

## 14. BIOMOLECULES

### Can you recall ?



- What are the constituents of a balanced diet ?
- What are the products of digestion of carbohydrates?
- Which constituent of diet is useful for building muscles?
- Which constituent of diet is a source of high energy?
- What is the genetic material of organisms?

### 14.1 Introduction : Principal molecules of the living world :

Bodies of living organisms contain large number of different molecules which constitute their structure. They are also part of various physiological processes taking place in them. Primary structural materials of organisms are proteins and cellulose. By means of the unique process of photosynthesis plants produce carbohydrates. Plants utilize the minerals absorbed by their roots to produce proteins. Lipids are the main ingredient of vegetable oils and milk fats. Nucleic acids constitute the genetic material of organisms.

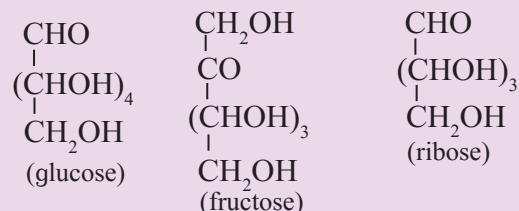
In this chapter we are going to study some aspects of three principal biomolecules, namely, carbohydrates, proteins and nucleic acids.

**14.2 Carbohydrates :** From the simple chemical reactions of many carbohydrates it is understood that carbohydrates are polyhydroxy aldehydes or ketones or compounds which give rise to such units on hydrolysis. Some carbohydrates like glucose, fructose are sweet in taste, and are called sugars. The most commonly used sugar is sucrose which is obtained from sugarcane or sugar beet. The sugar present in milk is called lactose.

### Try this...



Observe the following structural formulae carefully and answer the questions.



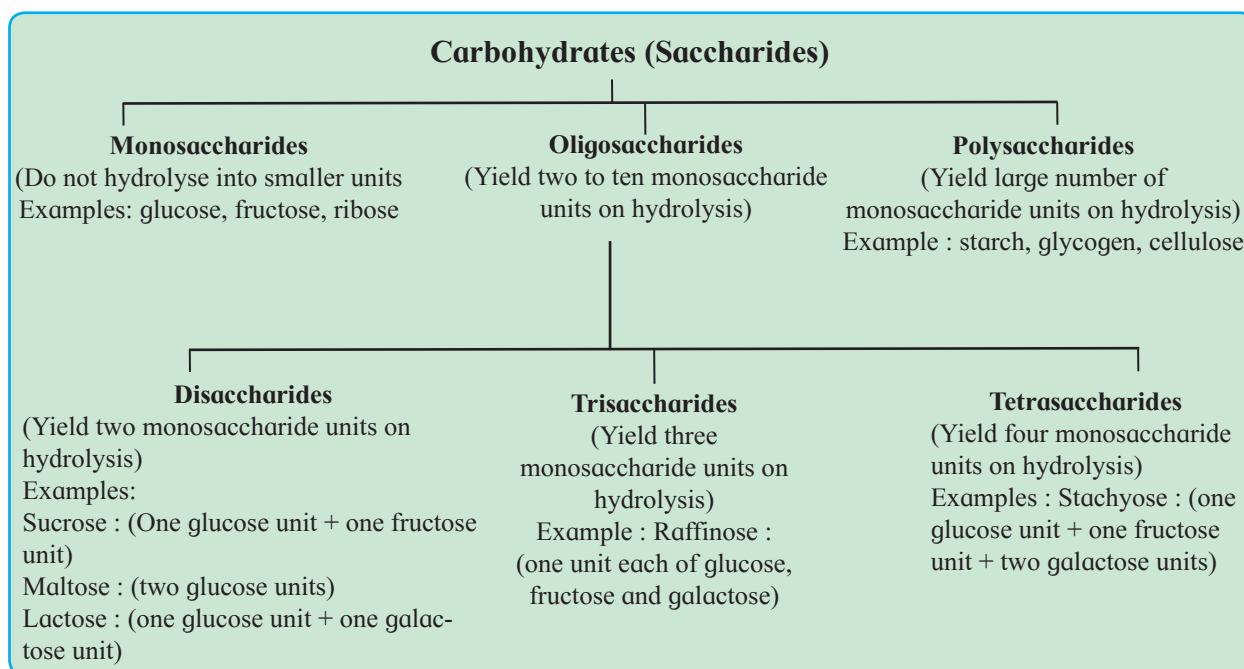
1. How many OH groups are present in glucose, fructose and ribose respectively?
2. Which other functional groups are present in these three compounds?

Greek word for sugar is sakkharon. Hence carbohydrates are also called **saccharides**. Origin of the term carbohydrate lies in the finding that molecular formulae of many of them can be expressed as C<sub>x</sub>(H<sub>2</sub>O)<sub>y</sub> (hydrates of carbon). For example: glucose (C<sub>6</sub>H<sub>12</sub>O<sub>6</sub> Or C<sub>6</sub>(H<sub>2</sub>O)<sub>6</sub>), sucrose (C<sub>12</sub>H<sub>22</sub>O<sub>11</sub> or C<sub>12</sub>(H<sub>2</sub>O)<sub>11</sub>), starch [(C<sub>6</sub>H<sub>10</sub>O<sub>5</sub>)<sub>n</sub> or [C<sub>6</sub>(H<sub>2</sub>O)<sub>5</sub>]<sub>n</sub>].

### 14.2.1 Classification of carbohydrates :

Carbohydrates are classified into three broad groups in accordance with their behaviour on hydrolysis. These are monosaccharides, oligosaccharides and polysaccharides (Fig. 14.1).

**Monosaccharides** do not hydrolyse further into smaller units of polyhydroxy aldehydes or ketones. **Oligosaccharides** on hydrolysis yield two to ten units of monosaccharides and accordingly they are further classified as **disaccharides**, **trisaccharides** and so on. **Polysaccharides** give very large number of monosaccharide units on complete hydrolysis.



**Fig. 14.1 : Classification of carbohydrates**

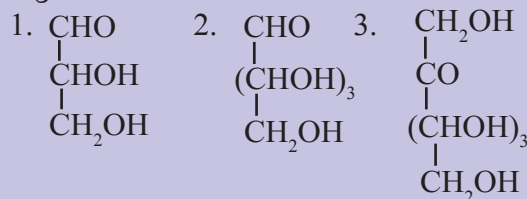
**Remember...**

- About **twenty different monosaccharides** are found in carbohydrates.
- **Disaccharides** are the most common oligosaccharides. The two monosaccharide units in disaccharides may be same or different.
- **Polysaccharides** : **Starch** is common ingredient of food grains. **Cellulose** is constituent of cell wall of plant cells. Animals store polysaccharides in their body in the form of **glycogen**.



**Use your brain power**

Give IUPAC names to the following monosaccharides.



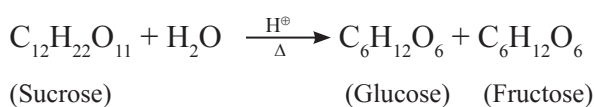
**14.2.2 Nomenclature of monosaccharides :**

According to IUPAC system of nomenclature, general name for monosaccharide is **glycose**. Monosaccharide with one aldehydic carbonyl group is called **aldose** while that with one ketonic carbonyl group is called **ketose**. These names are further modified in accordance with the total number of carbon atoms in the monosaccharide. For example, glucose ( $\text{C}_6\text{H}_{12}\text{O}_6$ ) is an aldose with six carbons, and is thereby, an **aldohexose**. Fructose ( $\text{C}_6\text{H}_{12}\text{O}_6$ ) is a ketose with six carbons, and is, thereby, a **ketoheptose**.

**14.2.3 Glucose :** Glucose occurs in nature in free as well as in combined state. Glucose can be obtained from sucrose or starch by acid catalysed hydrolysis as shown below.

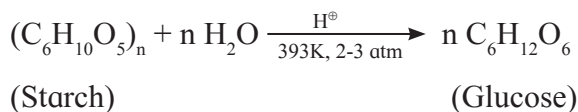
**a. Preparation of glucose from sucrose :**

Sucrose is hydrolysed by warming with dilute hydrochloric acid or sulfuric acid for about two hours. This hydrolysis converts sucrose into mixture of glucose and fructose. Glucose is separated from fructose by adding ethanol during cooling. Glucose being almost insoluble in alcohol crystallizes out first. The solution is filtered to obtain crystals of glucose.



### b. Preparation of glucose from starch :

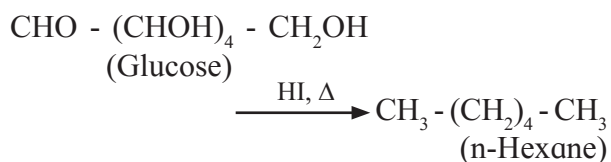
Commercially glucose is obtained by hydrolysis of starch by boiling it with dilute sulfuric acid at 393K under 2 to 3 atm pressure.



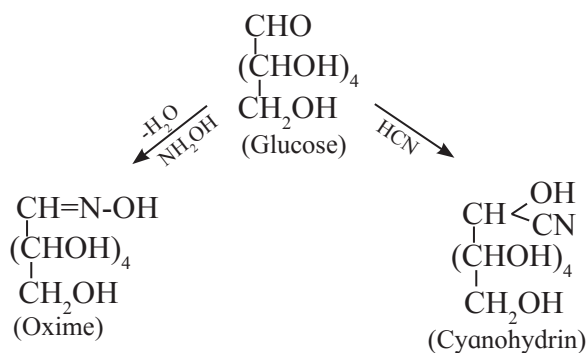
#### 14.2.4 Structure and properties of glucose

Glucose has an aldohexose structure. In other words, glucose molecule contains one aldehydic, that is, formyl group and the remaining five carbons carry one hydroxyl group (-OH) each. The six carbons in glucose form one straight chain. This aldohexose structure of glucose was established on the basis of the following **chemical properties**.

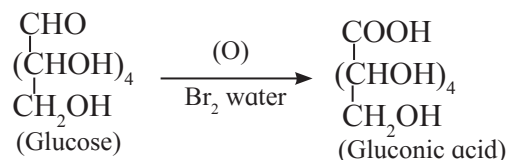
1. Molecular formula of glucose was found to be  $C_6H_{12}O_6$ , on the basis of its elemental composition and colligative properties.
2. The six carbons in glucose molecule form a straight chain. This was inferred from the following observation : Glucose gives n-hexane on prolonged heating with HI.



3. Glucose molecule contains one carbonyl group. This was inferred from the observation that glucose forms oxime by reaction with hydroxylamine and gives cyanohydrin on reaction with hydrogen cyanide.



4. The carbonyl group in glucose is in the form of aldehyde. This was inferred from the observation that glucose gets oxidised to a six carbon monocarboxylic acid called gluconic acid on reaction with bromine water which is a mild oxidizing agent.



#### Problem 14.1 :

An alcoholic compound was found to have molecular mass of 90 u. It was acetylated. Molecular mass of the acetyl derivative was found to be 174 u. How many alcoholic (-OH) groups must be present in the original compound?

**Solution :** In acetylation reaction H atom of an (-OH) group is replaced by an acetyl group (-COCH<sub>3</sub>). This results in an increase in molecular mass by [(12+16+12+3×1)-1], that is, 42 u.

