is treated with methotrexate. This drug inhibits the enzyme *dihydrofolate reductase* (DFR). In an attempt to develop resistance against this drug, the gene for the DFR gets amplified, resulting in increased production of the enzyme (up to 400-fold).

D. Promoter Insertion

The retroviral DNA is inserted immediately upstream of the cellular protooncogene and enhances its expression as shown in Figure 32.5a. The retroviral DNA has two identical promoter regions contained in the long terminal repeats (LTRs). The promoter located in the downstream LTR is used for enhancing the transcription of protooncogene. The transcription occurs at an abnormally high rate, without any negative controls that normally regulate the rate of transcription.

E. Enhancer Insertion

The viral enhancer (in the LTR of retrovirus) may also stimulate transcription of the neighbouring protooncogene

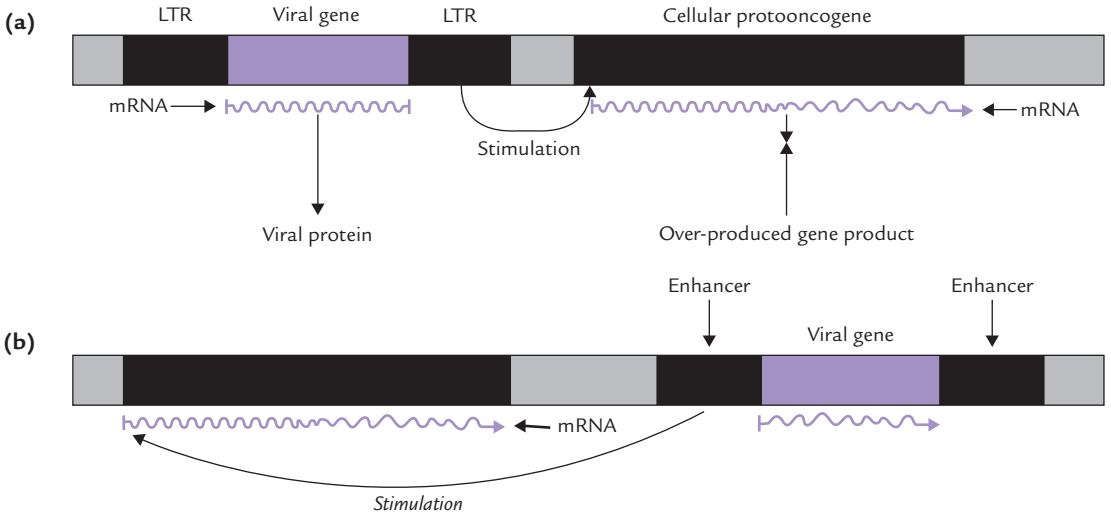


Fig. 32.5. Activation of protooncogene by integrated retrovirus. Mechanisms of activation shown are: **(a)** promoter insertion **(b)** enhancer insertion.

(Fig. 32.5b). The viral enhancer differs from the promoter as it can act over considerable distances; in some cases it is located 10,000 base pairs away from the protooncogene. The stimulation occurs even if the viral enhancer is:

- inserted downstream of the protooncogene
- inserted with opposite polarity (i.e. wrong 5′→3′ orientation).

Thus, the viral enhancers have similar actions as the normal cellular enhancers (Chapter 24).

A retrovirus that dose not carry an oncogene can cause cancer inserting next to a cellular protooncogene and activating it by:

(a) promoter insertion—The promoter in the downstream LTR stimulates transcription.

(b) Enhancer insertion—The enhancer stimulates transcription even if it is inserted downstream of the protooncogene or if inserted with opposite polarity.

V. Some Oncoproteins and Associated Cancers

The protein products of oncogenes are involved in transmission of mitogenic signals, thereby causing transformation and multiplication of cells. They can be divided into four categories: **growth factors**, **growth-factor receptors**, **GTP-binding protein** and **non-receptor *tyrosine-kinase***.

A. Growth Factors

Several growth factors that allow the cell to respond to signals from distant sources are known. Most growth factors stimulate cell proliferation, but some others cause their target cell to migrate, to differentiate, or to grow to a large size without dividing. Growth factors even can induce opposite effects in different cell types causing some cells to proliferate and others to stop growing, or to undergo terminal differentiation. Some of them act on a wide variety of cell types; others have narrow specificities (Table 32.5).

Growth factors transmit the signal across the plasma membrane to the interior of the cell, thereby regulating cell division. They act by binding to a protein receptor on the plasma membrane, which contains *tyrosine kinase* domain within its covalent structure. The above binding activates the receptor by phosphorylation, triggering a signalling cascade that leads to phosphorylation of intracellular transcription factors. The latter in turn regulate expression of genes that affect cell growth and proliferation.

Insulin-like growth factor (IGF-1), for example, is released from the liver in response to growth hormone. The growth promoting effects of growth hormone is mediated, in part, by IGF-1. Pygmies are said to have just as much growth hormone as taller people, but they release less of IGF-1.

B. Growth Factor Receptors

Growth factors bind to the extracellular domains of transmembrane receptors that have *tyrosine kinase* domains present within their intracellular domain. These *receptor-tyrosine-kinases* are monomeric and enzymatically inactive in the absence of the growth factor. The binding of the corresponding growth factor causes receptor to dimerize and undergo autophosphorylation and activation (Fig. 32.6).

The phosphotyrosine residues in turn cause phosphorylation of cytoplasmic proteins, which act as intracellular

Table 32.5. Some growth factors and their functions

Growth factor	Source(s)	Location	Functions
Platelet-derived growth factor (PDGE)	α-Granules of platelets	Chromosome 5	Acts on a large number of cell types, including fibroblasts and smooth muscles; participates in wound healing
Epidermal growth factor (EGF)	Salivary glands, and a wide variety of cells	Chromosome 7	Acts primarily in its tissues of origin. Most active on epithelial cells
Fibroblast growth factor	Fibroblasts	—	Acts mostly in tissues in which it is synthesized. Stimulates fibroblasts as well as many other cells
Insulin-like growth factor-1 (IGF-1)	Liver in response to growth hormone	Chromosome 15	Mediates growth promoting effect of growth hormone; stimulates incorporation of sulphates into cartilage; exerts insulin-like action on certain cells
Nerve growth factor (NGF)	Salivary gland	Chromosome 1	Stimulates growth and differentiation of post-ganglionic sympathetic neurons
Tumour necrosis factor (α)	Monocyte	Chromosome 6	Necrosis of tumour cells, promotes leukocyte proliferation
Transforming growth factor	Tumour cells, placenta	Chromosomes 5, and 6	Functions similar to epidermal growth factor

messenger and stimulate cell division by still unknown mechanism.

Over-expression and/or structural alterations in growth factor receptors are observed in certain cancers. For instance, the over expression of the *erb B* gene (encodes epidermal growth factor; EGF) is associated with lung cancer.

C. GTP-binding Proteins

These signal transducing proteins transfer the signal beyond *tyrosine kinase*. They have been detected in about 30% of the human cancers, and are believed to play a major role in cell growth and differentiation. An example is ras *protein* (or p^{21}), encoded by *ras-gene*. It is a small *GTPase* (MW 21000), which controls cell division, but when mutated, leads to cancer, as discussed here.

The ras-protein cycles between an active GTP-bound form and an inactive GDP-bound form (Fig. 32.6). Upon stimulation by growth factors, the inactive form is activated by exchange of GTP for GDP. The exchange is facilitated by a *guanine nucleotide exchange factor* (GEF), which opens a nucleotide-binding site, allowing the GDP to escape and GTP to enter in its place. The activated ras-protein causes activation of effectors (*protein kinases* and others), which phosphorylate specific targets that promote cell growth and differentiation. This effect is, however, transient only because the activated ras gets inactivated quickly. This is because of intrinsic *GTPase* activity of the ras-protein. This activity serves to terminate the signal and return the activated rasGTP to its original inactivated state (ras-GDP). Although the activity is slow, it is augmented by helper protein termed *GTPase activity promoter* (GAP).

Point mutation in *ras* gene results in the production of altered ras-protein, which **loses its *GTPase* activity**. This leads to *"locking" of the protein in its activated form* (i.e. persistent activation), causing uncontrolled cell multiplication.

D. Non-receptor Kinases

These proteins are located on the interior of the inner plasma membrane. They respond to external growth stimuli and cause phosphorylation of certain cytoplasmic proteins, which are involved in cell growth and differentiations. Mutations in the genes encoding these proteins cause increase in their *kinase* activity, and this in turn, increases phosphorylation of the cellular target protein, and this boosts up cell multiplication.

The proteins encoded by cellular oncogenes, such as growth factors, growth factor receptors, GTP-binding proteins, non-receptor *tyrosine kinases,* are all linked to the mitogenic signalling cascades, and hence implicated in cancer aetiology.

VI. Tumour Markers

A tumour marker is any substance that is produced by the tumour cell, and which is related to the presence or progress of a tumour.

It is known that metabolism of tumour cells differs from that of the normal progenitor cells. *Certain genes that are active during fetal period, but are suppressed in adult cells, are reactivated in the tumour cells.* Therefore, tumour cells synthesize certain abnormal proteins, such as surface

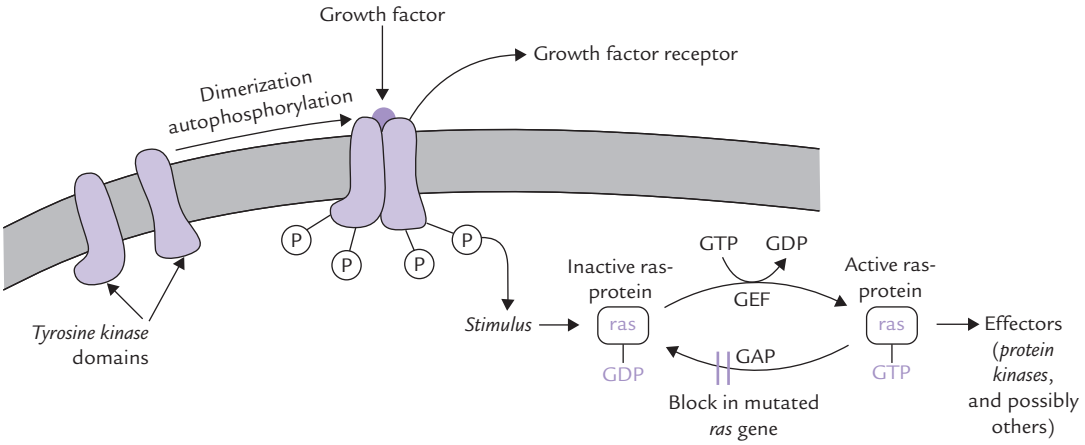


Fig. 32.6. Activation of ras-protein by growth factors. The sequence of events is: binding of GF → Dimerization of stimulated GF receptor → Receptor autophosphorylation → Activation of ras (stimulated by GEF) → Hydrolysis of the bound GTP by ras (stimulated by GAP) (GEF = guanine nucleotide exchange factor, GAP = *GTPase* activity promoter).

antigens, cytoplasmic proteins or enzymes, measurement of which (in plasma or serum) provides a useful marker for supporting cancer-diagnosis and follow up. Thus, these proteins serve as **biochemical markers of cancer**, and accordingly referred to as tumour markers.

The ***α-fetoprotein*** (AFP), for example, is secreted by fetal liver cells but the adult liver loses capability to synthesize this glycoprotein. In patients with hepatocellular carcinoma, the liver cells synthesize large amount of AFP and release it into the serum. Therefore, serum AFP levels increase markedly (Table 32.6). In malignant testicular teratoma also, elevation of AFP has been reported in about 90% of the patients. Thus, AFP estimation is a useful indicator of hepatic and testicular malignancies.

Characteristics of a Required Tumour Marker

1. Be clearly related to the malignant process.
2. Provide information on the type and location of the tumour and its stage.
3. Be of value in prognosis. For this the concentration of tumour a marker should correlate with tumour mass.
4. Be useful for monitoring treatment: the decline in concentration of a marker is an indication of the success of the treatment (surgery, chemotherapy, radiotherapy, or a combination of these).
5. Be useful in assessing follow up.

However, in reality most tumour markers have certain limitations.

Tumour marker, the substance related to presence or progress of a tumour, may be hormone, enzyme, or oncofetal protein. Tumour markers are of most value in monitoring treatment and assessing follow up, but are also used in diagnosis, prognosis and screening for the presence of a disease.

Table 32.6. Some important tumour markers

Cancer marker	Clinical utility
Tumour antigens	
α-Fetoprotein	Increases in tumours of liver and germ cells
Carcinoembryonic antigen	Increases in cancers of colon, pancreas, lung and breast
Enzymes	
Acid phosphatase	Increases in carcinoma prostate
Alkaline phosphatase	Increases in metastasis of liver, bone and carcinoma of head of pancreas
β-Glucuronidase (urinary)	Increases in cancers of pancreas and urinary bladder
Hormones	
Human chorionic gonadotropin	Increases in choriocarcinoma
Calcitonin	Increases in medullary carcinoma of thyroid
Others	
Prostate specific antigen (PSA)	Increases in prostate cancer
Paraprotein	Appears in multiple myeloma

A host of tumour markers has been described, which fall into one of the several groups: they may be **tumour antigens**, e.g. carcinoembryonic antigen (CEA) in colorectal carcinoma; or **hormones**, e.g. calcitonin secreted in medullary carcinoma of thyroid; or **enzymes**, e.g. *acid phosphatase* in carcinoma prostate. Some important ones are described here.

Carcinoembryonic antigen (CEA): Measurement of serum level of CEA, a complex glycoprotein (MW 180,000)

produced from embryonic tissues of liver and intestine, has been found useful in colorectal cancer. In about 65% of patients with this cancer, CEA can be detected in serum. In advanced stages of the disease, especially when the metastatic spread involves the liver, about 90–100% of the patients have elevated CEA levels.

However, lack of sensitivity and specificity limit utility of CEA for screening colorectal cancer. Elevated levels are also found in gastric, bronchial and ovarian carcinomas, as also in a number of non-malignant conditions, particularly liver diseases of various types, pancreatitis, and small bowel diseases. Heavy smokers (57%) also have high levels of CEA. Thus, CEA is not a reliable indicator of cancer detection and may lead to diagnostic ambiguities. Its utility in follow up of the known cases of colorectal cancer is also not established. The fact that it does not relate to the mass of the tumour limits its utility in monitoring the response to treatment.

α-Fetoprotein (AFP): It is glycoprotein (MW 7000) that is synthesized in yolk sac in early fetal life. Normal serum level is 12 U/ml. It helps in the diagnosis of liver and germline carcinomas along with indicating the presence of spina bifida, amencepalia and atresia of oesophagus.

Human chorionic gonadotropin (hCG): Normal serum level is 5 IU/ml and it is useful in mammary carcinoma and trophoblastic tumours.

Prostate specific antigen (PSA): It is a 32 HD glycoprotein, produced by secretory epithelium of prostate gland. It is secreted into seminal fluid where it exists either in a free form, or complexed with α_1-antitrypsin. PSA has *protease* activity, and is necessary for liquefaction of the seminal coagulum. The PSA level (especially of complexed form) is *elevated in prostate cancers*.

Presently, definition and scope of tumour markers has been expanded, and it includes not only those substances that are secreted in blood, but also those found on the surface of or within the cells. Such markers are not necessarily unique products of the malignant cells, but may simply be expressed by the tumours in greater amount than by normal cells.

Exercises

Essay type questions

1. Discuss some differences between malignant cell and the normal progenitor cell. Name some oncoproteins and their role in fostering malignant transformation of normal cells.
2. What are protooncogenes? Explain various mechanisms by which they are activated to cancer-producing oncogenes.

Write short notes on

1. Tumour markers
2. Oncogenes
3. Growth factors in cancer aetiology
4. Oncogenic retroviruses

CHAPTER

IMMUNOLOGY

33

Every living organism is confronted by continual assaults from its environment and, therefore, must develop some defense system to survive. The defense ranges from physical barriers, such as skin, to highly sophisticated specific immune system. The latter is a remarkably adaptive system in vertebrates, which protects against invasions by an apparently limitless variety of environmental agents that are foreign to the body (**antigens**) and that may disturb the body-homeostasis. These agents may be micro-organisms or their products, foods, chemicals, drugs, animal hair, etc. Immune system has a highly complex organization, comprising of cells, tissues, and molecules, which are involved in specific-recognition and interaction with the antigen followed by its inactivation/elimination from body. The immune system offers protection against apparently limitless types of antigens, especially proteins and polysaccharides, but not directly against small molecules **(Haptens)**, although such molecules may elicit an immune response if they are bound to a protein of the body.

Two essential functional elements of the immune system are recognition and response. The system effectively recognizes subtle chemical changes that discriminate foreign cells (**nonself**) from body's own cells and proteins (**self**) and mounts an effective response against the nonself by enlisting participation of a large number of cells, soluble factors and other molecules. The response, known as the **effector function of immune system**, eliminates the activating antigen with incredible precision by either cellular or humoral mechanisms. In the cellular mechanism, lymphocytes (called **T cells**) sensitized to specific antigen, target the antigen directly in a process called **cell-mediated response.** In addition to causing cytotoxic destruction of virus infected cells and bacteria in this manner, the T cells also cause activation of other cells of the immune system. In the **humoral mechanism**, a different type of lymphocytes (called **B cells**) manufacture specialized proteins (i.e. **immunoglobulins** or antibodies), which protect against organisms by a variety of mechanisms. These include neutralization of toxins, lysis of bacteria in the presence of complement proteins, opsonization of bacteria to facilitate phagocytosis, and preventing adherence of bacteria and viruses to cell surfaces.

In the present chapter an account is given of molecular and cellular biology of the immune system and the current concepts about the mechanisms by which it protects the body against the invading pathogens. Details have been simplified to reveal the essential structure and physiology of the immune system, in order that the clinical application of immunology may be appreciated.

By the end of this chapter the student should learn.

- ***Organization of immune system****:* the organs, cells and molecules involved in generating innate and specific immune responses.
- ***Antigens:*** Epitopes, haptens; antigen processing and presentation.
- ***Immune responses****:* humoral and cell mediated, role of B cells, accessory cells, and various populations of T cells in the generated immune responses.
- ***Antibodies***, complement, hybridoma and monoclonal antibodies.
- Vaccines, immunological techniques, immunodysregulation disorders.

I. Historical Milestones

Immunity is the body's ability to resist certain harmful substances; the Latin term *immunis* meaning "exempt" being the source of the English word immunity. The discipline of immunology, the science of man's resistance to disease, grew out of the observation that a person could catch certain infectious diseases only once in lifetime. It had been a common knowledge that the individuals who had recovered from certain infectious diseases were thereafter protected from the same disease. The concept of immunity was expressed in folklores dating back to 430 B.C., as evidenced by Thucydides' account of the *Peloponnesian War*. In this he noted that those who got infected with plague and survived, could safely nurse the sick without the risk of contracting the disease again. Even if they caught the disease a second time, the second attack was never fatal. In ancient times, Indians and Chinese were known to obtain protection against smallpox by inoculating live organisms from the diseased pustules.

However, such observations and concepts were translated successfully into a medically effective practice only in recent times. In 1718, **Lady Wortley Montagu**, a renowned English author and wife of the British Ambassador to Turkey, reported that the Turks induced immunity against smallpox by administering the crusts derived from smallpox pustules through small cuts in skin or into nostrils (the technique termed *variolation*). She tried to introduce this technique into England in 1718, but the method was considered too dangerous. **Edward Jenner** (1798), a British physician, improved this technique and discovered means of preventing smallpox. Jenner intentionally caused a mild cowpox infection in James Phipps, a healthy 8-year-old boy, in order to protect him against smallpox. Phipps developed cowpox, but when Jenner intentionally infected him with the smallpox matter, the boy did not develop smallpox. Thus, inoculation with the cowpox matter protected against smallpox.

Nearly a hundred years later, **Louis Pasteur** appreciated the role of vaccination (as the Frenchman called the procedure) in protecting against infectious diseases. He recognized that aging impairs virulence of pathogens, and that such an avirulent strain might be administered to protect against the disease. This followed his chance observation that the infectious agent in cholera of chickens (now known as *Pasteurella antiseptica*) became avirulent on ageing, but this avirulent form protected the chickens against subsequent exposure to fresh virulent preparations. This led Pasteur to realize that *it was possible to immunize individuals by using inactivated organisms*. Such organisms retained their ability to provoke an immune response (they were ***immunogenic***) while failing to induce the disease (they were **not *pathogenic***). He called this avirulent strain of organisms a vaccine (from Latin *vacca*, which means cow) in honour of Jenner's pioneering contribution with cowpox inoculation. These findings were subsequently extended to other diseases and demonstrated that *introduction of dead or weakened (i.e. attenuated) viruses or bacteria, or their toxins, into the body developed resistance to diseases*.

Discovery of Two Types of Immune Responses

Humoral and cellular: The first insight into the mechanism of immunity was provided by **Emil von Behring** (Nobel Prize in 1890) who recognized antibodies to diphtheria toxin in serum. The immune response mediated by the soluble antibodies present in serum was called the **humoral response** (the body fluids used to be called *humours*). The other type of immune response, the **cell-mediated response,** was elucidated around this time only from discovery of Elie Metchnikoff in 1883 who hypothesized that *cells, rather than antibodies, were responsible for the immune response*. These cells attacked immunogens directly rather than synthesizing special protective proteins (i.e. immunoglobulins).

