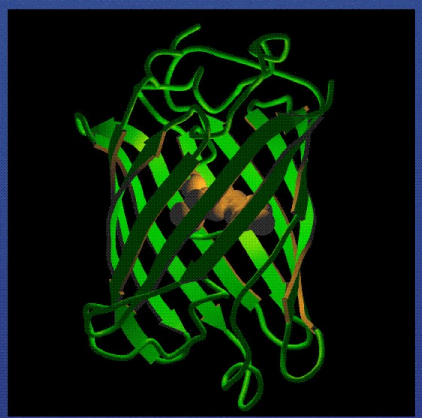
Biotechnology

A problem approach



Third edition

Pranav Kumar | Usha Mina

Biotechnology

A problem approach

Third edition

Pranav Kumar
Former faculty,
Department of Biotechnology
Jamia Millia Islamia,
New Delhi, India

Usha Mina
Scientist,
Division of Environmental Sciences
Indian Agricultural Research Institute (IARI),
New Delhi, India

Pathfinder Publication
New Delhi, India

Pranav Kumar
Former faculty,
Department of Biotechnology
Jamia Millia Islamia,
New Delhi, India

Usha Mina
Scientist,
Division of Environmental Sciences
Indian Agricultural Research Institute (IARI),
New Delhi, India

Biotechnology A problem approach, *Third edition*
ISBN: 978-93-80473-00-0 (paperback)

Publisher : Pathfinder Publication
Production editor : Ajay Kumar
Copy editor : Jomesh Joseph
Illustration and layout : Pradeep Verma
Cover design : Pradeep Verma
Marketing director : Arun Kumar
Production coordinator : Murari Kumar Singh
Printer : Ronit Enterprises, New Delhi, India

Pathfinder Publication
A unit of **Pathfinder Academy Private Limited**, New Delhi, India.
www.thepathfinder.in
09350208235

Preface

The present century has been considered as one that belongs to biotechnology. This branch of science has been viewed as something vital for life with numerous scientific applications in several fields of human endeavours. The branch of science is significant for mankind that many of the big discoveries of the second half of the last century and early this century would not have been possible in the absence of our accomplishments in this discipline. Biotechnology – A problem approach, covers fundamentals and techniques. This book provides a balanced introduction to all major areas of the subject. The chapters such as Biomolecules and catalysis, Bioenergetics and metabolism, Cell structure and functions, Immunology, Bioinformatics and Bioprocess engineering were selected in a sharply focused manner without overwhelming or excessive detail. Sincere efforts have been made to support textual clarifications and explanations with the help of flow charts, figures and tables to make learning easy and convincing. The chapters have been supplemented with self-tests and questions so as to check one's own level of knowledge.

Acknowledgements

Our students were the original inspiration for the first edition of this book, and we remain continually grateful to all of them, because we learn from them how to think about the life sciences and how to communicate knowledge in most meaningful way. We thank, Abhai Kumar, Rizwan Ansari, Lekha Nath, Harleen Kaur and Mr. Ajay Kumar, reviewers of this book, whose comment and suggestions were invaluable in improving the text. Any book of this kind requires meticulous and painstaking efforts by all its contributors. Several diligent and hardworking minds have come together to bring out this book in this complete form. This book is a team effort, and producing it would be impossible without the outstanding people of Pathfinder Publication. It was a pleasure to work with many other dedicated and creative people of Pathfinder Publication during the production of this book, especially Pradeep Verma.

Pranav Kumar

Usha Mina

This page intentionally left blank.

Contents

Chapter 2

Bioenergetics and Metabolism

Chapter 3

Cell Structure and Functions

Chapter 4

Prokaryotes and Viruses

Chapter 5

Immunology

Chapter 6

Genetics

Chapter 7

Recombinant DNA technology

Chapter 8

Bioprocess engineering

Chapter 9

Bioinformatics

This page intentionally left blank.

Chapter 01

Biomolecules and Catalysis

A *biomolecule* is an organic molecule that is produced by a living organism. Biomolecules act as building blocks of life and perform important functions in living organisms. More than 25 naturally occurring chemical elements are found in biomolecules. Most of the elements have relatively low atomic numbers. Biomolecules consist primarily of carbon, hydrogen, nitrogen, oxygen, phosphorus and sulfur. The four most abundant elements in living organisms, in terms of the percentage of the total number of atoms, are hydrogen, oxygen, nitrogen, and carbon, which together make up over 99% of the mass of most cells.

Nearly all of the biomolecules in a cell are carbon compounds, which account for more than one-half of the dry weight of the cells. Covalent bonding between carbon and other elements permit formation of a large number of compounds. Most biomolecules can be regarded as derivatives of hydrocarbons. The hydrogen atoms may be replaced by a variety of functional groups to yield different families of organic compounds. Typical families of organic compounds are the alcohols, which have one or more hydroxyl groups; amines, which have amino groups; aldehydes and ketones, which have carbonyl groups; and carboxylic acids, which have carboxyl groups. Many biomolecules are polyfunctional, containing two or more different kinds of functional groups. Functional groups determine chemical properties of biomolecules.

Sugars, fatty acids, amino acids and nucleotides constitute the four major families of biomolecules in cells. Many of the biomolecules found within cells are macromolecules and mostly are polymers (composed of small, covalently linked monomeric subunits). These macromolecules are proteins, carbohydrates, lipids and nucleic acids.

Small biomolecules	*Macromolecules*
Sugars	Polysaccharide
Fatty acids	Fats/Lipids
Amino acids	Proteins
Nucleotide	Nucleic acid

Nucleic acids and proteins are informational macromolecules. Proteins are polymers of amino acids and constitute the largest fraction (besides water) of cells. The nucleic acids, DNA and RNA, are polymers of nucleotides. They store, transmit and translate genetic information. The polysaccharides, polymers of simple sugars, have two major functions. They serve as energy-yielding fuel stores and as extracellular structural elements.

1.1 Amino acids and Proteins

Amino acids are compounds containing carbon, hydrogen, oxygen and nitrogen. They serve as monomers (building blocks) of proteins and composed of an amino group, a carboxyl group, a hydrogen atom, and a distinctive side chain, all bonded to a carbon atom, the α-carbon.

In an α-amino acid, the amino and carboxylate groups are attached to the same carbon atom, which is called the α-carbon. The various α-amino acids differ with respect to the side chain (R group) attached to their α-carbon. The general structure of an amino acid is:

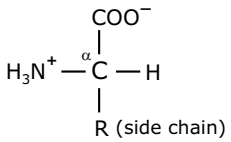


Figure 1.1 General structure of an amino acid.

This structure is common to all except one of the α-amino acids (*proline* is the exception). The R group or side chain attached to the α-carbon is different in each amino acid. In the simplest case, the R group is a hydrogen atom and amino acid is glycine.

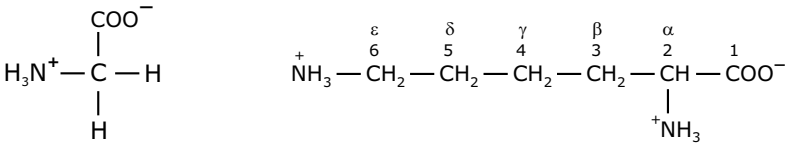


Figure 1.2 Structure of glycine and lysine.

In α-amino acids both the amino group and the carboxyl group are attached to the same carbon atom. However, many naturally occurring amino acids not found in protein, have structures that differ from the α-amino acids. In these compounds the amino group is attached to a carbon atom other than the α-carbon atom and they are called β, γ, δ, or ε amino acids depending upon the location of the C-atom to which amino group is attached.

Amino acids can act as acids and bases

When an amino acid is dissolved in water, it exists in solution as the *dipolar ion* or *zwitterion*. A zwitterion can act as either an acid (proton donor) or a base (proton acceptor). Hence, an amino acid is an **amphoteric molecule**. At high concentrations of hydrogen ions (low pH), the carboxyl group accepts a proton and becomes uncharged, so that the overall charge on the molecule is *positive*. Similarly at low concentrations of hydrogen ion (high pH), the amino group loses its proton and becomes uncharged; thus the overall charge on the molecule is *negative*.

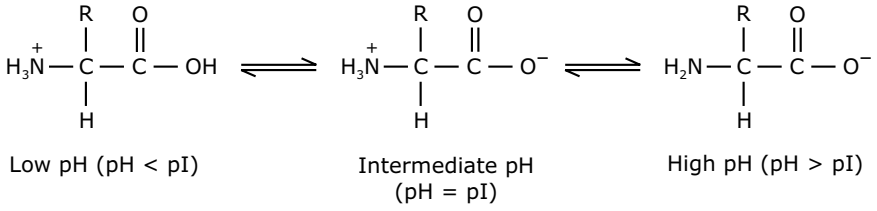


Figure 1.3 The acid-base behavior of an amino acid in solution. At low pH, the positively charged species predominates. As the pH increases, the electrically neutral zwitterion becomes predominant. At higher pH, the negatively charged species predominates.

1.1.1 Optical properties

All amino acids except glycine are optically active i.e. they rotate the plane of plane polarized light. Optically active molecules contain *chiral* carbon. A tetrahedral carbon atom with four different constituents are said to be chiral. All amino acids except glycine have chiral carbon and hence they are optically active.

This page intentionally left blank.

Problem

A solution of L-leucine (3.0 g/50 ml of 6 N HCl) had an observed rotation of +1.81° in a 20 cm polarimeter tube. Calculate the specific rotation of L-leucine in 6 N HCl.

Solution

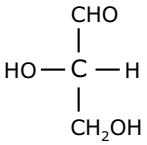
$$[\alpha]_D^T = \frac{\text{Å}}{l \times C}$$

$$[\alpha]_D^T = \frac{+1.81}{2 \times 0.06} \quad \text{where, } l = 20\,\text{cm} = 2\,\text{dm} \text{ and } C = \frac{3\,\text{g}}{50\,\text{ml}} = 0.06\,\text{g/ml}$$

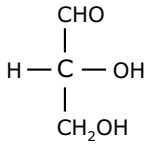
$[\alpha] = +15.1°$.

1.1.2 Absolute configuration

An amino acid with a chiral carbon can exist in two configurations that are non-superimposable mirror images of each other. These two configurations are called *enantiomers*. An enantiomer is identified by its absolute configuration. For example, glyceraldehyde has two absolute configurations. When the hydroxyl group attached to the chiral carbon is on the left in a Fischer projection, the configuration is L; when the hydroxyl group is on the right, the configuration is D.

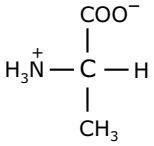


L-Glyceraldehyde

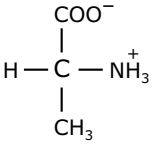


D-Glyceraldehyde

In the above figure, prefixes D- and L- refer to absolute configuration of glyceraldehyde. Similarly, absolute configuration of amino acids are specified by the D- and L- system. The designation of D or L to an amino acid refers to its absolute configuration relative to the structure of D- or L-glyceraldehyde, respectively.



L-Alanine

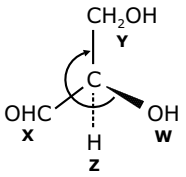


D-Alanine

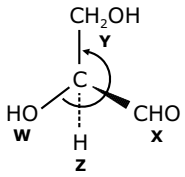
All amino acids except glycine exist in these two different enantiomeric forms. However, all the amino acids ribosomically incorporated into proteins exhibit L-configuration. Therefore, they are all L-α-amino acids. The basis for preference for L-amino acids is not known. D-form of amino acids are not found in proteins, although they exist in nature. D-form of amino acids are found in some peptide antibiotics and peptidoglycan cell wall of eubacteria.

A second absolute configuration notation using the symbols *R* (from *rectus*, Latin for right) and *S* (from *sinister*, Latin for left) can also be used. In this approach, the substituents on an asymmetric carbon (a chiral carbon with four different substituents) are prioritized by decreasing the atomic number. Atoms of higher atomic number bonded to a chiral centre are ranked above those of lower atomic number. For example, the oxygen atom of a —OH group takes precedence over the carbon atom of the $—CH_3$ group that is bonded to the same chiral carbon atom. If any of the first substituent atoms are of the same element, the priority of these groups is established from the

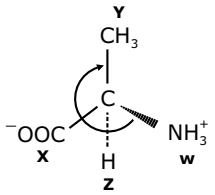
atomic number of the second, third etc, atoms outward from the chiral carbon atom. Hence, a CH_2OH group takes precedence over a CH_3 group. The prioritized groups are assigned the letters W, X, Y and Z with the order of priority rating is W > X > Y > Z. Configuration is assigned by looking down the bond to the lowest priority substituent, Z. If the order of the group W → X → Y is clockwise, then the configuration of the chiral centre is designated *R*. If the order of W → X → Y is counterclockwise, the chiral centre is designated *S*.



D-Glyceraldehyde
R-Glyceraldehyde



L-Glyceraldehyde
S-Glyceraldehyde



D-Alanine
R-Alanine

The absolute configuration of the amino acids at the α-carbon is typically described by D-L system rather than the more modern R-S system. According to the R-S system, all the L-amino acids from proteins are S-amino acids, with the exception of L-cysteine, which is R-cysteine.

1.1.3 Standard and non–standard amino acids

More than 300 amino acids are present in cells; however, only 22 amino acids participate in protein synthesis ribosomically. Such amino acids are called *standard* or *proteinogenic amino acids*. Some amino acids are more abundant in proteins than other amino acids. Four amino acids - leucine, serine, lysine, and glutamic acid- are the most abundant amino acid residues in a typical protein. Tryptophan and methionine are rare amino acids in a protein. Standard L-α-amino acids are specified by simple three letter codons. α-amino acids can be classified on the properties of their side chain (or R group), in particular, their polarity, or tendency to interact with water at physiological pH (near pH 7). The polarity of the side chain varies widely, from nonpolar and hydrophobic to highly polar and hydrophilic.

1. *Amino acids with nonpolar side chain*

   Among standard amino acids, nine amino acids contain nonpolar side chain or R group. These are glycine, alanine, valine, leucine, isoleucine, proline, methionine, phenylalanine and tryptophan. *Proline* differs from other members in having its side chain bonded to both the nitrogen and the α-carbon atoms. *Phenylalanine* and *tryptophan* have aromatic side chains. The side chain of phenylalanine contains a phenyl ring whereas tryptophan has an indole ring.

2. *Amino acids with uncharged polar side chain*

   Six amino acids contain uncharged polar side chain – serine, threonine, cysteine, asparagine, glutamine and tyrosine. Three amino acids, serine, threonine and tyrosine contain *hydroxyl groups* attached to a side chain. Cysteine is structurally similar to serine but contains a *sulfhydryl,* or *thiol group* (–SH) in place of the hydroxyl group.

3. *Amino acids with charged polar side chain*

   Positively charged R group : *Lysine and arginine* have side chains that contain *positively charged* groups at neutral pH or physiological pH. Lysine has an amino group whereas arginine contains a guanidinium group. *Histidine* contains an imidazole group, an aromatic ring. The imidazole group can be uncharged or positively charged near neutral pH, depending on its local environment.

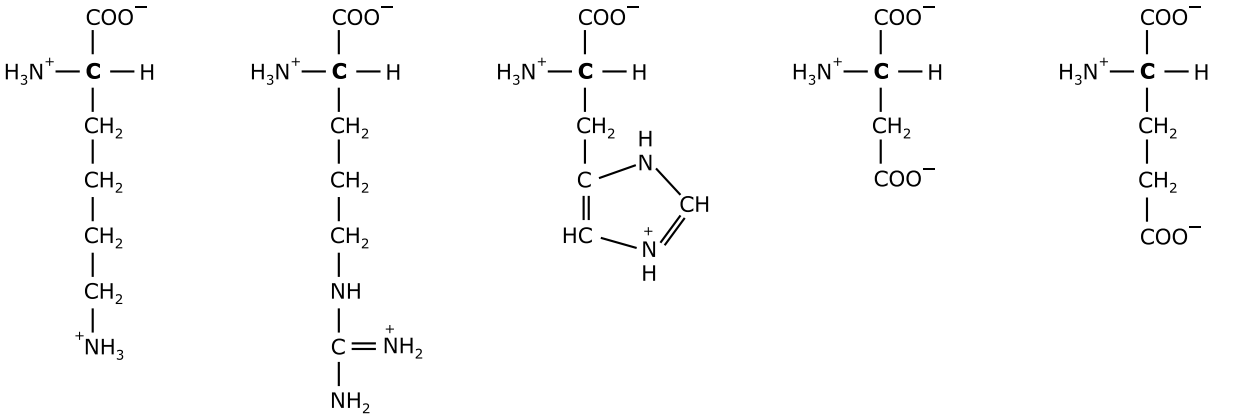
   Negatively charged R group : Amino acids *aspartate* and *glutamate* contain *acidic side chains* that contain negatively *charged* groups at physiological pH.

This page intentionally left blank.

Amino acids with charged polar side chain

Positively charged R groups *Negatively charged R groups*

Lysine Arginine Histidine Aspartate Glutamate

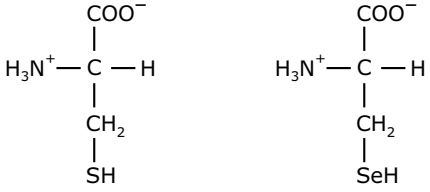


Nonstandard amino acids

More than three hundred amino acids have been found in cells. Twenty-two amino acids are ribosomically incorporated into proteins and are called *proteinogenic* or standard amino acids. Apart from the 22 standard amino acids, all other amino acids are not ribosomically incorporated into proteins are called *non-standard*. In addition to the standard amino acids, some proteins may contain non-standard amino acid residues formed by post-translational modification of standard amino acid residues already incorporated into a polypeptide. These modifications are often essential for the function or regulation of a protein. Examples of some of these amino acids are 4-Hydroxyproline (derivative of proline), 5-Hydroxylysine (derivative of lysine), desmosine (derivative of lysine), N-acetylserine, N-formylmethionine and γ-carboxyglutamate (found in the blood clotting protein prothrombin).

Besides their role in proteins, amino acids and their derivatives have many other biologically important functions. Many nonstandard amino acids are not found in proteins. These amino acids often occur as intermediates in the metabolic pathways for standard amino acids. For example, *ornithine* and *citrulline* are key intermediates in the biosynthesis of arginine and in the urea cycle. Similarly, *azaserine,* a nonstandard amino acid, acts as an antibiotic.

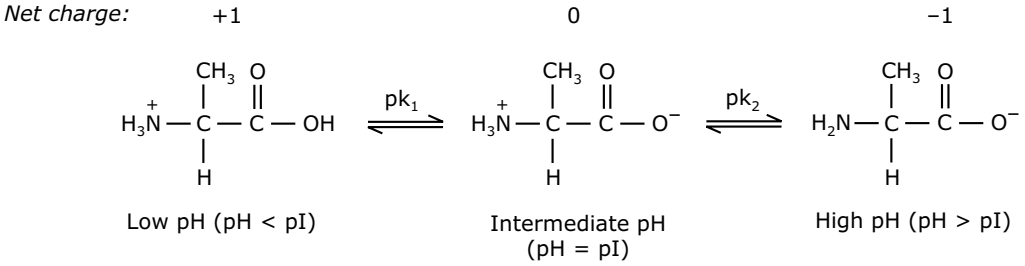
It was originally thought that all unconventional amino acids were made by modifying one of the standard amino acids after it was incorporated into protein, a process called a post-translational modification. But amino acids like *selenocysteine*, *pyrrolysine* are inserted into proteins by the translational machinery. **Selenocysteine** is introduced during protein synthesis rather than created through a postsynthetic modification. It contains selenium rather than sulphur of its structural analog, cysteine. Since selenocysteine is incorporated into polypeptides during translation, it is referred to as 21st amino acid. However, it is specified by a triplet codon, UGA (a stop codon). Selenocysteine has its own tRNA containing the anticodon UCA and it is formed by modifying a *serine* that has been attached to the selenocysteine tRNA. Enzymes like *glutathione peroxidase* and *formate dehydrogenase* contain selenocysteine in their catalytic center. **Pyrrolysine** is similar to lysine and is present in some bacterial proteins. It is coded by UAG codon.



Cysteine Selenocysteine

1.1.4 Titration of amino acids

Because amino acids contain ionizable groups, the predominant ionic form of these molecules in solution depends on the pH. Titration of an amino acid illustrates the effect of pH on amino acid structure. Consider alanine, a simple amino acid, which has two titrable groups (α-amino and α-carboxyl group). During titration with a strong base such as NaOH, alanine loses two protons in a stepwise fashion. In a strongly acidic solution, alanine is present mainly in the form in which the carboxyl group is uncharged. Under this condition the molecule's net charge is +1, since the ammonium group is protonated. However, an increase in the pH results in the deprotonation of α-carboxyl group. At this point, alanine has no *net* charge and is electrically neutral. The pH at which this occurs is called the **isoelectric point** (pI).



Because there is no net charge at the isoelectric point, amino acids are electrophoretically non-mobile and least soluble at this pH. Further increase in pH i.e. lowering of the H^+ concentration results in the deprotonation of the charged amino group and an uncharged amino group forms. So at high pH, the net charge on the molecule is −1, since the ammonium group is deprotonated and a net negative charge develops due to the presence of the carboxylate group.

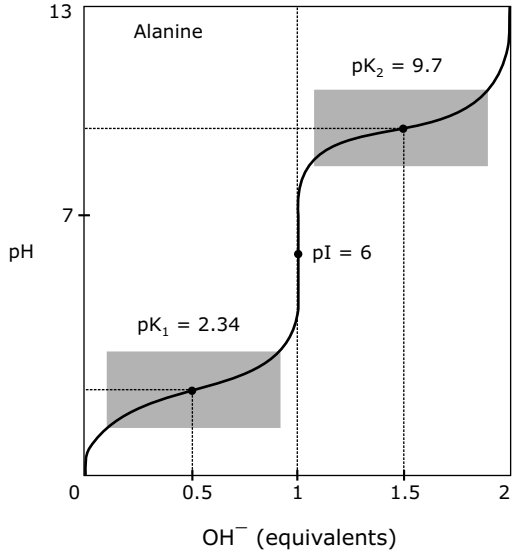


Figure 1.6 Titration curve of alanine (monoamino and monocarboxylic acid). A plot of the dependence of the pH on the amount of OH^- added is called a *titration curve.*

The isoelectric point for alanine may be calculated as follows:

$$pI = \frac{pK_1 + pK_2}{2}$$

This page intentionally left blank.

Table 1.2 pK_a values for the ionizing groups of the standard amino acids

Amino acid	**pK_1** (—COOH)	**pK_2** ($—NH_3^+$)	**pK_R**
Glycine (Gly, G)	2.34	9.6	
Alanine (Ala, A)	2.34	9.69	
Valine (Val, V)	2.32	9.62	
Leucine (Leu, L)	2.36	9.6	
Isoleucine (Ile, I)	2.36	9.6	
Serine (Ser, S)	2.21	9.15	
Threonine (Thr, T)	2.11	9.43	
Methionine (Met, M)	2.28	9.21	
Phenylalanine (Phe, F)	1.83	9.13	
Tryptophan (Trp, W)	2.83	9.39	
Asparagine (Asn, N)	2.02	8.8	
Glutamine (Gln, Q)	2.17	9.13	
Proline (Pro, P)	1.99	10.6	
Cysteine (Cys, C)	1.71	10.78	8.33
Histidine (His, H)	1.82	9.17	6.04
Aspartic acid (Asp, D)	2.09	9.82	3.86
Glutamic acid (Glu, E)	2.19	9.67	4.25
Tyrosine (Tyr, Y)	2.2	9.11	10.46
Lysine (Lys, K)	2.18	8.95	10.54
Arginine (Arg, R)	2.17	9.04	12.48

Note: Seven of the 20 amino acids have ionizable side chains. These 7 amino acids are able to donate or accept protons.

Absorption of UV radiation by aromatic amino acids

Aromatic amino acids such as tryptophan, tyrosine and phenylalanine absorb ultraviolet (UV) light. The aromatic side chains of these amino acids are responsible for UV absorption. Tryptophan and tyrosine absorb maximum near a wavelength of 280 nm. However phenylalanine absorbs maximum at 257.4 nm. Absorbance at 280 nm is used for detection and quantification of purified proteins. The absorbance of each protein depends on the number and positions of its aromatic amino acid residues.

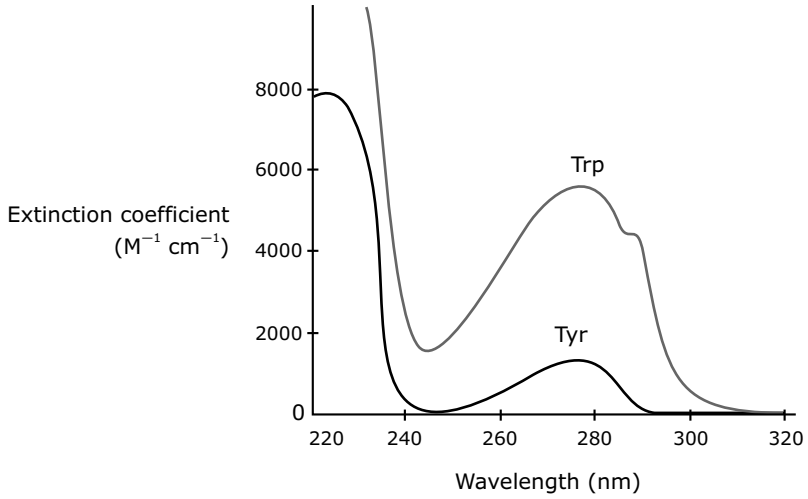


Figure 1.9 Absorption of UV-light. Maximum radiation absorption for both tryptophan and tyrosine occur near a wavelength of 280 nm and absorbance of tryptophan is as much as four times that of tyrosine.

This page intentionally left blank.

1.1.6 Peptide bond

Peptides and polypeptides are linear and unbranched polymers composed of amino acids linked together by peptide bonds. Peptide bonds are amide linkages formed between α-amino group of one amino acid and the α-carboxyl group of another. This reaction is a dehydration reaction, that is, a water molecule is removed and the linked amino acids are referred to as *amino acid residues*. Peptide bond formation is an endergonic process, with $\Delta G \sim +21$kJ/mol.

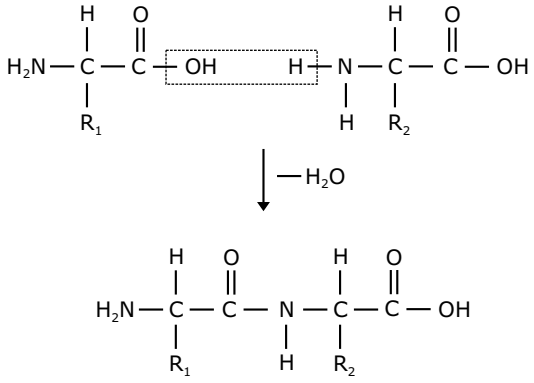


Figure 1.11 The formation of a peptide bond (also called an *amide bond*) between the α-carboxyl group of one amino acid to the α-amino group of another amino acid is accompanied by the loss of a water molecule.

The peptide C—N bond has a partial double bond character that keeps the entire six-atom peptide group in a rigid planar configuration. Consequently, the peptide bond length is only 1.33 Å, shorter than the usual C—N bond length of 1.45 Å. The peptide bond appears to have approximately 40 percent double-bonded character. As a result, the rotation of this bond is restricted.

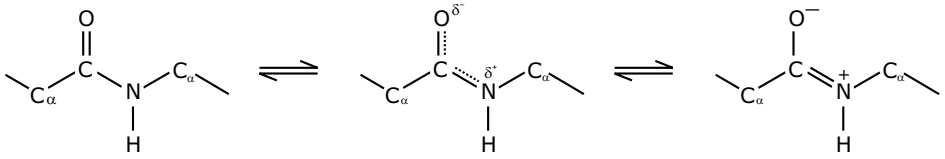
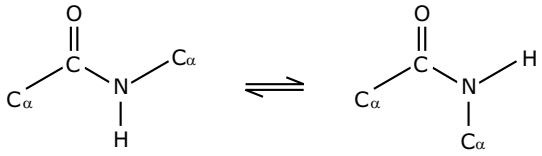


Figure 1.12 The peptide bond has some double-bond character due to resonance. The carbonyl oxygen has a partial negative charge and the amide nitrogen a partial positive charge, setting up a small electric dipole. Virtually all peptide bonds in proteins occur in *trans* configuration.

The angle of rotation around the peptide bond, ω, usually has the value ω = 180° (*trans*) and occasionally ω = 0° (*cis*). The *trans* form is favoured by a ratio of approximately 1000:1 over the *cis* form because in the *cis* form the C_α atom and the side chains of neighbouring residues are in close proximity.



trans *cis*

However, the rotation is permitted about the N–C_α and the C_α–C bonds. Rotation about bonds are described as *torsion* or *dihedral* or *conformational* angle. By convention, the bond angles resulting from rotations at C_α are labeled ϕ (phi) for the N–C_α bond and ψ (psi) for the C_α–C bond.

isoelectric precipitation. Salting out is dependent on the hydrophobic nature of the surface of the protein. Hydrophobic groups predominate in the interior of the protein, but some are located at the surface. Water is forced into contact with these groups. When salts are added to the system, water solvates the salt ions and as salt concentration increases water is removed from around the protein, eventually exposing the hydrophobic groups. Hydrophobic groups on one protein molecule can interact with those of another, resulting in aggregation and thus precipitation.

Effect of solvent

Organic solvents such as acetone, ethanol, decrease the *dielectric constant* of the aqueous solution, which in effect allows two proteins to come close together through electrostatic force of attraction. These solvents due to their low dielectric constants lower the solvating power of aqueous solutions.

1.1.10 Simple and conjugated proteins

On the basis of composition, proteins are classified as *simple* or *conjugated*. **Simple proteins**, such as serum albumin, contain only amino acids. In contrast, **conjugated protein** consists of a simple protein combined with a non-protein component. The non-protein component is called a *prosthetic group.* A conjugated protein without its prosthetic group is called an *apoprotein.* Apoprotein combined with its prosthetic group is referred to as a **holoprotein**. Conjugated proteins are further classified according to the nature of their prosthetic groups. For example, glycoproteins contain a carbohydrate component, lipoproteins contain lipid molecules, and metalloproteins contain metal ions. Similarly, *phosphoproteins* contain phosphate groups and *hemoproteins* possess heme groups.

Table 1.5 Examples of few conjugated proteins

Class	
Glycoproteins	Fibronectin
	Cadherin
Lipoproteins	Chylomicron
	High Density Lipoprotein (HDL)
Metalloproteins	Ferritin (Iron)
	Alcohol dehydrogenase (Zinc)
	Cytochrome oxidase (Copper and iron)
	Nitrogenase (Molybdenum and iron)
Hemoprotein	Hemoglobin (Transport of oxygen in blood)
	Myoglobin (Storage of oxygen in muscle)
	Cytochrome C (Involvement in electron transport chain)
	Cytochrome P450 (Hydroxylation of xenobiotics)
	Catalase (Degradation of hydrogen peroxide)

1.2 Fibrous and globular proteins

Proteins are also classified into two categories – fibrous and globular proteins, on the basis of shape and solubility. **Fibrous proteins** are long, rod-shaped molecules that are insoluble in water and physically tough. Fibrous proteins, such as keratins have structural and protective functions. These proteins usually consist largely of a single type of secondary structure. **Globular proteins** are compact spherical molecules that are usually water-soluble. These proteins often contain several types of secondary structure. In globular proteins, the nonpolar residues *Val, Leu, Ile, Met* and *Phe* largely occur in the interior of a protein, out of contact with the aqueous solvent. The charged polar residues *Arg, His, Lys, Asp* and *Glu* are largely located on the surface of the protein in the contact with the aqueous solvent. Uncharged polar residues *Ser, Thr, Asn, Gln* and *Tyr* are usually present on the protein surface.

Table 1.6 The major functions of proteins and their examples

Function	*Class of protein*	*Example*
Structure	Fibers	Collagen, Keratin, Fibrin
Metabolism	Enzymes	Lysosomes, Proteases, Polymerase, Kinases
Membrane transport	Channels/Carriers	Proton pump, Anion channels
Cell recognition	Cell surface antigens	MHC proteins, ABO blood group
Osmotic regulation	Albumin	Serum albumin
Regulation of gene action	Repressors	*l*ac repressor
Regulation of body functions	Hormones	Insulin, Vasopressin, Oxytocin
Transport throughout body	Globins	Hemoglobin, Myoglobin, Cytochromes
Storage	Ion-binding	Ferritin, Casein, Calmodulin
Contraction	Muscle	Actin, Myosin
Defense	Ig, Toxins	Antibodies, Snake venom

1.2.1 Collagen

Collagen is the major structural protein in the extracellular matrix. It is the most abundant protein in vertebrates. Collagens are a large family of proteins containing at least 19 different members. A typical collagen molecule is long, inelastic, stiff, triple stranded helical structure. The fundamental unit of collagen is **tropocollagen** (length ~300nm). Tropocollagen consists of 3-coiled polypeptides called α-chains. The α-chains are left-handed polypeptide helices and have 3.3 amino acid residues per turn. Three α-chains wind around one another in a characteristic right-handed triple helix. Vertebrates have about 25 different kinds of α-chains, each coded by different genes and has its own unique amino acid sequence. These different types of α-chains combine in various ways to form at least 19 different types of collagen molecules. Types I, II, and III represent 90% of collagens.

The amino acid sequence in α-chain is generally a repeating tripeptide unit, *Gly–X–Y*, where *X* is often proline and *Y* is often 3- or 4-hydroxyproline or 5-hydroxylysine. Glycine constitutes approximately one-third of the amino acid residues. Proline and hydroxyproline confer rigidity on the collagen molecule. The hydroxylation is carried out during post-translational modifications of α-chains by two enzymes: *prolyl hydroxylase* and *lysyl hydroxylase*. Ascorbate (vitamin C) acts as cofactors for these enzymes and hence is essential for hydroxylation of proline and lysine residues. Hydroxylation results in formation of interchain H-bonds. It also allows the glycosylation of hydroxylysine residues. Deficiency of ascorbic acid causes **scurvy**, a disease that affects the structure of collagen. It occurs due to impaired synthesis of collagen as a result of deficiencies of prolyl and lysyl hydroxylases.

The α-chains are synthesized on membrane-bound ribosomes and enter into the lumen of the endoplasmic reticulum (ER). In the lumen of the ER selected proline and lysine residues are hydroxylated to form hydroxyproline and hydroxylysine, respectively, and some of the hydroxylysine residues are glycosylated. Each α-chain then combines with two others to form a hydrogen-bonded, triple-stranded helical molecule known as tropocollagen (or collagen molecule). Tropocollagen is secreted into the extracellular space. After secretion the tropocollagens assemble in the extracellular space to form *collagen fibrils*.

Cross-linking of collagen

Covalent cross-links are formed both within a tropocollagen molecule and between different molecules. *Intramolecular* cross-links form through the action of *lysyl oxidase*, a copper-dependent enzyme that oxidatively deaminates the ε-amino groups of lysine residues, yielding reactive aldehydes of *allysine* residues. Such aldehydes of two side chains then link covalently in a spontaneous nonenzymatic aldol condensation. Histidine may also be involved in certain cross-links.

This page intentionally left blank.

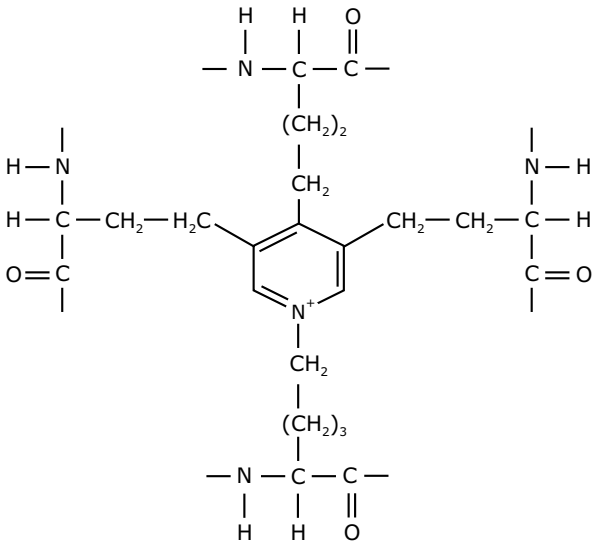


Figure 1.20 Intramolecular desmosine cross-links in elastin.

1.2.3 Keratins

Keratins are fibrous proteins present in eukaryotes. They form a large family, with about 30 members being distinguished. Keratins have been classified as either α-keratins or β-keratins.

Proteins	*α-keratin*	*β-keratin*
Characteristics	Tough, insoluble	Soft, flexible
Conformation	Helical	Extended chain
Basic unit	Protofibril	Antiparallel β-pleated sheet

α-keratins are intermediate filament proteins present only in many metazoans, including vertebrates. In vertebrates, α-keratins constitute almost the entire dry weight of hair, wool, feathers, nails, claws, scales, horns, hooves, and much of the outer layer of skin. The α-keratin polypeptide chain which forms polymerized α-keratin structure, is a right-handed α-helix and rich in hydrophobic amino acid residues Ala, Val, Leu, Ile, Met and Phe. Every α-keratin polypeptide chain dimerizes to form heterodimer. The heterodimer is made up of type I (acidic) and the type II (neutral/basic) α-keratin polypeptide chains. The two chains in heterodimer have a parallel arrangement. Two heterodimers join in an antiparallel manner to form the fundamental tetrameric subunit (a *protofilament*). Two protofilaments constitute a *protofibril*. Four protofibrils constitute a *microfibril*, which associates with other microfibrils to form a *macrofibril.*

1.2.4 Myoglobin

Myoglobin (Mb), a globular protein, contains a single polypeptide chain of 153 amino acid residues (molecular weight 17,800), and a single heme group. The inside of myoglobin consists almost exclusively of nonpolar residues, whereas the outside contains both polar and nonpolar residues. About 75% of the polypeptide chain is α-helical. There are eight helical segments. These eight helical segments are commonly labeled A–H, starting from the NH_2-terminal end. The interhelical regions are designated as AB, BC, CD,..., GH, respectively. The iron atom of the heme is directly bonded to a nitrogen atom of a histidine side chain of globin.

Heme

Globin of Mb binds a single *heme* group by forming a co-ordinate bond. The heterocyclic ring system of heme is a porphyrin derivative. The porphyrin in heme is known as *protoporphyrin IX.* It is made up of 4-pyrrole ring and 4-pyrroles are linked by methylene bridges to form a tetrapyrrole ring. The *Fe atom* is present either in Fe^{2+} or Fe^{3+} oxidation state in the center of the porphyrin ring.

the native structure and lacks the proper packing interactions in the interior of the protein. The interior side chains remain mobile, more closely resembling a liquid than the solid-like interior of the native state.

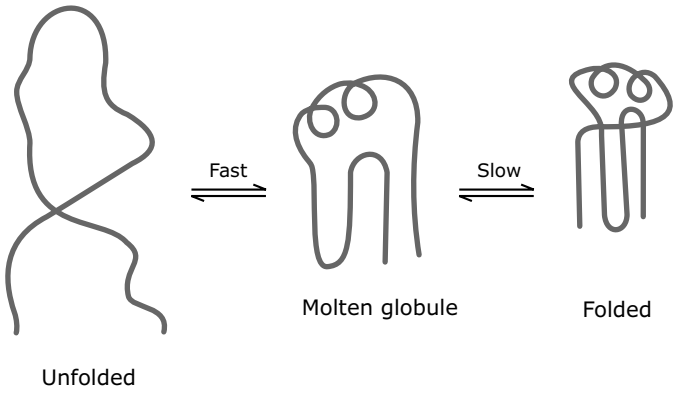


Figure 1.32 The molten globule state is an intermediate state in the folding pathway when a polypeptide chain converts from an unfolded to a folded native state.

1.3.1 Molecular chaperones

Not all proteins fold spontaneously after or during synthesis in the cell. Folding of many proteins requires molecular chaperones. Molecular chaperones are a class of proteins which bind to incompletely folded or assembled proteins in order to assist their folding or prevent them from aggregating. Chaperones function mainly by preventing formation of incorrect structures rather than by promoting formation of correct structures. Chaperones may also be required to assist the formation of oligomeric structures and for the transport of proteins through membranes.

Molecular chaperones were first identified in bacteria *E. coli* but are present in both prokaryotes and eukaryotes (ubiquitous). Several molecular chaperones are included among the *heat-shock proteins* (hence their designation as *Hsp),* because they are synthesized in increased amounts after a brief exposure of cells to an elevated temperature (for example, 42°C for cells that normally live at 37°C).

Eukaryotic cells have at least two major families of molecular chaperones known as the **Hsp60** and **Hsp70** families. Hsps have been classified by molecular mass, for example: Hsp70 for the 70 kDa hsp. The members of these two chaperone families function differently. The members of Hsp70 (Hsp70, Hsc70, Hsp40 and GrpE) act early in the life of many proteins, binding to a string of about seven hydrophobic amino acids before the protein leaves the ribosome. The Hsp70 polypeptide chain is divided into two functional regions, one that binds and hydrolyses ATP and a second that binds hydrophobic segments of unfolded polypeptide chains. The polypeptide binding domain is an antiparallel C-terminal region. Hsp70 is induced by stress (e.g. heat shock) whereas Hsc70 is constitutively expressed in cells. Cytosolic Hsp70s prevent misfolding and maintain the polypeptide chain in unfolded condition. Cytosolic Hsp70s are also necessary for normal translocation of protein from cytosol into either ER or mitochondria. The Hsp70 family is found in bacteria, eukaryotic cytosol, in the endoplasmic reticulum, and in chloroplasts and mitochondria.

In contrast, the Hsp60 family of molecular chaperones (sometime also called **chaperonins**) forms a large barrel-shaped structure that acts later in a protein's life, after it has been fully synthesized. Chaperonins bind unfolded, partly folded and incorrectly folded protein molecules but not protein in their native state. This type of chaperone forms an *isolation chamber* into which misfolded proteins are fed, preventing their aggregation and providing them a favorable environment to refold. The typical structure is a ring of many subunits, forming a cylinder. Hsp60 itself (known as GroEL in *E. coli*) forms a structure consisting of 14 subunits that are arranged in two heptameric rings stacked on top of each other in an inverted orientation. This structure associates with a ring shaped heptamer formed of subunits of Hsp10 (GroES in *E. coli*), also described as **co-chaperonin**.

This page intentionally left blank.

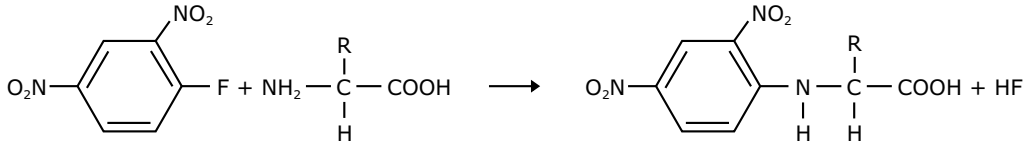
1.4 Protein sequencing and assays

Determination of amino acid compositions

Peptide bonds of proteins are hydrolyzed by either strong acid or strong base. In acid hydrolysis, the peptide can be hydrolyzed into its constituent amino acids by heating it in 6 M HCl at 110°C for 24 hours. Base hydrolysis of polypeptides is carried out in 2 to 4 M NaOH at 100°C for 4 to 8 hours. A mixture of amino acids in hydrolysates can be separated by ion exchange chromatography or by reversed phase HPLC. The identity of the amino acid is revealed by its elution volume and quantified by reaction with *ninhydrin.*

N-terminal analysis

Reagent *1-fluoro-2,4-dinitrobenzene* (FDNB) and *Dansyl chloride* are used for determination of N-terminal amino acid residue. FDNB reacts in alkaline solution (pH 9.5) with the free amino group of the N-terminal amino acid residue of a peptide to form a characteristic yellow dinitrophenyl (DNP) derivative. It can be released from the peptide by either acid or enzymic hydrolysis of the peptide bond and subsequently identified. Sanger first used this reaction to determine the primary structure of the polypeptide hormone insulin. This reagent is often referred to as *Sanger's reagent*.



FDNB Yellow-coloured derivative

Figure 1.34 FDNB reacts with free amino group to produce dinitrophenyl (DNP) derivative of amino acid.

Similarly, *Dansyl chloride* reacts with a free amino group of the N-terminal amino acid residue of a peptide in alkaline solution to form strongly fluorescent derivatives of free amino acids and N-terminal amino acid residue of peptides.

Edman degradation

Edman degradation method for determining the sequence of peptides and proteins from their N-terminus was developed by Pehr Edman. This chemical method uses phenylisothiocyanate (also termed Edman reagent) for sequential removal of amino acid residues from the *N-terminus* of a polypeptide chain.

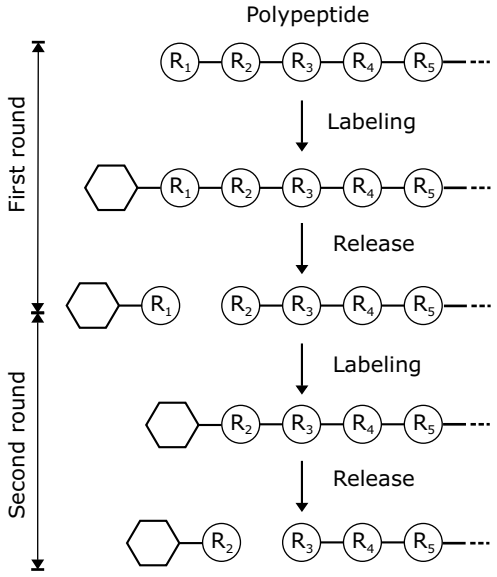
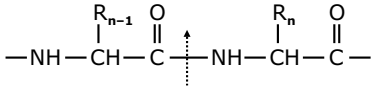


Figure 1.35
Edman degradation sequentially removes one residue at a time from the amino end of a peptide. The labeled amino-terminal residue (R_1) can be released without hydrolyzing the rest of the peptide bonds. Hence, the amino-terminal residue of the shortened peptide (R_2—R_3—R_4—R_5) can be determined in the second round.

This page intentionally left blank.

trypsin, chymotrypsin, elastase, thermolysin and pepsin. Various other chemicals also cleave polypeptide chains at specific locations. The most widely used is *cyanogen bromide* (CNBr), which cleaves peptide bond at C-terminal of Met residues. Similarly hydroxylamine cleaves the polypeptide chain at Asn-Gly sequences.

Table 1.8 Specificities of proteolytic enzymes.



Agents	*Site of Cleavage*
Trypsin	Carboxyl side of Lys or Arg, $R_n \neq$ Pro
Chymotrypsin	Carboxyl side of aromatic amino acid residues, $R_n \neq$ Pro
Pepsin	Amino side of aromatic amino acids like Tyr, Phe and Trp, $R_{n-1} \neq$ Pro
Elastase	Carboxyl side of Ala, Gly and Ser, $R_n \neq$ Pro

Carboxypeptidases and *aminopeptidases* are exopeptidases that remove terminal amino acid residues from C and N-termini of polypeptides, respectively. Carboxypeptidase A cleaves the C-terminal peptide bond of all amino acid residues except Pro, Lys and Arg. Carboxypeptidase B is effective only when Arg or Lys are the C-terminal residues. Carboxypeptidase C acts on any C-terminal residue. Aminopeptidases catalyze the cleavage of amino acids from the amino terminus of the protein. Aminopeptidase M catalyzes the cleavage of all free N-terminal residues.

Cleavage of disulfide bonds

If protein is made up of two or more polypeptide chains and held together by noncovalent bonds then denaturing agents, such as urea or guanidine hydrochloride, are used to dissociate the chains from one another. But polypeptide chains linked by disulfide bonds can be separated by two common methods. These methods are used to break disulfide bonds and also to prevent their reformation.

Oxidation of disulfide bonds with performic acid produces two cysteic acid residues. Because these cysteic acid side chains are ionized SO_3^- groups, electrostatic repulsion prevents S-S recombination. The second method involves the reduction by β-mercaptoethanol or dithiothreitol (Cleland's reagent) to form cysteine residues. This reaction is followed by further modification of the reactive –SH groups to prevent reformation of the disulfide bond. Acetylation by iodoacetate serves this purpose which modifies the –SH group.

Protein assays

To determine the amount of protein in an unknown sample is termed as *protein assays*. The simplest and most direct assay method for proteins in solution is to measure the absorbance at 280 nm (UV range). Amino acids containing aromatic side chains (i.e. tyrosine, tryptophan and phenylalanine) exhibit strong UV-light absorption. Consequently, proteins absorb UV-light in proportion to their aromatic amino acid content and total concentration. Several colorimetric, reagent-based protein assay techniques have also been developed. Protein is added to the reagent, producing a color change in proportion to the amount added. Protein concentration is determined by reference to a standard curve consisting of known concentrations of a purified reference protein. Some most commonly used colorimetric, reagent-based methods are:

Biuret method : Biuret method is based on the direct complex formation between the peptide bonds of the protein and Cu^{2+} ion. This method is not highly sensitive since the complex does not have a high extinction coefficient.

Folin method : The Folin assay (also called *Lowry* method) is dependent on the presence of aromatic amino acids in the protein. First, a cupric/peptide bond complex is formed and then this is enhanced by a phosphomolybodate complex with the aromatic amino acids.

Bradford method : Bradford method is based on a blue dye (Coomassie Brilliant Blue) that binds to free amino groups in the side chains of amino acids, especially Lys. This assay is as sensitive as the Folin assay.

1.5 Nucleic acids

Nucleic acid was first discovered by Friedrich Miescher from the nuclei of the pus cells (Leukocytes) from discarded surgical bandages and called it **nuclein**. Nuclein was later shown to be a mixture of a basic protein and a phosphorus-containing organic acid, now called **nucleic acid**. There are two types of nucleic acids (*polynucleotides*): ribonucleic acid (RNA) and deoxyribonucleic acid (DNA).

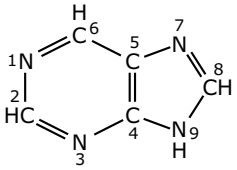
1.5.1 Nucleotides

The monomeric units of nucleic acids are called *nucleotides*. Nucleic acids therefore are also called *polynucleotides*. Nucleotides are phosphate esters of nucleosides and made up of three components:

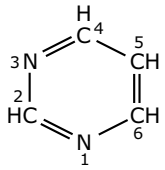
1. A base that has a nitrogen atom (nitrogenous base)
2. A five carbon sugar
3. An ion of phosphoric acid

Nitrogenous bases

Nitrogenous bases are heterocyclic, planar and relatively water insoluble aromatic molecules. There are two general types of nitrogenous bases in both DNA and RNA, pyrimidines and purines.



Purine



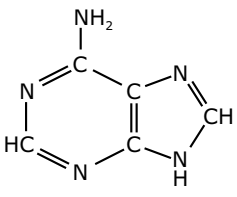
Pyrimidine

Purines

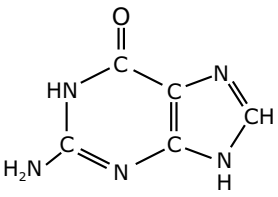
Two different nitrogenous bases with a purine ring (composed of carbon and nitrogen) are found in DNA. The two common purine bases found in DNA and RNA are *adenine* (6-aminopurine) and *guanine* (6-oxy-2-aminopurine). Adenine has an amino group ($-NH_2$) on the C6 position of the ring (carbon at position 6 of the ring). Guanine has an amino group at the C2 position and a carbonyl group at the C6 position.

Pyrimidines

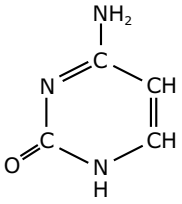
The two major pyrimidine bases found in DNA are *thymine* (5-methyl-2,4-dioxypyrimidine) and *cytosine* (2-oxy-4-aminopyrimidine) and in RNA they are uracil (2,4-dioxypyrimidine) and cytosine. Thymine contains a methyl group at the C5 position with carbonyl groups at the C4 and C2 positions. Cytosine contains a hydrogen atom at the C5 position and an amino group at C4. Uracil is similar to thymine but lacks the methyl group at the C5 position. Uracil is not usually found in DNA. It is a component of RNA.



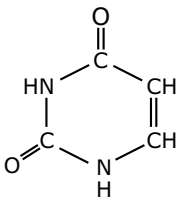
Adenine



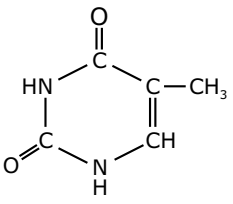
Guanine



Cytosine



Uracil



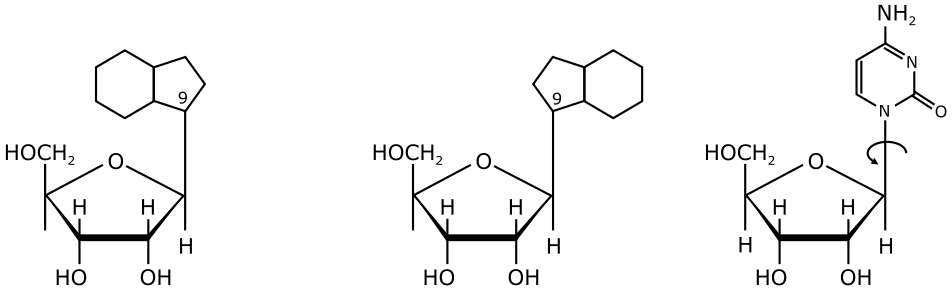
Thymine

Sugars

Naturally occurring nucleic acids have two types of pentose sugars: Ribose and deoxyribose sugar. All known sugars in nucleic acids have the D-stereoisomeric configuration.

This page intentionally left blank.

The base is free to rotate around the glycosidic bond. Due to rotation of the glycosidic bond, two different conformations are possible. The two standard conformations of the base around the glycosidic bond are *syn* and *anti.* Pyrimidines tend to adopt the *anti* conformation almost exclusively, because of steric interference between O2 and C5' in the *syn*-conformation, whereas purines are able to assume both forms (*syn* as well as *anti*).



Syn (purine) *Anti* (purine) *Anti* (pyrimidine)

Nucleotides

The nucleotides are phosphoric acid esters of nucleosides, with phosphate at position C5′. The nucleotide can have one, two, or three phosphate groups designated as α, β and γ for the first, second and third, respectively.

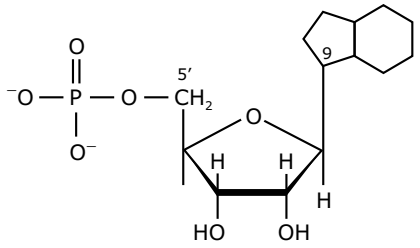


Figure 1.40 Structure of nucleotide.

Nucleotides are found primarily as the monomeric units comprising the major nucleic acids of the cell, RNA and DNA. However, they also are required for numerous other important functions within the cell. These functions include:

- Formation of energy currency like ATP, GTP.
- Act as a precursor for several important coenzymes such as NAD^+, $NADP^+$, FAD and coenzyme A.
- Serving as a precursor for secondary messengers like cyclic AMP (cAMP), cGMP.