In the given alcohol,

$$\begin{aligned} \text{increase in molecular mass} &= 174 \text{ u} - 90 \text{ u} \\ &= 84 \text{ u} \end{aligned}$$

$$\therefore \text{Number of -OH groups} = \frac{84 \text{ u}}{42 \text{ u}} = 2$$

#### Can you recall ?

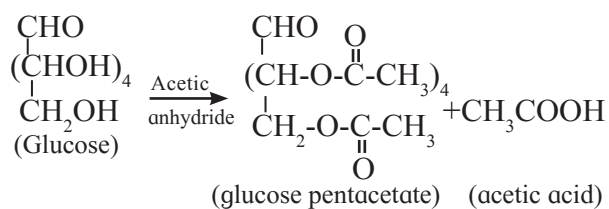
What are the products of reaction of

- i.  $\text{CH}_3 - \text{CO} - \text{CH}_3$  with  $\text{NH}_2 - \text{OH}$  ?
- ii.  $\text{CH}_3 - \text{CHO}$  with  $\text{HCN}$ ?
- iii.  $\text{CH}_3 - \text{OH}$  with  $\text{CH}_3 - \text{CO} - \text{O} - \text{CO} - \text{CH}_3$ ?

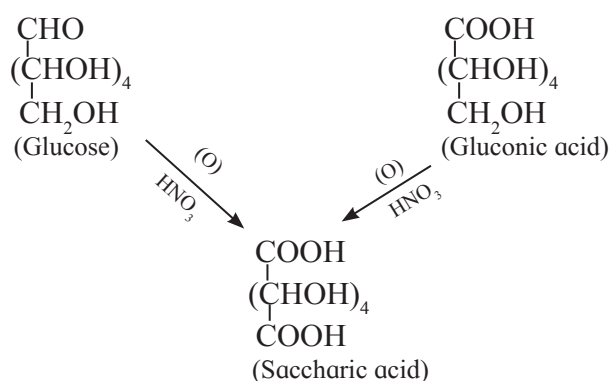


### 5. Glucose contains five hydroxyl groups :

This was inferred from the observation that Glucose reacts with five moles of acetic anhydride to form glucose pentaacetate. As glucose is a stable compound, it was further inferred that the five hydroxyl groups are bonded to five different carbon atoms in glucose molecule.



**6. Glucose contains one primary alcoholic (-CH<sub>2</sub>OH) group :** This was inferred from the following observation : Glucose and gluconic acid both on oxidation with dilute nitric acid give the same dicarboxylic acid called saccharic acid.



#### Use your brain power

- Write structural formula of glucose showing all the bonds in the molecule.
- Number all the carbons in the molecules giving number 1 to the (-CHO) carbon.
- Mark the chiral carbons in the molecule with asterisk (\*).
- How many chiral carbons are present in glucose?



#### 14.2.5 Optical isomerism in glucose :

Structural formula of glucose shows that it contains four chiral carbon atoms. You have learnt that every chiral carbon can have two distinct spatial arrangements of groups around it (section 10.5.1). In other words, two distinct configurations are possible for each of the four chiral carbons of glucose. Stereostructure of glucose is therefore one out of several possible stereostructures of an aldohexose.

#### Do you know ?

A structural formula containing 'n' number of chiral carbon can have maximum '2<sup>n</sup>' numbers of stereostructures or optical isomers. An aldohexose therefore, can exist as sixteen (2<sup>4</sup> = 16) optical isomers, and glucose is one of them.

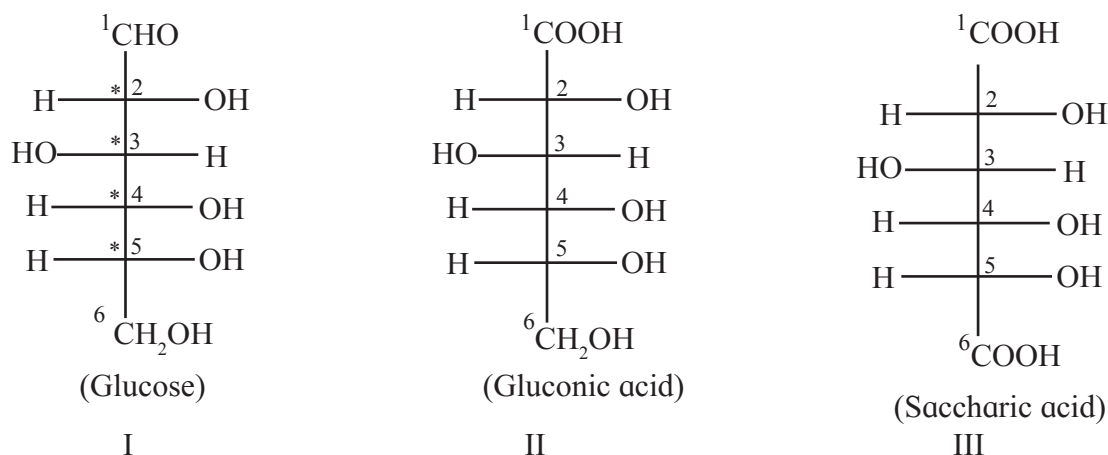


#### Can you recall ?

- What are the ways to represent three dimensional structure of an organic molecule?
- How is a Fischer projection formula drawn?



On the basis of very elaborate chemical evidence and measurement of optical activity of various chemicals involved, Emil Fischer, a German Nobel laureate (1902), determined the configuration of the four chiral carbons (C-2, C-3, C-4, C-5) in glucose.

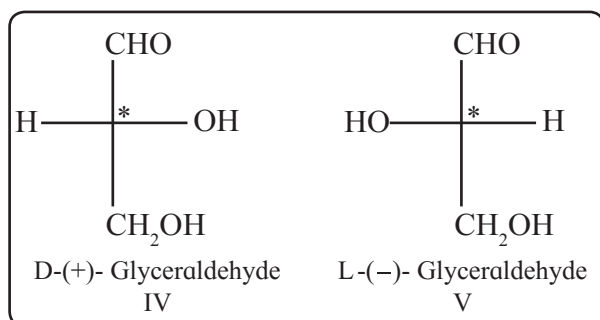


**Fig 14.2 : Fischer projection formulae of glucose, gluconic acid and saccharic acid**

Figure 14.2 shows the Fisher projection formulae of glucose (I), gluconic acid (II) and saccharic acid (III).

Glucose is an optically active compound and has its specific rotation,  $[\alpha]_D^{20}$ , equal to  $+52.7^\circ$ . Due to its dextrorotation glucose is also called **dextrose**. The designations **(+)-glucose** or **d-glucose** imply the dextrorotatory nature of glucose. **D-glucose** is another designation of glucose, which is more common. This designation indicates the configuration of glucose rather than the sign of its optical rotation.

**D/L configuration system :** The prefix D- or L- in the name of a compound indicates relative configuration of a stereoisomer. It refers to a particular enantiomer of glyceraldehyde. **Glyceraldehyde** has one chiral carbon (C-2) and exists as two enantiomers. These are represented by two Fischer projection formulae (see Fig. 14.3).



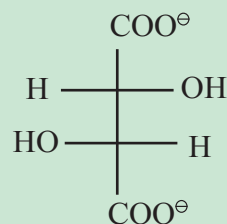
**Fig. 14.3 : Enantiomers of glyceraldehyde**

Conventionally (+)-glyceraldehyde is represented by the Fischer projection formula having OH group attached to C-2 on right side (IV) and this configuration is denoted by symbol 'D'. Similarly, configuration of (-) glyceraldehyde (V) is denoted by symbol 'L'. All the compounds which can be correlated by a series of chemical reactions to (+) - glyceraldehyde are said to have D-configuration. The compounds which are chemically correlated to (-) - glyceraldehyde are said to have L- configuration. This is the system of **relative configuration** of chiral compounds.

### Do you know ?

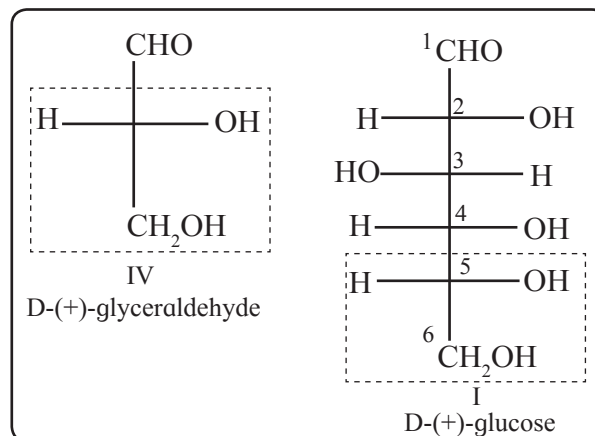


Optical rotation is an experimentally measurable property of a compound. Configuration of chiral carbon, on the other hand, is difficult to observe by simple experiment. In 1951 X-ray crystallographic studies of (+) - sodium rubidium tartarate established its configuration as :



This was the first instance of determining absolute configuration.

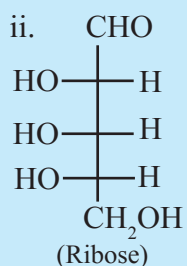
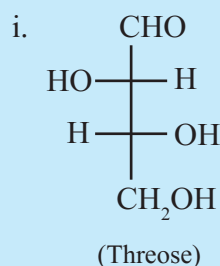
A monosaccharide is assigned D/L configuration on the basis of the configuration of the lowest chiral carbon in its Fischer projection formula. Figure 14.4 illustrates the D-configuration of (+) - glucose.



**Fig. 14.4 : Relative configuration of (+) - glucose**

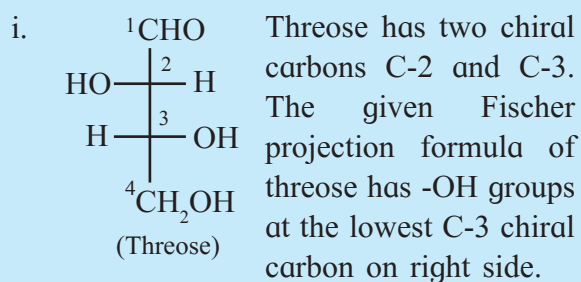
**14.2.6 Ring structure of glucose :** On the basis of chemical evidence stereostructure of D-glucose was represented by the Fischer projection formula **I** (Fig. 14.2 and Fig. 14.4). Glucose, however, was found to exhibit some more chemical properties which could not be explained on the basis of the structure **I**. It was necessary to write another structure for glucose which will explain all the properties. Ring structure of glucose fulfils this requirement.

**Problem 14.2 :** Assign D/L configuration to the following monosaccharides.



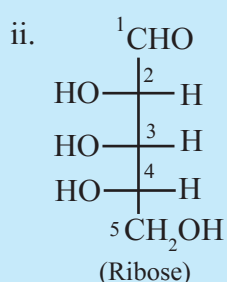
**Solution :**

D/L configuration is assigned to Fischer projection formula of monosaccharide on the basis of the lowest chiral carbon.



Threose has two chiral carbons C-2 and C-3. The given Fischer projection formula of threose has -OH groups at the lowest C-3 chiral carbon on right side.

∴ It is D-threose.