Subsequently it was shown that inter-related roles of both types—humoral and cellular activities—were essential for development of effective immune response. *Lymphocytes were identified as the cells playing a major role in both types of immune responses.*

Selective and instructional theories: Several theories were proposed to account for the distinct specificity of antibody molecule against the non-self. The **Selective theory** (Paul Ehrlich, 1900) proposed that the antigen is selected and bound to preformed cell surface antibody having a complementary structure. This in turn would enhance the cellular production of this antibody. The selective theory further suggested that specificity of the antibody had been determined prior to antigen exposure; the role of antigen appeared merely to select the appropriate antibody and induce its proliferation.

The **instructional theory** suggested that the major role in determining the specificity of the antibody molecule was played by antigens. A portion of a particular antigen was suggested to serve as a template, and an antibody complementary to the antigen was formed. Thus, *this theory envisages antigen to be instructing synthesis of the antibody.* In the 1930s and the 1940s various forms of instructional theories were proposed. However, in the 1950s attention was once again reverted to selection theories and currently a refined form of it, called **Clonal Selection theory,** is widely accepted in modern immunology. This theory will be dealt with in more details later in the chapter.

II. Non-specific and Specific Immune Responses

The immune responses include specific and non-specific components. **Non-specific immunity**, sometimes referred to as ***innate, inborn, unchanging*** or ***natural*** derives from all those elements with which an individual is born and that are always present and available at short notice. The **specific immunity** (also called acquired or adaptive) occurs after exposure to an agent and is mediated by antibodies and by lymphoid cells. It came late in evolutionary terms and is present only in vertebrates.

Immune response to nonself elements, the antigens, involves a number of components. Some of these show unique specificity for the particular stimulating antigen(s) and comprise the adaptive immune response.

A joint operation of innate and adaptive responses provides an excellent all round protection against a variety of challenges from the environment.

A. Non-specific (Innate) Immune Response

It is considered more primitive than the specific immune response and can be envisioned as the basic state of resistance or the first-line of defense that a species possesses. It is *innate* in that it is not affected by any prior contact with the infectious agent although its activity can be up- and down-regulated by a number of factors. It mainly uses **mechanical barriers** to infectious agents, e.g. skin and associated secreted products—mucus, tears, etc. It also involves **physiological barriers**, e.g. temperature and pH. *These purely physical barriers* to *infection are supplemented by certain cells* (e.g. natural killer cells), *certain proteins (e.g. the complement cascade and interferons) and other processes such as phagocytosis and inflammation (Table 33.1).*

Some of the important factors enlisted in Table 33.1 are discussed here.

1. ***Temperature and pH***: Many species display innate immunity against certain infections simply because their body temperature inhibits survival of the invading pathogen. Body temperature of chickens, for example, is too high to permit growth of anthrax bacteria, thereby providing an innate physiologic barrier to this infection. Low pH in our stomach is also highly unfavourable for pathogen growth.

2. ***Lysozyme***, an enzyme first described by *Alexander Fleming* and found in tears and mucous secretions, is capable of attacking and catalyzing the hydrolytic destruction of the peptidoglycan layer of the bacterial cell wall. Thus, *Fleming is credited with discovery of both endogenous and exogenous antibiotics* (lysozyme and penicillin respectively).

3. ***Phagocytosis and endocytosis:*** In phagocytosis the plasma membrane of the phagocytotic cell expands around the particulate material; the particulate material includes the whole pathogenic microorganisms. Intracellularly, the ingested particle is entrapped in phagocytic vacuole into which lysosomes release their enzymes, which then digest the particle into small breakdown products, e.g. peptides, nucleotides and sugars. *Phagocytosis differs from endocytosis in that the expansion of membrane requires participation of microfilaments, which do not take part in endocytosis.* Moreover, only specialized cells are capable of phagocytosis, whereas endocytosis is carried out by virtually all cells.

Phagocytosis is enhanced by a variety of factors that make the foreign cell an easy target. These factors, collectively called **opsonins,** consist of antibodies and components of complement system.

Specialized phagocytotic cells include tissue macrophages and granulocytes, which make up 60% of the leukocytes in the blood. In addition to using lysosomal enzymes, they kill the internalized invading pathogen by

Table 33.1. The innate immunity

Mechanism	Factor
1 Mechanical barrier	Keratin layer of intact skin, mucous
2 Physiological barriers	Low pH of stomach and vagina, body temperature
3 Cells	Natural killer cells
4 Proteins	Interferon (inhibit virus replication) Complement cascade Lysozymes in tears and other secretions
5 Processes	Phagocytosis, and endocytosis Inflammatory response Normal flora of throat, colon and vagina (inhibit colonization by pathogen)

1 and 2 restrict entry of pathogen, whereas 3, 4 and 5 inhibit growth and survival of pathogen.

use of cytotoxic free radical attack also. In the case of granulocytes, the oxidative attack on invading organisms is so vigorous that the cells themselves rarely survive, whereas macrophages can carry out many rounds of killing.

4. ***Inflammatory response:*** Inflammation is the body's response to injury or tissue damage, which is provoked by presence of invading pathogen. Its purpose is to limit and then repair the damage wrought by the injurious agent. The cardinal signs of inflammation are redness or ***rubor***, swelling or ***tumour***, heat or ***calor*** and pain or ***dolor***. These signs are due to increased blood flow, increased capillary permeability, and the escape of fluid and cells (macrophages and lymphocytes) from the engorged capillaries into the tissue spaces. The increased permeability is due to several chemical mediators, of which *histamine*, *prostaglandins and leukotrienes* are the most important. *Bradykinin* is an important mediator of pain.

The fluid that accumulates in tissues, called **exudate,** has high content of *serum proteins* which possess antibacterial properties. As part of the inflammatory process, *neutrophils* and *macrophages*, both of which arrive early, supplement the antibacterial action by engulfing the bacterial cells. Certain proteins, known collectively as the **acute-phase reactants,** are also produced early in acute inflammation (within 1–2 days). The best known of these is C-reactive protein, which is synthesized in the liver and is thought to play a role in activating the alternative pathway of complement activation by binding to the surface of bacteria.

The non-specific immune response can be envisioned as the basic state of resistance or the first-line defense. The cells and soluble mediators involved are primarily those associated with the processes of acute inflammation.

It is clear from the foregoing discussion that the so called more ancient and primitive innate response is still an important aspect of protection against infection in humans. Its impairment can lead to life-threatening infections (**Case 33.1**). Moreover, it freely interacts with the more recent, more specific adaptive immune response (described below) to provide an excellent protection against a variety of invading pathogens.

B. Specific Immune Response

If the non-specific defenses are unsuccessful, e.g. due to persistence of the triggering agent, the specific immune response is brought into play, which is capable of specifically recognizing and selectively eliminating foreign microorganisms and molecules. Unlike innate immunity, which is inborn, the acquired immunity is a consequence of an encounter with a foreign substance. The antigenicity (ability of the foreign substance to induce an immune response) is determined by one or more specific molecular groups on its structures, known as **antigenic determinants** (or **epitopes**). When lymphocytes (B- or T-cells) recognize a foreign substance (antigen), they respond by transforming into lymphoblasts and then into antibody secreting plasma cells; or into cytokine secreting helper-cells or activated cytotoxic cells, as depicted in Figure 33.1.

The second distinctive feature of specific immunity is that (unlike innate immunity) it displays ***specificity, diversity, memory*** and ***self/non-self recognition***. These four features characterize all immune responses:

1. ***Specificity*:** The immune system can distinguish subtle differences among antigens. A single base mutation, resulting in a single amino acid substitution, may be recognized effectively.
2. ***Diversity*:** The immune system is capable of generating a vast array of structurally different recognition molecules of almost any specificity, and therefore can respond to billions of antigens.

   ***Memory*:** The immune system employs a mechanism to remember the specific encounter so that if the same foreign (or nonself) antigen is encountered again it can be dealt with more quickly and effectively. This is achieved by induction of a *heightened state of immune reactivity in secondary encounter*.

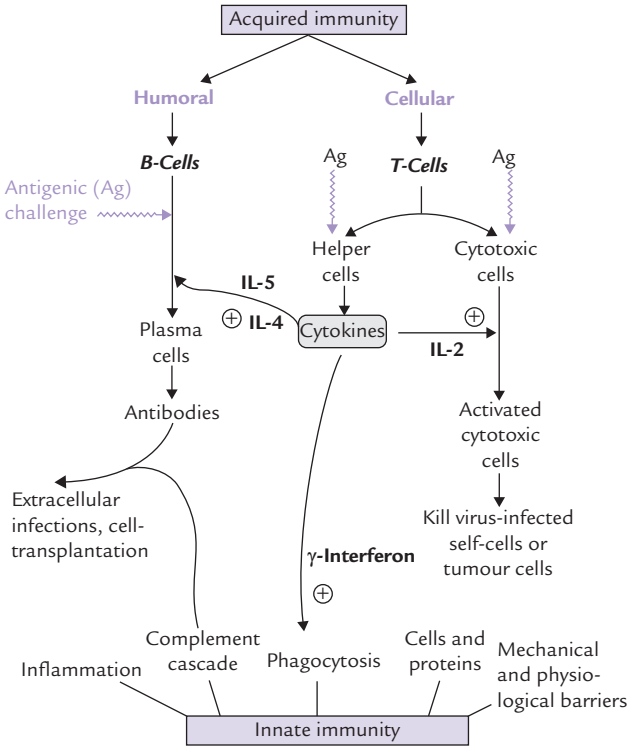


Fig. 33.1. Introduction to the interactions and functions of major components of immune system (IL = Interleukin).

3. ***Self/nonself recognition*:** The immune system responds only to those antigens that are not "self ". Because of this attribute, self-antigens are spared, which is essential because response to self-antigens can result in fatal auto-immune diseases.

The more refined specific immune response is usually elicited only in the face of either failure/continued stimulation of the non-specific response.

The specific immune responses are mediated by humoral and cellular elements, which complement each other. As indicated earlier, (a) if the response to an immunogen is the production of soluble antibodies in the body fluids (i.e. humours), it is called **humoral immunity,** and (b) if the response is through cytotoxic or killer cells, then the immunity is known as **cell-mediated.** The cells responsible are lymphocytes, that develop and reside in the lymphoid tissues and are of two major types:

1. ***B-lymphocytes***: They are independent of the thymus but dependent on the ***bursa of Fabricius*** in birds; its *equivalent in man is bone marrow*. B lymphocytes are responsible for the humoral immunity. The antibodies produced in this type recognize and destroy specific foreign antigens, either alone or along with complement.
2. ***T-lymphocytes***: They are dependent upon **thymus** for becoming immunologically competent from non-competent predecessor originating in the bone marrow, and are responsible for cell-mediated immunity. The activated T lymphocytes (**cytotoxic cells**) act on self-cells infected by virus, destroying the self-cell as well as the viral antigen. *A subset of T-helper cells lie at the heart of all immune responses*: they release a group of proteins (cytokines), which activates different arms of immune system (Fig. 33.1).

Currently, immune response is thought to consist of two major lymphoid systems: the thymus-dependent (cell-mediated system) or the bursa equivalent-dependent (humoral system).

A coordinated interplay of the innate and the specific immune response successfully defends against most invading pathogens. However, these pathogens develop various means of evading the host defenses, thereby intensifying the challenge (Box 33.1).

Though in the foregoing discussion, the two arms of immune response—innate and specific—have been discussed separately, they do not operate independently of each other. They interact with each other, as evidenced by the following examples:

- Cells of phagocytic systems, most notably macrophages, capture the foreign antigen and present them (antigens) to cells of specific immune response, leading to activation of the latter.
- The soluble factors (cytokines), produced during a specific immune response, augment the activity of these phagocytic cells.

A well-orchestrated and carefully regulated interplay of the two arms of immune system, as outlined in Figure 33.1, effectively eliminates the foreign invaders. Detailed mechanisms of these cooperative interactions are presented throughout the rest of this chapter.

BOX 33.1

The Response-evasion Interaction: A See-saw Struggle between Pathogen and Host

Most invading microorganisms are effectively combated and often eliminated by innate response, chiefly by phagocytic cells, which *engulf* them, and then, with the aid of the degradative enzymes in their lysosomal granules, which *break down* the trapped material into simpler molecules for extrusion. However, some pathogenic organisms have evolved a strategy of **evasion** in that they have become able to reside within phagocytes, where they can easily survive and multiply. Some of these pathogens prevent fusion of the lysosome with the phagosome and are able to proliferate within the phagosomes; others produce inhibitors of lysosomal enzymes; some others can resist free radical attack; and still others escape from the phagosome and proliferate within the cytoplasm of the infected macrophage. Not only these pathogens have developed defense against the non-specific phagocytic defense system, but also their intracellular location prevents production of antibodies against them. The **host response** to this challenge has been development of system of cellular plasma membrane scrutiny. Cells infected with intracellular pathogens—either bacteria or virus—acquired the ability to present fragments of proteins from the invading microorganisms on their surfaces by means of the MHC molecules. Cells with foreign peptides bound to their MHC molecules were then recognized as being infected and killed by groups of T lymphocytes (i.e. the cytotoxic T cells). In addition, intercellular signalling molecules (cytokines) were released by another group of T cells, which promoted activation of macrophages. The activated macrophages, with enhanced functional capacity, were then able to deal with some of these pathogens more effectively.

III. Antigens and Immunogens

Antigens are molecules that react with antibodies, whereas immunogens are molecules that induce an immune response. In most cases antigens are immunogens and the terms are used interchangeably. The features of molecules that determine immunogenicity are as follows:

1. ***Foreignness*:** The molecule should be recognized as nonself by the immune system
2. ***Molecular size*:** Macromolecules having molecular weight over 100,000 are the most potent immunogens; those below 10,000 are weakly immunogenic, whereas small ones, e.g. an amino acid are nonimmunogenic. Exceptions are seen in haptens.
3. ***Structural/chemical complexity*:** A certain amount of complexity is required for immunogenicity. For example, homopolymers are less immunogenic than heteropolymers.
4. ***Antigenic determinants (Epitopes):*** Epitopes are small chemical groups on the antigen molecule that can elicit immune response and react with antibody. An antigen can have one (monovalent) or more (multivalent) antigenic determinants.
5. ***Others*:** Dosage, route and timing of antigen administration and the genetic constitution of the host also affect immunogenicity.

Haptens: Generally, antigens are immunogens but there are certain important exceptions, e.g. haptens which are not able to provoke immune response but which can bind to antibodies. Haptens are usually small molecules, but some high molecular weight nucleic acids are haptens as well. They cannot induce the immune response by themselves but do become immunogenic when bound to a "carrier" protein. The carrier protein may activate T-helper cell, while the hapten interacts with the B cell, bearing surface receptor antibody (i.e. IgM) specific for the hapten. The activated T-helper cell stimulates the B cell to produce antibodies to the hapten.

Adjuvants enhance immune response to an immunogen though they are unrelated to the latter. Some human vaccines contain adjuvants.

IV. Role of Lymphocytes

Lymphocytes are the cells **primarily involved in generation of an immune response**. There are approximately 10^{12} lymphocytes in the human body, the total cellular-mass of which is equivalent to that of the liver. The four features which characterize all immune responses—specificity, diversity, memory and self/nonself recognition—are mediated by lymphocytes.

Lymphocytes are produced from haematopoietic stem cells in the bone marrow and differentiate in the ***central lymphoid organs***: B cells in bone marrow and T cells in the thymus. Through this process, called ***maturation***, lymphocytes become immunocompetent (Fig. 33.2). From central lymphoid organs, the immunocompetent lymphocytes migrate through the bloodstream to the ***peripheral lymphoid tissues*** (lymph nodes, spleen, tonsils, etc.) which are sites of lymphocyte activation by antigen.

B cells and **T cells** comprise respectively 50–70% and 10–20% of the total lymphocyte population. Both recognize antigens by means of membrane receptors: specificity of the immune response is through the recognition of the antigen by such receptors.

A third population, termed **natural killer (NK) cells,** make up the remainder of the lymphocytes. NK cells are so called because they demonstrate the ability to kill without prior exposure or sensitization. Despite their different functions, discussed below, *it is not possible to discern any morphologic features that can be used to distinguish B cells and T cells, and NK cells from one another*. Instead, their identification is based on immunophenotypic or functional studies.

The cells responsible for the specific immune response are the lymphocytes: B, T and NK. The specificity they show for the inciting antigen is achieved via the use of specific antigen receptors, expressed on their cell surface.

A. The B Lymphocytes

B cells, the mediators of humoral immune response, undergo their early development in the bone marrow. This process, termed ***maturation***, is antigen independent, leading to appearance of unique antigen binding membrane receptors. The B-cell receptor is a membrane bound immunoglobulin (IgM, in its monomeric form), termed **surface immunoglobulin (sIg)**. There are approximately 10^5 sIg molecules on the membrane of a single B cell. Each sIg molecule has an identical antigen binding specificity, which enables the B cell to recognize specifically "its" antigen (Fig. 33.2).

Structure of a sIg: It consists of two identical heavy polypeptide chains and two identical light chains, held together by disulphide bonds. The amino terminal ends of each heavy and light chain fold together to form the antigen-binding cleft (Chapter 5). The sIg differs from the secreted immunoglobulin in the amino acid sequences at the carboxy end of the heavy chain: the sIg contains a transmembrane helix that is lacking in the secreted immunoglobulin. This structural difference results from the optional use of a small exon for the membrane attachment region (Chapter 24).

Activation of B-cell: Upon encountering an antigen, for which its surface immunoglobulin is specific, the B cell gets activated in a peripheral lymphoid tissue. It undergoes rapid cell division and its clonal progeny differentiates into two cell types: (a) the antibody secreting ***plasma cells,*** and (b) the ***memory cells*** (Fig. 33.2).

- Plasma cells are the effector cells, which produce an enormous amount of one of the five classes of antibodies in a short lifespan of a few days. A single plasma cell can secrete more than 2000 antibody molecules per second; all clonal progeny from a given B cell being of the same antigen-binding specificity. The secreted antibodies are the major effector molecules of humoral immunity (discussed later).
- Memory cells, on the other hand, have a longer lifespan (circulate for several years) and continue to express membrane bound immunoglobulins with the same specificity as the original parent cell. They remain in the body to respond to the same antigen during a subsequent encounter.

Immunoglobulins are produced by a class of appropriately stimulated B cells, called the plasma cells. A proportion of the stimulated B cells become long lived memory cells, which are the basis of long-term immunity from a repeat infection.

B. The T Lymphocytes

T cells derive their name from the site of maturation in the thymus. Roles of T cells are three-fold:

- They kill host cells infected with pathogens, which, because of their intracellular locations, are sequestered from antibodies.

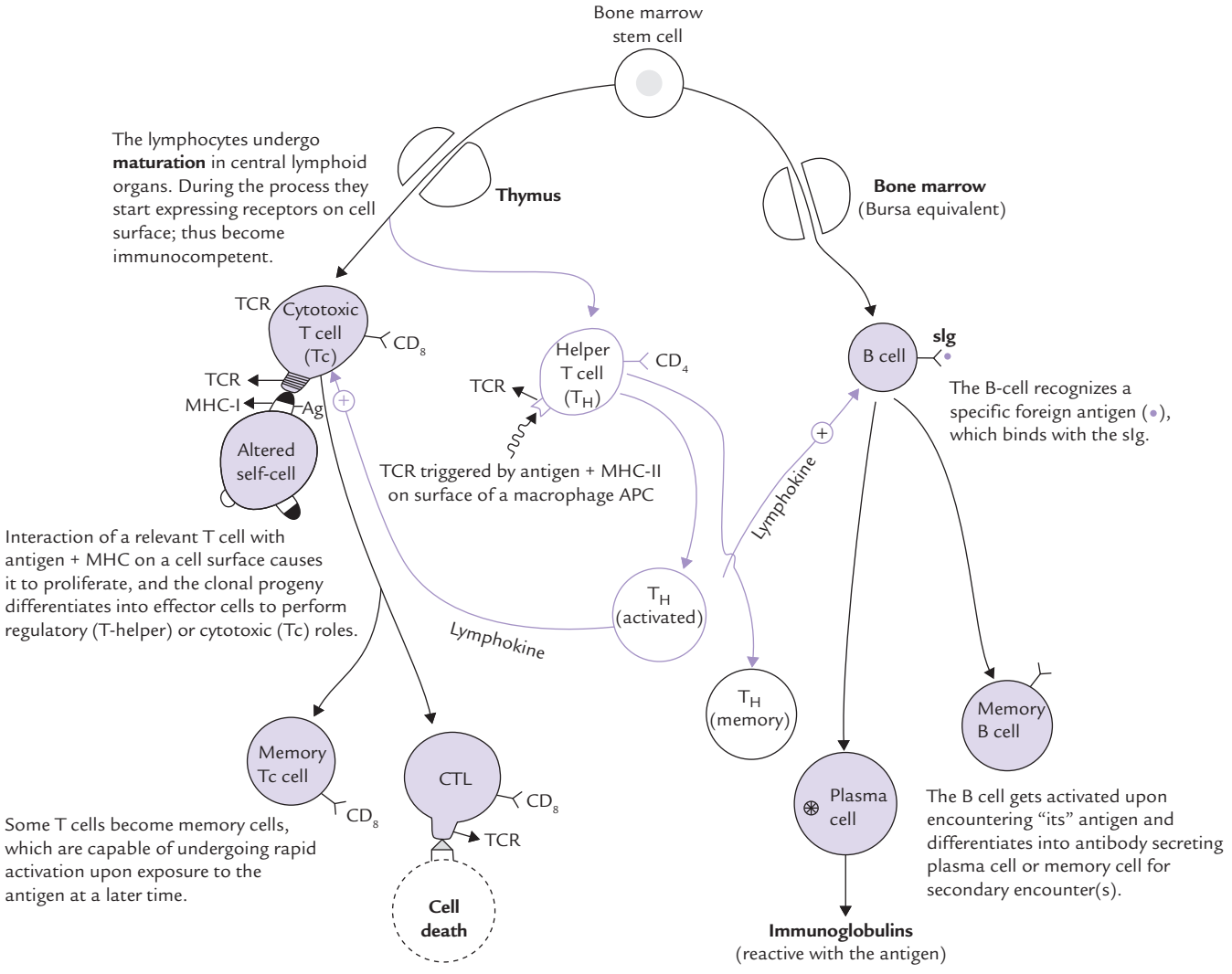


Fig. 33.2. Origin of B and T cells and their role in induction of humoral and cell-mediated immunity (APC = antigen-presenting cell, TCR = T cell receptor, CTL = cytotoxic T lymphocyte, slg = surface immunoglobulin, Ag = antigen).