ATP

ATP is the chemical link between catabolism and anabolism. It is the energy currency of the living cells. It acts as a donor of high energy phosphate. ATP consists of an *adenosine* moiety to which three *phosphoryl groups* ($—PO_3^{2-}$) are sequentially linked via a *phosphoester bond* followed by two *phosphoanhydride bonds*, referred to as a high energy bond. The active form of ATP is usually a complex of ATP with Mg^{2+} or Mn^{2+}.

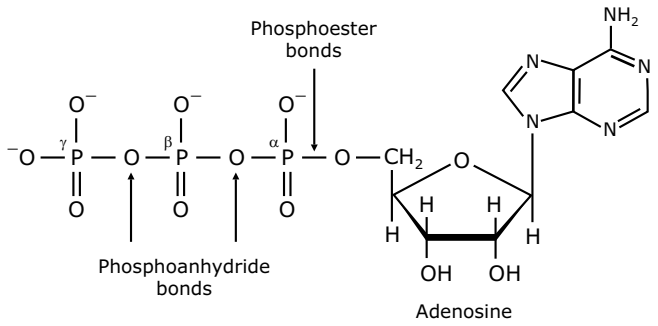


Figure 1.41 Structure of ATP indicating phosphoester and phosphoanhydride bonds.

Pages 45 to 48 are not shown in this preview.

1.6.2 Z–DNA

Left-handed Z-DNA has been mostly found in alternating purine-pyrimidine sequences $(CG)_n$ and $(TG)_n$. Z-DNA is thinner (18 Å) than B-DNA (20 Å), the bases are shifted to the periphery of the helix, and there is only one deep, narrow groove equivalent to the minor groove in B-DNA. In contrast to B-DNA where a repeating unit is a 1 base pair, in Z-DNA the repeating unit is a 2 base pair. The backbone follows a zigzag path as opposed to a smooth path in B-DNA. The sugar and glycosidic bond conformations alternate; C2′ endo in *anti* dC and C3′ endo in *syn* dG. Electrostatic interactions play a crucial role in the Z-DNA formation. Therefore, Z-DNA is stabilized by high salt concentrations or polyvalent cations that shield interphosphate repulsion better than monovalent cations.

Z-DNA can form in regions of alternating purine-pyrimidine sequence; $(GC)_n$ sequences form Z-DNA most easily. $(GT)_n$ sequences also form Z-DNA but they require a greater stabilization energy. $(AT)_n$ sequences generally does not form Z-DNA since it easily forms cruciforms.

Table 1.10 Comparisons of different forms of DNA

Geometry attribute	*A-form*	*B-form*	*Z-form*
Helix sense	Right-handed	Right-handed	Left-handed
Repeating unit	1 bp	1 bp	2 bp
Rotation/bp (Twist angle)	33.6°	34.3°	60°/2
Mean bp/turn	10.7	10.4	12
Base pair tilt	20°	–6°	7°
Rise/bp along axis	2.3Å	3.32Å	3.8Å
Pitch/turn of helix	24.6Å	33.2Å	45.6Å
Mean propeller twist	+18°	+16°	0°
Glycosidic bond conformation	*Anti*	*Anti*	*Anti* for C, *Syn* for G
Sugar pucker	C3'-endo	C2'-endo	C:C2'-endo, G:C3'-endo
Diameter	23Å	20Å	18Å
Major groove	Narrow and deep	Wide and deep	Flat
Minor groove	Wide and shallow	Narrow and deep	Narrow and deep

1.6.3 Triplex DNA

In certain circumstances (e.g. low pH), a DNA sequence containing a long segment consisting of a polypurine strand, hydrogen bonded to a polypyrimidine strand and form a *triple helix*. The triple helix will be written as (*dT*).(dA).(dT) with the third strand in italics. Triple-stranded DNA is formed by laying a third strand into the major groove of DNA. A third strand makes a hydrogen bond to another surface of the duplex. The third strand pairs in a Hoogsteen base-pairing scheme. The central strand of the triplex must be purine rich. Thus, triple-stranded DNA requires a homopurine: homopyrimidine region of DNA. If the third strand is purine rich, it forms *reverse Hoogsteen* hydrogen bonds in an antiparallel orientation with the purine strand of the Watson-Crick helix. If the third strand is pyrimidine rich, it forms *Hoogsteen bonds* in a parallel orientation with the Watson-Crick-paired purine strand.

Triple helix can be intermolecular or intramolecular. In the *intermolecular Pu*.Pu.Py triple helix, the poly-purine third strand is organized antiparallel with respect to the purine strand of the original Watson-Crick duplex. In the intermolecular *Py*.Pu.Py triplex, the polypyrimidine third strand is organized parallel with respect to the purine strand and the phosphate backbone is positioned.

This page intentionally left blank.

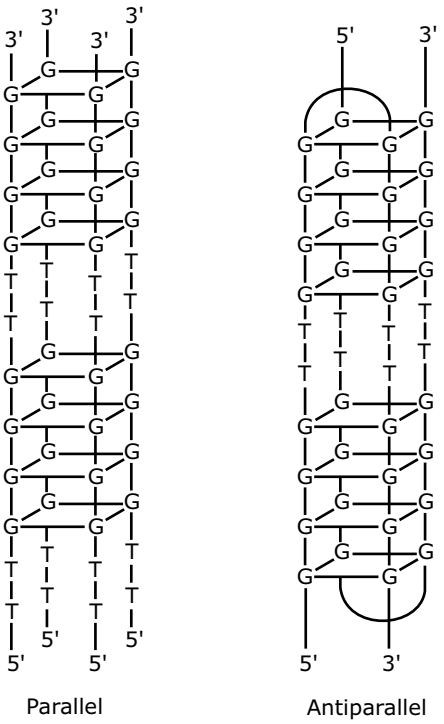


Figure 1.47 G-Quadruplex DNA. Quadruplex structures may be parallel or antiparallel.

1.6.5 Stability of the double helical structure of DNA

First : Internal and external hydrogen bonds stabilize the double helix. The two strands of DNA are held together by H-bonds that form between the complementary purines and pyrimidines, two H-bonds in an A:T pair and three H-bonds in a G:C pair, while the polar atoms in the sugar-phosphate backbone form external H-bonds with surrounding water molecules.

Second : The core of the helix consists of the base pairs, which, in addition to being H-bonded, stack together through *stacking interactions*. These interactions include *hydrophobic interactions* and van der Waals interactions between base pairs that contribute *significantly* to the overall stability. Base stacking helps to minimize contact of the bases with water.

Third : The negatively charged phosphate groups are all situated on the exterior surface of the helix in such a way that they have minimal effect on one another and are free to interact electrostatically with cations in solution such as Mg^{2+}.

1.6.6 Thermal denaturation

When duplex DNA molecules are subjected to specific conditions of pH, temperature or ionic strength that disrupt the hydrogen bonds and stacking interactions, the strands are no longer held together. That is, the double helix is **denatured** and the strands separate as individual random coils. If temperature is the denaturing agent, the double helix is said to have *melted*. DNA denaturation is a co-operative process. Denaturation process is accompanied by a change in the DNA's physical properties. Denaturation increases the relative absorbance of the DNA solution at 260 nm (as much as 40%). This increase in the absorbance is known as **hyperchromic shift**. The increased absorbance is due to the fact that the aromatic bases in DNA interact via their π-electron clouds when stacked together in the double helix. Because the absorbance of the bases at 260 nm is a consequence of π-electron

Pages 52 to 62 are not shown in this preview.

1.8 Carbohydrates

Carbohydrates are polyhydroxy aldehydes or polyhydroxy ketones, or compounds that can be hydrolyzed to them. In the majority of carbohydrates, H and O are present in the same ratio as in water, hence also called as *hydrates of carbon*. Carbohydrates are the most abundant biomolecules on Earth. Carbohydrates are classified into following classes depending upon whether these undergo hydrolysis and if so on the number of products form:

Monosaccharides are simple carbohydrates that cannot be hydrolyzed further into polyhydroxy aldehyde or ketone unit.

Oligosaccharides are polymers made up of two to ten monosaccharide units joined together by glycosidic linkages. Oligosaccharides can be classified as di-, tri-, tetra- depending upon the number of monosaccharides present. Amongst these the most abundant are the disaccharides, with two monosaccharide units.

Polysaccharides are polymers with hundreds or thousands of monosaccharide units. Polysaccharides are not sweet in taste hence they are also called *non-sugars*.

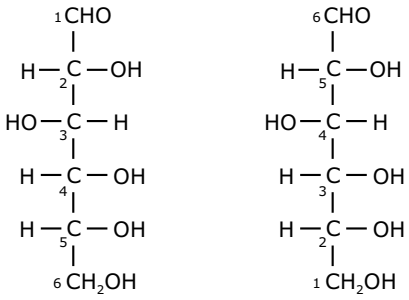
1.8.1 Monosaccharide

Monosaccharides consist of a single polyhydroxy aldehyde or ketone unit. Monosaccharides are the simple sugars, which cannot be hydrolyzed further into simpler forms and they have a general formula $C_nH_{2n}O_n$. Monosaccharides are colourless, crystalline solids that are freely soluble in water but insoluble in nonpolar solvents. The most abundant monosaccharide in nature is the D-glucose. Monosaccharides can be further sub classified on the basis of:

The number of the carbon atoms present

Monosaccharides can be named by a system that is based on the number of carbons with the suffix-*ose* added. Monosaccharides with four, five, six and seven carbon atoms are called *tetroses, pentoses, hexoses* and *heptoses,* respectively.

System for numbering the carbons : The carbons are numbered sequentially with the aldehyde or ketone group being on the carbon with the lowest possible number.



Correct Incorrect

Presence of aldehyde or ketone groups

Aldoses are monosaccharides with an aldehyde group.

Ketoses are monosaccharides containing a ketone group.

The monosaccharide *glucose* is an *aldohexose;* that is, it is a six-carbon monosaccharide (-hexose) containing an aldehyde group (aldo-). Similarly *fructose* is a *ketohexose;* that is, it is a six-carbon monosaccharide (-hexose) and containing a ketone group (keto-).

Trioses are simplest monosaccharides. There are two trioses– dihydroxyacetone and glyceraldehyde. Dihydroxyacetone is called a *ketose* because it contains a *keto* group, whereas glyceraldehyde is called an *aldose* because it contains an *aldehyde* group.

Leukotrienes are hydroxy fatty acid derivatives of arachidonic acid and do not contain a ring structure. Leukotrienes are distinguished by containing a conjugated triene double-bond arrangement. They are involved in chemotaxis, inflammation, and allergic reactions.

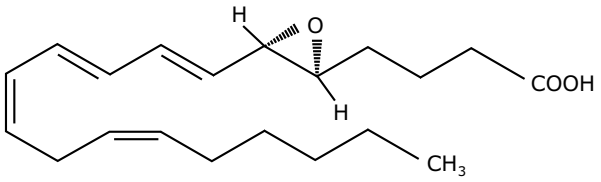


Figure 1.73 Structure of leukotriene A.

Table 1.17 Biological effects of eicosanoids

Type	*Major functions*
Prostaglandins	Mediation of inflammatory response
	Regulation of nerve transmission
	Inhibition of gastric secretion
	Sensitization to pain
	Stimulation of smooth muscle contraction
Thromboxanes	Platelet aggregation
	Aorta constriction
Prostacyclins	Thromboxane antagonists
Leukotrienes	Bronchoconstriction
	Leukotaxis

1.9.7 Plasma lipoproteins

Triacylglycerols, phospholipids, cholesterol and cholesterol esters are transported in human plasma in association with proteins as **lipoproteins**. Blood plasma contains a number of soluble *lipoproteins,* which are classified, according to their densities, into four major types. These lipid-protein complexes function as a lipid transport system because isolated lipids are insoluble in blood. There are four basic types of lipoproteins in human blood : *chylomicrons, very low density lipoproteins* (VLDL), *low density lipoproteins* (LDL), and *high density lipoproteins* (HDL). A lipoprotein contains a core of neutral lipids, which includes triacylglyerols and cholesterol esters. This core is coated with a monolayer of phospholipids in which proteins (called *apolipoprotein*) and cholesterol are embedded.

Table 1.18 Some properties of major classes of human plasma lipoproteins

Lipoprotein	*Density (g/mL)*	*Protein*	*Phospho-lipids*	*Free cholesterol*	*Cholesterol esters*	*Triacyl-glycerols*	*Apolipo-protein*
Chylomicrons	<1.006	1.5–2.5	7–9	1–3	3–5	85	A-I, C-I, B-48
VLDL	0.95–1.006	5–10	15–20	5–10	10–15	50	B-100, C-I, C-II
LDL	1.006–1.063	20–25	15–20	7–10	35–40	7–10	B-100
HDL	1.063–1.210	50–55	20–25	3–4	15	3–4	A-I, A-II, C-I

1.10 Vitamins

Vitamins are organic compounds required by the body in trace amounts to perform specific cellular functions. They can be classified according to their solubility and their functions in metabolism. The requirement for any given vitamin depends on the organisms. Not all vitamins are required by all organisms. Vitamins are not synthesized by humans, and therefore must be supplied by the diet. Vitamins may be water soluble or fat soluble. Nine vitamins (thiamines, riboflavin, niacin, biotin, pantothenic acid, folic acid, cobalamin, pyridoxine, and ascorbic acid) are classified as water soluble, whereas four vitamins (vitamins A, D, E and K) are termed fat-soluble. Except for vitamin C, the water soluble vitamins are all precursors of coenzymes.

1.10.1 Water–soluble vitamins

Thiamine (vitamin B_1)

Thiamine pyrophosphate (TPP) is the biologically active form of the vitamin, formed by the transfer of a pyrophosphate group from ATP to thiamine. Thiamine is composed of a substituted *thiazole ring* joined to a substituted *pyrimidine* by a methylene bridge.

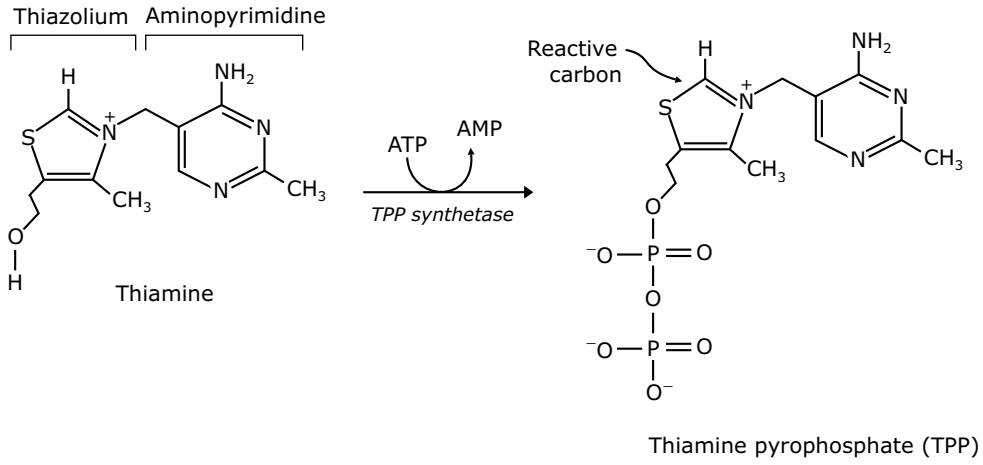


Figure 1.74 Structure of thiamine and thiamine pyrophosphate.

TPP serves as a coenzyme in the oxidative decarboxylation of α-keto acid, and in the formation or degradation of α-ketols (hydroxy ketones) by transketolase.

$$\text{Pyruvate } (\alpha\text{-keto acid}) \xrightarrow{\textit{Pyruvate decarboxylase}} \text{Acetaldehyde} + CO_2$$

$$\text{Xylulose-5-Phosphate} + \text{Ribose-5-Phosphate} \xrightarrow{\textit{Transketolase}} \text{Glyceraldehyde–3–Phosphate} + \text{Sedoheptulose–7-Phosphate}$$

Beri-Beri is a severe thiamine-deficiency syndrome found in areas where polished rice is the major component of the diet.

Riboflavin (vitamin B_2)

Riboflavin is a constituent of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). FMN is synthesized after the addition of phosphate in riboflavin and FAD formed by the transfer of an AMP moiety from ATP to FMN. FMN and FAD are each capable of reversibly accepting two hydrogen atoms, forming $FMNH_2$ or $FADH_2$. The oxidized form of the isoalloxazine structure absorbs light around 450 nm. The colour is lost, when the ring is reduced.

This page intentionally left blank.

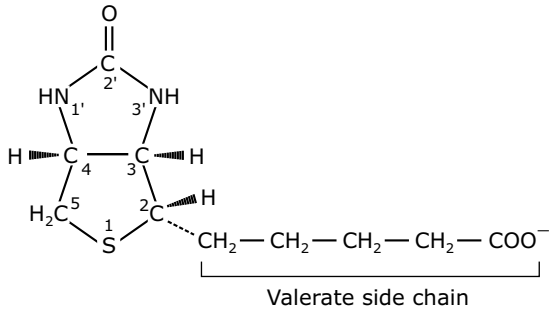


Figure 1.77 Structure of Biotin.

Most biotin-dependent carboxylations use bicarbonate as the carboxylating agent and transfer the carboxyl group to a substrate carbanion. Examples of some important biotin-dependent carboxylations are given below:

$$ATP + HCO_3^- + Pyruvate \xrightarrow{\textit{Pyruvate carboxylase}} Oxaloacetate + ADP$$

$$ATP + HCO_3^- + Acetyl\text{-}CoA \xrightarrow{\textit{Acetyl-CoA carboxylase}} Malonyl\text{-}CoA + ADP$$

$$ATP + HCO_3^- + Propionyl\text{-}CoA \xrightarrow{\textit{Propionyl-CoA carboxylase}} Methylmalonyl\text{-}CoA + ADP$$

Biotin deficiency does not occur naturally because the vitamin is widely distributed in foods. Raw egg white contains a glycoprotein, *avidin*, which tightly binds biotin and prevents its absorption from the intestine. The avidin homolog *streptavidin*, which is secreted by the *Streptomyces avidinii*, also has high affinity for biotin.

Pantothenic acid

Pantothenic acid is a component of **coenzyme A**, which is responsible for the transfer of acyl groups. Coenzyme A contains a thiol group that carries acyl compounds as activated thiol esters. Pantothenic acid is also a constituent of *acyl carrier protein* (ACP). Coenzyme A performs two main functions:

- Activation of acyl groups for transfer by nucleophilic attack.
- Activation of the α-hydrogen of the acyl group.

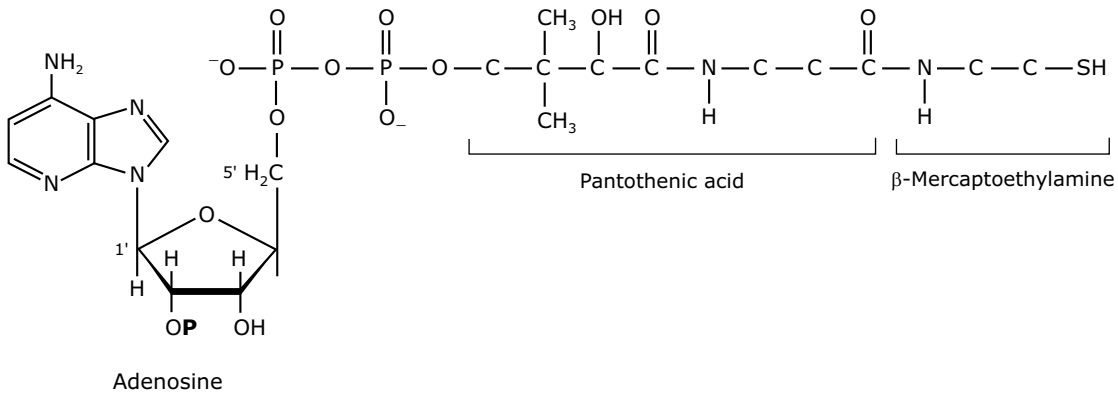


Figure 1.78 Structure of Coenzyme A (CoA-SH).

1.11.5 Enzyme inhibition

Inhibition of enzyme activity may be *irreversible* or *reversible*. Irreversible inhibitors usually bound covalently to the enzyme and destroy the functional group in the active site. Most of irreversible inhibitors are toxic substances. The antibiotic penicillin acts as an irreversible inhibitor of the enzyme *glycopeptide transpeptidase* (also known as glycoprotein peptidase). Penicillin exerts its effects by covalently reacting with an essential serine residue in the active site of glycopeptide transpeptidase, an enzyme that acts to cross-link the peptidoglycan chains during the synthesis of bacterial cell walls. Once the cell wall synthesis is blocked, the bacterial cells undergo osmotic lysis and bacterial growth is halted.

Table 1.25 Examples of irreversible enzyme inhibitors

Name	*Mode of action*
Cyanide	Reacts with enzyme metal ions (i.e. Fe, Zn, Cu); respiratory chain enzymes are primary targets.
Diisopropyl phosphofluoridate (DIPF)	Inhibits enzymes with serine at active site e.g. acetylcholinesterase.
Sarin (nerve gas)	Like DIPF
Physostigmine	Like DIPF
Parathion (insecticide)	Like DIPF, but especially inhibitory to insect acetylcholinesterase.

In *reversible inhibition*, the inhibitor can dissociate from the enzyme. Reversible inhibitors involve the non-covalent binding with enzymes. Three common types of reversible inhibition are competitive, uncompetitive and noncompetitive inhibition.

Competitive inhibition

The structure of a competitive inhibitor closely resembles that of the enzyme's normal substrate. Because of its structure, a competitive inhibitor binds reversibly to the enzyme's active site. The inhibitor forms an enzyme-inhibitor complex (EI) that is equivalent to the ES complex. The effect of a competitive inhibitor on activity can be reversed by increasing the concentration of substrate. At high [S], all the active sites are filled with substrate, and reaction velocity reaches the value observed without an inhibitor.

$$E + S \rightleftharpoons ES \longrightarrow E + P$$

$$+$$

$$I$$

$$\Updownarrow K_I$$

$$EI$$

$$K_I = \frac{[E]\,[I]}{[EI]}$$

In the presence of a competitive inhibitor, the Michaelis-Menten equation becomes

$$V = \frac{V_{max}\,[S]}{\alpha K_m + [S]} \qquad \text{Where, } \alpha = 1 + \frac{[I]}{K_I}$$

In a double-reciprocal form the equation will be

$$\frac{1}{V} = \left(\frac{\alpha K_m}{V_{max}}\right)\frac{1}{[S]} + \frac{1}{V_{max}}$$

In competitive inhibition, V_{max} stays same and K_m increases, but the inhibitor does not affect the turnover number of the enzyme. Clinical treatment of methanol poisoning is a classical example of the exploitation of competitive inhibitory mechanism. In the case of methanol poisoning, methanol in the body is converted to harmful formaldehyde by alcohol dehydrogenase. A high dose of ethanol is used to alleviate the effect of methanol because ethanol competitively binds with the active site of alcohol dehydrogenase.

Pages 103 to 116 are not shown in this preview.

Chapter 02

Bioenergetics and Metabolism

2.1 Bioenergetics

Bioenergetics is the quantitative study of the energy transductions that occur in living cells and of the nature and functions of the chemical processes underlying these transductions.

Thermodynamic principles

The *First law of thermodynamics* states that the energy is neither created nor destroyed, although it can be transformed from one form to another i.e. the total energy of a system, including surroundings, remains constant.

Mathematically, it can be expressed as:

$$\Delta U = \Delta q - \Delta w$$

ΔU is the change in internal energy,

Δq is the heat exchanged from the surroundings,

Δw is the work done by the system.

If Δq is positive, heat has been transferred to the system, giving an increase in internal energy. When Δq is negative, heat has been transferred to the surroundings, giving a decrease in internal energy. When Δw is positive, work has been done by the system, giving a decrease in internal energy. When Δw is negative, work has been done by the surroundings, giving an increase in internal energy.

The *Second law of thermodynamics* states that the total entropy of a system must increase if a process is to occur spontaneously. Mathematically, it can be expressed as:

$$\Delta S \geq \frac{\Delta q}{T}$$ where, ΔS is the change in entropy of the system

Entropy is unavailable form of energy and it is very difficult to determine it, so a new thermodynamic term called *free energy* is defined.

Free energy

Free energy or Gibb's free energy indicates the portion of the total energy of a system that is available for useful work (also known as chemical potential). The change in free energy is denoted as ΔG.

Under constant temperature and pressure, the relationship between free energy change (ΔG) of a reacting system and the change in entropy (ΔS) is expressed by following equation:

$$\Delta G = \Delta H - T\Delta S$$

Where, ΔH is the change in enthalpy and T is absolute temperature. ΔH is the measure of change in heat content of reactants and products. The change in the free energy, ΔG, can be used to predict the direction of a reaction at constant temperature and pressure.

If ΔG is negative, the reaction proceeds spontaneously with the loss of free energy (**exergonic**),

ΔG is positive, the reaction proceeds only when free energy can be gained (**endergonic**),

ΔG is 0, the system is at **equilibrium**; both forward and reverse reactions occur at equal rates,

ΔG of the reaction A $\rightarrow$ B depends on the concentration of reactant and product. At constant temperature and pressure, the following relation can be derived:

$$\Delta G = \Delta G^0 + RT \ln \frac{[B]}{[A]}$$

Where, ΔG^0 is the standard free energy change;

R is the gas constant;

T is the absolute temperature;

[A] and [B] are the actual concentrations of reactant and product.

Standard free energy change

The actual change in free energy (ΔG) during a reaction is influenced by temperature, pressure and the initial concentrations of reactants and products, and usually differs from *standard free energy change*, ΔG^0.

The chemical reaction has a characteristic standard free energy change and it is constant for a given reaction. It can be calculated from the equilibrium constant of the reaction under standard conditions i.e., at a solute concentration of 1.0M, at temperature of 25°C and at 1.0 atm pressure. The free energy change which corresponds to this standard state is known as standard free energy change, ΔG^0.

Relationship between ΔG^0 and K_{eq}

In a reaction A $\rightarrow$ B, a point of equilibrium is reached at which no further *net* chemical change takes place–that is, when A is being converted to B, B is also being converted to A, as fast as A into B. In this state, the ratio of [B] to [A] is constant, regardless of the actual concentrations of the two compounds:

$$K_{eq} = \frac{[B]_{eq}}{[A]_{eq}}$$

where K_{eq} is the equilibrium constant, and $[A]_{eq}$ and $[B]_{eq}$ are the concentrations of A and B at equilibrium. The concentration of reactants and products at equilibrium define the *equilibrium constant*, K_{eq}. The equilibrium constant K_{eq} depends on the nature of reactants and products, the temperature and the pressure. Under standard physical conditions (25°C and 1 atm pressure, for biological systems), the K_{eq} is always the same for a given reaction, whether or not a catalyst is present.

If the reaction A $\rightleftharpoons$ B is allowed to go to equilibrium at constant temperature and pressure, then at equilibrium the overall free energy change (ΔG) is zero. Therefore,

$$\Delta G^0 = -RT \ln \frac{[B]_{eq}}{[A]_{eq}}$$

So, $\Delta G^0 = -RT \ln K_{eq}$

This equation allows some simple predictions:

K_{eq}	ΔG^0	Reaction
> 1.0	Negative	proceeds forward
1.0	Zero	is at equilibrium
< 1.0	Positive	proceeds in reverse

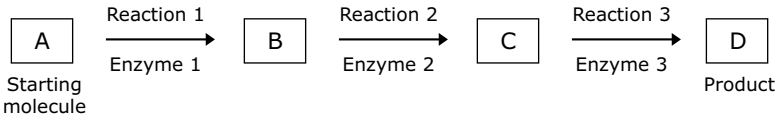
As we know, the ionic composition of an acid or base varies with pH. So, the standard free energy calculated according to the biochemistry convention is valid only at pH=7. Hence, under biochemistry convention, ΔG^0 is symbolized by $\Delta G^{0'}$ and likewise, the biochemical equilibrium constant is represented by K'_{eq}.

So, $\Delta G^{0'} = -RT \ln K'_{eq}$

Pages 119 to 121 are not shown in this preview.

2.2 Metabolism

Metabolism (derives from the Greek word for *change*) is a series of interconnected chemical reactions occurring within a cell; the chemical compounds involved in this process are known as *metabolites*. It consists of hundreds of enzymatic reactions organized into discrete pathways. These pathways proceed in a stepwise manner, transforming substrates into end products through many specific chemical *intermediates*. Each step of metabolic pathways is catalyzed by a specific enzyme.



Metabolic pathways can be linear (such as glycolysis), cyclic (such as the citric acid cycle) or spiral (such as the biosynthesis of fatty acids). Metabolism serves two fundamentally different purposes: generation of energy to drive vital functions and the synthesis of biological molecules. To achieve these, metabolic pathways fall into two categories: anabolic and catabolic pathways. *Anabolic pathways* are involved in the synthesis of compounds and endergonic in nature. *Catabolic pathways* are involved in the oxidative breakdown of larger complex molecules and usually exergonic in nature. The basic strategy of catabolic metabolism is to form ATP and reducing power for biosyntheses. Some pathways can be either anabolic or catabolic, depending on the energy conditions in the cell. They are referred to as *amphibolic pathways*. Amphibolic pathways occur at the 'crossroads' of metabolism, acting as links between the anabolic and catabolic pathways, e.g. the citric acid cycle.

Characteristics of metabolic pathways are:

1. They are irreversible.
2. Each one has a first committed step.
3. Those in eukaryotic cells occur in specific cellular locations.
4. They are regulated. Regulation occurs in following different ways:
   I. Availability of substrate; the rate of reaction depends on substrate concentration.
   II. Allosteric regulation of enzymes by a metabolic intermediate or coenzyme.
   III. By extracellular signal such as growth factors and hormones that act from outside the cell in multicellular organisms; changes the cellular concentration of an enzyme by altering the rate of its synthesis or degradation.

A number of central metabolic pathways are common to most cells and organisms. These pathways, which serve for synthesis, degradation, interconversion of important metabolites, and energy conservation, are referred to as the *intermediary metabolism*.

Metabolic pathways involve several enzyme-catalyzed reactions. Most of the reactions in living cells fall into one of five general categories: oxidation-reductions; reactions that make or break carbon–carbon bonds; group transfers; internal rearrangements, isomerizations and eliminations; and free radical reactions.

Feedback inhibition and feedback repression

In *feedback inhibition* (or end product inhibition), the end product of a biosynthetic pathway inhibits the *activity* of the first enzyme that is unique to the pathway, thus controlling production of the end product. The first enzyme in the pathway is an allosteric enzyme. Its allosteric site will bind to the end product of the pathway which alters its active site so that it cannot mediate the enzymatic reaction.

The feedback inhibition is different from feedback repression. An inhibitory feedback system in which the end product produced in a metabolic pathway acts as a co-repressor and represses the *synthesis* of an enzyme that is required at an earlier stage of the pathway is called *feedback repression*.

2.3 Respiration

Living cells require an input of free energy. Energy is required for the maintenance of highly organized structures, synthesis of cellular components, movement, generation of electrical currents and for many other processes. Cells acquire free energy from the oxidation of organic compounds that are rich in potential energy.

Respiration is an oxidative process, in which free energy released from organic compounds is used in the formation of ATP. The compounds that are oxidized during the process of respiration are known as *respiratory substrates,* which may be carbohydrates, fats, proteins or organic acids. Carbohydrates are most commonly used as respiratory substrates.

During oxidation within a cell, all the energy contained in respiratory substrates is not released free in a single step. Free energy is released in multiple steps in a controlled manner and used to synthesise ATP, which is broken down whenever (and wherever) energy is needed. Hence, ATP acts as the energy currency of the cell.

During cellular respiration, respiratory substrates such as glucose may undergo *complete or incomplete* oxidation. The complete oxidation of substrates occurs in the presence of oxygen, which releases CO_2, water and a large amount of energy present in the substrate. A complete oxidation of respiratory substrates in the presence of oxygen is termed as *aerobic respiration*.

Although carbohydrates, fats and proteins can all be oxidized as fuel, but here processes have been described by taking glucose as a *respiratory substrate.* Oxidation of glucose is an exergonic process. An exergonic reaction proceeds with a net release of free energy. When one mole of glucose (180 g) is completely oxidized into CO_2 and water, approximately 2870 kJ or 686 kcal energy is liberated. Part of this energy is used for synthesis of ATP. For each molecule of glucose degraded to carbon dioxide and water by respiration, the cell makes up to about 30 or 32 ATP molecules, each with 7.3 kcal/mol of free energy.

$$C_6H_{12}O_6 + 6O_2 \longrightarrow 6CO_2 + 6H_2O + \text{Energy (ATP + Heat)}$$

The incomplete oxidation of respiratory substrates occurs under anaerobic conditions i.e. in the absence of oxygen. As the substrate is never totally oxidized, the energy generated through this type of respiration is lesser than that during aerobic respiration.

2.3.1 Aerobic respiration

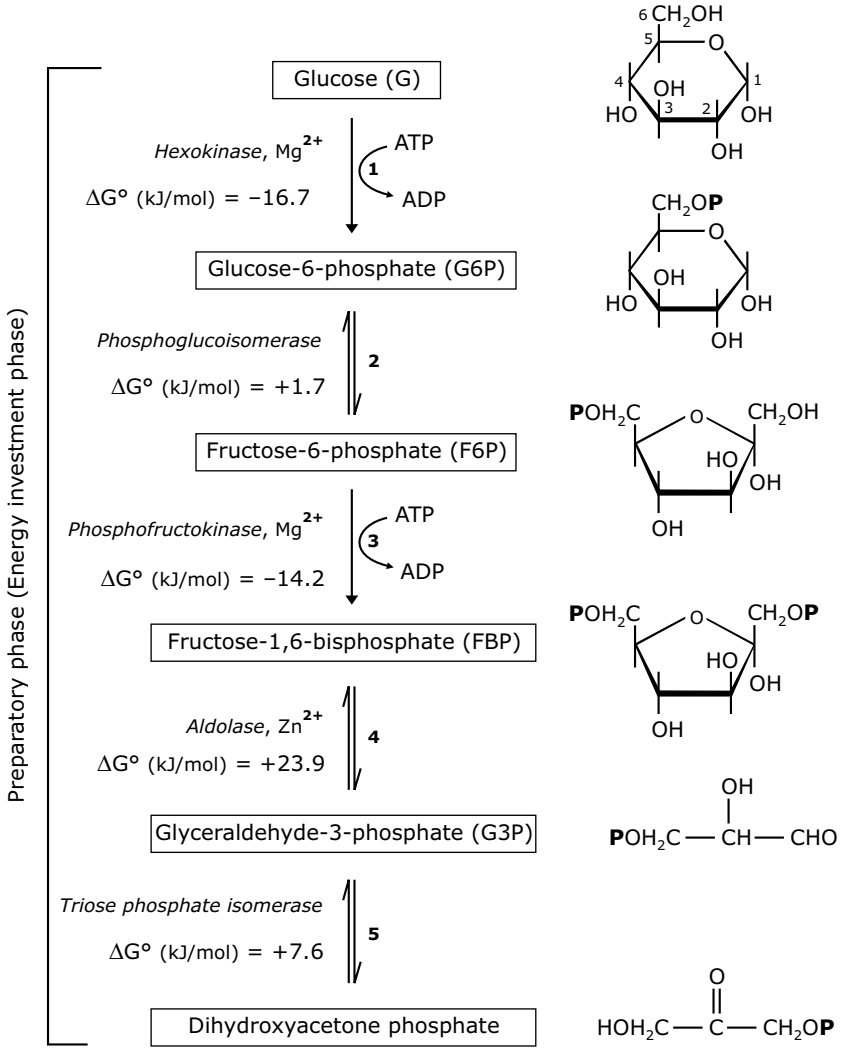
Enzyme catalyzed reactions during aerobic respiration can be grouped into three major processes: glycolysis, citric acid cycle and oxidative phosphorylation. Glycolysis takes place in the cytosol of cells in all living organisms. The citric acid cycle takes place within the mitochondrial matrix of eukaryotic cells and in the cytosol of prokaryotic cells. The oxidative phosphorylation takes place in the inner mitochondrial membrane. However, in prokaryotes, oxidative phosphorylation takes place in the plasma membrane.

Table 2.3 Intracellular location of major processes of aerobic respiration

In eukaryotes,		
Glycolysis	–	Cytosol
Citric acid cycle	–	Mitochondrial matrix
Oxidative phosphorylation	–	Inner mitochondrial membrane
In prokaryotes,		
Glycolysis	–	Cytosol
Citric acid cycle	–	Cytosol
Oxidative phosphorylation	–	Plasma membrane

2.3.2 Glycolysis

Glycolysis (from the Greek *glykys*, meaning *sweet*, and *lysis*, meaning *splitting*) also known as *Embden-Meyerhof pathway*, is an oxidative process in which one mole of glucose is partially oxidized into the two moles of pyruvate in a series of enzyme-catalyzed reactions. Glycolysis occurs in the cytosol of all cells. It is a unique pathway that occurs in both aerobic as well as anaerobic conditions and does not involve molecular oxygen.



Step 1 : (**Phosphorylation**) Glucose is phosphorylated by ATP to form a glucose 6-phosphate. The negative charge of the phosphate prevents the passage of the glucose 6-phosphate through the plasma membrane, trapping glucose inside the cell. This *irreversible* reaction is catalyzed by *hexokinase*. Hexokinase is present in all cells of all organisms. Hexokinase requires divalent metal ions such as Mg^{2+} or Mn^{2+} for activity. Hepatocytes and β-cells of the pancreas also contain a form of hexokinase called **glucokinase** (hexokinase D). Hexokinase and glucokinase are isozymes. Glucokinase is present in liver and beta-cells of the pancreas and has a high K_m and V_{max} as compared to hexokinase.

Step 2 : (**Isomerization**) A readily reversible rearrangement of the chemical structure (isomerization) moves the carbonyl oxygen from carbon 1 to carbon 2, forming a ketose from an aldose sugar. Thus, the isomerization of glucose 6-phosphate to fructose 6-phosphate is a conversion of an aldose into a ketose.

Pages 125 to 130 are not shown in this preview.

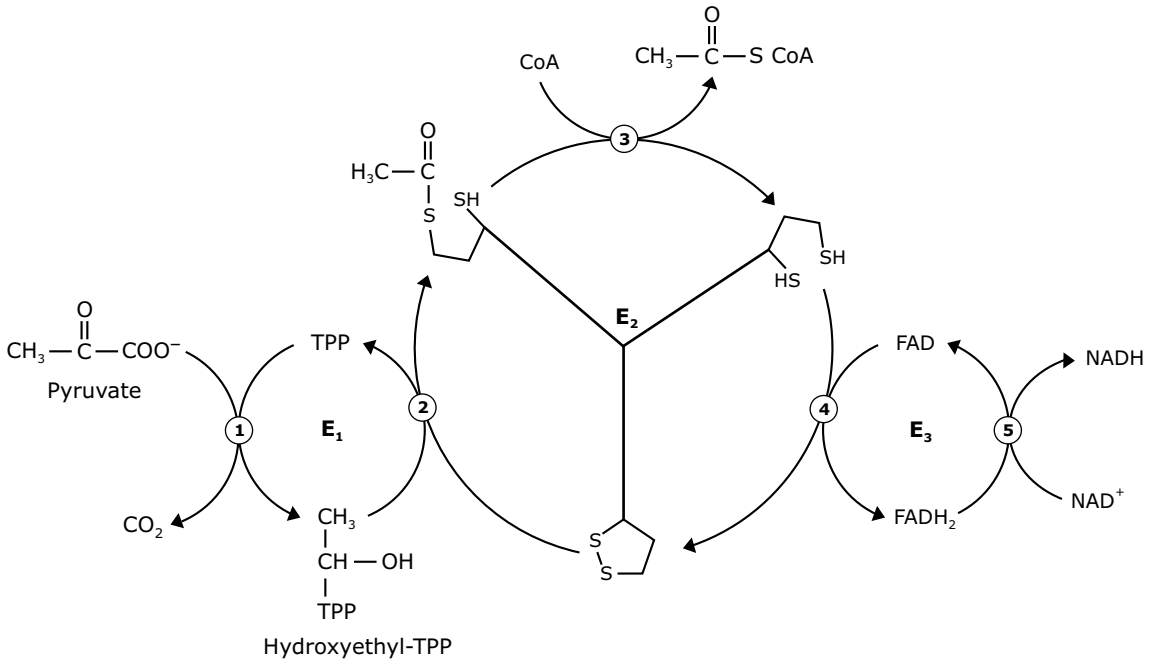


Figure 2.6 Structure of pyruvate dehydrogenase and its catalytic activities. Catalytic activities occur in four steps:

Step 1 : Decarboxylation of pyruvate occurs with formation of hydroxy ethyl – TPP.

Step 2 : Transfer of the two carbon unit to lipoic acid.

Step 3 : Formation of acetyl-CoA.

Step 4 : Lipoic acid is re-oxidized.

2.3.4 Krebs cycle

Krebs cycle (also known as the citric acid cycle or tricarboxylic acid cycle) was discovered by H. A. Kreb, a German born British Biochemist, who received the Nobel prize in 1953. This cycle occurs in the matrix of mitochondria (cytosol in prokaryotes). The whole cycle is explained in the following figure. The net result of Krebs cycle is that for each acetyl group entering the cycle as acetyl-CoA, two molecules of CO_2 are produced.

Step 1 : The Krebs cycle begins with the condensation of an oxaloacetate (four carbon unit), and the acetyl group of acetyl-CoA (two-carbon unit). Oxaloacetate reacts with acetyl-CoA and H_2O to yield citrate and coenzyme A. This reaction, which is an aldol condensation followed by a hydrolysis, is catalyzed by *citrate synthase.*

Step 2a and 2b : An isomerization reaction, in which water is first removed and then added back, moves the hydroxyl group from one carbon atom to its neighbour. The enzyme catalyzing this step, *aconitase* (nonheme iron protein), is the target site for the toxic compound **fluoroacetate** (used as a pesticide). Fluoroacetate blocks the citric acid cycle by its metabolic conversion of *fluorocitrate*, which is a potent inhibitor of aconitase.

Step 3 : Isocitrate is oxidized and decarboxylated to α-ketoglutarate. In the first of four oxidation steps in the cycle, the carbon carrying the hydroxyl group is converted to a carbonyl group. The immediate product is unstable, losing CO_2 while still bound to the enzyme. The oxidative decarboxylation of isocitrate is catalyzed by *isocitrate dehydrogenase.*

Step 4 : A second oxidative decarboxylation reaction results in the formation of succinyl-CoA from α-ketoglutarate. *α-ketoglutarate dehydrogenase* catalyzes this oxidative step and produces NADH, CO_2, and a high-energy thioester bond to coenzyme A.

Pages 132 to 153 are not shown in this preview.

integrity requires a plentiful supply of reduced glutathione (GSH), a *Cys*-containing tripeptide (γ-glutamyl-cysteinylglycine). A major function of GSH in the erythrocyte is to eliminate H_2O_2 and organic hydroperoxides. H_2O_2, a toxic product of various oxidative processes, reacts with double bonds in the fatty acid residues of the erythrocyte cell membrane to form organic hydroperoxides. These, in turn, result in premature cell lysis. Peroxides are eliminated through the action of glutathione peroxidase, yielding glutathione disulfide (GSSG). So, G6PD deficiency results in hemolytic anemia caused by the inability to detoxify oxidizing agents.

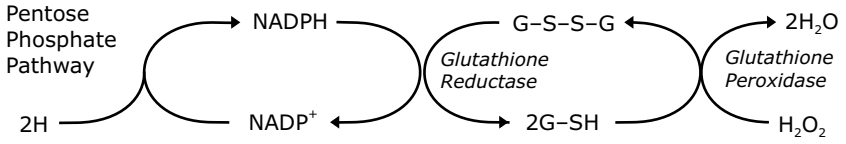


Figure 2.30 Role of the pentose phosphate pathway in the reduction of oxidized glutathione.

2.6 Entner–Doudoroff pathway

Entner-Doudoroff pathway is an alternative pathway that catabolizes glucose to pyruvate using a set of enzymes different from those used in either glycolysis or the pentose phosphate pathway. This pathway, first reported by Michael Doudoroff and Nathan Entner, occurs only in prokaryotes, mostly in gram-negative bacteria such as *Pseudomonas aeruginosa, Azotobacter, Rhizobium*.

In this pathway, glucose phosphate is oxidized to 2-keto-3-deoxy-6-phosphogluconic acid (KDPG) which is cleaved by *2-keto-3-deoxyglucose-phosphate aldolase* to pyruvate and glyceraldehyde-3-phosphate. The latter is oxidized to pyruvate by glycolytic pathway where in two ATPs are produced by substrate level phosphorylations. This process yields one ATP as well as one NADH and one NADPH for every glucose molecule.

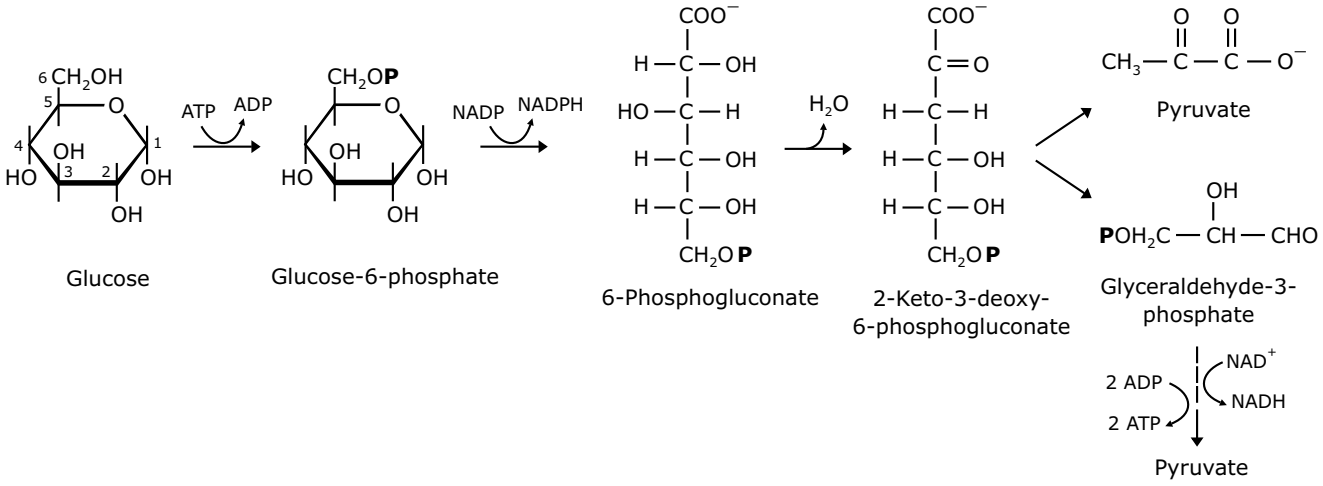


Figure 2.31 Entner-Doudoroff pathway.

2.7 Photosynthesis

Photosynthesis is a physiochemical process by which photosynthetic organisms convert light energy into chemical energy in the form of reducing power (as NADPH) and ATP, and use these chemicals to drive carbon dioxide fixation.

Pages 155 to 161 are not shown in this preview.

molecules act together as one **photosynthetic unit** in which only one member of the group- the *reaction center chlorophyll*- actually transfers electrons to an electron acceptor. The majority of the *chlorophyll* molecules serve as an *antenna complex,* collecting light and transferring the energy to the *reaction center*, where the photochemical reaction takes place.

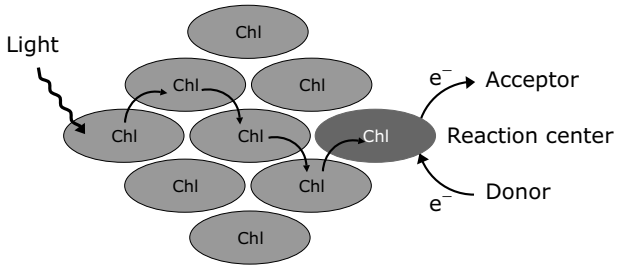


Figure 2.40 Simplified representation of the photosynthetic unit consisting of the light-harvesting antenna chlorophyll molecules and the reaction center, small arrows in the chlorophyll antenna represent transfer of excitation energy.

2.7.5 Hill reaction

In 1937, Robert Hill found that in the presence of light, isolated chloroplast from green leaves reduce a variety of compounds. Hill's isolated chloroplast did not evolve O_2 when illuminated, but did so when the suitable electron acceptor (oxidants) like potassium ferrioxalate or potassium ferricyanide was added to the illuminated suspension. This phenomenon is known as *Hill reaction*. It is a light-driven transfer of electrons from water to non-physiological oxidants (Hill reagent). One of the non-physiological oxidants used by Hill was the dye 2,6-dichlorophenolindophenol (DCPIP), now called a *Hill reagent*, which in its oxidized form is blue and in its reduced form is colourless. Later S. Ochoa showed that $NADP^+$ is the biological electron acceptor in the chloroplast. Light reduces $NADP^+$ which in turn serves as the reducing agent for carbon fixation in the Calvin cycle.

$$2H_2O + 2NADP^+ \xrightarrow{\text{Light}} 2NADPH + 2H^+ + O_2$$

2.7.6 Oxygenic and anoxygenic photosynthesis

In anoxygenic photosynthesis, light energy is captured and converted into ATP, without the production of oxygen. Water is, therefore, not used as an electron donor. In oxygenic photosynthesis, light energy is captured and converted into ATP, with the production of oxygen. Here, synthesis of oxygen occurs due to photooxidation or photolysis of water.

Before 1930, reserchers considered carbon dioxide as source of oxygen in oxygenic photosynthetic organisms. This idea was challenged in the 1930s by C. B. van Niel of Stanford University. According to him, the O_2 produced by plants is derived from H_2O and not from CO_2. van Niel found that photosynthetic bacteria *Chromatium vinosum* assimilates CO_2 in light without evolving O_2. Such bacteria use H_2S, instead of water as an electron donor and forms sulphur instead of O_2.

$$CO_2 + 2H_2S \xrightarrow[\text{Bchl}]{\text{Light}} (CH_2O) + H_2O + 2S$$

The chemical similarity between H_2S and H_2O led van Niel to propose the general photosynthetic reaction:

$$CO_2 + 2H_2A \xrightarrow{\text{Light}} (CH_2O) + 2A + H_2O$$

where H_2A is H_2O in green plants and cyanobacteria and H_2S in photosynthetic sulfur bacteria. Thus, he hypothesized that in oxygenic photosynthetic organisms water acts as a source of oxygen.

The validity of van Niel's hypothesis was established in the year 1941 by Ruben and Kamen. They directly demonstrated by *isotopic study* using ^{18}O labeled water and CO_2 on green alga *chlorella* that the source of oxygen formed in photosynthesis is water. In the following equations, *bold* letter denotes labeled atom of oxygen (^{18}O).

Experiment 1: $CO_2 + 2H_2\mathbf{O} \longrightarrow [CH_2O] + H_2O + \mathbf{O}_2$

Experiment 2: $C\mathbf{O}_2 + 2H_2O \longrightarrow [CH_2\mathbf{O}] + H_2\mathbf{O} + O_2$

2.7.7 Concept of pigment system

In 1943, Robert Emerson and Charlton Lewis examined the action spectrum in the visible region for oxygen evolution in the green algae *Chlorella pyrenoidosa*. They found that the quantum yield remained fairly constant upto 680 nm, beyond which it declined sharply. This drop in quantum yield in the far–red region of the spectrum was called the *red drop phenomenon*. This suggests that light with a wavelength greater than 680 nm is much less efficient than light of shorter wavelengths. *Quantum yield* is the number of oxygen molecules produced per photon absorbed. The reciprocal of quantum yield is quantum requirement i.e. number of photons needed for each oxygen molecule produced.

In the another experiment, Emerson and his colleagues set up two beams of light – one in the red region (wavelengths less than 680 nm) and the other in the far red region (wavelengths greater than 680 nm). Emerson found that when the two beams were applied simultaneously, the rate of photosynthesis was 2–3 times greater than the sum of the rates obtained with each beam separately. This phenomenon is known as a **Emerson enhancement effect**. The enhancement effect suggests that photosynthesis involves two photosystems – one driven by light of long wavelength (greater than 680 nm) and other driven by light of short wavelength (less than or equal to 680 nm).

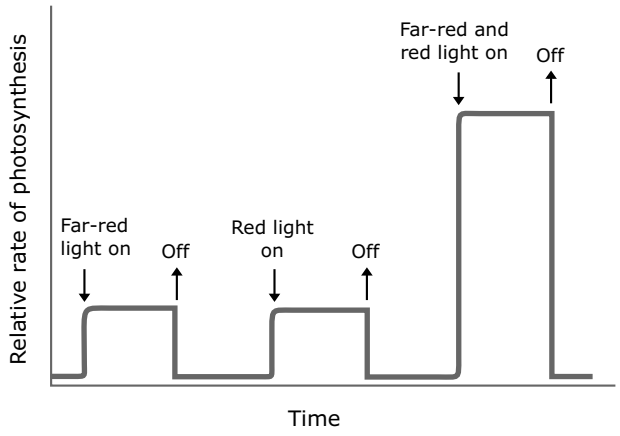


Figure 2.41 Emerson enhancement effect. The rate of photosynthesis when red and far-red light are given together is greater than the sum of the rates when they are given apart. The enhancement effect provided essential evidence in favor of the concept that photosynthesis is carried out by two different photosystems working in series but with slightly different wavelength optima.

Pigment systems

In all natural photosynthetic systems, pigment molecules are bound to proteins forming pigment–protein complexes called *pigment system* (or photosystem). The pigment systems have two components: *Photochemical reaction* center and *antenna complex.*

Photochemical reaction center carries out photochemical reaction. In all oxygenic photosynthetic organisms, the reaction center contains the special pair of *chlorophyll a* molecules associated with specific proteins that participate in photochemical reactions.

This page intentionally left blank.

2.7.8 Stages of photosynthesis

Photosynthesis is a two-stage process: one stage is dependent on the light and another independent of it.

The **light reactions**, a *light-dependent reactions* which occur in the grana of chloroplast, and require the direct energy of light to make NADPH and ATP that are used in the dark reaction. A process of formation of ATP from ADP and inorganic phosphate by utilizing light energy is called *photophosphorylation*.

The **dark reaction**, a *light-independent reaction,* occurs in the stroma of the chloroplasts when the products of the light reaction, ATP and NADPH, are used to make glyceraldehyde 3-phosphate (a triose phosphate) from reduction of carbon dioxide.