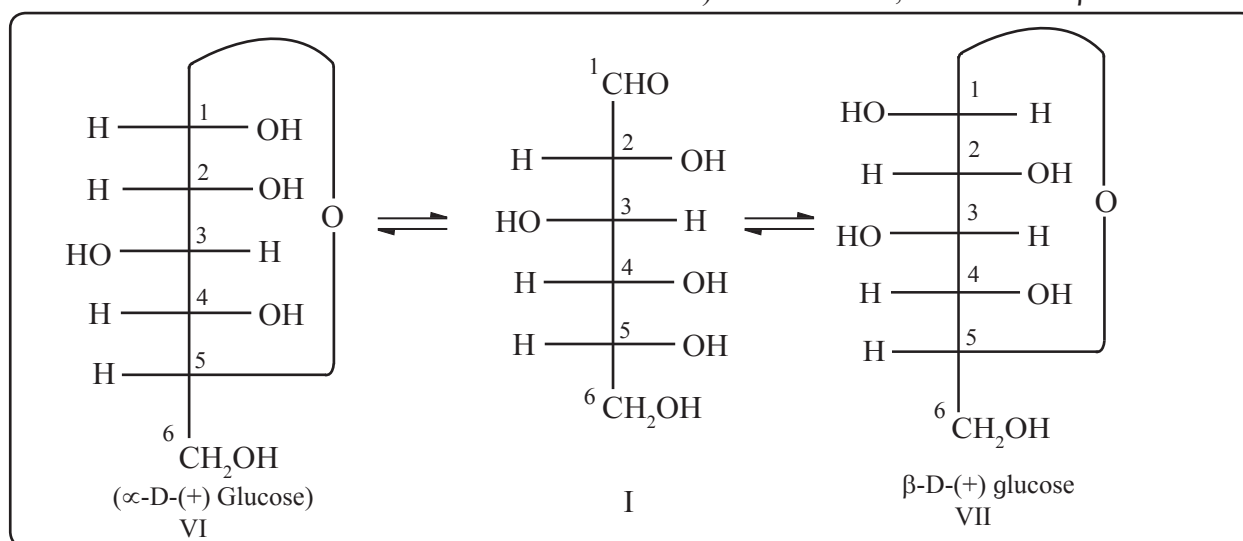


Ribose has three chiral carbons C-2, C-3 and C-4. The given Fischer projection formula of ribose has -OH group at the lowest C-4 chiral carbon on left side.

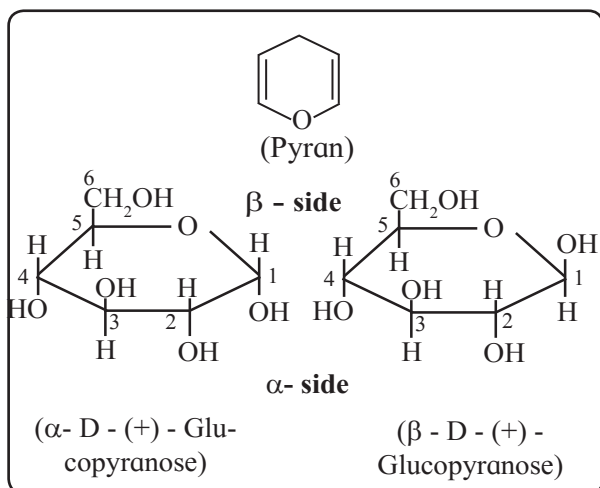
∴ It is L-ribose

Glucose is found to have two cyclic structures (VI and VII) which are in equilibrium with each other through the open chain structure (I) in aqueous solution (Fig.14.5).

The ring structure of glucose is formed by reaction between the formyl (-CHO) group and the alcoholic (-OH) group at C-5. Thus, the ring structure is a **hemiacetal** structure (section 12.8.2 c). The two hemiacetal structures (VI and VII) differ only in the configuration of C-1 (Fig. 14.5), the additional chiral centre resulting from ring closure. The two ring structures are called **α-** and **β- anomers** of glucose and C-1 is called the anomeric carbon. The ring of the cyclic structure of glucose contains five carbons and one oxygen. Thus, it is a six membered ring. It is called pyranose structure, in analogy with the six membered heterocyclic compound pyran (Fig. 14.6). Hence glucose is also called **glucopyranose**. Haworth formula is a better way than Fischer projection formula to represent structure of glucopyranose (Fig. 14.6). In the **Haworth formula** the pyranose ring is considered to be in a perpendicular plane with respect to the plane of paper. The carbons and oxygen in the ring are in the places as they appear in Fig. 14.6. The lower side of the ring is called **α-side** and the upper side is the **β-side**. The α-anomer has its anomeric hydroxyl (-OH) group (at C-1) on the α-side, whereas the β-anomer has



**Fig. 14.5 Ring structures of glucose: Fischer projection formulae**



**Fig. 14.6 : Haworth formula of anomers of glucopyranose**

its anomeric hydroxyl (-OH) group (at C-1) on the β-side. The groups which appear on right side in the Fischer projection formula appear on α-side in the Haworth formula, and viceversa.

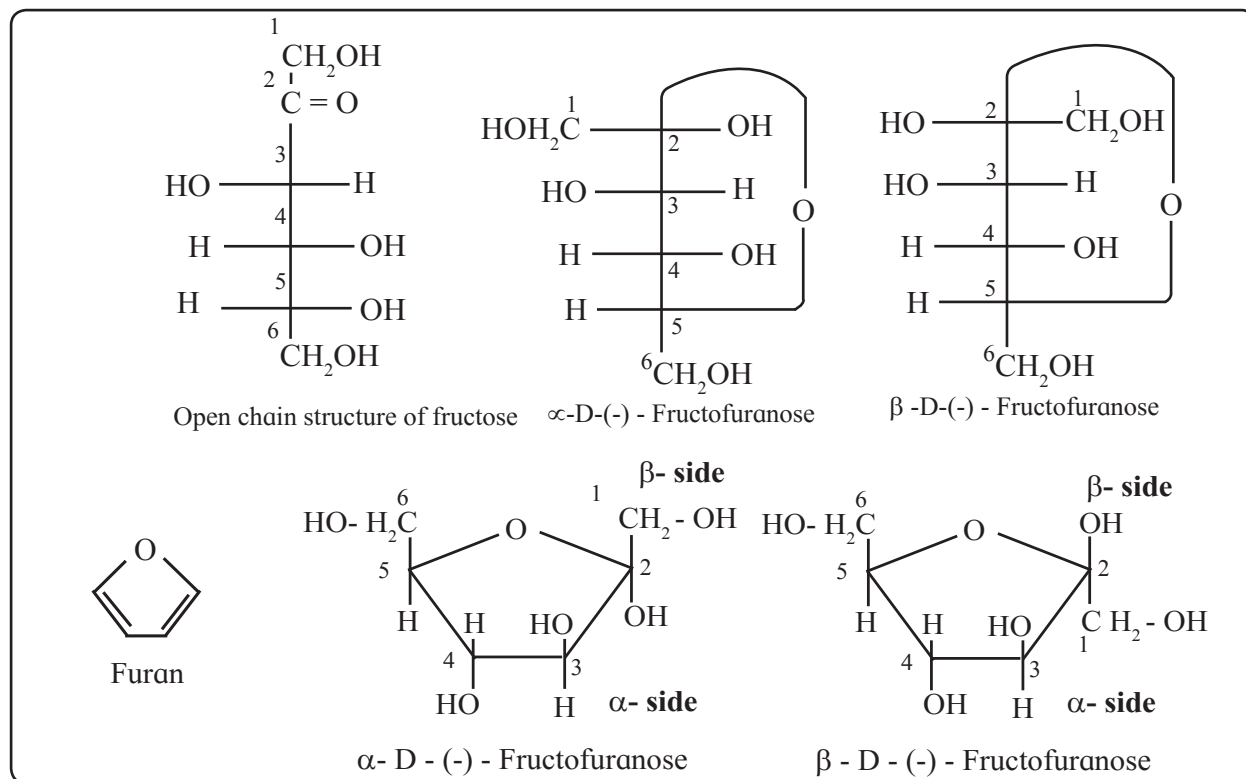
#### 14.2.7 Reducing nature of glucose :

Hemiacetal group of glucopyranose structure is a **potential aldehyde group**. It imparts reducing properties to glucose. Thus, glucose gives **positive Tollens test and positive Fehling test** (Section 12.8.1 a).

#### 14.2.8 Representation of Fructose structure

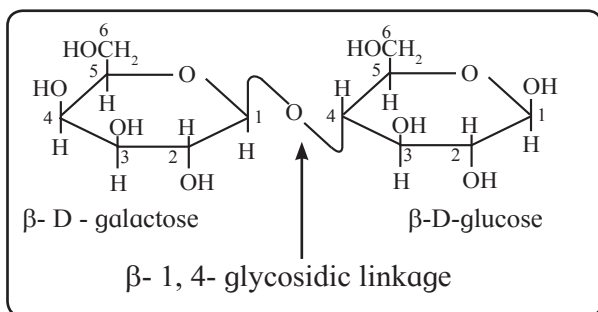
Fructose ( $C_6H_{12}O_6$ ) is a laevorotatory ketohexose. Fructose is also called laevulose due to its laevorotation  $[\alpha]_D^{20} = -92.4^\circ$ . Being an α-hydroxy keto compound fructose is a **reducing sugar**. In free state it exists as mixture of fructopyranose (major) and fructofuranose. In combined state fructose is found in the form of **fructofuranose** ring structure (as in sucrose, see section 14.2.9). The name furanose is given by analogy with furan, a five membered heterocyclic compound. Figure 14.7 shows representations of open chain structure of fructose and ring structures of α- and β- anomers of fructofuranose. Ring structure of fructose is a hemiketal (section 12.8.2 c).

**14.2.9 Disaccharides :** Disaccharides give rise to two units of same or different monosaccharides on hydrolysis with dilute acids or specific enzymes. The two monosaccharide units are linked together by an ether oxide linkage (-O-), which is termed as **glycosidic linkage** in carbohydrate chemistry. Glycosidic linkage is formed by removal of a water molecule by reaction of two hydroxyl



**Fig. 14.7 : Representations of fructose structure**





**Fig. 14.10 : Haworth formula of lactose**

**c. Lactose** : Lactose ( $C_{12}H_{22}O_{11}$ ) is a disaccharide present in milk. It is formed from two monosaccharide units, namely D-galactose and D-glucose. The glycosidic linkage is formed between C-1 of  $\beta$ -D-galactose and C-4 of glucose. Therefore the linkage in lactose is called  $\beta$ -1,4-glycosidic linkage. The hemiacetal group at C-1 of the glucose unit is not involved in glycosidic linkage but is free. Hence lactose is a **reducing sugar**. Figure 14.10 shows Haworth formula of lactose.

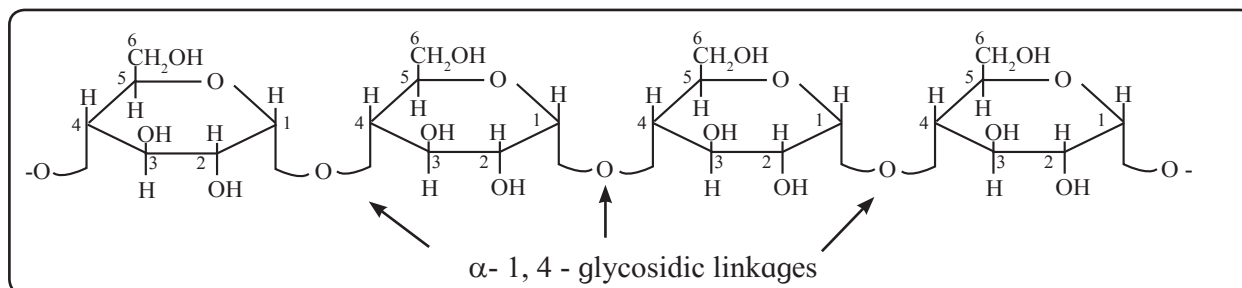
**14.2.10 Polysaccharides** : Polysaccharides are formed by linking large number of monosaccharide units by glycosidic linkages.