- Where infection persists, they help in maintaining an inflammatory response.
- Finally, they secrete cytokines which are regulators of various immune effector cells.

Like B cells, the T lymphocytes are also produced in the bone marrow, but unlike B cells which mature within the bone marrow, T cells migrate to the thymus gland to mature (Table 33.2). The term *maturation implies acquiring immunocompetence versus antigen of a given specificity*. During maturation, the T cell starts expressing a unique membrane receptor for antigen, called **T cell receptor (TCR).**

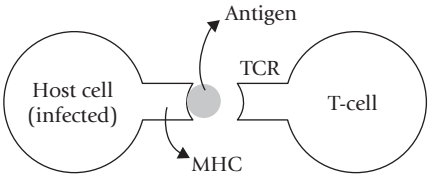
Structure of TCR: The receptor is a heterodimer, comprising two non-identical polypeptide chains, linked by disulphide bonds.

These chains are termed the TCR-α, β, γ, and δ chains and the functional combinations are αβ and γ δ chains. Each chain comprises two domains—one constant and one variable. The amino terminals of the two chains pair to form a cleft within which antigen binds. Although TCR is structurally distinct from immunoglobulin, it is in some respects the counterpart of the latter, most notably with regards to structure of its antigen-binding site.

What distinguishes the TCR from membrane-bound antibody on B cells (i.e. the sIg) is that it recognizes antigen only when antigen is complexed with a cell-membrane protein, i.e. the major histocompatibility complex (MHC) molecule. It is as if MHC is a signpost, guiding the T-cytotoxic cell to its target.

Table 33.2. Comparative features of T and B cells

Feature	T cells	B cells
Site of maturation	Thymus	Bone marrow
Mediator	Cellular immunity	Humoral immunity
Immunoglobulin synthesis	No	Yes
Synthesis of cytokines	Yes	No
Antigen receptor on surface	Heterodimer	IgM (monomer)
CD_3 protein on surface	Yes	No



This points to the fundamental difference between the humoral and the cell-mediated branches of the immune system. *Whereas the B cell is capable of binding soluble antigen, the T-cell system is restricted to binding antigen displayed on self-cells with a background of self-molecules, in the form of MHC*. The antigen in association with MHC may be presented on the surface of the antigen-presenting cells or on virus infected cells, cancer cells and grafts (Box 33.2).

T cell subpopulations: Two major subpopulations are known:

- immunoregulatory T-helper cells, and
- cytotoxic T cells (Fig. 33.2).

These two subpopulations of T cells can be distinguished by their display of either the one or the other of the two membrane glycoproteins, CD_4 or CD_8. The T-helper cells display CD_4 and recognize antigens associated with MHC molecules of class II. The cytotoxic T cells display CD_8 and recognize antigens associated with class I-MHC molecules. *Thus, the T-helper cells (Tc) are class II restricted and T-cytotoxic cells are class I restricted.*

Activation of T cells: Interaction of T_H cells with antigen + MHC-II results in activation of the former, so that it proliferates extensively and starts secreting various growth factors, known collectively as **cytokines,** or more specifically

BOX 33.2

Antigen-presenting Cells Ensure Appropriate Stimulation of the T-cell Activity

Antigen-presenting cells (APCs) are specialized cells which are involved in antigen trapping and concentration of antigen at the cell surface for effective presentation to the T lymphocytes, so as to start off the antigen-specific immune responses. These include interdigitating cells in the thymus, the Langerhans cells of the skin, interdigitating cells in the T areas of lymph nodes and follicular dendritic cells in B areas of lymph nodes. The APCs are distinguished by their ability to express a particular type of MHC molecule on the cell surface. They internalize antigen, either by phagocytosis or endocytosis, and then re-express part of that antigen together with the MHC molecule on their membrane. The combination of antigenic fragment with the MHC molecule is then recognized by the relevant T cell. This is an important mechanism to ensure appropriate stimulation. It limits the possibility of a single microorganism being able to trigger the immune system for no good reason, because there is a huge potential for binding to at least some component (i.e. epitope) of the microbe. In this way, MHC molecules presented on the cell surface (of antigen presenting cell) restrict activation to a limited number of relevant T cells. Without such restriction inappropriate T-cell response may result with hazardous consequences.

BOX 33.3

Cluster of Differentiation (CD)

Ratio of TH cells and Tc cells is approximately 70 in normal human peripheral blood, as ascertained by determining the number of CD_4-bearing and CD_8- bearing T cells. In certain diseases, such as immunodeficiency diseases or autoimmune diseases, this ratio may show significant alterations. Presently many techniques are available to detect the lymphocyte membrane molecules, e.g. CD_4 or CD_8. A number of other such membrane molecules, or surface markers as they are now called, have been detected. For example, B cells express CD_{19} and CD_{20}, the T cell receptor is complexed with CD_3, and the NK cells are currently identified by the expression of CD_{16} and CD_{56} in combination. These markers indicate different lineages or maturational stages of lymphocytes, macrophages and other cells, and thus allow subdivision of the main morphological categories. Formerly called T1, T3, etc. they have been renamed as CD_1, CD_3, etc. (CD_5 cluster of differentiation). CD testing is done with monoclonal antibodies: all monoclonal antibodies that react with a particular membrane molecule had been grouped as a cluster of differentiation (CD) during the First International Workshop on Human Leukocyte Differentiation Antigens. New determinants are being described each year. However, some functional groups of cells still remain, which cannot be distinguished by any monoclonal antibodies; for example, T-suppressor cells, that can be only identified by their functional characteristics.

lymphokines. Secretion of lymphokines is a crucial event which brings about activation of B cells, T_c cells, phagocytic cells and various other cells that participate in the immune response (e.g. macrophages).

The T_c cell on the other hand is activated by interaction with an antigen-MHC-I complex on the surface of an altered self cell in the presence of appropriate lymphokines. The activated Tc cells proliferate and differentiate into effector cells, called **cytotoxic T lymphocytes** (CTL), which mediate the killing of the altered self cells (Fig. 33.2).

In addition to T_H and T_c, another type of T cell called ***T-suppressor*** cell has been postulated, but it is still not clear whether its lineage is distinct from the T_H and T_c subpopulations. Unlike the T_H and T_c, they do not display any CD surface markers (Box 33.3).

T cells' response, called the cellular immune response, is initiated when T cells recognize processed antigen presented by MHC-bearing cells via the TCR interaction, leading to generation of the so-called effector functions. The functions include secretion of lymphokines and T-cell-mediated cytotoxicity brought about by the T-helper and T-cytotoxic subsets, respectively.

C. Natural Killer Cells

These are large granular cells, important in resistance to virus infections and probably also malignancies. When a cell becomes infected by a virus or undergoes malignant transformation, its surface molecules are altered in that it starts expressing high molecular weight glycoproteins on the membrane. The latter are recognized by NK cells. These cells bear NK cell-receptors which recognize these glycoproteins on the surface of the altered self-cells, at an early stage before the virus has had a chance to reproduce. The NK cells are activated by the cytokine, interferon-γ, as discussed later. When activated, they release their granular contents into the space between the target and the NK cell, including a molecule called perforin, which acts, as its name implies, by punching a hole in the cell membrane of the target cell leading to its death.

NK cells also have receptors for immunoglobulin, the latter links the NK cell and the target cell closely. This phenomenon called *antibody-dependent cell-mediated cytolysis* (ADCC) has been described earlier (Chapter 5).

Lymphocytes are able to recruit and utilize components of the non-specific response in a more effective and targeted manner with the aim of eliminating or eradicating the antigen.

V. Organs of the Immune System

Organs of the immune system, or the lymphoid organs, are classified as primary or secondary on the basis of their respective functions. The **primary lymphoid organs** provide appropriate environment wherein maturation of naive lymphocytes into antigen-recognizing immunocompetent lymphocytes takes place, the most notable of them are T and B cells. The **secondary lymphoid organs** trap antigen from defined tissues or vascular spaces and provide sites where these antigens interact with the immunocompetent lymphocytes. An antigen-driven proliferation and differentiation of the mature lymphocytes follows to generate an immune response.

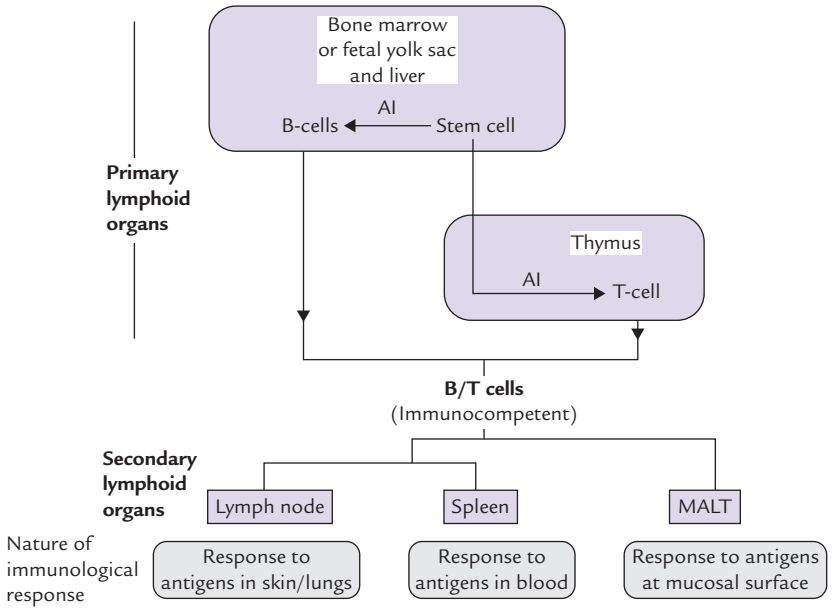


Fig. 33.3. Organization of the organs of immune system (AI = antigen independent).

The organization of primary and secondary lymphoid organs is illustrated in Figure 33.3.

Conversion of naive lymphocytes into immunocompetent T- or B-cells (maturation) occurs in primary lymphoid organs. In secondary lymphoid organs the immunocompetent cells interact with relevant antigen(s) and get activated.

A. Primary Lymphoid Organs

These include the *fetal* ***yolk sac****, the* ***liver*** *in the embryo, and the* ***bone marrow*** *and* ***thymus*** *after birth*. Lymphocytes originate and undergo early development and differentiation in the primary lymphoid tissues. Their location alters during gestation: in early embryo the main site of lymphocyte production is the fetal yolk sac but later in embryonic life this shifts to the liver. Still later this shifts to the bone marrow and the thymus where it will remain throughout the rest of an individual's life.

Stem cells in the ***bone marrow*** *either differentiate to give rise to* ***B cells*** *or they migrate to the* ***thymus*** *where their development into various types of* ***T cells*** *is fostered* (Fig. 33.3). The B cell progenitors initially formed in the marrow pass through a maturation process to become mature B cells; the process results in changing a (virgin) B cell into an immunocompetent B cell (i.e. capable of mounting an immune response). An immunocompetent cell is committed to a particular antigenic specificity. The cells that migrate to the thymus gland (i.e. the immature thymocytes) mature there to become antigen-committed, immunocompetent T cells. Maturation of both B and T cells are **antigen-independent processes**, occurring prior to the antigenic challenge.

Thymus

The thymus is a bilobed structure found in the anterior chest or mediastinum above the heart wherein the T-cell progenitors enter as immature thymocytes, and mature into functional T cells. The **maturation** involves:

- Expression of T-cell receptor (TCR) on cell surface.
- Commitment of such TCR expressing cell to recognize and respond to a given antigen.

Structure of thymus: Thymus has epithelial and mesenchymal (supporting tissue) components in addition to lymphoid contributions. At the microscopic level, each lobule is organized into two compartments: an outer *cortex* which is densely packed with thymocytes, and an inner *medullary area* which is sparsely populated with thymocytes.

Though the actual maturation sequence is not known, the T-cell maturation appears to progress as the immature thymocytes migrate from the cortex to the medulla. During their migration, these cells interact with three-dimensional network of the thymic stromal cells (composed of epithelial cells, interdigitating dendritic cells and macrophages) which make up the framework of the thymus and contribute to the thymocyte maturation. Many of these cells physically interact with the developing thymocytes and also secrete hormonal factors (e.g. ***thymosin***, ***thymopoietin*** and ***thymulin***) necessary for the differentiation and maturation of the T lymphocytes. The thymic stromal cells secrete a cytokine, **IL-7**, which also plays a role in T-cell maturation within thymus.

In the course of thymocyte maturation, the cells start expressing antigen-binding receptors, called the T-cell receptors (TCRs); each cell expressing a TCR of a unique specificity. (The antigenic diversity of these receptors is generated by a series of random rearrangements of gene segments encoding the receptor polypeptide chains.

Selection processes during maturation: The mature, immunocompetent T cells so formed would be subsequently interacting with foreign antigens. However, T cell is capable of interacting with an antigen only when the latter is displayed in association with a self-MHC molecule (Fig. 33.2). Therefore, it is subjected to a rigorous selection process so that only those T cells which (a) can recognize antigenic peptide in the context of self-MHC, and (b) which are *self-tolerant* (non-reactive with self antigen) are released from the thymus. Thymic stromal cells play an important role in the selection process, which has both positive and negative elements:

1. ***Positive selection*:** The thymic stromal cells express high levels of class I and class II MHC molecules to which the developing thymocyte is exposed all through the maturation process. Only the *T cells bearing receptors that recognize foreign peptides in association with self-MHC will be selected and allowed to mature further*. Any developing thymocyte, incapable of recognizing self-MHC plus foreign-antigen will not be selected and eliminated by apoptosis or programmed cell death.
2. ***Negative selection***: Among the positively selected cells, some will be potentially self-reactive, i.e. ***reacting with self-antigen***. These cells must be **eliminated**. Negative selection targets for destruction of any such cell bearing a high affinity receptor for self-antigen (or self-antigen associated with self MHC).

So rigorous and unsparing are these processes that 95–99% of all thymocyte progeny dies within the thymus, without ever fully maturing. Rest of the $< 5\%$ cells, which are self tolerant with diverse TCRs on their surface, appear in circulation. Subsequently, these cells would migrate to secondary lymphoid organs where they form various effector population of T cells (Fig. 33.3).

Selection processes result in elimination of those thymocytes whose receptors cannot recognize foreign antigen in association with self-MHC (positive selection), or whose receptors recognize self-antigen, either alone or associated with self-MHC (i.e. negative selection).

Bone Marrow

In birds, a lymphoid organ called the bursa of Fabricius is the primary site of B-cell maturation, whereas in humans bone marrow serves as the "bursa equivalent" where the B cell matures. The process is antigen-independent, leading to the production of immunocompetent (and self-tolerant cells) with diverse monomeric IgM and IgD molecules expressed on their surface. These molecules constitute the **B-cell receptors** or surface immunoglobulins **(sIg)**.

Subsequently, when B cells reach the secondary lymphoid tissue and encounter foreign antigen, they are stimulated to proliferate to yield two types of cells: plasma cells and memory cells (Fig. 33.2). ***Plasma cells*** *secrete the soluble antibody corresponding to the sIg found on the stimulated parent cell,* and **memory cells** persist until a second encounter with that antigen occurs, when they mediate the secondary immune response (Chapter 5).

Lymphocytes originate in bone marrow; undergo early development and differentiation in primary lymphoid organs, and become committed to a B- or T-cell lineage.

B. Secondary Lymphoid Organs

The adaptive immune response occurs mainly in the secondary lymphoid organs: the lymph nodes, the spleen, and the mucosal-associated lymphoid tissues (MALT) (Fig. 33.3). These tissues are uniquely suited to *trap pathogens and their secreted antigens, and to present these to the naive lymphocytes that constantly pass through.*

- **Lymph nodes:** Microorganisms that enter the body through the skin or the lungs drain to regional lymph nodes where they stimulate an immune response.
- **Spleen:** Microbes that enter the bloodstream stimulate an immune response in the spleen.
- **MALT:** Microorganisms and food antigens that enter the gastrointestinal tract are collected in the MALT.

Common to all lymphoid tissues is the degree of compartmentalization, with specific areas for T cells and B cells, and areas of overlap where they interact.

Lymph Nodes

Structurally, the lymph node is divided in three concentric regions called ***cortex***, ***paracortex*** and ***medulla***. The outermost cortex contains follicular structures of two types: the unstimulated primary follicles and the stimulated secondary follicles. The latter are characterized by the presence of germinal centre—the sites of intense B-cell activation and differentiation into plasma and

memory cells. The paracortex, which lies inner, to the cortex, contains primarily T cells. The innermost medulla contains several antibody-secreting plasma cells.

Spleen

The spleen contains non-lymphoid tissue populated by macrophages and erythrocytes, where the old and defective red blood cells are destroyed and removed (the **red pulp**); as well as lymphoid areas, termed the **white pulp**. The latter contains T cells and several interdigitating dendritic cells. The *blood-borne antigens are trapped* by their interdigitating dendritic cells, which present these antigens to T_H cells, thereby initiating immune response.

MALT

It comprises lymphoid elements found at various locations along mucous membrane surfaces where they trap antigen and provide localized sites for lymphocyte interaction with that antigen. Structurally these tissues range from loose clusters of lymphoid cells with little organization in the lamina propria mucosa of the intestinal villi to organized structures such as the **tonsils, adenoids, appendix** and **Peyer's patches**. Lymphocytes are also found singly throughout the mucosal epithelium.

Secondary lymphoid organs include spleen, lymph nodes and mucosa-associated lymphoid tissues. It is at these sites that immune response is actually generated.

Lymphocyte traffic around lymphoid tissue: There is a one way traffic of T and B cells from the primary lymphoid organs into the bloodstream and continuous re-circulation of these cells between the secondary lymphoid organs, tissues and bloodstream. Initially, these cells migrate via blood from the primary to the secondary lymphoid organs (lymph nodes, spleen and MALT), where they are stimulated by antigens. Lymphatics drain extracellular fluid as lymph through the lymph nodes and into the thoracic duct, which returns the lymph to the bloodstream by emptying into the left subclavian vein (Fig. 33.4). Thus, there is a continuous re-circulation.

As the lymphocytes re-circulate, they make contact with antigens presented on the surface of antigen-presenting cells in the secondary lymphoid organs. However, only a very small percentage (about one in 10^3 to 10^6) of the lymphocytes can recognize a particular antigen within a relatively short period of time in order to generate specific immune response. The odd of small percentage of immunocompetent lymphocytes actually making contact with the antigen is greatly overcome by the extensive lymphocytes traffic into and out of lymphoid tissues, especially after an antigenic challenge and by cell adhesion molecules (Box 33.4).

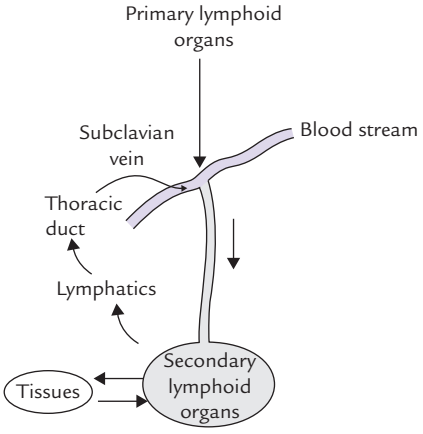


Fig. 33.4. General routes of lymphocyte circulation.

BOX 33.4

The Cell Adhesion Molecules (CAMs)

The whole process of immune response is dependent on the expression of the molecules and ligands, on external surfaces of cells that mediate adhesion between cells. CAM is a collective term for such molecules and ligands that have the function of binding the cell to other cells or to the connective tissue elements. They are found on a wide variety of cell types and significantly influence immune response. Their expression is increased by cytokines. They have important roles in mediating interactions between lymphocytes and endothelia lining the small blood vessels. A major determinant of this interaction is the prevailing cytokine environment.

The CAMs are grouped into the following families based on their structural features: selectins, integrins, intracellular adhesion molecules (ICAM), and immunoglobulin supergene family. Their role in leukocyte migration and homing is exemplified as below. At the site of infection, expression of selectins on the endothelial cell surface is induced by cytokines. Selectins interact with carbohydrate ligands on leucocytes. This binding retards flow on leucocytes, so that they sluggishly roll on the endothelium. This facilitates binding of ICAM-1 on endothelial cell with its counter receptors on leucocytes, which further arrest rolling and allow leucocytes to squeeze between the endothelial cells. As a result, leucocytes are able to form a protective wall on the blood vessel.