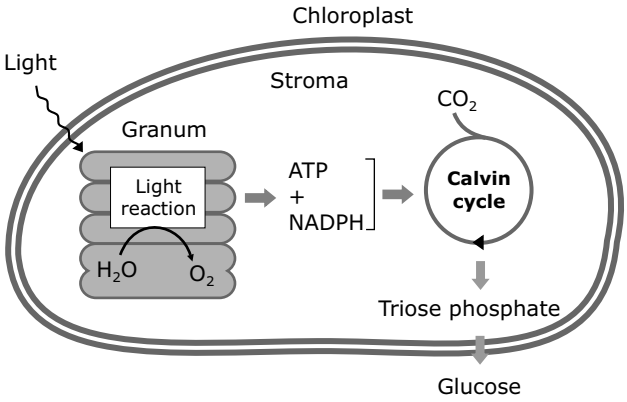


Figure 2.44 The chemical reactions in which water is oxidized to oxygen, NADP is reduced, and ATP is formed are known as the thylakoid reactions because almost all the reactions up to NADP reduction take place within the thylakoids. The carbon fixation and reduction reactions are called the stroma reactions because the carbon reduction reactions take place in the aqueous region of the chloroplast, the stroma.

Table 2.11 Differences between light and dark reactions in plants

Light reaction	*Dark reaction*
Light-dependent phase	Light-independent phase
Occurs in the grana of chloroplast	Occurs in the stroma of the chloroplasts
Photochemical reaction occurs	Chemical reaction occurs
Formation of ATP and NADPH occurs	Utilization of ATP and NADPH occurs
Oxidation of H_2O occurs	Reduction of CO_2 occurs

2.7.9 Light reactions

Light reaction (photochemical reaction) in the photosystem starts electron flow. In oxygenic photosynthetic organisms, flow of electron is of two types: *non-cyclic* as well as *cyclic*.

Noncyclic electron flow

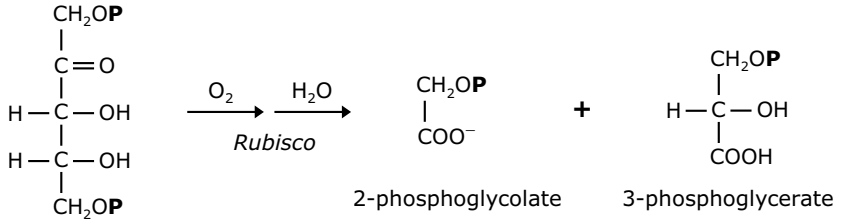
It is a light-induced electron transport from water to $NADP^+$ and a concomitant evolution of oxygen. It involves a collaboration of two photosystems: PSII and PSI. Electrons move from water through PSII to PSI and then to $NADP^+$. Electron transport leads to generation of a proton-motive force and synthesis of ATP. *Formation of ATP due to light-induced non-cyclic electron flow is called non-cyclic photophosphorylation.*

The diagram in figure 2.45 often called the *Z scheme* because of its overall form, outlines the pathway of electron flow between the two photosystems and the energy relationship in the light reactions.

Pages 166 to 177 are not shown in this preview.

2.8 Photorespiration

Otto Warburg made an observation that O_2 inhibits photosynthesis in C_3 plants. This phenomenon, originally known as the **Warburg effect**, was later recognized as the light-dependent release of CO_2 due to oxygenase activity of RuBisCo. RuBisCo is a bifunctional enzyme. It catalyzes both the carboxylation and the oxygenation of RuBP. At low CO_2 concentration, RuBisCo performs oxygenase activity. Oxygenation of RuBP leads to the production of one molecule of 3-phosphoglycerate and one of 2-phosphoglycolate.



Ribulose-1,5-bisphosphate

2-phosphoglycolate

3-phosphoglycerate

2-Phosphoglycolate produced by RuBisCo when it oxygenates RuBP cannot be utilized within the Calvin cycle. The pathway by which 2-Phosphoglycolate is further metabolized is described as **glycolate pathway** (*C_2 cycle* or *oxidative photosynthetic carbon cycle*).

The pathway involves three subcellular compartments, the chloroplasts, peroxisomes and mitochondria. The key features of the pathway are the conversion of two-carbon molecule, 2-phosphoglycolate to glycine and decarboxylation of two molecules of glycine to serine, CO_2 and NH_3. The three-carbon molecule, serine, is then converted into 3-phosphoglycerate, which re-enters the Calvin cycle. Release of CO_2 in this process *decreases* the photosynthetic output and limits the plant biomass production.

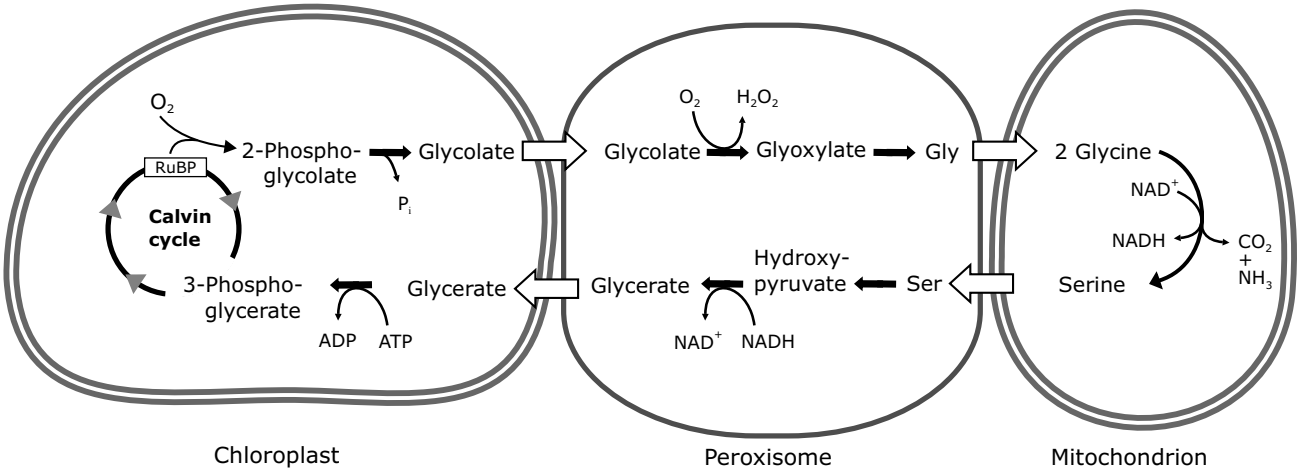


Figure 2.56 Photorespiration. Operation of the C2 cycle involves the cooperative interaction among three organelles: chloroplasts, mitochondria and peroxisomes. Glycolate transported from the chloroplast into the peroxisome are converted to glycine, which in turn is exported to the mitochondrion and transformed to serine with the concurrent release of carbon dioxide. Serine is transported to the peroxisome and transformed to glycerate. The latter flows to the chloroplast where it is phosphorylated to 3-phosphoglycerate and incorporated into the Calvin cycle.

Effect of temperature

Apart from the ambient concentrations of O_2 and CO_2, the factor influencing the enzyme's oxygenase activity most is the temperature. High temperatures promote oxygenation, and hence the photorespiration, because the solubility of CO_2 in water declines more rapidly than that of O_2 as the temperature is increased. Second the specificity factor of RuBisCo also decreases with increasing temperature.

This page intentionally left blank.

In fact, the C_4 cycle concentrates CO_2 in the bundle sheath cells, keeping the CO_2 concentration in the bundle sheath cells high enough for RuBisCo to bind carbon dioxide rather than oxygen. In this way, C_4 photosynthesis minimizes photorespiration and enhances carbohydrate production.

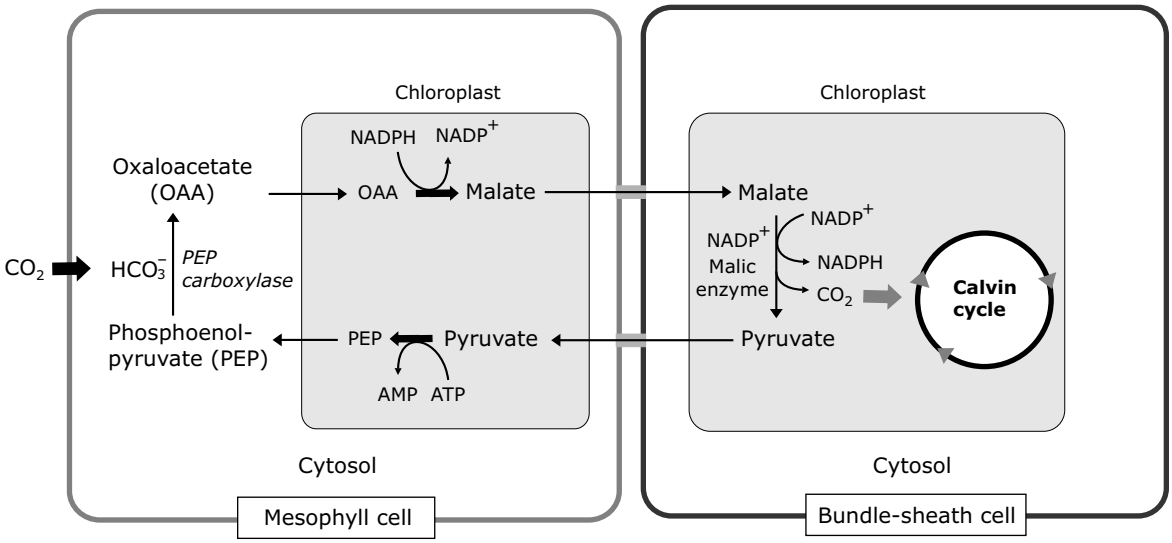


Figure 2.58 There are three distinct biochemical subtypes of C_4 cycle. These are classified on the basis of the enzyme which is employed to decarboxylate C_4 acids in the bundle sheath. These are $NADP^+$–malic enzyme type, NAD^+–malic enzyme type and PEP carboxykinase type. In the figure, reactions of the $NADP^+$–malic enzyme type C_4 cycle is described. CO_2 is transported from mesophyll cells into the bundle sheath cells by coupling it to phosphoenolpyruvate, forming oxaloacetate. Oxaloacetate is then reduced to malate, which is passed to bundle sheath cells and decarboxylated. The pyruvate product is returned to the mesophyll cells, where it is phosphorylated to regenerate phosphoenolpyruvate.

Leaf anatomy of C_4 plants

C_4 plants are unique in possessing two types of photosynthetic cell. A layer of cells surrounding the vascular bundle, the bundle-sheath, is a common structural feature, but only in C_4 plants it contains chloroplasts. The bundle-sheath is thick-walled, sometimes suberized and there is no direct access from the intercellular spaces of the mesophyll. The appearance of a wreath of cells surrounding the vasculature gives rise to the term *Kranz* (German: wreath) anatomy. The distance between bundle-sheath cells are normally only two or three mesophyll cells, so that no mesophyll cell is more than one cell away from a bundle-sheath cell. Mesophyll cells are also connected to the bundle-sheath cells by large numbers of plasmodesmata.

2.8.2 CAM pathway

Crassulacean Acid Metabolism (CAM) is a photosynthetic adaptation in succulent plants. *Succulent plants*, also known as fat plants, are xerophytic plants adapted to arid climates or soil conditions. Succulent plants store water in their leaves and stems. The storage of water often gives succulent plants a swollen or fleshy appearance than other plants, a characteristic known as succulence. The best-known succulents are cacti. These plants open their stomata during the night and close them during the day. Closing stomata during the day helps succulent plants conserve water, but it also prevents CO_2 from entering the leaves. During the night, when their stomata are open, these plants take up CO_2. Assimilation of CO_2 occurs into malic acid at night which is stored in the vacuole. This mode of carbon fixation is called crassulacean acid metabolism, or CAM, after the plant family Crassulaceae, the succulents in which the process was first discovered.

During the day time, when the light reactions can supply ATP and NADPH for the Calvin cycle, CO_2 is released from the malate for fixation through Calvin cycle. This cycle differs from the C_4 cycle. In C_4 plants, formation of the

Pages 181 to 191 are not shown in this preview.

Glycogen storage diseases

Glycogen storage diseases are caused by a genetic deficiency of one or another of the enzymes of glycogen metabolism. Many diseases have been characterized that result from an inherited deficiency of the enzyme. These defects are listed in the table.

Table 2.17 Glycogen storage diseases

Name	*Enzyme deficiency*
Von Gierke's disease	Liver glucose-6-phosphatase
Pompe's disease	Lysosomal $\alpha 1 \rightarrow 4$ and $\alpha 1 \rightarrow 6$ glucosidase (acid maltase)
Hers' disease	Liver phosphorylase
Tarui's disease	Muscle and erythrocyte phosphofructokinase 1
McArdle's disease	Muscle glycogen phosphorylase
Andersen's disease	Amylo (1,4 → 1,6) transglycosylase (Branching enzyme)

2.10 Lipid metabolism

2.10.1 Synthesis and storage of triacylglycerols

All animals and plants have the ability to synthesize triacylglycerol (TAG). In animals, many cell types and organs have the ability to synthesise triacylglycerols, but the liver and intestines are most active. Within all cell types, even those of the brain, triacylglycerols are stored as cytoplasmic *lipid droplets* (also termed *fat globules*, *oil bodies*, *lipid particles*, *adiposomes*, etc.) enclosed by a monolayer of phospholipids and hydrophobic proteins, such as the perilipins in adipose tissue or oleosins in seeds. Two main biosynthetic pathways are known, the *sn-glycerol-3-phosphate pathway*, which predominates in liver and adipose tissue, and a *monoacylglycerol pathway* in the intestines. The most important route to triacylglycerol biosynthesis is the *sn*-glycerol-3-phosphate or Kennedy pathway.

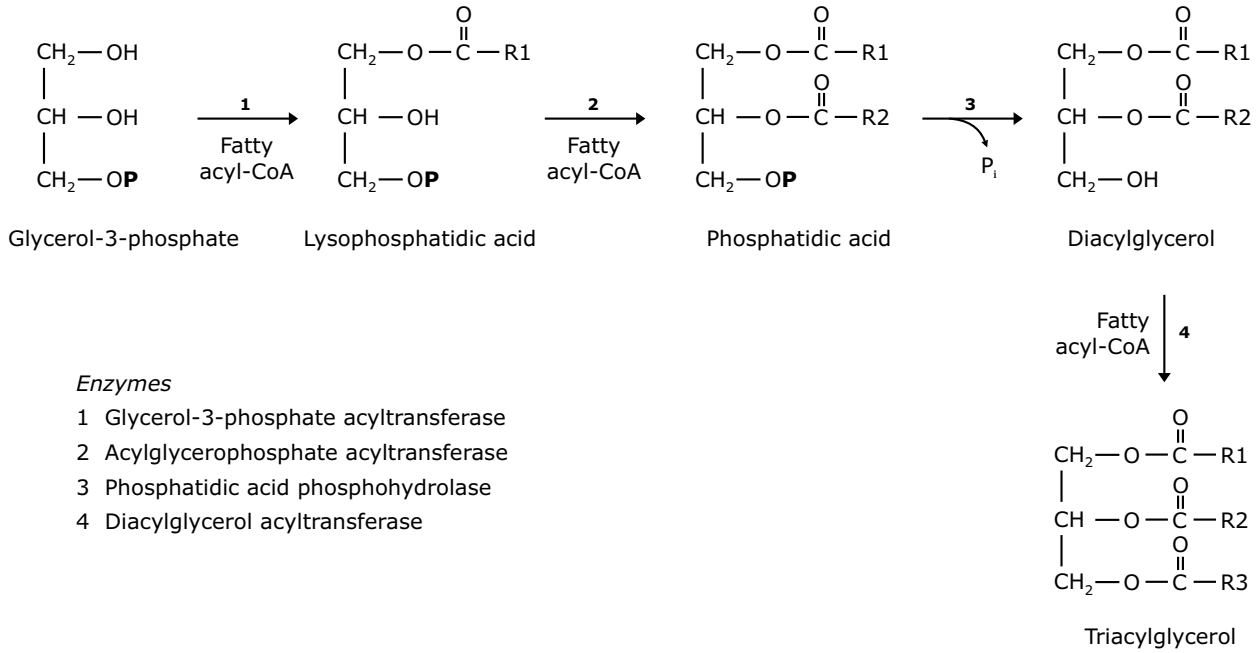


Figure 2.68 Triacylglycerol biosynthetic pathway.

Pages 193 to 210 are not shown in this preview.

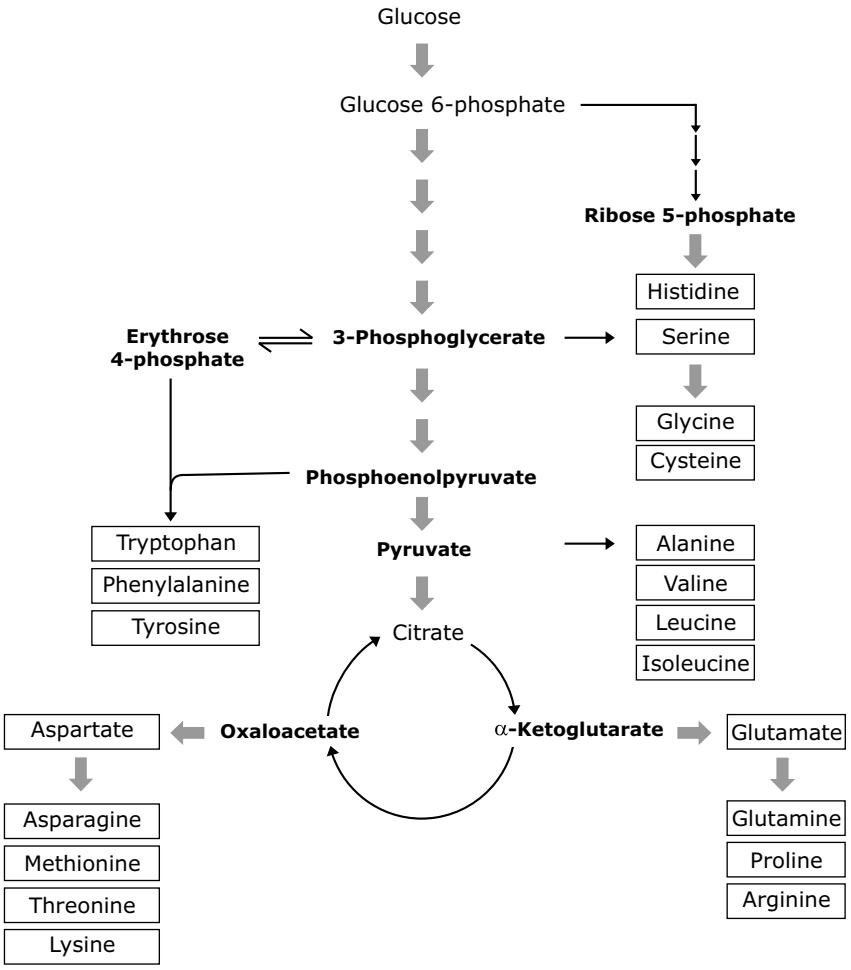


Figure 2.85 Overview of amino acid biosynthesis. The carbon skeleton precursors are derived from three sources—glycolysis, TCA cycle and pentose phosphate pathway.

2.11.2 Biological nitrogen fixation

Nitrogen is present in many forms in the biosphere. The atmosphere contains about 78% (by volume) molecular nitrogen. Acquisition of nitrogen from the atmosphere requires the breaking of an exceptionally stable triple covalent bond between two nitrogen atoms to produce ammonia (NH_3) or nitrate (NO_3^-). Conversion of molecular nitrogen to nitrate or ammonia is termed as *nitrogen fixation*, which can be accomplished by both industrial and natural processes. Natural processes of nitrogen fixation includes *lightning*, *photochemical reactions* and *biological nitrogen fixation*. Approximately 90% of nitrogen fixation is biological nitrogen fixation, in which prokaryotic organisms fix molecular nitrogen into ammonium ions. It is a reductive biosynthetic process. Few prokaryotic organisms (termed as nitrogen fixing organisms or *diazotroph*) are capable of biological nitrogen fixation only. Eukaryotic organisms are unable to fix nitrogen. The biological reaction of nitrogen fixation generates at least one mole of H_2 in addition to two moles of NH_3 for each mole of nitrogen molecule. Hence, total eight electrons are required in reduction of one mole of nitrogen to two moles of NH_3.

$$N_2 \overset{8e^- + 8H^+}{\rightleftharpoons} 2NH_3 + H_2$$

Pages 212 to 233 are not shown in this preview.

Chapter 03

Cell Structure and Functions

3.1 What is a Cell?

The basic structural and functional unit of cellular organisms is the *cell*. It is an aqueous compartment bound by cell membrane, which is capable of independent existence and performing the essential functions of life. All organisms, more complex than viruses, consist of cells. Viruses are *noncellular* organisms because they lack cell or cell-like structure. In the year 1665, Robert Hooke first discovered cells in a piece of cork and also coined the word *cell*. The word cell is derived from the Latin word *cellula,* which means small compartment. Hooke published his findings in his famous work, *Micrographia*. Actually, Hooke only observed cell walls because cork cells are dead and without cytoplasmic contents. Anton van Leeuwenhoek was the first person who observed living cells under a microscope and named them animalcules, meaning *little animals.*

On the basis of the internal architecture, all cells can be subdivided into two major classes, *prokaryotic cells* and *eukaryotic cells*. Cells that have unit membrane bound nuclei are called eukaryotic, whereas cells that lack a membrane bound nucleus are prokaryotic. Eukaryotic cells have a much more complex intracellular organization with internal membranes as compared to prokaryotic cells. Besides the nucleus, the eukaryotic cells have other membrane bound **organelles** (*little organs*) like the endoplasmic reticulum, Golgi complex, lysosomes, mitochondria, microbodies and vacuoles. The region of the cell lying between the plasma membrane and the nucleus is the **cytoplasm**, comprising the cytosol (or cytoplasmic matrix) and the organelles. The prokaryotic cells lack such unit membrane bound organelles.

Cell theory

In 1839, Schleiden, a German botanist, and Schwann, a British zoologist, led to the development of the *cell theory* or *cell doctrine*. According to this theory *all living things are made up of cells* and *cell is the basic structural and functional unit of life.* In 1855, Rudolf Virchow proposed an important extension of cell theory that *all living cells arise from pre-existing cells* (*omnis cellula e cellula*). The cell theory holds true for all cellular organisms. Non-cellular organisms such as virus do not obey cell theory. Over the time, the theory has continued to evolve. The modern cell theory includes the following components:

- All known living things are made up of one or more cells.
- The cell is the structural and functional unit of life.
- All cells arise from pre-existing cells by division.
- Energy flow occurs within cells.
- Cells contain hereditary information (DNA) which is passed from cell to cell.
- All cells have basically the same chemical composition.

Evolution of the cell

The earliest cells probably arose about 3.5 billion years ago in the rich mixture of organic compounds, the *primordial soup*, of prebiotic times; they were almost certainly chemoheterotrophs. Primitive heterotrophs gradually acquired

the capability to derive energy from certain compounds in their environment and to use that energy to synthesize more and more of their own precursor molecules, thereby becoming less dependent on outside sources of these molecules-less extremely heterotrophic. A very significant evolutionary event was the development of photosynthetic ability to fix CO_2 into more complex organic compounds. The original electron (hydrogen) donor for these photosynthetic organisms was probably H_2S, yielding elemental sulfur as the byproduct, but at some point, cells developed the enzymatic capacity to use H_2O as the electron donor in photosynthetic reactions, producing O_2. The cyanobacteria are the modern descendants of these early photosynthetic O_2 producers.

One important landmark along this evolutionary road occurred when there was a transition from small cells with relatively simple internal structures - the so-called prokaryotic cells, which include various types of bacteria - to a flourishing of larger and radically more complex *eukaryotic* cells such as are found in higher animals and plants. The fossil record shows that earliest eukaryotic cells evolved about 1.5 billion years ago. Details of the evolutionary path from prokaryotes to eukaryotes cannot be deduced from the fossil record alone, but morphological and biochemical comparison of modern organisms has suggested a reasonable sequence of events consistent with the fossil evidence.

Three major changes must have occurred as prokaryotes gave rise to eukaryotes. *First*, as cells acquired more DNA, mechanisms evolved to fold it compactly into discrete complexes with specific proteins and to divide it equally between daughter cells at cell division. These DNA-protein complexes called chromosomes become especially compact at the time of cell division. *Second,* as cells became larger and intracellular membrane organelles developed. Eukaryotic cells have a *nucleus* which contains most of the cell's DNA, enclosed by a double layer of membrane. The DNA is, thereby, kept in a compartment separate from the rest of the contents of the cell, the cytoplasm, where most of the cell's metabolic reactions occur.

Finally, primitive eukaryotic cells, which were incapable of photosynthesis or of aerobic metabolism, pooled their assets with those of aerobic bacteria or photosynthetic bacteria to form symbiotic associations that became permanent. Some aerobic bacteria evolved into the mitochondria of modern eukaryotes, and some photosynthetic cyanobacteria became the chloroplasts of modern plant cells.

3.2 Structure of eukaryotic cells

3.2.1 Plasma membrane

Plasma membrane is a dynamic, fluid structure and forms the external boundary of cells. It acts as a *selectively permeable* membrane and regulates the molecular traffic across the boundary. The plasma membrane exhibits selective permeability; that is, it allows some solutes to cross it more easily than others. Different models were proposed to explain the structure and composition of plasma membranes. In 1972, Jonathan Singer and Garth Nicolson proposed **fluid-mosaic model**, which is now the most accepted model. In this model, membranes are viewed as quasi-fluid structures in which proteins are inserted into lipid bilayers. It describes both the *mosaic* arrangement of proteins embedded throughout the lipid bilayer as well as the *fluid* movement of lipids and proteins alike.

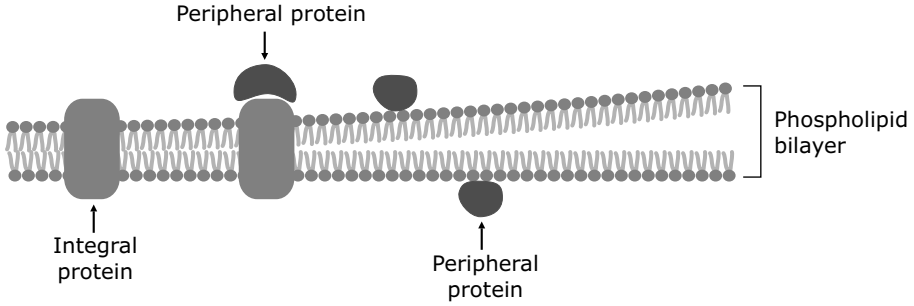


Figure 3.1 Fluid mosaic model for membrane structure. The fatty acyl chains in the lipid bilayer form a fluid, hydrophobic region. Integral proteins float in this lipid bilayer. Both proteins and lipids are free to move laterally in the plane of the bilayer, but movement of either from one face of the bilayer to the other is restricted.

This page intentionally left blank.

Glycolipids

Glycolipids contain carbohydrate (either monosaccharide or oligosaccharide) covalently attached to the lipid. These can derive from either glycerol or sphingosine. The simplest glycolipid, called a *cerebroside*, contains a single sugar residue, either glucose or galactose. *Gangliosides* are more complex glycolipids, containing a branched chain of as many as seven sugar residues. The glycolipids are found exclusively in the outer leaflet of the plasma membrane, with their carbohydrate portions exposed on the cell surface.

Sterols

The basic structure of sterol is a *steroid nucleus,* consisting of four fused rings, three with six carbons and one with five. It is planar, and relatively a rigid structure. *Cholesterol* is the major sterol present in the plasma membrane of animal cells. The plasma membrane of plant cells lacks cholesterol, but they contain other sterols like stigmasterol, sitosterol. With rare exceptions like *Mycoplasma*, bacterial plasma membrane also lacks cholesterol.

Table 3.1 Major lipid components of plasma membranes

Source	PC	PE + PS	SM	Cholesterol
Plasma membrane (human RBC)	21	29	21	26
Plasma membrane (*E. coli*)	0	85	0	0
Myelin membrane (human neurons)	16	37	13	34

Composition in mol %

PC – phosphatidylcholine; PE – phosphatidylethanolamine; PS – phosphatidylserine; SM – sphingomyelin.

Lipids are not randomly mixed in each leaflet of a bilayer. Certain lipids in the plasma membrane, particularly cholesterol and sphingolipids, are organized into aggregates called **lipid rafts**. Lipid rafts are membrane microdomains that are enriched with cholesterol and glycosphingolipids. These microdomains also contain specific proteins. In mammalian cells, lipid rafts termed *caveolae* are marked by the presence of *caveolin* proteins. The rafts in cells appear to be heterogeneous both in terms of their protein and lipid content, and can be localized to different regions of the cell. Lipid rafts have been implicated in processes as diverse as signal transduction, endocytosis and cholesterol trafficking.

When amphipathic lipids are mixed with water, three types of lipid aggregates can form. In the case of fatty acid salt, which contains only one fatty acid chain, the molecules form a small and spherical micellar structure (diameter usually <20nm) in which the hydrophobic fatty acid chains are hidden inside the **micelle**. A second type of lipid aggregate in water is the bilayer, in which two lipid monolayers combine to form a two-dimensional sheet. In the third type of lipid aggregate, lipid bilayer forms a hollow sphere called a **liposome**. Liposomes are closed, self sealing, solvent filled vesicles that are bound by only a single bilayer.

Asymmetry of lipid bilayer

The phospholipids in plasma membranes are asymmetrically distributed across the bilayer; the amine-containing phospholipids are enriched on the cytoplasmic surface of the plasma membrane, while the choline-containing and sphingolipids are enriched on the outer surface. This asymmetry is usually not absolute, except for glycolipids. In the human erythrocyte, for example, the phospholipids, such as sphingomyelin and phosphatidylcholine are mostly found in the extracytoplasmic leaflet, whereas phosphatidylserine and phosphatidylethanolamine are preferentially located on the cytoplasmic face.

The maintenance of transbilayer lipid asymmetry is essential for normal membrane function. Once lipid asymmetry has been established, it is maintained by a combination of slow transbilayer diffusion, protein-lipid interactions and protein-mediated transport.

Pages 238 to 251 are not shown in this preview.

3.3 Membrane potential

Electrical character of ion transport may be **electroneutral** i.e. electrically silent either by symport of the oppositely charged ions or antiport of similarly charged ions or **electrogenic** i.e. result in charge separation across the membrane. Electrogenic transport affects and can be affected by the membrane potential. For example, the Na^+–K^+ pump imports $2K^+$ and simultaneously exports $3Na^+$; that is, it moves 1 positive charge out of the cell. Its electrogenic operation directly contributes to the negative inside membrane potential, which is evidenced by the fact that stopping the pump using an alkaloid inhibitor, *ouabain*, causes an immediate and slight depolarization of the cell membrane.

All cells have an electrical potential difference, or *membrane potential*, across their plasma membrane. Electrical potential across cell membranes is a function of the electrolyte concentrations in the intracellular and extracellular solutions and of the selective permeabilities of the ions. Active transport of ions by ATP-driven ion pumps, generate and maintain ionic gradients. In addition to ion pumps, which transport ions against concentration gradients, plasma membrane contains channel protein that allows ions to move through it at different rates down their concentration gradient. Ion concentration gradients and selective movements of ions create a difference in electric potential or voltage across the plasma membrane. This is called *membrane potential*.

How membrane potentials arise?

To help explain how an electric potential across the plasma membrane can arise, we first consider a set of simplified experimental systems in which a membrane, which is only permeable for K^+ separates a 1 M KCl solution on the left from a 1 M KCl solution on the right. Because the concentrations of K^+ across the membrane are equal, there is no net flow of ions across the membrane and thus no electric potential is generated. If the concentration of K^+ ions across the membrane is different as shown in the figure, then K^+ ions tend to move down their concentration gradient from the left side to the right, leaving an excess of negative Cl^- ions compared with K^+ ions on the left side and generate an excess of positive K^+ ions compared with Cl^- ions on the right side. The resulting separation of charge across the membrane constitutes an electric potential, or voltage, with the left side of the membrane having excess negative charge with respect to the right. However, continued left-to-right movement of the K^+ ions eventually is inhibited by the mutual repulsion between the excess positive charges accumulated on the right side of the membrane and by the attraction of K^+ ions to the excess negative charges built up on the left side. The system soon reaches an equilibrium point at which the two opposing factors that determine the movement of K^+ ions—the membrane electric potential and the ion concentration gradient—balance each other out. At equilibrium, no net movement of K^+ ions occurs across the membrane.

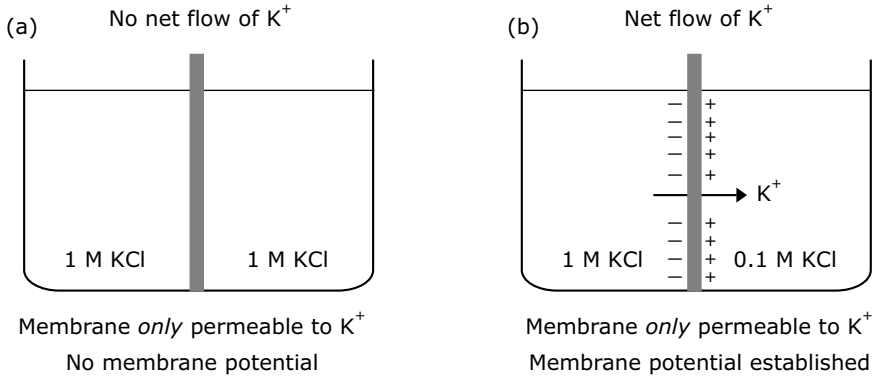


Figure 3.16 Two compartments are separated by a membrane permeable only to K^+ ions. (a) Because the concentrations in the two compartments are equal, there is no net flow of ions across the membrane and no electrical potential. (b) A difference in concentration causes K^+ ions to move from the left compartment to the right one. At equilibrium, an electrical potential is established across the membrane due to an accumulation of negative charges on the left side and positive charges on the right.

3.4 Transport of macromolecules across plasma membrane

The plasma membrane is a dynamic structure that functions to segregate the chemically distinct intracellular milieu (the cytoplasm) from the extracellular environment by regulating and coordinating the entry and exit of small and large molecules. Essential small molecules, such as amino acids, sugars and ions, can traverse the plasma membrane through the action of integral membrane protein pumps or channels. Macromolecules must be carried into the cell in membrane bound vesicles derived by the invagination and pinching-off of pieces of the plasma membrane in a process termed *endocytosis*.

3.4.1 Endocytosis

The term *endocytosis* was coined by Christian de duve in the year 1963. Endocytosis is a process whereby eukaryotic cells internalize material from their surrounding environment. Internalization is achieved by the formation of membrane-bound vesicles at the cell surface that arise by progressive invagination of the plasma membrane, followed by pinching off and release of free vesicles into the cytoplasm.

Classically, endocytosis has been divided into *phagocytosis* (cellular eating) and *pinocytosis* (cellular drinking).

Phagocytosis or cell eating (first reported by Metchnikoff) describes the internalization of large particles following particle binding to specific plasma membrane receptors and by the formation of large endocytic vesicles (generally >250 nm in diameter) called *phagosomes*. Phagocytosis occurs in specialized mammalian cells (macrophage, monocytes, neutrophils). It is an active and highly regulated process involving specific cell-surface receptors and signalling cascades mediated by Rho-family GTPases.

Pinocytosis or cell drinking (also termed as *fluid-phase endocytosis*) involves the ingestion of fluid and solutes via small vesicles (<150 nm in diameter). Uptake of material dissolved in extracellular fluid during pinocytosis occurs both selectively as well as non-selectively. Selective and efficient uptake occurs when solutes are captured by specific high-affinity receptors (receptor mediated endocytosis). In receptor-mediated endocytosis, a specific receptor on the cell surface binds tightly to the extracellular macromolecule (the ligand) that it recognizes. The plasma membrane region containing the receptor-ligand complex then undergoes endocytosis, becoming a transport vesicle. Receptor ligand complexes are selectively incorporated into the intracellular transport vesicles. Pinocytosis occurs in all cells by at least four basic mechanisms: macropinocytosis, clathrin-mediated endocytosis, caveolae-mediated endocytosis and clathrin- and caveolae independent endocytosis.

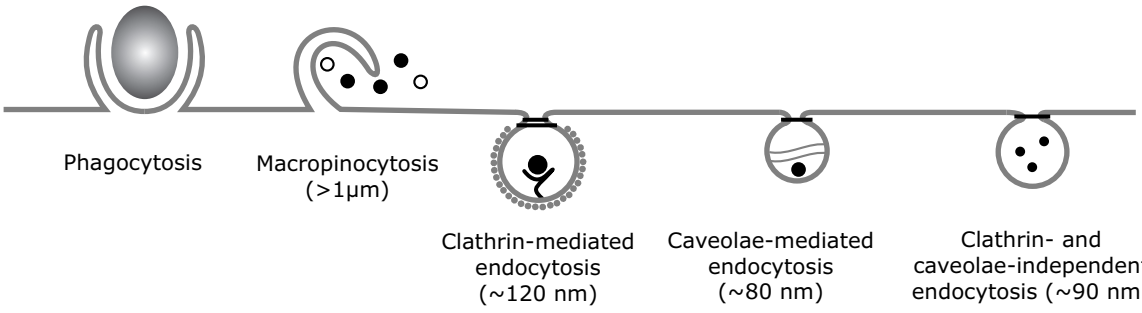


Figure 3.22 The endocytic pathways differ with regard to the size of the endocytic vesicle, the nature of the cargo (ligands, receptors and lipids) and the mechanism of vesicle formation.

Macropinocytosis

Macropinocytosis involves the membrane ruffling that is induced in many cell types upon stimulation by growth factors or other signals. Like phagocytosis, the signalling cascades that induce macropinocytosis involve Rho-family GTPases, which trigger the actin-driven formation of membrane protrusions. However, unlike phagocytosis,

Pages 263 to 266 are not shown in this preview.

plasma membrane at the opposite side. An example of transcytosis is the movement of maternal antibodies across the intestinal epithelial cells of the newborn rat. A newborn rat obtains antibodies from its mother's milk by transporting them across the epithelium of its gut. The lumen of the gut is acidic, and, at this low pH, the antibodies in the milk bind to specific receptors on the apical (absorptive) surface of the gut epithelial cells. The receptor-antibody complexes are internalized via clathrin coated vesicles and are delivered to early endosomes. The complexes remain intact and are retrieved in transport vesicles that bud from the early endosome and subsequently fuse with the basolateral domain of the plasma membrane. On exposure to the neutral pH of the extracellular fluid that bathes the basolateral surface of the cells, the antibodies dissociate from their receptors and eventually enter the newborn's bloodstream.

3.4.3 Exocytosis

Transport vesicles destined for the plasma membrane undergo fusion with the plasma membrane and release the contents outside the cell in the process called exocytosis. It may be a *constitutive secretory pathway* (carried out by all cells) or *regulated secretory pathway* (carried out by specialized cells). Examples of proteins released by such constitutive (or continuous) secretion include collagen by fibroblasts, serum proteins by hepatocytes, and antibodies by activated B-lymphocytes.

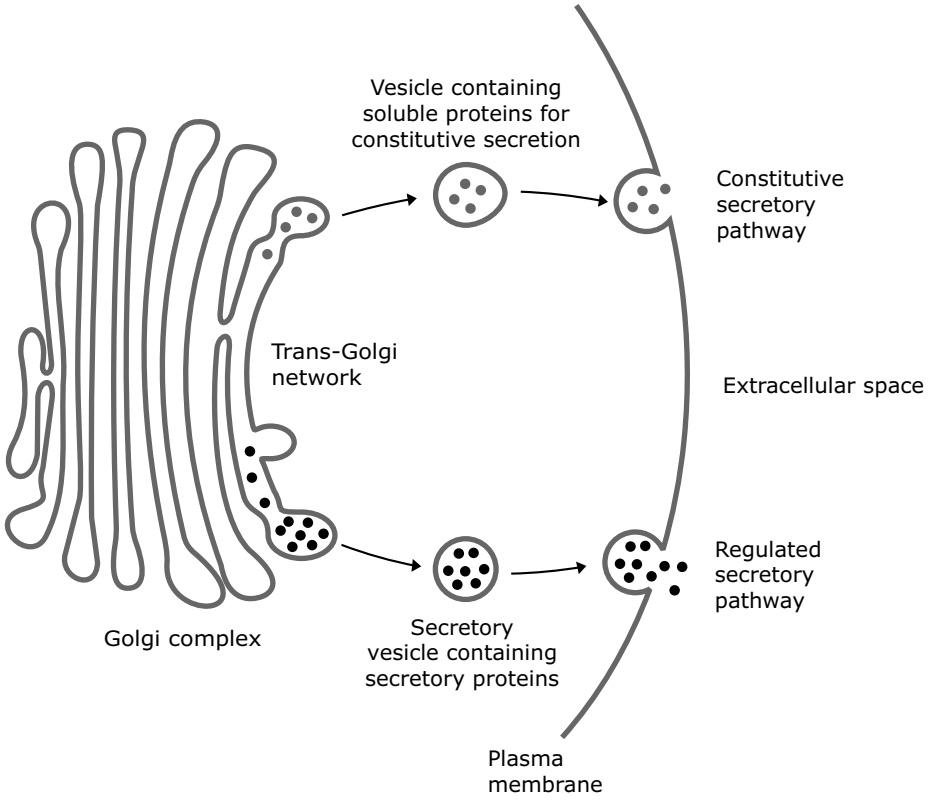


Figure 3.27 Constitutive and regulated secretory pathways. The two pathways diverge in the *trans* Golgi network. The constitutive secretory pathway operates in all cells. Many soluble proteins are continually secreted from the cell by this pathway. This pathway also supplies the plasma membrane with newly synthesized lipids and proteins. Specialized secretory cells also have a regulated secretory pathway, by which selected proteins in the *trans* Golgi network are diverted into *secretory vesicles*, where the proteins are concentrated and stored until an extracellular signal stimulates their secretion. The regulated secretion of small molecules, such as histamine and neurotransmitters, occurs by a similar pathway.

The regulated secretory pathway is found mainly in cells specialized for secreting products rapidly on demand such as hormones, neurotransmitters, or digestive enzymes. In this secretory pathway, secretory vesicles form from the *trans* Golgi network, and they release their contents to the cell exterior by exocytosis in response to specific signals. The secreted product can be either a small molecule (such as histamine) or a protein (such as a hormone or digestive enzyme). Proteins destined for secretion (called *secretory proteins*) are packaged into appropriate secretory vesicles in the *trans* Golgi network. The signal that directs secretory proteins into such vesicles is not known.

3.5 Ribosome

The ribosomes are large *ribonucleoproteins* consisting of RNAs and proteins, ubiquitous in all cells, that translate genetic information stored in the messenger RNA into polypeptides. The ribosome is approximately globular structure, its average diameter ranging from 2.5 nm (*Escherichia coli*) to 2.8 nm (mammalian cells). The functional ribosomes consist of two subunits of unequal size, known as the large and small subunits. Ribosomes consist of *rRNA* and *r-proteins*. The r-proteins are termed as L or S depending on whether the protein is from the large or small subunit.

Table 3.5 Ribosome structure and chemical composition

Property	*Prokaryote*	*Eukaryote*
Overall size	70S	80S
Small subunit	30S	40S
Number of proteins	~21	~30
RNA size (number of bases)	16S (1500)	18S (2300)
Large subunit	50S	60S
Number of proteins	~34	~50
RNA size (number of bases)	23S (2900)	28S (4200)
	5S (120)	5.8S (160)
		5S (120)

'S' stands for the *sedimentation coefficient*. It is the ratio of a velocity to the centrifugal acceleration. The sedimentation coefficient has units of second. A sedimentation coefficient of 1×10^{-13} second is defined as one Svedberg, S.

rDNA organization

In prokaryotes such as *Escherichia coli*, there are three ribosomal RNAs (16S, 23S and 5S), which are organized as a single transcription unit. In all eukaryotes studied so far, the organization of the ribosomal RNA genes is recognizably similar to that of prokaryotes, but with major differences; the size of the small subunit RNA has increased from 16S to 18S, and that of the large subunit from 23S and 28S; a new small 5.8S rRNA has become interspersed between the 18S and the 28S rRNA, and the 5S rRNA has become separated from the other rRNAs in a different transcription unit. The former transcription unit is generally referred to as the rRNA gene or the ribosomal DNA (rDNA). 5S genes are transcribed by a different RNA polymerase from rRNA genes (RNA polymerase III rather than RNA polymerase I). There are generally more copies of the 5S genes than of the rRNA genes. The human genome contains about 100 copies of rRNA genes per haploid set. Many other species, including most plants, have several thousand copies. The rRNA gene is transcribed to give a precursor the 45S pre-rRNA, which is processed in a series of post-transcriptional modifications to the mature rRNA species.

Table 3.6 Different types of ribosomes and their rRNAs

Ribosome source	*Sedimentation coefficient*	*rRNA* (large subunit/small subunit)
Bacterial	70S	5S, 23S/16S
Chloroplast	70S	5S, 23S/16S
Mitochondria (human)	55S	16S/12S
Archaebacteria	70S	5S,23S/16S
Eukaryotes (cytosol)	80S	5S, 5.8S, 28S/18S

This page intentionally left blank.

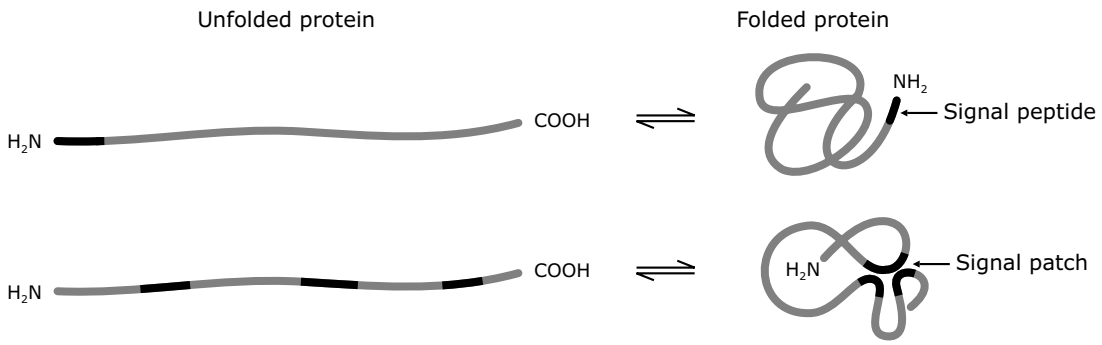


Figure 3.28 Signal peptide and signal patch.

Protein translocation describes the movement of a protein across a membrane. Within the cell, translocation of proteins from cytosol to specific organelle or organelle to cytosol and from one organelle to another occur in three different ways:

1. *Gated transport* : The protein translocation between the cytosol and nucleus occurs through the nuclear pore complexes. This process is called gated transport because the nuclear pore complexes function as selective gates that can actively transport specific macromolecules.
2. *Transmembrane transport* : In transmembrane transport, membrane-bound *protein translocators* directly transport specific proteins across a membrane from the cytosol into a organelle. The transport of selected proteins from the cytosol into the ER lumen or into mitochondria is an example of *transmembrane transport*.
3. *Vesicular transport* : In vesicular transport, proteins move from one organelle to another through transport vesicles. The transfer of proteins from the endoplasmic reticulum to the Golgi apparatus, for example, occurs in this way.

Protein translocation may occur co-translationally or post-translationally. Proteins synthesized by membrane bound ribosomes are translocated co-translationally. All proteins synthesized by membrane free ribosomes are translocated post-translationally.

3.6 Endoplasmic reticulum

Endoplasmic reticulum (ER) is the largest single membrane bound intracellular compartment. It is an extensive network of closed and flattened membrane-bound structure. The enclosed compartment is called the ER *lumen*. ER membranes are physiologically active, interact with the cytoskeleton and contain differentiated domains specialized for distinct functions.

ER membranes are differentiated into rough and smooth regions (RER and SER, respectively), depending on whether ribosomes are associated with their cytoplasmic surfaces. Regions of ER that lack bound ribosomes are called SER (sometime also called *transitional* ER). The membranes and luminal spaces of the ER are normally continuous throughout the cell and that RER and SER form an interconnected membrane system. When cells are disrupted by homogenization, the ER breaks into fragments and reseals into small vesicles called **microsomes**. Microsomes derived from RER are studded with ribosomes on the outer surface and are called *rough microsomes*. Microsomes lacking attached ribosomes are called *smooth microsome*.

Pages 271 to 280 are not shown in this preview.

Table 3.8 Coated vesicles found within eukaryotic cells

Coated vesicle	*Coat proteins*	*Transport*
Clathrin	Clathrin, AP1	Golgi complex to endosome
Clathrin	Clathrin, AP2	Plasma membrane to endosome
COPI	COPI	Golgi complex to the ER or intra Golgi complex
COPII	COPII	ER to Golgi complex

The coat proteins surrounding transport vesicles that move from the late endosome to lysosomes and to the plasma membrane have not yet been identified.

ER-resident proteins often are retrieved from the *Cis*-Golgi

As we have mentioned in the previous section that proteins entering into the lumen of the ER are of two types- *resident* proteins and *export* proteins. How, then, are resident proteins retained in the ER lumen to carry out their work?

The answer lies in a specific C-terminal sequence present in resident ER proteins. Most ER-resident proteins have a *Lys-Asp-Glu-Leu* (KDEL in the one-letter code) sequence at their C-terminus. Several experiments demonstrated that the *KDEL sequence* which acts as sorting signal, is both necessary and sufficient for retention in the ER. If this *ER retention signal* is removed from BiP, for example, the protein is secreted from the cell; and if the signal is transferred to a protein that is normally secreted, the protein is now retained in the ER. The KDEL sorting signal is recognized and bound by the *KDEL receptor* found on the ER and the *cis*-Golgi. The KDEL receptor acts mainly to retrieve proteins with the KDEL sorting signal that have escaped to the *cis*-Golgi network and returns them to the ER. The finding that most KDEL receptors are localized to the membranes of small transport vesicles shuttling between the ER and the *cis*-Golgi also supports this concept. The KDEL receptor acts mainly to retrieve soluble proteins containing the KDEL sorting signal. The retention of transmembrane proteins in the ER is carried out by short C-terminal sequences that contain two lysine residues (KKXX sequences).

The affinity of the KDEL receptor for proteins with KDEL sorting signal changes in different compartments. How can the affinity of the KDEL receptor change depending on the compartment in which it resides? The answer may be related to the differences in pH. In the low-pH environment of *cis*-Golgi and transport vesicles, the KDEL receptor has greater binding affinity with the KDEL sorting signal whereas in the neutral-pH environment of the ER, the ER proteins dissociate from the KDEL receptor due to lesser affinity.

Clearly, the transport of newly synthesized proteins from the RER to the Golgi cisternae is a highly selective and regulated process. The selective entry of proteins into membrane-bound transport vesicles is an important feature of protein targeting as we will encounter them several times in our study of the subsequent stages in the maturation of secretory and membrane proteins.

3.7 Golgi complex

The Golgi complex was first discovered in 1897 by Italian physician Camillo Golgi. The Golgi complex, also termed as *Golgi body* or *Golgi apparatus,* is a single membrane bound organelle and part of endomembrane system. It consists of five to eight flattened membrane-bound sacs called the **cisternae**. Each stack of cisternae is termed as *Golgi stack* (or dictyosome). The cisternae in *Golgi stack* vary in number, shape and organization in different cell types. The typical diagrammatic representation of three major cisternae (cis, medial and trans) as shown in the figure 3.42 is actually a simplification. In some unicellular flagellates, however, as many as 60 cisternae may combine to make up the Golgi stack. The number of Golgi complexes in a cell varies according to its function. A mammalian cell typically contains 40 to 100 stacks. In mammalian cells, multiple Golgi stacks are linked together at their edges.

Each Golgi stack has two distinct faces: a ***cis* face** (or entry face or forming face) and a ***trans* face** (or maturing face). Both *cis* and *trans* faces are closely associated with special compartments: the ***cis* Golgi network** (CGN)

Pages 282 to 285 are not shown in this preview.

After the v-SNAREs and t-SNAREs have mediated the fusion of a vesicle on a target membrane, the **NSF** (NEM Sensitive Factor) binds to the SNARE complex via adaptor proteins, **SNAPs** (Soluble NSF Attachment Proteins) proteins. NSF, a hexamer of identical subunits, and SNAPs are not necessary for actual membrane fusion but rather are required for regeneration of free SNARE proteins. NSF is a soluble ATPase and hydrolyzes ATP to dissociate the SNAREs apart.

3.9 Lysosome

Lysosomes are membrane-enclosed compartments filled with hydrolytic enzymes that are used for the controlled *intracellular* digestion of macromolecules. They contain about 40 types of hydrolytic enzymes, including proteases, nucleases, glycosidases, lipases, phospholipases, phosphatases and sulfatases. All are *acid hydrolases* because for optimal activity they require an acid environment and the lysosome provides this by maintaining a pH of about 5.0 in its interior. A H^+ pump in the lysosomal membrane uses the energy of ATP hydrolysis to pump H^+ into the lysosome, thereby maintaining the acidic pH of lumen. Lysosomes greatly vary in size and shape. There are two types of lysosomes: *Primary lysosomes* (do not contain particle or membrane for digestion) and *Secondary lysosomes* (contain particles or membranes in the process of being digested).

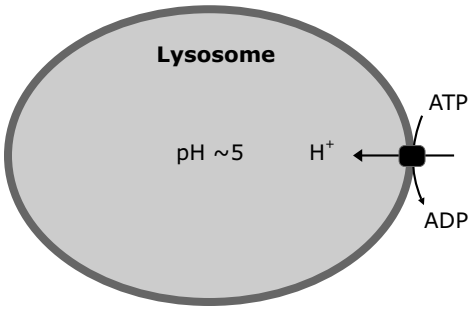


Figure 3.46 The interior of lysosomes has a pH of about 5.0. To create the low pH environment, transport proteins located in the lysosomal membrane pump hydrogen ions into the lysosome using energy supplied from ATP. All the lysosomal enzymes work most efficiently at acidic pH and collectively are termed acid hydrolases.

Lysosomes are responsible for the degradation of large particles taken up by phagocytosis and for the gradual digestion of the cell's own components by autophagy. On this basis lysosome can be divided into:

Heterophagic vacuoles (or heterolysosomes or phagolysosomes): They are formed by the fusion of primary lysosome with cytoplasmic vacuoles containing *extracellular* substances brought into the cell by an endocytic process.

Autophagic vacuoles (or autolysosomes): Autophagic vacuoles contain particles isolated from the cells own cytoplasm including mitochondria, microbodies etc.

Autophagy: A process of self-digestion

During autophagy, sequestration begins with the formation of a *phagophore.* Phagophores form *de novo* in the cytoplasm from a cup-shaped membrane that expands into a double-membrane bound *autophagosome* surrounding a portion of the cytoplasm. The autophagosome may fuse with an endosome. The product of the endosome-autophagosome fusion is called an *amphisome*. The completed autophagosome or amphisome fuses with a lysosome, which supplies acid hydrolases. The enzymes in the resulting compartment, an autolysosome, break down the inner membrane from the autophagosome and degrade the cargo. The resulting macromolecules are released and recycled in the cytosol.

