

### 1.3.4.5 Cyanide and Isocyanide Groups

In isocyanide ( $\text{R}-\text{N}=\text{C}$ ) both carbon and nitrogen atom of  $-\text{CN}$  group are  $\text{sp}$  hybridized and include one sigma bond which is formed by overlapping of  $\text{sp}$  hybrid orbitals and two pi-bonds formed by the sideways overlap of  $2\text{p}$  atomic orbitals. Nitrogen atom containing a lone pair of electrons in  $\text{sp}$  hybrid orbital points away from the molecule.

### 1.3.4.6 Nitro Group

It is an ambient nucleophile. When the linkage of a group is through oxygen atom ( $\text{R}-\text{O}-\text{N}=\text{O}$ ), the group is known as nitrite, and when the linkage is through nitrogen atom ( $\text{R}-\text{NO}_2$ ), it is known as nitro.

### 1.3.4.7 Amines

These are derivatives of ammonia in which one H atom is replaced by alkyl or aryl group and they are classified as  $1^\circ$ ,  $2^\circ$  and  $3^\circ$  amines. The simplest aromatic amine is aniline. The nitrogen atom in amine is  $\text{sp}^3$  hybridized forming tetrahedral arrangement. Three vertices of tetrahedron are occupied by three substituents and fourth by lone pair of electrons. This geometry is known as pyramidal.

1.4

## Carbohydrates

Carbohydrates or saccharides are an important non-nitrogenous group of natural products. They are widely distributed in nature, especially in plants. Most commonly known as sugars, they constitute the most abundant class of biomolecules. Carbohydrates are called the hydrates of carbon, because they contain hydrogen and oxygen in the same proportion as in water, having the general formula  $\text{C}_n(\text{H}_2\text{O})_n$ . Although a large number of compounds correspond to this formula, still there are several exceptions, for example 2-deoxyribose,  $\text{C}_5\text{H}_{10}\text{O}_4$ . Moreover all compounds conforming to the formula  $\text{C}_n(\text{H}_2\text{O})_n$  are not necessarily carbohydrates, for example formaldehyde ( $\text{HCHO}$ ) and acetic acid ( $\text{CH}_3\text{COOH}$ ). This definition, obviously, is not appropriate for such compounds, which are really not carbohydrates, even though these fall under the general formula. Therefore, the carbohydrates may be defined as polyhydroxy aldehyde or ketone or substances that yield these on hydrolysis.

### 1.4.1 Classification

Carbohydrates are divided into two main classes: sugars and polysaccharides. Sugars are further subdivided into two major groups:

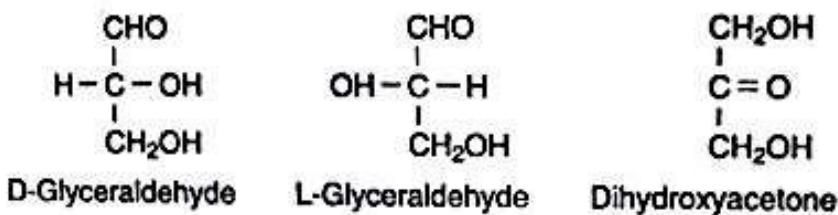
1. Monosaccharides: These are the simplest sugars which cannot be hydrolyzed into smaller molecules. The general formula is  $\text{C}_n\text{H}_{2n}\text{O}_n$ , where  $n$  varies from 3 to 7. The most important are the pentoses and hexoses.
2. Oligosaccharides: They consist of short chain (2–10) monosaccharides joined together by glycosidic bonds. The most abundant are the disaccharides, which consist of two monosaccharide units. Sucrose or cane sugar is one typical disaccharide, which consists of the six-carbon sugars, D-glucose and D-fructose.

Polysaccharides, on the other hand, are carbohydrates which yield a large number of monosaccharides on hydrolysis. The most widely spread polysaccharides have general formula  $(\text{C}_6\text{H}_{10}\text{O}_5)_n$ . A group of polysaccharides which are not widely spread in nature is the pentosans  $(\text{C}_5\text{H}_8\text{O}_4)_n$ .

### 1.4.1.1 Monosaccharides

These are polyhydroxy aldehyde (aldose) or polyhydroxyketone (ketose). Glyceraldehyde, the simplest sugar, is a three-carbon aldotriose and dihydroxyacetone is a ketotriose. Aldoses with four, five, six and seven carbon atoms are called aldötetroses, aldopentoses, aldohexoses and aldoheptoses, respectively. Similarly ketoses can be named ketotetroses, ketopentoses, ketohexoses and ketoheptoses.

All monosaccharides contain one or more asymmetric (chiral) carbon atom. Glyceraldehyde contains one chiral carbon; therefore, it has two different optical isomers designated as D-glyceraldehyde and L-glyceraldehyde, nonsuperimposable mirror image. On the other hand, dihydroxy acetone does not contain a chiral carbon atom; therefore, it exists only in one form.



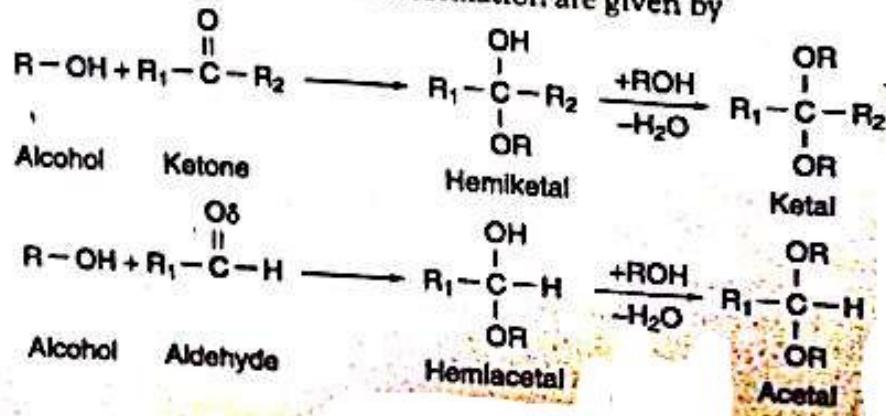
Six-carbon sugars are most abundant in nature; two five-carbon sugars, ribose and deoxyribose, are important structural components of RNA and DNA, respectively. Four-carbon and seven-carbon sugars play important roles in photosynthesis and other metabolic pathways.

**Cyclic Structure of Monosaccharides:** Five- and six-carbon sugars occur predominantly as cyclic structures rather than an open structure. The cyclization takes place due to interaction between the functional group at C-1 and C-5 to form a cyclic hemiacetal in case of aldohexoses. Ketohexoses forms a cyclic hemiketal due to interaction between C-2 and C-5. In both aldohexoses and ketohexoses carbonyl carbon becomes a new chiral center called the anomeric carbon. The cyclic sugars exist in two different forms designated as  $\alpha$  and  $\beta$ , called anomers of each other.

The two forms are identical in chemical composition. The six-membered ring compounds are called pyranoses because they resemble the six-membered ring pyran. The formation of pyranose structure of D-glucose is the result of hemiacetal formation, which contains an asymmetric carbon and thus can exist in two stereoisomeric forms.

Ketohexoses also occurs in  $\alpha$  and  $\beta$  anomeric forms. In such compounds carbon atom 5 reacts with carbonyl group at carbon atom 2 forming a five-membered rings known as furanoses, containing a hemiketal linkage, because they resemble five-membered ring compound furan.

General reactions for ketal and acetal formation are given by



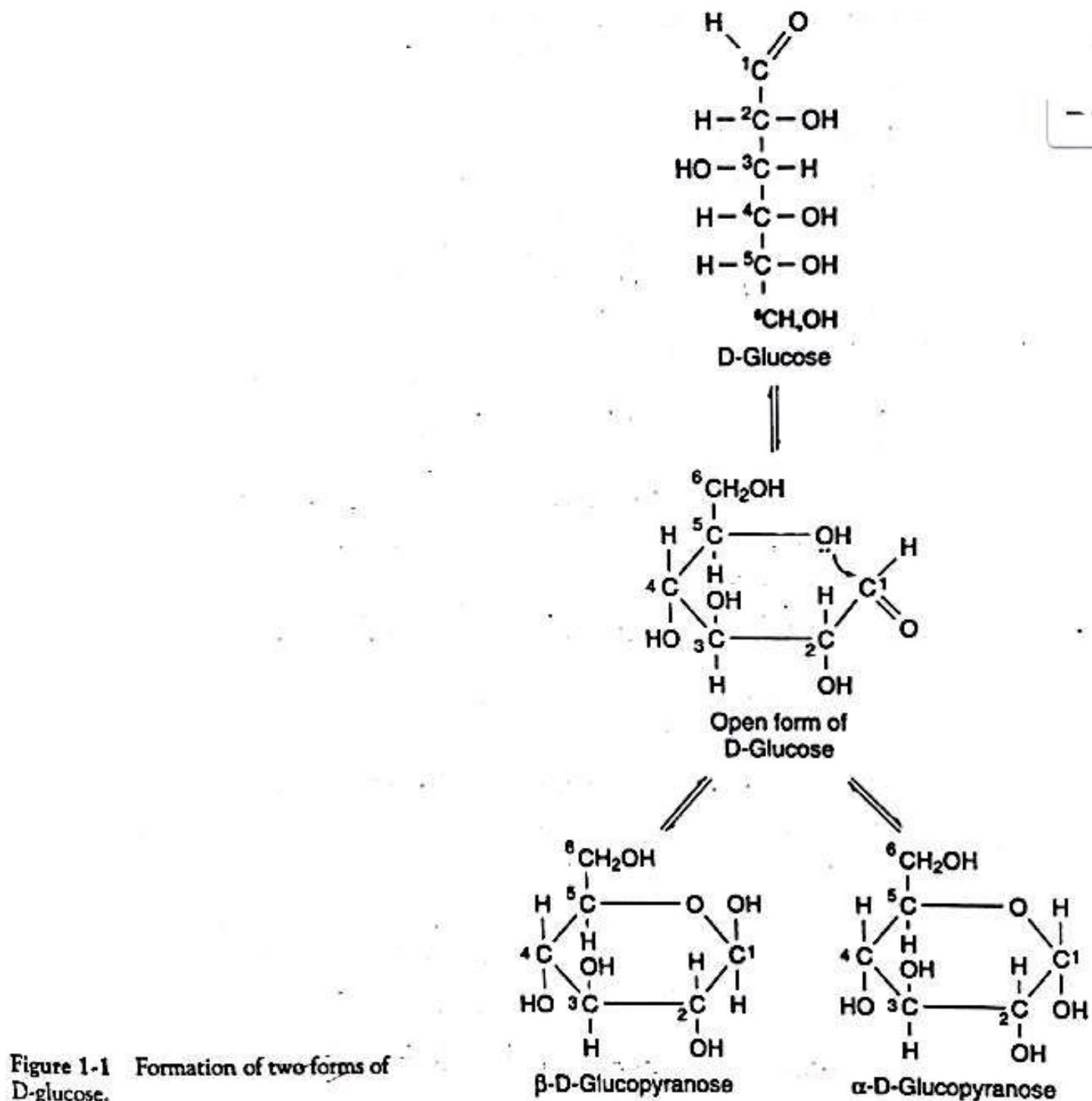


Figure 1-1 Formation of two forms of D-glucose.

The six-membered pyranose structure is not planar. Generally, the Haworth projection formula is used for showing different forms of monosaccharides. In most sugars it occurs in the chair conformation, while in some it exists as boat form. These are represented by conformational formulas.

It is difficult to isolate  $\alpha$  and  $\beta$  forms of a sugar. Both forms show mutarotation. When monosaccharide is dissolved in water, the optical rotatory power of the solution gradually changes until it reaches a constant value. A freshly prepared solution of glucose has a specific rotation of  $+112^\circ$  and when this solution is allowed to stand, rotation falls to  $+52.7^\circ$  and remains constant at this value. This corresponds to about 38% of the  $\alpha$ -isomer and 62% of  $\beta$ -isomer.