### Use your brain power

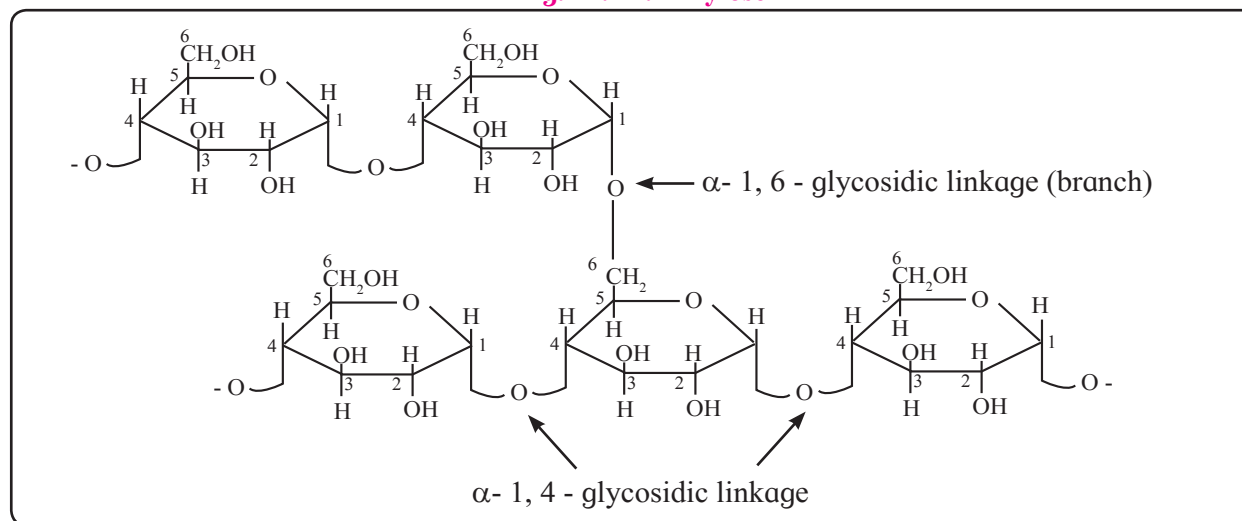


- Is galactose an aldohexose or a ketohexose?
- Which carbon in galactose has different configuration compared to glucose?
- Draw Haworth formulae of  $\alpha$ -D-galactose and  $\beta$ -D-galactose.
- Which disaccharides among sucrose, maltose and lactose is/are expected to give positive Fehling test?
- What are the expected products of hydrolysis of lactose?

Starch, cellulose and glycogen are the most common natural polysaccharides. Starch is storage carbohydrate of plants and important nutrient for humans and other animals. Cellulose is the main constituent of cell wall of plant and bacterial cells. It is also main constituent of wood and cotton. Glycogen constitutes storage carbohydrate of animals and is present in liver, muscles and brain. It is also found in yeast and fungi.



**Fig. 14.11 : Amylose**



**Fig. 14.12 : Amylopectin**

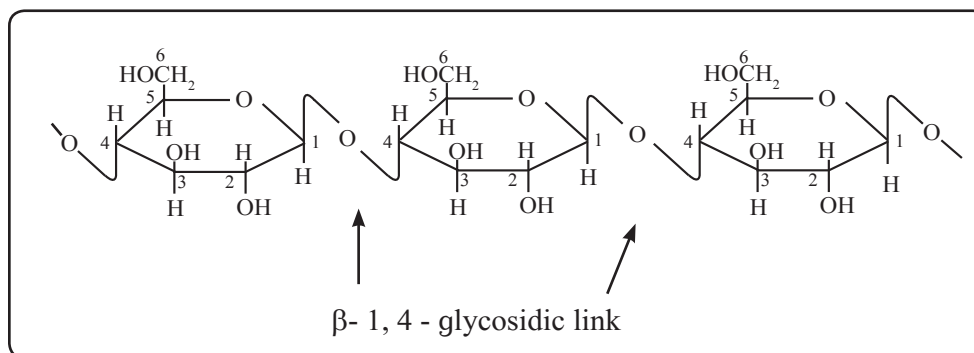
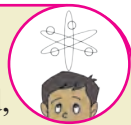


Fig. 14.13 : Cellulose

### Can you think ?

When you chew plain bread, *chapati* or *bhaakari* for long time, it tastes sweet. What could be the reason ?



**a. Starch :** Starch is a polymer of  $\alpha$ -D-glucose. Starch has two components, namely, **amylose** (15-20%) and **amylopectin** (80-85%). Amylose is soluble in water and forms blue coloured complex with iodine. It contains 200-1000  $\alpha$ -glucose units linked by  $\alpha$ -1,4- glycosidic linkages giving rise to unbranched chain of variable length (Fig. 14.11). Amylopectin is water insoluble component of starch which forms blue-violet coloured complex with iodine. It is a branched chain polysaccharide. In amylopectin, chains are formed by  $\alpha$ -1,4- glycosidic linkages between  $\alpha$ -glucose units, whereas branches are formed by  $\alpha$ -1,6- glycosidic linkages (Fig. 14.12).

**b. Cellulose :** Cellulose is a straight chain polysaccharide of  $\beta$ -glucose units linked by  $\beta$ -1,4- glycosidic bonds. Chemical hydrolysis of cellulose requires use of concentrated strong acids at high temperature and pressure. This implies that the  $\beta$ -1,4- glycosidic bond is very strong and difficult to hydrolyse. Humans do not have enzymes which can hydrolyse this linkage. Hence cellulose cannot be digested by human beings; it serves as the fibrous content of food useful for bowel movement. Figure 14.13 shows the Haworth formula of cellulose.

**c. Glycogen :** Glycogen has its structure similar to that of amylopectin, but it is more highly branched.

### Do you know ?

The symbiotic bacteria in guts of insects called termites have enzymes that can hydrolyse  $\beta$ -1,4- glycosidic linkage in cellulose.



## 14.3 Proteins

### Can you recall ?

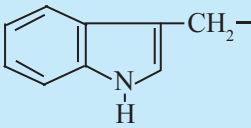

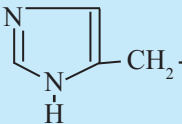
- What is the product of reaction of acetic acid with ammonia ?
- Write the structural formula of N-methyl acetamide. What is the name of the functional group in this compound?
- What are the nitrogenous nutrients in human diet?

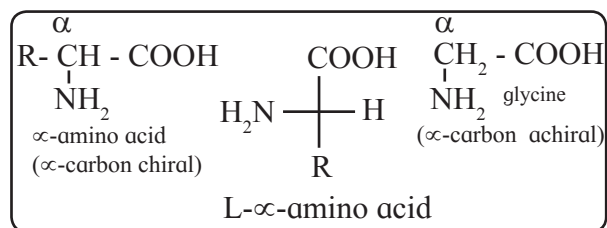


Proteins are the fundamental structural materials of animal bodies. Proteins in the form of enzymes play prime role in all the physiological reactions. The name protein is derived from the Greek word, 'proteios' which means 'primary' or 'of prime importance'. Nutritional sources of proteins are milk, pulses, nuts, fish, meat, etc. Chemically proteins are polyamides which are high molecular weight polymers of the monomer units called  $\alpha$ -amino acids.

**14.3.1  $\alpha$ -Amino acids :** Proteins on complete hydrolysis give rise to a mixture of  $\alpha$ -amino acids.  $\alpha$ -Amino acids are carboxylic acids having an amino ( $-\text{NH}_2$ ) group bonded to

**Table 14.1 Natural  $\alpha$  - amino acids : L - RCH (NH<sub>2</sub>) COOH (\* Essential  $\alpha$  - amino acids)**

Type	Name	R	Three letter symbol	One letter symbol
Neutral $\alpha$ -amino acids	1. Glycine	H-	Gly	G
	2. Alanine	CH <sub>3</sub> -	Ala	A
	3. Valine*	Me <sub>2</sub> CH-	Val	V
	4. Leucine*	Me <sub>2</sub> CH-CH <sub>2</sub> -	Leu	L
	5. Isoleucine*	CH <sub>3</sub> -CH <sub>2</sub> -CH(Me)-	Ile	I
	6. Asparagine	H <sub>2</sub> N-CO-CH <sub>2</sub> -	Asn	N
	7. Glutamine	H <sub>2</sub> N-CO-CH <sub>2</sub> -CH <sub>2</sub> -	Gln	Q
	8. Serine	HO-CH <sub>2</sub> -	Ser	S
	9. Threonine*	CH <sub>3</sub> -CHOH-	Thr	T
	10. Cysteine	HS-CH <sub>2</sub> -	Cys	C
	11. Methionine*	Me-S-CH <sub>2</sub> -CH <sub>2</sub> -	Met	M
	12. Phenylalanine*	Ph-CH <sub>2</sub> -	Phe	F
	13. Tyrosine	p-HO-C <sub>6</sub> H <sub>4</sub> -CH <sub>2</sub> -	Tyr	Y
	14. Tryptophan*		Trp	W
	15. Proline	 (entire structure)	Pro	P
Acidic	16. Aspartic acid	HOOC-CH <sub>2</sub> -	Asp	D
	17. Glutamic acid	HOOC-CH <sub>2</sub> -CH <sub>2</sub> -	Glu	G
Basic	18. Lysine*	H <sub>2</sub> N-(CH <sub>2</sub> ) <sub>4</sub> -	Lys	K
	19. Arginine*	HN = C(NH <sub>2</sub> ) - NH - (CH <sub>2</sub> ) <sub>3</sub> -	Arg	R
	20. Histidine*		His	H



**Fig. 14.14 : Natural  $\alpha$ -amino acids**

the  $\alpha$ -Carbon, that is, the carbon next to the carboxyl (-COOH) group (Fig. 14.14).  $\alpha$ -carbon in all the  $\alpha$ -amino acids (except glycine) is chiral. It is found that the  $\alpha$ -carbon in  $\alpha$ -amino acids obtained by hydrolysis of proteins has 'L' configuration. The L- $\alpha$ -amino acids are represented by the Fischer projection formula as shown in Fig. 14.14.

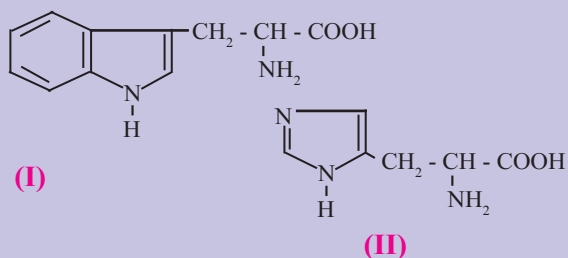
The symbol 'R' in the structure of  $\alpha$ -amino acids represents side chain and may contain additional functional groups. If 'R' contains a carboxyl (-COOH) group the amino acid is **acidic amino acid**. If 'R' contains an amino (1<sup>o</sup>, 2<sup>o</sup> or 3<sup>o</sup>) group, it is called **basic amino acid**. The other amino acids having neutral or no functional group in 'R' are called **neutral amino acids**.