This page intentionally left blank.

3.10 Vacuoles

Most plants and fungal cells contain one or several very large, fluid-filled vesicles called **vacuoles**. They are surrounded by single membrane called *tonoplast* and related to the lysosomes of animal cells, containing a variety of hydrolytic enzymes, but their functions are remarkably diverse. Like a lysosome, the lumen of a vacuole has an acidic pH, which is maintained by similar transport proteins in the vacuolar membrane. The plant vacuole contains water and dissolved inorganic ions, organic acids, sugars, enzymes and a variety of secondary metabolites. Solute accumulation causes osmotic water uptake by the vacuole, which is required for plant cell enlargement. This water uptake generates the turgor pressure.

The vacuole is different from contractile vacuole. A contractile vacuole is an organelle involved in osmoregulation. It pumps excess water out of the cell. It is found predominantly in protists (such as *Paramecium, Amoeba*) and in unicellular algae (*Chlamydomonas*). It was previously known as pulsatile or pulsating vacuole.

3.11 Mitochondria

Mitochondria (term coined by C. Benda) are *energy-converting organelles*, which are present in virtually all eukaryotic cells. They are the sites of aerobic respiration. They produce cellular energy in the form of ATP, hence they are called 'power houses' of the cell. Mitochondria are membrane-bound mobile as well as plastic organelle. Each mitochondrion is a double membrane-bound structure with outer and inner membranes. The outer membrane is fairly smooth. But the inner membrane is highly convoluted; forming folds called *cristae.* The inner membrane is also very impermeable to many solutes due to very high content of a phospholipid called *cardiolipin*. The cristae greatly increase the inner membrane's surface area. The two faces of this membrane are referred to as the *matrix side* (N-side) and the *cytosolic side* (P-side). Inner membrane contains enzyme complex called ATP synthase (or F_0-F_1 ATPase or oxysome) that makes ATP. The outer membrane protects the organelle, and contains specialized transport proteins such as *porin* which allows free passage for various molecules into the *intermitochondrial space* (the space between the inner and outer membranes) of the mitochondria. Mitochondrial porins, or voltage-dependent anion-selective channels (VDAC) allow the passage of small molecules across the mitochondrial outer membrane.

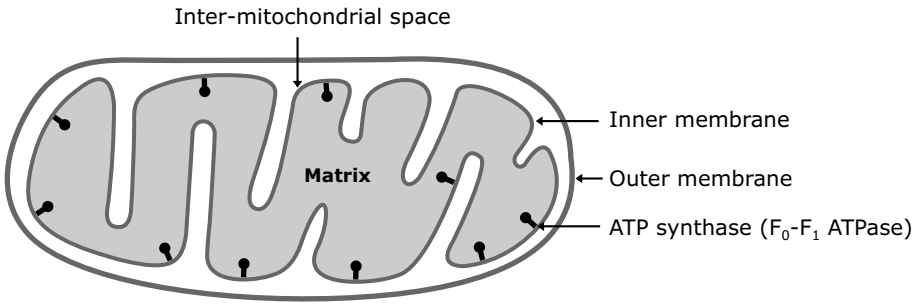


Figure 3.48 A mitochondrion has double-membraned organization and contains: the outer mitochondrial membrane, the intermembrane space (the space between the outer and inner membranes), the inner mitochondrial membrane, and the matrix (space within the inner membrane).

The matrix (large internal space) contains several identical copies of the dsDNA (as genetic material), mitochondrial ribosomes (ranging from 55S-75S), tRNAs and various proteins. Mitochondrial dsDNA is *mostly* circular. The size of mitochondrial DNA also varies greatly among different species.

Pages 289 to 305 are not shown in this preview.

3.14.8 Intermediate filaments

Intermediate filaments are rope-like cytoplasmic filaments of about 10-nm diameter. These filaments are found in many metazoans, including vertebrates, nematodes and molluscs but not in plants and fungi. Unlike the actin and tubulin proteins, the intermediate filament proteins are chemically heterogenous and show species-specific variations in molecular weight. The principal functions of intermediate filaments are structural to reinforce cells and to organize cells into tissues. Unlike microfilaments and microtubules, intermediate filaments do not participate in cell motility. All intermediate filaments share a common structural organization. The individual polypeptide of intermediate filament is an elongated molecule consisting of a non-α-helical N-terminal head domain, a central α-helical *rod domain* and a non-α-helical C-terminal tail domain. The central rod domain consists of long tandem repeats of a distinctive seven amino acid sequence called the *heptad repeat*. Polypeptide chain forms a parallel coiled coil dimeric structure with another. Two dimers then line up side by side to form an *antiparallel* tetramer of four polypeptide chains. Tetramer, the soluble subunit of intermediate filament, further organizes to form higher level organization. Tetramer is analogous to the $\alpha\beta$-tubulin heterodimer or G-actin. Unlike the actin or tubulin subunits, the intermediate filament subunits do not contain a binding site for a nucleoside triphosphate. The antiparallel arrangement of dimers implies that the tetramer, and hence the intermediate filament that it forms, is a non-polarized structure.

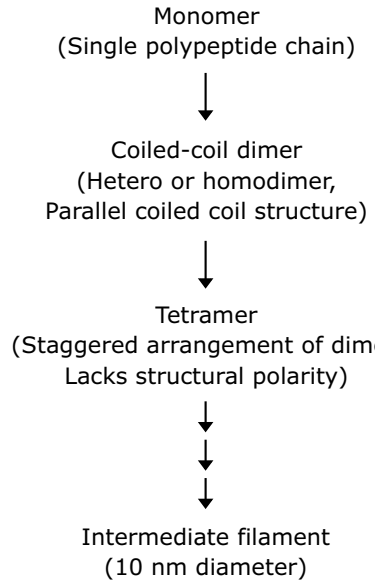


Figure 3.62 A model of intermediate filament construction.

Intermediate filament proteins are classified into four major types based on their sequences and tissue distribution: *nuclear, vimentin-like, epithelial* and *axonal.*

Types	Component polypeptides	Features
Nuclear	Lamins A, B and C	Most ubiquitous group of intermediate filaments and found exclusively in the nucleus. Lamins form a network structure that lines the inside surface of the inner nuclear membrane termed *nuclear lamina*.
Vimentin-like	Vimentin	Most widely distributed of all intermediate filament proteins is vimentin, which is typically expressed in leukocytes, blood vessel endothelial cells, some epithelial cells, and mesenchymal cells such as fibroblasts.

	Desmin	Desmin expressed in skeletal, cardiac and smooth muscles.
Epithelial	Type I keratins (acidic)	The largest group of intermediate filament proteins.
	Type II keratins (basic/neutral)	Keratins are obligatory heterodimers containing equimolar amounts of type I plus type II keratin polypeptide chains.
Axonal	Neurofilament	Forms primary cytoskeletal component in mature nerve cells.
	(NF-L, NF-M and NF-H)	In mammals, three different neurofilament proteins have been recognized: NF-L, NF-M and NF-H, for low, middle, and high molecular weight, respectively. All three are usually found in each neurofilament.

3.15 Cell junctions

Many cells in tissues are linked to one another and to the extracellular matrix at specialized contact sites called cell junctions. The cell junctions are critical to the development and functions of multicellular organisms. Cell junctions can be classified into three functional groups: occluding junctions, anchoring junctions and communicating junctions.

1. Occluding junctions

Occluding junctions seal cells together in an epithelium in a way that prevents even small molecules from leaking from one side of the sheet to the other (i.e. forms permeability barrier across epithelial cell sheets). These junctions are of two types– tight junction and septate junction.

Tight junctions (or zonula occludens) are cell-cell occluding junctions mediated by two major transmembrane proteins-*claudins and occludin.* Claudins and occludins associate with intracellular peripheral membrane proteins called ZO proteins. Tight junctions make the closest contact between adjacent cells and prevent the free passage of molecules (including ions) across an epithelial sheet in the spaces between cells. They also maintain the polarity of epithelial cells by preventing the diffusion of molecules between the apical and the basolateral regions of the plasma membrane. **Septate junctions** are the main occluding junctions in invertebrates.

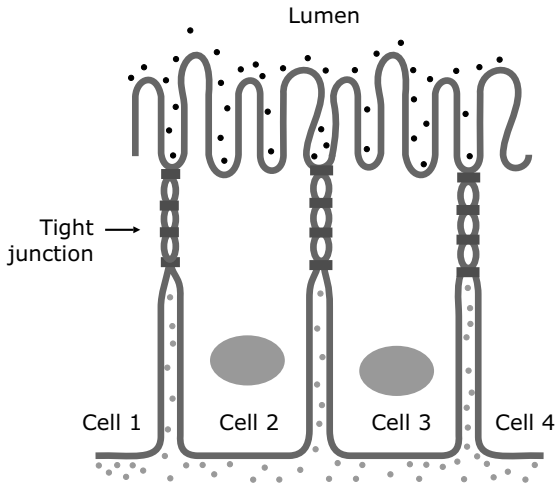


Figure 3.63 Tight junctions allow cell sheets to serve as barriers to solute diffusion. Schematic drawing showing how a small extracellular molecule present on one side of an epithelial cell sheet cannot traverse the tight junctions that seal adjacent cells together.

Pages 308 to 313 are not shown in this preview.

3.19 Nucleus

The nucleus is the controlling center of eukaryotic cell. It contains most of the genetic materials of cell. Most eukaryotic cells have one nucleus (*uninucleate*) each, but some have many nuclei (*multinucleate*) and certain cells, such as mature red blood cells, do not have it. Paramecium (unicellular ciliate protozoa) have two nuclei - a *macronucleus* and a *micronucleus*. Genes in the macronucleus control the everyday functions of the cell, such as feeding, waste removal, and maintenance of water balance. Micronucleus controls the sexual reproduction.

Nuclei differ in size depending on the cell type. Most nuclei are spherical, but multilobed nuclei are also common, such as those found in polymorphonuclear leukocytes or mammalian epididymal cells. A nucleus has four components: *Nuclear envelope, nucleolus, nucleoplasm* and *chromosomes*.

Nuclear envelope

The nuclear envelope consists of two concentric membranes called the inner and outer nuclear membranes. The outer nuclear membrane is continuous with the endoplasmic reticulum, so the space between the inner and outer nuclear membranes, the perinuclear space, is directly connected with the lumen of the endoplasmic reticulum. In addition, the outer nuclear membrane is functionally similar to the membranes of the endoplasmic reticulum and has ribosomes bound to its cytoplasmic surface. In contrast, the inner nuclear membrane carries unique proteins that are specific to the nucleus.

A network of intermediate filaments present on the nuclear side of the inner membrane is known as nuclear lamina. The nuclear lamina is made up of lamin proteins. The nuclear lamina provides the mechanical support to the nucleus. The critical function of the nuclear membrane is to act as a barrier that separates the contents of the nucleus (nucleoplasm) from the cytoplasm. The nuclear matrix or the nucleoplasm contains nucleolus and chromatin.

The nuclear envelope contains **nuclear pores** for transport of macromolecules between the cytoplasm and nucleus. Each nuclear pore is formed from an elaborate structure termed the **nuclear pore complex**. Each nuclear pore complex is a cylindrical structure comprised of eight spokes surrounding a central channel. The inner and outer membranes fuse at the nuclear pore complexes. Nuclear pore complexes are made up of some 50 to 100 different proteins. The proteins that make up the nuclear pore complex are known as **nucleoporins**. The nucleus of a typical mammalian cell contains about 3000 to 4000 pores.

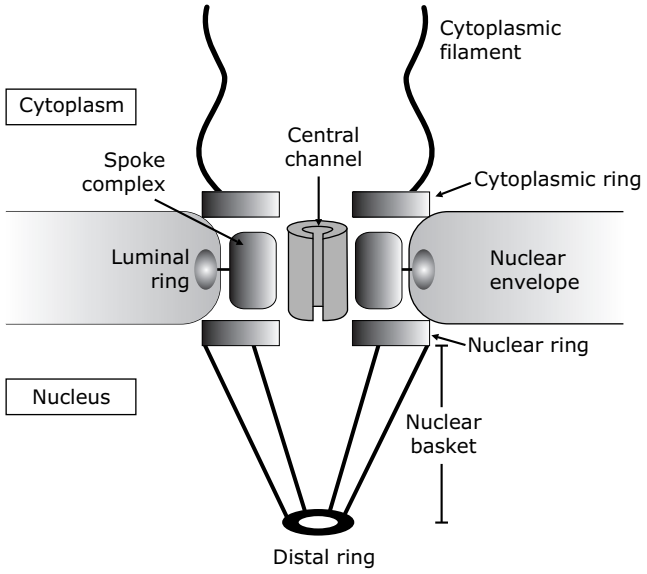


Figure 3.69 The nuclear pore complex is cylindrical and displays an octagonal symmetry. At the center of the pore is a spoke assembly of 8 annular units anchored to the membrane by luminal ring. Attached by column subunits are two rings, one facing the nucleus and the other the cytoplasm. The nucleoplasmic side of the nuclear pore complex is associated with fibrils. On the nucleoplasmic side, a nuclear basket is attached.

Pages 315 to 336 are not shown in this preview.

interacts with iron bound to the active site of the enzyme guanylyl cyclase. This increases enzymatic activity, resulting in the synthesis of the second messenger cyclic GMP, which induces muscle cell relaxation and blood vessel dilation.

Effect of Viagra

Concentration of cGMP decreases because a specific phosphodiesterase convert cGMP to the inactive 5'-GMP. Sildenafil (Viagra) causes cGMP levels to remain high by inhibiting the activity of phosphodiesterase.

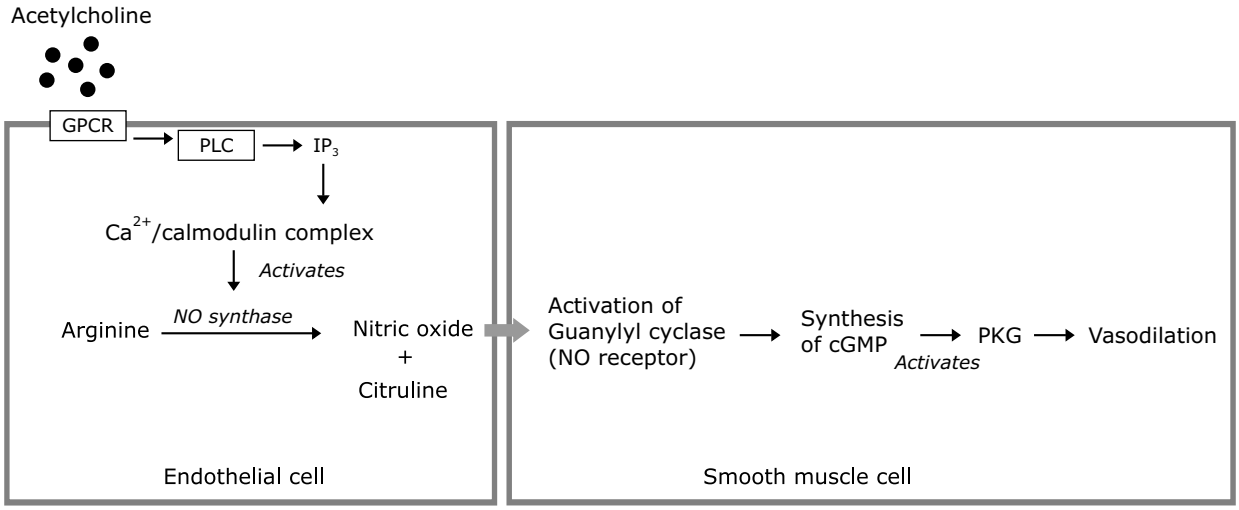


Figure 3.86 Action of nitric oxide.

3.20.7 Two–component signaling systems

Two-component signaling system is the most common form of signaling pathway that responds to extracellular events in bacteria and plants. The canonical two-component system in bacteria consists of a *sensor* that is an autophosphorylating *histidine kinase* and a *response regulator,* which transfers the phosphate from sensor kinase to a conserved aspartate within itself.

The sensor histidine kinase is located in the membrane. It can be activated by binding a ligand that is in the extracellular medium. Activation causes the kinase to autophosphorylate. The reaction transfers the phosphate from ATP on to a histidine residue in the kinase domain. The sensor interacts with an effector protein i.e. *response regulator*. The response regulator has two domains - conserved receiver domain and effector domain. The receiver domain catalyzes transfer of the phosphate group from the histidine on the sensor to an aspartic acid residue in its own domain. This activates the effector domain. The usual end target of a two-component pathway is the regulation of gene transcription.

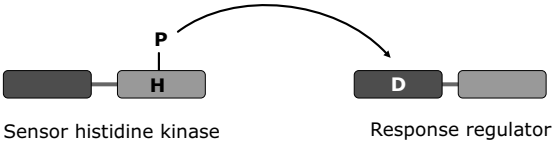
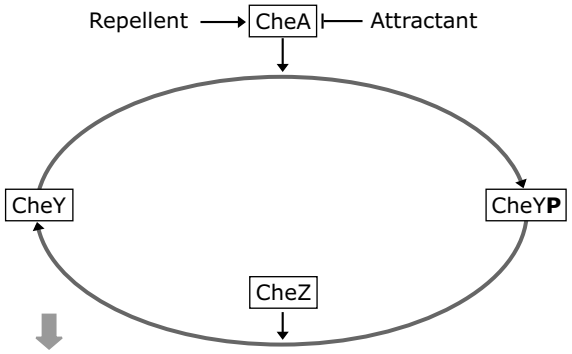


Figure 3.87 Two component system.

This page intentionally left blank.

concentration of an attractant or repellent is only transient, even if the higher level of ligand is maintained, as the bacteria *desensitize*, or *adapt*, to the increased stimulus. The adaptation is mediated by the covalent methylation. A methyltransferase, CheR, catalyzes methylation of the MCP. [In other species of bacteria such as *B. subtilis,* attractants may stimulate and repellents inhibit CheA activity].



Flagellum	Counterclockwise rotation	Clockwise rotation
Sensor	Degree of methylation increases	Degree of methylation decreases
Behavior	Run	Tumbling

Chemotaxis proteins

1. CheA, a cytoplasmic *sensor kinase*.
2. CheW, an adaptor protein linking the sensor protein with CheA.
3. CheY, the *response regulator* controlling the flageller motor.
4. CheZ, an Asp–specific protein phosphatase for signal termination.
5. CheR, a methyltransferase catalyzing methylation of the MCP.

3.20.9 Quorum sensing

The term *quorum sensing* describes a bacterial communication phenomenon that allows bacteria to communicate using secreted signal molecules to assess their population density. This process enables a population of bacteria to collectively regulate gene expression and, therefore, behaviour. In quorum sensing, bacteria assess their population density by detecting the concentration of a particular signal molecule termed autoinducer, which is correlated with cell density.

Quorum sensing is the regulation of gene expression in response to fluctuations in cell-population density. Bacteria that use quorum sensing constantly produce and secrete certain signaling molecules (called *autoinducers*). These bacteria also have a receptor for the autoinducer. When the inducer binds to the receptor, it activates the transcription of a set of genes, including those responsible for the synthesis of the autoinducer itself. The concentration of the autoinducer in the surrounding medium depends on cell-population density. As the bacterial population grows, the concentration of the autoinducer in the surroundings increases, causing more autoinducer molecules to be synthesized. The detection of a minimal threshold stimulatory concentration of an autoinducer leads to an alteration in gene expression. Both gram-positive and gram-negative bacteria use quorum sensing communication circuits to regulate a diverse array of physiological activities. These processes include symbiosis, virulence, competence, conjugation, antibiotic production, motility and sporulation. In general, Gram-negative bacteria use *N*-Acyl-L-Homoserine Lactones (AHLs) as autoinducers, and Gram-positive bacteria use processed *oligo-peptides* to communicate.

Pages 340 to 347 are not shown in this preview.

The proteasomal degradation of p53 results from its polyubiquitination by a ubiquitin ligase called *Mdm2.* In the case of DNA damage, DNA dependent protein kinase ATM (Ataxia Telangiectasia Mutated) and ATR (ATM and Rad3-related, Rad3 is a DNA dependent protein kinase of yeast) are activated. ATM primarily is a sensor of DNA defects caused by ionizing radiation, while ATR is more specialized for UV induced DNA damage and inhibitors of DNA replication. In the case of DNA damage, the rapid degradation of p53 is inhibited by ATM, which phosphorylates p53 at a site that interferes with binding to Mdm2. Hence, in response to DNA damage, p53 levels increase and arrest the cell at the G1-phase of the cell cycle. If all of the repairs have been made to the DNA, the cell divides normally and completes the cycle. However, if the cell still contains mutated or duplicated DNA sequences, it dies by a suicidal *apoptotic* mechanism to prevent its proliferation. In those cells that have mutated or lost p53 genes, the arrest at G1 does not occur, and the cells that have mutated genomes proliferate and become cancerous.

3.21.3 Replicative senescence

Normal eukaryotic cells have only a limited capacity for cell division. The process that limits the cell division has been termed *replicative senescence*. It appears to be a fundamental feature of somatic cells, with the exception of most tumor cells and possibly certain stem cells. In the early 1960s, Leonard Hayflick observed that human cells placed in tissue culture stop dividing after a limited number of cell divisions. The number of mitosis a cell is capable of undergoing in tissue culture before it stops dividing is described as the **Hayflick limit**. Telomere shortening is considered as the main causal mechanism of replicative senescence. When telomeres become critically short, the nucleus signals the cell to cease proliferation, provoking cell senescence or cell death. The telomeres are short tandemely repetitive DNA sequences that cap the ends of eukaryotic chromosomes. It serves a dual role in protecting the chromosome ends and in intracellular signaling for regulating cell proliferation. A complex of six telomere-associated proteins has been identified – the *telosome* or *shelterin complex* - that is crucial for both the maintenance of telomere structure and its signaling functions. The length of telomeric DNA is maintained by the enzyme *telomerase*. It is a reverse transcriptase that maintains the length of telomeres by overcoming 'end replication problem'. Most human somatic cells are telomerase negative and thereby experience progressive telomere shortening, at each cell division, due to the end replication problem. As human cells proliferate in culture, their telomeres get progressively shorter and shorten down to a point when they elicit a DNA-damage response. It leads to an irreversible growth arrest, a phenomenon called cellular senescence. Telomerase therefore constitutes a telomere maintenance mechanism conferring infinite replicative potential. However, telomerase is normally expressed in stem cells, tumor cells and germ-line cells, but is nearly absent in most somatic cells. Induction of telomerase synthesis bypasses normal cellular senesce in cancer cells and endows them with unlimited replicative potential.

A number of human tumors and cell immortalized in culture maintain their telomeres by a *telomerase independent* mechanism termed Alternative Lengthening of Telomeres (ALT). The available data indicate that ALT involves homologous recombination-mediated DNA replication.

3.22 Mechanics of cell division

In eukaryotes, two types of cell divisions partition the genetic material into progeny, or daughter cells. In one process called **mitosis**, a parent cell divides into two daughter cells and each receives an exact copy of the chromosomes (genetic material) in the parent cell. Since the number of chromosomes in the parent and progeny cells is the same, it is also called as equational division. In the other partitioning process, the genetic material must precisely halve so that fertilization will restore the diploid complement. This cellular process is termed **meiosis**.

3.22.1 Mitosis

Mitosis is the process that partitions newly replicated chromosomes equally into two daughter cells. The term mitosis (derived from the Greek word, meaning *thread*) was introduced by Walther Flemming in 1882. During mitosis, one round of DNA replication is followed by a single round of chromosome segregation and generate two genetically identical daughter cells.

Pages 349 to 364 are not shown in this preview.

pro-apoptotic. Mammalian Bcl2 family of proteins regulate the intrinsic pathway of apoptosis mainly by controlling the release of cytochrome *c* and other intermembrane mitochondrial proteins into the cytosol.
The anti-apoptotic Bcl2 proteins include Bcl2 itself and Bcl-X_L. Bcl2 was the first protein shown to cause an inhibition of apoptosis. It is the mammalian homologue of the CED-9 in *C. elegans*. The pro-apoptotic Bcl2 proteins consist of two subfamilies - the BH123 proteins and the BH3-*only* proteins. The main BH123 proteins are Bax and Bak, which are structurally similar to Bcl2. Important members of the BH3-*only* proteins are Bid, Bim, Bik, Bad and Bmf.
When an apoptotic stimulus triggers the intrinsic pathway, the pro-apoptotic BH123 proteins become activated and induces the release of cytochrome c and other intermembrane proteins by an unknown mechanism. In the absence of an apoptotic stimulus, anti-apoptotic Bcl2 proteins bind to and inhibit the BH123 proteins on the mitochondrial outer membrane and in the cytosol. In the presence of an apoptotic stimulus, BH3-*only* proteins are activated and bind to the anti-apoptotic Bcl2 proteins so that they can no longer inhibit the BH123 proteins. Some activated BH3-*only* proteins may stimulate mitochondrial protein release more directly by binding to and activcting the BH123 proteins.

3.24 Cancer

A normal cell undergoes regulated division, differentiation and apoptosis (programmed cell death). When normal cells have lost the usual control over their division, differentiation and apoptosis they become tumor cells. So, a tumor is the result of an abnormal proliferation of cells without differentiation and apoptosis. Tumor or neoplasm (any abnormal proliferation of cells) may be of two types: *Benign tumor* and *Malignant tumor*.

Benign and malignant tumor

In benign tumor, neoplastic cells remain clustered together in a single mass and cannot spread to other sites. It contains cells that closely resemble normal cells and that may function like normal cells.

Neoplastic cells that don't remain localized and encapsulated and becomes progressively invasive and malignant are described as *malignant tumors*. They invade surrounding normal tissues (called *invasiveness*) and spread throughout the body through circulatory or lymphatic systems (called *metastasis*). The term *cancer* refers specifically to malignant tumors.

Table 3.20 Comparison of benign and malignant tumours

Characteristics	*Benign*	*Malignant*
Differentiation	Well differentiated	Lack differentiation
Rate of growth	Slow	Rapid
Invasiveness	Absent	Present
Metastasis	Absent	Present

Both benign and malignant tumors are classified according to the type of cell from which they arise. Most cancers fall into three main groups– *Carcinomas* (tumors that arise from endodermal or ectodermal tissues), *Sarcomas* (malignancies of mesodermal connective tissues) and *Leukemia/Lymphomas* (from blood forming tissues and from cells of the immune system).

Most cancers originate from single abnormal cell i.e. *monoclonal origin.* Cancers are probably initiated by changes in the cell's DNA sequence (*genetic changes*) or change in pattern of gene expression without a change in DNA sequences (*epigenetic changes*).

Most cancers are initiated by genetic changes and majority of them are caused by changes in somatic cells and therefore are not transmitted to the next generation. About 1% of all cancers is due to genetic changes in germinal cells and is therefore inherited. About 80% of these inherited cancers are dominant in nature.

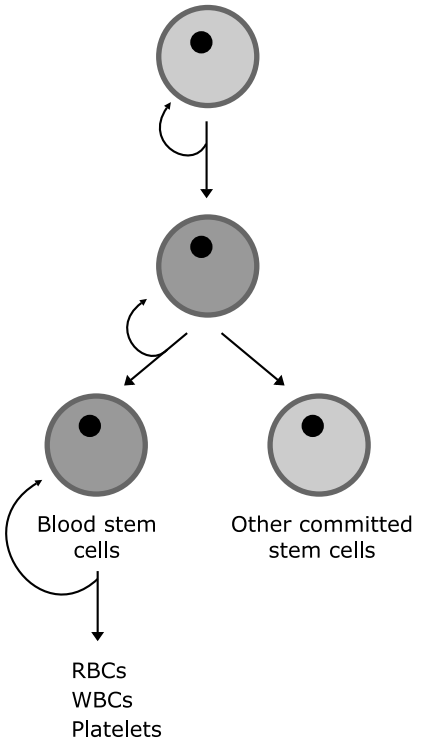
The transition of a normal cell into a tumor cell is referred to as *transformation.* The transition from a normal to a transformed state is a multisteps process involving genetic/epigenetic changes and selection of cells with the

Pages 366 to 371 are not shown in this preview.

3.25 Stem cells

Stem cells are unspecialized (undifferentiated) cells that have the ability to differentiate into other cells and self-regenerate. These cells divide to produce one daughter cell that remains a stem cell and one that divides and differentiates. Because the division of stem cells produces new stem cells as well as differentiated daughter cells, stem cells are self renewing populations of cells that can serve as a source for the production of differentiated cells throughout life. Typically, stem cells generate an intermediate cell type or types before they achieve their fully differentiated state.

The intermediate cell is called a *precursor* or *progenitor cell*. The ability to differentiate is the potential to develop into other cell types. Depending on the ability to differentiate into other cell types, stem cells can be classified as *totipotent*, *pluripotent* and *multipotent* stem cells. Totipotent stem cells are cells that can give rise to a fully functional organism as well as to every cell type of the body. Pluripotent stem cells can differentiate into nearly all cell types. Multipotent stem cells can differentiate into a limited number of closely related families of cells.



Totipotent stem cell
These cells have unlimited capability, and have the ability to form extra embryonic membranes and tissues, the embryo itself, and all post embryonic tissues and organs.

Pluripotent stem cell
These cells are capable of giving rise to most, but not all, tissues of an organism. An example is inner mass cells.

Multipotent stem cell
These cells are committed to give rise to cells that have a specific function. An example is blood stem cell.

There are two broad types of stem cells: *embryonic stem cells*, which are isolated from the inner cell mass of blastocysts, and *adult stem cells*, which are found in various tissues. Embryonic stem cells can become all cell types of the body because they are pluripotent. An *adult stem cell* (also termed as *somatic stem cell*) is an undifferentiated cell found among differentiated cells in a tissue or organ, can renew itself and differentiate to yield the major specialized cell types of the tissue or organ. The primary roles of adult stem cells in a living organism are to maintain and repair the tissue in which they are found. Unlike embryonic stem cells, which are defined by their origin (the *inner cell mass* of the blastocyst), the origin of adult stem cells in mature tissues is unknown. Most adult stem cells are multipotent. The bone marrow contains two kinds of stem cells. One population, called *hematopoietic stem cells*, forms all the types of blood cells in the body. A second population called *bone marrow stromal cells* generates bone, cartilage, fat and fibrous connective tissue. The adult brain also contains stem cells that are able to generate the brain's three major cell types—*astrocytes* and *oligodendrocytes*, which are non-neuronal cells and *neurons* or nerve cells.

Pages 373 to 376 are not shown in this preview.

Chapter 04

Prokaryotes and Viruses

4.1 General features of Prokaryotes

Prokaryotes (*pro* means before and *karyon* means kernel or nucleus) consist of *eubacteria* and the *archaea* (also termed as archaebacteria or archaeobacteria). The term eubacteria refer specifically to *bacteria*. *The informal name bacteria are occasionally used loosely in the literature to refer to all the prokaryotes, and care should be taken to interpret its meaning in any particular context.* Prokaryotes can be distinguished from eukaryotes in terms of their cell structure and molecular make-up. Prokaryotic cells have a simpler internal structure than eukaryotic cells. Although many structures are common to both cell types, some are unique to prokaryotes. Most prokaryotes lack extensive, complex, internal membrane systems. The major distinguishing characteristics of prokaryotes and eukaryotes are as follows:

Features of prokaryotic organisms

True membrane bound nucleus	–	Absent
DNA complexed with histone	–	Absent
Number of chromosomes	–	One (mostly)
Mitosis and meiosis	–	Absent
Genetic recombination	–	Partial (unidirectional transfer of DNA)
Sterol in plasma membrane	–	Absent (Except *Mycoplasma*)
Ribosome	–	70S
Unit membrane bound organelles	–	Absent
Cell wall	–	Present in *most* of prokaryotic cells. In eubacteria, it is made up of peptidoglycan.

Features of eukaryotic organisms

True membrane bound nucleus	–	Present
DNA complexed with histone	–	Present
Number of chromosomes	–	More than one
Mitosis and meiosis	–	Present
Genetic recombination	–	By crossing over during meiosis
Sterol in the plasma membrane	–	Present
Ribosome	–	80S (in cytosol) and 70S (in organelles)
Unit membrane bound organelles	–	Present
Cell wall	–	Made up of cellulose in plant and chitin in fungi. Absent in animal cells.

Prokaryotic cells show similarities with eukaryotic organelles like mitochondria and chloroplast. The endosymbiotic theory (Margulis, 1993) proposes that the mitochondria and chloroplasts of eukaryotic cells originated as symbiotic prokaryotic cells. The presence of circular, covalently closed DNA and 70S ribosomes in mitochondria and chloroplast support this theory.

Table 4.1 Similarities between prokaryotic cells and eukaryotic organelles

	Prokaryotic cells	*Eukaryotic organelles*
Nature of DNA	ds circular	ds circular
Histone protein	Absent	Absent
Ribosome type	70S	70S
Growth	Binary fission	Binary fission

4.2 Phylogenetic overview

Historically, prokaryotes were classified on the basis of their phenotypic characteristics. Prokaryotic taxonomy therefore involved measuring a large number of characteristics, including morphology and biochemical characteristics (e.g. ability to grow on different substrates, cell wall structure, antibiotic sensitivities, and many others). This contrasts with the classification of eukaryotic organisms, for which phylogenetic (evolution-based) classification was possible through the availability of fossil evidence.

A major revolution occurred with the realization that evolutionary relationships could be deduced on the basis of differences in gene sequence. The most important gene for prokaryote phylogeny is the 16S ribosomal RNA (rRNA) gene, which is present in all cells. The gene is approximately 1500 bp in length and possesses *signature sequences.* These sequences are conserved and found in the organisms of one taxonomic group but not in other groups.

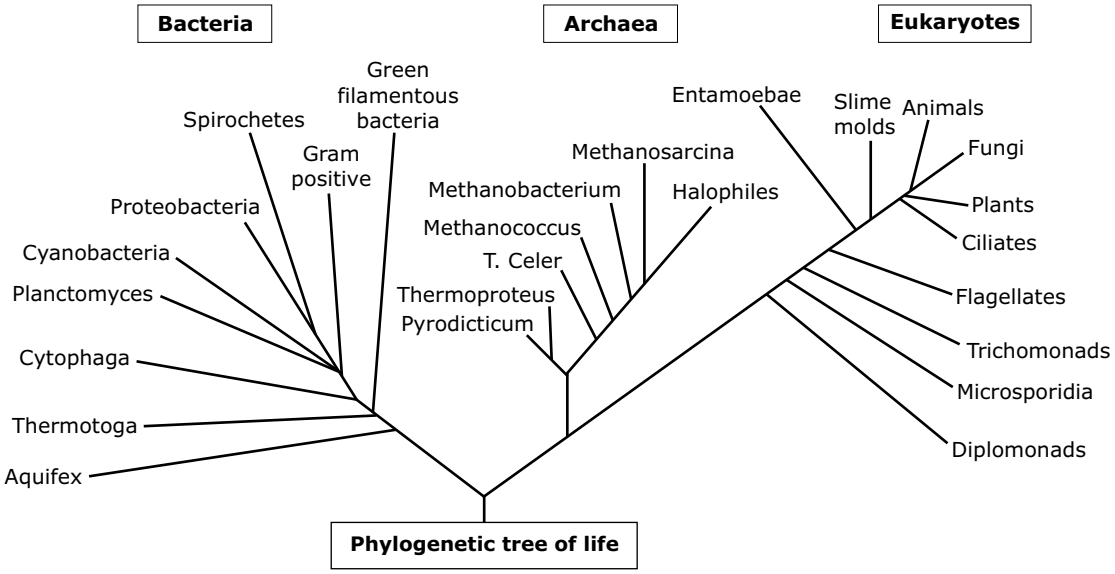


Figure 4.1 A phylogenetic tree of living things, based on RNA data (proposed by Carl Woese), showing the separation of bacteria, archaea, and eukaryotes from a common ancestor.

Based on ribosomal RNA signature sequences, Carl Woese proposed a radical reorganization of the five kingdoms into three domains. In his classification system, Woese placed all four eukaryotic kingdoms (protista, fungi, plantae, animalia) into a single domain called *Eukarya*, also known as the eukaryotes. He then split the former kingdom of Monera into the *Eubacteria* and the *Archaea* domains. Unlike Whittaker's five kingdom system, Woese's three domain system organizes biodiversity by evolutionary relationships.

4.3 Structure of bacterial cell

Bacteria (eubacteria) are microscopic, relatively simple, prokaryotic organisms whose cells lack a nucleus. Prokaryotes can be distinguished from eukaryotes in terms of their cell structure and molecular make-up. Prokaryotic cells are

This page intentionally left blank.

Bacterial staining protocols can be divided into three basic types – simple, differential, and specialized. *Simple stains* react uniformly with all cell types and only distinguish the organisms from their surroundings. *Differential stains* do not stain all types of cells with the same colour. It discriminates different cell types depending upon the chemical or physical composition of the cells. The differential stains most frequently used for bacteria are the Gram stain and the acid-fast stain. *Specialized stains* detect specific structures of cells such as flagella and endospores. These stains are used to color and isolate specific parts of organisms.

Gram staining

Gram staining (or Gram's method) is a *differential* staining method for differentiating bacterial species into two groups based on the physical properties of their cell walls. The method is named after the inventor, the Danish scientist Hans Christian Gram, who developed the technique in 1884.

The gram staining procedure involves four basic steps:

1. The bacteria are first stained with the basic dye *crystal violet*. Crystal violet (it is referred to as a *primary stain*) imparts purple colour to all cells.
2. The bacteria are then treated with Gram's iodine solution. This allows the stain to be retained better by forming an insoluble crystal violet-iodine complex. Iodine is used as a *mordant*. A mordant is used to increase the affinity of a stain for a biological specimen.
3. Gram's decolorizer, a mixture of ethyl alcohol and acetone, is then added. A decolorizer or *decolouring agent* removes the stain from the specimen. This is the differential step. After this step some bacteria retain the purple colour while some other loose purple colour. Bacteria that retain colour are classified as *gram-positive* and bacteria that lose the colour after decolorization are classified as *gram negative*.
4. Because gram-negative bacteria are colourless after the treatment with decolorizer, they are no longer visible. Thus, the counterstain safranin (also a basic dye) is applied. Since the gram-positive bacteria are already stained purple, they are not affected by the counterstain. Gram-negative bacteria, that are now colourless, become directly stained by the safranin. Thus, gram-positive appear purple, while gram-negative appear red or pink.

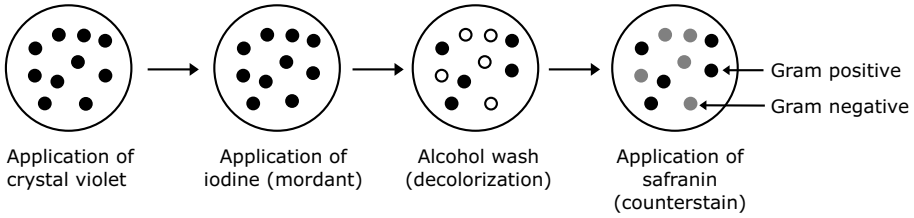


Figure 4.2 The Gram-staining procedure. In the first step of the Gram-staining procedure, the smear is stained with the basic dye crystal violet, the primary stain. It is followed by treatment with an iodine solution functioning as a mordant. The decolorization with ethanol or acetone removes crystal violet from gram-negative cells but not from gram-positive cells. The gram-negative cells then turn pink to red when counterstained with safranin.

Acid-fast staining

The acid-fast stain is a differential stain used to identify acid-fast organisms such as members of the genus *Mycobacterium*. The acid-fast staining procedure involves heating of bacteria with a mixture of basic fuchsin and phenol (also known as *Ziehl-Neelsen stain*). The presence of phenol and heat treatment helps the stain to penetrate the cell wall. Once basic fuchsin has penetrated the cell wall, acid-fast cells are not easily decolorized by an acid-alcohol treatment and hence remain red. It occurs due to the presence of large amounts of **mycolic acid**, a branched chain hydroxy fatty acid. Non-acid-fast bacteria are decolorized by acid-alcohol. Because non-acid-fast bacteria are colourless after the treatment with decolorizer, they are no longer visible. Finally, the counterstain methylene blue is applied. Methylene blue colours non-acid-fast bacteria blue.

Partitioning

Partitioning is an active process which assures that after cell division each daughter cell gets a copy of plasmid. For plasmids present in high copy numbers (50 to 100 copies per cell), random diffusion may be enough to get at least one copy of the plasmid to each daughter cell. However, random segregation of low-copy-number plasmids (only 1 to 2 copies per cell) would most likely mean that, following cell division, one of the daughter cells would not receive a plasmid. The plasmid would eventually be diluted from the population. Consequently, regulated partitioning mechanisms are essential for these plasmids. The mechanism used for partitioning differs depending on the plasmid. Partitioning, especially of low-copy-number plasmids, is regulated by *par* genes present on plasmids. The par systems consist of a *cis-acting* centromere-like site, often called *par*S and two genes termed *par*A and *par*B, which encode *trans-acting* proteins, Par A and B.

Functions encoded by plasmids

Depending on their size, plasmids can encode a few or hundreds of different proteins. However, plasmids rarely encode gene products that are essential for growth, such as RNA polymerase, ribosomal subunits, or enzymes of the tricarboxylic acid cycle. Instead, plasmid genes usually give bacteria a selective advantage under only some conditions. Gene products encoded by plasmids include enzymes for the utilization of unusual carbon sources such as toluene, resistance to substances such as heavy metals and antibiotics, synthesis of antibiotics, and synthesis of toxins and proteins that allow the successful infection of higher organisms. A plasmid that confers no identified functions or phenotypic properties is termed as *cryptic plasmid*.

Table 4.8 List of some plasmid-coded traits

Trait	*Organisms in which trait is found*
Antibiotic resistance	*E. coli, Salmonella sp., Staphylococcus sp.*
Pilus synthesis	*E. coli*, *Pseudomonas sp*.
Tumor formation in plants	*Agrobacterium tumefaciens*
Nitrogen fixation (in plants)	*Rhizobium sp.*
Oil degradation	*Pseudomonas sp.*
Gas vacuole production	*Halobacterium sp.*
Insect toxin synthesis	*Bacillus thuringiensis*
Plant hormone synthesis	*Pseudomonas sp.*
Antibiotic synthesis	*Streptomyces sp.*
Increased virulence	*Yersinia enterocolitica*

Plasmids in eukaryotic organisms

Plasmids are not limited to prokaryotes only. Plasmids are also found in eukaryotic organisms like yeast. One yeast plasmid is called the 2μ circle. The 2μ circle is a 6.3 kb circular, extrachromosomal element found in the nucleus of most *Saccharomyces cerevisiae* strains. It is stably maintained at about 50 to 100 copies per haploid genome of the yeast cell. Like the nuclear chromosomes, the 2μ circle is coated with nucleosomes and replication is initiated by host replication enzymes once per cell cycle.

4.5 Bacterial nutrition

Nutrients are substances used in biosynthesis and energy production and therefore are required for bacterial growth. All bacteria require several macro- and micronutrients. *Macronutrients* (C, O, H, N, S, P, K, Ca, Mg and Fe) are needed in relatively large quantities; *micronutrients* (e.g. Mn, Zn, Co, Mo, Ni and Cu) are used in very small amounts. In addition to the need for carbon, hydrogen and oxygen, all organisms require sources of energy and electrons for growth to take place. On the basis of sources of carbon, energy and hydrogen/electrons bacteria can be categorized into following types:

Pages 394 to 413 are not shown in this preview.

Actinomycetes

Actinomycetes are aerobic, gram-positive, mold-like bacteria that form branched, septate hyphae and asexual spores. The thin walled asexual spores are *conidiospores* or conidia (located at the tip of hyphae) and *sporangiospores* (located in a sporangium). Classification of actinomycetes is primarily based on the properties like conidia arrangement, the presence or absence of the sporangium, cell wall type. Most actinomycetes are nonmotile. When motility is present, it is confined to flagellated spores only.

Streptomyces is the largest genus of actinomycetes. Members of the genus are strict aerobes, mostly non pathogenic saprophytes and bear the chains of nonmotile conidia. The natural habitat of most streptomycetes is the soil. In fact, the odor of moist soil is largely due to the production of volatile substances such as *geosmin* from streptomycetes. Streptomycetes are best known for their synthesis of a vast array of antibiotics like amphotericin B, chloramphenicol, erythromycin, neomycin, nystatin, streptomycin and tetracycline.

Spirochetes

Spirochetes are a group of gram-negative, chemoheterotrophic bacteria. They are slender, long organisms with a flexible, helical shape. Spirochetes lack external rotating flagella. They exhibit creeping or crawling movements. Their unique pattern of motility is due to an unusual morphological structure called the *axial filament*. The central *protoplasmic cylinder*, which contains cytoplasm and the nucleoid, is bounded by a plasma membrane and gram-negative type cell wall. Two or more than a hundred prokaryotic flagella, called *axial filaments* or *periplasmic flagella*, extend from both ends of the cylinder often overlap. The whole complex of periplasmic flagella lies inside a flexible outer membrane. The outer membrane contains lipid, protein, and carbohydrate. *Treponema pallidum* (causes syphilis) and *Borrelia burgdorferi* (responsible for Lyme disease) are examples of spirochetes.

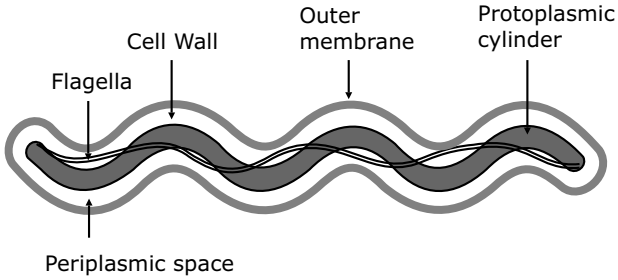


Figure 4.30 The most peculiar feature of spirochetes may be the location of their flagella. Flagella present in the periplasmic space between the plasma membrane and outer membranes. *Spirochete Borrelia burgdorferi* has 7-11 flagella attached near each end of the 'protoplasmic' or cell cylinder, with each flagellum extending through the periplasm towards the center of the spirochete.

Mycoplasmas

Mycoplasmas are the smallest and simplest self-reproducing gram negative bacteria. Mycoplasmas lack cell walls and thus placed in a separate class *Mollicutes* (*mollis*, soft; *cutis,* skin). Formerly, Mycoplasmas were called *pleuropneumonia-like organisms* (PPLO) because it was first isolated from cattle suffering from pleuropneumonia. The trivial term *mollicutes* is frequently used as a general term to describe any member of the class, replacing in this respect the older term mycoplasmas. Mycoplasmas are pleomorphic (vary in shape) and mostly non-motile. General metabolic nature is chemoorganoheterotrophic and require cholesterol for growth. They can be saprophytes or parasites and usually facultative anaerobes. Characteristically, mycoplasmas growing on solid media produce *fried egg* colonies with a central dense region surrounded by a lighter peripheral zone.

Cyanobacteria

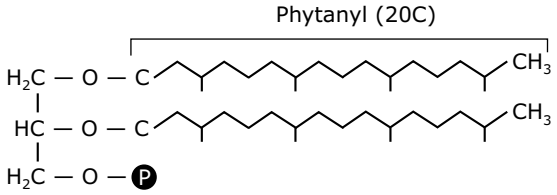
Cyanobacteria are gram negative bacteria. They are oxygenic photosynthetic and obligate photolitho-autotrophs. Photosynthetic pigments present in cyanobacteria are chlorophyll a, carotenoids and phycobilins (phycocyanin and

This page intentionally left blank.

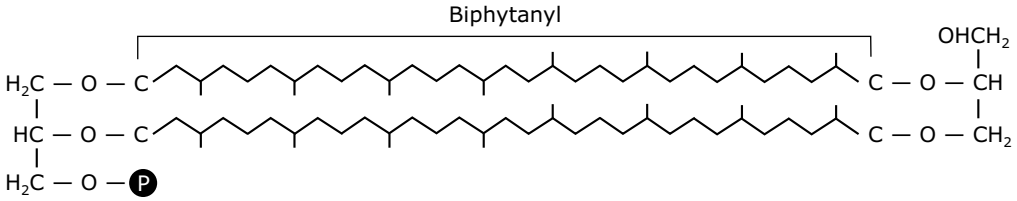
Features of cell wall and plasma membrane lipid

The cell walls of archaebacteria are distinctive from those of eubacteria. Archaebacterial cell walls are composed of different polysaccharides, glycoproteins or proteins, with no peptidoglycan. Many archaebacteria have cell walls made of the polysaccharide *pseudomurein* (a modified peptidoglycan lacking D-amino acids and containing N-acetyltalosaminuronic acid instead of N-acetylmuramic acid; found in methanogenic bacteria). All archaebacteria are resistant to lysozyme and beta-lactam antibiotics such as penicillin.

The nature of archaebacterial plasma membrane lipids differs from both eubacteria and eukaryotes. Archaebacterial membrane lipids contain branched chain hydrocarbons attached to glycerol by *ether links.* Sometimes two glycerols are linked to form a long tetraether. Usually the diether chains are 20 carbons in size and the tetraether chains are 40 carbons. Eubacterial and eukaryotic lipids have glycerol connected to fatty acids by ester bonds.



Glycerol diether



Diglycerol tetraether

Major groups of archaebacteria

There are three major known groups of archaebacteria: *methanogens, extreme halophiles,* and *thermophiles*.

Methanogens

Methanogens are methane producing obligate anaerobes. They comprise the largest group of archaebacteria.

Extreme halophiles

Extreme halophiles are aerobic chemoorganoheterotrophs. They thrive in very high salt concentrations. The best-studied member of the family is *Halobacterium salinarium. H. salinarium* can carry out photosynthesis without chlorophyll or bacteriochlorophyll by using *bacteriorhodopsin*.

Thermophiles

Thermophiles are heat-loving archaebacteria found near hydrothermal vents and hot springs. The optimum growth temperature is between 70–110°C. They are gram-negative and usually strict anaerobes.

4.10 Bacterial toxins

A **toxin** (Latin *toxicum,* poison) is a specific substance, often a metabolic product of the organism that damages the host. Toxins can even induce disease in the absence of the organism that produced them. Diseases that result from the entrance of a specific toxin into the body of a host is called *intoxication*. The term *toxemia* refers to the condition caused by toxins that have entered the blood of the host. Toxins produced by organisms can be divided into two main categories: **exotoxins** and **endotoxins**.

Pages 417 to 421 are not shown in this preview.

4.12 Virus

Viruses are simple, noncellular entities consisting of one or more molecules of either DNA or RNA enclosed in a coat of protein. They can reproduce only within living cells and are *obligate intracellular parasites*. Viruses are smaller than prokaryotic cells ranging in size from 0.02 to 0.3 μm (smallpox virus is largest virus about 200 nm in diameter and polio virus is the smallest virus about 28 nm in diameter). A fully assembled infectious virus is called a **virion**. The main function of the virion is to deliver its DNA or RNA genome into the host cell so that the genome can be expressed (transcribed and translated) by the host cell. Each viral species has a very limited *host range;* i.e. it can reproduce in only a small group of closely related species.

Viral structure

The structure of virions are very diverse, varying widely in size, shape and chemical composition. All viruses have a nucleocapsid composed of nucleic acid surrounded by a protein capsid.

A protein coat, the **capsid**, which functions as a shell to protect the viral genome from nucleases and which during infection attaches the virion to specific receptors exposed on the prospective host cell. Capsids are formed as single or double protein shells and consist of only one or a few structural protein species. The proteins used to build the capsid are called *capsomeres*. The nucleic acid together with the genome forms the nucleocapsid. Some viruses have a membranous envelope that lies outside the nucleocapsid. Those virions having an envelope are called **enveloped viruses**; whereas those lacking an envelope are called *naked viruses*. In enveloped viruses, the nucleocapsid is surrounded by a lipid bilayer and glycoprotein derived from the modified host cell membrane. Enveloped viruses often exhibit a fringe of glycoprotein **spikes**, also called *peplomers*. In viruses that acquire their envelope by budding through the plasma membrane or another intracellular cell membrane, the lipid composition of the viral envelope closely reflects that of the particular host membrane.

Viral genomes are smaller in size. The largest known viral genome, that of bacteriophage G, is 670 kbs. The genome of a virus may consist of DNA or RNA, which may be single stranded (ss) or double stranded (ds), linear or circular. The genomic RNA strand of single-stranded RNA viruses is called **sense** (positive sense, *plus* sense) in orientation if it can serve as mRNA, and **antisense** (negative sense, *minus* sense) if a complementary strand synthesized by a viral RNA transcriptase serves as mRNA.

RNA genomes of certain viruses may be *segmented* in nature. The segmented genomes are those which are divided into two or more physically separate molecules of nucleic acid, all of which are then packaged into a single viral particle. The segmented genome is different from *multipartite* genome. Multipartite genomes are also segmented, but each genome segment is packaged into a separate virus particle. These discrete particles are structurally similar and may contain the same component proteins, but often differ in size depending on the length of the genome segment packaged. Multipartite viruses are only found in plants.

Table 4.18 Types of viral nucleic acids

Nucleic acid type	*Nucleic acid structure*
DNA	
Single stranded	Linear, single-stranded DNA
	Circular, single-stranded DNA
Double stranded	Linear, double-stranded DNA
	Linear double-strand DNA with single chain breaks
	Circular, double-strand DNA
RNA	
Single stranded	Linear, single-stranded, positive-strand RNA
	Linear, single-stranded, negative-strand RNA
	Linear, single-stranded, segmented RNA
Double stranded	Linear, double-stranded, segmented RNA

Pages 423 to 442 are not shown in this preview.

Protease inhibitors work by blocking the activity of the *HIV protease* and thus interfere with virion assembly. Examples include indinavir (Crixivan), ritonavir (Norvir), nelfinavir (Viracept), and saquinavir (Invirase).

However, the most successful treatment approach is to use drug combinations. An effective combination is a cocktail of AZT, lamivudine, and a protease inhibitor such as ritonavir.

Hepatitis virus

Hepatitis is a liver inflammation commonly caused by an infectious agent. Hepatitis sometimes results in destruction of functional liver anatomy and cells, a condition known as *cirrhosis.* Some forms of hepatitis may lead to liver cancer. Although many viruses and a few bacteria can cause hepatitis, a restricted group of viruses is often associated with liver disease termed hepatitis viruses. Hepatitis viruses are diverse, and none of these viruses are genetically related, but all infect cells in the liver, causing hepatitis.

Characteristics of hepatitis viruses

	Features	*Incubation period*
Hepatitis A	ssRNA; No envelope	2–6 week
Hepatitis B	dsDNA; enveloped	4–26 week
Hepatitis C	ssRNA; enveloped	2–22 week
Hepatitis D	ssRNA; enveloped	6–26 week
Hepatitis E	ssRNA; No envelope	2–6 week

The genome of hepatitis B virus (hepadnavirus) is among the smallest known of any viruses, 3-4 kb. Like retroviruses, hepatitis B virus uses reverse transcriptase during replication cycle. However, unlike retroviruses the DNA genome of hepatitis B virus is replicated through an RNA intermediate, the opposite of what occurs in retroviruses. Hepatitis D virus, classified as a *hepatitis delta virus*, is considered to be a subviral satellite because it can propagate only in the presence of the hepatitis B virus. Transmission of hepatitis D virus can occur either via simultaneous infection with hepatitis B virus (coinfection) or via infection of an individual previously infected with hepatitis B virus (*superinfection*). The hepatitis D virus genome consists of a single stranded, negative sense, circular RNA.