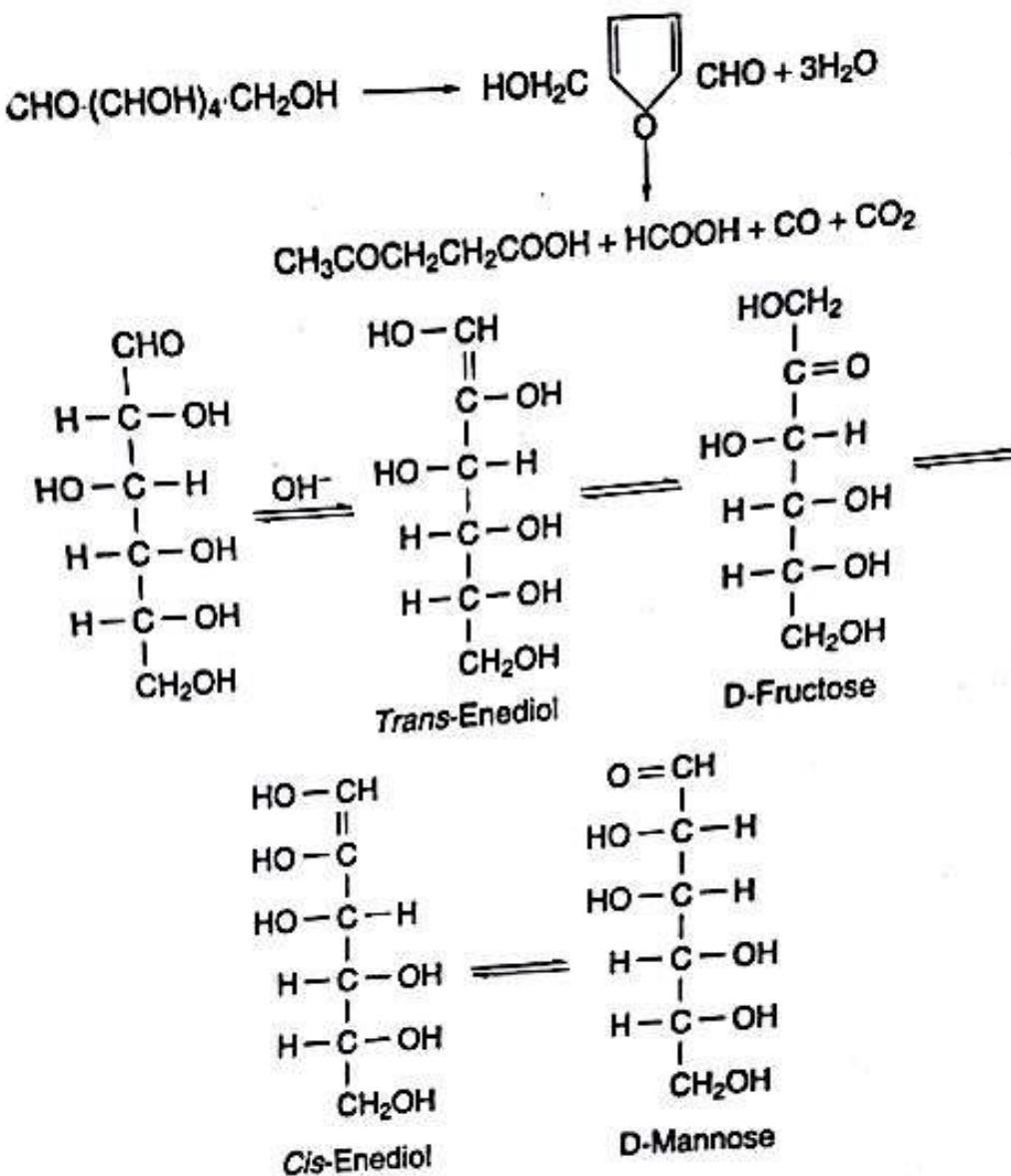


Figure 1-3 Isomerization of D-glucose by dilute alkali.

#### 1.4.1.2 Oligosaccharides

Oligosaccharides contain 3–10 monosaccharides joined together covalently by glycosidic linkages. Those of importance are disaccharides. Disaccharides are composed of two monosaccharide units joined together covalently by glycosidic bond. Glycosidic bond is formed when a hydroxyl group of one monosaccharide reacts with anomeric carbon of the other monosaccharide. There are two types of natural disaccharides, reducing and nonreducing.

##### Reducing Disaccharides:

1. Maltose: It is a disaccharide composed of two D-glucose residues joined together by  $\alpha$  (1 → 4) glycosidic bond. The  $\alpha$  refers to the configuration of the anomeric carbon C(1) of the left glucose residue, which exists as an acetal and has a fixed configuration. The second residue is a hemiacetal and has a variable configuration. The hydroxyl group attached to its anomeric carbon can have either  $\alpha$  or  $\beta$  configuration. The left glucose represents the nonreducing end while the right glucose has a potential aldehydic group, therefore, constitutes the reducing end. Since the disaccharide has a potential aldehydic group, therefore, the disaccharide is a reducing sugar. It reduces Fehling's solution, forms oxime and osazone. It also shows mutarotation. On oxidation with bromine water, it gives maltobionic acid  $C_{11}H_{22}O_{12}$ . On acetylation, two isomeric octa-acetyl derivatives are formed. This shows the presence of eight OH group in the molecule:

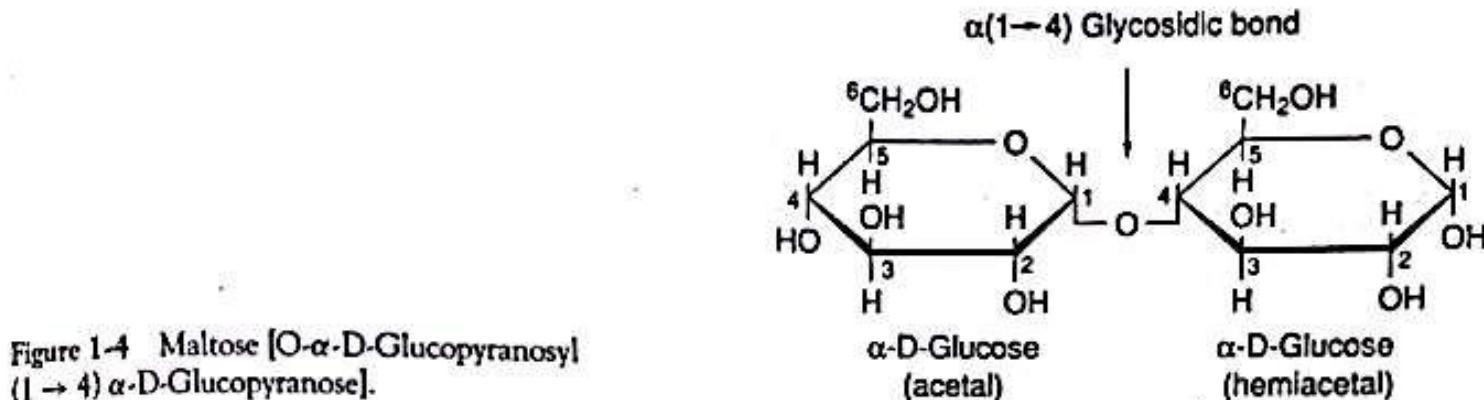


Figure 1-4 Maltose [ $O\text{-}\alpha\text{-D-Glucopyranosyl}$  ( $1 \rightarrow 4$ )  $\alpha\text{-D-Glucopyranose}$ ].

2. Lactose: Lactose is the main carbohydrate in milk. It is composed of galactose and glucose. These are linked through  $\beta$  ( $1 \rightarrow 4$ ) glycosidic bond, where  $\beta$  refers to the configuration of the anomeric carbon of galactose which exists as an acetal. There exists two forms of lactose ( $\alpha$  and  $\beta$ ) based on the configuration of the anomeric carbon of glucose which exists as hemiacetal. Because of the property of the glucose, lactose has a potential aldehyde group and is a reducing sugar. It reduces Fehling's solution and forms an osazone and oxime also undergoes mutarotation. On oxidation with bromine water, lactobionic acid  $C_{12}H_{22}O_{12}$  is formed which on hydrolysis yields, D-gluconic acid and D-galactose.

Lactase, an enzyme present in the intestinal villi, catalyzes the hydrolysis of lactose. In some adults the deficiency of this enzyme causes a build up of lactose when milk products are ingested. This is because the enzyme is necessary to degrade lactose so that it can be absorbed in the intestine and can be acted upon by the lactase of intestinal bacteria producing hydrogen, carbon dioxide and organic acids that lead to digestive problems, such as bloating and diarrhea. This is known as lactose intolerance.

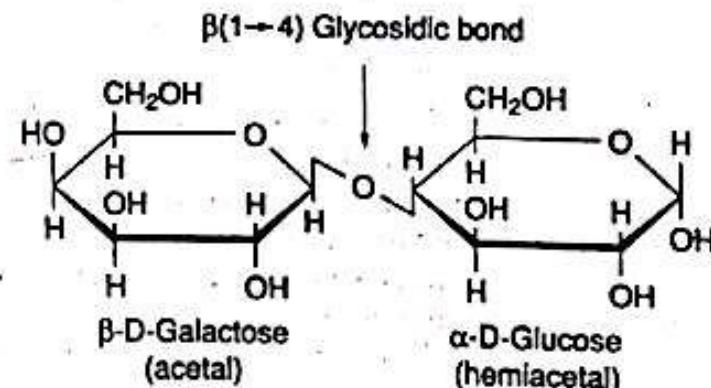


Figure 1-5 Lactose [ $O\text{-}\beta\text{-D-Galactopyranosyl}$  ( $1 \rightarrow 4$ )  $\alpha\text{-D-Glucopyranose}$ ].

**Nonreducing Disaccharides:** Sucrose (cane sugar) is made of  $\alpha$ -D-glucose and  $\beta$ -D-fructose. Glucose is a pyranose and fructose is a furanose. The  $\alpha$  C-1 of glucose is linked to  $\beta$  C-2 of fructose through a glycosidic bond. Sucrose is nonreducing because both anomeric groups are involved in the glycosidic linkage.

#### 1.4.1.3 Polysaccharides

These are carbohydrates which yield a large number of monosaccharides on hydrolysis. These have been further classified as:

### $\alpha(1 \rightarrow 2)$ Glycosidic bond

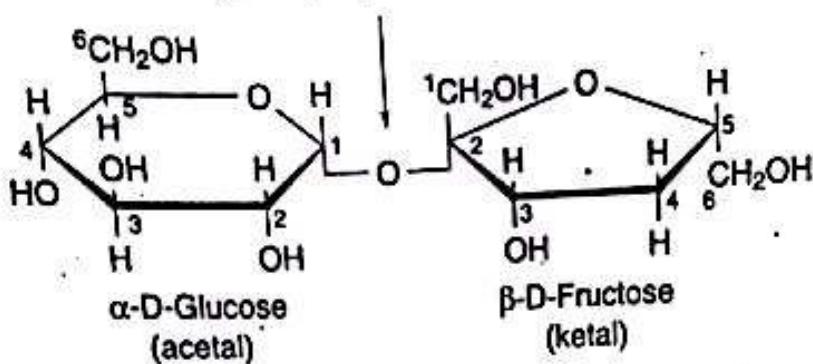


Figure 1-6 Sucrose [ $O\text{-}\alpha\text{-D-Glucopyranosyl}$   
 $(1 \rightarrow 2)\beta\text{-D-fructofuranoside}$ ]

1. Homopolysaccharides: They contain only a single type of monosaccharide. Some homopolysaccharides serve as storage forms (starch and glycogen) and others serve as structural component in plant cell walls and animal exoskeletons (cellulose and chitin).
2. Heteropolysaccharides: They contain more than one kind of monosaccharide, hemicellulose, xyloglucan, pectin and agarose in plants and hyaluronic acid, chondroitin, dermatan, keratan and heparin sulfates in animals.