VI. Cells of the Immune System

A wide range of cells participate in non-specific and specific immunity. As described earlier, *lymphocytes are the pivotal and critically important cellular components of the immune system; all other cells play only accessory/supportive roles*, such as serving either to activate lymphocytes or to increase effectiveness of antigen clearance by phagocytosis, or secretion of cytokines.

The cells of immune system can be divided in two major groups:

(a) the **antigen-presenting cells:** macrophages, interdigitating cells, Langerhans cells, follicular dendritic cells, etc. and
(b) **white blood cells** or leucocytes, divided into five categories: neutrophils, eosinophils, basophils, lymphocytes and monocytes (the first three cell types listed are referred to under the general name of granulocytes).

Many of the cells involved in immunity fulfil more than one function; for example, macrophages, as the name ("**larger eater**") suggests are the cells chiefly involved in engulfment of microorganisms (i.e. phagocytosis). In addition, they are also known to play a role in induction of immune response by presenting antigens on their surface to specific T lymphocytes.

A. Lymphocytes

As discussed earlier lymphocytes are the motile, non-phagocytic cells which play a central role in immune response, being responsible for attributes of specificity, diversity, memory and self/nonself recognition—hallmarks of an immune response. The constantly circulating lymphocytes between tissues, lymphoid organs, blood and lymph provide a high degree of cellular integration to the immune system as a whole. Different lineages of lymphocytes and their functions have been already dealt with, in detail.

B. Macrophages

The macrophages in tissues, together with the circulating monocytes in the blood constitute the mononuclear phagocytic system. Macrophages derive from bone marrow promonocytes which, after differentiation to blood monocytes, finally settle in tissues as mature macrophages. Monocytes remain for about 8 hours in bloodstream, and then enlarge and migrate to tissues as they differentiate into macrophages.

Many organs contain characteristic population of macrophages, e.g. liver Kupffer cells, lung alveolar macrophages, brain microglial cells, splenic lymphoid macrophages, and kidney mesangial cells.

Promonocyte (Bone marrow) → Monocytes (Blood) → Macrophages (Tissues)

Functions

Macrophages serve dual role being involved in **phagocytosis** and in **presenting antigens** on their surface to specific T-cells.

Phagocytosis: Cells of the mono-nuclear phagocytic system—macrophages and monocytes—with their impressive anti-microbial potential, present formidable weaponry against the invading pathogen. In fact, macrophages are one of the two cell types involved in engulfment and digestion of exogenous antigens, other ones being microphages (small eaters) or polymorphonuclear neutrophils, discussed later.

In addition to the whole pathogenic microorganisms, the antigens engulfed by these phagocytic cells include cellular debris, insoluble particles, injured and dead host cells, and activated clotting factors. Engulfed by cytoplasmic processes, the antigen initially comes to lie in a vacuole termed a phagosome, and subsequently digested by lysosomal contents. The latter include *hydrogen peroxide, active oxygen metabolites, peroxidase, lysozyme, lactoferrin, cationic proteins* and a rich variety of proteolytic and other hydrolytic enzymes. The digested material is then extruded from the cell by exocytosis.

Note: Aggregates of macrophages, **granulomas**, are characteristic of many chronic infectious and idiopathic inflammatory diseases such as tuberculosis, leprosy, and sarcoidosis.

Antigen presentation and processing: Following phagocytosis, the macrophages process the ingested antigen and then present it appropriately in association with MHC molecules. In general, when antigen is taken up by macrophages, a proportion is degraded by phagocytic digestion while part is fixed to the cell surface where it is thought to be in a strongly immunogenic state in association with MHC class II molecules. *Antigen presentation in association with MHC-II is an essential requirement for activation of T_H cells*, which cannot recognize an antigen alone. As discussed earlier, this presentation is a crucial event in the development of humoral and cell-mediated immune responses.

Secretory role: Following phagocytosis, the macrophages release a number of protein factors, which are central to the development of the immune response. **Interleukin 1 (IL-I)** is the first one to be released after macrophage phagocytose antigen. IL-I is not only required for activation of

a variety of cells (e.g. T and B cells, neutrophils, fibroblasts), but it also effects vascular endothelium to influence the inflammatory response and acts as endogenous pyrogen: it acts on thermo-regulatory centre in the hypothalamus, leading to fever response. The other factors released by macrophages and their function in immune response are as below:

1. ***Interferon alpha:*** Activates cellular genes resulting in the production of proteins that confer an **antiviral** state on cell.
2. ***Tumour necrosis factors (TNF):*** Cause necrosis of a variety of cells including tumour cells. Along with IL-I, the TNFs play a critical role in initiation of inflammatory response. TNF-alpha, also known as **cachectin,** has profound effects on general cellular metabolism; it causes weight loss, fever, acute phase reaction, etc.
3. ***Hydrolytic enzymes:*** Can be secreted by activated macrophages in tissues where they promote inflammatory response.
4. ***Complement proteins:*** Assist in elimination of foreign pathogens and in the amplifying inflammatory reaction.

Finally, activated macrophages secrete a number of cytokines (Interleukin-6, γ-interferon, GM-CSF, G-CSF and M-CSF), which are considered in detail subsequently in this chapter.

C. Granulocytes

These cells have granulated cytoplasm that stains with acid or basic dyes. On the basis of cytoplasmic staining characteristics and cellular morphology, they have been divided in three major types: eosinophils staining with acidic dye (eosin Y); basophils staining with basic dye (methylene blue); and neutrophils staining with both acidic and basic dyes. Both neutrophils and eosinophils are phagocytic, whereas basophils have a major role in allergic responses.

Functions

Neutrophils, which are much more numerous than other granulocytes, constitute 50–70% of the circulating white blood cells (basophils 1–3% and eosinophils $<0.1\%$). They are produced in the bone marrow during haematopoiesis and released into the peripheral blood where they circulate for 7–10 days, and then migrate into the tissues where they have a 3-days-lifespan. Process of phagocytosis by neutrophils is same as that for macrophages, except that the lytic enzymes contained in the cytoplasmic granules mediate killing of pathogen mainly by a cytotoxic free radical attack. The attack is so vigorous that the granulocyte cells themselves rarely survive, whereas in case of macrophages several rounds of killing can be carried out by the same cell. Failure of polymorphs to fulfill the above task leads to severe infections, though clinical features may be minimal (**Case 33.2**).

Eosinophils are also motile phagocytic cells like neutrophils but their phagocytic role is mainly to aid human host defenses against parasites and worms. They are also implicated in allergic diseases such as asthma. Their presence in tissues, and a high count of eosinophils in the blood (eosinophilia), is therefore, often a marker of an allergic disease or parasitic infection.

Basophils are not phagocytic as mentioned earlier. They have a large number of cytoplasmic granules containing pharmacologically active substances, which play important role in development of allergic responses.

D. Dendritic Cells

These cells have peculiar structural features, such as an unusual shape resembling dendrites of a nerve cell and the cell surface being covered with a maze of long membrane processes. *They express high level of MHC molecules on their surface, therefore suited to function as important antigen-presenting cells for T-cell activation.* After capturing antigen in tissues, dendritic cells migrate to various lymphoid organs where they present the antigen to T cells. They are widely distributed in non-lymphoid organs and tissues, in lymphoid organs (e.g. interdigitating dendritic cells and follicular dendritic cells) and in blood and lymph (e.g. veiled cells).

Though the dendritic cells at these different locations have several morphologic and functional differences, they appear to have arisen from a common progenitor cell and may represent various stages of a single lineage.

E. Mast Cells

These cells have large cytoplasmic granules containing histamine and other pharmacologically active substances. Like basophils, they play major role in allergic responses during which they release the contents of their granules.

VII. Molecules of Immune Response

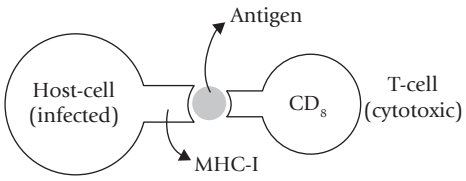
A. Proteins Encoded by Major Histocompatibility Complex (MHC)

The MHC complex is a large complex of genes, with multiple loci, present on the short arm of chromosome 6. Although the MHC was originally identified by its role in transplant rejection, it is now recognized that proteins

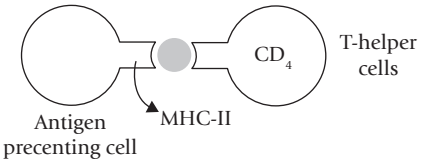
encoded by this gene cluster are involved in many aspects of immunological recognition.

General organization: The MHC gene cluster is divided into three regions termed **class I** (also designated HLA-ABC), **class II** (also designated HLA-D) and **class III**. The same nomenclature is applied to the respective polypeptide products: the polypeptide products of class I, II, and III genes are termed respectively MHC class I, II and III molecules (Fig. 33.5). The first two (i.e. class I and II molecules) are collectively known as **human leucocyte antigens (HLA antigens)** which are found on many cells, not only leucocytes. They determine the tissue type of an individual.

Class I molecules: These are glycoproteins found on the membrane of nearly **all cells of the** body except red blood cells. They comprise one peptide chain, encoded by the MHC called α-*chain*, associated with a different polypeptide, β_2-microglobulin (the latter is encoded elsewhere in the genome). The class I molecules are involved in immune recognition by binding the antigenic peptides synthesized by intracellular pathogens (bacteria or viruses). The peptides bound to MHC-I molecule is presented on the cell surface, and is recognized by receptors on CD_8 T cells.



Class II molecules: These polypeptide products of the class II MHC genes have more limited tissue distribution, being constitutively expressed on **antigen-presenting cells**, including *macrophages, dendritic cells*, and *B lymphocytes*. They perform the same function as the class I molecules in immune recognition (and have similar structural motif), but the peptides presented by MHC class II molecules can only be recognized by receptors on CD_4 T-helper cells and not by those of CD_8 T-cells.



Class III molecules: These are involved with the inflammatory response by virtue of coding for soluble mediators, including complement components and TNF.

The MHC is 'how T cells see antigen against the background of self'. They can restrict stimulation to a limited number of cell types through differing properties of class I and class II molecules.

MHC loci and allelic variants: The class I genes encoding the α-chains are organized in several loci, the most

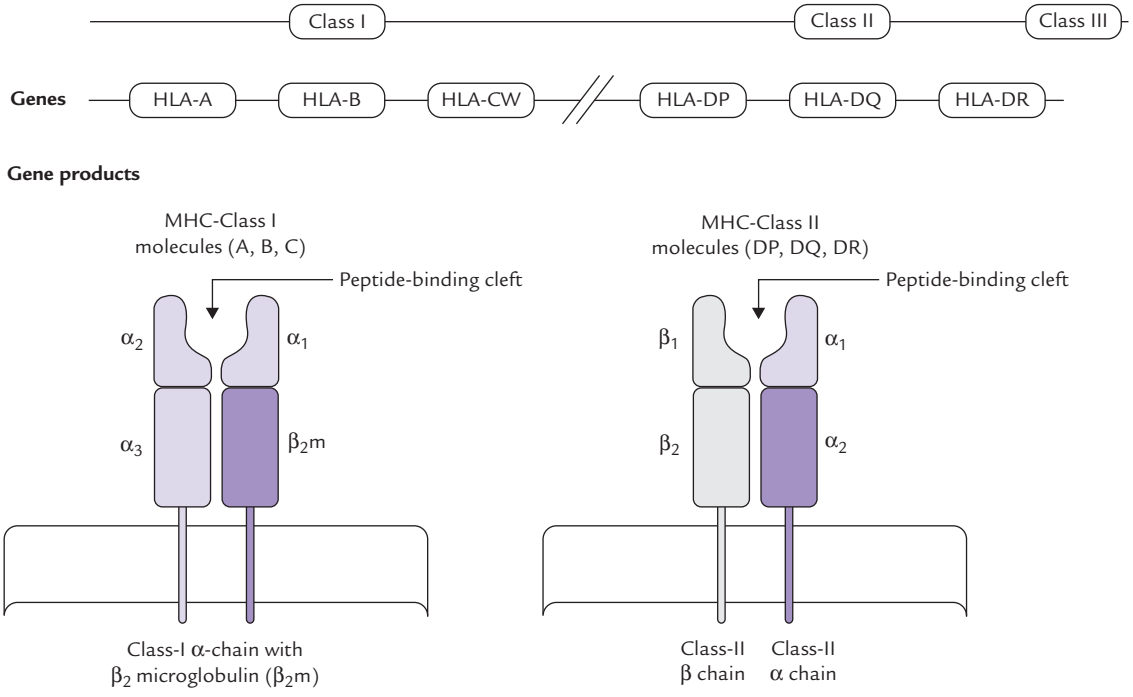


Fig. 33.5. Schematic representation of the domain organization of the major histocompatibility complex on chromosome 6, and the gene products.

important of which in humans are HLA-A, HLA-B and HLA-CW (there are two loci in mice-K and D). These genes are highly polymorphic and so there are numerous variants of HLA-A, HLA-B and HAL-C within the population. Each locus has a large number of different alleles (different forms of the same gene), which are transmitted and expressed in a co-dominant fashion. Because of their closeness on the chromosome, they are inherited 'enbloc' as parts of a haplotype. Since a person inherits one allele from each parent for each locus, a multiple class I MHC molecules are expressed on each of his nucleated cells.

The principal class II molecules are designated DP, DQ, and DR, and like the MHC class I molecules and they are highly polymorphic. *As in the case of the class I MHC, there are large number of different alleles for each class II locus.* Owing to inheritance of one allele from each parent for each locus, the surface of the antigen-presenting cell has multiple class II MHC molecules.

Intensely polymorphic nature of MHC system is illustrated in Box 33.5.

Subunit structure of MHC molecules: Schematic representation of a human class I molecule (Fig. 33.5) shows that it is a heterodimer composed of one transmembrane α-chain bound non-covalently to β_2 microglobulin. The α-chain is folded into three protein domains (α_1, α_2 and α_3), the first two of which form, a cleft into which the peptide antigen binds.

The class II molecule is composed of two transmembrane glycoproteins chains, α and β, each folded into two-protein domains. The antigen peptide binds in a cleft between the two chains.

B. Cytokines

Cytokines are the proteins secreted from a variety of cells and tissues that act as soluble mediators of the inflammatory and immune responses. They orchestrate the immune response by transmitting signals to and between cells engaged in the response (Table 33.3). They also regulate several biological processes, such as cell growth, tissue repair, fibrosis and morphogenesis.

General characteristics: A number of cytokines originating from various sources are known, which are grouped into different families, but all have several common features and characteristics.

1. They are peptide or glycoprotein in nature, are antigenic non-specific, transiently secreted and have short half-lives. Their orbit of influence is also of limited range.
2. Cytokines bring about their effects by interacting with **receptors** on surfaces of their target cells. They are active at concentrations between 10^{-9} and 10^{-15} moles, and the majority act within short distances of the site of their production (***paracrine***) or on the same cell that produced them (***autocrine***). A few are capable of acting on cells distant from the site of their production.
3. The cell types on which they act may be multiple (***pleiotropy***). They show significant overlap in their functions (***redundancy***) and have potential for interaction via the effects they mediate. As will be discussed later, the specific immune response is driven by direct cellular interactions together with the effects of cytokines.

Biological effects: The cytokines are given special names to indicate their source or actions:

- ***Interleukins (IL):*** ILs are secreted by one leukocyte to influence another. Currently, 18 interleukins have been recognized. They play important role in regulation of activities of various cells of specific and innate immune responses. Source and functions of common

BOX 33.5

Intensely Polymorphic Nature of MHC System

Many different allelic variants can be identified in each of the loci associated with antigen presentation. Likelihood of two individuals within the same species having identical set of allelic variants is remote. This is because in each of the six major loci there are between 10 and 60 alleles and each parent passes only one set or haplotype on each chromosome. The genetic polymorphism is predominantly in exon 2 of the β-chain of class II molecules and α-chain of class I molecules. These encode amino acids in the floor and sides of the peptide grooves and thus have a significant bearing on the capacity to bind a large number of differing antigenic fragments.

Table 33.3. Source and action of some cytokines

Cytokine	Major source	Important functions
Interleukin-1	Activated macrophages and other antigen-presenting cells	Activates T-helper cells leading to secretion of interleukin-2, causes fever.
Interleukin-2	Activated T-helper cells	Activates helper and cytotoxic T cells.
Interleukin-4	Activated T-helper cells	Stimulates B-cell growth and promotes Ig isotype switching.
Interleukin-5,6	Activated T-helper cells	Stimulates B-cell differentiation into antibody-secreting plasma cells.
Gamma interferon	Activated T-helper cells	Stimulates phagocytes, especially macrophages.
Tumour necrosis factor	Activated macrophages	Activates neutrophils mediated septic shock, causes necrosis of tumours, cytotoxic to normal tissue also (but have short half-life).

interleukins (1, 2, 4–6) are outlined in Table 33.3, and a detailed account of their roles in making an immune response is given subsequently (see cellular cooperation in immune response).

- ***Interferons (IFNs)***: IFNα and IFNβ are produced altruistically by virus-infected cells. They bind to nearby cells and induce a generalized antiviral state by inhibiting viral replication. Interferon γ is secreted by certain activated T_H cells and promotes killing by macrophages and natural killer cells. It also plays important part in regulating the specific immune response.
- ***Tumour necrosis factor***: These chemicals released from activated macrophages, are toxic to tumour cells, as discussed earlier. Two major types: TNF-α (MW 17 kDa) and TNF-β (MW 25 kDa) are known. Though they are cytotoxic to normal tissues also, their effect is limited due to their short half-life.
- ***Chemokines***: These are relatively recently discovered family of mediators that bring about chemokinesis (i.e. movement in response to chemical stimuli). IL-8 is emerging as an extremely important neutrophil chemotactic agent in vivo. Receptors for these mediators appear to act as co-receptors for the infection of lymphocytes by human immunodeficiency virus.

Cytokines are antigen non-specific glycoproteins or peptides that mobilize the entire protective system. They also regulate several biological processes, e.g. cell-growth, repair, morphogenesis, etc.

Cytokines **are more practically classified based on the principal effect or role they mediate**:

1. *Pro-inflammatory cytokines:* Tumour necrosis factor (TNF)-α, IL-1, IL-6, IL-8 and other chemokines, IL-12 and IL-15.
2. *Anti-inflammatory*: Transforming growth factor (TGFβ) and IL-10.
3. *Immuno-stimulatory cytokines:* For cellular responses IL-2 and IFNγ; and for humoral responses IL-4, IL-13, TGFβ and IL-10.

Note: *Arachidonic acid metabolites*, such as prostaglandins and leukotrienes form another group of inflammatory mediators.

C. Molecules Involved in Antigen Recognition

Specific receptors present on T and B cells are involved in recognition of the antigen. As indicated, *the T-cell receptors (TCR) are heterodimeric polypeptide chains and the B-cell receptors are membrane bound antibodies or surface immunoglobulins* (IgM in monomeric form) (Fig. 33.6). These receptors are too small to recognize antigens in their entirety, which are generally very large and complex. They recognize discrete sites on antigens (i.e. **antigenic determinants**, or **epitopes**), which are the immunologically active regions on a complex antigen.

Although both T and B-cell receptors belong to the same immunoglobulin super-gene family and exhibit several structural similarities in terms of folds and domains, there are some significant differences also.

The important **similarities** are: (a) both have hand-in-glove relationship with the corresponding antigenic element, (b) both have a limited number of genes that rearrange to form them, and (c) both show marked variability in the sequence of amino acids that actually come into contact with the epitope (the rest of the molecule being relatively constant in their amino acid sequence).

The important **differences** between the two types are that (a) TCR has two chains rather than four, and (b) TCR recognizes an epitope only when it is displayed together with molecules of the major histocompatibility complexes on self cell, whereas B-cell receptors recognize free antigen.

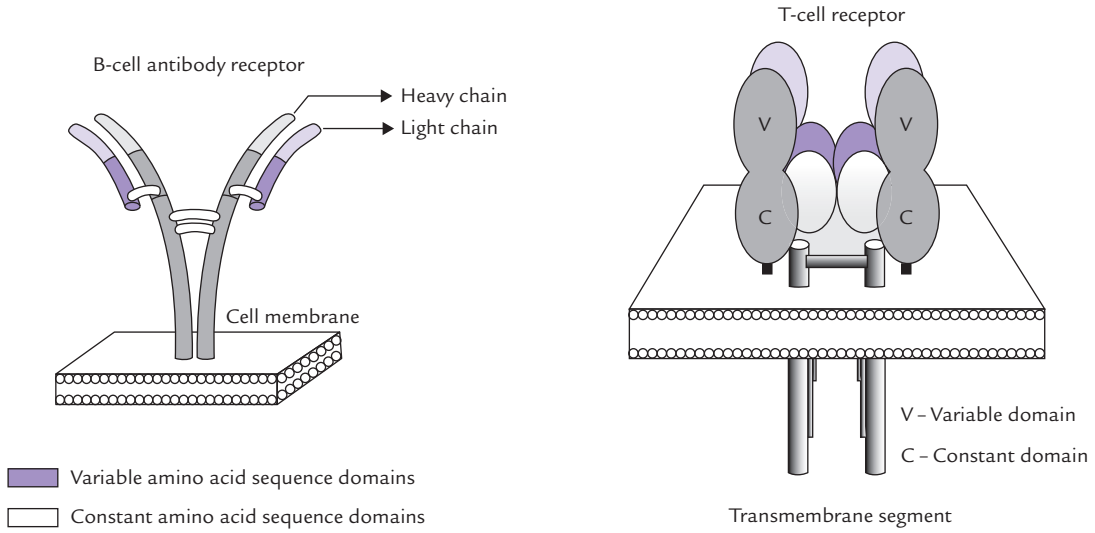


Fig. 33.6. Structures of receptors on B-cell and T-cell.