4.12.6 Plant viruses

Plant viruses exist in rod and polyhedral shape. Most plant viruses have genomes consisting of a single RNA strand of plus (+) sense type. The best-known plant virus is the rod-shaped tobacco mosaic virus (TMV). Relatively few plant viruses have DNA genomes. There are only two classes of DNA containing plant viruses. The cauliflower mosaic virus belongs to the first class, which contains a double-stranded DNA genome in a polyhedral capsule. The second class of DNA containing plant viruses are the geminiviruses (gemini = twins), characterized by a connected pair of capsids, each containing a circular, single-stranded DNA molecule of about 2500 nucleotides.

Tobacco Mosaic Virus (TMV) causes leaf mottling and discoloration in tobacco and many other plants. It was the first virus to be discovered (by Dmitri Iwanowasky) and first virus to be crystallized (by W. Stanley). TMV is a rod shaped virus with ~2130 capsomeres arranged in a hollow right handed helix. It contains a single genetic RNA (ss, plus sense) of ~6400 nucleotides.

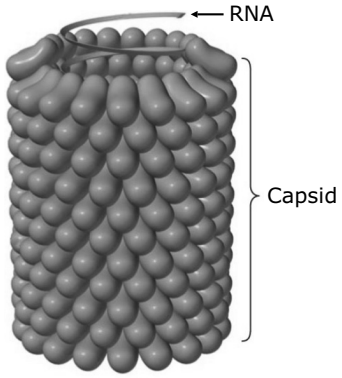


Figure 4.51
Tobacco mosaic virus has a rod-like appearance. Its capsid is made of ~2130 capsomeres. One molecule of genomic ssRNA, 6400 nucleotides long, present in the centre of the capsid. The capsomere self-assembles into the rod like helical structure (16.3 capsomeres per helical turn) around the RNA.

4.13 Prions and Viroid

Prions are proteinaceous infectious agents that are responsible for neurodegenerative diseases in animals including human. Prions are devoid of nucleic acid. The word prion, coined in 1982 by Stanley B. Prusiner, is derived from the words protein and infection. Prion proteins are designated as PrP. The endogenous, normal form is denoted PrP^{C} (for Cellular) while the disease-causing, misfolded form is denoted PrP^{Sc} (for Scrapie, after one of the diseases first linked to prions).

The normal cellular form, PrP^{C}, is converted into PrP^{Sc} through a process whereby a portion of its α-helical and coil structure is refolded into a β-sheet. This structural transition is accompanied by profound changes in the physicochemical properties of the PrP. PrP^{C} is sensitive to proteases whereas PrP^{SC} is protease resistant. High content of β-sheet in PrP^{Sc} results in the formation of amyloid fibrillous structure that is absent from the PrP^{C} form. Both proteins are glycosylated and linked to the membrane by a GPI-linkage. The PrP^{Sc} form can perpetuate itself by causing the newly synthesized PrP protein to take up the PrP^{Sc} form instead of the PrP^{C} form.

Prions are novel transmissible pathogens causing a group of neuro-degenerative diseases that can be perpetuated by inoculating animal with tissue extracts from infected one. Collectively, prion diseases are described as **spongiform encephalopathies**. No prion diseases of plants are known. In 1997, American scientist Stanley B. Prusiner won the Nobel Prize for this pioneering work with these diseases and with the prion proteins. **Kuru** was the first naturally occurring spongiform encephalopathy of humans shown to be caused by prions. It was first described by *Gajdusek* and *Zigas* in 1957. Kuru is characterized by cerebellar ataxia and a shivering-like tremor that produces complete motor incoordination.

Table 4.23 Prion disease of human/animals

Disease	*Organism*
Creutzfeldt-Jakob	Human
Kuru	Human
Bovine spongiform encephalopathy (Also known as *Mad cow disease*)	Cow
Scrapie	Sheep
Chronic wasting disease	Mule deer

Viroid and virusoid

Viroid is an infectious agent of plants that is a single-stranded, covalently closed circular RNA (about 250 to 400 nucleotides long) not associated with any protein. Viroid RNA does not code for any proteins. Viroids (discovered and named by Otto Diener) have so far been shown to infect plants only. A few well-studied viroids include *coconut cadang-cadang viroid* and Potato Spindle-Tuber Viroid (PSTV). No viroid diseases of animals are known, and the precise mechanisms by which viroids cause plant diseases remain unclear. Although the viroid encodes no protein enzymes, the viroid RNA itself acts as a ribozyme.

Table 4.24 Comparison of viruses and viroids

Features	*Virus*	*Viroid*
Nucleic acid	DNA or RNA (ss or ds)	RNA (ss)
Protein	Present	Absent
Capsid	Present	Absent
Host	Bacteria, animal and plants	Plants

Virusoid are satellite nucleic acids. Satellite nucleic acids may be single stranded RNA, single-stranded DNA, or double-stranded RNA. Most of the characterized satellites are associated with plant viruses, and most are single-stranded RNA. Satellite nucleic acids are always functionally dependent on specific helper viruses and are encapsidated

Pages 445 to 447 are not shown in this preview.

Chapter 05

Immunology

Immunology is the science that is concerned with immune response to foreign challenges. **Immunity** (derived from Latin term *immunis*, meaning *exempt*), is the ability of an organism to resist infections by pathogens or state of protection against foreign organisms or substances. The array of cells, tissues and organs which carry out this activity constitute the **immune system**. Immunity is typically divided into two categories—*innate* and *adaptive* immunity.

5.1 Innate immunity

Innate (native/natural) immunity is present since birth and consists of many factors that are relatively nonspecific—that is, it operates against almost any foreign molecules and pathogens. It provides the *first line of defense* against pathogens. It is not specific to any one pathogen but rather acts against all foreign molecules and pathogens. It also does not rely on previous exposure to a pathogen and response is functional since birth and has no memory.

Elements of innate immunity

Physical barriers

Physical barriers are the *first line of defense* against microorganisms. It includes skin and mucous membrane. Most organisms and foreign substances cannot penetrate intact skin but can enter the body if the skin is damaged. Secondly, the acidic pH of sweat and sebaceous secretions and the presence of various fatty acids and hydrolytic enzymes like lysozyme inhibit the growth of most microorganisms. Similarly, respiratory and gastrointestinal tracts are lined by mucous membranes. Mucous membranes entrap foreign microorganisms. The respiratory tract is also covered by cilia, which are hair like projections of the epithelial-cell membranes. The synchronous movement of the cilia propels mucus-entrapped microorganisms out of these tracts. Similarly, the conjunctiva is a specialized, mucus-secreting epithelial membrane that lines the interior surface of each eyelid. It is kept moist by the continuous flushing action of tears (lacrimal fluid) from the lacrimal glands. Tears contain *lysozyme, lactoferrin, IgA* and thus provide chemical as well as physical protection.

Microorganisms do occasionally breach the epithelial barricades. It is then up to the innate and adaptive immune systems to recognize and destroy them, without harming the host. In case of innate immune response several antimicrobial chemicals and phagocytic cells provide protection against pathogens.

Chemical mediator

A variety of chemicals mediate protection against microbes during the period before adaptive immunity develops. The molecules of the innate immune system include complement proteins, cytokines, pattern recognition molecules, acute-phase proteins, cationic peptides, enzyme like lysozyme and many others.

Complement proteins

The complement proteins are soluble proteins/glycoproteins that are mainly synthesized by liver and circulate in the blood and extracellular fluid. They were originally identified by their ability to amplify and complement the

This page intentionally left blank.

neutrophils, macrophages, monocytes and dendritic cells. In vertebrates, macrophages reside in tissues throughout the body. Macrophages are long lived cells, which patrol the tissues of the body. The second major type of phagocytic cells in vertebrates, the neutrophils, are short lived cells which are abundant in blood but are not present in normal healthy tissues. *Phagocytosis* is the ingestion of invading foreign particles, such as bacteria by individual cell. Phagocytosis may be enhanced by a variety of factors collectively referred to as **opsonins** (Greek word meaning 'prepared food for'), which consist of antibodies and various serum components of complement. The process by which particulate antigens are rendered more susceptible to phagocytosis is called **opsonization**. After ingestion, the foreign particle is entrapped in a phagocytic vacuole (phagosome), which fuses with lysosomes forming the *phagolysosome*. The antimicrobial and cytotoxic substances present within the lysosome destroy the phagocytosed microorganisms in the following ways:

Oxygen dependent killing mechanisms

During phagocytosis, a metabolic process known as the *respiratory burst* occurs in activated phagocytes. It results in a transient increase in oxygen consumption by cell. Activated phagocytes generate a number of toxic products such as reactive oxygen intermediates (such as hydroxyl radicals, hypochlorite anion, superoxide anions, hydrogen peroxide) and reactive nitrogen intermediates (like NO, NO_2, $HNO_2^{\bullet}$) which have potent antimicrobial activity.

Oxygen independent killing mechanisms

Activated macrophages also synthesize *lysozyme*, *defensins* (cysteine rich cationic peptides containing 29–35 amino acid residues) and various hydrolytic enzymes/cytotoxic peptides whose degradative activities do not require oxygen.

Inflammatory barriers

Inflammation is an important nonspecific defense reaction to cell injury. The hallmark signs of inflammation are pain, redness (erythema), swelling (edema) and heat. Each of these is the result of specific changes in the local blood vessels. Erythema is caused by *increased vascular diameter*, which leads to *increased blood flow*, thereby causing heat and redness in the area. The blood vessels become permeable to fluid and proteins, leading to local swelling and an accumulation of blood proteins that aid in defense. At the same time, the endothelial cells lining the local blood vessels are stimulated to express cell adhesion proteins that facilitate the attachment and *extravasion* (movement of blood cells through the vessel wall into the surrounding tissue) of white blood cells, including neutrophils, lymphocytes, and monocytes.

The inflammatory response is mediated by a variety of signaling molecules. Activated macrophages produce chemoattractants (known as **chemokines**). Some of these attract neutrophils, which are the *first* cells recruited in large numbers to the site of the new infection. Others later attract monocytes and dendritic cells. The dendritic cells pick up antigens from the invading pathogens and carry them to nearby lymph nodes, where they present the antigens to lymphocytes to marshal the forces of the adaptive immune system. Two principal mediators of the inflammatory response are **histamine** (released by a variety of cells in response to tissue injury) and **kinins** (present in blood plasma in an inactive form). Both cause vasodilation and increased permeability of capillaries. Kinins are also very potent nerve stimulators and are the molecules most responsible for pain associated with inflammation.

5.2 Adaptive immunity

Adaptive immunity, also known as *specific* or *acquired* immunity, is capable of recognizing and selectively eliminating specific foreign antigens. It does not come into play until there is an antigenic challenge to the organism. Adaptive immunity displays four characteristic features:

1. *Antigenic specificity* : It is the ability to discriminate among different epitopes/antigens.
2. *Immunologic memory* : It is the ability to recall previous contact with a foreign molecule and respond to it in a learned manner-that is, with a more rapid and larger response.

This page intentionally left blank.

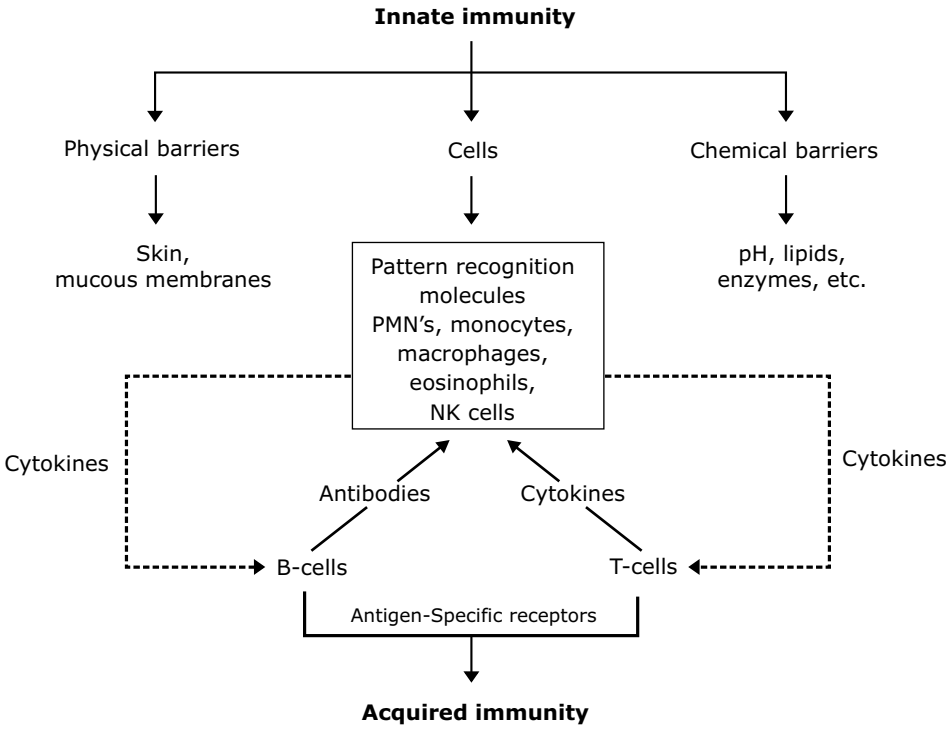
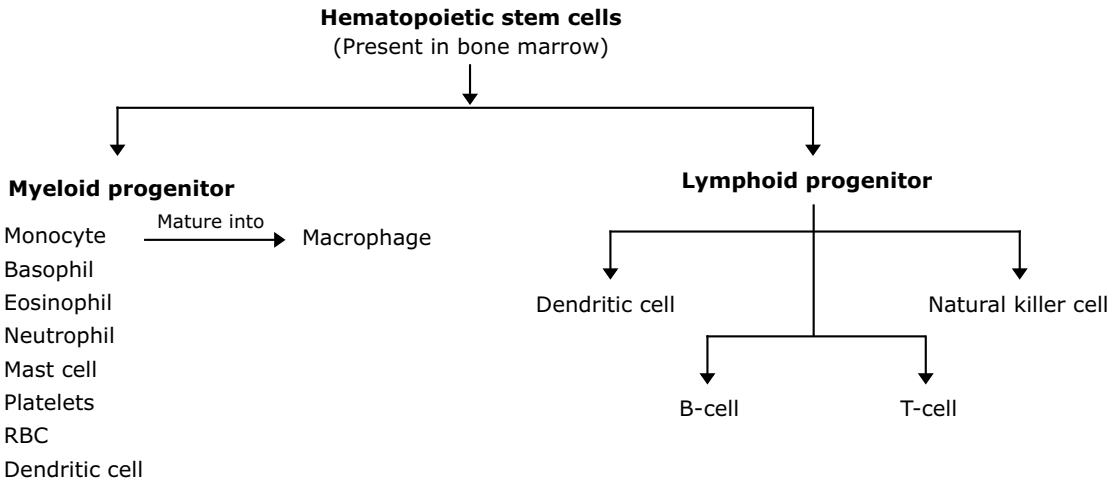


Figure 5.2 The interrelationship between innate and acquired immunity.
Adapted and modified from Immunology A short Course, R. Coico et. al, Wiley-Liss Publication.

The encounter between macrophages and microbes can generate 'danger' signals that stimulate and direct adaptive responses. It may increase the ability of macrophages to display antigen for recognition by antigen specific T-cells. Macrophage stimulated by encounters with microbes also secrete immunoregulatory molecules, called **cytokines**. These cytokines guide adaptive immune response. Vice-versa, adaptive immune system also produces signals and components which stimulate and increase the efficacy of innate response. The cells of adaptive system (e.g. T-cells) secrete cytokines and increase the ability of macrophage to kill the ingested microbes. By binding to the pathogens, antibodies mark it as a target for attack by complement.

5.3 Cells of the immune system

The immune system is a defensive system in a host consisting of widely distributed cells, tissues and organs that recognize foreign substances and microorganisms and acts to neutralize or destroy them. The cells responsible for both nonspecific and specific immunity are the leukocytes or white blood cells. All leukocytes arise from a type of cell called the *hematopoietic stem cell.* A hematopoietic stem cell is *multipotent cell*. During hematopoiesis, hematopoietic stem cell differentiates along one of two pathways, giving rise to either a common **lymphoid progenitor cell** or a common **myeloid progenitor cell**. Common lymphoid progenitor cells give rise to B-cells, T-cells and natural killer cells and some dendritic cells. The common myeloid progenitor cells give rise to red blood cells (erythrocytes), white blood cells (neutrophils, eosinophils, basophils, monocytes, mast cells, dendritic cells) and platelets.



5.3.1 Lymphoid progenitor

Lymphocytes (responsible for adaptive immune response) are mononuclear leukocytes which constitute 20 to 40% of total white blood cells (or leukocytes). They occur in large numbers in the blood and lymph and in lymphoid organs such as the thymus, lymph nodes, spleen and appendix. Up to 99% of lymphocytic cells are found in lymph. Lymphocytes are of three main types:

1. B-lymphocytes or B-cells
2. T-lymphocytes or T-cells
3. Natural killer (NK) cells

B-lymphocytes

The B-lymphocyte matures in the bone marrow in many mammalian species, including humans (in birds it is *Bursa of Fabricius*) and expresses membrane-bound antibody. After interacting with antigen, it differentiates into antibody-secreting *plasma cells* and *memory cells*. They are the only cell type capable of producing antibody molecules and therefore the central cellular component of humoral immune responses. B-cells also serve as Antigen Presenting Cells (APCs).

Properties of B-cells

Origin	—	Bone marrow
Maturation	—	Bone marrow (Bursa of Fabricius in bird)
Expression of Ag receptor	—	Bone marrow
Differentiation	—	In lymphoid tissue
Surface immunoglobulin	—	Present
Immunity	—	Humoral
Distribution	—	Spleen, Lymph nodes, Bone marrow and other lymphoid tissue
Secretory product	—	Antibodies and cytokines
Complement receptors	—	Present

T-lymphocytes

T-lymphocytes arise in the bone marrow. Unlike B-cells, which mature within the bone marrow, T-cells migrate to the thymus gland to mature. During its maturation within the thymus, the T-cell comes to express a unique antigen-binding molecule, called the *T-cell receptor*, on the membrane. T-cells do not make antibodies but perform various effector functions when APC bring antigens into the secondary lymphoid organ. T-cells help in eliminating APCs, cancer cells, virus-infected cells or grafts which have altered self-cells.

Pages 454 to 456 are not shown in this preview.

Thymus

Thymus is the site where T-cells mature. Progenitor cells from the bone marrow migrate into the thymus gland, where they differentiate into T-cells. It is a flat, bilobed organ situated above the heart. Each lobe is surrounded by a capsule and is divided into lobules, which are separated from each other by strands of connective tissue called trabeculae. Each lobule is organized into two compartments: the outer compartment, or cortex, and the inner compartment, or *medulla*. T-lymphocytes mature in the cortex and migrate to the medulla, where they encounter macrophages and dendritic cells. Here, they undergo thymic selection, which results in the development of mature, functional T-cells, which then leave to enter the peripheral blood circulation, through which they are transported to the secondary lymphoid organs. It is in these secondary lymphoid organs where the T-cells encounter and respond to foreign antigens.

5.4.2 Secondary lymphoid organs/tissues

Mature B and T-lymphocytes migrate from bone marrow and thymus, respectively, through the bloodstream to the *secondary (peripheral) lymphoid organs.* These secondary (peripheral) lymphoid organs are those organs in which antigen-driven proliferation and differentiation take place.

The major secondary lymphoid organs are the **spleen**, the **lymph nodes** and **mucosa associated lymphoid tissue** (MALT). Spleen and lymph nodes are the highly organized secondary lymphoid organs. The secondary lymphoid organs have two major functions: They are highly efficient in trapping and concentrating foreign substances, and they are the main sites of production of antibodies and the induction of antigen-specific T- lymphocytes.

Spleen

The spleen is the *largest* of the secondary lymphoid organs. It is highly efficient in trapping and concentrating foreign substances carried in the blood. It is the major organ in the body in which antibodies are synthesized and from which they are released into the circulation.

The interior of the spleen is a compartmentalized structure. The compartments are of two types – *Red pulp* and *white pulp. Red pulp is the site where old and defective RBCs are destroyed and removed*, whereas white pulp forms **PALS** (Periarteriolar lymphoid sheath) which are rich in T-cells. The marginal zone, located peripheral to the PALS, is rich in lymphocyte and macrophage. Approximately 50% of spleen cells are B-lymphocytes; 30-40% are T-lymphocytes.

Lymph nodes

Lymph nodes are small encapsulated bean shaped structures (normally <1cm in diameter) found in various regions throughout the body. The lymph nodes are composed of a *medulla* and a *cortex*, which is surrounded by a capsule of connective tissue. They are packed with lymphocytes, macrophages, and dendritic cells. The cortical region contains primary lymphoid follicles. After antigenic stimulation, these structures enlarge to form secondary lymphoid follicles with germinal centers containing dense populations of lymphocytes (mostly B-cells). The deep cortical area or paracortical region contains T-cells and dendritic cells. Antigens are brought into these areas by dendritic cells, which present antigen fragments to T-cells. The medullary area of the lymph node contains antibody-secreting plasma cells that have traveled from the cortex to the medulla via lymphatic vessels.

Lymph nodes are highly efficient in trapping antigen that enters through the afferent lymphatic vessels. In the node, the antigen interacts with macrophages, T-cells, and B-cells, and that interaction brings about in immune response, manifested by the generation of antibodies and antigen-specific T-cells.

Mucosa associated lymphoid tissue

The majority of secondary lymphoid tissue in the human body is located within the lining of respiratory, digestive and genitourinary tracts. These are collectively called as Mucosa Associated Lymphoid Tissue (MALT). There are several types of MALT. Two major MALT includes Bronchial Associated Lymphoid Tissue (BALT) and Gut-Associated Lymphoid Tissue (GALT). GALT includes the tonsils, adenoids, and specialized regions in the small intestine called **Peyer's patches**.

5.5 Antigens

Adaptive immune responses arise as a result of exposure to foreign compounds. The compound that evokes the response is referred to as **antigen**, a term initially coined due to the ability of these compounds to cause *anti*body responses to be *gene*rated. An antigen is any agent capable of binding specifically to T-cell receptor (TCR) or an antibody molecule (membrane bound or soluble). The ability of a compound to bind with an antibody or a TCR is referred to as **antigenicity**. There is a functional distinction between the term antigen and immunogen. An **immunogen** is any agent capable of inducing an immune response and is therefore **immunogenic**. The distinction between the terms is necessary because there are many compounds that are incapable of inducing an immune response, yet they are capable of binding with components of the immune system that have been induced specifically against them. Thus all immunogens are antigens, but not all antigens are immunogens.

Requirements for immunogenicity

A substance must possess the following characteristics to be immunogenic:

1. *Foreignness*

   The most important feature of an immunogen is that an effective immunogen must be *foreign* with respect to the host. The adaptive immune system recognizes and eliminates only foreign (nonself) antigens. Self antigens are not recognized and thus individuals are *tolerant* to their own self molecules, even though these same molecules have the capacity to act as immunogens in other individuals of the same species.

2. *Size*

   The second requirement for being immunogenic is that the compound must have a certain minimal molecular weight. There is a relationship between the size of immunogen and its immunogenicity. In general, small compounds with a molecular weight <1000 Da (e.g. penicillin, aspirin) are not immunogenic; those of molecular weight between 1000 and 6000 Da (e.g. insulin, adrenocorticotropic hormone) may or may not be immunogenic; and those of molecular weight >6000 Da (e.g. albumin, tetanus toxin) are generally immunogenic. The most active immunogens tend to have a molecular mass of 100,000 Da or more. In short relatively small substances have decreased immunogenicity, whereas large substances have increased immunogenicity.

3. *Chemical complexity*

   The third characteristic necessary for a compound to be immunogenic is a certain degree of chemical complexity. For example, homopolymers of amino acids or sugars are seldom good immunogens regardless of their size. Similarly, a homopolymer of poly-γ-D-glutamic acid (the capsular material of *Bacillus anthracis*) with a molecular weight of 50,000 Da is not immunogenic. The absence of immunogenicity is because these compounds, although of high molecular weight, are not sufficiently chemically complex.

   Virtually all proteins are immunogenic. Furthermore, the greater the degree of complexity of the protein, the more vigorous will be the immune response to that protein. Carbohydrates are immunogenic only if they have a complex polysaccharide structure or part of complex molecules such as glycoproteins. Nucleic acids and lipids are poor immunogens by themselves, but they become immunogenic when they are conjugated to protein carriers.

4. *Dosage and route of administration*

   The insufficient dose of immunogen may not stimulate an immune response either because the amount administered fails to activate enough lymphocytes or because such a dose renders the responding cells unresponsive. Besides the need to administer a threshold amount of immunogen to induce an immune response, the number of doses administered also affects the outcome of the immune response generated.

   The route of administration also affects the outcome of the immunization because this determines which organs and cell populations will be involved in the response. Immunogens can be administered through a number of common routes: *Intravenous* (into a vein); *intradermal* (into the skin); *subcutaneous* (beneath the skin);

Pages 459 to 466 are not shown in this preview.

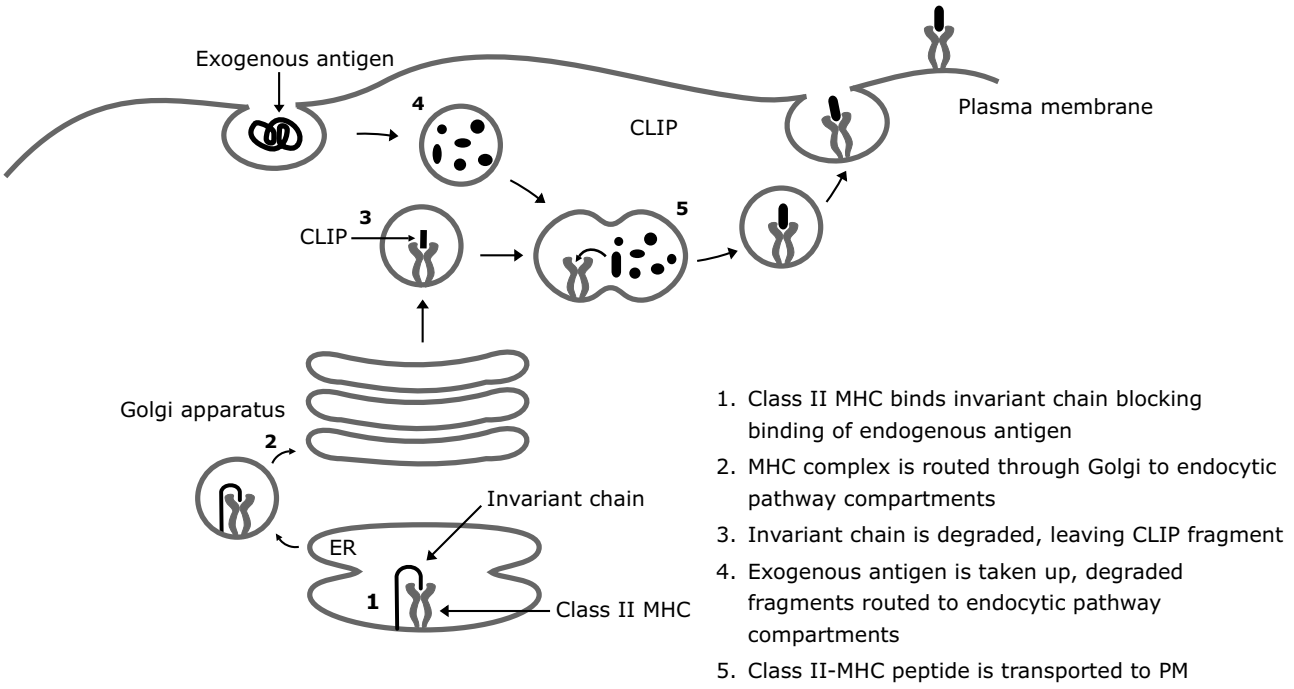


Figure 5.11 The processing of an exogenous protein antigen for presentation to a helper T-cell.

5.6.3 Laboratory mice

Mice are the most commonly used mammalian research model. They are common experimental animals in biology, primarily because they are mammals, are relatively easy to maintain and handle, reproduce quickly, and share a high degree of homology with humans. Laboratory mice include several inbred, outbred, knockout and transgenic mice strains. Many laboratory strains are *inbred*. An inbred strain is one that is produced using at least 20 consecutive generations of sister and brother or parent and offspring matings. The mating of two genetically related parents is called *inbreeding*. Inbreeding results in increased homozygosity. In contrast to inbred mice, outbred mice are usually heterozygous at many loci.

If mice are inbred (that is, have identical alleles at all loci), each H-2 locus will be homozygous because the maternal and paternal haplotypes are identical, and all offspring therefore express identical haplotypes. Inbred mouse strains are *syngeneic* or identical at all genetic loci. Two strains are considered *congenic* if they are genetically identical except at a single genetic locus.

Some inbred mouse strains have been designated as *prototype* strains and the MHC haplotype expressed by these strains is designated by an arbitrary italic superscript (e.g. H-2^a, H-2^b). If another inbred strain has the same set of alleles as the prototype strain, its MHC haplotype is the same as the prototype strain.

Table 5.5 H-2 haplotypes of some mouse strains

			H-2 alleles				
Prototype strain	*Other strains with the same haplotype*	*Haplotype*	*K*	*IA*	*IE*	*S*	*D*
CBA	AKR, C3H, C57BR	*k*	*k*	*k*	*k*	*k*	*k*
DBA/2	BALB/c, SEA, YBR	*d*	*d*	*d*	*d*	*d*	*d*
C57BL/10 (B10)	C57BL/6, C57L	*b*	*b*	*b*	*b*	*b*	*b*
A	A/He, A/Sn	*a*	*k*	*k*	*k*	*d*	*d*

5.7 Immunoglobulins : Structure and function

Antibodies, the antigen-binding *glycoproteins* are synthesized exclusively by B-cells and in billions of forms, each with a different amino acid sequence and a different antigen binding site. Collectively called *immunoglobulins* (Ig), they are among the most abundant protein components in the blood, constituting about 20% of the total protein components in the blood plasma. Antibodies (Ab) are present on the B-cell membrane and also are secreted by plasma cells.

5.7.1 Basic structure of antibody molecule

The simplest antibodies are Y-shaped molecules with two identical antigen-binding sites, one at the tip of each arm of the Y. Because of their two antigen-binding sites, they are described as *bivalent*.

Antibody has a common structure of 4 polypeptide chains. It is a heterodimer and consists of two identical **light (L) chains** (each containing about 220 amino acids residues, about 25000 MW) and two identical **heavy (H) chains** (each usually containing about 440 amino acids residues, about 50000 MW). Each light chain is bound to heavy chain by disulfide bridges and other non-covalent linkages. Thus, antibody is a dimer of H—L chain.

All species studied have the two major classes of light chains: κ and λ. Any one individual of a species produces both types of light chain. However, in any one immunoglobulin molecule, the light chains are always either both κ or both λ, never one of each. While there are two types of light chains, the immunoglobulins of virtually all species have been shown to consist of five different types of heavy chains- α, γ, δ, ε and μ. These five different types of heavy chains are called *isotypes*. The heavy-chains of a given antibody molecule determine the class of that antibody: IgM (μ), IgG (γ), IgA (α), IgD (δ) or IgE (ε). Each class can have either κ or λ light chains. Any individual of a species makes all heavy chains, but in any one antibody molecule, both heavy chains are identical. Thus an antibody molecule of the IgG class could have the structure $\kappa 2\gamma 2$ with two identical κ light chains and two identical γ heavy chains. Alternatively, it could have the structure $\lambda 2\gamma 2$ with two identical λ light chains and two identical γ heavy chains.

Immunoglobulin heavy chain isotypes

Isotype	*Heavy chain*
IgM	μ
IgD	δ
IgG	γ
IgA	α
IgE	ε

Minor differences in the amino-acid sequences of the α and the γ heavy chains led to further classification of the heavy chains into subclasses. In humans, there are two subclasses of α heavy chains (α_1 and α_2) and four subclasses of γ heavy chains (γ_1, γ_2, γ_3 and γ_4).

Both light and heavy chains have a variable sequence at their N-terminal ends but a constant sequence at their C-terminal ends. Light chains have a **constant region** (C_L) about 110 amino acids long and a **variable region** (V_L) of the same size. The variable region (V_H) of the heavy chains (at their N-terminus) is also about 110 amino acids long, but the heavy-chain constant region (C_H) is about three to four times longer (330 or 440 amino acids), depending on the class. It is the N-terminal ends of the light and heavy chains that come together to form the antigen-binding site.

The diversity in the variable regions of both light and heavy chains is for the most part restricted to three small *hypervariable regions* (each ~10 amino acid residues long) in each chain called **complementarity determining regions** (CDR); the remaining parts of the variable region, known as **framework regions**, are relatively constant. Proceeding from either the V_L or V_H amino terminus, these regions are called CDR1, CDR2 and CDR3. The CDR3 is the most variable of the CDRs.

Pages 469 to 479 are not shown in this preview.

Clonal selection theory

The clonal selection theory is a central paradigm of adaptive immunity. The most remarkable feature of the adaptive immune system is that it can respond to millions of different foreign antigens in a highly specific way. B-cells, for example, make antibodies that react specifically with the antigen that induced their production. The clonal selection theory (formulated by Sir Macfarlane Burnet) explains how the adaptive immune system can respond to millions of different antigens in a highly specific way.

According to this theory, an animal first randomly generates a vast diversity of lymphocytes, and then those lymphocytes that can react against the foreign antigens that the animal actually encounters are specifically selected for proliferation. As each lymphocyte develops in a central lymphoid organ, it becomes committed to react with a particular antigen before ever being exposed to the antigen. It expresses this commitment in the form of cell-surface receptor proteins that specifically fit the antigen.

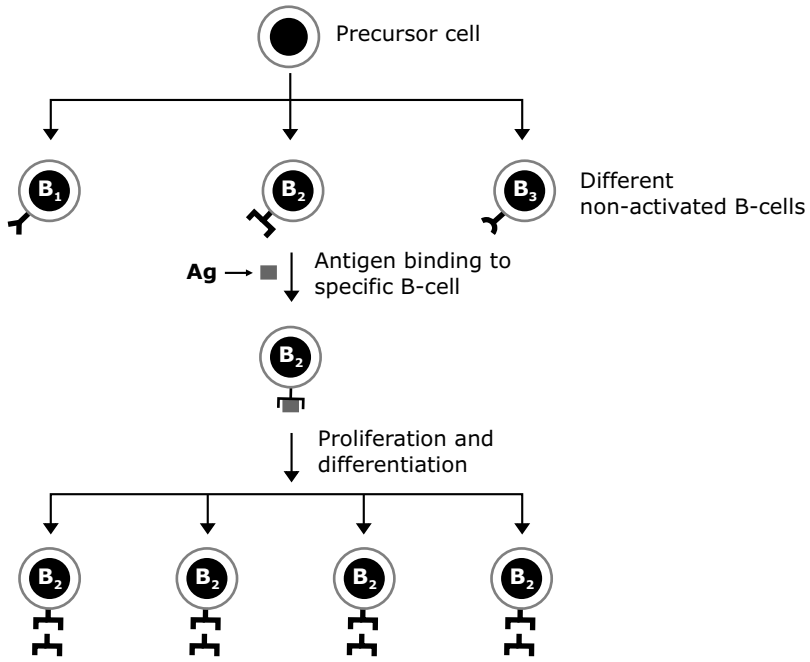


Figure 5.23 The clonal selection theory of B-cells leading to antibody production.
Adapted from Molecular Biology of the cell, Albert et al., Garland Science.

When a lymphocyte encounters its antigen in a peripheral lymphoid organ, the binding of the antigen to the receptors activates the lymphocyte, causing it both to proliferate and to differentiate into an effector cell. An antigen therefore selectively stimulates those cells that express complementary antigen-specific receptors and are thus already committed to respond to it. This arrangement is what makes adaptive immune responses antigen-specific. According to the clonal selection theory, then, the immune system functions on the *ready-made* principle rather than the *made-to-order* one.

The term *clonal* in clonal selection theory derives from the postulate that the adaptive immune system is composed of millions of different families, or clones, of lymphocytes, each consisting of T or B-cells descended from a common ancestor. Each ancestral cell was already committed to make one particular antigen-specific receptor protein, and so all cells in a clone have the same antigen specificity.

5.9 Kinetics of the antibody response

Humoral immunity is mediated by serum antibodies which are the proteins secreted by the B-cells. B-cells are initially activated to secrete antibodies after the binding of antigens to specific membrane immunoglobulin molecules (B-cells receptors), which are expressed by these cells. Once bound, the B-cell receives signals to begin making the secreted form of this immunoglobulin, a process that initiates the full-blown antibody response whose purpose is to eliminate the antigen from the host. Antibodies are a heterogeneous mixture of serum globulins, all of which share the ability to bind individually to specific antigens.

Primary and secondary responses

The first exposure of an individual to an immunogen is referred to as the primary immunization, which generates a **primary response**. The primary antibody response may be divided into several phases, as follows:

1. *Lag or latent phase*: It is the immediate stage following antigenic stimulus during which no antibody is detectable in circulation. The length of this period is generally one to two weeks.
2. *Log or exponential phase*: In this phase there is a steady rise in the titer of antibody and the concentration of antibody in the serum increases exponentially.
3. *Plateau or steady state*: During this phase there is an equilibrium between antibody synthesis and degradation.
4. *Declining phase*: The concentration of antibody in serum declines rapidly.

A second exposure to the same immunogen results in a **secondary response**. This second exposure may occur after the response to the first immune event has leveled off or has totally subsided. The secondary response is also called the *memory* or *anamnestic* response and the B-and T-lymphocytes that participate in the memory response are termed **memory cells**.

The primary response is slow and short lived with a long lag phase and low titer of antibodies that do not persist for long. However the secondary response is prompt, powerful and prolonged, with a short or negligible lag phase and a much higher level of Ab that lasts for long periods.

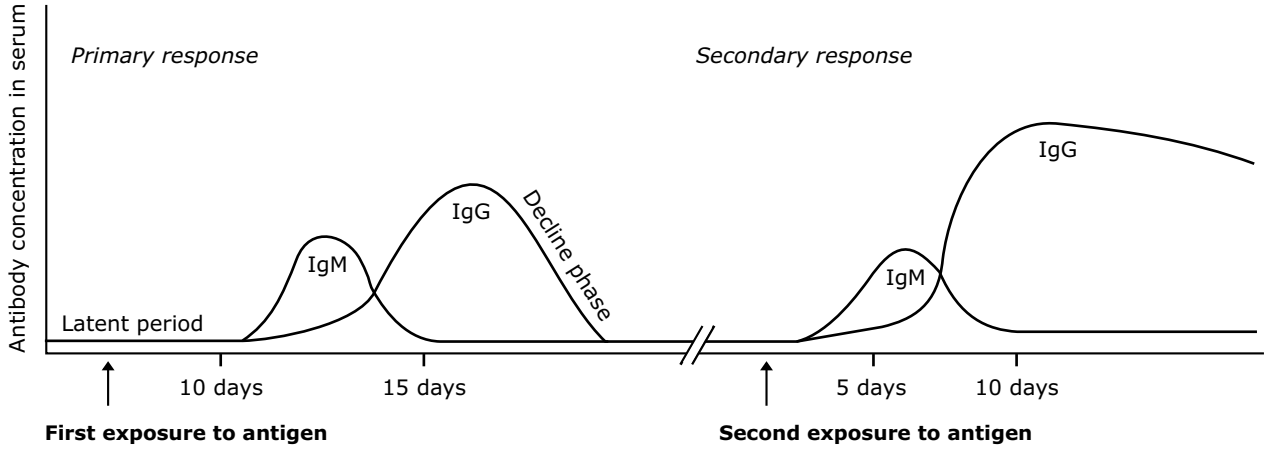


Figure 5.24 Antibody production and kinetics.

In the primary response, the first class of antibody detected is generally IgM, then IgG, or another antibody class. There is a marked change in the type and quality of antibody produced in the secondary response. There is a shift in class response, known as **class switching**, with IgG antibodies appearing at higher concentrations and with greater persistence than IgM, which may be greatly reduced or disappear altogether. This may be also accompanied by the appearance of IgA and IgE. The IgG, IgE, and IgA molecules are collectively referred to as secondary classes of antibodies because they are thought to be produced only after antigen stimulation and because they dominate secondary antibody responses.

This page intentionally left blank.

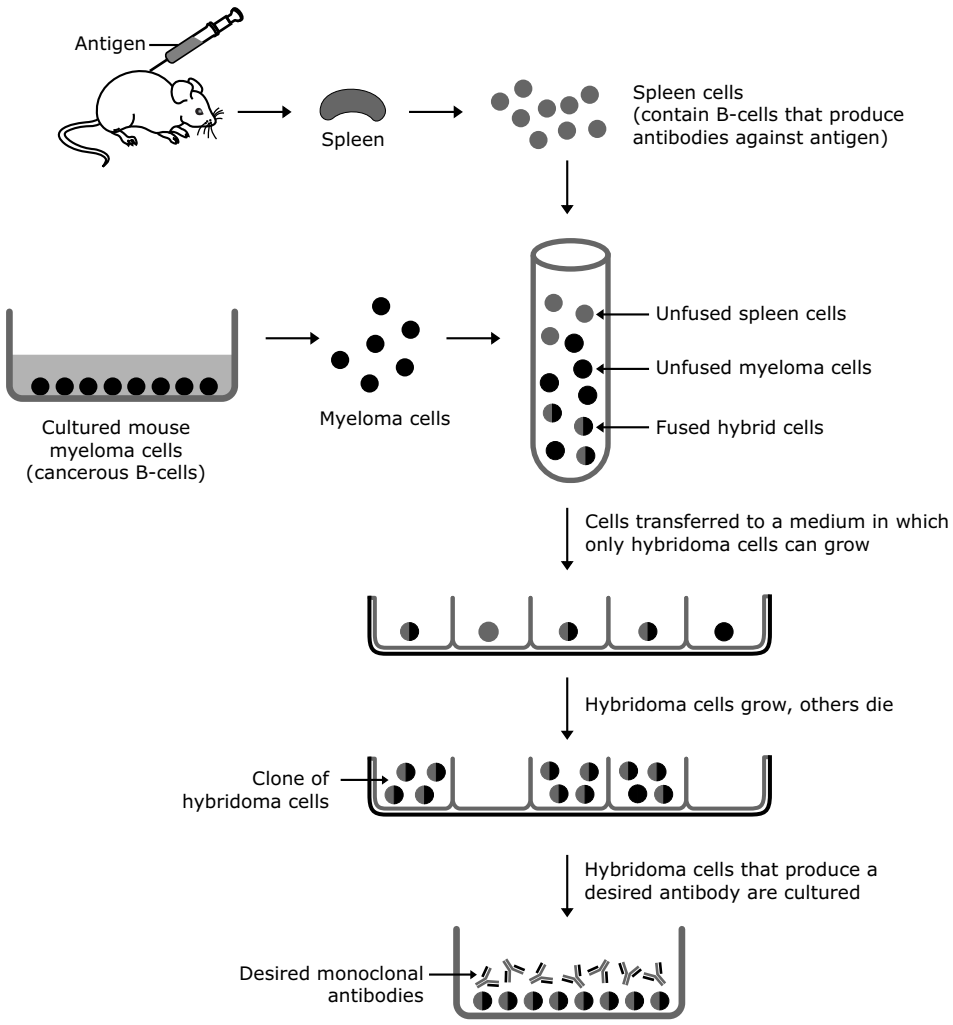


Figure 5.25 Production of monoclonal antibodies.

In the procedure, myeloma cells are engineered to be deficient in enzyme HGPRT. After fusion of lymphocytes with HGPRT-negative myeloma cells, aminopterin-containing medium, supplemented with hypoxanthine and thymidine to ensure and adequate supply of substrates for the salvage pathway (HAT medium) is added, which kills myeloma cells but allows hybridomas to survive as they inherit HGPRT from the lymphocyte parent. Unfused lymphocytes die after a short period of culture, which results in a pure preparation of hybridomas.

5.10.1 Engineered monoclonal antibodies

Immunotoxins

Immunotoxins are protein-based drugs contain two functional domains, one allowing them to bind specific target cells (target-specific binding domain), and one that kills the cells following internalization (cytotoxic domain). An immunotoxin is prepared by replacing the target-specific binding domain of toxin with a monoclonal antibody that is specific for a particular antigen. The toxins used may be *bacterial toxins* such as diphtheria toxins or *plant toxins* like ricin, abrin, etc. Toxins used to prepare immunotoxins include ricin, *Shigella* toxin and diphtheria toxin, all of which inhibit protein synthesis.

Pages 484 to 493 are not shown in this preview.

Immunological tolerance of B-cells are also mediated by the process of *clonal anergy* or *inactivation*. When a mature B-cell escapes tolerance in primary lymphoid organ and bind self antigens in the peripheral lymphoid organ, self-reactive B-cell may either die by apoptosis or be functionally inactivated and cannot amplify the immune response.

5.13 T−cells and CMI

Ig molecules secreted from B-cells, play a critical role in interacting with antigens when they are present *outside* the cells; for example, when viruses are encountered in blood plasma or at mucosal surfaces. Once an antigen gets into a cell however, antibodies do not generally have access to it, and so antibodies are ineffective in dealing with antigens *inside* cells. T-cells deal with pathogens – such as viruses, bacteria, and parasites – that resides inside the cells of the host. T-cells responses differ from B-cell responses in at least two crucial ways.

First, T-cells are activated by foreign antigen to proliferate and differentiate into effector cells only when the antigen is displayed on the surface of antigen-presenting cells/target cells in peripheral lymphoid organs.

The *second* difference is that, once activated, effector T-cells act only at short range, either within a secondary lymphoid organ or after they have migrated into a site of infection. They interact directly with another cell in the body, which they either kill or signal in some way.

T-cell receptor

T-cells, like B-cells, express antigen specific receptors. The T-cell receptor (TCR) is a heterodimer and composed of two transmembrane glycoprotein chains, α and β. The extracellular portion of each chain consists of two domains, resembling immunoglobulin variable (V) and constant (C) domains, respectively. Both chains are glycosylated and connected with each other with the help of interchain disulfide bond. The transmembrane helices of both chains are unusual in containing positively charged (basic) amino acid residues within the hydrophobic transmembrane segment. The α-chains carry two such residues; the β-chains have one.

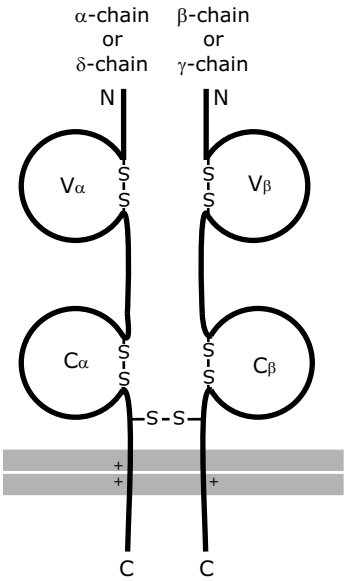


Figure 5.38 The predominant form of the antigen-binding chains of TCR.

Pages 495 to 503 are not shown in this preview.

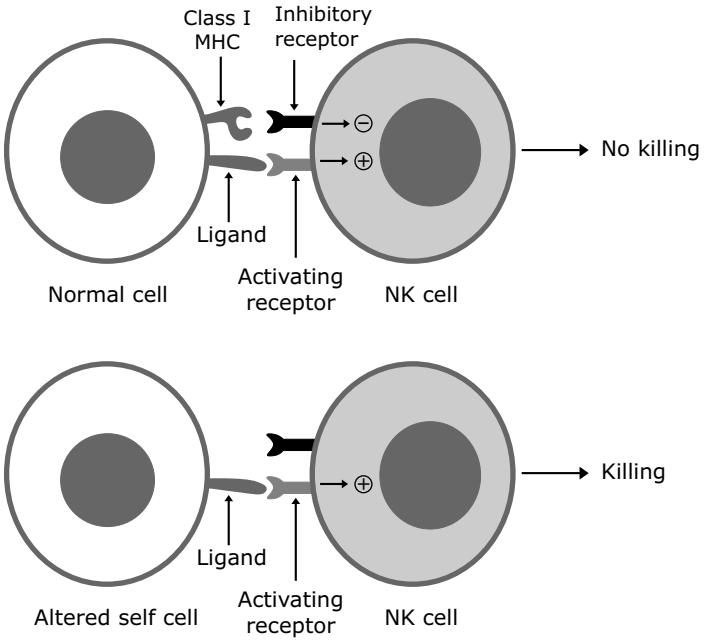


Figure 5.46 An activating receptor on NK cells interacts with its ligand on normal and altered self cells, inducing an activation signal that results in killing. However, interaction of inhibitory NK-cell receptors with class I MHC molecules delivers an inhibition signal that counteracts the activation signal. Expression of class I molecules on normal cells thus prevents their destruction by NK cells. Because class I expression is often decreased on altered self cells (virus infected cells and tumor cells), the killing signal predominates, leading to their destruction.

5.13.1 Superantigens

Superantigens are viral or bacterial proteins that bind simultaneously to the variable domain of β of a T-cell receptor (TCR) and to the α-chain of a class II MHC molecule (i.e. outside the peptide-binding groove). Because of their unique binding ability, superantigens can activate large numbers of T-cells irrespective of their antigenic specificity. Superantigens can be exogenous and endogenous. *Exogenous* superantigens are soluble proteins secreted by bacteria whereas *endogenous* superantigens are cell-membrane proteins encoded by certain viruses that infect mammalian cells.

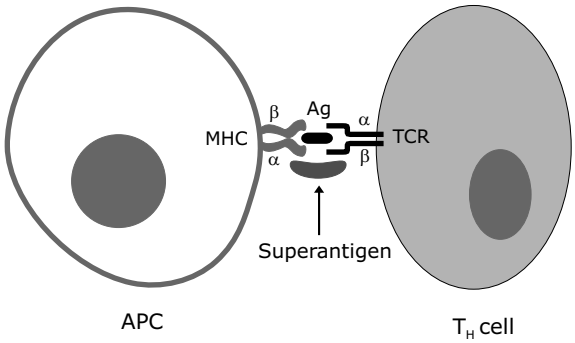


Figure 5.47 Superantigen-mediated cross-linkage of T-cell receptor (TCR) and class II MHC molecules. Superantigen binds to class II MHC molecule and a part of the Vβ chain of the T-cell receptor that is outside the normal antigen-binding site and this binding is sufficient to trigger T-cell activation. A superantigen binds to all TCRs bearing a particular V sequence regardless of their antigen specificity.

5.14 Cytokines

Cytokines are low-molecular-mass (generally less than 30 kDa) soluble proteins/glycoproteins, non-immunoglobulin in nature, secreted by a variety of cell types and act nonenzymatically through specific receptors to regulate host cell function. They do not include the peptide and steroid hormones of the endocrine system. Cytokines play major roles in the development of cellular and humoral immune responses, induction of the inflammatory response, regulation of hematopoiesis, control of cellular proliferation and differentiation.

Cytokines can affect the same cell responsible for their production (an *autocrine* function) or nearby cells (a *paracrine* function), or they can be distributed by the circulatory system to distant target cells (an *endocrine* function). They are highly potent hormone-like substances, active even at femto molar concentration. However, they differ from endocrine hormones as being not produced by glands but by widely distributed cells. Cytokines produce biological actions only when they bind to specific, high-affinity receptors on the surface of target cells. The biological activities of cytokines exhibit *pleiotropy* (a given cytokines that has different biological effect on different target cells), *redundancy* (two or more cytokines that mediates similar functions), *synergy* (combined effect of two cytokines on cellular activity is greater than the additive effect of the individual cytokines) and *antagonism* (effect of one cytokines inhibit the effect of another cytokines).

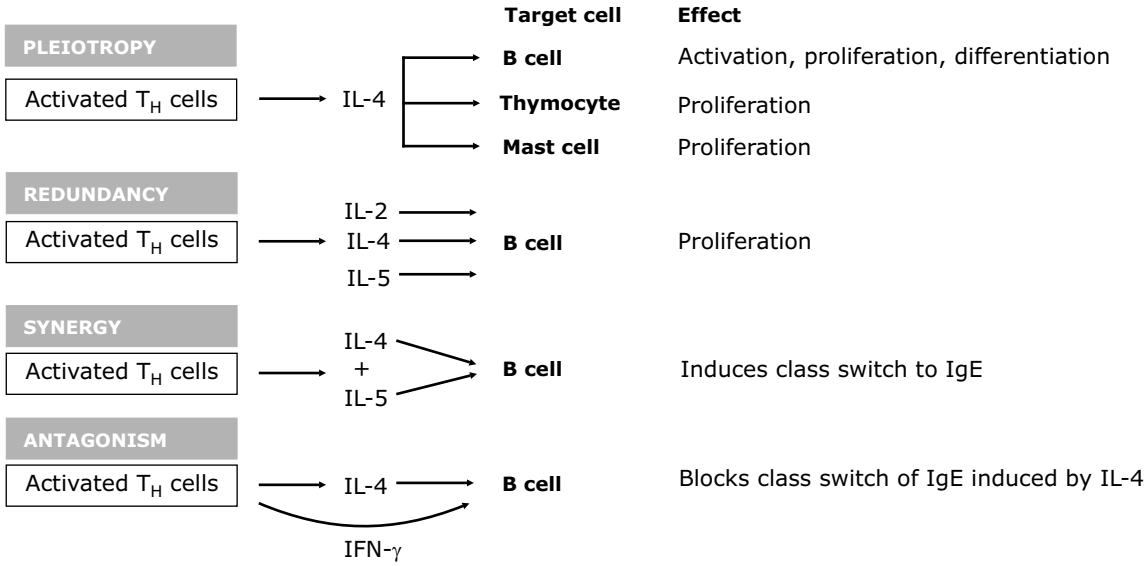


Figure 5.48 Cytokine attributes of pleiotropy, redundancy, synergy (synergism), antagonism.

Cytokines differ from hormones and growth factors. All three are secretory proteins that elicit their biological effects at very low concentrations by binding to receptors on target cells. Growth factors tend to be produced constitutively, whereas cytokines and hormones are secreted in response to discrete stimuli. Unlike hormones, which generally act long range in an endocrine fashion, most cytokines act over a short distance in an autocrine or paracrine fashion. In addition, most hormones are produced by specialized glands and tend to have a unique action on one or a few types of target cell. In contrast, cytokines are often produced by, and bind to, a variety of cells.