$\alpha$ -Amino acids have trivial names and are generally represented by three letter symbols or sometimes by one letter symbol. Table 14.1 lists the twenty  $\alpha$ -amino acids, often referred to as simply amino acids, commonly found in proteins with their symbols and also their

types as neutral, acidic or basic. Ten  $\alpha$ -amino acids from this list cannot be synthesised in human body and have to be obtained through diet. These are called **essential amino acids** and are marked with asterisk (\*) in Table 14.1.

### Use your brain power

Tryptophan and histidine have the structures (I) and (II) respectively. Classify them into neutral/acidic/basic  $\alpha$ -amino acids and justify your answer. (Hint : Consider involvement of lone pair in resonance).



### Can you think ?

Compare the molecular masses of the following compounds and explain the observed melting points.

Formula	Molecular mass	Melting point
$\text{CH}_3 - \underset{\text{NH}_2}{\text{CH}} - \text{COOH}$	89	293.5°C
$\text{C}_5\text{H}_{11} - \text{NH}_2$	87	-55°C
$\text{C}_3\text{H}_7 - \text{COOH}$	88	-7.9°C

$\alpha$ -Amino acids are high melting, water soluble crystalline solids, unlike simple amines or carboxylic acids. These properties are due to a peculiar structure called **zwitter ion structure** of  $\alpha$ -amino acids. An  $\alpha$ -amino acid molecule contains both acidic carboxyl (-COOH) group as well as basic amino (-NH<sub>2</sub>) group. Proton transfer from acidic group to basic group of amino acid forms a salt, which is a dipolar ion called zwitter ion (Fig. 14.15).

Amino acid can exist in different forms depending upon the pH of the aqueous solution in which it is dissolved. Consider,

for example, zwitter ion and the other forms of alanine (Fig. 14.16).

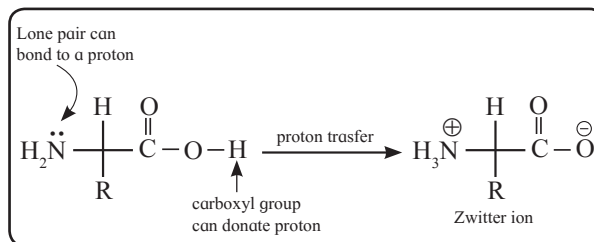


Fig. 14.15 : Zwitter ion

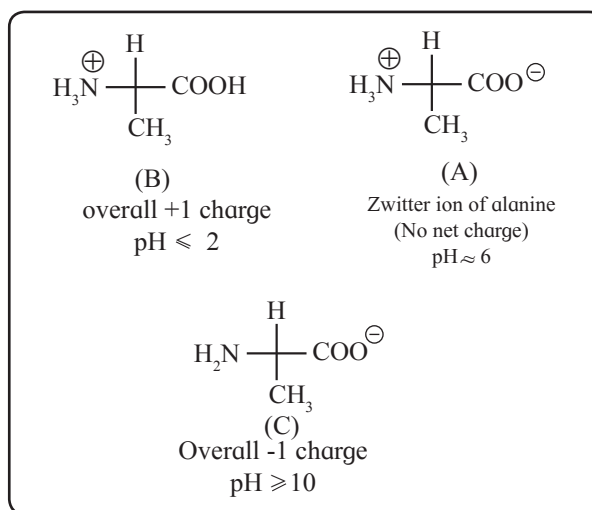


Fig. 14.16 : Three forms of alanine

### Do you know ?

At the physiological pH of 7.4, neutral  $\alpha$ -amino acids are primarily in their zwitterionic forms. On the other hand, at this pH acidic  $\alpha$ -amino acids exist as anion (due to deprotonation of the carboxyl group), while basic  $\alpha$ -amino acids exist as cation (due to protonation of the amino groups). Ionic structures of constituent  $\alpha$ -amino acids result in ionic nature of proteins.

### Can you recall ?

- What does the enzyme pepsin do?
- What are the initial and final products of digestion of proteins?

### 14.3.2 Peptide bond and protein :

Proteins are known to break down into peptides in stomach and duodenum under the influence of enzymes, pepsin being one of them which is secreted by stomach. Polypeptides are further broken down to  $\alpha$ -amino acids. This implies that proteins are formed by connecting  $\alpha$ -amino acids to each other. The bond that connects  $\alpha$ -amino acids to each other is called **peptide bond**. Consider, for example, linking of a molecule of glycine with that of alanine. One way of doing this is to combine carboxyl group of glycine with  $\alpha$ -amino group of alanine. This results in elimination of a water molecule and formation of a **dipeptide** called **glycylalanine** in which the two amino acid units are linked

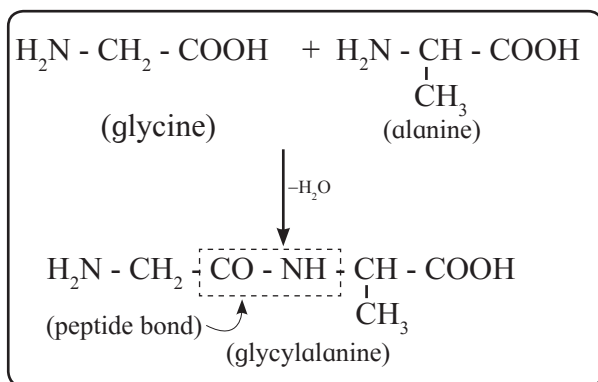


Fig. 14.17 : Peptide bond

by a peptide bond (Fig. 14.17). It can be seen that a peptide bond or peptide linkage is same as what is described as **secondary amide** in organic chemistry. Combination of a third molecule of an  $\alpha$ -amino acid with a dipeptide would result in formation of a **tripeptide**. Similarly linking of four, five or six  $\alpha$ -amino acids results in formation of tetrapeptide, pentapeptide or hexapeptide

#### Use your brain power

- Write the structural formula of dipeptide formed by combination of carboxyl group of alanine and amino group of glycine.
- Name the resulting dipeptide.
- Is this dipeptide same as glycylalanine or its structural isomer?



respectively. When the number of  $\alpha$ -amino acids linked by peptide bonds is more than ten, the products are called **polypeptides**. The -CHR- units linked by peptide bonds are referred to as '**amino acid residues**'. Proteins are polypeptides having more than hundred amino acid residues linked by peptide bonds. It may be, however, noted that distinction between proteins and polypeptides is not sharp. The two ends of a polypeptide chain of protein are not identical. The end having free carboxyl group is called **C-terminal** while the other end having free amino group is called **N-terminal**. In the dipeptide glycylalanine glycine residue is N-terminal and alanine residue is C-terminal.

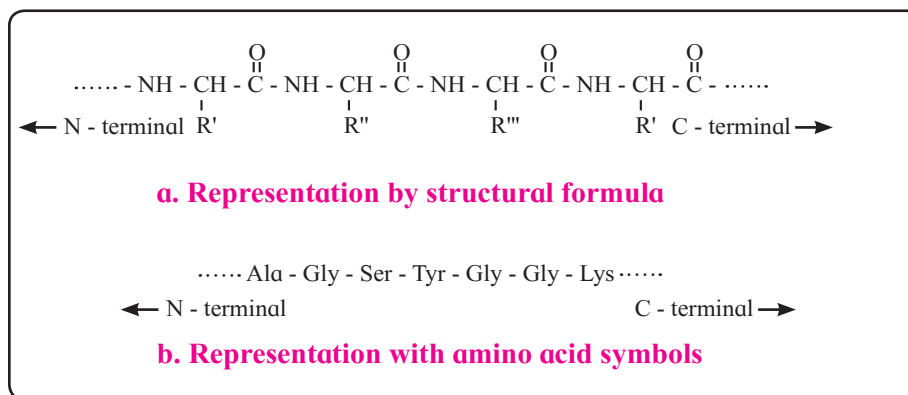
**14.3.3 Types of proteins :** Depending upon the molecular shape proteins are classified into two types.

**a. Globular proteins :** Molecules of globular proteins have spherical shape. This shape results from coiling around of the polypeptide chain of protein. Globular proteins are usually soluble in water. For example : insulin, egg albumin, serum albumin, legumelin (protein in pulses)

**b. Fibrous proteins :** Molecules of fibrous proteins have elongated, rod like shape. This shape is the result of holding the polypeptide chains of protein parallel to each other. Hydrogen bonds and disulfide bonds are responsible for this shape. Fibrous proteins are insoluble in water. For example : keratin (present in hair, nail, wool), myosin (protein of muscles).

The shapes of protein molecules are the result of **four level structure of proteins**.

**14.3.4 Structure of proteins :** Proteins are responsible for a variety of functions in organisms. Proteins of hair, muscles, skin give shape to the structure, while enzymes are proteins which catalyze physiological reactions. These diverse functions of proteins can be understood by studying the **four**



**Fig. 14.18 : Representation of primary structure of protein**

**level structure of proteins**, namely primary, secondary, tertiary and quaternary structure of proteins.

**a. Primary structure of proteins :** Primary structure of proteins is **the sequence of constituent  $\alpha$ -amino acid residues linked by peptide bonds**. Any change in the sequence of amino acid residue results in a different protein. Primary structure of proteins is represented by writing the three letter symbols of amino acid residues as per their sequence in the concerned protein. The

symbols are separated by dashes. According to the convention, the N-terminal amino acid residue as written at the left end and the C-terminal amino acid residue at the right end (Fig. 14.18).

**Problem 14.3**

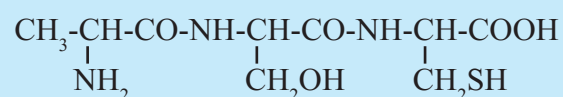
Chymotrypsin is a digestive enzyme that hydrolyzes those amide bonds for which the carbonyl group comes from phenylalanine, tyrosine or tryptophan. Write the symbols of the amino acids and peptides smaller than pentapeptide formed by hydrolysis of the following hexapeptide with chymotrypsin.