In general, four related but distinct cell membrane molecules are responsible for antigen recognition by immune system: (a) the T-cell antigen receptors, (b) the B-cell receptors, (c) class I MHC molecules present on all nucleated cells, and (d) class II MHC molecules present on antigen-presenting cells.

T-cell receptors: Four types of polypeptide chains (α, β, γ and δ) make up the T-cell receptors (TCR) in two functional combinations (αβ and γδ). Each chain comprises one constant and one variable domain and a C-terminal transmembrane region. The domains in each chain are organized into folds. The chains are joined by a disulphide bridge. There are gene-families encoding the two chains; these are segmented in the germ line and undergo developmental diversification.

The antigen-binding site of the TCR is in the cleft formed by the adjoining single variable domains of the constituent alpha (V_α), beta (V_β), gamma (V_γ), or delta (V_δ) chains.

B-cell receptors: It is the membrane form of the immunoglobulin serving as receptor for antigen. These surface immunoglobulins (**sIgs**) are Y-shaped molecules made up of four polypeptide chains—a pair of heavy chains, each of approximate molecular weight of 50 kDa and a pair of light chains each of approximate molecular weight of 23 kDa. There are areas of constant and variable sequences of amino acids in both the heavy and the light chains. The variably sequenced amino terminal domains of both the heavy (V_H-variable heavy) and the light (V_L-variable light) chains form a pocket. The latter constitutes the antigen binding site, termed the fragment antigen binding (Fab) portion sites, at the end of the Fab sites. The remaining relatively constant amino acid sequence domains of the chains, termed constant heavy (C_H) and constant light (C_L), form the stem that provides transduction effects.

Note: Unlike B cells which secrete immunoglobulins (the secreted form are B-cell receptors), no secreted version of the TCR is made.

MHC I and MHC II molecules: These are expressed respectively on all nucleated cells and on antigen-presenting cells, and are encoded by different loci within the MHC complex. *The T-cell receptors can interact with antigen only in association of MHC molecules*. The T-helper cells generally recognize antigen expressed together with the class II MHC molecule (class II restricted) and the T-cytotoxic cells antigen together with class I MHC molecules.

T and B cells recognize enormous number of possible antigenic configurations and also account for specificity of immune response.

VIII. Immunoglobulins

Immunoglobulins are the secreted forms of the B cell receptors, which react specifically with the antigen that stimulated their production. They are directed at extracellular infections, especially bacteria and their products, the extracellular phase of viral infection and individual cell transplantation. There are three ways by which they participate in host defense:

1. ***Neutralization:*** Immunoglobulins bind to and neutralize a bacterial toxin, preventing it from interacting with host cells and causing harmful effect. Unbound

toxin reacts with receptors on the host cell, whereas the toxin–immunoglobulin complex cannot. Immunoglobulins also neutralize complete viral particles and bacterial cells by binding to them and inactivating them.

2. ***Opsonization:*** Antibodies can coat a foreign antigen, and this makes the latter an easier target for phagocytes (macrophages and polymorphonuclear leukocytes). In this process, called opsonization, the antibodies coating an antigen render it recognizable as foreign by phagocytes (Chapter 5).
3. ***Complement activation*:** Antibodies coating a bacterial cell form a receptor for the complement system. The latter eventually forms a protein complex on the surface of the bacterium that favours its uptake and destruction by phagocytes and can, in some cases, directly kill the bacterium. Thus antibodies target pathogens and their products for disposal by phagocytes.

In view of pivotal role played by antibodies in body's defense, it is evident that primary antibody deficiency results in hazardous consequences **(Case 33.3).**

Functions of individuals classes of antibodies and their structural features has been dealt with in detail in Chapter 5. Some additional information about immunoglobulins is presented here.

Antibodies illustrate excellent diversity of the immune system: An animal can synthesize a limitless number of antibodies, each one with the ability to interact with a specific antigen, and together these millions of antibodies have the ability to recognize any and all nonself elements they come into contact with. How can the genes for some millions of different antibodies find space in the human genome. The mechanism by which such diversity arises from a basic template presents a good example of special genetic mechanisms, like DNA rearrangement and RNA splicing, being used for diversity (Chapter 24).

For each type of immunoglobulin chain, i.e. kappa light chain (κL), lambda light chain (λL) and the five heavy chains (αH, γH, μH, εH and δH), there is a **separate pool of gene segments in germline located on different chromosomes** (genes for κL, λL and five heavy chains are on chromosomes 2, 22 and 14 respectively). Each pool contains a set of different genes for the variable and constant domains. During development of the B lymphocyte, these separate genes are combined into a single transcription unit that codes for a complete L or H chain.

Theoretically, a thousand different L chains and a thousand different H chains, encoded by a total of 2000 genes, would be sufficient to make a million different antibodies.

Assembly of light chain (κL) gene (Fig. 33.7): Most of the variable domain (reaching from amino acid 1 to 95) is encoded by **V genes** (*V* for variable), and the rest of the variable domain (amino acids 96–108) is encoded by ***J* genes** (*J* for joining). The constant region (starting from amino acid 109) is encoded by a **single C gene** (*Cκ*). The germline contains a large repertoire of *V* and *J* genes: about 50 *V* genes and 5 *J* genes are present.

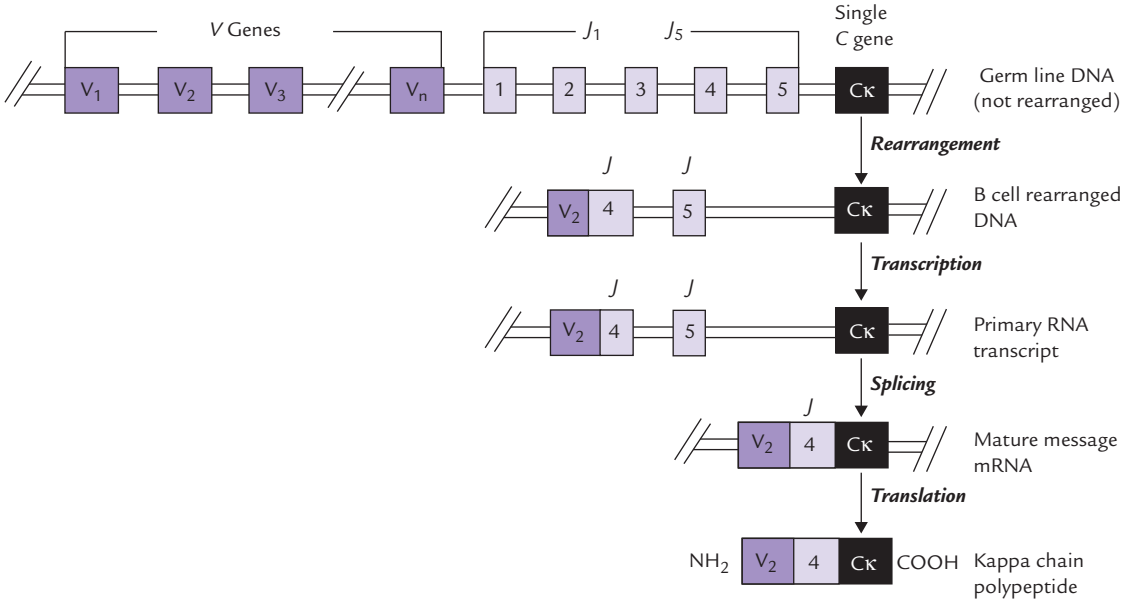


Fig. 33.7. Genetic events leading to assembly of a light chain gene. The kappa light chain (κL) shown here. It is encoded by separate gene clusters. During development of B-cell, these separate genes are combined into a single transcription unit, that codes for a complete light chain.

In course of B-cell differentiation, one of the *V* genes is selected at random from the library of *V* genes and spliced to the *J* genes to form a complete variable domain gene ($V + J$). Thus, two different segments are united into one functional gene by DNA rearrangements. This variable domain gene is then transcribed together with the constant domain gene. Transcription of the entire sequence of V + J + CR—one *V* gene spliced to one *J* sequence, together with the single *Cκ* gene—yields a large primary transcript. These events, occur in the nucleus, as does the splicing of the RNA transcript, from which all the introns and the unused *J* chains are excised, and the remaining coding sequences are then joined into a continuous strand of mRNA. The latter is then transported into the cytoplasm where it is translated into the κ-chain polypeptide.

Assembly of the heavy chain gene (Fig. 24.12): The heavy chain genes are assembled in the same way but the heavy chain gene cluster contains a set of approximately 30 *D* genes (*D* for diversity), in addition to the *V, J* and C_H genes. During B-cell differentiation, the first translocation brings a *V* region close to a *D* segment and *J* segments to form a complete variable domain gene of the heavy chain (V_H): $V + D + J$. This V_H is then assembled with the *C* gene and transcribed.

Thus, it is observed that *L* and *H* genes of virtually unique structure are constructed by a recombination process that randomly selects one out of each set of gene segments and assembles them with *C* gene.

The mechanism which permits the assembly of a very large number of genes depends on **(a)** a large repertoire of gene segments, **(b)** their rearrangements into different sequences, i.e. somatic recombinations, and finally **(c)** the combining of different L and H chains in the assembly of immunoglobulin molecules.

Immunoglobulin class switching (isotype switching): During its development, the B cell is able to change the class of its antibody without changing the antigen-binding specificity. Initially all B cells carry IgM specific for an antigen and produce IgM antibody in response to exposure to that antigen. IgM often is followed by IgD and eventually IgG, IgA or IgE may appear. This phenomenon is known as class switching and takes place either before or (more commonly) after exposure to antigen. In class switching, same assembled V_H gene can sequentially associate with different C_H genes so that the immunoglobulins produced later (IgG, IgA, IgE) have the same specificity as the original IgM but have different characteristics. The process is stimulated by lymphokines.

Another example of 'switching' is seen in haemoglobin synthesis, as described in Chapter 24.

Monoclonal antibodies and hybridoma technology: Administration of a protein or polysaccharide to an animal (that is of a species unrelated to the source of the immunogen) results in production of heterogenous antibodies. This is because the antibodies are formed by several different clones of cells, i.e. they are polyclonal. A large protein immunogen usually has several antigenic sites (i.e. epitopes), each of which gives rise to a distinct antibody, hence the name **polyclonal antibody.**

In contrast to polyclonal antibodies, a **monoclonal antibody** reacts with only a single epitope. The monoclonal antibodies arise from a single clone of cells, e.g. in a plasma cell tumour (myeloma) and are homogenous. The monoclonal antibodies can be synthesized in the laboratory by fusing a myeloma cell with an antibody-producing cell. Such *hybridoma cell* produces virtually unlimited quantities of monoclonal antibodies that are useful in diagnostic tests and in research.

Steps in synthesis: The monoclonal antibodies-synthesizing-hybrid-cells are made in the following manner:

1. An animal, e.g. mouse is injected with the antigen of interest. The antigen bears more than one epitope.
2. The spleen of the mouse is removed and the spleen cells are grown in a culture dish.
3. The mouse myeloma cells are added to the culture dish. These cells have an important attribute of immortality: they have **unlimited lifespan** in culture. They lack the salvage enzyme, *HGPRT*.
4. The myeloma cells are fused with the antibody-secreting (spleen) cells. The fusion is promoted by adding certain chemicals, e.g. polyethylene glycol. The hybrid cells now contain the gene of normal mouse as well as of myeloma cells.
5. ***Selection of hybrid cells*** is now carried out in HAT medium. By culturing in this medium (HAT = hypoxanthine-amethopterin-thymidine), the hybrid cells are separated from the other two cell types, i.e. the unfused myeloma cells and the antibody-producing cells.
   (a) The unfused myeloma cells do not survive in HAT because they can not salvage hypoxanthine (due to lack of *HGPRT*) and the *de novo* synthesis is blocked by amethopterin.
   (b) The antibody secreting (spleen) cells do not survive long as they have limited lifespan.

Thus, after sometime only the hybrid cells are left in the medium. *They have acquired the property of immortality from myeloma cells and the property of secreting a particular antibody from the antibody-secreting cell.*

The resulting clones are screened for the production of antibody to the antigen of interest.

Monoclonal antibodies, the single antibody species, are produced by fusion of spleen cells from an immunized mouse with a mouse myeloma cell-line. The hybridoma cell-line so produced is cloned to produce a cell-line that secretes monoclonal antibodies indefinitely.

Monoclonal antibodies were first produced by Georges Kohler and Cesal Milstein in 1975, and they were awarded Nobel Prize in 1984.

Technically, production of human monoclonal antibodies is also possible. This is accomplished by fusion of human lymphocytes with human plasmacytoma cell lines.

Uses of monoclonal antibodies: Being more specific, the monoclonal antibodies bind antigen with increased avidity. Therefore, smaller quantity of monoclonal antibody is required. They target an antigen specifically, and so the reaction is more specific.

Chimeric monoclonal antibodies consisting of mouse variable regions and human constant regions are being made for use in treating human diseases such as leukaemia. Presence of the human constant region gives following advantages:

1. Activation of complement is possible with the human-derived constant region (whereas it is not if the constant region is mouse-derived).
2. Antibodies are not formed against the chimeric monoclonal antibody (whereas antibodies are formed if the constant region is mouse-derived).

Lately, chimeric antibodies have been found useful for eliminating tumour cells by complement mediated cytotoxicity.

IX. More about Complement Cascade

The complement cascade is an amplifying cascade, similar to those responsible for blood clotting and fibrinolysis. As discussed earlier in Chapter 5, the complement system comprises a group of about 20 proteins, some of which are enzymes. The complement cascade involves stepwise activation, via proteolysis of these proteins till the formation of a lytic complex. Further information about complement components, activation pathway, general principles of reactions, and their clinical significance will be discussed in the present section. Generally, the term "complement" refers to the ability of this effector mechanism to complement, i.e. augment the effects of other components of the immune system, e.g. antibody. In fact, activation of complement system (a component of non-specific immune response) is one of the most important antibody effector functions of the specific immune response.

Activation of the complement cascade: This can be triggered in two ways: (a) by antigen-antibody complexes (i.e. the *classical pathway*), or (b) by bacteria in combination with several serum factors. (i.e. the *alternative pathway*). Both these lead to a common sequence of events, which result in formation of a lytic complex called **membrane attack complex.** The latter attacks the membrane with the help of tiny tubules, which results in appearance of holes in the cell membrane, and cell death ensues. In addition to the *lytic effect*, the complement system initiates specific cellular functions that mediate *inflammatory response*. These functions include attracting neutrophils and macrophages to an area where the foreign material is located, promoting phagocytosis and opsonization, etc.

Note: In addition to the classical and the alternative pathways, a mannose-binding ligand (MBL) pathway to complement activation has also been defined, which, like the alternative pathway, is triggered directly by mannose, found in cell walls of fungi, bacteria and viruses.

Complement components: Approximately 20 proteins, that are present in normal human serum, comprise the complement system. They act in concert, and in an orderly sequence, to generate biologically active molecules, such as enzymes, opsonins, anaphylatoxins and chemotaxins. The complement components are designated as C1, C2, C3, C4, etc. but they do not act in the same sequence as their identification numbers. Most components are β-globulins made in liver, and have *proteinase* activity when activated.

Sequence of events in complement activation: An outline of the important stages of complement activation is shown in the Figure 33.8. The *activation of C3, the pivotal and critically important component, is an absolute requirement for full complement activation*. There are three possible pathways to this activation: classical, alternative and MBL.

The **classical pathway** is triggered when immune complexes (IgG or IgM bound to its specific antigen) attracts the complement component C1 (made up of C1q, Clr and Cls). The latter binds to the constant region of the immunoglobulin and gets activated. The activated form is depicted as $C1\overline{qrs}$: the horizontal bar denotes activated form of a complement fragment. The activated C1 acquires proteolytic activity. It in turn cleaves and activates the next component, C4, splitting it into C4a

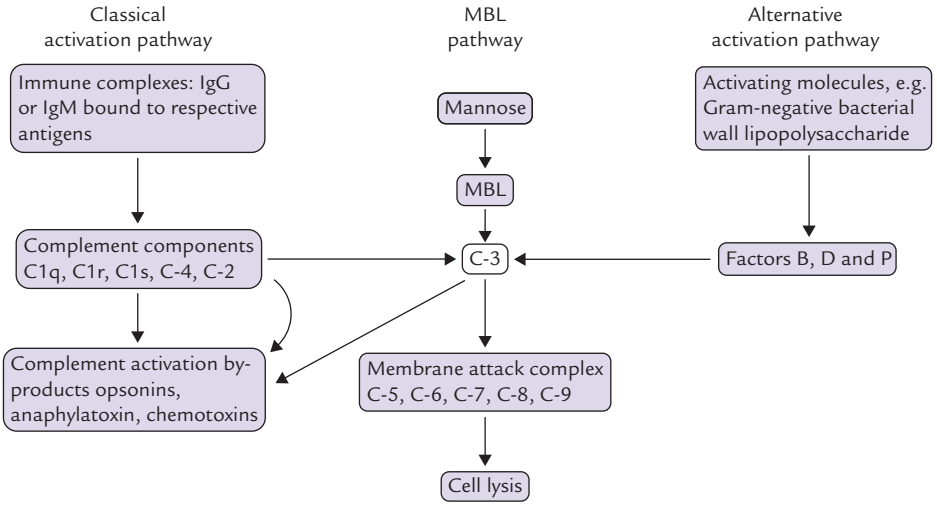


Fig. 33.8. Complement activation pathways (MBL = mannose-binding ligand).

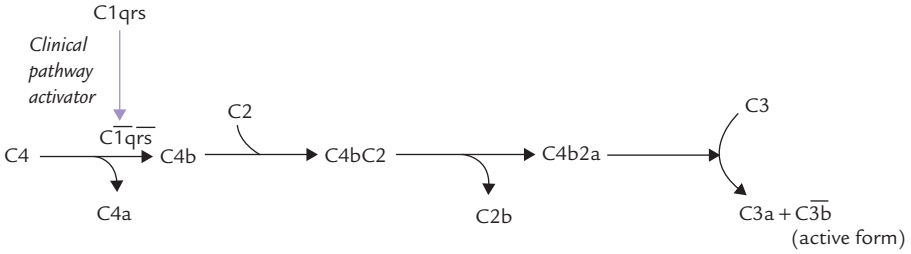


Fig. 33.9. C3 activation by classical pathway.

and C4b. The latter then binds with C2 and cleaves it into C2a and C2b. C4b and C2a then act together on C3 and split it into C3a and C3b. The latter is the activated form of C3. These events are illustrated in Figure 33.9.

In the alternative pathway (and the MBL pathway), activation of C3 is achieved by materials such as bacterial cell walls and endotoxins. In their presence, C3 is slowly hydrolyzed to C3a and C3b (other proteins, e.g. factor B, P and D also participate in the activation). The alternative pathway may therefore, be particularly relevant before a primary immune response has been mounted.

Once activation of C3 is achieved, the terminal membrane attack complex, which comprises the components C5, C6, C7, C8 and C9 is generated. This complex eventually generates the polymeric ring structure that inserts into the cell membrane of bacteria and is responsible for cell lysis.

Biological activities of complement activation by-products: A number of biologically active molecules are generated during complement activation. Some of the important biological activities of these are as below:

1. *C3a*: Smooth muscle contraction, increase in vascular permeability, platelet aggregation.
2. *C3b*: Opsonization and phagocytosis.
3. *C4a*: Smooth muscle contraction and vascular permeability increase.
4. *C5a*: Smooth muscle contraction, vascular permeability increase, platelet aggregation, polymorph and monocyte chemotaxis, neutrophil hydrolytic enzyme release.

Control: Various specific inhibitors, e.g. C1 inhibitor, control the complement pathway meticulously and tightly. Moreover, a number of activated components being inherently unstable decay fast, so their action persists for a limited duration.

X. The Recognition of and Response to Nonself

The challenge for the immune system is to be able to provide protection against the antigenic elements that arise as a result of the presence of an immunogen, or against foreign cells (both known as **nonself**) that invade the body. It must not react in this way with the body's own proteins (known as **self**). Thus, a nonself structure

must be recognized effectively for mounting an immune response against it.

Generation of specificity and diversity: Attributes of specificity and diversity that characterize all immune responses, apply to both B and T cells. In a given B-cell, for example, a unique specificity is generated by a combinatorial process, which involves *random rearrangements of a series of gene segments encoding the sIg molecules*. As a result, the B cell comes to express a single gene arrangement for the antibody's heavy and light chains (Fig. 33.7), and so *each B lymphocyte synthesizes antibody molecules of a single specificity, some of which are displayed as receptors on its membrane*. Fine specificity of the antibody molecule is reflected in the incredible precision by which it can discriminate the protein antigens that may differ by only a single amino acid.

The distinct specificity of the antibody molecule is coupled to an *enormous diversity*—estimated to exceed 10^8 different antibody specificities. Diversity is also accounted by gene rearrangements, discussed above.

Clonal selection: How do antibodies arise? Does the antigen "instruct the B cell to make an antibody, or does the antigen "select" a B cell endowed with the pre-existing capacity to make the antibody. It appears that the latter alternative, i.e. clonal selection accounts for antibody formation. As discussed, the *specificity of each T and B cell is determined prior to contact with the antigen:* each lymphocyte expresses a unique receptor specificity, which is determined prior to the appearance of an antigen.