### Homopolysaccharides

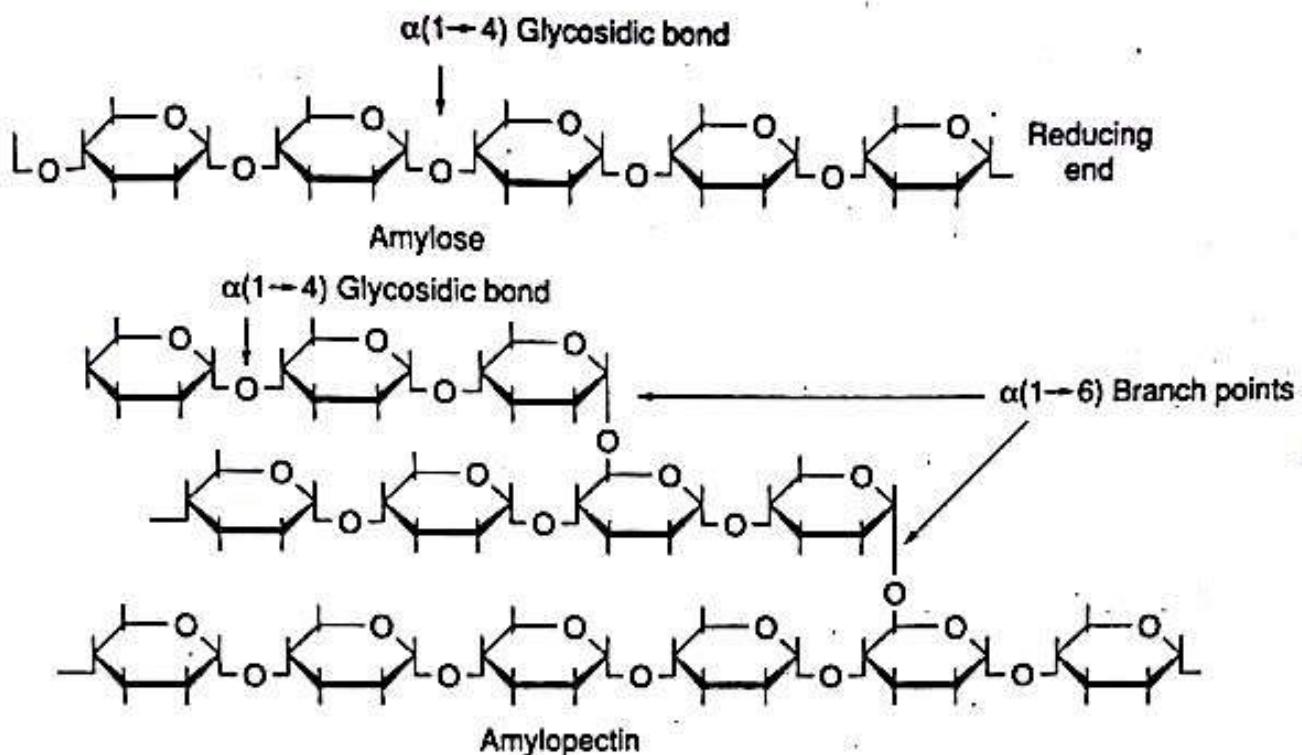
#### Storage Polysaccharides

1. Starch: It is the most abundant polymer of glucose, synthesized and stored in most of the green plants in special starch granules. These granules, which are formed within the membrane-bound plastids, have characteristic appearance and shape. The size and shape of starch granules vary from plant to plant. Starch occurs both as a linear polymer, amylose, and a branched polymer, amylopectin. Both types of starch consist entirely of D-glucose.
2. Amylose: It consists of residues linked by means of  $\alpha(1 \rightarrow 4)$  glycosidic bonds. Such chains vary in molecular weight from a few thousand to over a million. Owing to  $\alpha(1 \rightarrow 4)$  glycosidic bonds, each glucose residue is angled with respect to next, causing a turning in the chain and forming a regular helical configuration.

In amylopectin glucose residues are linked together by  $\alpha(1 \rightarrow 4)$  glycosidic bonds within each chain and by  $\alpha(1 \rightarrow 6)$  glycosidic bonds at branch points. Branching in amylopectin occurs about every 12–25 residues along the  $\alpha(1 \rightarrow 4)$  chain. Most starches contain 10–30% amylose and 70–90% amylopectin but there are special varieties of cereals which produce starches with 50–70% amylose. Waxy maize lacks amylose and contains only amylopectin. Differences in cereal starches arise mainly due to difference in the relative amounts of amylose and amylopectin and also due to chain length of these two polymers.

### Glycogen

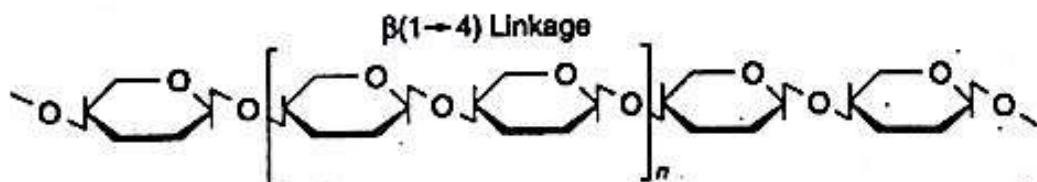
It is the main storage polysaccharide in animal cells (liver and skeletal muscles). The structure of glycogen is similar to that of amylopectin. It is a polymer of  $\alpha(1 \rightarrow 4)$  linked subunits of glucose with  $\alpha(1 \rightarrow 6)$  branching. Glycogen is more extensively branched compared to amylopectin. Branch points occur after about every 8–12 residues and like starch, glycogen is found in animal cells in granules. These granules are observed in well-fed liver and muscle cells and are tightly bound to the enzymes responsible for the synthesis and degradation.



**Figure 1-7** Linear  $\alpha$ -amylose and branched amylopectin.

## Cellulose

Cellulose is a fibrous, water-insoluble substance and is mainly responsible for the structure of cell walls of plants. It is a linear polymer consisting of long chains of 10,000 to 15,000  $\beta$ -D-glucose units. Glucose residues in cellulose are joined together by  $\beta$  (1  $\rightarrow$  4) linkages. These chains are closely attached to one another by intermolecular and intramolecular hydrogen bonds and van der Waals' forces. On complete hydrolysis with strong mineral acids, cellulose yields only D-glucose while on partial hydrolysis, cellobiose (a reducing sugar) is obtained. The most stable structure of cellulose converts the chains into sheets which are stacked to form fiber.



**Figure 1-8 Cellulose.**

## Chitin

Chitin resembles cellulose in both structure and function. It consists of a linear chain of a derived monosaccharide, N-acetyl- $\beta$ -D-glucosamine and is joined together by  $\beta$  (1 → 4) glycosidic linkage instead of  $\beta$ -D-glucose unit found in cellulose. Chitin plays a structural role and has got a fair amount of mechanical strength, as the individual chains are held together by hydrogen bonds. Chitin is a major constituent of the hard, horny exoskeletons of insects, lobsters and crabs. Like cellulose it is also a very abundant polysaccharide found in nature.

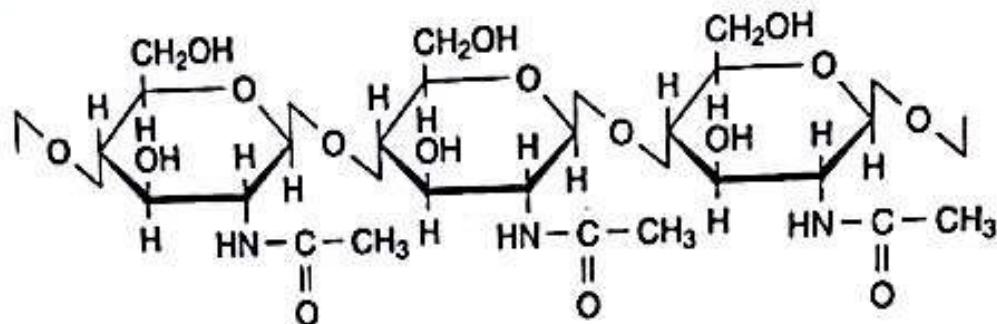


Figure 1-9 Structure of Chitin.

**Heteropolysaccharides:** These type of polysaccharides contain repeating units of different kind of monosaccharides. The most important heteropolysaccharides found in animals are: hyaluronate, chondroitin, dermatan, ketatan and heparin sulfates.

Hyaluronate is mostly found in connective tissue, the synovial fluid of joints, the vitreous humor of eye and cartilage. It is a polymer of glucuronic acid and N-acetylglucosamine, and has a molecular weight of several millions. Chondroitin and dermatan sulfates are similar polymers but with substitution of N-acetylgalactosamine and  $\alpha$ -L-iduronic acid and with sulfate ester group.

Keratan sulfates have no uronic acid and their sulfate content is variable. Heparin, a widespread component of cell surface, has a more variable composition, and is composed of few N-acetyl groups and fewer N- and O-sulfate groups.

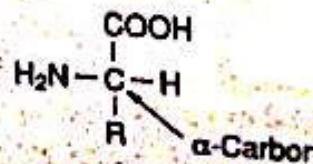
#### Plant heteropolysaccharides

One of the heteropolysaccharides present in plants is gums, with branched structure containing D-galactose, L-arabinose, L-rhamnose and D-glucuronic acid. Hemicellulose is another heteropolysaccharide which is composed of D-glucose, D-xylose and L-fucose. Pectin of higher plants contain  $\beta$  (1  $\rightarrow$  4) linked polygalacturonates intercepted by occasional  $\beta$  (1  $\rightarrow$  2) linked L-rhamnose residues. Agarose is composed of D-galactose and 3,6-anhydro-L-galactose joined together by  $\beta$  (1  $\rightarrow$  4) linkage.

## 15 Amino Acids and Peptides

Amino acids are the low molecular weight biomolecules which are the basic monomeric units of proteins. All proteins whether from bacteria or animal origins, consist of 20 amino acids. These amino acids are covalently linked in characteristic manner; the short chains consisting of two or more amino acids are known as peptides. However, long chains, containing large number of amino acids, are known as polypeptides or proteins. Besides proteins, amino acids are also building blocks of a large number of compounds, namely, enzymes, hormones, antibiotics and many other compounds having specific biological activity.

As indicated by name, amino acids are bifunctional compounds, consisting of both acidic and amine groups. Both carboxyl (acidic) and amino (amine) group are attached to the same  $\alpha$ -carbon atom. The remaining part of the molecule, R group, constitutes the sidechain of the amino acid. At pH 7.0, the physiological pH, both the  $\alpha$ -carboxyl and the  $\alpha$ -amino group are ionized.



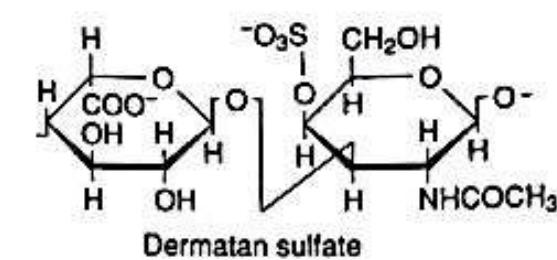
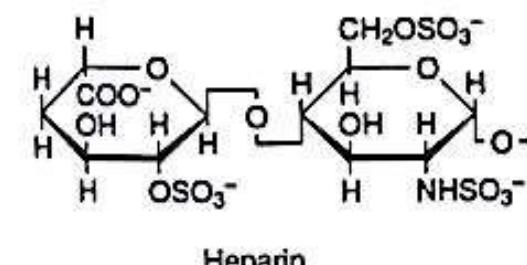
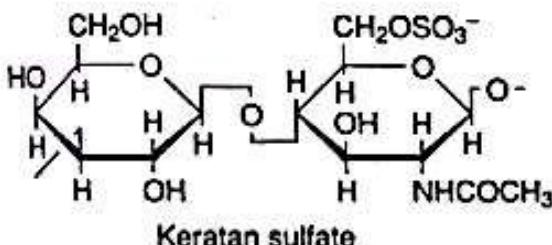
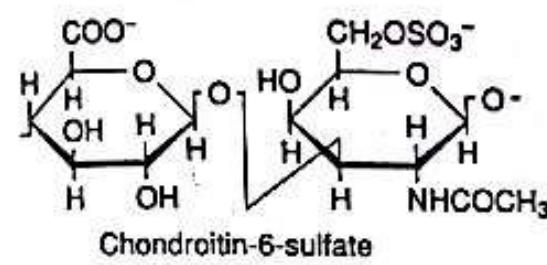
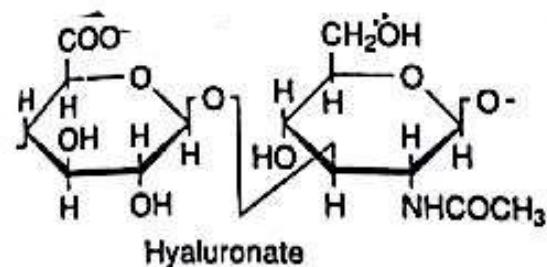


Figure 1-10 Repeating disaccharide units of heteropolysaccharides.

and glutamine is due to amide group and that of cysteine is due to thiol group. Cysteine occurs in protein in two forms either as cysteine itself or as cystine with a disulfide bridge.