Gly-Tyr-Gly-Ala-Phe-Val

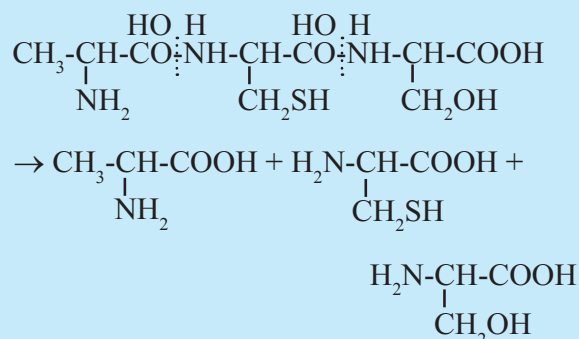
**Solution :** In the given hexapeptide hydrolysis by chymotrypsin can take place at two points, namely, Phe and Tyr. The carbonyl group of these residues is towards right side, that is, toward the C-terminal. Therefore the hydrolysis products in required range will be :

Gly-Tyr, Gly-Ala-Phe and Val  
(a dipeptide) (a tripeptide) ( $\alpha$ -amino acid)

**Problem 14.4 :** Write down the structures of amino acids constituting the following peptide.

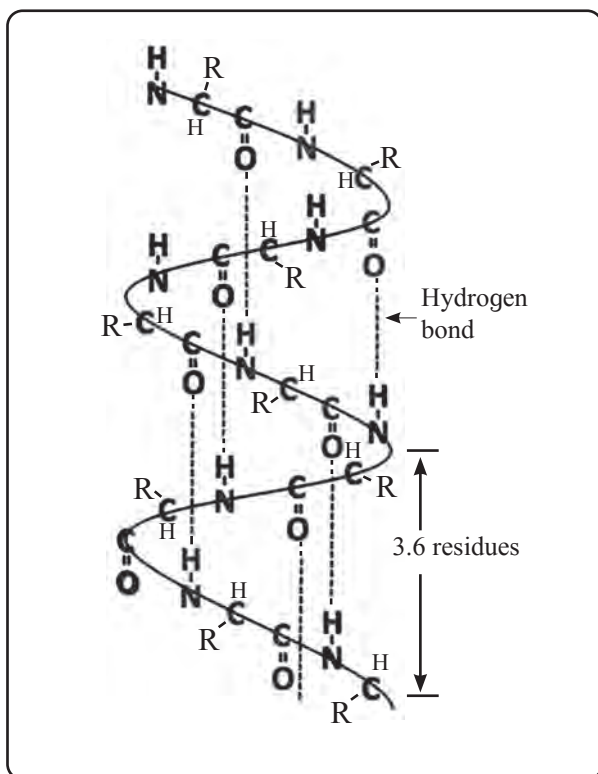


**Solution :** The given peptide has two amide bonds linking three amino acids. The structures of these amino acids are obtained by adding one  $\text{H}_2\text{O}$  molecule across the amide bond as follows :



**b. Secondary structure of proteins :**

**The three-dimensional arrangement of localized regions of a protein chain** is called the secondary structure of protein. Hydrogen bonding between N-H proton of one amide linkage and C=O oxygen of another gives rise to the secondary structure. Two types of secondary structures commonly found in proteins are  $\alpha$ -helix and  $\beta$ -pleated sheet.



**Fig. 14.19 : Backbone of  $\alpha$  - Helix**

**$\alpha$ -Helix** : The  $\alpha$ -helix forms when a polypeptide chain twists into a **right handed or clockwise spiral** (Fig. 14.19). Some characteristic features of  $\alpha$ -helical structure of protein are:

- Each turn of the helix has 3.6 amino acids.
- A C=O group of one amino acid is hydrogen bonded to N-H group of the fourth amino acid along the chain.
- Hydrogen bonds are parallel to the axis of helix while R groups extend outward from the helix core.

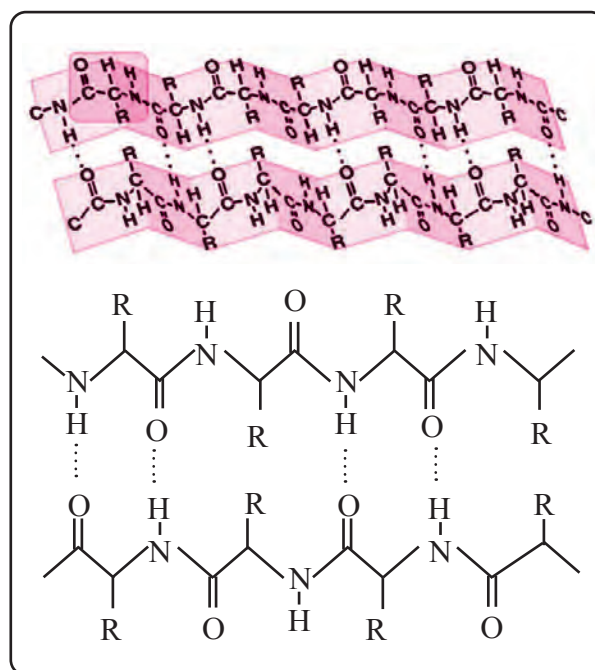
Myosin in muscle and  $\alpha$ -keratin in hair are proteins with almost entire  $\alpha$ -helical secondary structure.

#### Do you know ?

In collagen, the protein of connective tissue, the polypeptide chains have **unusual left-handed helix structure**. Three strands of these chains wind around each other in a right-handed triple helix.



**$\beta$ -Pleated sheet** : The secondary structure is called  $\beta$ -pleated sheet when **two or more polypeptide chains, called strands, line up side-by-side** (Fig. 14.20). The  $\beta$ -pleated sheet structure of protein consists of extended strands of polypeptide chains held together by hydrogen bonding. The characteristics of  $\beta$ -pleated sheet structure are :



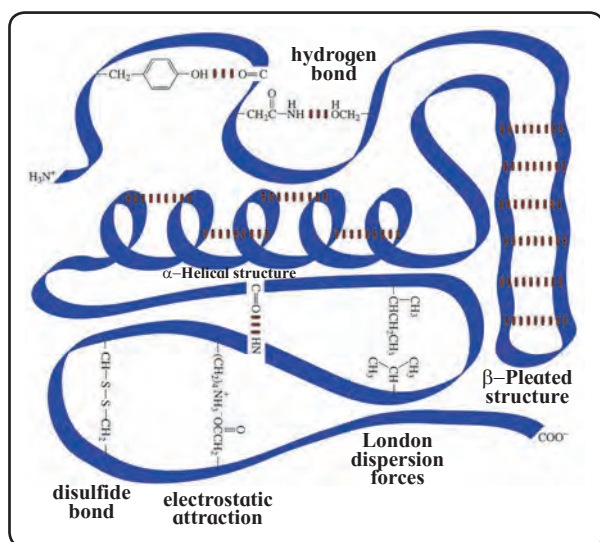
**Fig. 14.20 :  $\beta$  - pleated sheet**

- The C=O and N-H bonds lie in the plane of the sheet.
- Hydrogen bonding occurs between the N-H and C=O groups of nearby amino acid residues in the neighbouring chains.
- The R groups are oriented above and below the plane of the sheet.

The  $\beta$ -pleated sheet arrangement is favoured by amino acids with small R groups.

Most proteins have regions of  $\alpha$ -helix and  $\beta$ -pleated sheet, in addition to other random regions that cannot be characterised by either of these secondary structures. For example: Spider dragline silk protein is strong due to  $\beta$ -pleated sheet region, yet elastic due to  $\alpha$ -helical regions in it.

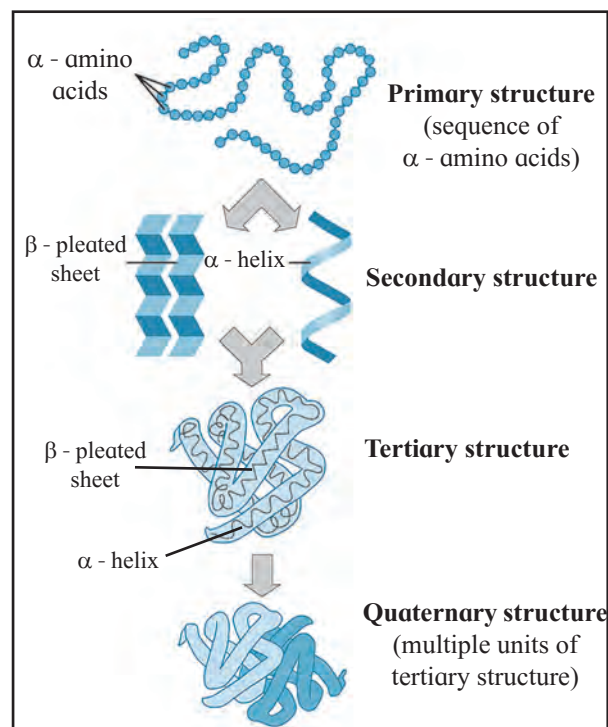
**c. Tertiary structure of proteins :** The three-dimensional shape adopted by the entire polypeptide chain of a protein is called its tertiary structure. It is the result of folding of the chain in a particular manner that the structure is itself stabilized and also has attractive interaction with the aqueous environment of the cell. The **globular** and **fibrous** proteins represent two major molecular shapes resulting from the tertiary structure. The forces that stabilize a particular tertiary structure include **hydrogen bonding**, **dipole-dipole attraction** (due to polar bonds in the side chains), **electrostatic attraction** (due to the ionic groups like  $-\text{COO}^-$ ,  $\text{NH}_3^+$  in the side chain) and also **London dispersion forces**. Finally, **disulfide bonds** formed by oxidation of nearby  $-\text{SH}$  groups (in cysteine residues) are the **covalent bonds** which stabilize the tertiary structure (Fig. 14.21).



**Fig. 14.21 : Tertiary structure of protein**

**d. Quaternary structure of proteins :** When two or more polypeptide chains with folded tertiary structures come together into one protein complex, the resulting shape is called quaternary structure of the protein. Each individual polypeptide chain is called a subunit of the overall protein. For example: **Haemoglobin** consists of **four subunits** called haeme held together by intermolecular forces in a compact three dimensional shape. Haemoglobin can do its function of oxygen

transport only when all the four subunits are together. Figure 14.22 summarizes schematically the four levels of protein structure.

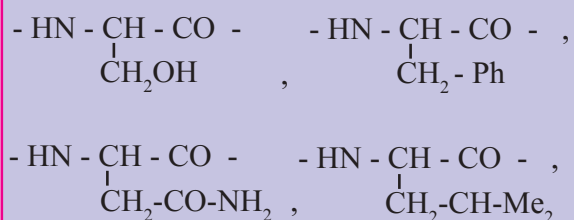


**Fig. 14.22 : Four levels of protein structure**

### Use your brain power



A protein chain has the following amino acids residues. Show and label the interactions that can be present in various pairs from these giving rise to tertiary level structure of protein.



### Can you tell ?



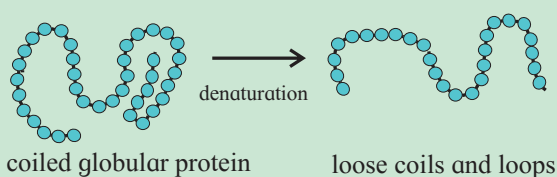
What is the physical change observed when (a) egg is boiled, (b) milk gets curdled on adding lemon juice ?

### 14.3.5 Denaturation of proteins

High temperature, acid, base and even agitation can disrupt the noncovalent interactions responsible for a specific shape of protein. This is denaturation of protein. **Denaturation is the process by which the molecular shape of protein changes without breaking the amide/peptide bonds that form the primary structure.**

#### Do you know ?

Globular proteins are typically folded with hydrophobic side chains in the interior and polar residues on the outside, and thereby are water soluble. Denaturation exposes the hydrophobic region of globular proteins and makes them water insoluble.



Denaturation results in disturbing the secondary, tertiary or quaternary structure of protein. This causes change in properties of protein and the biological activity is often lost.

### 14.3.6 Enzymes :

#### Can you recall ?

- Which parameter, equilibrium constant or activation energy, decides the rate of a chemical reaction?
- What is the influence of a catalyst on activation energy?