The antigen comes into picture only when it interacts with a (B or T) cell, which carries a matching receptor on its surface. After the antigen binds, cell is stimulated to proliferate and form a clone of cells, each with the same immunologic specificity as the original parent cell and the process is termed **clonal selection.**

The clonal selection is so called because an immunocompetent cell with a given specificity is selected for expansion, rather than induction of de novo synthesis of new antibody-secreting cells.

Clonal selection occurs within both the humoral and cell-mediated branches of the immune system. In the humoral branch, antigen induces clonal proliferation of the antigen-reactive B cells into a clone of memory B cells and effector cells, called *plasma cells*. The plasma cells secrete antibodies reactive with the activating antigen.

Similar process occurs in T-lymphocyte population where antigen-MHC complex induces clonal proliferation into T-memory and effector cells (Fig. 33.2).

Clonal selection provides framework for understanding three aspects of specific immunity: specificity, memory and self/nonself discrimination.

Specificity is shown because only those lymphocytes are selected and clonally expanded which possess a receptor specific for the antigen.

Memory is accomplished because some of the clonally expanded cells, which have a longer lifespan, continue to express the membrane antibody of the same specificity as the parent cell.

Self/nonself discrimination is accomplished because of clonal elimination, meaning that formation of the cells reacting with self antigen is halted.

XI. Cellular Cooperation in the Immune Response

For eliciting an effective immune response, interaction between various cell types is required during various events discussed below.

Antigen Processing and Presentation

Antigen is too large to be recognized in its entirety by either T- or B-cell receptors, therefore it is essential that antigen is processed and presented with a class II MHC molecule on the membrane of an antigen-presenting cell. Macrophages, interdigitating cells, dendritic cells and Langerhans cells are the major types of antigen presenting cells.

Activation and Proliferation of T-helper Cells

The antigen-dependent activation of T-helper (T_H) cells is the central event in both cellular and humoral immune response. In fact, activation of the naive T_H requires **two signals**, (i) binding with antigen-MHC complex, presented on surface of antigen-presenting cell (APC), and (ii) stimulation by IL-1.

As shown in Figure 33.10, binding of an antigen-MHC complex to a naive T_H causes the APC to secrete interleukin-1 (IL-1). The T_H cell is then induced (by IL-1) to undergo transition from G_0 to G_1 of the mitotic cycle and to secrete another cytokine, interleukin-2 (IL-2). The activated T_H cell now starts expressing surface receptors of IL-2. The IL-2 binds to the receptor protein and induces further proliferation of T_H. Interestingly, expression of IL-2 receptors is induced by IL-2 itself, which thereby acts in an autocrine fashion. A **positive feedback loop** is set up, which leads to a burst of proliferative activity of the T_H cells. Pivotal role of IL-2 in this process has important clinical applications (Box 33.6).

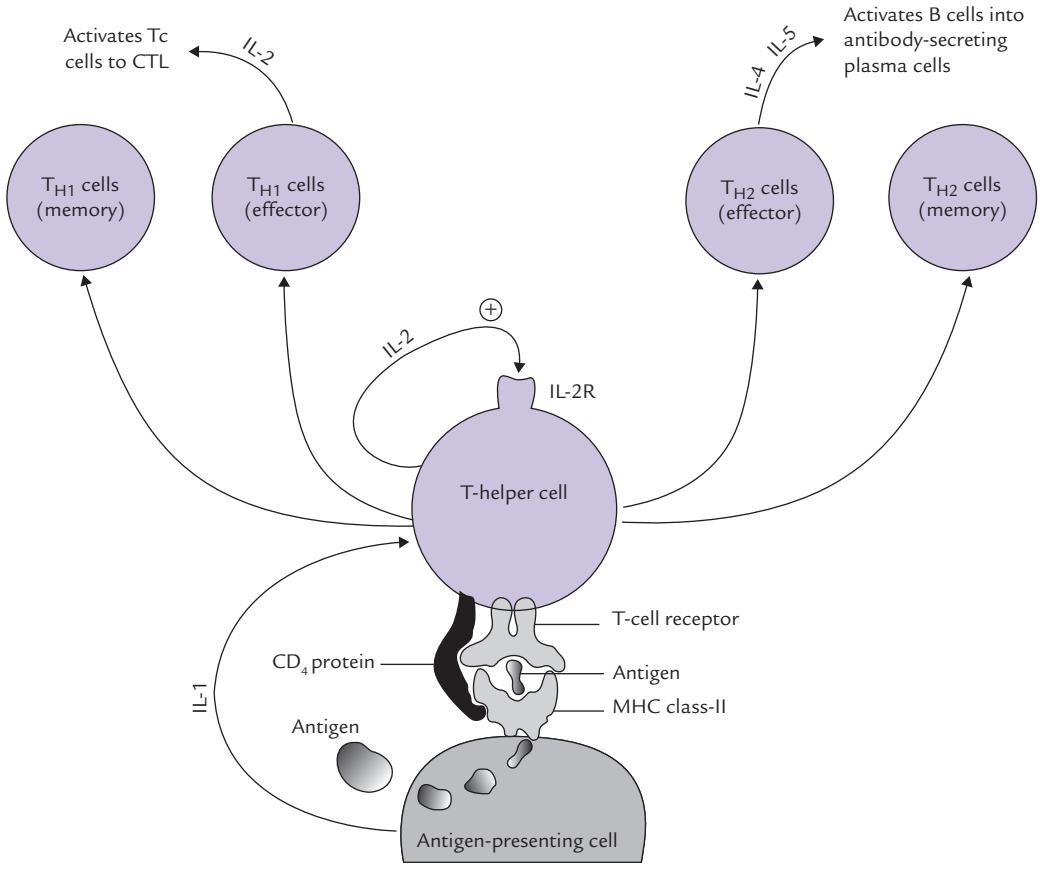


Fig. 33.10. Activation and subsequent effector roles of T-helper cell (T_H) (IL-1 = interleukin, IL-2R = receptor for interleukin-2, CTL = cytotoxic T-lymphocyte).

BOX 33.6

Elucidation of IL-2 Receptors has Helped Understanding of Certain Medical Facts

Ability of IL-2, secreted by activated T-helper cell to interact with the IL-2 receptor on its own surface, thereby establishing a positive *feedback loop*, is an essential feature of all immune responses. This observation is reinforced by two recent findings. First, individuals with the disease, severe combined immunodeficiency (SCID), have inherited a gene that encodes a defective subunit of the IL-2 receptor. The same protein also forms part of the receptors for IL-4 and IL-7, thus explaining why both B and T cells are defective in this disease (the maturation of T cells requires IL-7 and the differentiation of virgin B cells is prompted by IL-4). This also explains severity of the disease. Second, immunosuppressive drugs, such as cyclosporin and FK506, which inhibit the enhanced production of IL-2 and its receptor by T_H cell, find useful for preventing the rejection of grafts (since they suppress immune response against the transplanted tissue, recognized as nonself after the transplant surgery).

The expanded population of the T_H cells develops one of the **two phenotypes** (T_{H1} and T_{H2}) characterized by the cytokines produced. The T_{H1} cells continue to secrete IL-2, and also begin to release interferon gamma, while T_{H2} cells switch to the production of IL-4 and IL-5. These changes relate to different functions as:

- T_{H2} cells promote humoral immunity (because IL-4 and -5 activate B cells).
- T_{H1} cells promote cellular immunity (because IL-2 activates the cytotoxic T cells).

T cells are involved in many aspects of cellular immune response. They influence activity of several components of innate immunity also.

Generation of the Humoral Response

The naive B cell captures antigen with its surface immunoglobulin (sIg) as discussed earlier (Fig. 33.2). This is important for the activation of the B cell. In addition, IL-4 and

IL-5 secreted by T_{H2} cells are also required for activation. The following series of events are known to occur:

1. The antigen bound to its surface immunoglobulin receptor (sIg) is endocytosed by the B cell. Intracellularly, it is chopped into peptides, which are then represented on the cell membrane together with a class II MHC molecule (Fig. 33.11).
2. The antigen/MHC complex is then recognized by the antigen-specific T_H cells (in this way, the B cells themselves serve as antigen-presenting cells) and the interaction prompts the clonal expansion and activation of this T_H into cytokine-secreting T_{H2} cells.
3. The cytokines IL-4 and IL-5 secreted by T_{H2} then activate the B cell. (The IL-1 secreted by macrophages also serves as a growth factor for B cell).
4. The activated B cell then undergoes a series of cell divisions over approximately a 5-day period, ultimately resulting in its transformation into a population of both antibody-secreting plasma cells and memory cells.

The requirement of lymphokines-induced B-cell proliferation helps explain why it takes 5 days for a primary immune response to be established. In secondary immune response there is advantage of having memory T_H cells, which may initiate the response straight off.

The activated and clonally expanded population of T_H cells secrete interleukins (-4 and -5) which transform the B-cells into antibody secreting plasma cells. Memory cells arise possibly by partial triggering of the B-cells.

Generation of the Cell-mediated Response

As in case of humoral response, an activated and clonally expanded population of antigen-activated, T-helper cells is crucial for generation of the cell-mediated response. Macrophages, which express the processed antigen in association with MHC II on their cell surface, serve as antigen-presenting cells (APCs) in the cell-mediated response. The T-helper cells recognize the processed antigen + MHC II, get transformed into effector T_{H1} cells, which secrete interleukin-2 (Fig. 33.10). The latter acts on the cytotoxic T cells (Tc) and transforms them into cytotoxic T lymphocytes (CTLs), which attack the infected self cell, e.g. virus infected cell (recall that such infected cell would present viral antigen along with MHC class I molecule on its cell surface).

The cell-mediated response to a virus infected cell is shown in Figure 33.12.

The foregoing discussion makes it clear that Tc cells must receive at least two signals before they are activated to CTL: **(a)** they must be bound to the infected target cell via the MHC class I/processed antigen TCR complex (signal 1), and **(b)** they must be activated by IL-2 (signal 2) secreted by the T_{H1} cells.

The targeted cell is then killed by the secretion of several destructive proteins from CTL, the chief of which are the perforins and granzymes. The former undergo calcium-dependent polymerization in the extracellular fluid and this allows them to form channels in the

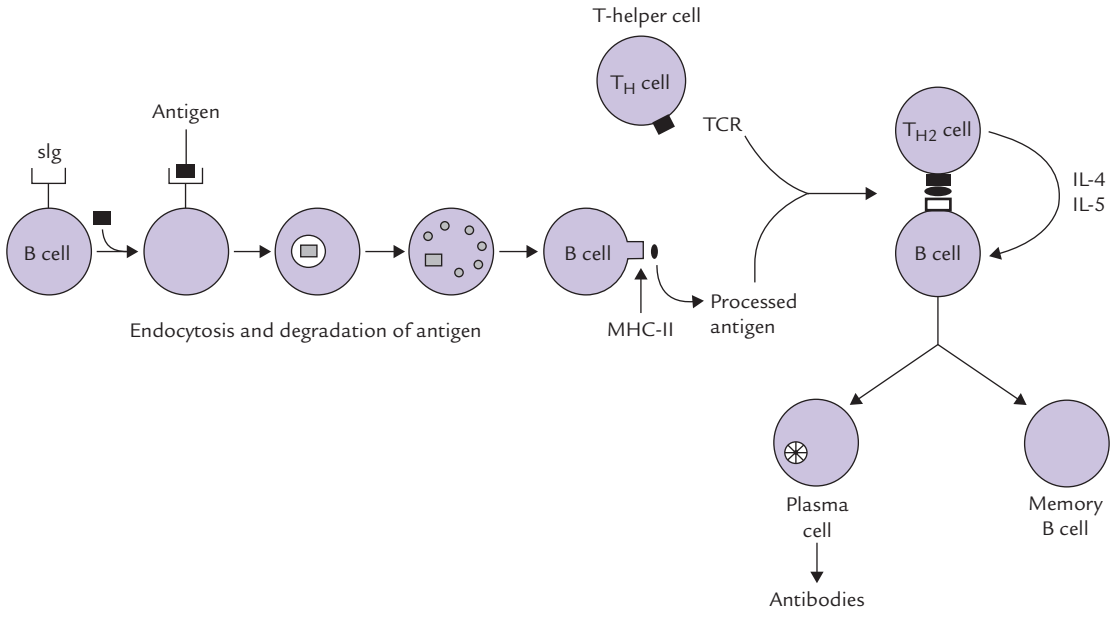


Fig. 33.11. Steps in conversion of naive B-cell into the antibody secreting plasma cell and the memory B cell (sIg = surface immunoglobulin, IL = interleukin, MHC = major histocompatibility complex, TCR = T-cell receptor).

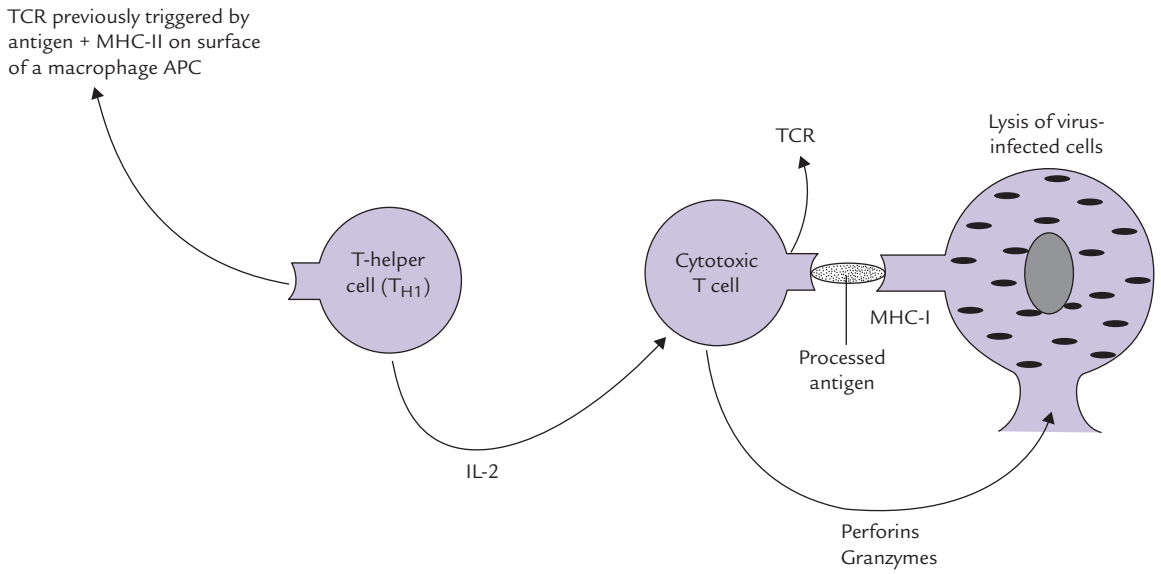


Fig. 33.12. Cooperation between T-helper cell and cytotoxic T cell in causing lysis of virus-infected cell (IL = interleukin, MHC = major histocompatibility complex, TCR = T-cell receptor).

plasma membrane of target cells. This in turn provides a route by which the proteolytic granzymes gain access to cells and destroy them.

Cytotoxic T lymphocytes kill cells that carry a recognizable antigen, i.e. the antigen for which killer cells carry a receptor.

XII. Non-specific Effector Cells are Influenced by T_H Cells and Immunoglobulins

T_H Influences activity of various non-specific effector cells: Activities of various non-specific effector cells, such as macrophages and natural killer cells, are enhanced by cytokines secreted by T_H cells. These cells in turn play important roles in cell-mediated responses. Following examples will make this point clear.

1. Cytokines like interleukin-2 and interferon gamma, secreted by T_H have been shown to activate macrophages, which in response exhibit enhanced ability to phagocytose the invading pathogens.
2. The natural killer cell is another non-specific effector cell whose activity is enhanced by interleukin-2 and gamma interferon. Natural killer cells, the large lymphocytes that lack the specific antigen-binding receptors of either B or T cells, is shown to kill tumour cells also.

Immunoglobulins provide specificity to non-specific effector cells: Both macrophages and NK cells have membrane receptors that can bind the Fc portion of the immunoglobin molecule. The Fab portion of the immunoglobulin can bind with the foreign cell. Thus an antibody can bind both to the foreign cell and the macrophage/NK, thereby acting as a bridge.

Simply stated, the C-terminal provides specificity and the N-terminal engages a non-specific cell which mediates the actual killing. For example, if immunoglobulin forms a complex with the corresponding antigen on surface of a solid particle (e.g. a virus or a bacterium) the antibody-coated particle can be phagocytosed by macrophages (or neutrophils). Such stimulation of phagocytosis, called **opsonization** has been discussed earlier in chapter 5.

Likewise, in antibody-dependent cell-mediated cytolysis (ADCC), the immunoglobulin bound to target cell acts as a tag for the natural killer cell. **ADCC and opsonization show that the innate and the adaptive immune responses interact in a cooperative manner.** More examples of such cooperation are illustrated in Box 33.7.

XIII. Specific Immune Responses to Some Infections

An interplay of both arms of specific immune system is often required for defense against the invading pathogen. However, the relative importance of humoral versus cell-mediated immunity varies with the type of infection. In infections with pyogenic bacteria, humoral immunity is more important; whereas protection against viral and intracellular bacterial infections is dependent on cell-mediated immunity. Individuals with antibody deficiencies

BOX 33.7

Interaction between Innate and Adaptive Immune Responses

Both B and T cells and their products are able to recruit and utilize components of the non-specific response in an effective and targeted manner, and converse is also true.

The mechanistic bases of interactions between the innate and adaptive arms of immune response have been described appropriately, but at different places, in this chapter. Some important examples are assembled and outlined as below:

1. Opsonization: Enhancement of phagocytosis by a variety of factors called opsonins that coat the antigen, thereby making it an easier target (Chapter 5).
2. Antibody-dependent cell-mediated cytolysis: Antibody (Fab) binding to a target microorganism or tumour cell while Fc is engaging a killer ADCC effector cell (e.g. the natural killer cell), thereby focussing killer on its target.
3. Complement activation: Complex formation between antigen, antibody and complement destroying the antigen.
4. Cytokines: The intercellular signalling molecules acting as regulators of cells that mount both innate and adaptive immune responses.
5. Cells of phagocytic system, most notably macrophages, take up foreign antigens, process them by denaturation or partial digestion, and present them on their surface to specific cells of adaptive immune response, leading to activation of the latter.

are therefore prone to recurrent infections with pyogenic bacteria and those with defective T-cell functions suffer from severe limitations in controlling viral infections. They also show increased susceptibility to mycobacteria. Since it is impossible to discuss immune response to all pathogens, two illustrative examples to discuss immune responses to mycobacteria and HIV are given, and immunopathogenesis of these infections outlined. (The undergraduates may skip rest of this section.)

A. Mycobacterial Infection

Mycobacterium tuberculosis is an obligate intracellular, slow growing bacteria, which produces no toxins and causes disease only by challenging body's immune defense mechanisms coupled with an inexorable increase in number. Around one third of the world's population is infected with this pathogen, which is responsible for causing 3 million deaths worldwide per year. Infection commonly occurs by inhalation. On entry into the body, mycobacteria are promptly taken up by macrophages, processed and presented to the T cells (CD_{4+} and CD_{8+}).

Presentation of mycobacterial antigens to T cells triggers clonal expansion and release of TH_1 cytokines, e.g. Y-IF, TNF and IL-2. This leads to macrophage activation and **granuloma** formation (granuloma is made up largely of macrophages, which have fused to form **giant cells** and also differentiated to epithelioid cells). These changes enable the immunocompetent individuals to contain and effectively eliminate the organism.

Mycobacteria have evolved clever strategies to evade the host immune response and survive inside macrophages, which thereby shelter bacteria and act as reservoirs of infection. Sheltered within macrophages, the mycobacteria can survive for several years. Development of active disease (tuberculosis) occurs either when the host response is attenuated or when exposure to certain risk factors occurs, or both. Some important risk factors are advanced age, alcoholism, drug abuse, malnutrition or impaired cellular immunity (e.g. by immunosuppressive therapy or HIV virus infection). The T-cell response is inadequate for the infection caused by the reactivated pathogen, and the granulomatous response is only partially able to contain it. An area of progressive granulomatous inflammation develops which slowly distorts the surrounding structures by fibrosis. In severe cases, macrophages are induced (by mycobacteria and by ineffective T-cell response) to persistently secrete cytokines, resulting in systemic ill health.

B. Human Immunodeficiency Virus (HIV) Infection

Disease associated with retroviruses HIV types 1 and 2 infection is characterized by major defects in immunity, and **selective elimination of CD_{4+} lymphocytes** and macrophages. Profound weakening of immune system because of depletion of the CD_{4+} cells leaves the patient vulnerable to a spectrum of disorders from a transient, acute glandular fever-like illness to life-threatening vascularized neoplasms, and severe, disseminated, opportunistic infections. Acquired immune deficiency syndrome (**AIDS**) is the final stage in the progression of infectious disease caused by HIV. It has been suggested that as many as 8–10 million persons have been infected with HIV, 70% of which are accounted by heterosexual transmission, with severe implications for the infants born to HIV-positive mothers.