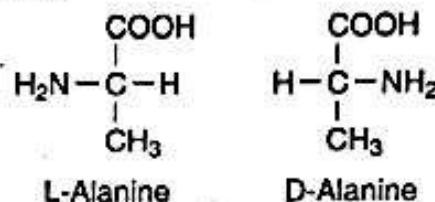
#### 1.5.1.3 Negatively Charged (Acidic) Amino Acids

The class includes amino acids in which both the groups are negatively charged.

#### 1.5.1.4 Positively Charged (Basic) Amino Acids

The amino acids of this class have a net positive charge. Lysine has a second amino group known as  $\epsilon$ -amino group on aliphatic chain. Arginine and histidine have positively charged guanidino and imidazole groups.

All the standard amino acids except glycine have an asymmetric carbon atom, that is the  $\alpha$ -carbon is attached to four different groups, therefore, the  $\alpha$ -carbon is a chiral center. Like the three carbon sugar glyceraldehyde, which exists in D and L configuration. All amino acids exist in two stereoisomeric forms. However, in protein all amino acids except glycine are present in L-stereoisomers.



#### 1.5.1 Classification of Amino Acids

All the standard amino acids have  $\alpha$ -carboxyl and  $\alpha$ -aminogroups, the two functional groups. Additionally, these may have a third functional group as sidechain, its chemical nature varies widely on the basis of sidechain (R). The amino acids may be categorized in four classes:

1. nonpolar or hydrophobic;
2. polar but uncharged (neutral);
3. negatively charged;
4. positively charged.

##### 1.5.1.1 Nonpolar or Hydrophobic Amino Acids

This class includes four amino acids with aliphatic sidechain (alanine, valine, leucine, isoleucine) and two with aromatic moiety (phenylalanine and tryptophan). Methionine contains sulfur while proline has a cyclic sidechain; these are also in this group.

##### 1.5.1.2 Polar but Uncharged (Neutral) Amino Acids

Glycine, serine, threonine, cysteine, tyrosine, asparagine and glutamine are included in this class. The polarity of serine, threonine and tyrosine is due to the presence of hydroxyl groups. The polarity of asparagine

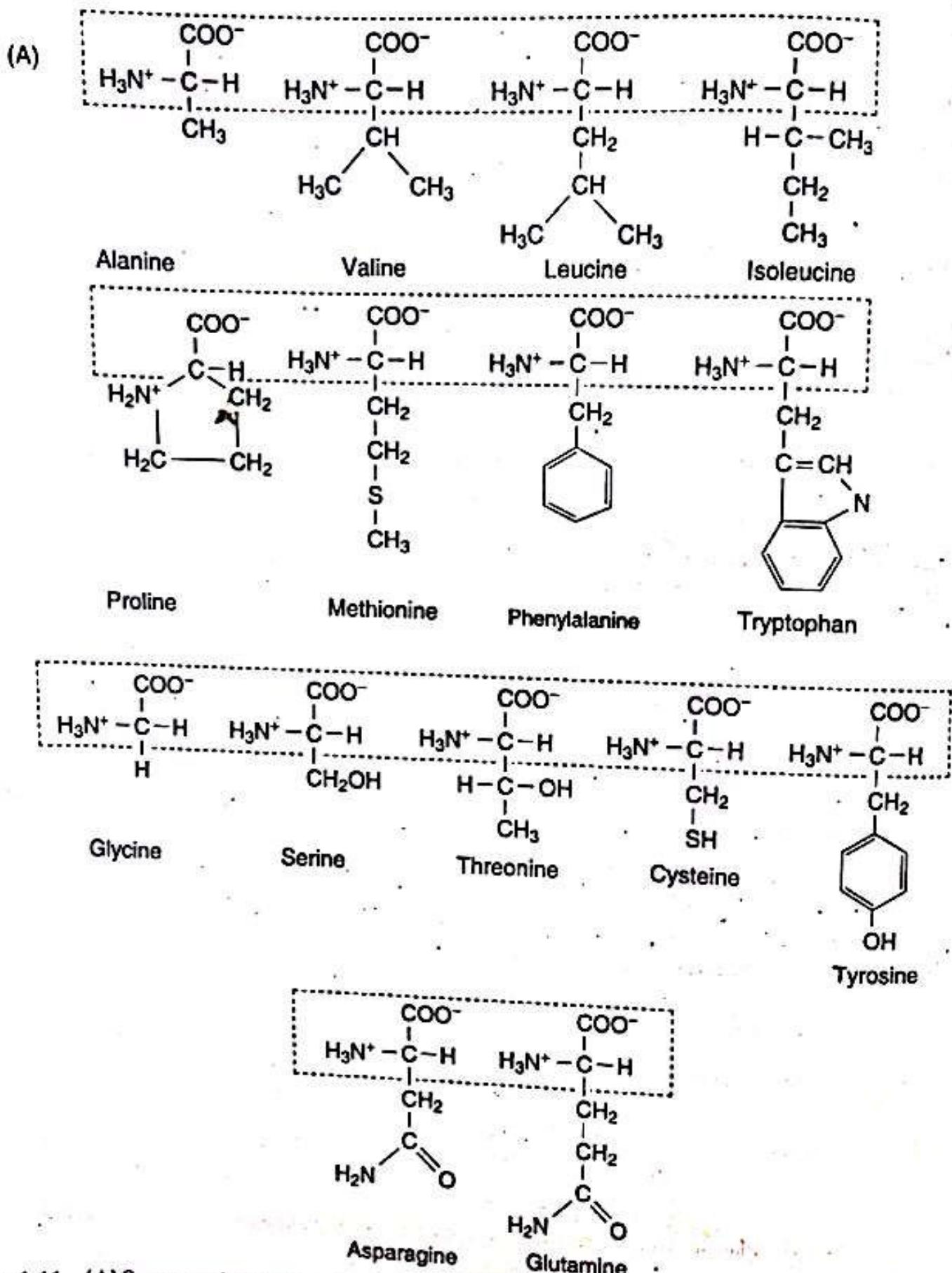


Figure 1-11 (A) Structure of nonpolar amino acids at pH 7.0. (B) Structure of polar amino acids at pH 7.0. Portion outside the box is the sidechain.

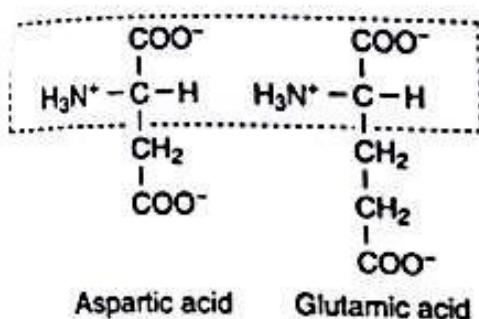
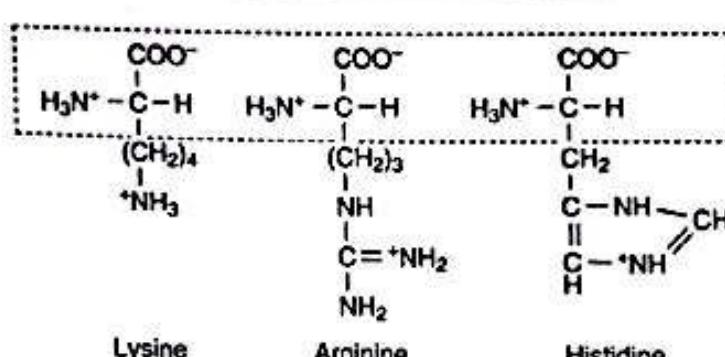


Figure 1-12 Structure of negatively charged amino acids at pH 7.0. Portion outside the box is sidechain.



**Figure 1-13** Structure of positively charged amino acids at pH 7.0. Portion outside the box is the sidechain.

Although there are 20 standard amino acids but some amino acids undergo modification after their incorporation into protein. The number of such modified amino acids is quite large. Some of the prominent of these modified amino acid residues are 4-hydroxyproline, 5-hydroxylysine, phosphoserine and N-formylmethionine.

On the basis of nutritional value, the amino acids have been categorized as essential and nonessential amino acids. In fact all amino acids are essential as building blocks of proteins. However, the term essential means the amino acids which cannot be synthesized by an organism or cannot be synthesized at an adequate rate or are in insufficient quantity and are obtained through diet. Adults require nine (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan and valine) essential amino acids, while infants and growing children require 10 amino acids (the extra amino acid is arginine).

## 5.2 Physical Properties

### **5.2.1 Solubility**

Amino acids contain multiple charged groups, therefore, with certain exceptions are generally soluble in polar solvents (water and ethanol), but are insoluble in nonpolar solvents (chloroform, benzene, hexane and ether). The high melting point ( $>200^{\circ}\text{C}$ ) also reflects the high energy needed to disrupt the ionic forces.

### **.5.2.2 Optical Isomerism**

All amino acids except glycine have an asymmetric  $\alpha$ -carbon, therefore, they exist in two stereoisomeric forms (D and L isomers) depending on whether the spatial structure is identical to that of D-glyceraldehyde or L-glyceraldehyde, and show optical activity.

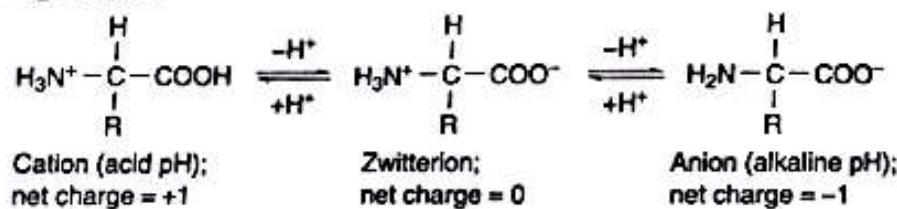
### **1.5.2.3 Absorption in Ultraviolet Region**

All amino acids show an appreciable absorption in the far ultraviolet (<220 nm) region. However, some of them absorb between 250 and 300 nm owing to the presence of chromophore like the phenyl radical (tyrosine) or the indole ring (tryptophan), permitting the rapid estimation of protein spectrophotometrically.

#### 1.5.2.4 Acid-Base Behavior

Amino acids contain ionizable functional groups, therefore, the net charge of the molecule is a function of pH. Both the  $\alpha$ -amino and  $\alpha$ -carboxyl groups are protonated at low pH giving net charge +1, but at high pH both groups lose a proton giving a molecule

net-1 charge.  $\alpha$ -carboxyl group is a stronger acid than amino group; it first loses its proton as the pH increases. From this, it may be concluded that an intermediate form of amino acid is produced which has both a negative and a positive charge. The intermediate compound is known as Zwitterion (dipolar form). Zwitterion is amphoteric in nature and the net charge is zero.



#### 1.5.2.5 Isoelectric Point

The pH at which a molecule has zero net charge is known as isoelectric point (pI). All amino acids, peptides have net positive charge at pH below the pI and a net negative charge at a pH above pI. The solubility is minimal and do not migrate in an electric field.

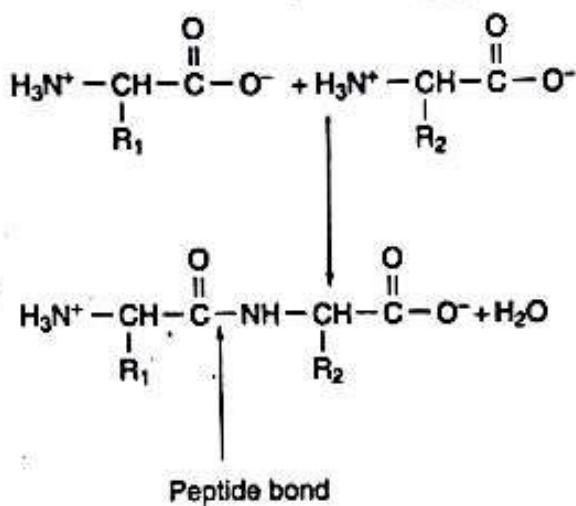
The dissociation of various amino acids may be studied by adding HCl or NaOH to the solution and measuring the pH after each addition. Plots (titration curves) of varying shapes are obtained depending on whether the amino acid is acidic, basic or neutral.