There are over 100 different cytokines. The generic name of cytokines includes all proteins with a small molecular weight, released by cells of the immune system, especially by monocytes and T-lymphocytes. But they are also secreted by many cells in addition to those of the immune system, such as endothelial cells and fibroblasts. They used to have different names depending either on their origin, such as *lymphokines* (produced by lymphocytes), *monokines* (substances produced by monocytes or macrophages) or on their activity: *chemokines, interleukins, interferons*.

Pages 506 to 511 are not shown in this preview.

Biologically active functions mediated by complement products

Activation of complement results in the production of several biologically active molecules which contribute to killing of cell, opsonization, chemotaxis, anaphylaxis and inflammation.

The most important action of complement is to facilitate the uptake and destruction of pathogens by phagocytes. This occurs by the specific recognition of bound complement components by complement receptors on phagocytes. These complement receptors bind pathogens opsonized with complement components: opsonization of pathogens is a major function of C3b and its proteolytic derivatives. C4b also acts as an opsonin but has a relatively minor role. There are several different complement receptors (CRs). CR1 and CR3 are especially important in inducing phagocytosis of bacteria with complement components on their surface.

The small complement fragments C3a, C4a and C5a act on specific receptors to produce local inflammatory responses. When produced in large amounts they induce a shocklike syndrome similar to that seen in a systemic allergic reaction involving IgE antibodies. Such a reaction is termed *anaphylactic shock* and these small fragments of complement are therefore often referred to as *anaphylotoxins*. Of the three, C5a is the most stable and has the highest specific biological activity.

Biologically active functions	*Complement component*
Cell lysis	C5b–9 (membrane-attack complex)
Inflammatory response	C3a, C4a, and C5a (anaphylatoxins)
Chemotaxis of leukocytes	C3a, C5a
Opsonization of particulate antigens	C3b, C4b
Viral neutralization	C3b, C5b–9 (membrane-attack complex)
Solubilization and immune clearance	C3b

5.16 Hypersensitivity

Hypersensitivity is an exaggerated immune response that results in tissue damage and is manifested in the individual on a second or subsequent contact with an antigen.

Hypersensitivity has been traditionally classified into *immediate* and *delayed* types based on the time required for a sensitized host to develop clinical reactions on re-exposure to the antigen. Later, Gell and Coombs proposed a classification scheme which defined four types of hypersensitivity reactions.

Type I Hypersensitivity

Type I hypersensitivity (also known as *allergic reaction*) is induced by antigens referred to as *allergens*. The term allergen refers specifically to nonparasitic antigens capable of stimulating type I hypersensitive responses. Type I hypersensitive reactions are IgE-mediated humoral antibody responses. These IgE-mediated reactions are stimulated by the binding of IgE (via its Fc region) to high-affinity IgE-specific Fc receptors expressed on mast cells and basophils. When cross linked by antigens, the IgE antibodies trigger the mast cells and basophils to release *primary mediators,* vasoactive amines, stored in the granules (degranulation). The most significant primary mediators are histamine, proteases, eosinophil chemotactic factor, neutrophil chemotactic factor, and heparin.

These mediators cause all the normal consequences of an acute inflammatory reaction - increased vascular permeability, smooth muscle contraction, granulocyte chemotaxis and extravasation etc. Mast cell activation via Fc also leads to the production of two other types of mediators. These *secondary mediators*, unlike the stored granule contents, must be synthesized *de novo* and comprise arachidonic acid derivatives (prostaglandins and leukotrienes), platelet-activating factor, bradykinins, and various cytokines.

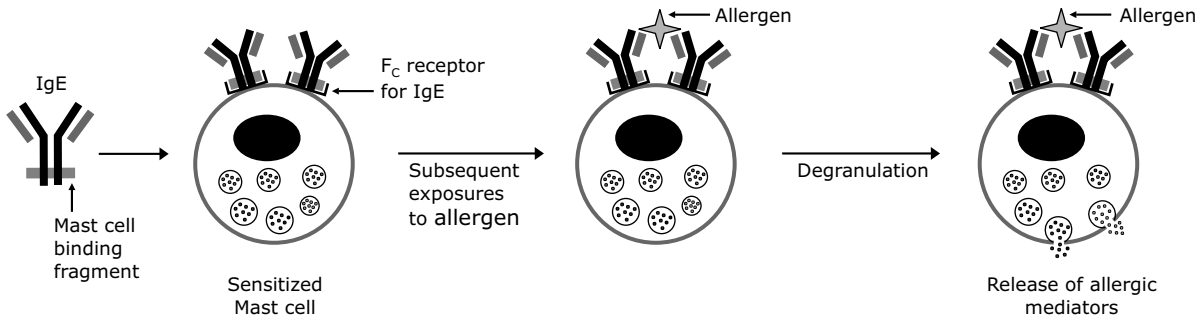


Figure 5.53 Ag induces crosslinking of IgE bound to mast cells and basophils with release of vasoactive mediators.

Type I hypersensitivity can be *anaphylaxis* or *atopy*. Anaphylaxis is a very rapid, life-threatening, severe whole body allergic reaction. It is caused by re-exposure to a previously encountered antigen. Atopy (atopic allergy) is a hereditary tendency to develop allergic reaction to substances such as pollen, food, insect venom etc.

Type II Hypersensitivity

Type II hypersensitivity is generally called a *cytolytic* or *cytotoxic reaction* because it results in the destruction of host cells, either by lysis or toxic mediators. Type II Hypersensitivity is caused by antibodies binding to cells or tissue antigens. The antibodies are of the IgM or IgG classes and cause cell destruction by Fc dependent mechanisms either directly or by recruiting complement via the classical pathway. Classical examples of type II hypersensitivity reactions are the response exhibited by a person who receives a transfusion with blood from a donor with a different blood group and erythroblastosis fetalis.

Two different antibody-mediated mechanisms are involved in these cytotoxic reactions. In *complement-mediated hypersensitivity* reactions, the antibodies react with a cell membrane component, leading to complement fixation. This activates the complement cascade and leads either to lysis of the cell or opsonization. Blood cells are most commonly affected by this mechanism.

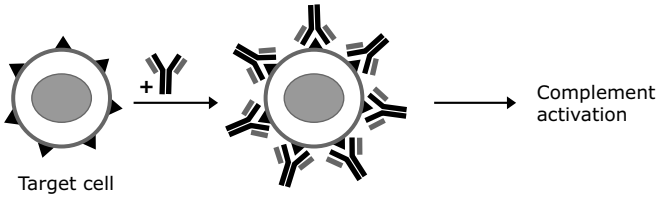


Figure 5.54 Antibody subclasses activate the complement system, creating pores in the membrane of a foreign cell.

Antibody-dependent cell mediated cytotoxicity (ADCC) used Fc receptors expressed on many cell types (e.g. natural killer cells, macrophages, neutrophils, eosinophils) as a means of bringing these cells into contact with antibody-coated target cells. Lysis of these target cells requires contact but does not involve phagocytosis or complement fixation. Instead, ADCC lysis of target cells is analogous to that of cytotoxic T cells and involves the release of cytoplasmic granules containing *perforin* and *granzymes* that activate events leading to apoptosis. ADCC reactions involve IgG and IgG Fc receptors.

This page intentionally left blank.

5.17 Autoimmunity

The body is normally able to distinguish its own self-antigens from foreign nonself antigens and does not mount an immunologic attack against itself. This phenomenon is called *immune tolerance*. **Autoimmunity** is a condition in which structural or functional damage is produced by the action of immunologically competent cells or Ab against self antigen. Autoimmunity literally means *protection against self*, but actually it implies *injury to self*, and therefore sometimes the term is also under criticism.

Autoimmune disease results from the activation of self-reactive T and B-cells that, following stimulation by genetic or environmental triggers, cause actual tissue damage. Four factors influence the development of autoimmune disease. These factors are genetic, viral, hormonal and psycho-neuro-immunological (the influence of stress and neurochemicals). All four of these factors can affect gene expression, which directly or indirectly interferes with important immunoregulatory actions. Based on the site of involvement and nature of lesions autoimmune diseases may be classified as *hemocytolytic*, *localized* (or organ specific), *systemic* (or non-specific) and *transitory* diseases. Important examples of autoimmune diseases in human and their respective autoantigen are given below in the table.

Table 5.14 Some autoimmune diseases in humans

Disease	*Autoantigen*
Autoimmune hemolytic anemia	Rh blood group
Graves disease	Thyroid-stimulating hormone receptor
Multiple sclerosis	Myelin basic protein
Myasthenia gravis	Acetylcholine receptor
Rheumatoid arthritis	Unknown synovial joint antigen
Systemic lupus erythematosus	DNA, histones, snRNP
Type 1 diabetes mellitus	Pancreatic beta cell antigen

5.18 Transplantation

The immune system has evolved as a way of discriminating between self and non-self. This discriminating power of the immune system between self and non-self is undesirable in the case of tissue transplant from one individual to another for therapeutic purposes. Indeed, result of transplants culminates in the phenomenon of *graft rejection*. Before the discussion about the immunological mechanisms associated with graft rejection, it is important to understand the various gradations in relationship from donor to recipient.

Isograft : Graft between genetically identical individuals (syngeneic). In humans, an isograft (or syngraft) can be performed between monozygotic twins.

Allograft : Transplants between genetically different individuals within a species.

Xenograft : A graft between individuals from different species.

Autograft : A graft or transplant from one body part to another on the same individual.

Transplanting tissue that is not immunologically privileged generates the possibility that the recipient's cells will recognize the donor's tissue as foreign. This triggers the recipient's immune mechanisms, which may destroy the donor tissue. Such a response is called a *graft rejection reaction*. Some transplanted tissues do not stimulate an immune response. For example, a transplanted cornea is rarely rejected because lymphocytes do not circulate into the anterior chamber of the eye. This site is considered an immunologically privileged site. Another example of a privileged tissue is the heart valve.

A tissue rejection reaction can occur by two different mechanisms. *First*, foreign class II MHC molecules on transplanted tissue, or the graft is recognized by host T-helper cells, which aid cytotoxic T-cells in graft destruction. Cytotoxic T-cells then recognize the graft through the foreign class I MHC molecules. This response is much like the activation

Pages 516 to 524 are not shown in this preview.

Chapter 06

Genetics

All living organisms reproduce. Reproduction results in the formation of offspring of the same kind. However, the resulting offspring need not and, most often, does not totally resemble the parent. Several characteristics may differ between individuals belonging to the same species. These differences are termed *variations.* The mechanism of transmission of characters, resemblances as well as differences, from the parental generation to the offspring, is called *heredity.* The scientific study of heredity, variations and the environmental factors responsible for these, is known as *genetics* (from the Greek word *genno* = give birth). The word *genetics* was first suggested to describe the study of inheritance and the science of variation by prominent British scientist William Bateson.

Genetics can be divided into three areas: classical genetics, molecular genetics and evolutionary genetics. In *classical genetics,* we are concerned with Mendel's principles, sex determination, sex linkage and cytogenetics. *Molecular genetics* is the study of the genetic material: its structure, replication and expression, as well as the information revolution emanating from the discoveries of recombinant DNA techniques. *Evolutionary genetics* is the study of the mechanisms of evolutionary change or changes in gene frequencies in populations (population genetics).

Classical genetics

6.1 Mendel's principles

Gregor Johann Mendel (1822–1884), known as the *Father of Genetics,* was an Austrian monk. In 1856, he published the results of *hybridization experiments* titled *Experiments on Plant Hybrids* in a journal "The proceeding of the Brunn society of natural history" and postulated the principles of inheritance which are popularly known as **Mendel's laws**. But his work was largely ignored by scientists at that time. In 1900, the work was independently rediscovered by three biologists - Hugo de Vries of Holland, Carl Correns of Germany and Erich Tschermak of Austria. Mendel did a statistical study (he had a mathematical background). He discovered that individual traits are inherited as discrete *factors* which retain their physical identity in a hybrid. Later, these factors came to be known as *genes*. The term was coined by Danish botanist Wilhelm Johannsen in 1909. A gene is defined as a *unit of heredity* that may influence the outcome of an organism's traits.

Mendel's experiment

Mendel chose the garden pea, *Pisum sativum*, for his experiments since it had the following advantages.

1. Well-defined discrete characters
2. Bisexual flowers
3. Predominant self fertilization
4. Easy hybridization
5. Easy to cultivate and relatively short life cycle

Characters studied by Mendel

The characteristics of an organism are described as *characters* or *traits*. Traits studied by Mendel were clear cut and discrete. Such clear-cut, discrete characteristics are known as *Mendelian characters.* Mendel studied seven characters/ traits (all having two variants) and these are:

		Dominant	*Recessive*
1.	Stem length	Tall	Dwarf
2.	Flower position	Axial	Terminal
3.	Flower colour	Violet	White
	Seed coat colour	Grey	White
4.	Pod shape	Inflated	Constricted
5.	Pod colour	Green	Yellow
6.	Cotyledon colour	Yellow	Green
7.	Seed form	Round	Wrinkled

Flower colour is positively correlated with seed coat colours. Seeds with white seed coats were produced by plants that had white flowers and those with gray seed coats came from plants that had violet flower.

Allele

Each gene may exist in alternative forms known as *alleles*, which code for different versions of a particular inherited character. We may also define alleles as genes occupying corresponding positions on homologous chromosomes and controlling the same characteristic (e.g. height of plant) but producing different effects (tall or short). The term *homologous* refers to chromosomes that carry the same set of genes in the same sequence, although they may not necessarily carry identical alleles of each gene.

Wild-type versus Mutant alleles

Prevalent alleles in a population are called *wild-type alleles*. These alleles typically encode proteins that are made in the right amount and function normally. Alleles that are present at less than 1% in the population and have been altered by mutation are called *mutant alleles.* Such alleles *usually* result in a reduction in the amount or function of the wild-type protein and are most often inherited in a recessive fashion.

Dominant and Recessive alleles

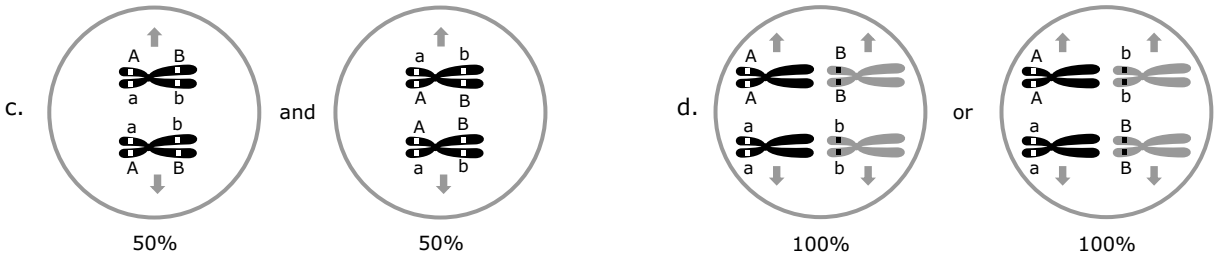
A *dominant allele* masks or hides expression of a recessive allele and it is represented by an *uppercase letter.* A *recessive allele* is an allele that exerts its effect only in the *homozygous state* and in heterozygous condition its expression is masked by a dominant allele. It is represented by a *lowercase letter*.

Homozygous and Heterozygous

Each parent (diploid) has two alleles for a trait — they may be:

1. **Homozygous**, indicating they possess two identical alleles for a trait.
   a. *Homozygous dominant* genotypes possess two dominant alleles for a trait (TT).
   b. *Homozygous recessive* genotypes possess two recessive alleles for a trait (tt).
2. **Heterozygous** genotypes possess one of each allele for a particular trait (Tt).

Pages 527 to 538 are not shown in this preview.



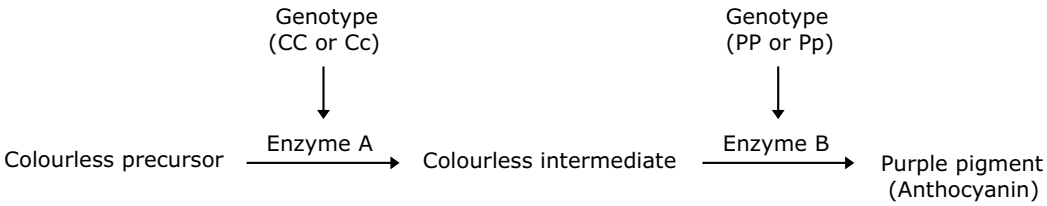
Solution

Choice 'a' represents the correct chromosomal arrangement in meiotic metaphase I. In this situation, as a result of independent assortment – 50% meiotic product will be AB + ab and 50% will be Ab + aB.

6.3 Gene interaction

According to Mendel, genes are functioning independently of each other i.e. each of seven traits considered was controlled by a single gene. But many traits of an organism are determined by the complex contribution of many different genes. When two or more different genes (non-allelic) influence the outcome of single trait, this is known as a *gene interaction*.

The first case of two different genes interacting to affect a single trait was discovered by William Bateson and Reginald Punnett in 1906. They discovered an unexpected gene interaction when they studied crosses involving the sweet pea, *Lathyrus odoratus*. When they crossed true breeding purple flowered plant to a true breeding white flowered plant, the F1 generation was all purple flowered plants and the F2 generation (produced by self fertilization of the F1 generation) contained purple and white flowered plants in a 3 : 1 ratio. But when they crossed two different varieties of white flowered plants then all F1 generation plants had purple flowers. When these purple flower plants were allowed to self fertilized, the F2 generation contained purple and white flowers in a ratio of 9 purple : 7 white. How can this unexpected result be explained? This surprising result was explained by Bateson and Punnett by considering the involvement of two different (non-allelic) genes; because the F2 9 : 7 ratio is a variation of the 9 : 3 : 3 : 1 ratio. Let us consider the formation of the purple pigment in which products of two different genes are involved.



C (purple colour producing) allele is dominant to c (white)

P (purple colour producing) allele is dominant to p (white)

In the above pathway, a colourless precursor molecule must be acted on by two different enzymes to produce the purple pigment. Gene C encodes a functional enzyme A, which converts the colourless precursor into a colourless intermediate and finally gene P encodes enzyme B, which gives purple colour by converting colourless intermediate. If any of these two genes will be in homozygous recessive condition (cc or pp) then purple colour will not appear. Thus the genotype cc can hide or mask the phenotype expression of genotype PP or Pp.

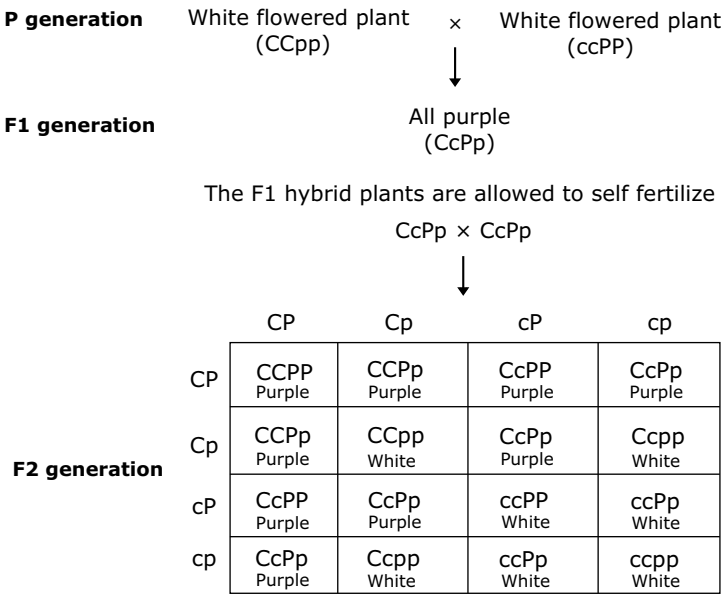


Figure 6.8 9 : 7 phenotypic ratio in F2 generation.

The purple colour appears only when dominant alleles of both genes are present. When one or both genes have only recessive alleles, the colour will be white.

Epistasis

The term *epistasis* (Greek for *standing upon*) describes a type of gene interaction when one gene masks or modifies the expression of another gene at distinct locus. Any gene that masks the expression of another non-allelic gene is *epistatic* to that gene. The gene suppressed is *hypostatic*. In the pathway discussed for formation of purple colour, when either is homozygous recessive (cc or pp) that gene is epistatic to the other.

Epistasis is different from dominance. Epistasis is the interaction between different genes (non-alleles). Dominance is the interaction between different alleles of the same gene i.e. intraallelic.

Table 6.3 Comparison between dominance and epistasis

Dominance	*Epistasis*
Allelic suppression.	Non-allelic suppression.
It involves a single pair of alleles.	It involves two pairs of alleles.
A gene suppresses the expression of its allele.	A gene suppresses the expression of its non-allele.
The effect of a recessive allele is suppressed.	Epistatic allele suppresses the effect of both dominant and recessive non-allele.
The effect is only due to dominant allele.	It may be due to dominant or recessive allele.

Now the term epistasis has come to be synonymous with almost any type of gene interaction that involves the masking or modifying of one of the gene effects. When epistasis is operative between two gene loci, the number of phenotypes appearing in the offspring will be less than four (normal F2 phenotypic classes in case of dihybrid crosses is four, 9 : 3 : 3 : 1). Such bigenic (two genes) epistatic interactions may be of several types.

6.3.1 Dominant epistasis

When the dominant allele of one gene masks the effects of either allele of the second gene, it is termed as dominant epistasis. When the dominant allele at one locus, for example, the *A* allele produces a certain phenotype regardless of the allelic condition of the other locus, then the A locus is said to be epistatic to the B locus. Furthermore, since

Pages 541 to 555 are not shown in this preview.

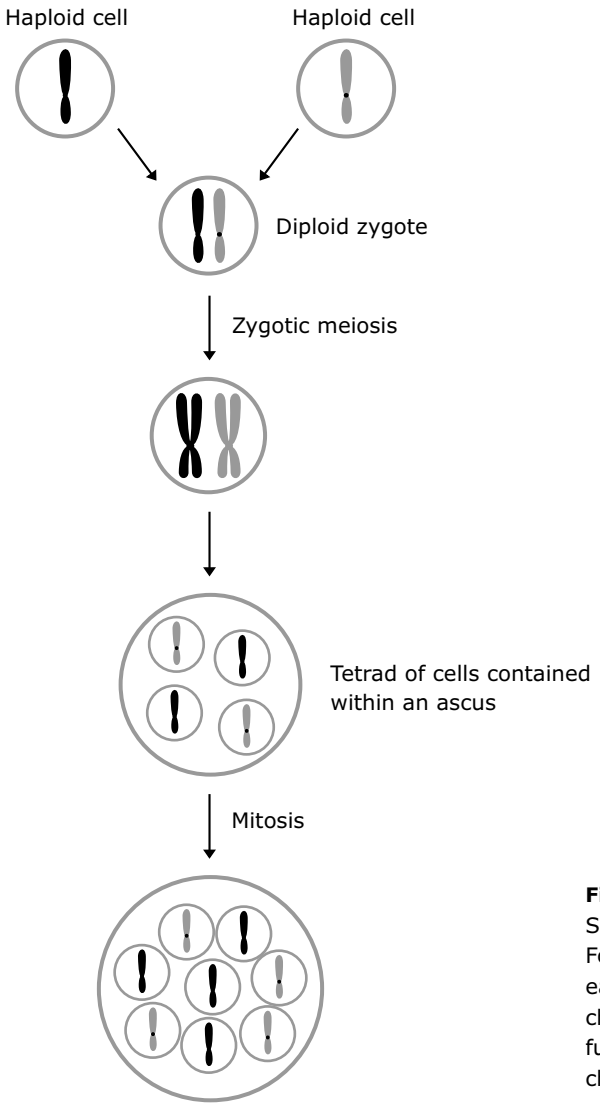


Figure 6.15
Sexual reproduction in ascomycetes. For simplicity, this diagram shows each haploid cell as having only one chromosome per haploid set. However, fungal species actually contain several chromosomes per haploid set.

Ordered *or* unordered tetrad/octad

The arrangement of spores within an ascus varies from species to species. In some cases, the ascus provides enough space for the tetrads or octads of spores to randomly mix together. This is known as an *unordered tetrad* or *octad*. These occur in fungal species such as *S. cerevisiae*. By comparison, other species of fungi produce a very tight ascus that prevents spores from randomly moving around. This can create a *linear tetrad or octad* found in *N.crassa.*

Saccharomyces cerevisiae *Neurospora crassa*

Figure 6.16 Different arrangements of fungal spores.

A key feature of linear tetrads or octads is that the position and order of spores within the ascus reflects their relationship to each other as they were produced by meiosis and mitosis. This idea is schematically shown in figure 6.17.

After the original diploid cell has undergone chromosome replication, the first meiotic division produces two cells that are arranged next to each other within the sac. The second meiotic division then produces four cells that are also arranged in a straight row. Due to the tight enclosure of the sac around the cells, each pair of daughter cells is forced to lie next to each other in a linear fashion. Likewise, when each of these four cells divides by mitosis, each of the daughter cells is located next to each other.

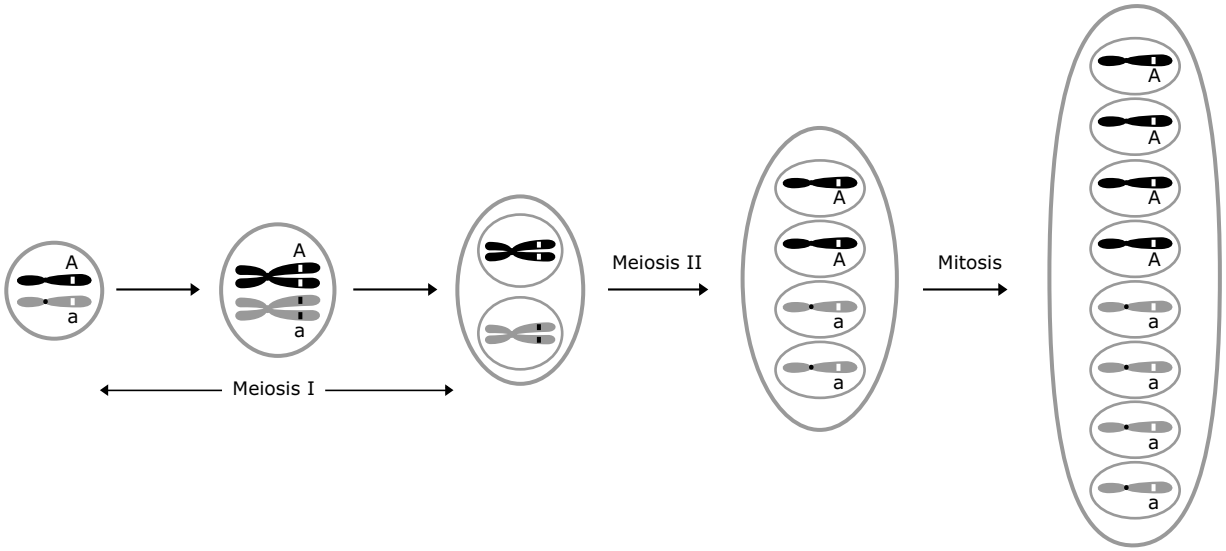


Figure 6.17 Formation of a linear octad in *N. crassa*.

6.5.1 Analysis of ordered tetrad

Linear tetrad analysis can be used to map the distance between a gene and the centromere. This approach has been extensively exploited in *N. crassa.* In *N. crassa,* the products of meiosis are contained in an ordered array of spores. Each mature ascus contains eight ascospores in four pairs, each pair representing one of the products of meiosis. The ordered arrangement of meiotic product makes it possible to map each gene with respect to its centromere; i.e. to determine the recombination frequency between a gene and its centromere. Two cases are possible depending on whether or not there is a crossover between the locus and its centromere.

First case

In the absence of crossing over between a gene and its centromere, the alleles of the gene (for example A and a) must separate in the first meiotic division, this separation is called **First Division Segregation** (FDS). Octad contains a linear arrangement of four haploid cells carrying the A allele, which are adjacent to four haploid cells that contain an allele i.e. 4:4 arrangement of spores within the ascus (figure 6.18).

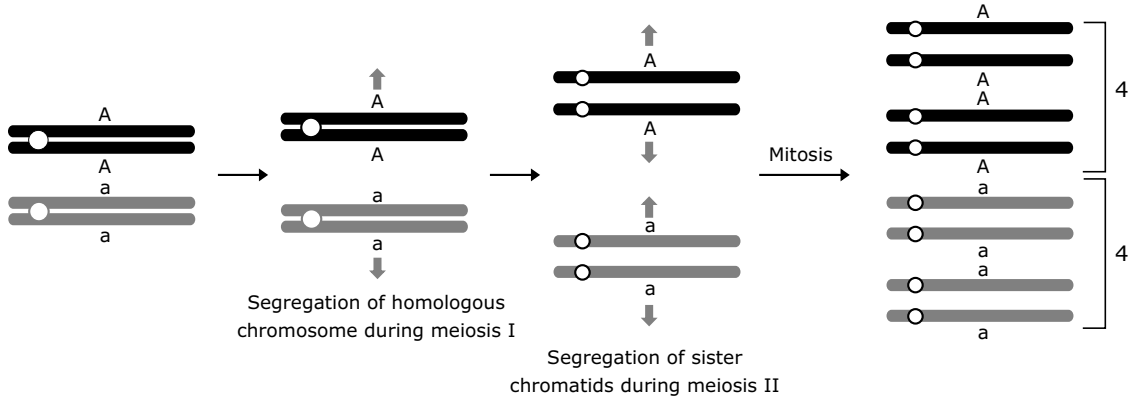


Figure 6.18 First Division Segregation (FDS) : No crossing over produces a 4 : 4 arrangement.

Pages 558 to 563 are not shown in this preview.

6.6.7 Mosaicism

Mosaicism is a condition in which cells within the same individual have a different genetic makeup. Individuals showing mosaicism are referred to as *mosaics*. Mosaicism can be caused by DNA mutations, epigenetic alterations of DNA, chromosomal abnormalities (change in chromosome number and structure) and the spontaneous reversion of inherited mutations. Mosaicism can be associated with changes in either nuclear or mitochondrial DNA. An individual with two or more cell types, differing in chromosome number or structure is either a mosaic or a chimera. If the two cell types originated from a single zygote, the individual is a *mosaic*, and when originated from two or more zygotes that subsequently fused, the individual is a *chimera*.

Mosaicism can exist in both somatic cells (*somatic mosaicism*) and germ line cells (*germline mosaicism*). As their names imply, somatic and germ line mosaicism refer to the presence of genetically distinct groups of cells within somatic and germ line tissues, respectively. If the event leading to mosaicism occurs during development, it is possible that both somatic and germ line cells will become mosaic. In this case, both somatic and germ line tissue populations would be affected, and an individual could transmit the mosaic genotype to his or her offspring. Conversely, if the triggering event occurs later in life, it could affect either a germ line or a somatic cell population. If the mosaicism occurs only in a somatic cell population, the phenotypic effect will depend on the extent of the mosaic cell population; however, there would be no risk of passing on the mosaic genotype to offspring. On the other hand, if the mosaicism occurs only in a germ line cell population, the individual would be unaffected, but the offspring could be affected.

How is somatic mosaicism generated? There are many possible reasons, including somatic mutations, epigenetic changes in DNA, alterations in chromosome structure and/or number, and spontaneous reversal of inherited mutations. In all of these cases, a given cell and those cells derived from it could exhibit altered function.

6.6.8 Sex–linked traits and sex–linked inheritance

In an XY-chromosomal system of sex determination, both X and Y-chromosomes are sex chromosomes. In general, genes on sex chromosomes are described as sex linked genes. However, the term *sex linked* usually refers to loci found only on the X-chromosome; the term *Y linked* is used to refer to loci found only on the Y-chromosome, which control holandric traits (traits found only in males).

Cytogeneticists have divided the X and Y-chromosomes of some species into homologous and non-homologous regions. The latter is called *differential* regions. These differential regions contain genes that have no counterparts on the other sex chromosome. Genes in the differential regions are said to be **hemizygous** (half zygous). Genes in the differential region of the X show an inheritance pattern called **X-linkage**; those in the differential region of the Y show **Y-linkage**. Genes in the homologous region show what might be called *X-and-Y linkage*.

Another important feature of sex linked genes in XY-chromosomal system of sex determination is that females have two X-chromosomes, they can have normal homozygous and heterozygous allelic combinations. But males, with only one copy of the X-chromosome can be neither homozygous nor heterozygous. Hence the term *hemizygous* is used for X-linked genes in males. Since only one allele is present, a single copy of a recessive allele can determine the phenotype, a phenomenon called **pseudodominance**. This is the same way that one copy of a dominant autosomal allele would determine the phenotype of a normal diploid organism; hence the term pseudodominance.

The genes on the differential regions of the sex chromosomes show patterns of inheritance related to sex. The inheritance patterns of genes on the autosomes produce male and female progeny in the same phenotypic proportions, as typified by Mendel's data (for example, both sexes might show a 3:1 ratio). However, crosses following the inheritance of genes on the sex chromosomes often show male and female progeny with different phenotypic ratios. T.H.Morgan demonstrated the X-linked pattern of inheritance in *Drosophila* in 1910, when a white eyed male appeared in a culture of wild type (red-eyed) flies.

Let's look at an example from *Drosophila.* When white-eyed males are crossed with red-eyed females, all the F1 progeny have red eyes, showing that the allele for white is recessive. Crossing the red-eyed F1 males and females produces a 3:1 F2 ratio of red-eyed to the white-eyed flies, but all the white-eyed flies are males.

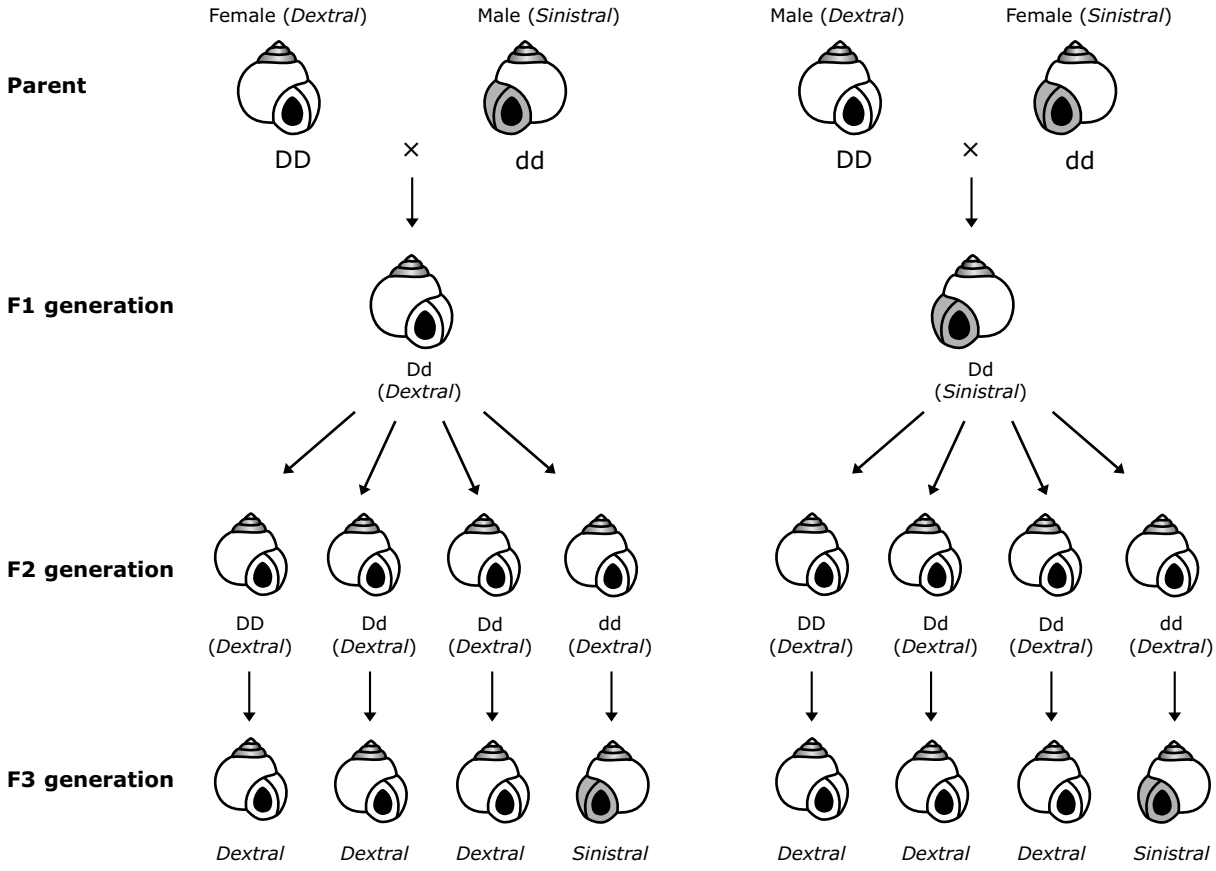


Figure 6.33 Inheritance of the direction of shell coiling in the snail *Lymnaea*. Sinistral coiling is determined by recessive allele d and dextral coiling by dominant allele D. The F2 and F3 generations are obtained by self-fertilization.

The next observation is that the phenotype of the F1 generation is always that of the female parent. One hypothesis would suggest that the genotype of the female controls the genotype of its offspring. *Can these results be confirmed in the subsequent generations*? If the genotypes we assigned to the parents are correct, then the genotype of F1 individuals from each cross are Dd (from DD×dd and dd×DD). If the female genotype does control the phenotype of its offspring, then we would predict that all the F2 snails would have right coils. This is the exact result that is seen. But what would the genotypes of the F2 snails be? If we intermate snails with the genotype Dd, the genotypic ratio should be 3 D_ to 1 dd. These genotypes would not be expressed as a phenotype until the F3 generation. These are the results that were obtained. A general conclusion from all traits that express a *maternal effect* is that the normal Mendelian ratios are expressed one generation than expected. Cytological analysis of developing eggs has provided the explanation of above mentioned result: the genotype of the mother determines the orientation of the mitotic spindle during the second cleavage (mitotic) division in the zygote, and this, in turn, controls the direction of shell coiling of the offspring.

6.9 Cytogenetics

A chromosome is an organized structure of DNA and protein that is found in the nucleus of a eukaryotic cell. The study of the structure, function and abnormalities of chromosome is called *cytogenetics*, a discipline that combines cytology with genetics.

Pages 579 to 593 are not shown in this preview.

6.11 Genome

Genome is the sum total of all genetic material of an organism which store biological information. The nature of the genome may be either DNA or RNA. All eukaryotes and prokaryotes always have a DNA genome, but viruses may either have a DNA genome or RNA genome. The eukaryotic genome consists of two distinct parts: Nuclear genome and organelles (mitochondrial and chloroplast) genome. The nuclear genome consists of linear dsDNA. In a few lower eukaryotes, double-stranded circular plasmid DNA (for example, 2-micron circle in yeast) is also present within the nucleus.

The amount of DNA present in the genome of a species is called a **C-value**, which is characteristic of each species. The value ranges from $<10^6$ bps as in smallest prokaryote, *Mycoplasma* to more than 10^{11} bps for eukaryotes such as amphibians. The genomes of higher eukaryotes contain a large amount of DNA.

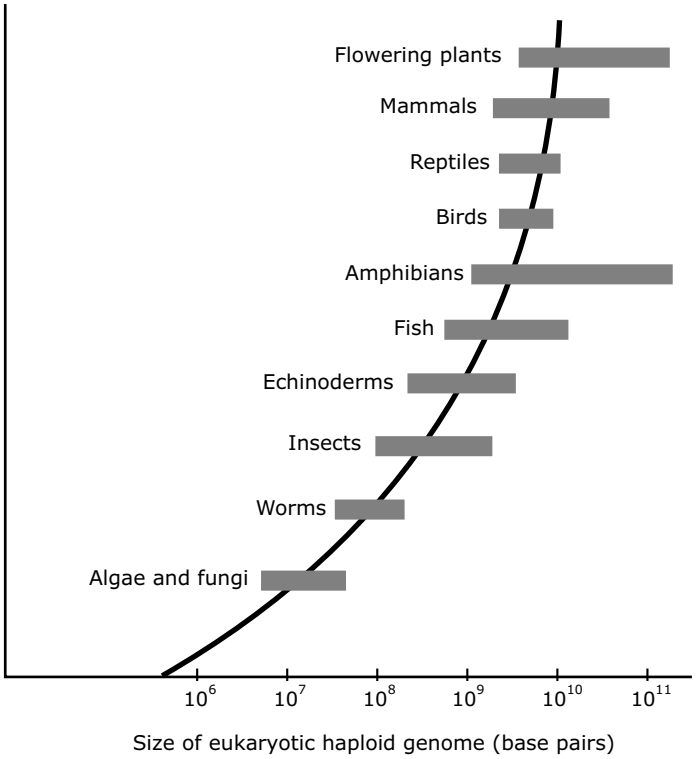


Figure 6.43 The DNA content of the haploid genome of a range of phyla. The range of values within a phylum is indicated by the shaded area.

The DNA content of the organism's genome is related to the morphological complexity of lower eukaryotes, but varies extensively among the higher eukaryotes. In lower eukaryotic organisms like yeast, amount of DNA increases with increasing complexity of organisms. However, in higher eukaryotes there is no correlation between increased genome size and complexity. This lack of correlation between genome size and genetic complexity refers to **C-value paradox**. For example, a man is more complex than amphibians in terms of genetic development, but some amphibian cells contain 30 times more DNA than human cells. Moreover, the genomes of different species of amphibians can vary 100-fold in their DNA contents.

Pages 595 to 611 are not shown in this preview.

6.11.10 Yeast *S. cerevisiae* genome

The yeast genome consists of 16 linear chromosomes, each containing a centromeric region required for chromosome segregation. The nucleotide sequence of the entire *S. cerevisiae* genome has been determined and found to contain 12,068 kb of DNA. Sequence analysis has identified 5885 potential protein coding genes and another 45S RNA coding genes (rRNA, snRNA and tRNA genes). Almost 70% of the yeast genome is devoted to protein coding sequences. Interestingly, unlike most other eukaryotic genes, only about 4% of the about 6000 yeast genes have introns, and even then, most of these genes contain only a single intron within the coding sequence.

6.11.11 *E. coli* genome

E. coli genome comprises single main chromosome and plasmids. The main chromosome is made up of circular dsDNA with a homogeneous distribution of genes. Computer analysis of the *E. coli* DNA sequence identified 4288 actual and proposed gene-coding sequences. It was found that approximately 88% of the genome encodes proteins or RNAs, ~11% appears to be utilized for gene regulatory functions, and <1% consists of repetitive DNA sequences. The average distance between *E. coli* genes is only 120 bp.

6.12 Eukaryotic chromatin and chromosome

A *chromatin* is an organized structure of DNA and protein that is found in the nucleus of eukaryotic cells. It contains a single dsDNA in coiled and condensed form. Chromatin and chromosomes are basically the same thing. The difference is that chromatin is less condensed, extended DNA while chromosomes are highly condensed DNA. The word *chromosome* comes from the Greek word *chroma*, color and *soma,* body due to their property of being very strongly stained by particular dyes. The extent of chromatin condensation varies during the life cycle of the cell. In non-dividing as well as interphase stages of cell, most of the chromatin remain relatively decondensed. The light-staining, less condensed portions of chromatin is termed *euchromatin*. The darkly stained and highly condensed regions of chromatin is termed as *heterochromatin.* In interphase nuclei, chromatin appears to be attached to a nuclear matrix, a proteinaceous structure. DNA sequence attached to nuclear matrix are called MAR (matrix attachment regions). MAR are usually ~70% A·T-rich, but lack any consensus sequences. A chromatin DNA molecule contains three specific nucleotide sequences: *Centromere, Telomere* and *Origin of replication*.

Centromere

The *centromere* is a constricted region of a eukaryotic chromatin/chromosome where the *kinetochore* is assembled and sister chromatids are held together. Although this constriction is termed as centromere, it is usually not located exactly in the center of the chromosome and, in some cases, is located almost at the chromosome's end. The regions on either side of the centromere are referred to as the chromosome's arms. Kinetochore associated with the centromere is a complex of proteins where spindle fibers attach to the chromosome during mitosis/meiosis and help in the proper segregation of sister chromatids or homologous chromosomes. The centromere has no defined DNA sequence. It typically consists of large arrays of tandemly repeated DNA sequences. In humans, the centromeric sequences are made up of 171 bp repeating unit and are called **alphoid DNA**. In the yeast, *Saccharomyces cerevisiae,* the centromeric sequence (CEN) is about 110 bp long and it consists of three types of sequence element:

- CDE-I - 9 bp sequence;
- CDE-II - >90% A·T-rich sequence of 80–90 bp;
- CDE-III - 11 bp highly conserved sequence.

Chromosomes can be classified into following types based on the position of the centromere:

Metacentric: If the centromere is located exactly in the middle of the chromosome, the two arms of the chromosome are nearly equal (median centromere). The chromosome appears V-shaped during anaphasic movement.

Submetacentric: If the centromere is situated some distance away from the middle (submedian centromere), one arm of the chromosome will be shorter than the other, such a chromosome will appear L-shaped during anaphasic movement.

Acrocentric: If the centromere is situated near the end of the chromosome, one arm will be extremely short and other very long (subterminal centromere). These chromosomes appear rod shaped during anaphase.

Telocentric: If the centromere is truly terminal, i.e. situated at the tip of the chromosome, the chromosome is said to be telocentric (terminal centromere).

In chromosome that is not metacentric, *p* represents the short arm of chromosome and *q* represents the long arm of chromosome. Most eukaryotic chromosomes are **monocentric**, having a single centromere, but some are **holocentric** (holokinetic or polycentric) and have diffused centromere. Every point along the length of the chromosome exhibits centromeric activity. The nematode *C. elegans* has holocentric chromosome. In holocentric chromosome, spindle fibers attach along the entire length of chromosome.

During interphase stage of cell cycle, chromatin replicates, resulting in the formation of two copies of each chromatin. As the cell enters M-phase, chromatin condensation leads to the formation of metaphase chromosomes consisting of two identical *sister chromatids*. These sister chromatids are held together at the centromere, which is seen as a constricted chromosomal region. A *cohesin* protein play a role in linking together sister chromatids immediately after replication and keeping them together at centromere.

Telomeres

Telomeres are specialized structures which cap the ends of eukaryotic chromosomes. They have several likely functions – maintaining the structural integrity of a chromosome (if a telomere is lost, the resulting chromosome end is unstable) and ensuring complete replication of the extreme ends of chromosomes. Eukaryotic telomeres consist of a long array of short and tandemly repeated sequences. There may be 100–1000 repeats, depending on the organism. One unusual property of the telomeric sequence is the presence of the G-rich single strand 3′ overhang, measuring between 50 to 300 nucleotides. The G-rich sequence is generated because there is a limited degradation of the C-rich complementary strand. Unlike centromeres, the sequence of telomeres has been highly conserved in evolution – there is considerable similarity in the simple sequence repeat, for example TTGGGG (*Paramecium*), TAGGG (*Trypanosoma*), TTTAGGG (*Arabidopsis*) and TTAGGG (*Homo sapiens*). Two sequence-specific DNA binding proteins – telomeric repeat binding factor 1 (TRF1) and telomeric repeat binding factor 2 (TRF2) bind directly with telomeric sequences, which in turn interact with a larger number of proteins.

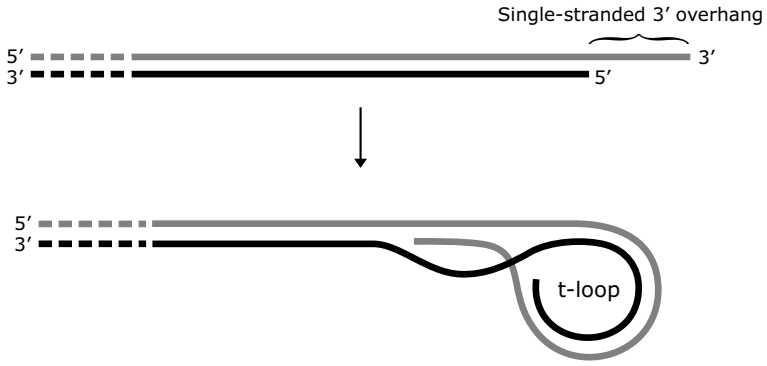


Figure 6.59 DNA at the telomeres consists of G-rich tandem sequences. The G-strand overhangs are important for telomeric protection by formation of a duplex loop. Telomeric duplex DNA forms a loop (t-loop), thus avoiding the sticky end problem. The loop formation is mediated by the TRF2, which bind to telomere repeats and the loop is anchored by the insertion of the G-strand overhang into a proximal segment of duplex telomeric DNA.

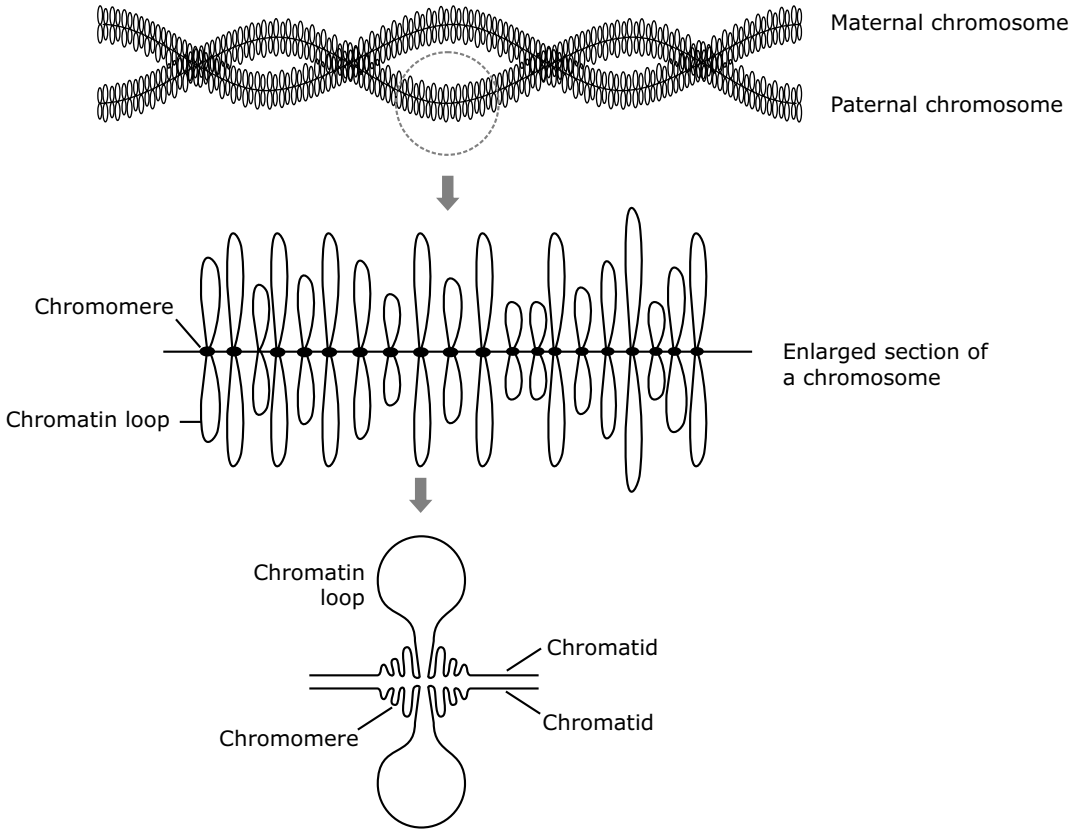


Figure 6.71 Lampbrush chromosome structure. Most of the DNA in each chromosome remains highly condensed in the chromomeres. Each of the two chromosomes shown consists of two closely apposed sister chromatids. This four stranded structure is characteristic of diplotene stage of meiosis.

6.12.6 B–chromosomes

The B-chromosomes (also referred to as *supernumerary* or *accessory chromosomes*) are additional (extra) chromosomes that are present in some individuals in some species. In eukaryotic cells normal chromosomes are termed as A-chromosomes. Most B-chromosomes are mainly or entirely heterochromatic and genetically inert. They are thought to be selfish genetic elements with no defined functions. The evolutionary origin of B-chromosomes is not clear, but presumably they must have been derived from heterochromatic segments of normal A-chromosomes.

6.13 DNA replication

Transmission of chromosomal DNA from generation to generation is crucial to cell propagation. This can only be achieved when chromosomal DNA is accurately replicated, providing two copies of the entire genome for faithful distribution into each daughter cell.

6.13.1 Semiconservative replication

It is crucial that the genetic material is reproduced accurately. When Watson and Crick worked out the double-helix structure of DNA in 1953, they recognized that the complementary nature of the two strands-A paired with T and G paired with C-might play an important role in its replication. Because the two polynucleotide strands are joined only

This page intentionally left blank.

6.13.2 Replicon and origin of replication

DNA replication does not start at random locations but at particular sites, called the *origins* of DNA replication. A unit of DNA in which replication starts from an origin and proceeds bidirectionally or unidirectionally to terminus site is called a **replicon**, a unit of DNA replication. Replicon can be linear or circular. Prokaryotic replicons are usually circular.

In bacterial cells, the circular chromosome contains a unique origin and DNA replication proceeds bidirectionally from the origin to the terminus. Therefore, the whole bacterial genome (~4.6 Mbp for *Escherichia coli*) is a single replicon (*monorepliconic*). On the other hand, eukaryotic cells contain multiple replication origins on single chromosome and hence many replicons (*multirepliconic*). Individual replicons in eukaryotic genomes are relatively small and generally 40 to 100 kb in size.

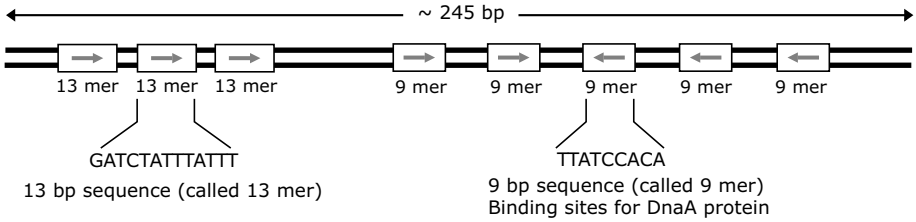


Figure 6.73 *E. coli* origin of replication, oriC. oriC contains repetitive 9-bp and A.T rich 13-bp sequences, referred to as 9-mers and 13-mers, respectively. Multiple copies of DnaA protein bind to the 9-mer and then 'melt' the 13-mer segments.

The origin of replication is a *cis* acting sequence. In *E. coli* single origin of the replication present in the chromosome is referred to as *oriC*. It spans approximately 245 bp of DNA. It contains two short repeat motifs, one of nine nucleotides and the other of 13 nucleotides. The nine-nucleotide repeat, five copies of which are dispersed throughout *oriC*, is the binding site for a protein called DnaA. The result of DnaA binding is that the double helix opens up ('melts') within the tandem array of three AT-rich, 13-nucleotide repeats located at one end of the *oriC* sequence.