A very large number of chemical reactions take place in our bodies. These are brought about at the physiological pH of 7.4 and the body temperature of 37° C with the help of biological catalysts called enzymes. For example : insulin, hormone secreted by pancreas, controls blood sugar level; amylase, an enzyme present in saliva, hydrolyzes starch.

Chemically enzymes are proteins. Every living cell contains at least 1000 different enzymes. Most enzymes catalyse only one reaction or one group of similar reactions. Thus, enzyme catalysis is **highly specific**. You have learnt that a mineral acid can catalyse hydrolysis of many types of compounds such as esters, acetals and amides. In contrast, an enzyme that catalyses hydrolysis of amide will not work on ester or acetal.

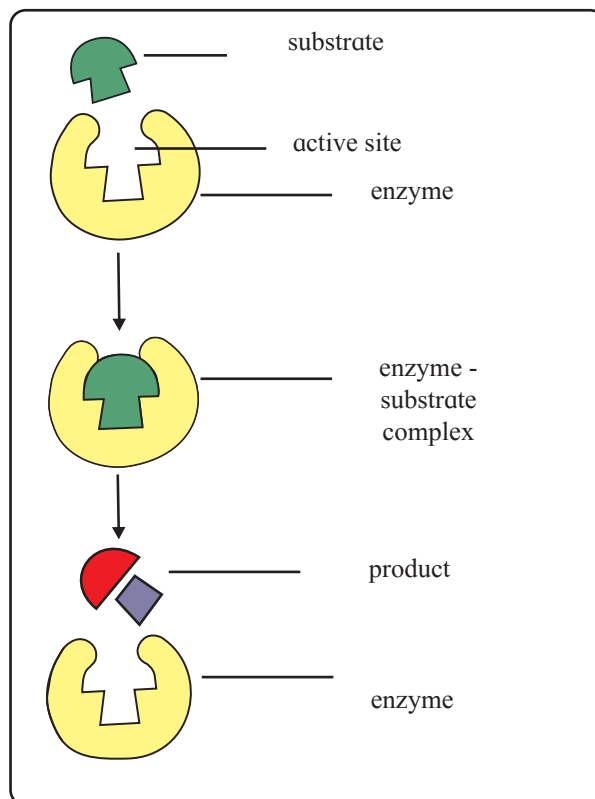


Fig. 14.23 : Enzyme catalysis

#### Mechanism of enzyme catalysis

Action of an enzyme on a substrate is described as **lock-and-key mechanism** (Fig.14.23). Accordingly, an enzyme has an active site on its surface. A substrate molecule can attach to this active site only if it has the right size and shape. Once in the active site, the substrate is held in the correct orientation to react and forms the products of reaction.

The products leave the active site and the enzyme is then ready to act as catalyst again. Formation of enzyme-substrate complex has very low activation energy. That is how the rate of the reaction is very high. Some enzymes are so **efficient** that one enzyme molecule can catalyse the reaction of 10000 substrate molecules in one second.

Several enzymes have been isolated from organisms (such as bacteria), purified and crystallised, and amino acid sequences of many of them have been determined. In many **industrial processes** specific reactions are carried out by use of enzymes extracted from organisms, and also by use of **new enzymes** made using genetic engineering.

Some examples of **industrial application** of enzyme catalysis are :

- Conversion of glucose to sweet-tasting fructose, using glucose isomerase.
- Manufacture of new antibiotics, using penicillin G acylase.
- Manufacture of laundry detergents, using proteases.
- Manufacture of esters used in cosmetics, using genetically engineered enzyme.

#### 14.4 Nucleic acids :

##### Can you tell ?

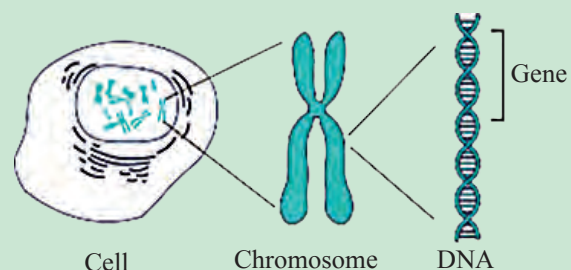
- What is the single term that answers all the following questions ?
- What decides whether you are blue eyed or brown eyed ?
- Why does wheat grain germinate to produce wheat plant and not rice plant ?
- Which acid molecules are present in nuclei of living cells ?

One of the most remarkable properties of living cells is their ability to produce their replicas through thousands of generations. This becomes possible because certain type

##### Do you know ?



Synthesis of protein, the fundamental structural material of body, is the process in which genetic information is transferred. DNA governs this process. DNA is present in the chromosomes of the cell nucleus. Each chromosome has a different type of DNA. An individual chromosome is composed of many **genes**. Gene is a portion of DNA molecule responsible for synthesis of a single protein. **DNA stores the genetic information**, while **RNA translates** this into synthesis of proteins needed by cells for proper function and development.

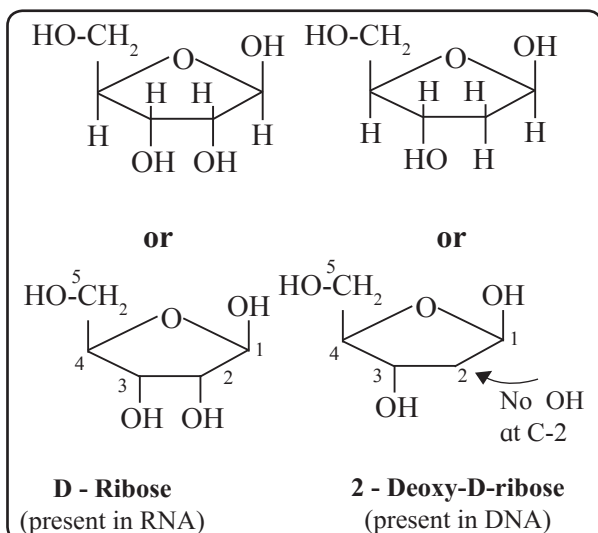


of information is passed unchanged from one generation to the next. Such information is called **genetic information** and its transfer to new cells is accomplished by **nucleic acids**.

There are two types of nucleic acids : **ribonucleic acids (RNA)** and **deoxyribonucleic acids (DNA)**. RNA are found mainly in the fluid of living cells (cytoplasm) while DNA are found primarily in the nuclei of living cells.

Knowledge of structure of nucleic acids is essential to understand their biological functions. In this chapter we are going to look at the **structural aspects of nucleic acids**.

**14.4.1 Nucleotides :** Nucleic acids are unbranched polymers of **repeating monomers** called nucleotides. In other words, nucleic acids have a polynucleotide structure. DNA molecules contain several million nucleotides while RNA molecules contain a few thousand



**Fig 14.24 : Sugar Components of nucleic acids**  
nucleotides.

The nucleotide monomers consist of three components : a **monosaccharide**, a **nitrogen-containing base** and a **phosphate group**.

Nucleotides of both RNA and DNA contain five-membered ring monosaccharide (furanose), often called simply sugar component. In RNA, the sugar component of nucleotide unit is D-ribose, while in DNA, it is 2-deoxy-D-ribose (Fig. 14.24).

Total five nitrogen - containing bases are present in nucleic acids. Three bases with one ring (**cytosine**, **uracil** and **thymine**) are derived from the parent compound **pyrimidine**. Two

### Remember...

RNA contains : D-ribose, A, G, C, U

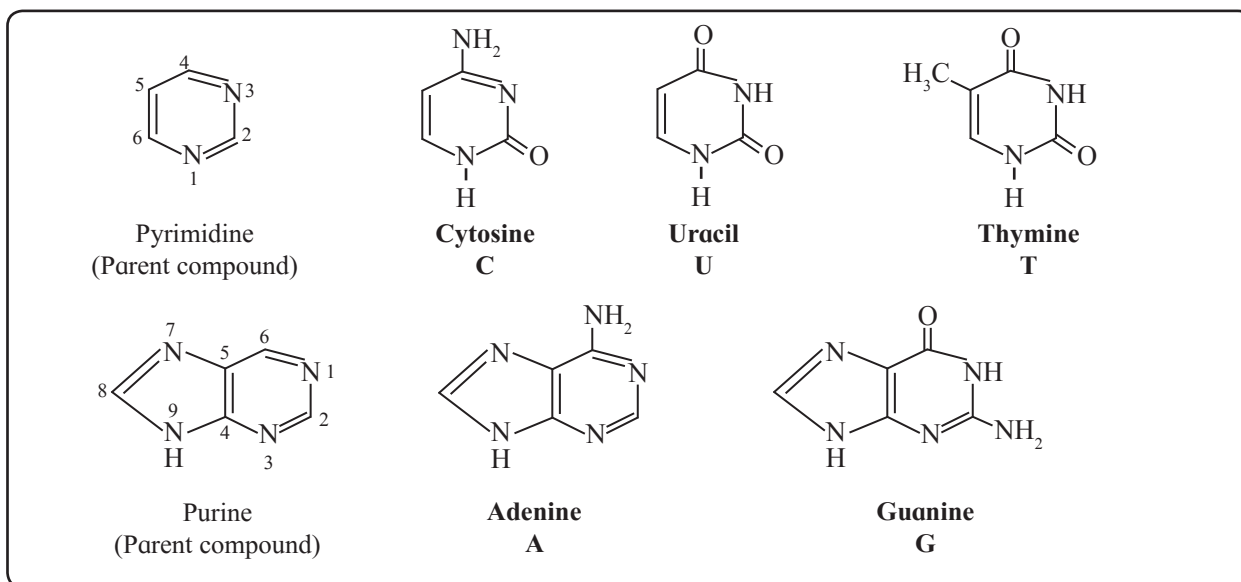
DNA contains : D-2-deoxyribose, A, G, C, T



bases with two rings (**adenine** and **guanine**) are derived from the parent compound **purine**. Each base is designated by a one-letter symbol (Fig. 14.25). Uracil (U) occurs only in RNA while thymine (T) occurs only in DNA.

A **nucleoside** is formed by joining the anomeric carbon of the furanose with nitrogen of a base. While numbering the atoms in a nucleoside, primes (') are used for furanose numbering to distinguish them from the atoms of the base (Fig. 14.26). With pyrimidine bases, the nitrogen atom at the 1 position bonds with the 1' carbon of the sugar. With purine bases, the nitrogen atom at the 9 position bonds with 1' carbon of the sugar.

**Nucleotides** are formed by adding a phosphate group to the 5'-OH of a nucleoside (Fig. 14.27). Thus, **nucleotides are monophosphates of nucleosides**. Abridged names of some nucleotides are AMP, dAMP, UMP, dTMP and so on. Here, the first capital letter is derived from the corresponding base.