HIV viruses enter susceptible cells though binding of viral envelope glycoprotein to specific receptors on the cell surface. Any cell bearing the CD_{4+} antigen is capable of being infected by HIV; typically these are T-helper cells, but some B cells, macrophages and glial cells of the central nervous system also express CD_4, even in low amount. HIV can also enter through binding of beta-chemokine receptor (CCR5) or by fusion between the viral lipid envelope and the target cell membrane. The RNA viral genome is copied into a DNA form by a viral enzyme termed *reverse transcriptase* and the latter is integrated into the genome of the infected cell, where it may remain latent. Once reactivated, the DNA is used as a template for the RNA required for virus production. New virions are produced at exceedingly high rate of 109–1010 virions per day. The host immune response finds itself unequal to the challenge posed by HIV because of the following reasons:

1. Precipitate fall in the absolute number of the T_H cells follows the HIV infection. This results in profound immunosuppression.
2. HIV has intrinsically unstable epitopes for antibody to bind to. It makes the production of functionally neutralizing antibody by B cell extremely difficult. The same applies to the T cells as well.
3. Exceedingly fast rate of replication of HIV is coupled with as many as 108 mutations per day.

Finally, laboratory evaluation of AIDS patients typically demonstrates lymphopenia, selective depletion of T lymphocytes, and marked reduction in T- to B-cell ratio. Serum immunoglobulin levels are often elevated, but specific antibody production in response to antigen challenge is deficient. Once AIDS has developed, prognosis is very poor but median survival varies with viral load (i.e. the HIV-RNA in the blood) at presentation.

XIV. Vaccines

Since the first use of vaccine by Jenner against small pox two centuries ago, deliberate stimulation of an immune response by immunization or vaccination has led to virtual elimination of several diseases like small pox, and has been one of the most effective measures in controlling infectious diseases. Yet suitable vaccines for a large number of infectious diseases like malaria, tuberculosis, AIDS, etc. are still to develop. Rapid advances in immunology and recombinant DNA technology are brought into use for the development of safe and effective vaccines and important breakthrough towards these goal have been made in recent years.

Vaccines are attenuated or killed organism, antigenic component of disease causing microbe, toxins produced by bacteria or DNA coding for an antigen which are deliberately administered to an individual to protect him from a host of infectious disease. The process of deliberate administration of a vaccine is known as vaccination. An ideal vaccine should mimic the immunological stimulus associated with natural infection, have no side effects, be readily available, cheap, easily administered and should provide long lasting immunity. There are certain immunological requirements of an ideal vaccine. It should activate antigen-presenting cells to initiate antigen processing followed by cytokines production and activation of both T and B cells, leading to formation of neutralizing antibody and development of memory cells responsible for resisting subsequent infections. Generation of T cells to several epitopes in order to overcome antigenic variation of pathogen and persistence of antigen on follicular dendritic cells in lymphoid tissue where memory B cells are recruited are the other characteristics of an ideal vaccine.

Vaccination depends on the integrated specific immune response.

Types of Vaccines

Inactivated or killed vaccine: Killed or inactivated microbes have been widely used as vaccines. Vaccines for cholera, pertussis, typhoid, polio (Salk), etc. are made out of killed bacteria or virus. Although these vaccines are safe, very often they fail to evoke immune response as observed during natural infection. Killed microbes cannot produce proteins in the cytosol and so antigen processing does not take place and the cytotoxic T cells are not generated by these vaccines.

Live attenuated vaccine: Organisms attenuated to decrease virulency either by growing in culture medium so that they replicate poorly in human cells or by recombinant DNA technology where the segment of gene for virulency is either deleted or mutated. These attenuated microbes on administration do not cause any disease, yet they are highly immunogenic because the replication of the living microbes provides the host with larger dose of antigen required for developing cytotoxic T cell memory through antigen presentation. BCG, polio (Salk), MMR (measles, mumps and rubella) vaccines are made out of live attenuated microorganism. In spite of the general acceptance of attenuated vaccine, it has some demerits as well. It may induce disease to immune compromised

hosts and also has the possibility of the virus strain to reacquire virulence.

Toxoid vaccine: Toxins released by microbes like *C. diphtheria, B. pertussis, C. tetani* are mostly responsible for the disease stage. Purified inactivated toxins are not very effective as vaccines. Mixed with suitable adjuvant (e.g. tetanus toxoid along with aluminium salt) they serve as effective vaccines. DPT triple vaccine is a mixture of diphtheria, tetanus and pertussis toxins where the latter also serves as an adjuvant.

Subunit vaccine: Vaccination with whole organism at times may elicit poor immune response and even hypersensitivity due to the presence of large number of antigens which are not associated with protective immunity. Identification and isolation of protective antigen from an organism and the use of these immunogenic subunits as vaccines is known as subunit vaccine. When such an immunogenic peptide is formed artificially by the use of recombinant DNA technique they are called recombinant vaccines. These vaccines are safe, and protective. Commercially available recombinant vaccines have been developed for hepatitis B.

DNA vaccine: Immune response can be induced by injecting DNA encoding a protective antigen and is known as DNA vaccine. A viral gene coding for a core or capsid protein is inserted into a plasmid. The plasmids are injected into the host cells from where the they migrate to the nucleus. Viral proteins are expressed in the cells producing both cell-mediated and humoral immune response.

XV. Immunological Techniques

A. Precipitin Reaction

When an antigen solution is mixed with a specific antiserum, they form an antigen-antibody complex which may precipitate out. The amount of precipitation depends on (a) relative concentrations of the antigen and the antibody, and (b) valency of the antigen. The concentration at which maximum precipitation occurs is known as equivalence point. The two common techniques which are based on precipitin reaction are immunodiffusion and immunoelectrophoresis.

In immunodiffusion antibody, mixed with agar is layered on a glass slide. A number of wells were cut on the gel and antigen at different concentration are added to it. Slowly, antigen diffuses from the well, its concentration falls and when it reaches the equivalence point it combines with antibody and forms precipitin ring. The diameter of the precipitin ring is proportional to the antigen concentration. This method is used routinely in clinical immunology for quantitation of immunoglobulins, complement components, Transferrin, C-reactive protein, α-fetoprotein, etc.

Electrophoresis of a mixture of antigen in gel and their identification by the use of specific antisera is the principle behind immunoelectrophoresis. In this technique, sample containing mixture of proteins is subjected to electrophoresis in agar gel. Proteins migrate to various positions in the gel. Now antisera to a specific protein is added into a horizontal trough made in agar gel. Development of precipitin arc identifies the presence of the protein of interest. A modified version of immunoelectrophoresis known as rocket immunoelectrophoresis is a quantitative technique which involves electrophoresis of antigen into a gel containing the antibody. The precipitin arc has the appearance of a rocket, the length of which is related to antigen concentration.

Antigen concentration can also be assayed by nephelometry. When an antigen is added to a solution of excess antibody, small aggregates of antigen-antibody complex (not precipitation which is formed when antigen and antibody are present at equivalence) is formed which creates a cloudiness or turbidity. This can be measured by forward scattering of a monochromatic incident light with the help of a nephelometer. Nowadays, nephelometry is replacing immunodiffusion for the measurement of immunoglobulins, CRP, C3, etc.

B. Agglutination

When an antigen is displayed of the surface of a large particle such as bacteria, RBC, etc., addition of antibodies cause antigen to clump or agglutinate. This phenomenon is known as agglutination and, is used in most qualitative serological assays. When RBC is the antigen, the term haemagglutination is used. Haemagglutination is the basis of determining blood groups.

C. Radioimmunoassay (RIA)

RIA is commonly used to measure the level of hormones in blood and tissue fluids. For estimation by RIA one should have a pure preparation of antigen or antibody. In RIA for antigen, pure antibody against it is radiolabelled usually with I125. The unlabelled antigen present in the sample is allowed to attach to a solid support such as plastic tube/wells of a microtiter plate. Now labelled antibody is added and allowed to react with the antigen

attached to the solid surface. Any unbound antibody is removed by washing. The amount of radioactivity retained in the tube/wells gives a measure of the amount of the antigen present in the sample.

D. Enzyme Linked Immunosorbent Assay (ELISA)

ELISA is used in clinical diagnostics for assaying hormones, cytokines, viral antigen/antibody, etc. In ELISA, an enzyme (usually *peroxidase*, *phosphatase*, etc.) is linked chemically to the antibody. Antigenic component to be assayed is allowed to attach to the solid surface of an ELISA plate. Now enzyme linked antibodies are allowed to bind to the antigen followed by washing of excess antibody. Addition of a suitable substrate is converted to a colour reaction product which can be read in an ELISA reader.

A modified form of ELISA known as sandwich ELISA is used to detect cytokines, etc. Instead of antigen being directly attached to ELISA plate, antigen specific antibodies are bound to the plate. These are able to bind antigen (component to be assayed) with high affinity. A separate enzyme linked antibody that recognizes a different epitope of immobilized first antibody is then used to detect bound antigen.

XVI. Immunologic Dysfunction

Immune system is mostly beneficial but some abnormalities in quality, quantity or direction of the response in certain situations may lead immunodysregulation, which may have hazardous consequences:

1. Too little response results in **immunodeficiency**.
2. Too much response results in **allergy**.
3. Misdirected response results in **auto-immunity**.

A. Immunodeficiency Disorders

These are an important group of immunologic dysfunctions where *a component of immune system, specific or non-specific, may be absent*. This makes the affected individual more susceptible to infections. The immunodeficiency may be **primary (inherited)** or **secondary (acquired)**.

The primary disorder may deplete humoral or cellular immunity, or both may be affected. In ***congenital agammaglobulinaemia***, an X-linked disease, the humoral immune response is impaired but the cell-mediated response is unaffected. An absence of B cells is compatible with a normal lifestyle so long as infusions of immunoglobulin G are maintained.

In ***DiGeorge syndrome***, cell-mediated response is impaired because the thymic processing of the T cells does not occur due lack of thymus. **Severe combined immunodeficiency disease** (SCID) severely impairs both humoral and cell-mediated responses because of *absence of functional T cells*. It is severe because it is fatal and it is combined because in humans B cells cannot function without help from T cells, so that even if the B cells are not primarily affected by the defect, humoral response is impaired.

B. Allergy

When immune response results in exaggerated or in appropriate reactions harmful to the host, the term ***hypersensitivity*** *or allergy* is used. Clinical manifestations of these reactions are typical in a given individual and occur on contact with the specific antigen to which the individual is hypersensitive. The first contact of the individual with the antigen sensitizes, i.e. induces the antibody, and then subsequent contacts elicit the allergic response.

Tendency to develop allergic response is inherited by some individuals. They have high level of **IgE** (also termed **reaginic antibody**) in their blood. This antibody mediates allergic reactions. It binds avidly to the surface of basophils and mast cells, where it acts as an antigen receptor. The binding of an allergen (antigen capable of generating allergic response) to surface IgE induces the release of stored histamine from the cell. Most of the allergic symptoms are mediated by the released histamine, and most antiallergic drugs act by blocking the effects of histamine on its target cells.

Hypersensitivity reactions can be subdivided into four major types: type I, II and III are antibody-mediated, whereas type IV is cell-mediated. These diseases are included in the discipline of clinical immunology and immunopathology. More information on these is beyond the scope of this book.

C. Autoimmunity

One particular aspect of immunodysregulatory disorders is that of self-reactivity (i.e. autoimmunity). Normally this is avoided because of intricate controlling and suppressor mechanisms. Included among such mechanisms is the process of thymic education by positive and negative selections. The other processes which eliminate the self-reactive clones are *clonal selection* and ***anergy*** (the T cells recognizing a self-antigen is likely to be rendered

incapable of further response; the process is referred to as anergy and constitutes an extension of self-tolerance initiated in the thymus). It is quite likely that an error may occur in one of these complex processes, which help a meticulous discrimination between the self and the non-self. The error may creep in spontaneously or be induced by some exogenous factor such as viral infection, often in a genetically predisposed individuals. The result will be an immune response directed against "self" and the inflammatory damage, and the ensuing condition is referred to as autoimmunity.

A wide spectrum of autoimmune diseases is known to range from the organ-specific to non-organ-specific. The reaction against ubiquitous antigens lead to non-organ-specific autoimmune diseases, whereas the reactions to unique components of individual tissues and organs result in organ-specific diseases. The former are best exemplified by **systemic lupus erythematosus** (SLE), in which the apparent target antigens are common to all nuclei. Consequently, damage is seen in a variety of tissues. The organ-specific disorders are exemplified by **autoimmune thyroiditis**, where the apparent target is thyroid *peroxidase*. Consequences may vary in severity from trivial to fatal.

There are a number of **diseases associated with particular MHC antigens**. For example, presence of DR2 is associated with increased frequency of *multiple sclerosis*, and DR3 or B8 are associated with *myasthenia gravis*, or *thyrotoxicosis*. Occasionally, the MHC status appears to be protective rather than associated with increased frequency of a disease. These associations are of great interest scientifically, but rarely of value in clinical practice.

Exercises

Essay type questions

1. Enumerate various types of lymphocyte populations and describe their roles in invoking specific immune response. Why are the T-helper cells said to occupy a central place in all immune responses.
2. Discuss the steps in synthesis of immunoglobulins by naive B cell following an antigenic challenge. Mention briefly the disorders related to immunoglobulins.
3. Discuss three ways by which immunoglobulins participate in the host defense. Explain the mechanisms by which human genome can direct synthesis of limitless variety of immunoglobulin molecules.
4. Explain with the help of suitable examples as to how innate and adaptive immune responses interact with each other. Why are both arms of immune responses—humoral as well as cell-mediated—affected in severe combined immunodeficiency disease?
5. What is a subunit vaccine? Discuss strategies involved in production of a bacterial vaccine.

Write short notes on

1. Clonal selection theory
2. Monoclonal antibodies and hybridoma technology
3. Cell-mediated immunity
4. Beneficial and harmful effects of immunity
5. Cytokines in immune regulation
6. Primary and secondary immune response
7. DNA vaccines
8. ELISA
9. Plasma cells

CLINICAL CASES

CASE 33.1 Tea coloured urine following treatment of burn wounds

A 13-year-old boy was brought to his general practitioner with burns in his left palm and forearm, after a fire-cracker accidently burst in his hand. He was promptly given first-aid including thorough cleaning and dressing of the burn-wounds and immunization against tetanus. The parents did not bring him to the doctor again and believed that the child was doing well with the home-remedies, when he abruptly developed high-grade fever and chills on the fifth day. The doctor noticed copious purulent discharge from the wounds together with signs of profound inflammation. Specimen for wound-culture and blood-culture were sent to the laboratory, and antibiotic therapy was started. Next day, the patient had to be rushed for hospitalization since the doctor observed signs of sepsis (cool, pale extremities with peripheral cyanosis, progressive hypotension, tachycardia and mental obtundation).

Reports of the wound-culture and blood-culture were meanwhile returned, and these showed growth of numerous bacterial colonies of *Pseudomonas aeruginosa*—the pathogen most commonly associated with wound infections—which were reported to be sensitive to the antibiotic being administered. Over the next week in hospital, the

patient showed remarkable symptomatic improvement: the blood culture became negative, the wounds healed significantly and the wound-cultures showed a progressive decline in bacterial colonies. However, he developed sore throat and inflammed cervical lymph nodes during his stay in the hospital. On examination, mucosa was reddened and tonsils were enlarged with white splotches on their surface. After starting apprepricate therapy for these, he was discharged with instructions to return for review after 5 days.

However, he did not report as instructed and was brought back after a fortnight when his mother noticed red colored urine. No residual abnormality in pharynx was detected, but urine examination showed haematuria and significant proteinuria. Further evaluation included estimation of antibodies to streptococcal antigens, specifically antistreptolysin and anti-DNAse B. High titres of these antibodies and hypocomplementaemia were demonstrated.

Q.1. Discuss immuno-pathogenesis of the signs and symptoms of the patient following the fire accident.
Q.2. Relate the series of events leading to inflammatory response after the burn injury.
Q.3. Discuss the changes in immune system leading to development of inflammatory response in tonsils and lymphadenopathy.
Q.4. Why did blood and proteins appear in urine? (Students are advised to attempt questions 3 and 4 only after the first reading of the entire Chapter).

CASE 33.2 Persistently pyrexial girl with repeated infections by unusual organisms

A 7-year-old girl was admitted to the hospital for investigation of pyrexia. She had history of recurrent episodes of fever and repeated infections (particularly bacterial and fungal) of skin, lung and lymph nodes, beginning during the first year of life. She was the fourth child of her parents; two elder brothers had died of infection at ages of 2 years and 3 years respectively, but parents and one sister were healthy. There was no family history of immunodeficiency. Past records of her numerous earlier hospital admissions for investigating her pyrexia showed that fever generally lasted from a few days to a few weeks, and was usually associated with severe respiratory infections and otitis media even by usually non-pathogonic organisms. On examination, she showed pallor with biaxillary and inguinal lymphadenopathy, and marked hepatosplenomegaly. There were multiple skin abscesses and small superficial ulcers in mouth. An extensive immunological evaluation was done; the most important findings were elevated WBC count (13.8×10^9/L), normal lymphocyte count (2.8×10^9/L), neutrophilia (9.7×10^9/L), and raised levels of serum immunoglobulins.

Quantitative serum immunoglobulins and some other lab reports are as below:

IgG 17.8; IgA 4.8; IgM 2.8 g/dL (normal values of IgG, IgA and IgM are 5.5–10.0, 0.3–0.8, and 0.4–1.8 g/dL, respectively)

- Haemoglobin 7.8 g/dL
- Abscess culture *Staphylococcus epidermidis*
- Urinalysis Normal
- Serum protein electrophoresis Normal
- Assays for complement, B cell and T cell were also normal, as evidenced respectively by CH_{50} assay, immunoglobulin assay and positive candida skin test (details omitted)
- Liver biopsy showed multiple granulomata

A tentative diagnosis was made and nitroblue tetrazolium (NBT) test was performed to detect abnormalities of chemotaxis, if any. An abnormal NBT test result (showing no reduction) was obtained. Quantitative NBT test performed in both her parents revealed half-normal values, though they never suffered from such episodes of infections.

One of the patient's genes, which codes for the large subunit of the cytochrome B_{558} was found to have undergone a point mutation

Codon No.	414	415	416
Patient	ACA	C**A**C	TTC
Normal	ACA	C**C**C	TTC

As a result, proline (encoded by CCC) is replaced by histidine (encoded by CAC)

Q.1. Comment on the case sheet and give your most probable diagnosis.
Q.2. Outline the principle of the NBT test and interpret the test results in this patient.
Q.3. In general terms, connect the above mutation with the patient's disease.
Q.4. Suggest treatment for this patient.
Q.5. Enlist some other defects in neutrophil functions.

CASE 33.3 Recurrent and prolonged infectious episodes in an emaciated child

An 18-month-old child had been suffering from severe recurrent bacterial infections, particularly bronchitis, otitis media and *Haemophilus influenzae* pneumonia since the age of about 6 months. He had also been treated for persistent fever, diarrhoea and scrotal cellulitis a few months earlier. These infections had responded well to the antibiotic therapy on each occasion, but did not completely resolve. Recurrent infections in spite of appropriate antibiotic therapy appeared to have taken their toll on his growth and development, which had plateaued and milestones got delayed. On examination, he was an emaciated and pale child, whose height and weight were below the third centile. Further questioning revealed that he had received full course of primary immunizations with tetanus and diphtheria toxoids, whole cell pertussis and oral polio (at 2, 3 and 4 months); and measles, mumps and rubella (MMR) vaccine at 12 months. Family history was remarkable : two of his mother's brothers had died before the age of 6 years, apparently with the same ailment; and, presently his younger brother (aged 8 months) had been suffering from recurrent chest infections for about last 2 months.

During the last hospital admission 10 days back, serum protein electrophoresis had shown weak staining in gamma-globulin region, suggesting virtual absence of gamma globulins. This had prompted a more extensive immunological evaluation into cause of his recurrent infections with the following results:

- White cell count 8.4 × 10^9/litre (4–11 × 10^9/L)
- Blood lymphocyte sub-populations ($\times 10^9$/L)
- Total leucocyte count 3.7 (2.5–5.0)
- T cells (CD_3) 3.07 (1.5–3.0)
- SIg positive (B) cells Not detected

Serum immunoglobulins (in G/L)

- IgG 0.12 (5.5–10.0)
- IgA Not detected (0.3–0.8)
- IgM Not detected (0.4–1.8)

(normal range for age 18 months is given in brackets)

Specific antibodies to

- Diphtheria toxoid: no detectable IgG antibodies
- Tetanus toxoid: no detectable IgG antibodies
- Polio, measels, rubella: no IgG detected

In view of these results (panhypo-gammaglobulinaemia, absent antibody production, and absence of mature B cells in peripheral blood) a bone marrow biopsy was done and the cells examined. The cell-antigen phenotying demonstrated pre-B cells but no B-lymphocytes.

Further examination of the pre-B cells at molecular level showed that the Btk gene, which is located on the long arm of chromosome 6 and codes for a *protein kinase* of the src family had undergone a point mutation. Sequences in vicinity of this gene are shown below:

Based on these reports, diagnosis was made and appropriate therapy started to which the patient responded well. His health steadily improved and he has had only one episode of chest infection in the last one year. Presently his weight and height are on the 10th centile.

Q.1. Suggest the most probable diagnosis and outline the treatment the child needs.

Q.2. Explain why the child remained protected from repeated infections until about 6 months of age.

Q.3. Interpret the result of bone marrow cell-antigen phenotyping.

Q.4. What is the most plausible explanation for failure to thrive?

Q.5. In general terms, try to connect the above mutation with the observed lack of mature B cells.

Q.6. Enumerate some other causes of primary antibody deficiencies.