### 1.5.3.3 Reactions of the R Group

The sidechains (R group) of amino acids show qualitative reactions. The thiol group of cysteine can be oxidized and cysteine–cystine system can act as hydrogen donor or acceptor. The disulfide bridge is important for three-dimensional structure of proteins. Hydroxyl group of tyrosine can be esterified by phosphoric acid to yield phosphotyrosine.

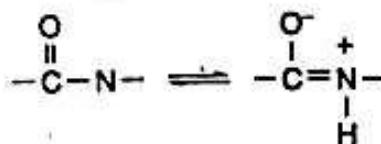
### 1.5.4 Peptides

Peptides and proteins are linear sequence of amino acids joined together by peptide bonds. It is a covalent bond formed between the  $\alpha$ -amino group of one amino acid and the  $\alpha$ -carboxyl group of other amino acid to yield a dipeptide:

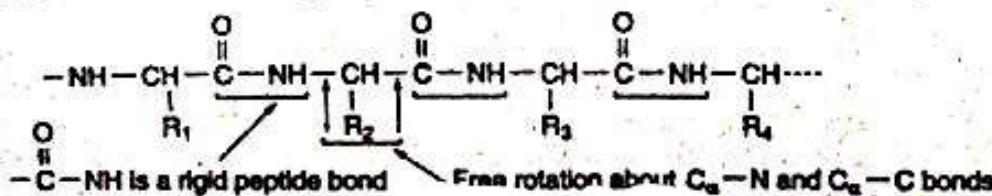


Once two amino acids join to yield dipeptide, it has got a free amino group at one end and free carboxyl group at the other end. These groups can in turn be linked to other amino acids by a peptide bond to yield a long unbranched chain of amino acids to form oligopeptides (upto 2–10 amino acid residues) and polypeptide ( $> 10$  amino acid residues). Since in polypeptides the  $\alpha$ -carbon exists as tetrahedron, therefore, the carbon backbone of these chains exists as a zig-zag structure.

The peptide bond exists as a resonance hybrid because the carbon and nitrogen exhibit partial double bond character due to the closeness of the carboxyl carbon–oxygen double bond. The configuration of atoms around double bond exists in two geometrical isomers (cis and trans). The peptide bond is made up of CO-NH, which is relatively rigid and planar, as the C–N bond length is shorter than normal C–N single bond, although free rotation can take place about  $\text{C}_{\alpha}\text{-N}$  and  $\text{C}_{\alpha}\text{-C}$  bonds, permitting adjacent peptide



units to be at different angles. The hydrogen of the amino group will be located *trans* to the oxygen of the carbonyl group rather than *cis*. The polypeptide chain can form different conformation in the protein and can form different structures of proteins.

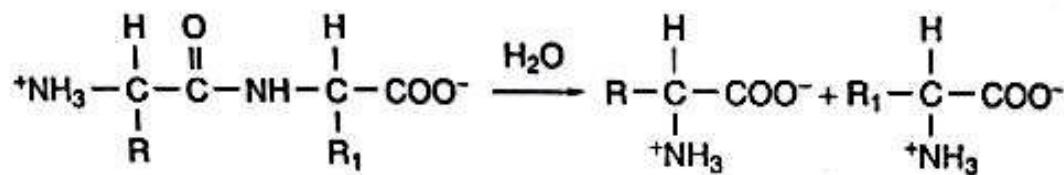


The polypeptide chain folds up to form different conformations in the protein giving different structures. Four levels of structures – primary, secondary, tertiary and sometimes quaternary – have been proposed.

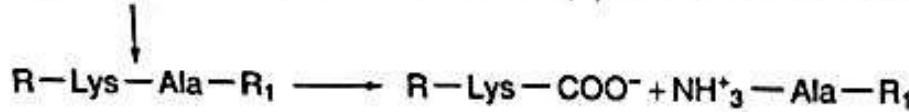
#### 1.5.4.1 Physical Properties of Peptides

Peptides have only one free  $\alpha$ -amino and one free  $\alpha$ -carboxyl at the terminal end. As in amino acids, these groups also ionize in peptides. They have characteristic titration curves and isoelectric pH.

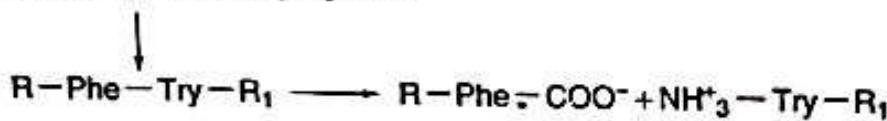
Peptides show characteristic properties of free amino, carboxyl and also of sidechain (R group). Peptide bonds are hydrolyzed by boiling with strong acid or base to yield free amino acids:



Peptide bonds can also be hydrolyzed by proteases, for example trypsin and chymotrypsin. Trypsin hydrolyzes peptide bonds on the carboxyl side of lysine and arginine:



Chymotrypsin hydrolyzes peptide bonds on the carboxyl side of aromatic residues (phenylalanine, tyrosine and tryptophan):



The other important reactions of peptides are with FDNB, dansyl chloride and phenylisothiocyanate. These chemicals are used in determining the N-terminal of polypeptides.

### 1.6 Heterocyclic Compounds and Secondary Metabolites

Heterocyclic compounds are cyclic compounds with the ring containing carbon and other elements, viz. nitrogen, sulfur and oxygen. Important amongst these are five- and six-membered heterocyclic compounds. Purine is a bicyclic compound of a six- and five-membered heterocyclic ring fused together, involving carbon atoms 4 and 5. However, secondary metabolites are specialized products (nucleotides, pigments, hormones, alkaloids, terpenoids and coenzymes) which are required in minute quantities by the cell and synthesized by specialized secondary pathways which are not so well understood.

#### 1.6.1 Purine and Pyrimidine Bases

There are three major pyrimidines – cytosine (C), thymine (T) and uracil (U) – and two major purines – adenine (A) and guanine (G) – in nucleic acids. Both pyrimidines and purines are planar due to the presence of conjugated double bonds and are weak bases. All these bases undergo (keto-enol tautomerism, involving the shift of a hydrogen). Owing to their aromatic character they absorb ultraviolet light. However, the absorption spectra vary with pH.

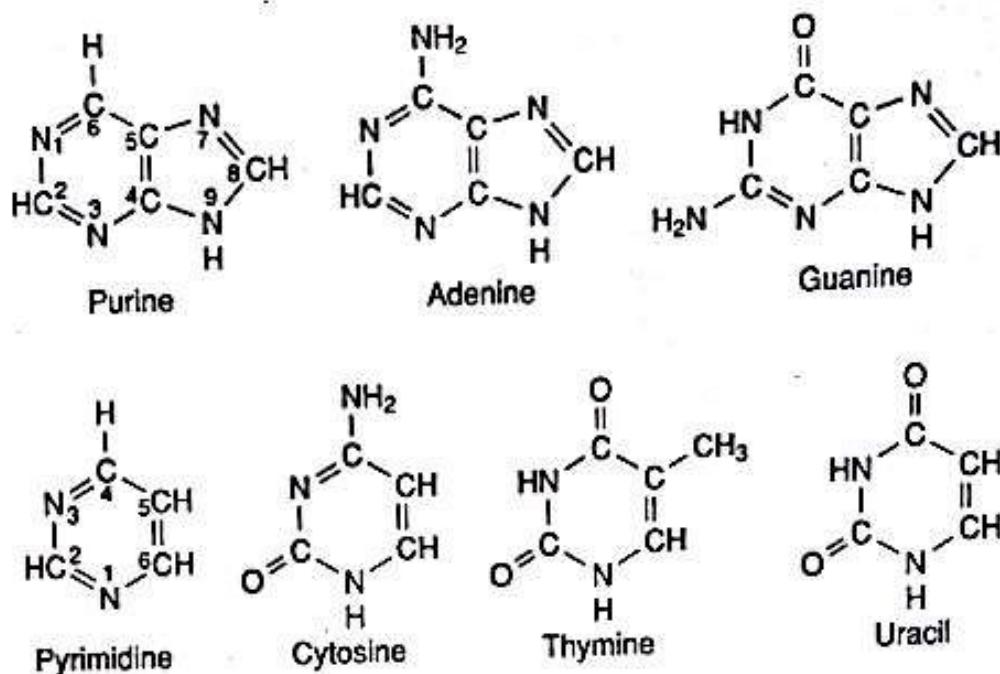


Figure 1-14 Major purines and pyrimidines in DNA and RNA.

#### 1.6.1.1 Nucleosides

Sugar and a base combines through a  $\beta$ -glycosidic bond involving anomeric carbon of sugar and N(9) of purine or N(1) of pyrimidine. AZT (3'-azido-2',3' dideoxy thymidine), a synthetically produced structural analog of dideoxy thymidine, is used to treat AIDS.

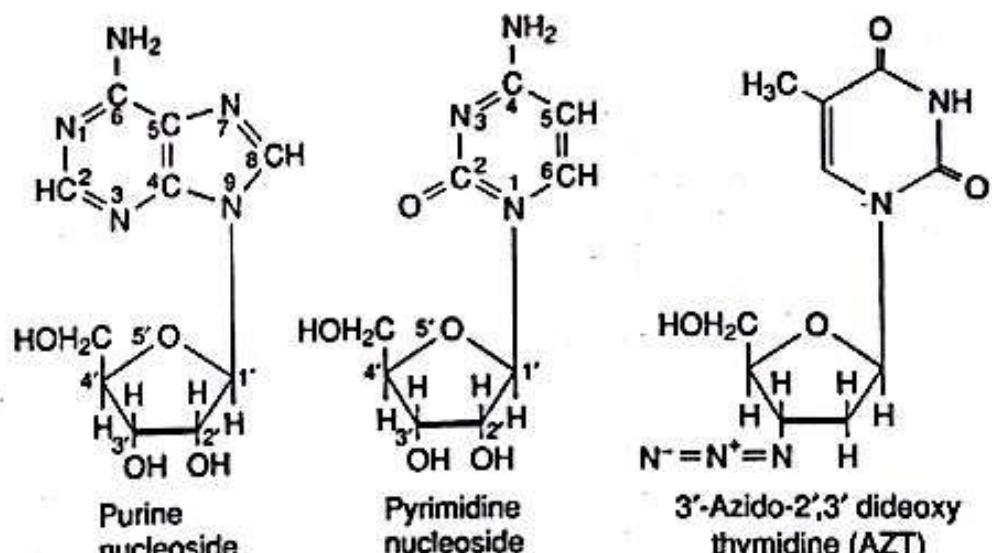
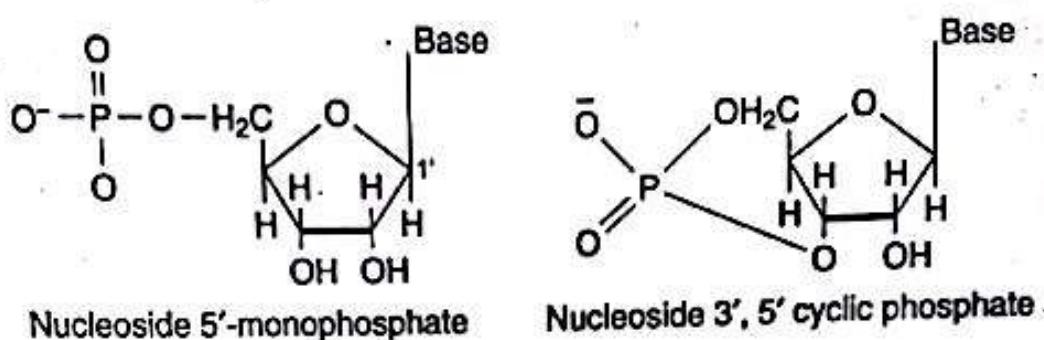