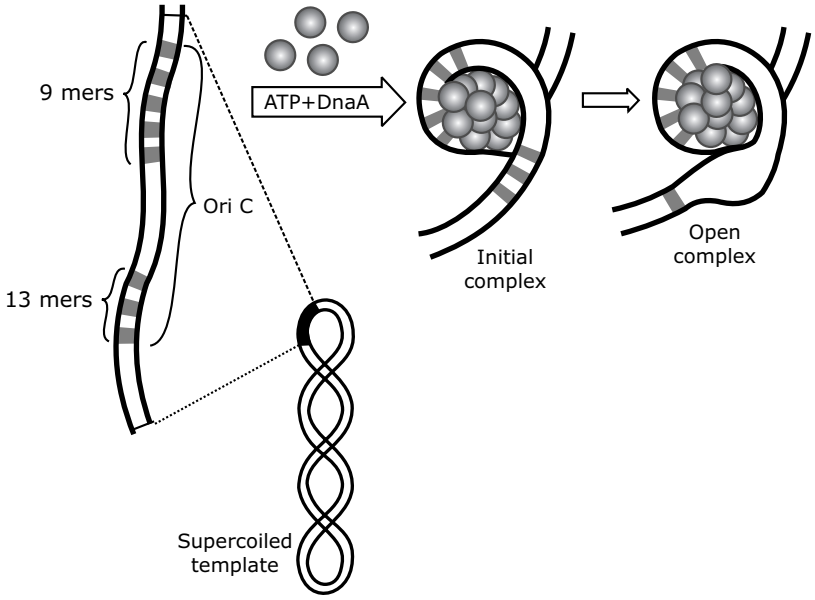


Figure 6.74 Initiation at oriC occurs after DnaA protein binds the five 9 mers. The 13 mer region is then denatured, and this open complex serves as a replication start site. Adapted and redrawn from D. Bramhill and A. Kornberg. Cell, 1988,54; 915-918.

Pages 627 to 639 are not shown in this preview.

6.13.6 Replication of mitochondrial DNA

Small and mostly circular mitochondrial and chloroplast DNA use a slightly different process of replication. Replication of circular double stranded mitochondrial DNA starts at a specific origin. But duplex DNA uses different origin sequences to initiate replication of each DNA strand. Initially, only one of the two parental strands is used as a template for synthesis of a new strand. Synthesis proceeds for only a short distance, displacing the original complementary strand, which remains single-stranded. This pattern of replication generates a **displacement** or **D loop** (hence, termed as *displacement replication*). A single D loop is found as an opening of 500–600 bases in mammalian mitochondria. Some mitochondrial DNAs possess several D loops which reflects the presence of multiple origins. Replication of the complementary strand is initiated when its origin is exposed by the movement of the first replication fork. The similar mechanism is employed in chloroplast DNA. Mammalian mitochondrial DNA is replicated by the DNA polymerase γ. The replisome machinery is formed by DNA polymerase, TWINKLE and mitochondrial SSB proteins. TWINKLE is a helicase, which unwinds short stretches of dsDNA in the 5′ to 3′ direction.

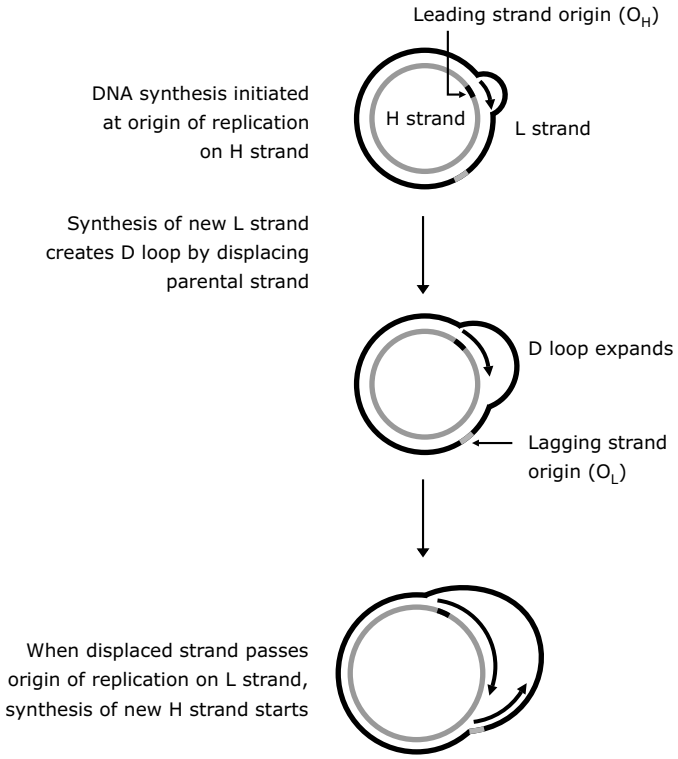


Figure 6.86 Replication of mammalian mitochondrial DNA. Replication starts at a specific origin in the circular duplex DNA. Initially only one of the two parental strands (the H strand in mammalian mitochondrial DNA) is used as a template for synthesis of a new strand. Synthesis proceeds for only a short distance, displacing the original partner (L) strand, which remains single-stranded. There is separate origins for L and H strand.

6.14 Recombination

Genomes are dynamic entities that change as a result of mutations and recombinations. Recombination is a large-scale rearrangement of a DNA molecule that involves the breakage and reunion of DNA. It was first recognized as the process responsible for crossing-over during meiosis of eukaryotic cells, and was subsequently implicated in the integration of the transferred DNA into bacterial genomes after conjugation, transduction or transformation. Genetic recombination events fall into two general classes:

6.22 RNA interference

RNA interference (abbreviated RNAi) is an evolutionarily conserved mechanism of gene regulation that is induced by *small silencing RNA* in a sequence-specific manner. In 1998, Fire and Mello first established this in *C. elegans*. Historically, RNA interference was known by other names, including *post transcriptional gene silencing* (PTGS), *transgene silencing* and *quelling*. RNAi has been observed in all eukaryotes, from yeast to mammals. RNA interference has an important role in post-transcriptional gene regulation, transposon regulation and defending cells against viruses. Two types of small silencing RNA molecules – small interfering RNA (siRNA) and microRNA (miRNA) – are central to RNA interference.

siRNAs mediated RNAi

In the siRNAs mediated RNAi pathway, the dsRNAs are processed into siRNAs duplexes comprised of two ~21 nucleotides long strands with two nucleotides overhangs at the 3' ends by an enzyme called Dicer. **Dicer** is a ~200 kDa multidomain, an RNase III family enzyme that functions in processing dsRNA to siRNA. The Dicer includes an ATPase/RNA helicase domain, catalytic RNase III domains, and dsRNA binding domain. Dicer and a dsRNA binding protein (together form the RISC loading complex) then load the RNA duplex into RISC. The siRNA is thought to provide target specificity to RISC through base pairing of the guide strand with the target mRNA. Only one of the two strands, which is known as the *guide strand*, directs the gene silencing. The other *anti-guide strand* or *passenger strand* is degraded during RISC activation. The active components of an RNA-induced silencing complex (RISC) are endonucleases called **argonaute** proteins, which cleave the target mRNA strand complementary to their bound siRNA.

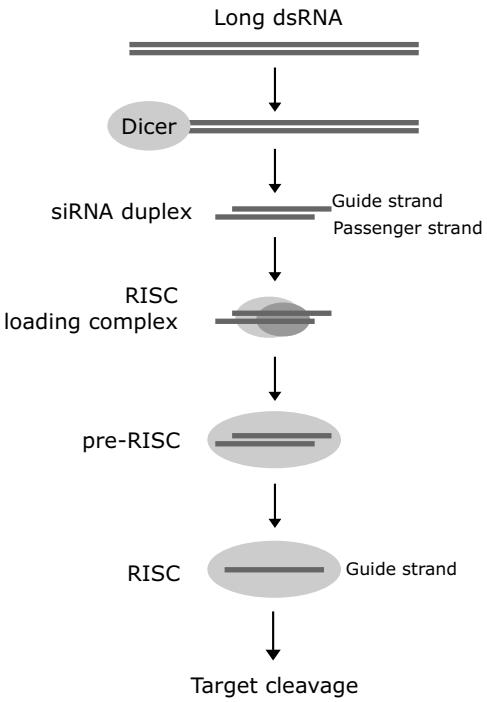


Figure 6.155 dsRNA precursors are processed by Dicer to generate siRNA duplexes containing guide and passenger strands. RISC-loading complex loads the duplex into RISC. The passenger strand is later destroyed and the guide strand directs RISC to the target RNA.

miRNAs mediated RNAi

miRNAs (microRNAs) are small, non-coding RNA molecules encoded in the genomes of plants, animals and their viruses. These highly conserved, 20–25 mer RNAs appear to regulate gene expression post-transcriptionally by

Pages 707 to 712 are not shown in this preview.

Problem

If poly-G is used as a messenger RNA in an incorporation experiment, glycine is incorporated into a polypeptide. If poly-C is used, proline is incorporated. If both poly-G and poly-C are used, no amino acids are incorporated into protein. Why?

Solution

We are mixing two RNA strands that are complementary; these strands will form a double-stranded RNA molecule. Since we observed the incorporation of no amino acids, the ribosome must not be able to read a double-stranded molecule.

Wobble hypothesis

It was first proposed that a specific tRNA anticodon would exist for every codon. If that were the case, at least 61 different tRNAs, possibly with an additional 3 for the chain-terminating codons, would be present. In 1966, Francis Crick devised the wobble concept to explain these observations. It states that the base at the 5′ end of the anticodon also shows non-standard base pairing with any of several bases located at the 3′ end of a codon. So, first base of anticodon and third base of codon is the wobble position. For example, U at the wobble position can pair with either adenine or guanine, while I can pair with U, C or A. However, the wobble rules do not permit any single tRNA molecule to recognize four different codons. Three codons can be recognized only when inosine occupies the first (5′) position of the anticodon.

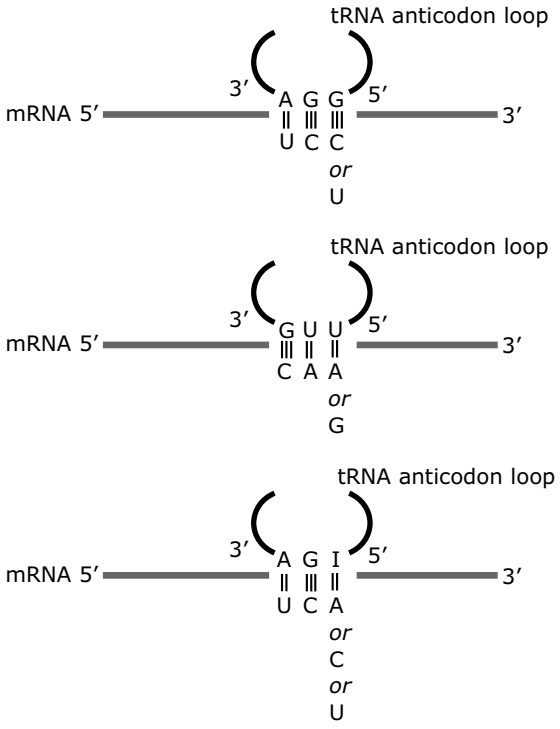


Figure 6.158 Wobble pairing between the anticodon on the tRNA and the codon in the mRNA.

Table 6.34 Pairing combinations with the Wobble concept

Base in anticodon	*Base in codon*
A	U
G	U or C
C	G
U	A or G
I	A, U, or C

6.24 Protein synthesis

The fundamental process of protein synthesis is the formation of a peptide bond between the carboxyl group of one amino acid or at the end of a growing polypeptide chain and a free amino group on an amino acid. Proteins are made up from a set of 20 amino acids called *standard* amino acids. Each of the 20 amino acids is specified by specific codon/s. One additional amino acid - selenocysteine present in some polypeptide is directed by a modified reading of the genetic code (5'UGA3'). Polypeptide synthesis proceeds from N-terminus to C-terminus and ribosome read mRNA in the 5' to 3' direction. Three kinds of RNA molecules perform different but cooperative functions in protein synthesis:

mRNA

mRNA carries the genetic information copied from DNA in the form of a series of three-base code words (codons), each of which specifies a particular amino acid.

Comparison of the structures of prokaryotic and eukaryotic mRNA

Eukaryotic mRNAs are mostly monocistronic; having an average size of 1500 to 2000 nucleotides. It has a 5' cap, which is recognized by the small ribosomal subunit. Protein synthesis, therefore, begins at an initiation codon near the 5' end of the mRNA. Upstream of the initiation codon contains a non-translatable sequences called 5' UTR (5'-untranslated region) or *leader* sequence. Similarly non translatable sequences at the 3' end after stop codon is termed as 3' UTR (3'-untranslated region) or *trailer* sequence, which varies in length and sequence.

In *prokaryotes*, most of the mRNAs are polycistronic. In contrast to eukaryotic mRNAs, the 5' end has no cap-like structure, and there are multiple ribosome-binding sites (called *Shine-Dalgarno sequences*) within the polycistronic mRNA chain, each resulting in the synthesis of a different protein. Just like prokaryotic mRNA, eukaryotic mRNA also contains 5' UTR and 3' UTR.

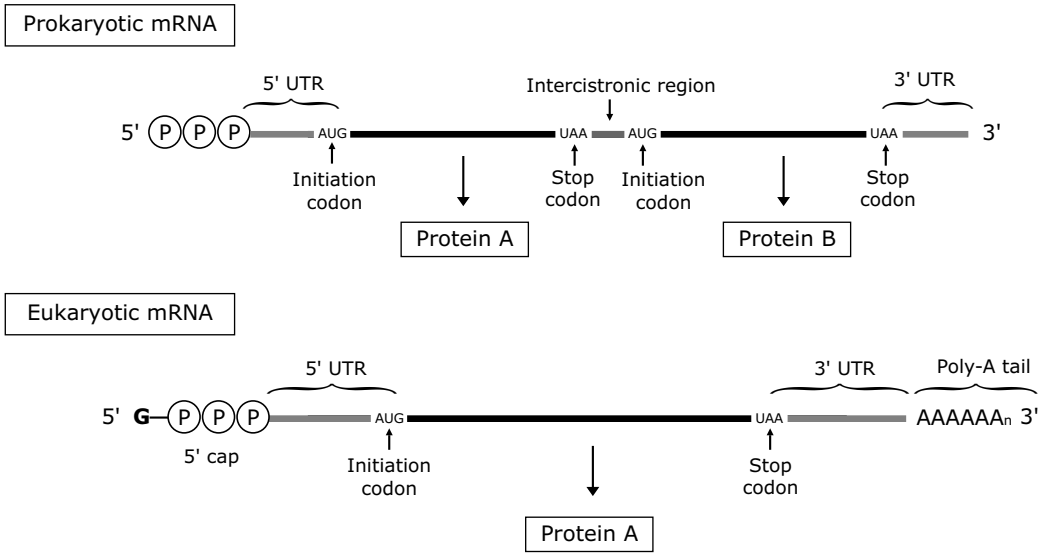


Figure 6.159 All mRNAs (monocistronic and polycistronic) contain two types of region – the coding region (which starts with initiation codon and ends with a stop codon) and untranslated region (5'- and 3'-UTR). A polycistronic mRNA also contains intercistronic regions. They vary greatly in size: they may be as long as 30 nucleotides.

An mRNA can be translated in three different **reading frames**, depending on where the decoding process begins. However, only one of the three possible reading frames in an mRNA encodes the required protein. Any sequence of bases (in DNA or RNA) that could, at least theoretically, encode a polypeptide, is known as an *open reading frame*,

Pages 715 to 730 are not shown in this preview.

Ubiquitination or ubiquitylation of substrates

The most well-established means of targeting proteins to proteasomes is by their modification with chains of ubiquitin. **Ubiquitin**, a highly conserved eukaryotic protein of 76 amino acid residues, is usually attached to substrates by an isopeptide bond between a substrate's lysine residue and the C-terminal glycine of ubiquitin. The process of ubiquitylation occurs in following steps:

Activation of ubiquitin: Ubiquitin is activated by an E1; *ubiquitin-activating enzyme.* E1 becomes covalently linked to free ubiquitin through the free C-terminal residue of ubiquitin, in an energy-dependent manner.

Transfer of ubiquitin from E1 to E2: The activated ubiquitin is subsequently transferred to a cysteine residue present on an E2; *ubiquitin-conjugating enzyme*.

Ligation of ubiquitin to target protein: Finally, E3; *ubiquitin ligases* (~500 in humans) transfer the activated ubiquitin from E2 to a Lys amino acid residue of its target protein, forming an isopeptide bond. A ubiquitinated protein is proteolytically degraded to short peptides in an ATP-dependent process mediated by proteasome.

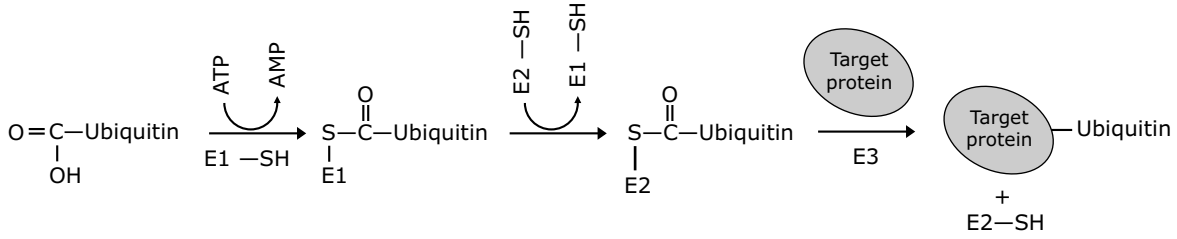


Figure 6.176 The reactions involved in the attachment of ubiquitin to a protein. In the first part of the process, ubiquitin's terminal carboxyl group is joined, via a thioester linkage, to E1 in a reaction driven by ATP hydrolysis. The activated ubiquitin is subsequently transferred to a sulfhydryl group of E2 and then in a reaction catalyzed by E3, to the amino group of a lysine residue on a target protein.

6.24.7 N-end rule

The half-lives of proteins in a living cell range from a few seconds to many days. The metabolic stability of proteins is guided by the presence of *degradation signals* (or degrons). The essential component of one degradation signal is a destabilizing N-terminal residue of a protein. This signal is called the **N-degron**. A set of N-degrons containing different destabilizing residues yields a rule, termed the **N-end rule** which relates the *in vivo* half-life of a protein based on its N-terminal residue and its post-translational modification. The N-end rule operates in all organisms examined, including the bacterium *E. coli*, the yeast *S. cerevisiae* and mammals. In eukaryotes, the N-degron comprises at least two determinants: a destabilizing N-terminal residue and an internal lysine (or lysines). The Lys residue is the site of ubiquitin ligation.

N-end rule pathway

Proteins with N-degron are degraded via the *N-end rule pathway*. In eukaryotes, the N-end rule pathway is a part of the ubiquitin system. This pathway is present in both the cytosol and the nucleus. In this pathway, proteins bearing a destabilizing amino acid residue at their N-terminus are degraded by the ubiquitin proteasome system (in bacteria by the ATP-dependent protease ClpAP, a functional counterpart of the eukaryotic 26S proteasome).

In bacteria, destabilizing residues are divided into two groups- primary and secondary destabilizing residues. In contrast to bacteria, destabilizing residues in eukaryotes can be classified into three hierarchical levels - primary, secondary and tertiary. The primary destabilizing residues fall into two categories – type 1 (basic N-terminal residues Arg, Lys and His) and type 2 (bulky hydrophobic N-terminal residues Ile, Leu, Phe, Tyr and Trp). In general, tertiary destabilizing residues (Asn, Gln and Cys) are first modified to generate a secondary destabilizing residue (Asp, Glu and oxidized Cys). Finally, the modified N-terminal amino acid is arginylated to create a substrate bearing a type 1 primary destabilizing residue (Arg). All primary destabilizing residues are recognized by an **N-recognins** (also called E3).

Pages 732 to 743 are not shown in this preview.

Loss- and gain- of function mutations

In principle, mutation of a gene might cause a phenotypic change in either of two ways:

- Loss of function *(null)* mutation : the product may have reduced or no function.
- Gain of function mutation : the product may have increased or new function.

Because mutation events introduce random genetic changes, most of the time they result in loss of function. Generally, loss of *function* mutations are found to be recessive. In a wild type diploid cell, there are two wild type alleles of a gene, both making normal gene product. In heterozygotes, the single wild type allele may be able to provide enough normal gene product to produce a wild type phenotype. In such cases, loss of function mutations are recessive. However, some loss of function mutations are dominant. In such cases, the single wild type allele in the heterozygote cannot provide the enough amount of gene product needed for the cells to be wild type. Gain of function mutations usually cause dominant phenotypes, because the presence of a normal allele does not prevent the mutant allele from behaving abnormally.

6.25.3 Fluctuation test

The fluctuation test was invented by *Luria and Delbruck* in 1943 to determine the randomness of mutation in bacteria. They grew a series of *E. coli* cultures in different flasks and then added T1 bacteriophage to each one. Most of the bacteria were killed by the phage, but a few T1 resistant mutants were able to survive. Luria and Delbruck measured the number of mutants resistant to bacteriophage T1 in a large number of replicate cultures of *E. coli*. If mutants occur after the culture is exposed to the phage, then little variation should occur among cultures in the number of mutants. However, if mutants arise at random during nonselective growth of cells, each culture would contain different number of resistant mutant. The numbers depend on how early during the growth period the first mutant cells arose. But the consequence of that mutation would depend on when during the growth of the population the mutation occurred. Thus a mutation during the early generations gives rise to a large clone of mutant cells, whereas a late mutation gives rise to a few mutant cells. Among a large set of identical cultures of dividing cells, the few cultures in which the mutation happened in the early generations have a large number of mutants, whereas the majority of the cultures have none or a few mutants. This is what Luria and Delbruck observed.

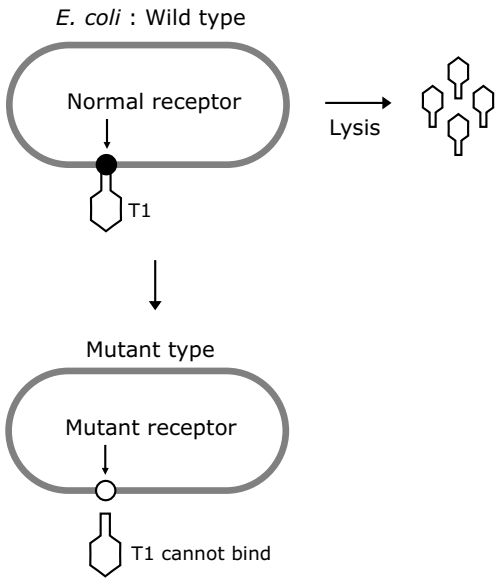


Figure 6.184 When bacteriophage T1 infects wild-type *E. coli*, it binds to a receptor in the outer membrane, protein TonB. After phage replication, the *E. coli* cell is lysed and new phages are released. A mutation in the tonB gene results in an altered receptor to which T1 can no longer bind and so the cells survive.

This page intentionally left blank.

6.25.5 Ames test

The Ames test, named for its developer, Bruce Ames, is a method to test chemicals for their cancer-causing properties. The use of the *Ames test* is based on the assumption that any substance that is mutagenic may also turn out to be a carcinogen; that is, to cause cancer.

The assay is based on the reversion of mutations in the histidine (*his*) operon in the genetically altered tester strains of bacterium *Salmonella typhimurium*. The *his* operon encodes enzymes required for the biosynthesis of the amino acid histidine. Strains with mutations in the *his* operon are histidine auxotrophs — they are unable to grow without added histidine. However, this mutation can be reversed, a back mutation, with the gene regaining its function. These revertants are able to grow on a medium lacking histidine. The tester strains are specially constructed to have both frameshift and point mutations in the genes required to synthesize histidine, which allows for the detection of mutagens acting via different mechanisms. The tester strains also carry mutations in the genes responsible for lipopolysaccharide synthesis, making the cell wall of the bacteria more permeable, and in the excision repair system to make the test more sensitive.

The Ames test can detect mutagens that work directly to alter DNA. In humans, however, many chemicals are promutagens, agents that must be activated to become true mutagens. Activation, involving a chemical modification, often occurs in the liver as a consequence of normal liver activity on unusual substances. Bacteria such as *S. typhimurium* do not produce the enzymes required to activate promutagens, so promutagens would not be detected by the Ames test unless they were first activated. An important part of the Ames test also involves mixing the test compound with enzymes from rat liver that convert promutagens into active mutagens. These potentially activated promutagens are then used in the Ames test. If the liver enzymes convert the agent to a mutagen, the Ames test will detect it, and it will be labeled as a promutagenic agent.

Problem

In the Ames test, auxotropic strains of *Salmonella* that are unable to produce histidine are mixed with a rat liver extract and a suspected mutagen. The cells are then plated on a medium without histidine. The plates are incubated to allow any revertant bacteria (those able to produce histidine) to grow. The number of colonies is a measure of the mutagenicity of the suspected mutagen. Why is the rat liver extract included?

Solution

Most mutagens cannot act unless they are converted to electrophile by liver enzymes called *mixed-function oxidase*, which include the cytochromes P-450s. The rat liver extract in the Ames test contains enzymes for converting suspected mutagens to compounds that would be physiologically relevant mutation-causing agents in a mammal.

6.25.6 Complementation test

If two recessive mutations arise independently and both have the same phenotype, how do we know whether they are both mutations of the same gene? The complementation test allows us to determine whether two mutations, both of which produce a similar phenotype are in the same gene i.e. whether they are alleles or represent mutations in separate genes, whose proteins are involved in the same function. In genetics, complementation occurs when two strains of an organism with different homozygous recessive mutations that produce the same phenotype produce offspring with the wild-type phenotype when mated or crossed. Complementation will occur only if the mutations are in different genes.

In a diploid organism the complementation test of allelism (allelism test) is performed by intercrossing homozygous recessive mutants two at a time and observing whether or not the progeny have a wild-type phenotype. If the two recessive mutations are in separate genes and are not alleles of one another, then following the cross, all F1 progeny are heterozygous for both genes. *Complementation is said to occur.* Because each mutation is in a separate gene and each F1 progeny is heterozygous at both loci, the normal products of both genes are produced. If the two mutations affect the same gene and are alleles of one another. *Complementation does not occur*. Because the two mutations affect the same gene, the F1 is homozygous for the two mutant alleles. No normal product of the gene is produced.

Pages 747 to 761 are not shown in this preview.

Chapter 07

Recombinant DNA technology

Recombinant DNA technology (also known as genetic engineering) is the set of techniques that enable the DNA from different sources to be identified, isolated and recombined so that new characteristics can be introduced into an organism. The invention of recombinant DNA technology—the way in which genetic material from one organism is artificially introduced into the genome of another organism and then replicated and expressed by that other organism—was largely the work of Paul Berg, Herbert W. Boyer, and Stanley N. Cohen, although many other scientists made important contributions to the new technology as well. Paul Berg developed the first recombinant DNA molecules that combined DNA from SV40 virus and lambda phage. Later in 1973, Herbert Boyer and Stanley Cohen develop recombinant DNA technology, showing that genetically engineered DNA molecules may be cloned in foreign cells.

One important aspect in recombinant DNA technology is *DNA cloning.* It is a set of techniques that are used to assemble recombinant DNA molecules and to direct their replication within host organisms. The use of the word cloning refers to the fact that the method involves the replication of a single DNA molecule starting from a single living cell to generate a large population of cells containing identical DNA molecules.

7.1 DNA cloning

DNA cloning is the production of a large number of identical DNA molecules from a single ancestral DNA molecule. The essential characteristic of DNA cloning is that the desired DNA fragments must be *selectively amplified* resulting in a large increase in copy number of selected DNA sequences. In practice, this involves multiple rounds of DNA replication catalyzed by a DNA polymerase acting on one or more types of template DNA molecule. Essentially two different DNA cloning approaches are used: *Cell-based* and *cell-free DNA cloning.*

Cell-based DNA cloning

This was the first form of DNA cloning to be developed, and is an *in vivo* cloning method. The first step in this approach involves attaching foreign DNA fragments *in vitro* to DNA sequences which are capable of independent replication. The recombinant DNA fragments are then transferred into suitable host cells where they can be propagated selectively.

The essence of cell-based DNA cloning involves following steps:

Construction of recombinant DNA molecules

Recombinants are hybrid DNA molecules consisting of autonomously replicating DNA segment plus inserted elements. Such hybrid molecules are also called *chimera*. Recombinant DNA molecules are constructed by *in vitro* covalent attachment (ligation) of the desired DNA fragments (target DNA) to a replicon (any sequence capable of independent DNA replication). This step is facilitated by cutting the target DNA and replicon molecules with specific restriction endonucleases before joining the different DNA fragments using the enzyme DNA ligase.

This page intentionally left blank.

Cell-free DNA cloning

The polymerase chain reaction (PCR) is a newer form of DNA cloning which is enzyme mediated and is conducted entirely *in vitro*. PCR (developed in 1983 by Kary Mullis) is a revolutionary technique used for selective amplification of specific target sequence of nucleic acid by using short primers. It is a rapid, inexpensive and simple method of copying specific DNA sequence.

7.2 Enzymes for DNA manipulation

The enzymes used in the recombinant DNA technology fall into four broad categories:

7.2.1 Template–dependent DNA polymerase

DNA polymerase enzymes that synthesize new polynucleotides complementary to an existing DNA or RNA template are included in this category. Different types of DNA polymerase are used in gene manipulation.

DNA polymerase I (Kornberg enzyme) has both the 3′-5′ and 5′-3′ exonuclease activities and 5′-3′ polymerase activity.

Reverse transcriptase, also known as RNA-directed DNA polymerase, synthesizes DNA from RNA.

Reverse transcriptase was discovered by *Howard Temin* at the University of Wisconsin, and independently by *David Baltimore* at about the same time. The two shared the 1975 Nobel Prize in Physiology or Medicine.

Taq DNA polymerase is a DNA polymerase derived from a thermostable bacterium, *Thermus aquaticus*. It operates at 72°C and is reasonably stable above 90°C and used in PCR. It has a 5′ to 3′ polymerase activity and a 5′ to 3′ exonuclease activity, but it lacks a 3′ to 5′ exonuclease (proofreading) activity.

7.2.2 Nucleases

Nucleases are enzymes that degrade nucleic acids by breaking the phosphodiester bonds that link one nucleotide to the next. Ribonucleases (RNases) attack RNA and deoxyribonucleases (DNases) attack DNA. Some nucleases will only attack single stranded nucleic acids, others will only attack double-stranded nucleic acids and a few will attack either kind. Nuclease are of two different kinds – exonucleases and endonucleases. *Exonucleases* remove nucleotides one at a time from the end of a nucleic acid whereas e*ndonucleases* are able to break internal phosphodiester bonds within a nucleic acid. Any particular exonuclease attacks either the 3′-end or the 5′-end but not both.

Mung bean nuclease

The mung bean nuclease is an endonuclease specific for ssDNA and RNA. It is purified from mung bean sprouts. It digests single-stranded nucleic acids, but will leave intact any region which is double stranded. It requires Zn^{2+} for catalytic activity.

S1 nuclease

The S1 nuclease is an endonuclease purified from *Aspergillus oryzae*. This enzyme degrades RNA or single stranded DNA, but does not degrade dsDNA or RNA-DNA hybrids in native conformation. Thus, its activity is similar to mung bean nuclease, however, the enzyme will also cleave a strand opposite a nick on the complementary strand.

RNase A

RNase A is an endonuclease, which digests ssRNA at the 3′ end of pyrimidine residues.

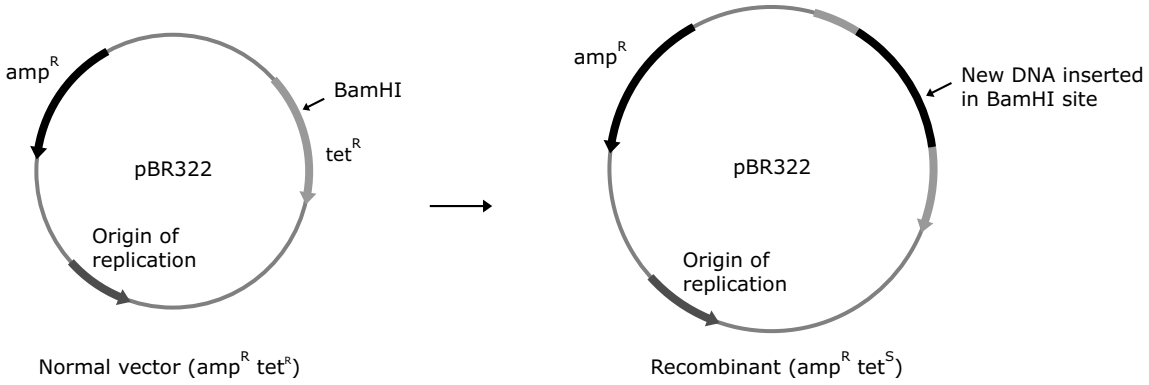
RNase H

It is an endonuclease which digests the RNA strand of an RNA-DNA heteroduplex. The enzyme does not digest ss or dsDNA.

7.7 Recombinant screening

A selective medium enables transformants to be distinguished from non-transformants. The next problem is to determine which of the transformed colonies comprise cells that contain recombinant DNA molecules, and which contain self-ligated vector molecules. With most cloning vectors insertion of a DNA fragment into the plasmid destroys the integrity of one of the genes present on the molecule. Recombinants can, therefore, be identified because the characteristic coded by the inactivated gene is no longer displayed by the host cells (called **insertional inactivation**).

Most commonly recombinant selection is carried out by insertional inactivation of antibiotic resistance gene. In this case the insertion of new DNA fragments (insert) occurs at the site within the gene that confers resistance towards a particular antibiotic.



Normal vector (amp^R tet^R) Recombinant (amp^R tet^S)

Insertional inactivation does not always involve antibiotic resistance genes. For example in pUC8, gene LacZ', which codes for part of enzyme β-galactosidase is used for insertional inactivation. Recombinant pUC8 involves insertional inactivation of the lac Z' gene, can be identified because of their inability to synthesize β-galactosidase. β-galactosidase, coded by lacZ gene, causes the breakdown of lactose to glucose plus galactose. lacZ', a modified lacZ gene, codes for the α peptide portion of β-galactosidase.

7.8 Expression vector

An expression vector contains regulatory elements allowing the expression of any foreign DNA it carries. A foreign gene present on expression vector can be efficiently transcribed and translated by the host cell. The simplest expression vectors, *transcription vectors*, allow transcription, but not a translation of cloned foreign DNA. Typical *protein expression vectors* allow both the transcription and translation of cloned DNA, and thus facilitate the production of recombinant protein. Such vectors are equipped with transcriptional regulatory sequences and sequences that control the RNA processing and protein synthesis.

For transcription, a *promoter site* and a *terminator site* are necessary. Transcription of the desired gene begins at the promoter site and ends at the terminator site. A *ribosome binding site* upstream from the start codon is also present in many of the expression vectors. This site is required for the efficient initiation of translation in bacteria. Promoter is the most critical component of an expression vector since it controls the very first stage of gene expression and also regulates the rate of transcription. An expression vector should carry a *strong promoter* so that the highest possible rate of gene expression could be achieved. Regulation of promoter is another important factor to be considered during construction of an expression vector. Two important ways of regulating a promoter in *E. coil* are:

Induction : Where transcription of a gene is switched on by the addition of a chemical.

Repression : Where gene transcription is switched off upon addition of a regulatory chemical.

Pages 790 to 801 are not shown in this preview.

Automated sequencing

The standard chain termination sequencing methodology employs radioactive labels, and the banding pattern in the polyacrylamide gel is visualized by autoradiography. *Fluorescent primers are the basis of automated sequencing.* The fluorolabel is attached to the ddNTPs, with a different fluorolabel used for each one. Chains terminated with A are therefore labeled with one fluorophore, chains terminated with C are labeled with a second fluorophore, and so on. Now it is possible to carry out the four sequencing reactions - for A, C, G and T - in a single tube and to load all four families of molecules into just one lane of the polyacrylamide gel, because the fluorescent detector can discriminate between the different labels and hence determine if each band represents an A, C, G or T. The sequence can be read directly as the bands pass in front of the detector and either printed out in a form readable by eye or sent straight to a computer for storage.

Genome sequencing

The first genome to be completely sequenced was the genome of bacteriophage ϕX174. Although sequencing can be performed directly on genomic DNA, this is generally impractical on a large scale. Hence genomes have to be split into fragments of a suitable size such that they can be maintained within a vector. Genomic DNA fragments are therefore cloned into a vector and each fragment is subsequently sequenced. The problem then is how to reconstruct the original genome sequence based on the small fragments that are cloned into individual vectors. Two different approaches have been developed for sequence assembly.

- The *clone contig approach* : The simplest way to generate overlapping DNA sequence is to isolate and sequence one clone, from a library, then identify (by hybridization) a second clone, whose insert overlaps with the first. The second clone is then sequenced and the information used to identify a third clone, whose insert overlaps with the second clone, and so on. This is used to build up large continuous DNA sequences (**contigs**) from small fragments cloned into vectors. This method is, however, laborious. A single clone has been isolated and sequenced before the next overlapping clone can be sought. Additionally, repetitive sequences within the genome can give rise to incorrect contig assignment.
- The *shotgun approach* : The fragments of the genome, which have been randomly generated, are cloned into a vector and each insert is sequenced. The sequence is then examined for overlaps (sequences that occur in more than one clone) and the genome is reconstructed by assembling the overlapping sequences together. This approach was first used to sequence the genome of the bacterium *Haemophilus influenzae*. The main advantage of the shotgun approach is that no prior knowledge of the sequence of the genome is required. The approach is, however, limited by the ability to identify overlapping sequences. Every sequence obtained must be compared with every other sequence in order to identify the overlaps.

Table 7.6 Genome sequencing of some model organisms

Genome sequenced	*Year*	*Genome size*	*Comment*
Bacteriophage ϕX174	1977	5.38 kb	First genome sequenced
Plasmid pBR322	1979	4.3 kb	First plasmid sequenced
Bacteriophage λ	1982	48.5 kb	
Yeast chromosome III	1992	315 kb	First chromosome sequenced
Haemophilus influenzae	1995	1.8 Mb	First genome of a cellular organism to be sequenced
Saccharomyces cerevisiae	1996	12 Mb	First eukaryotic genome to be sequenced
Ceanorhabditis elegans	1998	97 Mb	First genome of multicellular organism to be sequenced
Homo sapiens	2000	3000 Mb	First mammalian genome to be sequenced
Arabidopsis thaliana	2000	125 Mb	First plant genome to be sequenced

This page intentionally left blank.

7.12.1 Genetic marker

A gene or DNA sequence having a known location on a chromosome and associated with a particular trait or gene is used as a *genetic marker*. Genes were the first markers to be used to prepare the first genetic maps of fruit fly. There are three major types of genetic markers:

1. Morphological (also classical or visible) markers which are based on phenotypic traits or characters;
2. Biochemical markers, which are based on gene products; and
3. DNA (or molecular) markers, which reveal sites of variation in DNA.

Morphological markers are usually visually characterized phenotypic characters such as flower colour, seed shape, growth habits or pigmentation. Biochemical markers are differences in gene products that are detected by electrophoresis and specific staining. The major disadvantages of morphological and biochemical markers are that they may be limited in number and are influenced by environmental factors or the developmental stages. A *molecular* or *DNA marker* is defined as a particular segment of DNA that is representative of the differences at the genome level. Molecular markers should not be considered as normal genes, as they usually do not have any biological effect, and instead can be thought of as constant landmarks in the genome. They are identifiable DNA sequences, found at specific locations of the genome, and transmitted by the standard laws of inheritance from one generation to the next. An ideal molecular marker should have the following criteria:

1. be polymorphic and evenly distributed throughout the genome,
2. provide adequate resolution of genetic differences,
3. have linkage to distinct phenotypes.

7.12.2 Types of DNA markers

Various types of DNA markers have been described in the literature. They can be broadly divided into two classes based on the method of their detection: Hybridization-based (such as RFLP) and PCR based (such as RAPD, AFLP, SSLP). PCR-based techniques can further be subdivided into two subcategories: arbitrarily primed PCR-based techniques or sequence nonspecific techniques (such as RAPD, AFLP) and sequence targeted PCR-based techniques (such as SSLP, SNP).

DNA markers may be described as *codominant* or *dominant*. This description is based on whether markers can discriminate between homozygotes and heterozygotes. Codominant markers indicate differences in size whereas dominant markers are either present or absent.

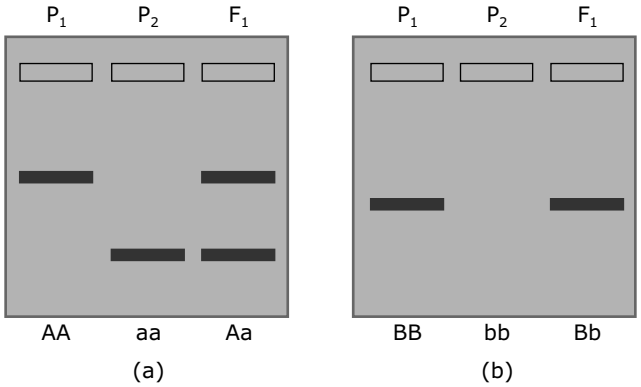


Figure 7.23 Comparison between (a) codominant and (b) dominant markers. Codominant markers can clearly discriminate between homozygotes and heterozygotes whereas dominant markers do not. Genotypes at two marker loci (A and B) are indicated below the gel diagrams.

Pages 805 to 816 are not shown in this preview.

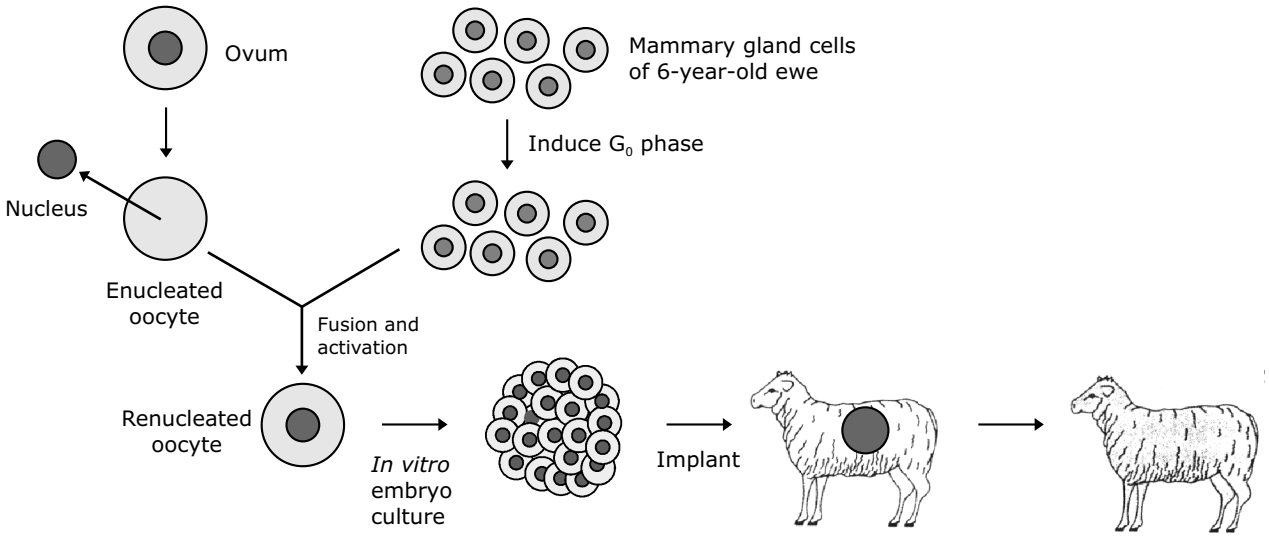


Figure 7.33 Cloning sheep by nuclear transfer. The nucleus of an ovum is removed with a pipette. Cells from the mammary epithelium of an adult are grown in culture, and the G_0 state is induced by inhibiting cell growth. A G_0 cell and an enucleated ovum are fused, and the renucleated ovum is grown in culture or in ligated oviducts until an early embryonic stage before it is implanted into a foster mother, where development proceeds to term.

7.16 Gene therapy

Gene therapy is a technique for correcting defective genes responsible for disease development. Gene therapy typically aims to supplement a defective mutant allele with a functional one. Scientist may use one of several approaches for correcting defective or abnormal genes:

- A normal gene may be inserted into a nonspecific location within the genome (*gene addition*). This is the most common approach.
- An abnormal gene can be replaced by a normal gene through homologous recombination (*gene replacement*).
- An abnormal gene can be repaired through selective reverse mutation, which returns the gene to its normal function.

Gene therapy may be *germ-line* or *somatic cell gene therapy*. Current gene therapy is exclusively *somatic* gene therapy which involves the introduction of genes into somatic cells of an affected individual. Germ-line gene therapy involves the permanent transmissible modification of the genome of a gamete, a zygote or an early embryo. The prospect of human germline gene therapy is currently not sanctioned.

Gene therapy may be c*lassical* and *nonclassical* gene therapy. In *classical* gene therapy genes are delivered to appropriate target cells with the aim of obtaining the optimal expression of the introduced genes. The idea of *nonclassical* gene therapy is to inhibit the expression of genes associated with the pathogenesis, or to correct a genetic defect for restoring the normal gene expression.

Potential use of somatic gene therapy

The potential use of this therapy is to cure genetic diseases. The first case of gene therapy occurred in 1990, at the NIH in Bethesda, Maryland. On that occasion, a four-year-old patient with a severe combined immunodeficiency

Pages 818 to 826 are not shown in this preview.

Secondly, the triggering of sense strand (mRNA) cleavage by incorporating ribozyme catalytic centres into antisense RNA. A number of ribozymes have been characterized, including the most studied form called the hammerhead ribozyme (first isolated from viroid RNA).

Thirdly, RNA interference induced by small interfering RNA molecules. This naturally occurring phenomenon, a potent sequence specific mechanism for post-transcriptional gene silencing, was first described for the nematode worm *Caenorhabditis elegans.*

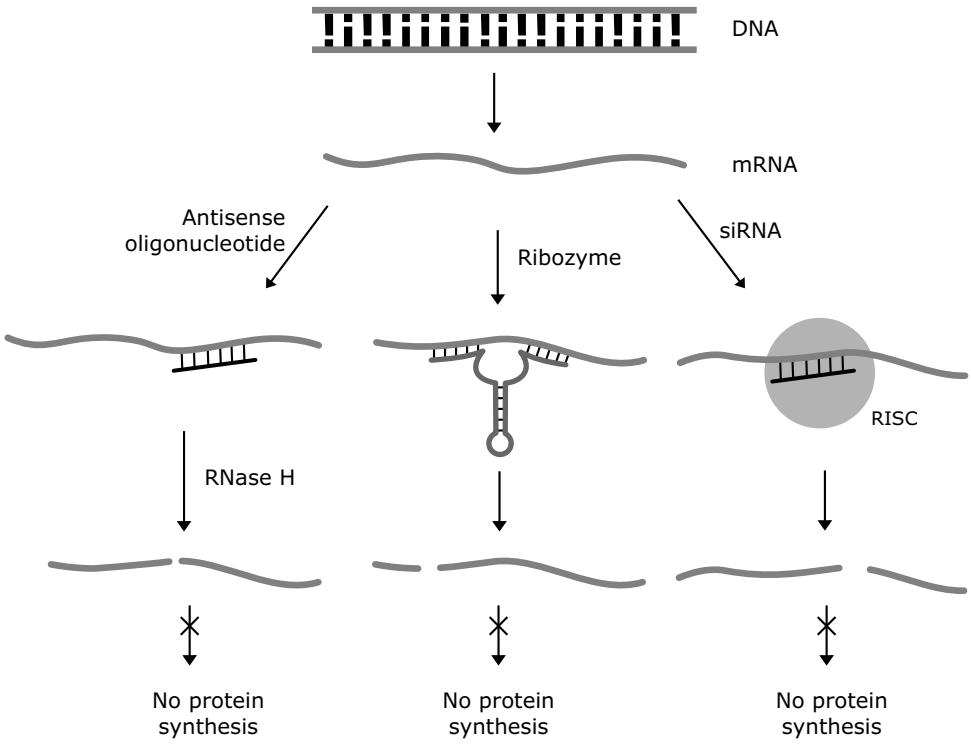


Figure 7.40 Comparison of different antisense strategies. Antisense-oligonucleotides block translation of the mRNA or induce its degradation by RNase H, while ribozymes possess catalytic activity and cleave their target RNA. RNA interference approaches are performed with siRNA molecules that are bound by the RISC and induce degradation of the target mRNA.

7.17.3 Molecular farming

It is an application of genetic engineering in which genes, primarily of human or animal origin are introduced into plants or farm animals for cost effective production of therapeutic products such as antibodies, blood products, cytokines, growth factors, hormones, recombinant enzymes and human and veterinary vaccines. Therapeutic compounds so produced are also known as biopharmaceuticals (pharmaceuticals from biological organisms).

The organisms in which genes coding for the target therapeutically active compound introduced are often referred to as expression system. Expression system studied so far include bacteria, yeast, plant viruses, animal cell culture, transgenic plants and transgenic animals. Initially bacteria were the most widely used expression systems but due to the complexity of the most therapeutic proteins to be produced and simplicity of the bacterial system, new expression systems were explored. As of now the plants are the preferred and most widely used expression system in comparison to other systems. The first recombinant pharmaceutical protein produced in the plant was human serum albumin, first produced in 1990 in transgenic tobacco and potato plants.

Table 7.10 Examples of some pharmaceutical recombinant human proteins expressed in plant systems

Tobacco, sunflower (plants)	Growth hormone
Tobacco, potato (plants)	Serum albumin
Tobacco (plants)	Epidermal growth factor
Rice (plants)	Alpha-interferon
Tobacco (cell culture)	Erythropoietin
Tobacco (plants)	Haemoglobin
Tobacco (cell culture)	Interleukins-2 and 4
Tobacco (root culture)	Placental alkaline phosphatase

7.18 Plant tissue culture

The field of plant tissue culture is based on the fact that plants can be separated into their component parts (organs, tissues or cells), which can be manipulated *in vitro* and then grown back into complete plants. Plant cells or tissues will continue to grow if supplied with the appropriate nutrients and conditions. The culture of plant cells, tissues and organs such as roots, shoot tips and leaves in artificial nutrient media aseptically is referred to as *plant tissue culture*.

Plant cells - *unique features*

A plant cell is a eukaryotic cell and shares similar features with the typical eukaryote cell. However some features are uniquely present in plant cells. Their distinctive features include:

- A cell wall outside the cell membrane which is composed of cellulose, hemicellulose, pectin and in many cases lignin.
- A large central vacuole enclosed by a membrane known as the *tonoplast which* maintains the cell's turgor, controls movement of molecules between the cytosol and sap, stores useful material and digests waste proteins and organelles.
- Specialized cell-cell communication through plasmodesmata, pores in the primary cell wall through which the plasmalemma and endoplasmic reticulum of adjacent cells are continuous.
- Plastids such as chloroplasts which contain chlorophyll for photosynthesis, amyloplasts for starch storage, elaioplasts for fat storage and chromoplasts for the synthesis and storage of pigments.
- A specialized peroxisome called glyoxysome for the operation of glyoxylate cycle.
- Cytokinesis by formation of a phragmoplast and cell plates.
- Absence of centrioles in MTOC that are present in animal cells.

7.18.1 Cellular totipotency

Totipotency is the ability of a single cell to divide and produce all the differentiated cells in an organism. In a multicellular organism, a cell after regulated division undergoes for cell differentiation. It is a process of specializing cells' functions. Isolated cells from differentiated tissues are generally non-dividing and quiescent; to express totipotency the differentiation process has to be reversed (called *de-differentiation*) and repeated again (called *re-differentiation*). A differentiated cell reverting to an undifferentiated state is termed *dedifferentiation*, whereas the ability of a dedifferentiated cell to form a whole organism or organs is termed *redifferentiation*. Theoretically, all living cells can revert to an undifferential status through this process. However, the more differentiated a cell has been, the more difficult it will be to induce its de-differentiation. In plants, even highly mature or differentiated cells have the ability to regress to a meristematic state as long as they are viable and express totipotency. This phenomenon of totipotency is an amazing developmental plasticity that sets plant cells apart from most of their animal counterparts. In animals the differentiation is irreversible.

Pages 829 to 838 are not shown in this preview.

7.19 Animal cell culture

Cells in animals exist in an organized tissue matrix which require for their controlled growth and differentiation. These cells from intact organisms may be isolated, maintained and grown *in vitro* in culture media aseptically containing a suitable mixture of nutrients and growth factors. This process is called *animal cell culture.*

7.19.1 Primary and secondary cultures

A *primary cell culture* is prepared by inoculating cells directly from tissues of an organism into culture media (that is, without cell proliferation *in vitro*). With the exception of some cells derived from tumors, most primary cell cultures have a limited lifespan. After a certain number of divisions, cells undergo the process of senescence and stop dividing. In these cells, the limited proliferation capacity reflects a progressive shortening of the cell's telomeres, the repetitive DNA sequences and associated proteins that cap the ends of each chromosome.

The primary cell culture is of two types depending on the kind of cells in culture – attachment culture and suspension culture. *Attachment culture* involves the adherent or anchorage dependent cells. To survive and grow, most cells require a surface to which they can attach, thus they are anchorage dependent. Without the surface attachment, these cells cannot survive. These adherent cells are usually derived from tissues of organs such as kidney, where they are immobile and embedded in connective tissue. *Suspension culture* involves non-adherent or *anchorage independent* cells which do not require attachment for growth or do not attach to the surface of the culture vessels. Lymphocytes are anchorage independent cells commonly grown in culture.

A *secondary culture* is prepared by subculturing a primary culture. *Subculture* (or *passage*) refers to the transfer of cells from one culture vessel to another. In most cases, cells in primary cultures can be removed from the culture dish and made to proliferate to form a large number of secondary cultures.

7.19.2 Cell line

When a primary cell culture is subcultured, it becomes a *cell line*. The cell lines may be *finite cell line* or *infinite cell line*. A *finite cell line* (or normal cell line) is a line of cells that will undergo only a finite number of divisions in cell culture and eventually undergoes senescence. It has a limited number of possible subcultures or passages. Normal mammalian cells generally have a finite life span in culture; that is, after a number of divisions characteristic of the species and cell type, the cells stop dividing. These cell lines exhibit the property of contact inhibition, density limitation and anchorage dependence.

A cell line that has the potential to be subcultured indefinitely is termed *infinite* (immortal or continuous) cell line. Tumor cells or normal cells that have undergone transformation induced by chemical carcinogens or viruses can be propagated indefinitely in tissue culture; thus, have unlimited number of possible subcultures.

Infinite cell lines are also known as *transformed* cell lines due to altered growth properties of immortalized cells. Transformed cells do not necessarily mean cancer or tumor cells. Transformed cell lines do *not* exhibit the property of contact inhibition, density-dependent inhibition of proliferation and anchorage dependence. They have a reduced requirement for serum or growth factors for optimal growth. A transformed cell line often has an abnormal chromosome number (aneuploid) and overproduces different proteins. Cancer cells are naturally immortal. Thus all cancerous cell lines are transformed, although it is not clear whether all transformed cell lines are cancerous.