	Normal								**Patient**							
DNA:	GCC	ATC	**GAG**	ATG	ATC	AAA	GAA	GGC	GCC	ATC	**AAG**	ATG	ATC	AAA	GAA	GGC
RNA:	GCC	AUC	GAG	AUG	AUC	AAA	GAA	GGC	GCC	AUC	AAG	AUG	AUC	AAA	GAA	GGC
Protein:	Ala	Ile	**Glu**	Met	Ile	Lys	Glu	Gly	Ala	Ile	**Lys**	Met	Ile	Lys	Glu	Gly

CHAPTER

34

TESTS FOR THYROID, ADRENAL AND KIDNEY FUNCTIONS

Endocrine organs like thyroid, adrenals, kidneys, liver, and pancreas perform important biological roles. The diseases involving these organs cause derangement of their functions, and cause several alterations in the metabolic milieu. Assessment of the functional status of these organs requires, in addition to clinical assessment, a number of laboratory investigations. No single test suffices to assess the functional status of a given organ and generally a battery of tests has to be performed. Some of the common function tests are discussed in this chapter.

After going through this chapter the student should be able to:

- Enumerate various tests for assessing the functional status of the above mentioned organs.
- Correctly interpret these test results for the diagnosis of an abnormality, if any.

I. Thyroid Function Tests

A. Assay of Hormones

Estimation of T_4 (thyroxine) and T_3 (triiodothyronine) by radioimmunoassay or ELISA is the mainstay of diagnosis of thyroid diseases. Both are decreased in hypothyroidism, but increased in hyperthyroidism. In general, changes in serum T_3 are proportionally greater than the changes of T_4 in most states of hypo- and hyperthyroidism.

In cases where the results of total serum T_4 and T_3 estimation are borderline or conflicting, ***free hormone levels*** are estimated. Recall that thyroid hormones in blood are in two forms: (a) protein bound, and (b) unbound or free form. Elevation or fall of free T_3 and free T_4 are true reflections of hyperthyroidism and hypothyroidism, because it is the free hormone that induces metabolic and biological effects in target cells.

B. Radioactive Iodide Uptake (RAIU)

The RAIU assay is used to measure the ability of the thyroid gland to trap iodide. A small dose of radioactive iodine (5 μc of $Na^{131}I$) is ingested by mouth and the radioactivity is counted at various lengths of time, usually 4–6 hours and again at 24 hours. Normally, 10–30% of the ingested dose can be detected within 24 hours. The uptake is increased in hyperthyroidism and decreased in hypothyroidism.

C. Serum TSH

Estimation of serum TSH is useful for differentiating the thyroid disorders of hypothalamic or pituitary origin from those of thyroid origin. For example, in hypothyroidism of thyroid origin (primary hypothyroidism), TSH level is elevated because of lack of feedback effect, but in secondary hypothyroidism (of pituitary or hypothalamic origin) the TSH level is low.

Likewise, in primary hyperthyroidism, serum TSH is decreased due to negative feedback by T_3 and T_4, whereas in hyperthyroidism of pituitary cause, high TSH is observed.

D. Thyroid Hormone Binding Proteins

Levels of hormone-binding proteins, such as thyroxine-binding globulins (TBG), effect the plasma concentration

of T_4 and T_3. Because most of the circulating T_3 and T_4 is bound to TBG, quantitative changes in this protein are clinically important (Table 34.1). The total amount of thyroid hormones transported in plasma will increase with increased TBG, or decrease with decreased TBG. Causes of abnormalities in TBG levels are summarized in Table 34.1.

In addition to absolute changes in TBG levels mentioned above (i.e. the quantitative changes), qualitative changes in this protein are also important. These changes refer to alterations of the hormone-binding affinity. Decreased affinity, for example, liberates thyroid hormones from TBG-binding sites, resulting in decreased total serum T_4 and T_3, and increased free T_4 and free T_3.

Quantitative as well as qualitative changes in thyroid hormone-binding proteins, such as thyroxine-binding globulin, affect total plasma concentration of T_4 and T_3.

E. TRH Stimulation Test

Administration of TRH causes increased release of TSH, and consequently of thyroid hormones, provided hypothalamo-pituitary-thyroid axis is normal. An abnormal response is obtained in the following conditions:

1. In hypothyroidism, higher-than-normal increments of serum TSH (after TRH stimulation) is observed. This is because of low feedback inhibition by deceased levels of circulating thyroid hormones.
2. In hyperthyroidism, feedback inhibition by elevated levels of thyroid hormones overrides the stimulating effect of TRH, resulting in *a blunted TSH response.*
3. When receptors for TRH on pituitary are defective, there is lower-than-normal response of serum TSH to TRH stimulation. In such cases, the TSH response (to TRH) is said to be blunted.

Table 34.1. Causes of abnormalities in thyroxine-binding globulins (TBG)

Quantitative	Qualitative
Increased TBG serum levels	**Genetic**
• Oral contraceptives	• Genetically determined increased binding-affinity
• Pregnancy	• Genetically determined decreased binding affinity
• Neonatal period	
• Genetic TBG excess	
• Oestrogen therapy	
Decreased TBG serum levels	**Induced by drugs competing for TBG-binding sites**
• Genetic deficiency (X-linked)	• Aspirin
• Anabolic agents	• Dicumarol
• Androgens	• Heparin
• Protein malnutrition	• Phenytoin
	• Phenylbutazone

F. Non-specific Indices

These include, (a) *basal metabolic rate*, which increases in hyperthyroidism and decreases in hypothyroidism, and (b) *serum cholesterol*, which is decreased in hyperthyroidism and increased in hypothyroidism.

G. Tests for Detecting Damage to Thyroid

This involves demonstration of **autoantibodies** to the thyroid gland, which may produce cell destruction and eventual hypothyroidism.

II. Adrenal Function Tests

Glucocorticoids (primarily cortisol) are mainly involved in the regulation of carbohydrate, lipid and protein metabolism, and the mineralocorticoids (primarily aldosterone) are chiefly concerned with regulation of the extracellular fluid volume and electrolyte metabolism.

A. Tests for Glucocorticoid Functions

Tests to Establish Diagnosis

Assessment of Diurnal Rhythm

Plasma cortisol levels are about 10 times higher in the early morning hours than at midnight. The normal range for plasma cortisol level is 8–26 µg/100 mL at 08.00 a.m. to less than 10 µg/100 mL at 12.00 a.m. Loss of this diurnal rhythmicity is an early indication of a lesion at any point in the hypothalamic-pituitary-adrenal axis. Stress such as trauma, pain, apprehension, fever and hypoglycaemia can also override this diurnal rhythm.

The hormone levels are most commonly estimated by radioimmunoassay (Box 34.1).

Estimation Urinary Free Cortisol

Another useful initial screening test is the estimation of urinary free cortisol in a 24-hour urine sample. The test should be performed on three different collection

BOX 34.1

Radioimmunoassay: Most Versatile Method for Determination of Hormone Levels

Diagnosis of endocrine disorders is mostly based on determination of circulating hormone levels. However, the circulating levels of hormones are extremely low and so their determination is difficult. Moreover, a number of plasma components may interfere with the hormonal assay, which makes the determination more difficult.

The most commonly applied procedure for hormone determination is radioimmunoassay (RIA), which requires (a) a specific antibody to the hormone, and (b) a radiolabelled version of the hormone that contains tritium, radioactive iodide, or some other suitable isotope. The two are mixed to obtain a labelled hormone-antibody complex. When patient's serum is added to it, the (unlabeled) hormone in patient's serum competes with the radiolabeled hormone for binding to the antibody. After incubating for some time, the free hormones and the antibody-hormone complexes are separated. The higher the hormone concentration in patient's serum, the more the radiolabeled hormone is displaced (from binding to the antibody) and appears as free unbound radioactivity. This is measured in a scintillation counter: larger the amount of hormone in the patient's serum, higher the specific radioactivity of the free hormone.

samples. Normal daily excretion of free cortisol in urine ranges from 10 μg to 100 μg. It rises in hyperadrenalism and falls in hypoadrenalism (**Case 34.1**).

Low Dose Dexamethasone Suppression Test

This test is recommended if results of urinary free cortisol tests are abnormal. Dexamethasone is a potent suppressor of pituitary ACTH secretion and cortisol level, causing about 50% fall in serum cortisol with a dose as low as 2 mg. The patient takes dexamethasone tablet at night and plasma cortisol is determined at 8.00 a.m. the following morning. A morning cortisol level less than 5 μg/100 mL usually excludes any cause of hypercortisolism. If the cortisol level is not suppressed, however, the next step is to identify the aetiology.

Tests to Establish Aetiology

Measurement of Plasma ACTH

Because cortisol and ACTH interact in a feedback loop, low levels of serum cortisol associated with high plasma ACTH level would indicate primary adrenocortical insufficiency. Likewise, increased plasma cortisol will be associated with suppression of ACTH in primary adrenal lesion.

In contrast, increased cortisol due to an ACTH-producing pituitary adenoma (Cushing's disease) or due to ectopic production of ACTH will be associated with increased plasma ACTH levels.

High Dose Dexamethasone Suppression Test

The test is carried out with administration of 2 mg dexamethasone every 6 hours for 2 days. This dose suppresses urinary 17-OH steroids as well as plasma cortisol in Cushing's disease. If these parameters are not suppressed, adrenal tumours producing high levels of cortisol or ectopic ACTH-producing tumours are usually the aetiology.

ACTH Stimulation Tests

These tests are useful in assessing adrenal reserve capacity and for documenting the existence of hormonal deficiency state. The test uses a synthetic form of ACTH (*synacthen*) consisting of the first 24 amino acids of ACTH, which is injected intravenously or intramuscularly. A **short ACTH stimulation test** (250 μg Synacthen administered intramuscularly or intravenously) is followed by a rise in serum cortisol within a few minutes, whereas a person with primary adrenal failure does not respond. **Long ACTH stimulation test** uses a lower daily dose of 1 μg administrated over several days. This dose is apparently more effective in stimulating adrenals than single administration of 250 μg which is considered supraphysiological. Thus, long ACTH test successfully assesses the chronic adrenal insufficiency.

Metyrapone Stimulation Test

This test is used to delineate cause of Cushing's syndrome. Metyrapone is a potent inhibitor of the *11-hydroxylase* enzyme, therefore oral administration of metyrapone blocks cortisol synthesis. This removes the negative feedback effect on pituitary. The excess of ACTH secretion continues to drive steroid biosynthesis, but with the pathway of cortisol synthesis blocked, the initially formed precursors (and progesterone) stop at 11-deoxycortisol (Fig. 32.8). Therefore, the serum levels of progesterone and these initial precursors rise following metapyrsone administration. Likewise, the urinary excretion of 17-hydroxycorticosteroids also increases.

In Cushing's syndrome caused by a pituitary tumour, the ACTH response remains intact, and 11-deoxycortisol levels show marked rise (> 200 nmol/L). Levels of 11-deoxycortisol that are less than this are consistent with adrenal tumour or ectopic ACTH.

CRH Stimulation Test

This test is now considered definitive in the differentiation of pituitary causes of Cushing's syndrome from other causes. If there is no ACTH response after a CRH stimulation test, the disease is of pituitary origin.

For determining the cause of Cushing's syndrome, metyrapone test may be better than the high-dose dexamethasone suppression test; however, most laboratories do not perform this protocol due to lack of ready availability of 11-deoxycortisol assay.

B. Tests for Mineralcorticoid Secretion

Plasma Aldosterone

Plasma aldosterone, estimated in blood collected from supine position, is 6–22 ng/dL for males and 5–30 ng/dL for females. These values increase in hyperaldosteronism and fall in hypofunction.

Plasma Renin Activity

This estimation helps to differentiate between primary and secondary hyperaldosteronism. The enzyme activity is high in secondary but low in primary hyperaldosteronism.

III. Renal Function Tests

The renal function tests may be broadly divided into (a) tests that assess glomerular functions, and (b) tests that assess tubular functions. In addition, *routine urine examination* is still held as an important aspect of the overall assessment.

A. Routine Urinalysis

It involves a careful physical examination and qualitative tests for detecting any abnormal urinary constituents, such as reducing sugars, proteins, ketone bodies, bile salts, bile pigments, or blood. A careful performance and interpretation of routine urine examination provides useful information about the presence of active lesion in the urinary tract. For example, presence of proteins (proteinuria) indicates glomerular lesion and a need for performing detailed glomerular function tests.

B. Glomerular Function Tests

These include measurement in blood of waste products (e.g. urea and creatinine) that accumulate once renal failure sets in, and the clearance tests.

Serum Urea

Urea makes up the majority (> 75%) of the non-protein waste products excreted daily as a result of the oxidative catabolism of proteins. Urea is excreted solely by kidneys: it is filtered at glomeruli and about 40–60% of filtered urea is reabsorbed in collecting ducts. Therefore, impaired glomerular filtration results in retention of urea and its concentration in blood rises (normal is 15–45 mg/dL).

However, it is not a sensitive test because serum urea concentration begins to rise only after the filtration rate has fallen substantially. It is non-specific as well because a rise in serum urea is seen in various non-renal conditions also, such as hypovolaemia and poor perfusion due to cardiovascular failure, prostatic hypertrophy, and prolonged starvation.

In spite of these limitations, serum urea is a widely used test.

Serum Creatinine

Creatinine is a small compound (MW 113) readily filtered by the glomerulus, and unlike urea, is not reabsorbed by the tubules or collecting ducts. Elevated serum creatinine concentration is a *more sensitive indicator of glomerular damage* than serum urea.

Increased creatinine is due to any cause of impaired renal function, or very large muscle mass (anabolic steroid users, acromegaly patients).

Clearance Tests

Clearance is that *volume of plasma from which a measured amount of substance can be completely eliminated into urine per unit time.* Measurement of clearance can detect much earlier stages of renal damage. This is a definite advantage over measurement of nitrogenous waste substances (urea and creatinine), because either of these substances begin to rise in blood only after renal failure is quite advanced, with only about 20–30% of the nephrons still functioning.

If a substance is completely filtered by glomeruli, but neither secreted into nor reabsorbed from the tubule, its clearance is equal to glomerular filtration rate (GFR). Inulin, and to a lesser extent creatinine, meet the above criteria. Hence, GFR can be estimated by measuring inulin clearance or creatinine clearance.

Creatinine Clearance

Creatinine is a nearly ideal substance for the measurement of clearance for various reasons:

- It is an endogenous metabolic product synthesized at a constant rate for a given individual.

- It is cleared essentially only by glomerular filtration (not reabsorbed and only slightly secreted by the proximal tubules).
- It can be analyzed inexpensively by colorimetric method.
- It is completely filtered at glomeruli, and there is little of tubular handling, therefore, its clearance is closer to GFR.

Specimen collection includes both a 24-hour urine specimen and a serum sample, preferably taken at the midpoint of the 24-hour urine collection. Volume of urine is carefully measured and concentration of creatinine in both serum and urine is determined. Creatinine clearance is calculated using the general **clearance formula UV/P**, where,

U = urine creatinine concentration,
V = urine volume excreted in 24 hours,
P = serum creatinine concentration

Although the general clearance formula (UV/P) is mostly used, correction for body surface area must be included in the formula because creatinine excretion varies with regard to lean body mass. A **normalization factor** (1.73/A) is used for the purpose, where 1.73 is the generally accepted average body surface in square meters and A is patient's body surface area. Accordingly, the creatinine clearance modified for surface area is calculated as:

UV/P × 1.73/A

Reference values: The normal range for creatinine clearance is 100–125 mL/min in males and 90–115 mL/min in females. When corrected for surface area, the creatinine clearance value becomes the same for both sexes, which is about 100 mL/min. 1.73 square meter.

Significance: Values below normal indicate impairment of glomerular functions. As mentioned earlier, creatinine clearance is a very sensitive indicator, being able to detect renal impairment in the early stages.

Creatinine clearance is widely used to approximate glomerular filtration rate.

Urea Clearance

Urea is filtered at glomeruli and 40–60% is reabsorbed by tubules and collecting ducts. For this reason it does not provide full clearance assessment. Moreover, urea clearance is influenced by protein content of the diet, and therefore, relatively insensitive for assessing glomerular functions. Though it was one of the first clearance tests performed, it has no relevance in modern medicine and is rarely, if ever, used.

Urea is completely filtered, but is reabsorbed at the tubular level, therefore, its clearance is less than GFR. Value of urea clearance varies with the rate of urinary output:

- If the output is more than 2 mL/min, maximum urea clearance is calculated by the formula UV/P. Its value is 75 mL/min.
- If the output is less than 2 mL/min, standard urea clearance is calculated by the formula UV/P. Its value is 54 mL/min.

Inulin Clearance

Inulin is a homopolysaccharide made up of fructose residues (Chapter 2). *It is neither reabsorbed nor secreted by renal tubules, so its clearance is equal to GFR.* It is administered intravenously to measure GFR. However, it is practically more convenient to estimate clearance of substances already present in blood, and for this reason creatinine clearance is preferred over inulin clearance.

C. Tubular Functions

Urine Concentration Test

One of the earliest manifestations of renal tubular damage is the inability to concentrate the urine. It involves measurement of specific gravity (SG) of urine from samples collected in the morning after an overnight fast. Generally, the first three urine specimens passed in the morning are collected, and their SG is measured. If it exceeds 1.022 in at least one of the samples, it implies that ability to concentrate urine is intact, and accordingly the tubular functions are considered as normal.

Dilution Test

Following intake of excess fluids, tubular reabsorption of water is decreased to get rid off the excess water load, and therefore dilute urine is excreted. Test is done in the morning after an overnight food and water deprivation. The bladder is emptied in the morning, a water load is given (1200 mL within 30 minutes) and urine samples are collected every hour for the next 4 hours. In normal subjects, most of the water load will be eliminated within these 4 hours and SG of at least one sample should fall to 1.003 or below. If SG of all samples is above 1.003, it indicates impairment of tubular function. *The dilution test is more sensitive than the concentration test.*

Checking urine specific gravity provides useful information about tubular functions and hydration.

PSP Excretion Test

Phenolsulphonaphthalein (PSP) is not filtered by glomeruli but secreted by proximal renal tubules. *PSP excretion*

in urine is used to test the efficiency of the tubular secretory capacity. The patient is fasted overnight and asked to drink two glasses of water in the morning (to ensure adequate urine output) and empty the bladder after 15 minutes. This is followed by an IV injection of 6 mg of PSP in 1 mL solution. Urine is collected after 15 minutes, 30 minutes, 60 minutes and 120 minutes after the injection and the amount of PSP in each sample is measured. In a normal subject, 20–25% of the injected dose should be excreted within 15 minutes, and 60–70% should be excreted in the first 2 hours. The test effectively detects early stages of renal tubular damage.

Para-amino Hippurate (PAH) Clearance Test

Renal clearance of PAH equals renal plasma flow (RPF). This is because this analyte is completely filtered by glomeruli and extensively secreted in tubules so much so that the entire PAH present in blood is removed by the kidneys in a single passage.

Acidification Test

The capacity of the kidneys to produce acidic urine is studied following stimulation by giving ammonium chloride (or hydrochloric acid). The following procedure is that of Davies and Wrong.

The subject is imposed an overnight fast but is encouraged to drink water till 8 a.m. Bladder is emptied and ammonium chloride, 0.1g per kilogram, in gelatin coated capsules is given orally. Bladder is emptied again and this sample is discarded. All the urine specimens passed during the next 6 hours are collected in sterile bottles, and the bladder is emptied at the end of that period. pH of the specimens and total ammonia of the combined urine is measured.

In healthy persons, urine pH falls below 5.3 in one of the specimens and total ammonia excretion is between 30–90 μEq/min. In most forms of renal failure, the pH falls in the same way but ammonia excretion is low. In renal tubular acidosis, the pH remains between 5.7 and 7.0; and the ammonia excretion is also low.

D. Other Blood Determinations in Renal Diseases

Calcium and Phosphorus

In chronic renal failure there is impairment of phosphate excretion, resulting in hyperphosphataemia, which causes a reciprocal fall in serum calcium concentration. The hypocalcaemia is aggravated by reduced calcium absorption from gut as a result of impaired calcitriol production.

Uric Acid

Like urea and creatinine, the renal retention of uric acid with renal insufficiency causes increased serum concentration of this analyte. Its concentration rises to a significant extent only in an advanced stage of chronic renal failure, but this rarely results in gout.

Serum Electrolytes

Serum concentration of potassium increases (hyperkalaemia) in acute or chronic renal failure due to decreased excretion. Decreased sodium concentration (hyponatraemia) can occur due to an increased extracellular fluid volume, resulting from the inability of the kidneys to excrete water. The concentration of chloride parallels that of sodium.

Exercises

Essay type question

1. Discuss various tests used in diagnosis of primary adrenal insufficiency (Addison's disease).

Write short notes on

1. Metyrapone stimulation test
2. TBG
3. Dexamethasone suppression
4. Insulin clearance
5. Pancreatic function tests

CLINICAL CASE

CASE 34.1 A mildly hypertensive man with central obesity and plethoric face

A 62-year-old man presents with long history of central obesity, plethoric face and abdominal bruising. He is mildly hypertensive. Excretion of free cortisol in urine is 400 μg, which is much higher than in a normal subject (< 100 μg). Plasma cortisol level at midnight is 110 μg/100 mL and his 08.00 h cortisol is 120 μg/100 mL after 1 mg of dexamethasone. Plasma ACTH is 100 ng/L (normal, 80 ng/L).

Q.1. What is your most likely diagnosis?

Q.2. What further test(s) would you recommend for differential diagnosis?

INDEX