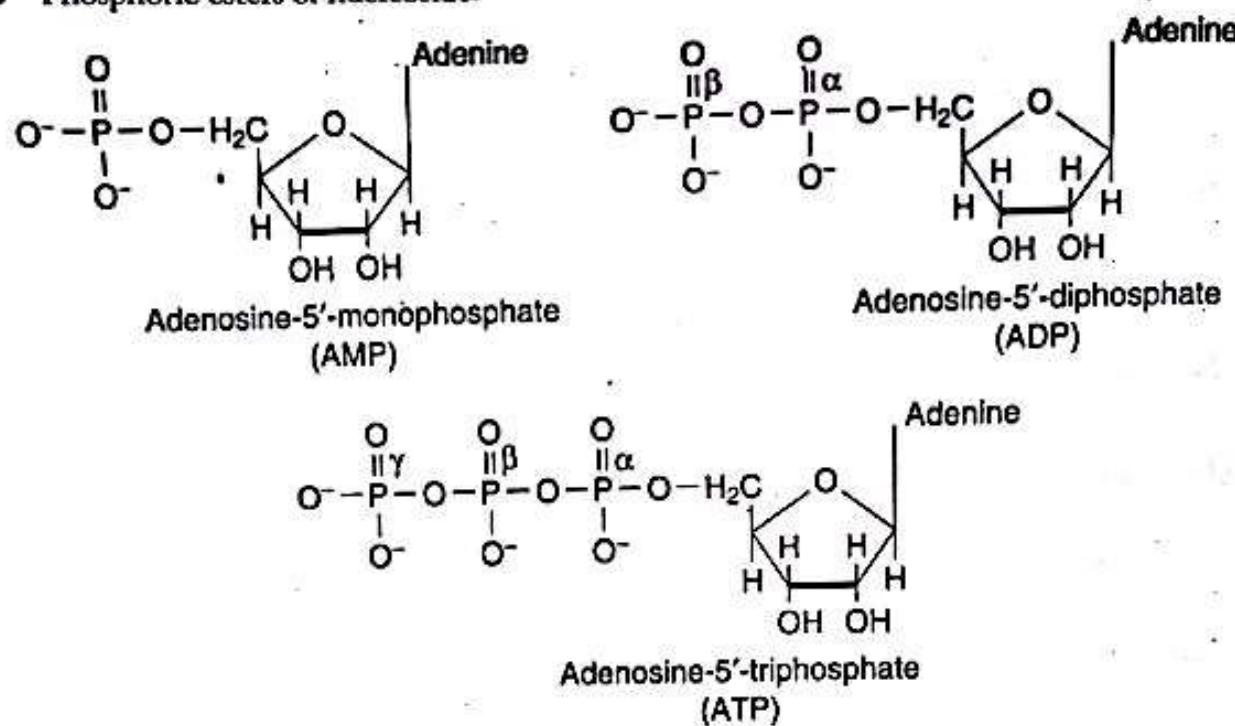
Figure 1-15 Structures of nucleosides (in case of deoxyribose OH at 2' is substituted by H) and AZT.

#### 1.6.1.2 Nucleotides

These are the phosphoric esters of nucleosides. Depending on the nature of sugar will get ribonucleotides and deoxyribonucleotides. The ribonucleoside has three positions (2', 3' and 5') which can be phosphorylated, while a deoxyribonucleoside can be phosphorylated at two positions (3' and 5'). Ester formation can occur at more than one position on the same ribonucleoside or deoxyribonucleoside. Esterification can also occur at two positions on sugar molecule yielding a cyclic phosphodiester. Besides purine and pyrimidine nucleotide, both DNA and RNA contain minor purine and pyrimidine bases.



**Figure 1-16** Phosphoric esters of nucleoside.



**Figure 1-17** Nucleotides formed by joining phosphate groups.

Phosphate group can be extended by anhydride linkages so that two or even three phosphate groups are linked at one position in the sugar molecule in the same nucleoside.

Besides these nucleotides a number of compounds that have same chemical parts of nucleotide in their structure, for example coenzymes nicotinamide adenine nucleotide (NAD), nicotinamide adenine dinucleotide phosphate (NADP) and flavin adenine dinucleotide (FAD).

Nucleoside-5'-diphosphates and 5'-triphosphates are strong acids, and form complexes with divalent metal  $\text{Mg}^{2+}$  and  $\text{Ca}^{2+}$ . ATP is a carrier of phosphate and pyrophosphate in many enzymatic reactions involved in energy transfer. ADP and UDP are specific carriers of sugar residues in the biosynthesis of polysaccharides. Nucleotide triphosphates serve as energy-rich precursors of mononucleotide units in the biosynthesis of nucleic acids.

### 1.6.2 Isoprenoides

Isoprenoid compounds are widely distributed in plant kingdom, with a wide variety of structures and functions. The high levels of some of these compounds present in turpentine oil gave them an alternative name – terpenoids; the two terms are used interchangeably. It is difficult to cover all aspects of the chemistry and biochemistry of terpenoids as they are widely distributed in large number of flowering plants. The thermal decomposition of almost all terpenoids gives isoprene, made up of five carbons, as one of the products. This led to the suggestion that all naturally occurring terpenoids are made

## Lipids – Classification, Structure and Function

Lipids are heterogenous organic compounds insoluble or sparingly soluble in water, but soluble in nonpolar solvents such as chloroform, ether and benzene. These solubility criteria are not absolute. Lipids were therefore, defined as compounds containing in their molecule an aliphatic chain (consisting of  $-\text{CH}_2-$ ) of at least eight carbon atoms. However, there are exceptions to this rule. Lipids are important dietary constituents, not only because of their high energy value, predominantly in the form of triacylglycerides, but also because both lipids and lipid derivatives serve as vitamins and hormones and provide essential fatty acids present in the fat of natural foods. Lipids also serve as

components of biological membranes and also has several protective roles in animals and birds.

## 7.1 Classification of Lipids

Lipids can be classified into three major categories:

1. Simple lipids: Esters of fatty acids with various types of alcohols.
2. Complex lipids: Esters of fatty acids with alcohols complexed with a non-lipid moiety.
3. Derived lipids: This includes diverse group of compounds (prostaglandins, fat-soluble vitamins, steroids, etc.).

The principal component of most lipids are the fatty acids which have a long hydrocarbon chain with a terminal carboxylic group at one end. Most of the fatty acids have an even number of carbon atoms arranged in an unbranched chain. The fatty acids have been further classified as saturated and unsaturated. Saturated fatty acids have no double bonds between the carbon atoms and have general formula  $\text{CH}_3(\text{CH}_2)_n\text{COOH}$ , where  $n$  is an even number. The most common fatty acids present in all animal and plant fats are palmitic acid,  $\text{CH}_3(\text{CH}_2)_{14}\text{COOH}$ , and stearic acid,  $\text{CH}_3(\text{CH}_2)_{16}\text{COOH}$ . Longer chain fatty acids, upto 36 carbon atoms, are present in numerous cells. Saturated fatty acids with less than eight carbon atoms are liquid at physiological pH, whereas those which contain more than ten carbon atoms are solid. Small amounts of branched and even cyclic fatty acids also occur in nature.

Unsaturated fatty acids contain one or more double bonds. They are numbered from the terminal carboxyl (C1) to the  $\text{CH}_3$  group. The double bond is indicated by the sign  $\Delta$ , accompanied by the number corresponding to first carbon atom participating in the double bond formation. Double bond may also be indicated by  $\omega$ . Unsaturated fatty acids can exist as geometric isomers (cis or trans conformations). However, in all biological unsaturated fatty acids, the double bond has a cis-configuration. The unsaturated fatty acids may further be classified as: monounsaturated for example palmitoleic and oleic acid have one site of unsaturation. Palmitoleic acid has 16 carbon while oleic acid has 18 carbon fatty acid with one unsaturation site between carbons 9 and 10.

Palmitoleic acid:  $-\text{CH}_3(\text{CH}_2)_5\text{CH}=\text{CH}(\text{CH}_2)_7\text{COOH}$  (C16 : 1 $\Delta^9$ ) or (16 $\omega^9$ )

Oleic acid:  $-\text{CH}_3(\text{CH}_2)_{17}\text{CH}=\text{CH}(\text{CH}_2)_7\text{COOH}$  (C18 : 1 $\Delta^9$ ) or (18 $\omega^9$ )

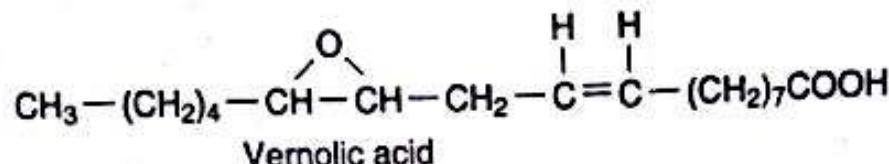
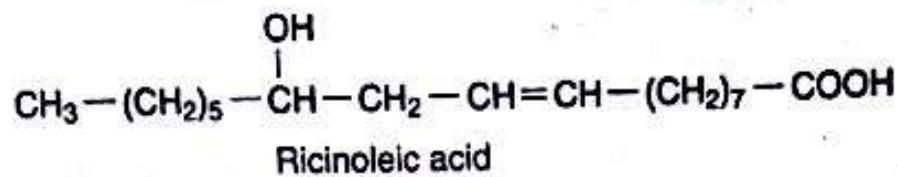
Polyunsaturated fatty acids have two or more double bonds which are separated at least by one methylene group.

Linoleic acid:  $\text{CH}_3(\text{CH}_2)_4\text{CH}=\text{CHCH}_2\text{CH}=\text{CH}(\text{CH}_2)_7\text{COOH}$  (C18 : 2 $\Delta^{9,12}$ )

Linolenic acid:  $\text{CH}_3\text{CH}_2\text{CH}=\text{CHCH}_2\text{CH}=\text{CHCH}_2\text{CH}=\text{CH}(\text{CH}_2)_7\text{COOH}$   
(C18 : 3 $\Delta^{9,12,15}$ )

Chain length and the number of double bonds are important in determining the properties of unsaturated fatty acids. Shorter the chain length, lower the melting point.

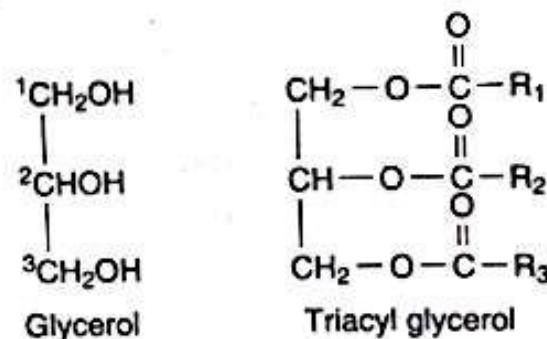
Plants can synthesize a number of unusual fatty acids, for example ricinoleic acid and vernolic acid.



## 1.1 Simple Lipids

In living cells the concentration of free fatty acids is generally low; they mostly occur as esters which constitute parts of simple and complex lipids. In simple lipids, esters of fatty acids with various alcohols are formed. If the alcohol is glycerol, we call the ester as neutral fats or simply fats, if the alcohol has long chain of hydrocarbon, instead of glycerol, the esters formed are known as waxes.

**Glycerides:** Glycerides or acylglycerol are formed by the esterification of alcohol groups of glycerol by fatty acids, these are mono-, di- and tri-acylglycerides depending upon the number of OH groups of glycerol esterified. Acylglycerols have further been classified as simple triacyl glycerol which contain same kind of fatty acid at all the three position, for example tristearin contain three stearic acid where as triolein has three oleic acid residues. Mixed triacylglycerols contain two or three different types of fatty acid residues and are named according to their attachment on the glycerol molecule.