The first cell line—the mouse fibroblast L cell—was derived from cultured mouse subcutaneous connective tissue by exposing the cultured cells to a chemical carcinogen. Another important cell line, the HeLa cell, was derived from a 31-year-old black woman named Henrietta Lacks, who died of cervical cancer in 1951. Since these early cell lines, hundreds of cell lines have been established.

This page intentionally left blank.

defined quantities of purified growth factors, lipoproteins and other proteins usually provided by the serum or extract supplement. Since the components (and concentrations thereof) in such culture media are precisely known, these media are generally referred to as *defined culture media* and often as *serum-free media* (SFM). A number of SFM formulations are commercially available, such as those designed to support the culture of endothelial cells, monocytes/macrophages, fibroblasts, neurons, lymphocytes, chondrocytes or hepatocytes.

Some extremely simple SFM, which consist essentially of vitamins, amino acids, organic and inorganic salts and buffers have been used for cell culture. Such media (often called *basal media*), however, are usually, seriously deficient in the nutritional content required by most animal cells. Accordingly, most SFM incorporate additional components into the basal media to make the media more nutritionally complex, while maintaining the serum-free and low protein content of the media. Examples of such components include serum albumin from bovine (BSA) or human (HSA), animal-derived lipids such as human excyte, sterols, etc., and certain growth factors or hormones derived from natural (animal) or recombinant sources.

7.19.4 Growth pattern

Animal cells growth in culture have a characteristic growth pattern similar to bacteria. The cell growth is typically divided into three phases: Lag phase, Log phase and Plateau phase.

Lag phase

The lag phase is a period of zero growth when cells are first inoculated into the growth medium. The length of this phase depends on the type of cells and their metabolic state at inoculation. It is a period of adaptation during which the cell replaces elements of the glycocalyx lost during trypsinization, attaches to the substrate and spreads out.

Log phase

The exponential growth phase is a period of continuous cell doubling. Animal cells normally exhibit a doubling time of between 15 and 25 hours. The length of the log phase depends on the seeding density, the growth rate of the cells and the density at which cell proliferation is inhibited by density.

Plateau (*or stationary*) *phase*

The *stationary phase* is a period after growth when there is no change in the culture cell density. The phase occurs when the nutrients have been depleted or inhibitory metabolites have accumulated in the culture. All the available growth surface is occupied and all the cells are in contact with surrounding cells. Further growth of cells can be obtained by subculturing the cells in fresh medium.

7.19.5 Application of animal cell culture

Cell culture has become one of the major tools used in cell and molecular biology. Some of the important areas where cell culture is currently playing a major role are briefly described below:

Model systems

Cell cultures provide a good model system for studying 1) basic cell biology and biochemistry, 2) the interactions between disease-causing agents and cells, 3) the effects of drugs on cells, 4) the process and triggers for ageing and 5) nutritional studies.

Toxicity testing

Cultured cells are widely used alone or in conjunction with animal tests to study the effects of new drugs, cosmetics and chemicals on survival and growth in a wide variety of cell types. Especially important are liver- and kidney-derived cell cultures.

Cancer research

Since both normal cells and cancer cells can be grown in culture, the basic differences between them can be closely studied. In addition, it is possible, by the use of chemicals, viruses and radiation, to convert normal cultured cells

Pages 842 to 846 are not shown in this preview.

Chapter 08

Bioprocess engineering

Bioprocess engineering is a specialization of chemical engineering that deals with the design and development of equipment and processes for the manufacturing of products such as food, pharmaceuticals and polymers from biological materials. It uses the capabilities of organisms in industrial, medical, environmental or agricultural processes in order to produce useful biological materials.

Application of bioprocess engineering includes:

- Design and operation of fermentation systems,
- Development of food processing systems,
- Application and testing of product separation technologies,
- Design of instrumentation to monitor and
- Control biological processes and many more.

Bioprocess engineers work at the frontiers of biological and engineering sciences to *Bring engineering to Life* through the conversion of biological materials into other forms needed by mankind. One of the main tasks of a bioprocess engineer is control and maintenance of a biological processes such as the production of beverages, pharmaceuticals, antibiotics, enzymes, biochemicals, enzyme-catalyzed reactions, food processing and biological waste treatment. These processes require a well-designed growth environment to obtain the maximum yield of the product and consequently these conditions need to be carefully controlled. Environmental design comprises the determination of the environment of the process, while fermentation engineering provides the means for meeting those requirements.

8.1 Concept of material and energy balance

8.1.1 Material balance

Material balances (mass balances) are based on the *law of conservation of mass*. The law of 'conservation of mass' states that mass cannot be created or destroyed. In performing the material balance we apply thermodynamic terms – *system* and *process.* A *system* is defined as that part of the universe that is under consideration. All space outside the system is known as the *surroundings.* A system is separated from the surrounding by a *system boundary*, which may be real or imaginary. If the boundary doesn't allow mass to pass from system to surroundings and vice versa, the system is considered as a *closed system* with constant mass. If the system boundary allows the mass to pass from system to surroundings and vice versa, then it is an *open system*. A *process* causes changes in the system or surroundings. In bioprocess, the process can be batch, fed batch and continuous processes. A *batch process* operates in a closed system. All materials are added to the system at the start of the process, the system is then closed and products removed only when the process is complete. A *fed-batch* process allows input of material to the system but not output and a *continuous* process allows matter to flow in and out of the system.

A cell or a bioreactor is defined as a system, and mass balance is performed on the system. Doing a mass balance is similar in principle to accounting. In accounting, accountants do balances of what happens to a company's money. In the process of mass balance, the first step is to look at the three basic categories: mass in, mass out and mass stored. The mass can be total mass, the mass of a particular molecular or atomic species, or biomass. Total mass balance for the system can be written in a general way:

Mass *in* (input) = Mass *stored* (accumulation) + Mass *out* (output)

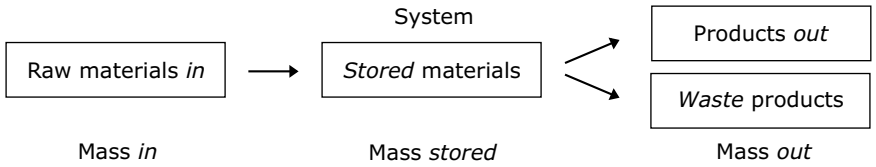


Figure 8.1 Mass balance.

Bioprocess engineers do a mass balance to account for what happens to each of the chemicals that is used in a chemical process. For example, in a plant that is producing sugar, if the total quantity of sugar going into the plant is not equalled by the total of the purified sugar and the sugar in the waste liquors, then there is something wrong. Sugar is either being burned (chemically changed) or accumulating in the plant or else it is going unnoticed down the drain somewhere. In this case the mass balance is;

Raw materials = Products + Waste products + Stored products + Losses

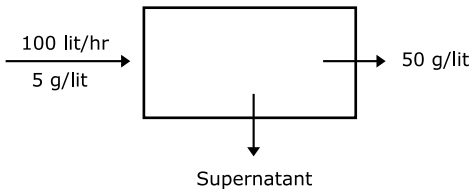
Mass balances can be based on total mass, mass of dry solids or mass of particular components, for example protein. If a mass balance is written using the total mass in each process stream, then it is called *total balance.* A separate mass balance can be written for a particular chemical component in the total mass. This is called *component balance.* Thus, for a component mass balance the simplest expression is:

Input – Output + Formation – Disappearance = Accumulation

Problem

In a filtration device, the input concentration of the cell is 5g/litre which is pumped in at 100 litre/hr. The desired output concentration is 50 g/litre. The system runs continuously so there is no accumulation. Calculate the rate of removal of the permeate supernatant.

Solution



In this case, the system runs continuously and there is no reaction, hence

Input = Output

Input cells = 5 gram/litre × 100 litre/hr = 500 gram/hr

If 'Y' litres is the output of concentrated cells then,

Output cells = 50 gram/litre × Y litre/hr = 50 Y gram/hr

So, 500 = 50 Y

Y = 10 litre/hr

Hence, supernatant volume = 100 – 10 = 90 litre. Thus, supernatant has to be removed at a rate of 90 litre/hr.

Pages 849 to 852 are not shown in this preview.

Concept of degree of reduction

The degree of reduction, γ, is defined as the number of equivalents of available electrons in the quantity of material containing 1 g atom (1 mole) carbon. Therefore, for substrate $C_wH_xO_yN_z$, the number of available electrons is (4w + x – 2y – 3z). The degree of reduction for the substrate, γ_s, is therefore (4w + x – 2y – 3z)/w.

The degree of reduction relative to CO_2, H_2O and NH_3 is zero.

Yield coefficient

The *yield coefficient* is the ratio of the amount produced to the amount consumed for any product/reactant pair. It is a ratio having no unit. A yield coefficient is often used to describe the conversion efficiency. Virtually any pair of output and input can be combined to give a yield coefficient. For example, the *yield coefficient of biomass,* $Y_{X/S}$, is the biomass of cells formed per unit of substrate consumed for biosynthesis. Yield coefficient can be related to ATP consumption. The ATP yield coefficient, $Y_{X/ATP}$, represents the amount of biomass synthesized per mole of ATP consumed.

Cell biomass and product formation can be described quantitatively by yield coefficients. Let's consider the overall stoichiometric equation for growth and production:

$$sS + nN + oO_2 \longrightarrow X + pP + wH_2O + eCO_2$$

where S, carbon source; N, nitrogen source; X, biomass; P, product and s, n, o, p, w, e are stoichiometric coefficients. The theoretical yield coefficients can be determined from the above stoichiometry with a known chemical formula for S, N, X and P. The cell biomass yield coefficient and the product yield coefficient are

$$Y_{X/S} = M_X / sM_S \quad \text{and} \quad Y_{P/S} = pM_P / sM_S$$

respectively, where M_x, M_p and M_s are the molecular weights of cell biomass, product and carbon source.

8.1.2 Energy balance

Energy balances are used to quantify the energy used or produced by a system. In bioprocessing, energy accounting system can be set up to determine the amount of steam or cooling water required to maintain optimum process temperatures. The principle underlying all energy-balance calculations is the *law of conservation of energy*, which states that the energy can be neither created nor destroyed. Although this law does not apply to nuclear reaction, conservation of energy remains a valid principle for bioprocesses.

The law of conservation of energy can be represented as:

(Energy in through system boundaries) – (Energy out through system boundaries) = (Energy accumulated within the system).

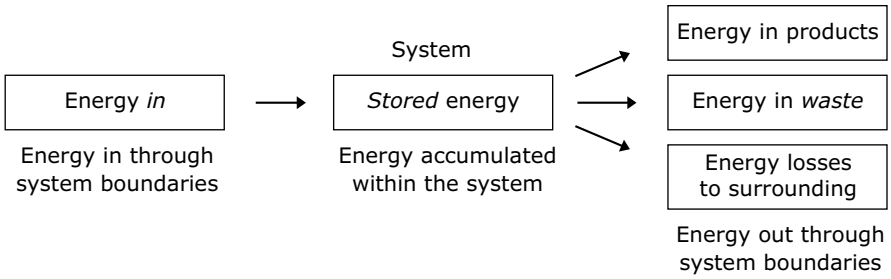


Figure 8.2 Energy balance.

Energy takes many forms, such as heat, kinetic energy, chemical energy and potential energy but because of interconversions it is not always easy to isolate separate constituents of energy balances. However, under some circumstances certain aspects predominate. In many heat balances in which other forms of energy are insignificant; in some chemical situations mechanical energy is insignificant.

This page intentionally left blank.

Phase change

When a substance changes from one phase of matter to another, we say that it has undergone a change of phase. These changes of phase always occur with a change of heat. Heat either comes into the material during a change of phase or heat comes out of the material during this change. However, the heat content of the material changes, the temperature does not. The amount of energy released or absorbed during a phase change is called latent heat. The latent heat for a different mass of the substance can be calculated using the equation:

$$Q = ML$$

where, Q = is the amount of energy released or absorbed during the change of phase of the substance

M = is the mass of the substance and

L = is the *specific latent heat* per gram for a particular substance ; substituted as L_f to represent the specific latent heat of fusion, L_v as the specific latent heat of vaporization.

Enthalpy change due to mixing

When compounds are mixed or dissolved, the bonds between molecules in the solvent and solute are broken and reformed so a net absorption or release of energy takes place due to which internal energy and enthaply of mixture change. The enthaply change during mixing of non-ideal two compounds A and B is given by

$$\Delta H_{mixing} = H_{mixture} - (H_A + H_B)$$

For an ideal solution or ideal mixture, $\Delta H_{mixing} = 0$

Enthalpy change due to reaction

Bioprocessing involves enzyme catalyzed reactions. During the reaction, relatively large changes in internal energy and enthalpy occur. Enthalpy of a reaction is the amount of heat released or absorbed during the reaction and equal to the difference in enthalpy of reactants and products. In case of an exothermic reaction, the enthalpy of a reaction is negative. On the other hand, enthalpy of a reaction is positive for an endothermic reaction.

$$\Delta H_{reaction} = H_{product} - H_{reactant}$$

8.2 Microbial growth kinetics

Microbial growth is a result of both cell division and change in cell size. Microorganisms can grow under a variety of physical, chemical and nutritional conditions. In a suitable nutrient medium, organisms extract nutrients from the medium and convert them into biological compounds. Part of these nutrients are used for energy production and part are used for biosynthesis and product formation. Thus microbial growth is an orderly increase in the quantity of cellular constituents (i.e. cell mass) and number. It depends on the ability of the cell to form new protoplasm from nutrients available in the environment. Microbial growth is a good example of an autocatalytic reaction.

Microbial batch growth

When a microbe (such as bacterial cell) is inoculated into a flask containing fresh culture medium and incubated, it enters into a rapid growth phase during which the microbe divides and increases its population in the flask medium. Since the microbes are not transferred to a new medium or no fresh nutrients are added to the medium, the increasing population of microbial cells, after sometime, enters into a stationary-phase with the exhaustion of the required nutrients and the accumulation of inhibitory end products in the medium. Eventually, the stationary phase of microbial population culminates into death-phase when the viable microbial cells begin to die. A batch culture can be considered to be a *closed system*.

Consequently, the residual substrate concentration in the reactor is controlled by the dilution. Any alteration to this dilution rate results in a change in the growth rate of the cells that will be dependent on substrate availability at the new dilution rate. Thus, growth is controlled by the availability of a rate-limiting nutrient. This system, where the concentration of the rate-limiting nutrient entering the system is fixed, is often described as a **chemostat** as opposed to operation as a **turbidostat**, where nutrients in the medium are not limiting. In turbidostat, turbidity of the culture is monitored and maintained at a constant value by regulating the dilution rate, i.e. cell concentration is held constant.

The concentration of biomass or microbial metabolites in a continuous fermenter under steady-state conditions can be related to the yield coefficient as described in the batch fermentation section. Inserting the equation for residual substrate into the biomass or a metabolic product yield coefficient equation gives, in this case, for steady-state biomass ($\bar{x}$),

$$\bar{x} = Y_{X/S}\left(S_R - \frac{DK_s}{\mu_{max} - D}\right)$$

where S_R is the substrate concentration of inflowing medium or

$$\bar{x} = Y_{X/S}(S_R - S_r)$$

Therefore, the biomass concentration under steady-state conditions is controlled by the substrate concentration of inflowing medium and the operating dilution rate. Under non-inhibitory conditions, where there is no substrate or product inhibition, the higher the feed concentration the greater the biomass concentration and residual substrate concentration remains constant. However, the higher the dilution rate, the faster the cells grow, which results in a simultaneous increase in the residual substrate concentration and a consequent reduction in the steady-state biomass concentration. As D approaches μ_{max}, the biomass concentration becomes even lower, yet the cells grow faster and there is a concurrent increase in the residual substrate concentration.

8.3 Fermentation

Fermentation (derived from the Latin verb *fervere*, to boil) is the production of carbon dioxide by the anaerobic catabolism of the sugars. The term fermentation has been used in a strict biochemical sense which mean an energy-generation process in which organic compounds act as both electron donors and terminal electron acceptors. However, industrial microbiologists have extended the term fermentation to describe any process for the production of the product by the mass culture of a microorganism.

8.3.1 Fermentation processes

On the commercial scale, there are five major groups of fermentation processes:

1. Produce microbial cells (or biomass) as a product.

   Bakers' yeast, used in the baking industry, is an example of a produced cell mass. Others include single-cell proteins for food sources.
2. Produce microbial enzymes.
3. Produce microbial metabolites.
4. Produce recombinant products.
5. Processes that modify a compound that is added to the fermentation process are referred to as *biotransformations*. Biotransformations occur using the inherent enzymatic capability of most cells. Cells of all types can be employed to biocatalyze a transformation of certain compounds via dehydration, oxidation, hydroxylation, amination or isomerization.

This page intentionally left blank.

For increased documentation and reproducibility in the fermentation industry, there is a trend towards application of *defined media* with a carbon and energy source (glucose, sucrose or starch), an inorganic nitrogen source (ammonia or urea), a mixture of minerals and perhaps a few vitamins.

8.4 Bioreactor

A *bioreactor* (in biochemical engineering, we also use terms like biochemical reactor, biological reactor, fermenter or microbial reactor, which are all synonymous) is a vessel in which the growth and metabolism of cells take place. What makes the bioreactors different from the chemical reactors is the presence of the living organisms. Bioreactors are commonly cylindrical, ranging in size from some liter to cubic meters and are often made of stainless steel. The process in *bioreactor* can either be aerobic or anaerobic. The term *bioreactor* is often used synonymously with *fermenter*; however, in the strict sense, a fermenter is a system in which anaerobic process is carried out.

Bioreactor design

Bioreactor design is a relatively complex engineering task. The goal of an effective bioreactor is to control, contain and positively influence the biological reaction. Suitable bioreactor design criteria include:

- Microbiological and biochemical characteristics of the cell systems (microbial, mammalian, plant cell).
- Hydrodynamic characteristics of the bioreactor.
- Mass and heat characteristics of the bioreactor.
- Kinetics of cell growth and product formation.
- Genetic stability characteristics of the cell system.
- Sterilization and maintenance of sterility.
- Agitation (for mixing of cells and medium) and aeration (aerobic fermenters; for O_2 supply).
- Process monitoring and control (regulation of factors like temperature, pH, pressure, aeration, nutrient).
- Implication of bioreactor design on downstream product separation.
- Capital and operating costs of the bioreactor.
- Potential for bioreactor scale-up.

In addition to controlling these, the bioreactor must be designed to both promote formation of the optimal morphology of the organism and eliminate or reduce contaminations by unwanted organisms or mutation of the organisms. There are a wide variety of bioreaction systems, and any attempt to categorize them by their various attributes will naturally result in some overlap of system characteristics.

8.4.1 Agitation and aeration

Agitation

Mixing is one of the most important operations in bioprocessing. Within a fermenter, there is a need to mix three different phases:

- Liquid phase, which contains dissolved nutrients and metabolites.
- Gaseous phase, which is predominantly oxygen and CO_2.
- Solid phase, which is made up of the cells and any solid substance that may be present.

Purpose of mixing

- Air bubble dispersion;
- Mass transfer from air bubbles (i.e. oxygen supply) to the liquid and then to the cells;
- Supply of the nutrient components to cells (more precisely, cell agglomerates);

Pages 865 to 880 are not shown in this preview.

During bioprocessing, it may reduce the overall rate of synthesis of plasmid-encoded products in bioreactor. Plasmid instability occurs as a result of DNA mutation or defective plasmid segregation. For segregational stability, the total number of plasmids present in the culture must double once per generation, and the plasmid copies must be equally distributed between mother and daughter cells.

A simple model has been developed for batch culture to describe changes in the fraction of plasmid-bearing cells as a function of time. The important parameters in this model are the probability of plasmid loss per generation of cells, and the difference in the growth rates of plasmid-bearing and plasmid-free cells. If x^+ is the concentration of plasmid-carrying cells and x^- is the concentration of plasmid-free cells, the rates at which the two cell populations grow are:

$r_X{}^+ = (1 - p)\mu^+ x^+$ and

$r_X{}^- = p\,\mu^+ x^+ + \mu^- x^-$

where, $r_X{}^+$ is the rate of growth of the plasmid-bearing population,

$r_X{}^-$ is the rate of growth of the plasmid-free population,

p is the probability of plasmid loss per cell division ($p \leq 1$),

μ^+ is the specific growth rate of plasmid-carrying cells, and

μ^- is the specific growth rate of plasmid free cells.

This model is based on the following assumptions:

1. Exponential growth of the host cells
2. All plasmid-containing cells are identical in growth rate and probability of plasmid loss:
3. All plasmid-containing cells have the same copy number.

8.8 Mass and Heat transfer

8.8.1 Mass transfer

Mass is transferred from one location to another under the influence of a concentration gradient in the system. Mass transfer takes place by two basic processes: diffusion and convection. In bioreactions, the transport of nutrients to the cell surface and the removal of metabolites from the cell surface to the bulk of the medium are rate processes with time constants. The driving forces for mass transfer are concentration, temperature or pressure gradients.

An example of mass transfer is the supply of oxygen in fermenters for aerobic culture. The concentration of oxygen at the surface of air bubbles is high compared with the rest of the fluid; this concentration gradient promotes oxygen transfer from the bubbles into the medium.

Diffusion

Diffusion is the movement of component molecules in a mixture under the influence of a concentration difference in the system. In single-phase systems, the rate of mass transfer due to molecular diffusion is given by *Fick's law of diffusion,* which states that mass flux is proportional to the concentration gradient.

$$J_A = -D_{AB}\frac{dC_A}{dx}$$

where, J_A = the mass flux of component A,

C_A = the concentration of component A,

X = distance,

$\frac{dC_A}{dx}$ = the concentration gradient, or change in concentration of A with distance,

D_{AB} = the *binary diffusion coefficient* or diffusivity of component A in a mixture of A and B.

Pages 882 to 894 are not shown in this preview.

Due to the typical fragility of the engineered microorganisms, large-scale fermentation vessels must be designed with the ability to:

- Remove the heat buildup that results from metabolic processes;
- Manage agitation and mixing with minimal shear damage;
- Effectively control the highly variable liquid flow rates and turndowns that are associated with batch fermentation;
- Execute safeguards and sterilization techniques to guard against potential contamination.

8.12 Downstream processing

Bioprocessing treats raw materials and generates useful products. A problem common to all biological processes, whether based on fermentation or cell culture technology, is the need to recover the product. Fermentation broths are complex, aqueous mixtures of cells, comprising the soluble extracellular, intracellular products and any unconverted substrates. The fermentation broth has to be processed and passed through several stages of separation and purification.

In the case of protein production especially human therapeutic proteins, these products must be recovered in a highly purified form, with the molecule in its proper 3-D configuration. The need for extremes in purity, retention of molecular configuration and efficiency in recovery are major challenges. *Downstream processing* refers to the recovery and purification of biosynthetic products. The problem of recovery depends very much on the type of cell and how the bioreactor is designed and operated. The selection of appropriate purification step depends on the nature of the end product, its concentration, the side product present, the stability of the biological materials and necessary degree of purification.

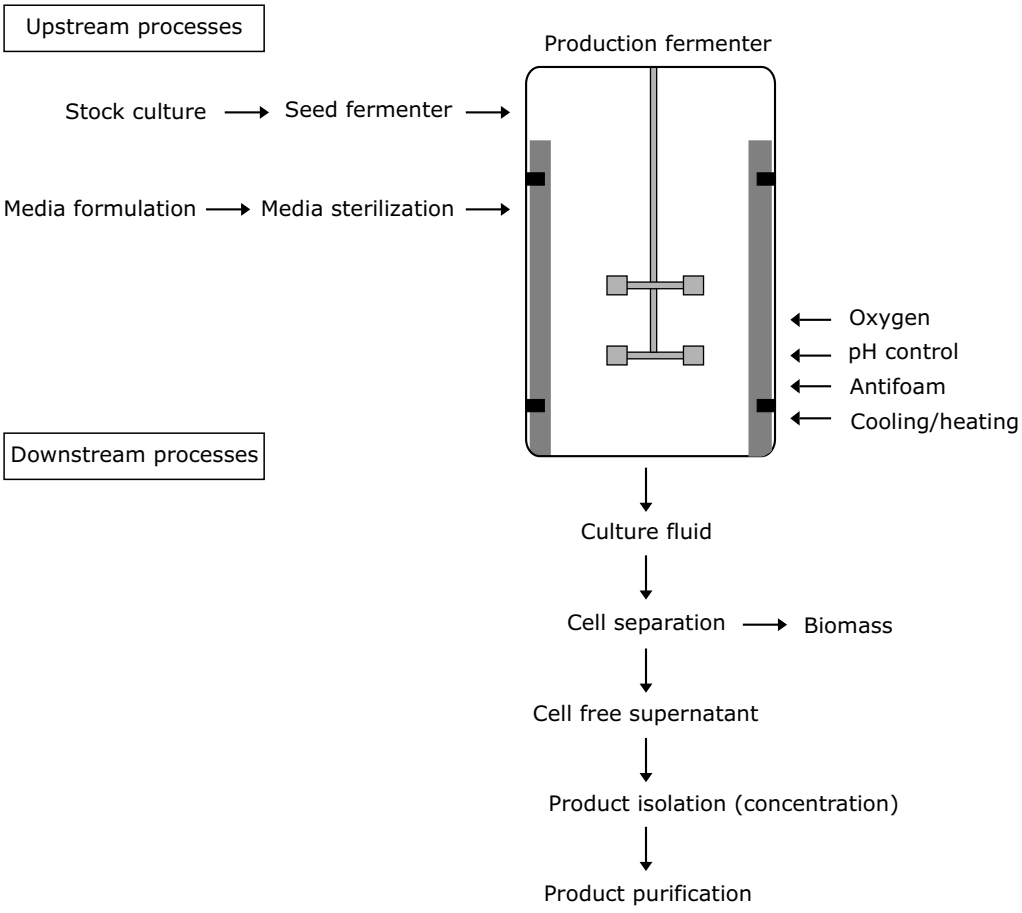


Figure 8.14 An outline of upstream and downstream processing operations.

Pages 896 to 902 are not shown in this preview.

Lactic acid	*Lactobacillus delbrueckii* (bacteria)
Propionic acid	*Propionibacterium* (bacteria)
Malic acid	*Leuconostoc brevis* (bacteria)
Penicillins	*Penicillium chrysogenum* (fungi)
Cephalosporins	*Acremonium chrysogenum* (fungi)
Bacitracin	*Bacillus licheniformis* (bacteria)
Gramicidin	*Bacillus brevis* (bacteria)
B_{12} (Cyanocobalamin)	*Pseudomonas denitrificans* (bacteria)
β-Carotene (Provitamin A)	*Blakeslea trispora* (fungi)
Ascorbic acid (vitamin C)	*Acetobacter suboxydans* (bacteria)
Alginate	*Azotobacter vinelandii* (bacteria)
Cellulose	*Acetobacter xylinum* (bacteria)
Dextran	*Leuconostoc mesenteroides* (bacteria)
Pullulan	*Aureobasidium pullulans* (fungi)
Xanthan	*Xanthomonas campestris* (bacteria)

8.14 Wastewater treatment

Waste materials generated in a society can be classified into three major categories: industrial wastes, domestic wastes and agricultural wastes. Each of these waste materials has its own characteristics, and thus treatment methods vary. Three major waste treatment methods are the following:

1. *Physical treatment*
   Physical treatment includes screening, flocculation, sedimentation and filtration, which are usually used for the removal of insoluble materials.
2. *Chemical treatment*
   Chemical treatment includes chemical oxidations and chemical precipitation.
3. *Biological treatment*
   Biological treatment includes the aerobic and anaerobic treatment of wastewater by a mixed culture of microorganisms.

Quantification of biodegradable material in wastewater

The biodegradable materials of a waste-water sample can be expressed in two ways: biological oxygen demand (BOD) and chemical oxygen demand (COD). The BOD test estimates the amount of oxygen required by aerobic microorganisms to oxidize biodegradable materials in the wastewater- over a fixed period of time (normally 5 days), at constant temperature (20°C) in the dark. A wastewater sample is saturated with oxygen and seeded with an inoculum containing a diverse range of microbes. Its oxygen concentration is measured before and after a 5 days incubation period and the results are expressed as milligrams of oxygen per litre of waste.

COD determines the amount of oxygen required to chemically oxidize any oxidizable organic material present in a waste water. Organic compounds are oxidized by a strong chemical oxidant, and using the reaction stoichiometry, the organic content is calculated. This test involves the addition of a known volume of sample to a mixture of oxygen-rich potassium dichromate and concentrated sulphuric acid. Almost all organic compounds present in waste water are oxidized by strong chemical oxidants. Therefore, the COD content of a waste-water sample usually exceeds the measured BOD (COD > BOD). The BOD:COD ratios for sewage are normally between 0.2:1 and 0.5:1.

This page intentionally left blank.

Oxidation ponds

Oxidation ponds are large, shallow ponds, typically 1-2 m deep. It acts as a shallow waste-treatment reactor where raw or partially treated sewage is decomposed by microorganisms. The conditions are similar to eutrophic lake. The ponds can be designed to maintain aerobic conditions. Oxidation ponds are also used to augment secondary treatment, in which case they are often called *polishing ponds*.

Advanced wastewater treatment

Advanced wastewater treatments are designed for the purpose of removing nitrogen and phosphorus. Nitrogen containing organic compounds are first oxidized biologically to ammonium ions which is further oxidized to nitrite and nitrate by *genera nitrosomonas* and *nitrobacter,* respectively. The second phase is anaerobic *denitrification* which releases nitrogen gas. A number of bacteria can act as denitrifiers such as *Pseudomonas, Alcaligenes, Arthrobacter.*

Phosphorus in wastewater exists in many forms but all of it ends up as orthophosphate. Removing phosphate is most often accomplished by adding a coagulant, usually alum or lime. Phosphate removal from wastewater by biological means involves assimilation or storage.

8.15 Bioremediation

Bioremediation is a biological process whereby organic wastes are biologically degraded under controlled conditions. This process involves the use of living organisms, primarily microorganisms, to degrade the environmental contaminants. In this process, contaminant compounds are transformed by living organisms through reactions that take place as a part of their metabolic processes. For bioremediation to be effective, microorganisms must enzymatically attack the contaminants and convert them to harmless products. Hence, it is effective only where environmental conditions permit microbial growth and activity. Thus, its application involves the manipulation of environmental parameters to allow microbial growth and degradation to proceed at a faster rate. The control and optimization of bioremediation processes is a complex phenomenon. Various factors influencing this process include: the existence of a microbial population capable of degrading the pollutants; the availability of contaminants to the microbial population; and the environment factors (type of soil, temperature, pH, the presence of oxygen or other electron acceptors, and nutrients).

Bioremediation strategies

Bioremediation strategies can be *in-situ* or *ex-situ*. *In-situ bioremediation* involves treating the contaminated material at the site while *ex-situ bioremediation* involves the removal of the contaminated material to be treated elsewhere. *In-situ* bioremediation techniques are generally the most desirable options due to lower cost and less disturbance since they provide the treatment at a site avoiding excavation and transport of contaminants. *Ex-situ* bioremediation requires transport of the contaminated water or excavation of contaminated soil prior to remediation treatments.

In-situ and *ex-situ* bioremediation strategies involve different technologies such as bioventing, biosparging, bioreactor, composting, landfarming, bioaugmentation and biostimulation.

Bioventing is an *in-situ* bioremediation technology that uses microorganisms to biodegrade organic constituents adsorbed on soils in the *unsaturated zone* (extends from the top of the ground surface to the water table). Bioventing enhances the activity of indigenous bacteria and stimulates the natural *in-situ* biodegradation of contaminated materials in soil by inducing air or oxygen flow into the unsaturated zone and, if necessary, by adding nutrients.

Biosparging is also an *in-situ* bioremediation technology that uses indigenous microorganisms to biodegrade organic constituents in the *saturated zone*. In biosparging, air (or oxygen) and nutrients (if needed) are injected into the saturated zone to increase the biological activity of the indigenous microorganisms.

Biostimulation involves the modification of the environment to stimulate the existing bacteria capable of bioremediation. This can be done by the addition of various forms of rate limiting nutrients and electron acceptors, such as phosphorus, nitrogen, oxygen or carbon (e.g. in the form of molasses).

Pages 906 to 911 are not shown in this preview.

Chapter 09

Bioinformatics

9.1 Introduction

Bioinformatics is a discipline at the intersection of biology, computer science, information technology and mathematics. It aims at integrating and analyzing a wealth of biological data with the aim of identifying and assigning a function to each. It is applied, for example, in the construction of genetic and physical maps of genomes, gene discovery, the inference of the molecular function and three-dimensional structure of their products, the interpretation of the effect of gene variations on the phenotype, the reconstruction of interaction and signal transduction pathways and the simulation of biological systems.

9.2 Biological databases

Bioinformatics is about exploring biological information. This information is kept safely in databases. A *database* consists of an organized collection of persistent data that provides a standardized way for locating, adding, and changing data. Biological data are available in the form of sequences and structures of proteins and nucleic acids. The biological information of nucleic acids is available as sequences while the data of proteins is available as sequences and structures. Sequences are represented in a single dimension whereas the structure contains the three dimensional data of sequences.

The first database was created after the insulin protein sequence was made available in 1956. Insulin (consists of 51 residues) is the first protein to be sequenced. Later, three dimensional structure of proteins were studied and the well known *Protein Data Bank* was developed as the first protein structure database.

Database classification

Biological databases can be classified into sequence and structure databases or primary and secondary databases. Primary and secondary databases are classified on the basis of source of data.

Primary databases

Databases consisting of data derived experimentally such as nucleotide sequences and three dimensional structures are known as *primary databases.* Examples of these include GenBank, EMBL and DDBJ for nucleotide sequences and the Protein Data Bank (PDB) for 3D-protein structures.

Secondary databases

A *secondary database* derives from the analysis or treatment of the primary database. A secondary sequence database contains information like the conserved sequence, signature sequence and active site residues of the protein families arrived by multiple sequence alignment of a set of related proteins.

This page intentionally left blank.

- ALN: a database of protein sequence alignments.
- RESID: a database of covalent protein structure modifications.

In 2002 PIR, along with its international partners, EBI (European Bioinformatics Institute) and SIB (Swiss Institute of Bioinformatics), were awarded a grant from NIH to create UniProt, a single worldwide database of protein sequence and function, by unifying the PIR-PSD, Swiss-Prot, and TrEMBL databases. The UniProt database has larger coverage than any one of the three databases while at the same time maintaining the original SWISS-PROT feature of low redundancy, cross-references and a high quality of annotation.

Protein Data Bank (PDB)

The PDB archive is the single worldwide repository of information about the 3D structures of large biological molecules, including proteins and nucleic acids. Understanding the shape of a molecule helps to understand how it works. The PDB was established in 1971 at Brookhaven National Laboratory and originally contained 7 structures. In 1998, the Research Collaboratory for Structural Bioinformatics (RCSB) became responsible for the management of the PDB. As of Aug 2012, around 83000 structures are deposited so far in PDB. In 2003, the wwPDB was formed to maintain a single PDB archive of macromolecular structural data that is freely and publicly available to the global community. It consists of organizations that act as deposition, data processing and distribution centers for PDB data.

Structural Classification of Proteins (SCOP)

The SCOP database provides a detailed and comprehensive description of the relationships of known protein structures. PDB contains many protein entries. These proteins have structural similarities with other proteins and, in many cases, share a common evolutionary origin. To facilitate access to this information, the Structural Classification of Proteins (SCOP) database was constructed. The classification of proteins in SCOP has been constructed by visual inspection and comparison of structures. The unit of classification is usually the protein domain. The classification of the proteins in SCOP is on hierarchical levels are as follows: Family, Superfamily, Common fold and Class. There are now a number of other databases which classify protein structures, such as CATH, FSSP, Entrez and DDBASE, however, the distinction between evolutionary relationships and those that arise from the physics and chemistry of proteins is a feature that is so far unique to SCOP. Because functional similarity is implied by an evolutionary relationship but not necessarily by a physical relationship, we believe that this classification level is of considerable value, for example as a way of reliably linking very distant sequence families.

Class, Architecture, Topology and Homology (CATH)

The CATH classification of protein domain structures was established in 1993 as a hierarchical clustering of protein domain structures into evolutionary families and structural groupings, depending on sequence and structure similarity. There are four major levels, corresponding to protein class, architecture, topology or fold and homologous family. CATH consists of both phylogenetic and phenetic descriptors for protein domain relationships.

Molecular Modeling Database (MMDB)

The three-dimensional structures of biomolecules provide a wealth of information on their biological function and evolutionary relationships. The MMDB, as part of the Entrez system, facilitates access to structure data by connecting them with associated literature, protein and nucleic acid sequences, chemicals, biomolecular interactions, and more. It is possible, for example, to find 3D structures for homologs of a protein of interest by following the 'Related Structures' link in an Entrez Protein sequence record.

Genome databases

Genome sequences form entries in the standard nucleic acid sequence databases. Many species like *Arabidopsis thaliana*, *C. elegans*, Rice etc., have special databases that bring together the genome sequence and its annotation with other data related to the species.

- Microbial genome database
  http://www.ncbi.nlm.nih.gov:80/PMGGifs/Genomes/micr.html
- TIGR: The comprehensive Microbial Resource
  http://www.tigr.org/tigr-scripts/CMR2:CMRHomepage.spl
- Arabidopsis thaliana genome displayer
  http://www.kazusa.or.jp/kaos
- *Caenorhabditis elegans* (worm) database
  http://www.wormbase.org/
- EBI genomes
  http://www.ebi.ac.uk/genomes/

Superspecialized databases

Many individuals or groups select, annotate, and recombine data focused on particular topics, and include links affording streamlined access to information about subjects of interest. The **protein kinase resource** is a specialized compilation that includes sequences, structures and functional information, laboratory procedures, list of interested scientists, tools for analysis, a bulletin board and links. The **HIV protease** database store structures of HIV1 proteinases, HIV2 proteinases and SIV proteinases, and their complexes and provides tools for their analysis and other links.

9.3 Sequence formats

The protein and nucleic acids sequences can be stored in computer files. Once in the computer, the sequences can be analyzed by a variety of methods. Most sequence analysis programs require that the information in a sequence file be stored in a particular format. *Format* refers to the arrangement of data within a document file that typically permits the document/data to be read or written by certain application. In other words, it is an organization of data in a particular order. Some of the commonly used sequence formats are discussed below:

GenBank sequence entry

It has the following features:

- LOCUS: Short name for this sequence (Maximum of 32 characters).
- DEFINITION: Definition of sequence (Maximum of 80 characters).
- ACCESSION: accession number of the entry.
- VERSION: Version of the entry.
- DBSOURCE: Shows the source, the date of creation and last modification of the database entry.
- KEYWORDS: Keywords for the entry.
- AUTHORS: Authors of the work.
- TITLE: Title of the publication.
- JOURNAL: Journal reference for the entry.
- MEDLINE: Medline ID.
- COMMENT: Lines of comments.
- SOURCE ORGANISM: The organism from which the sequence was derived.
- ORGANISM: Full name of organism (Maximum of 80 characters).
- AUTHORS: Authors of this sequence (Maximum of 80 characters).
- ACCESSION: ID Number for this sequence (Maximum of 80 characters).
- FEATURES: Features of the sequence.
- ORIGIN: Beginning of sequence data.
- // End of sequence

This page intentionally left blank.

FASTA sequence format

The FASTA sequence format includes three parts shown in the figure below:

- A comment line identified by a ">" character in the first column followed by the name and origin of the sequence;
- The sequence is standard one-letter symbol and
- An optional "*" which indicates the end of the sequence and which may or may not be present.

```
>MCHU - Calmodulin - Human, rabbit, bovine, rat, and chicken
ADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADGNGTID
FPEFLTMMARKMKDTDSEEEIREAFRVFDKDGNGYISAAELRHVMTNLGEKLTDEEVDEMIREA
DIDGDGQVNYEEFVQMMTAK*
```

Figure 9.3 FASTA sequence entry format.

NBRF/PIR sequence format

The NBRF (National Biomedical Research Foundation) format has the following features. The first line includes an initial ">" character followed by a two-letter code such as P for complete sequence or F for fragment, followed by a 1 or 2 to indicate type of sequence, then a semicolon, then a four- to six-character unique name for the entry. There is also an essential second line with the full name of the sequence, a hyphen, then the species of origin. The sequence terminates with an asterisk.

```
>P1;CRAB_ANAPL
ALPHA CRYSTALLIN B CHAIN (ALPHA(B)-CRYSTALLIN).
MDITIHNPLIRRPLFSWLAPSRIFDQIFGEHLQESELLPASPSLSPFLMRSPIFRMPSWL
ETGLSEMRLEKDKFSVNLDVKHFSPEELKVKVLGDMVEIHGKHEERQDEHGFIAREFN
RKYRIPADVDPLTITSSLSLDGVLTVSAPRKQSDVPERSIPITREEKPAIAGAQRK*
```

Figure 9.4 NBRF/PIR sequence entry format.

9.4 Biosequence analysis

The determination of the linear sequence of amino acids in proteins and the nucleotides in DNA and RNA leads to the requisite for compiling and analyzing sequence data. Sequence analysis is the process of investigating the information content of linear raw nucleic acid and protein sequence data.

Amino acid sequence analysis

Apart from maintaining the large database, mining useful information from these sets of primary and secondary databases is very important. Linear chains of amino acids, in proteins, the product of gene translation, are normally found in cells folded into functionally active structures. It is established that the primary sequence of the protein, that is, its amino acid sequence, determines the ultimate conformation of the protein and therefore its biological function. However, the flexibility of long-chain polypeptides can generate an almost infinite number of shapes, and the computational task of predicting correct structures is beyond the reach of current knowledge. Predicting the shape of a protein from its linear amino acid sequence is one of the important goals of computational biology.

A lot of efficient algorithms have been developed for data mining and knowledge discovery. These are computation intensive and need fast and parallel computing facilities for handling multiple queries simultaneously. It is these search tools that integrate the user and the databases. One of the widely used search program is BLAST (Basic Local Alignment Search Tool).

Nucleic acid sequence analysis

Nucleic acid sequence analysis includes assembling partially overlapping fragments, analyzing sequences, comparing sequences and detecting functional (RNA coding) regions. The bulk of genomic DNA does not code for proteins, and the protein-coding regions of human genes are not collinear but arranged with exons interspersed with introns. Therefore, an important question for computational biology is how to detect protein-coding regions within genomic DNA.

Current DNA sequencing technologies are not capable of generating a complete sequence of long nucleic acid molecules in a single sequencing run and so it is necessary to utilize computational methods to assemble contiguous sequences from individual short-sequence determinations. If a large DNA molecule is randomly broken into smaller pieces for the actual sequence determinations then a contiguous linear sequence can be reconstructed by aligning the overlapping portions from different random fragments.

A common question arising when new genes are cloned and sequenced is whether the sequence is already known or does not occur in current databases. Answering this question requires comparing the newly obtained sequence to every sequence in the database.

9.5 Sequence alignment

Sequence alignment refers to the procedure of comparing two or more sequences of nucleic acid or protein by looking for a series of individual characters or character patterns that are in the same order in the sequences. It is used to identify regions of similarity that may be a consequence of functional, structural or evolutionary relationships between the sequences.

Global alignments and local alignments

Computational approaches to sequence alignment generally fall into two categories: *global alignments* and *local alignments*.

G*lobal alignment* is an attempt to match as many characters as possible, from end to end, in a set of two or more sequences. It attempts to align every residue in every sequence. Sequences that are quite similar and approximately of the same length are suitable candidates for global alignment. A general global alignment technique is the *Needleman-Wunsch algorithm*, which is based on dynamic programming.

Local alignment searches for regions of local similarity need not include the entire length of the sequences. Local alignments are more useful for dissimilar sequences that are suspected to contain regions of similarity or similar sequence motifs within their larger sequence context. The *Smith-Waterman algorithm* is a general local alignment algorithm, also based on dynamic programming. With sufficiently similar sequences, there is no difference between local and global alignments.

Pairwise and multiple sequence alignments

Pairwise sequence alignment

Pairwise alignment is used between two query sequences at a time. It is used to identify regions of similarity that may indicate functional, structural and/or evolutionary relationships between two biological sequences (protein or nucleic acid). It involves matching of homologous positions in two sequences. Positions with no homologous pair are matched with a space '–' and a group of consecutive spaces is a *gap*.

CA−−GATTCGAAT
CGCCGATT−−−AT
gap

The three primary methods of producing pairwise alignments are dot-matrix methods, dynamic programming, and word methods. Although each method has its individual strengths and weaknesses, all three pairwise methods have difficulty with highly repetitive sequences.

Pages 919 to 921 are not shown in this preview.

The **E-value** is a parameter that describes the number of hits one can 'expect' to see by chance when searching a database of a particular size. It decreases exponentially with the score (S) that is assigned to a match between two sequences. Essentially, the E-value describes the random background noise that exists for matches between sequences. For example, an E-value of 1 assigned to a hit can be interpreted as in a database of the current size, one might expect to see one match with a similar score simply by chance. This means that the lower the E-value, or the closer it is to '0', the higher is the 'significance' of the match. However, it is important to note that searches with short sequences can be virtually identical and have relatively high E-value. This is because the calculation of the E-value also takes into account the length of the query sequence. This is because shorter sequences have a high probability of occurring in the database purely by chance.

BLAST family

There are a number of different versions of the BLAST program for comparing either nucleic acid or protein sequence with nucleic acid or protein sequence databases. These programes are:

- BLASTP compares an amino acid query sequence against a protein sequence database.
- BLASTN compares a nucleotide query sequence against a nucleotide sequence database.
- BLASTX compares a nucleotide query sequence translated in all reading frames against a protein sequence database.
- TBLASTN compares a protein query sequence against a nucleotide sequence database dynamically translated in all reading frames.
- TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

PSI-BLAST (Position-Specific-Iterated BLAST)

PSI-BLAST uses a method that involves a series of repeated steps or iterations. First, a database search of a protein sequence database is performed using a query sequence. Second, the results of the search are presented and can be assessed visually to see whether any database sequences that are significantly related to the query sequence are present. Third, if such is the case, the mouse is clicked on a decision box to go through another iteration of the search. The high-scoring sequence matches found in the first step are aligned, and from the alignment, a type of scoring matrix that indicates the variations at each aligned position is produced. The database is then again searched with this scoring matrix.

PHI-BLAST (Pattern-Hit Initiated BLAST)

This program functions much like PSI-BLAST except that the query sequence is first searched for a complex pattern provided by the investigator. The subsequent search for similarity in the protein sequence database is then focused on regions containing the pattern. PSI-BLAST like other programs are – sSEARCH, MAXHOM.

FASTA

FASTA is a software program for rapid alignment of pairs of protein and DNA sequences. FASTA is pronounced 'fast A', where A stands for All, because it works with any alphabet, an extension of 'FAST-P' (protein) and 'FAST-N' (nucleotide) alignment. It is a heuristic approximation to the Smith-Waterman algorithm. It is a two step algorithm. The first step is a search for highly similar segments in the two sequences. In this search a word with a specific word size is used to find regions in a two-dimensional table similar to the Smith-Waterman algorithm. These regions are a diagonal or a few closely spaced diagonals in the table which have a high number of identical word matches between the sequences. The second step is a Smith-Waterman alignment centered on the diagonals that correspond to the alignment of the highly similar sequence segments.

Version of FASTA

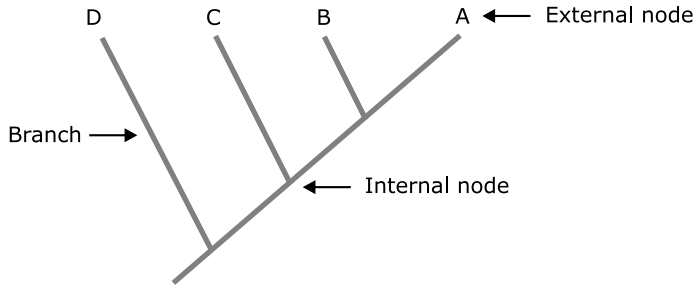
1. FASTA compares a query protein sequence to a protein sequence library to find similar sequences. FASTA also compares a DNA sequence to a DNA sequence library.

This page intentionally left blank.

between the sequences. Nucleic acids and proteins are linear molecules made of smaller units called nucleotides and amino acids, respectively. The nucleotide differences within a gene or amino acid differences within a protein reflect the evolutionary distance between two organisms. In other words, closely related organisms will exhibit fewer sequence differences than distantly related organisms.

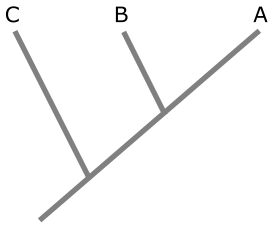
Phylogenetic trees

In phylogenetic studies, the most convenient way of visually presenting evolutionary relationships among a group of organisms is through illustrations called *phylogenetic trees*. Phylogenetic tree is represented by *lines* and *nodes*. Nodes can be internal or external (terminal). The different sequences of DNA/proteins compared are located at *external nodes* but connected via branches to *interior nodes* which represent ancestral forms for two or more sequences. The terminal nodes at the tips of trees represent operational taxonomic units (OTUs). *Branch* defines the relationship between the taxa in terms of descent and ancestry. The lengths of the branches indicate the degree of difference between the sequence represented by the nodes. The branch lengths are proportional to the predicted evolutionary time between organisms or sequences. The branching pattern of the tree is termed a *topology*.

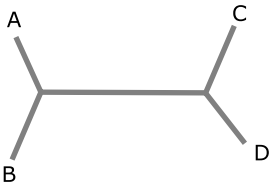


An evolutionary tree

A phylogenetic tree may be *rooted* or *unrooted.* A *rooted tree* infers the existence of a common ancestor and indicates the direction on the evolutionary process. A rooted tree in which every node has two descendants is called a *binary tree*. An *unrooted tree* does not infer a common ancestor and shows only the evolutionary relationships between the organisms.



Rooted tree



Unrooted tree

Gene trees versus species trees

A *gene tree* is a model of how a gene evolves through duplication, loss, and nucleotide substitution. It is constructed from comparisons between the sequences of orthologous genes. A *species tree* depicts the pattern of branching of species lineages via the process of speciation. When reproductive communities are split by speciation, the gene copies within these communities likewise are split into separate bundles of descent.

An internal node in a gene tree indicates the divergence of an ancestral gene into two genes with different DNA sequences, usually resulting from a mutation of one sort or another. An internal node in a species tree represents what is called a *speciation event*, whereby the population of the ancestral species splits into two groups that are no longer able to interbreed. These two events, mutation and speciation, do not always occur at the same time.

Pages 925 to 926 are not shown in this preview.

which reproduce the original data as closely as possible. An example of the distance method for a dataset of 4 nucleic-acid sequences is given below. The diagram below summarizes the calculation of pairwise distances between the gene sequences for four hypothetical species.

Proportional distance (p)

DNA site

Species	1	2	3	4	5	6	7	8	9	10
1	A	T	A	T	A	C	G	T	A	T
2	A	T	G	T	A	C	G	T	A	T
3	G	T	A	–	A	C	G	T	G	C
4	G	C	G	T	A	T	G	C	A	C

$$p = \frac{\text{Differences}}{\text{Sites}}$$

	1	2	3	4
1	–	0.1	0.4	0.6
2		–	0.5	0.5
3			–	0.6
4				–

The coefficients provide a simple summary of how similar (or different) each sequence is from the other. Sequence 1 and 2 are more alike to each other than either is to 3. In this example, we calculated the distances across the length of the whole sequence (10 bases); distances can be calculated for different sections of a sequence to see if some parts are more conserved than others.

Maximum likelihood approach

This method uses probability calculations to find a tree that best accounts for the variation in a set of sequences. All possible trees are considered, Hence, the method is only feasible for a small number of sequences. For each tree, the number of sequence changes or mutations that may have occurred to the given sequence variation is considered. Because the rate of appearance of new mutations is very small, the more mutations needed to fit a tree to the data, the less likely the tree.

The maximum likelihood method presents an additional opportunity to evaluate trees with variations in mutation rates in different lineages, and to use explicit evolutionary models such as the Jukes-Cantor and Kimura models. The method can be used to explore relationships among more diverse sequences and conditions that are not well handled by maximum parsimony methods.

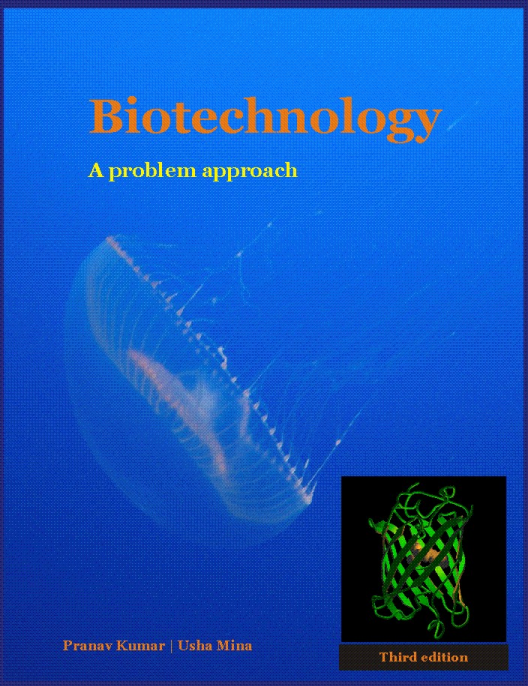
9.7 Protein structure prediction

Genome sequencing projects are producing linear amino acid sequences, but full understanding of the biological role of these proteins will require knowledge of their structure and function. One of the major goals of bioinformatics is to understand the relationship between amino acid sequence and the three dimensional structure in proteins. If these relationships are known then the structure of a protein could be reliably predicted from the amino acid sequence. Although experimental structure determination methods are providing high-resolution structure information about a subset of the proteins, computational structure prediction methods will provide valuable information for the large fraction of sequences whose structures will not be determined experimentally.

Methods for prediction of protein structure from amino acid sequence include:

- Attempts to predict *secondary structure* without attempting to assemble these regions in three dimensions.
- Homology modeling prediction of the three-dimensional structure of a protein from the known structures of one or more related proteins.

This page intentionally left blank.



Biotechnology

A problem approach

This covers essential fundamentals and their applications. This book provides a balanced introduction to all major areas of the subject. It is presented in a sharply focused manner without overwhelming or excessive details. Each chapter contains information in a systematic and logically organized manner. All chapters were carefully edited to maintain consistent use of terminology.

Key features of Biotechnology–A problem approach

- Focuses on fundamentals and principles with expanded coverage of critical topics.
- Enables the reader to grasp the subject quickly and easily
- Offers a structured approach to learning.
- Contains clear and simple illustrations.

Pathfinder Publication
www.thepathfinder.in
09350208235

978-93-80473-00-0

₹ 555/-