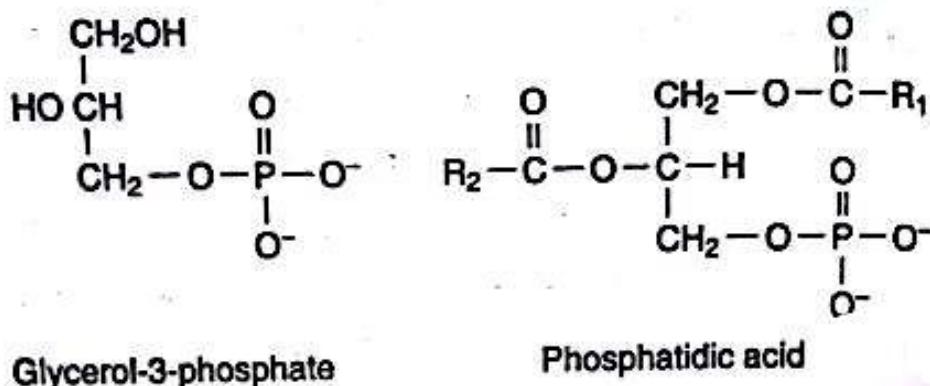


**Waxes:** In general waxes are esters of long-chain saturated or unsaturated fatty acids ( $\text{C}_{14}$  to  $\text{C}_{36}$ ) with long chain alcohols ( $\text{C}_{16}$  to  $\text{C}_{22}$ ). These are chemically inert. They serve as protectant to plant from water loss and from abrasive damage. Waxes also play an important role in providing a water barrier for insects, birds and animals.

### 1.7.1.2 Complex lipids

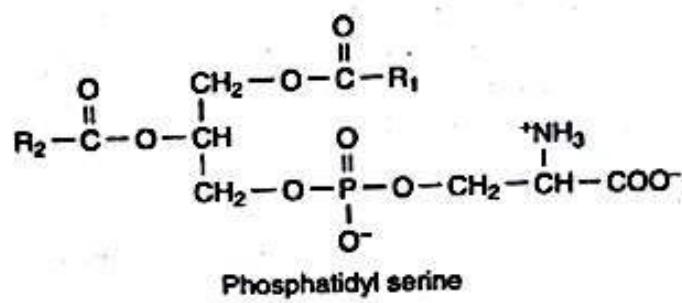
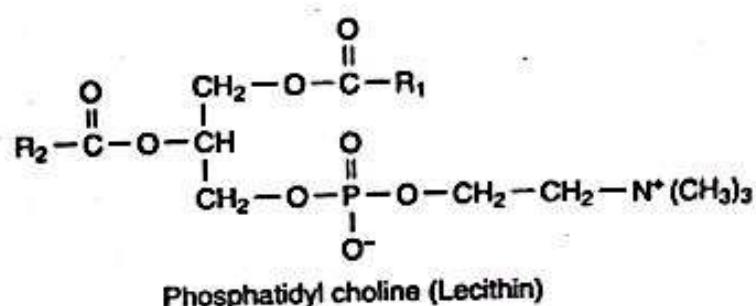
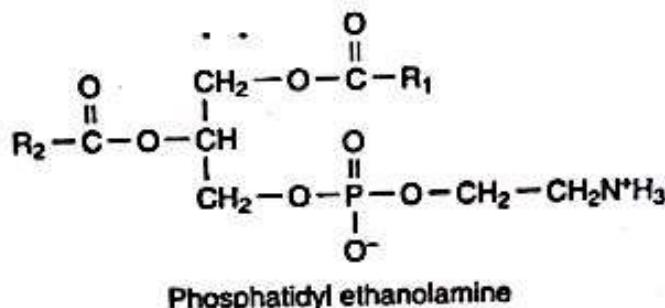
The class consists of a lipid moiety and a non-lipid component that may be a small molecule or a macromolecule. Major classes of complex lipids are glycerophospholipids, sphingophospholipids, glycolipids, lipoproteins and proteolipids.

**Glycerophospholipids:** These are found in high concentrations in the cellular and subcellular membranes of all living organisms, in which two fatty acids are attached in ester linkage to C-1 and C-2 of glycerols and a high polar group is attached to the C-3 (stereospecific number 3) through a phosphodiester linkage. The main building block of phospholipid is phosphatidic acid.



Hydrogen in phosphatidic acid can be substituted by a number of alcohols. The common ones are:

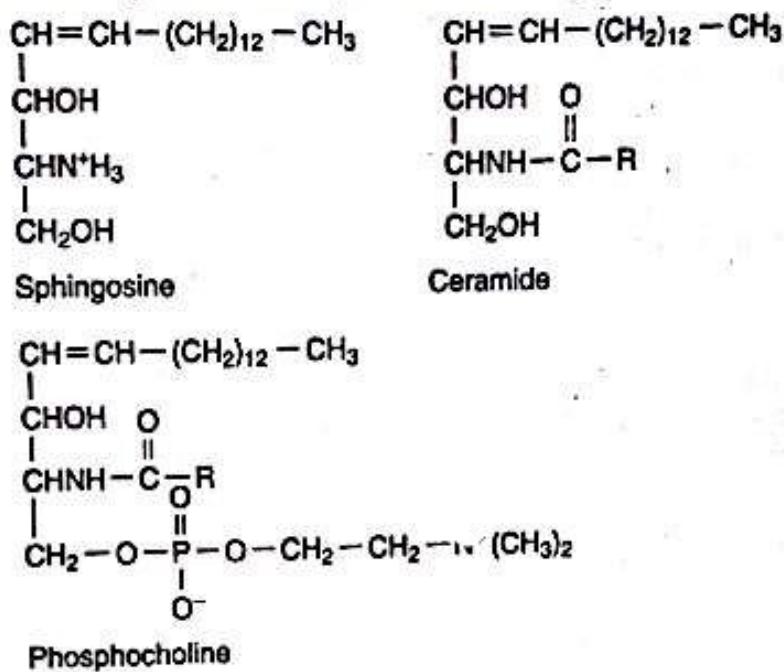
- (i) Ethanol amine:  $\text{HO}-\text{CH}_2-\text{CH}_2-\text{N}^+\text{H}_3$ , which yields phosphatidylethanolamine
- (ii) Choline:  $\text{HO}-\text{CH}_2-\text{CH}_2-\text{N}^+(\text{CH}_3)_3$ , the substitution yields phosphatidyl choline, also called lecithins.
- (iii) L-serine:  $\text{HO}-\text{CH}_2-\underset{\text{COO}^-}{\text{CH}}-\text{N}^+\text{H}_3$ . the substitution yields phosphatidylserine.



Myo-inositol (phosphatidyl inositol) can have a number of inositol alcohol that on phosphorylation yields polyphosphatidyl inositols, phosphatidyl glycerol, diphosphatidyl glycerol, which are commonly known as cardiolipins.

**Sphingophospholipids:** These are membrane lipids, which are composed of one molecule of sphingosine, a long chain amino alcohol, one molecule of long chain fatty acid and a polar head group, joined together by glycosidic linkage in some cases or by a

phosphodiester linkage. When a fatty acid is attached in amide linkage the resulting compound is ceramide. Sphingo phospholipids are formed from ceramide by esterification of phosphoric acid by either serine, ethanol amine or choline.



**Glycolipids:** Glycolipids are the derivatives of carbohydrate and glyceride and do not contain phosphate. A common example is galactosyldiacylglycerol found in higher plants and also in neural tissue of vertebrates. Glycolipids containing di and trisaccharides also exist. Gangliosides constitute oligosaccharide derivative of ceramide that contain at least one residue of N-acetyl neuraminic acid.

**Lipoprotein:** Lipoproteins consist of lipid components, triacyl glycerol, phospholipids and cholesterol or cholesterol ester. The protein components have a relatively high proportion of non-polar amino acid residues responsible for binding of lipids. Plasma proteins have been classified on the basis of their density in to five broad categories, chylomicrons, very low density lipoproteins (VLDL), intermediate density lipoproteins (IDL), low density lipo proteins (LDL) and high density lipoproteins (HDL).

**Plasmalogens:** These are phospholipids with ether linked fatty acids. These are present in animal tissues and unicellular organisms. The major difference between phospholipids and plasmalogens is that the fatty acid at C-1 (Sn1) of glycerol contains either an O-alkyl or O-alkenyl ether species. The most potent biological molecule is platelet activating factor (PAF) which is a choline plasmalogen in which C-2 (Sn2) position of glycerol is esterified with an acetyl group in place of a long chain fatty acid.

### 1.7.1.3 Derived Lipids

**Lipids**  
Derived lipids constitute compounds of diverse groups which include prostaglandins, isoprenoids, fat soluble vitamins (A, D, E and K) and steroids.

## 1.7.2 Functions of Lipids

**Functions of Lipids**  
Lipids participate directly or indirectly in several metabolic activities of animals as well as plants. They serve as a major source of energy in animals and plants. Triglycerides are broken down to generate energy, they also protect animals from mechanical injury. Triglycerides are stored in oil seeds, provides energy and biosynthetic precursors during germination. Derivatives of fatty acids such as prostaglandins, thromboxanes and

The minimum molecular weight of a protein is that of adrenocorticotropin hormone (4500)\*, insulin (bovine insulin - 5733) and bacterial ferredoxin (about 6000). Other common proteins are human haemoglobin (66500), enzyme urease (483000), iso-citrate dehydrogenase (1,000,000) and pyruvate dehydrogenase complex (4,600,000). Being macromolecules proteins are not freely soluble in water but form colloidal complex with the same. Chemically a protein is made of carbon, hydrogen, nitrogen, oxygen and sulphur. Some proteins additionally contain phosphorus, iron and other elements.

Proteins are variously folded linear polymers of amino acids. The linear polymers of amino acids are called polypeptides. A protein may have one polypeptide (monomeric proteins, e.g., myoglobin, ribonuclease), two polypeptides (e.g., insulin) or more (24 in ribulose biphosphate carboxylase and 72 in pyruvate dehydrogenase complex). A protein having two or more polypeptides is called oligomeric protein. A polypeptide contains from a few (22 and 31 in the two polypeptides of human insulin, 21 and 30 in bovine insulin, 97 in spinach ferredoxin) to a few hundred amino acid residues (582 in human serum albumin). The amino acids are linked serially by peptide bonds ( $-\text{CONH}-$ ) formed between amino group ( $-\text{NH}_2$ ) of one amino acid and carboxylic group ( $-\text{COOH}$ ) of the adjacent one. The sequence of amino acids present in a polypeptide is specific for a particular protein. The distinctive sequence of amino acid units is governed by the codon sequence of the gene or cistron that controls its formation. Only some 20 amino acids are used in the synthesis of all types of proteins. This is similar to the formation of innumerable words from a limited number of alphabets. In a polypeptide of only 100 amino acid residues, there is possibility of  $20^{100}$  arrangements or types of polypeptides. This accounts for thousands of specific proteins found in each of the living species as well as those of the past species.

### Organisation (Levels of Structure)

A protein can have upto four levels of organisation— primary, secondary, tertiary and quaternary.

**1. Primary Structure.** It is the description of basic structure of a protein. This includes number of polypeptides, number and sequence of amino acids in each polypeptide. The latter is determined genetically (by DNA) through transcription and translation. Specific amino acids determine the places where polypeptides are to bend or fold and where the different lengths will be attracted to each other. The distance between two adjacent peptide bonds is about 0.35 nm.

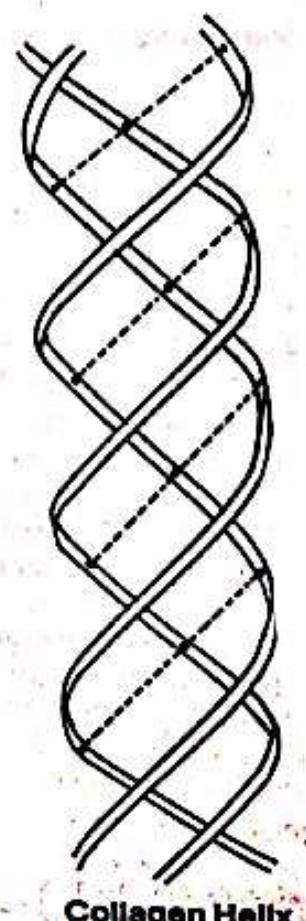
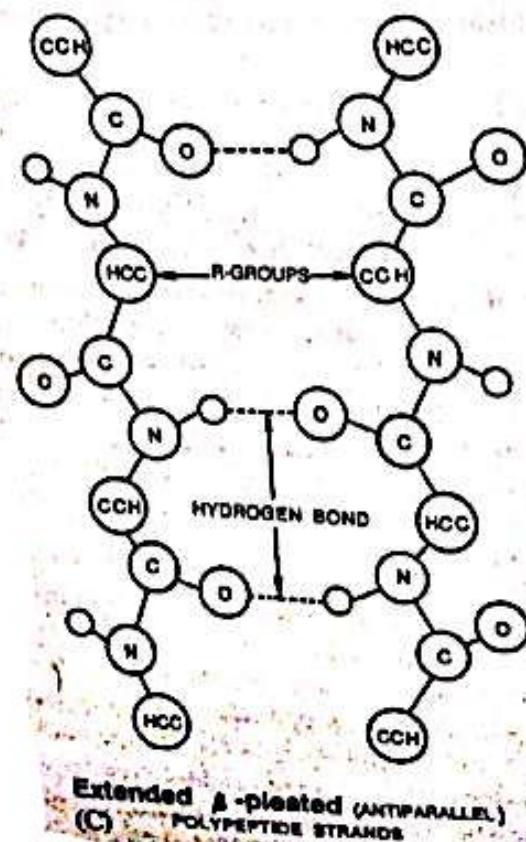
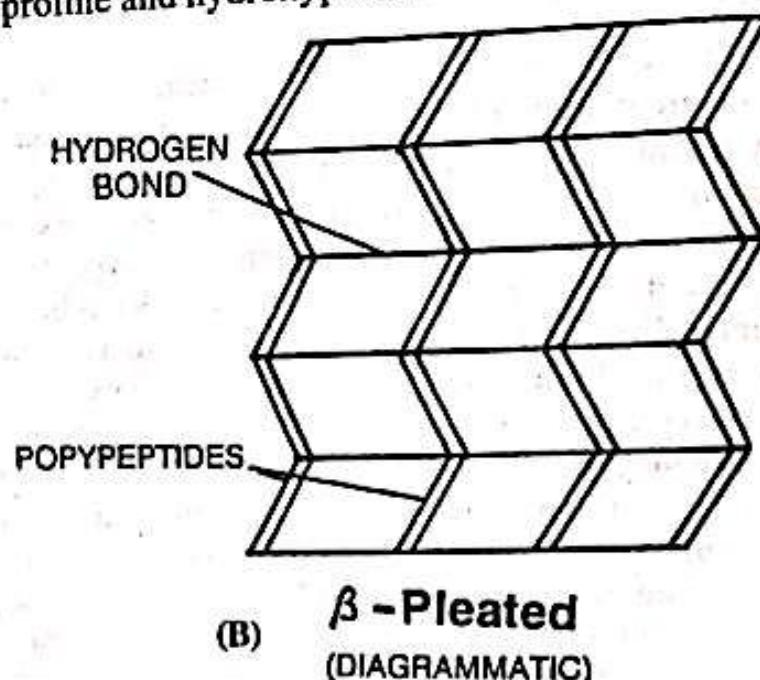
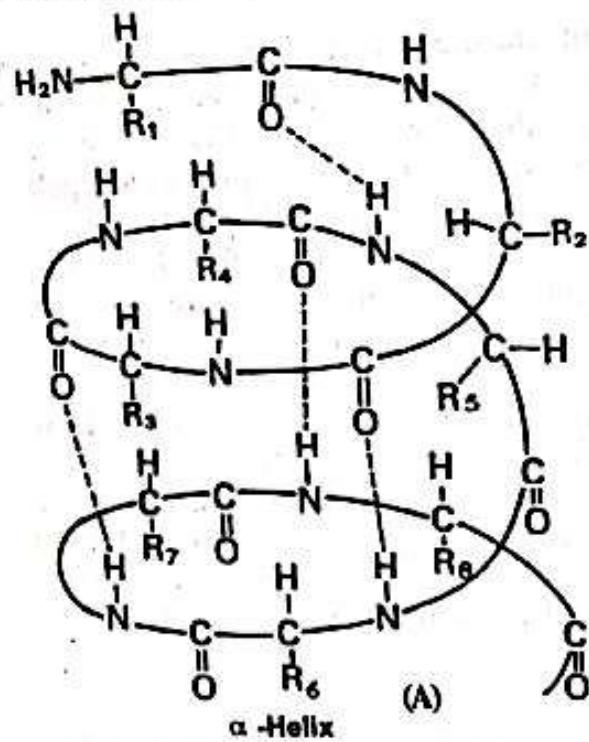
**2 Secondary Structure.** It is the development of new steric relationships of amino acids present in the linear sequence inside the polypeptides. Some of the new relationships are of regular nature and give periodicity to the structure. There are three types of secondary structures—  $\alpha$ -helix,  $\beta$ -pleated and collagen helix. The  $\alpha$  and  $\beta$  terms simply designate the first and the second type of secondary structures discovered in proteins. In  $\alpha$ -helix the polypeptide chain is coiled spirally, generally in right handed manner. At places the helix is less regular, forming random coils. The helix is stabilized by hydrogen bonds between oxygen of carboxylic group ( $-\text{CO}$  group) of one amino acid residue and  $>\text{NH}$  group of next fourth amino acid residue. Actually all the main chain  $-\text{CO}$  and  $>\text{NH}$  groups are hydrogen bonded.  $\alpha$ -helical coiled secondary structure is found in several proteins, e.g., keratin (hair), myosin, tropomyosin (both muscles), epidermin (skin), fibrin (blood clot). Two or more polypeptides can further coil around each other to form cables. This gives helical strand.

In  $\beta$ -pleated secondary structure two or more polypeptide chains get interconnected by hydrogen bonds. A sheet is produced instead of a fibre or rod in  $\alpha$ -helix. Therefore, this secondary structure is often called pleated sheet or  $\beta$ -pleated sheet. Adjacent strands of polypeptides may run in the same direction (parallel  $\beta$ -sheet, e.g.,  $\beta$ -keratin) or in opposite directions (antiparallel  $\beta$ -sheet, e.g., fibroin of silk). In some cases single polypeptide may show  $\alpha$ -helix in some

\*4600 according to one calculation.

portion and bent to form two or more parallel strands with  $\beta$ -pleated structure in other parts, e.g., ribonuclease.  $\beta$ -pleated proteins are more extended than the ones having  $\alpha$ -helix.

In collagen or tropo-collagen helix there are generally three strands or polypeptides coiled around one another. The coil is strengthened by the establishment of hydrogen bond between  $>\text{NH}-$  group of glycine residue of each strand with  $-\text{CO}$  group of the other two strands. There is also a locking effect with the help of proline and hydroxyproline residues (Fig. 6.6).



3. **Tertiary Structure.** There is bending and folding of various types to form spheres, rods or fibres. It further brings new steric relationships of amino acids specially those which are far apart in the linear sequence. The active sites (e.g., polar side chains) of the protein are often brought towards the surface. Certain other side chains (e.g., hydrophobic) are brought to the interior of the protein. Tertiary structure is stabilized by several types of bonds—hydrogen bonds, ionic bonds, Van der Waal's interactions, covalent bonds, hydrophobic bonds (Fig. 6.7). Tertiary structure gives the protein a three dimensional conformation (Fig. 6.8).

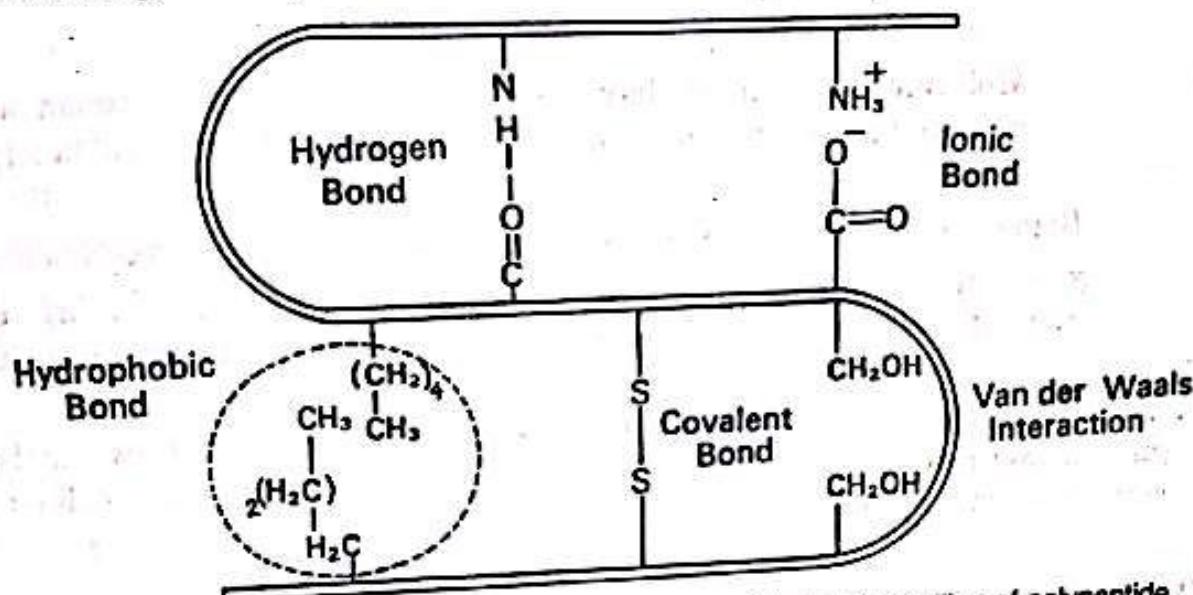


Fig. 6.7. Various types of bonds or interactions found during the coiling of polypeptide.

In protein structure, covalent bonds are the strongest. They are of two types, peptide bonds and  $\text{—S—S—}$  (disulphide) bonds. Ionic bonds or electrostatic bonds occur due to attractive force between oppositely charged ionised groups e.g.,  $\text{—NH}_3^+$  and  $\text{—COO}^-$ . Hydrogen bonds develop due to sharing of  $\text{H}^+$  or proton by two electronegative atoms. Van der Waals interactions develop with charge fluctuations between two closely placed groups (e.g.,  $\text{—CH}_2\text{OH}$  and  $\text{—CH}_2\text{OH}$ ). Hydrophobic bond is formed between two nonpolar groups. It helps in excluding water in that area and increasing compaction.

The bonds required to form tertiary structure can be easily broken by high energy radiations, high temperature, drastic changes in pH and salts of heavy metals. This process of degrading the tertiary structure is known as denaturation. Proteins can also be precipitated or coagulated by several chemicals and low temperature. In some cases removal of denaturing agent causes re-establishment of the bonds required for maintenance of tertiary structure. The phenomenon is called renaturation.

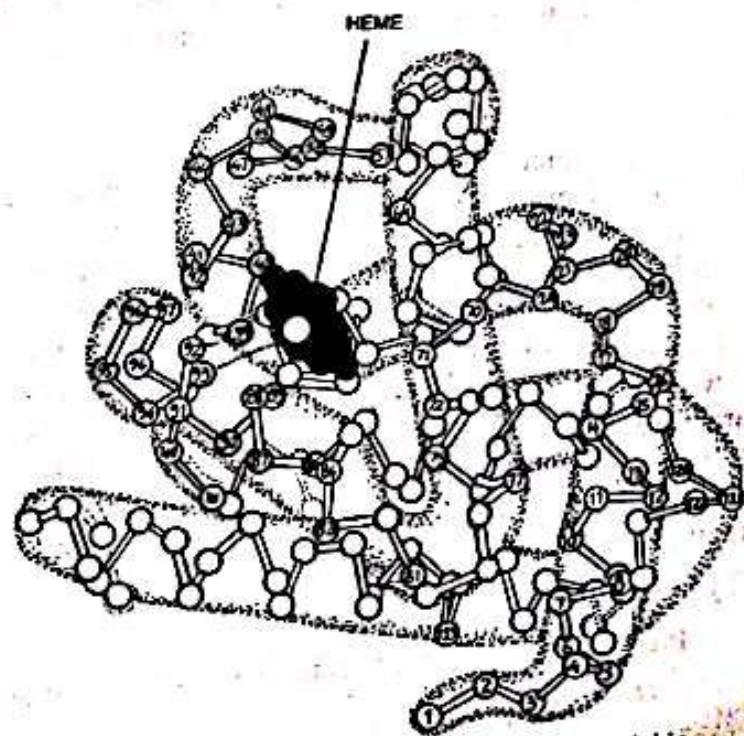


Fig. 6.8. Tertiary structure of protein myoglobin.