**Fig 14.25 : Bases in nucleic acids**

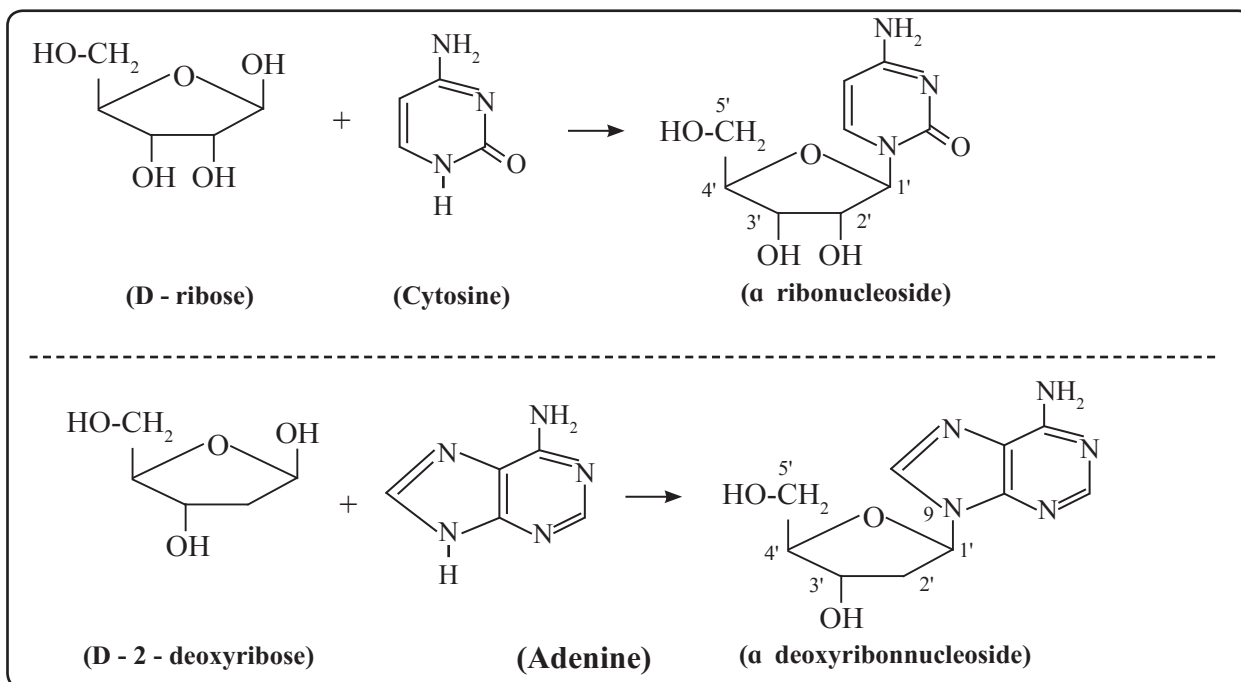


Fig 14.26 : Formation of nucleoside

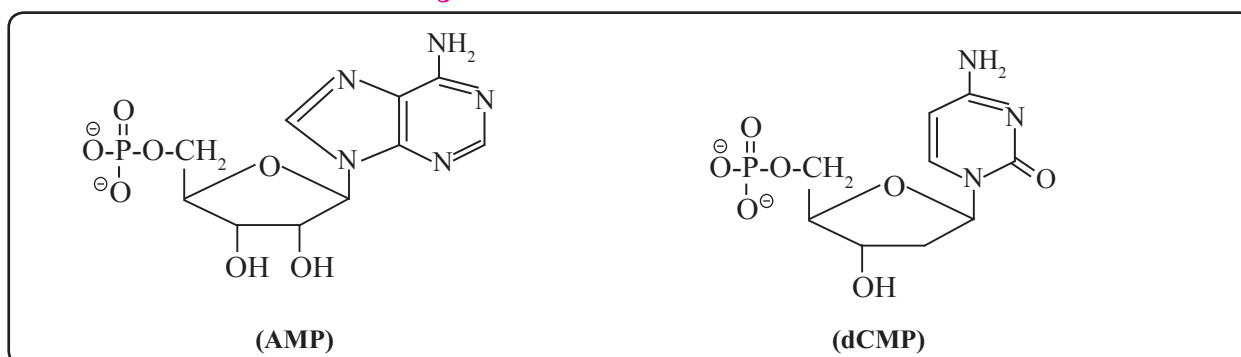


Fig 14.27 : Structures of nucleotides

### Use your brain power

Draw structural formulae of nucleosides formed from the following sugars and bases.

- D-ribose and guanine
- D-2-deoxyribose and thymine

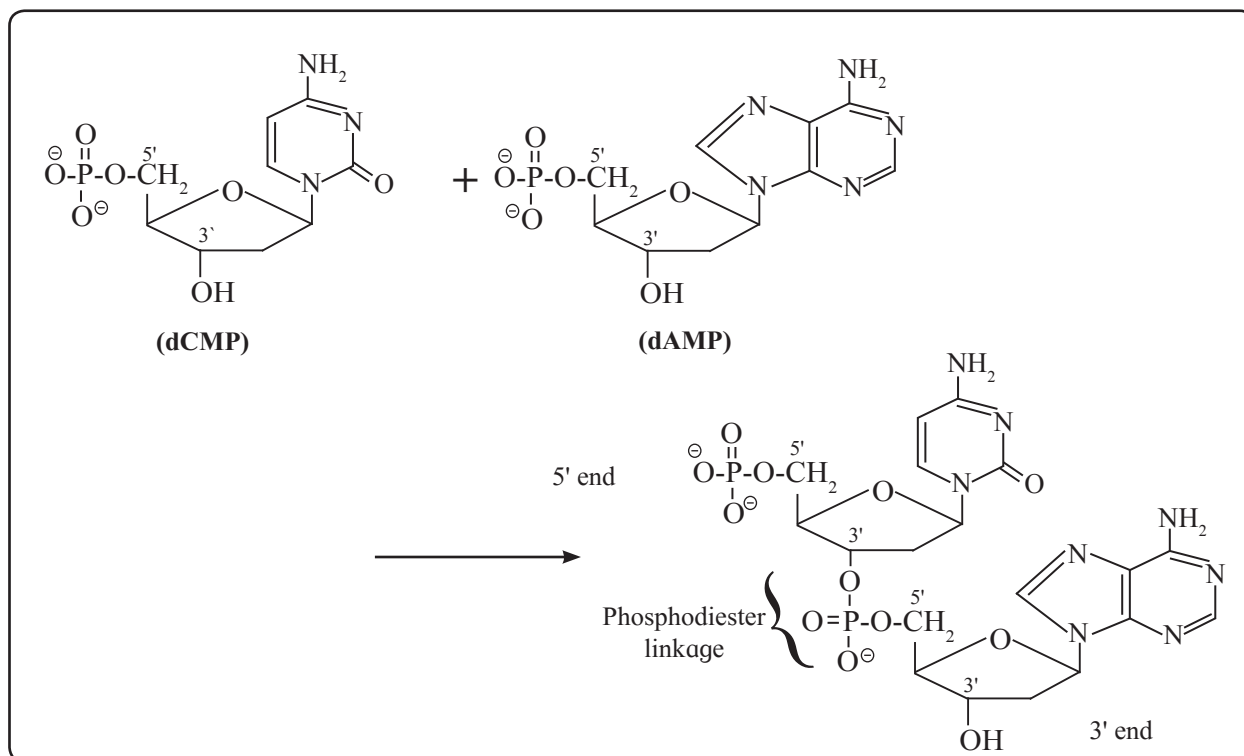
MP stands for monophosphate. Small letter 'd' in the beginning indicates deoxyribose in the nucleotide.

**14.4.2 Structure of nucleic acids :** Nucleic acids, both DNA and RNA, are polymers of nucleotides, formed by joining the 3' - OH group of one nucleotide with 5'-phosphate of another nucleotide (Fig. 14.28). Two ends of polynucleotide chain are distinct from each

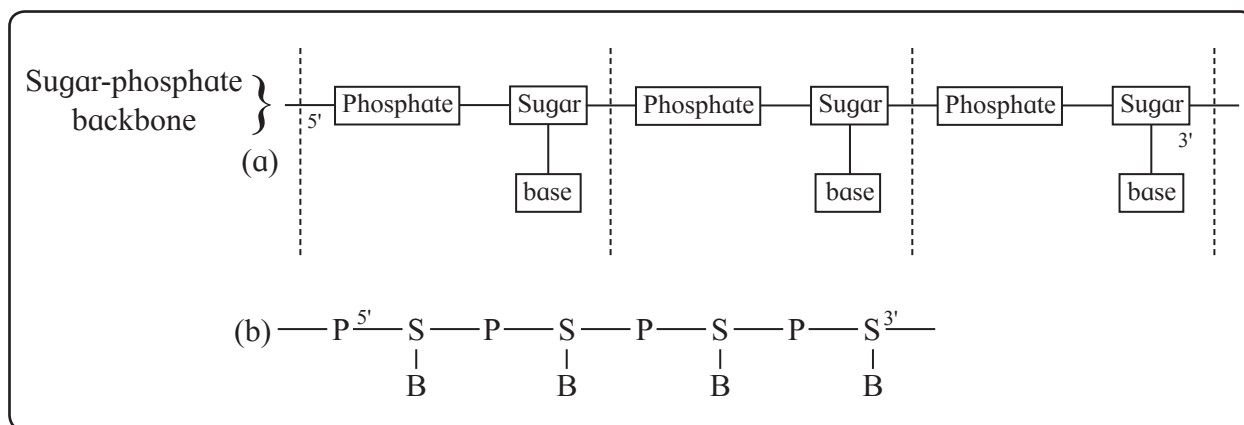
other. One end having free phosphate group of 5' position is called 5' end. The other end is 3' end and has free OH- group at 3' position.

The polynucleotide structure of nucleic acids can be represented schematically as in Fig. 14.29 (a and b).

**The primary structure** of nucleic acids is the sequence of the nucleotides in it. This, in turn, is determined by the identity of the bases in the nucleotides. Different nucleic acids have distinct primary structure. **It is the sequence of bases in DNA which carries the genetic information of the organism.** The polynucleotide chains of nucleic acids are named by the sequence of the bases, beginning at the 5' end and using the one letter symbols of the bases. For example the



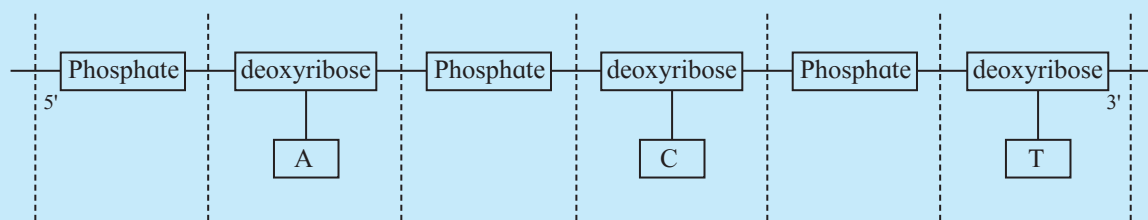
**Fig 14.28 : Formation of a dinucleotide**



**Fig 14.29 : Polynucleotide structure of nucleic acids : Schematic representations (a) and (b)**

**Problem 14.5 :** Draw a schematic representation of trinucleotide segment 'ACT' of a DNA molecule.

**Solution :** In DNA molecule sugar is deoxyribose. The base 'A' in the given segment is at 5' end while the base 'T' at the 3' end. Hence the schematic representation of the given segment of DNA is



name CATG means there are four nucleotides in the segment containing the bases cytosine, adenine, thymine and guanine, in the indicated order from the 5' end.

### Remember...

- A nucleic acid contains a backbone consisting of alternating sugar and phosphate groups.
- Backbone of all types of DNA contains the sugar 2-deoxy-D-ribose while that of RNA contains the sugar D-ribose.
- The identity and sequence of bases distinguish one polynucleotide from the other.
- A polynucleotide has one free phosphate group at the 5' end.
- A polynucleotide has a free OH group at the 3' end.

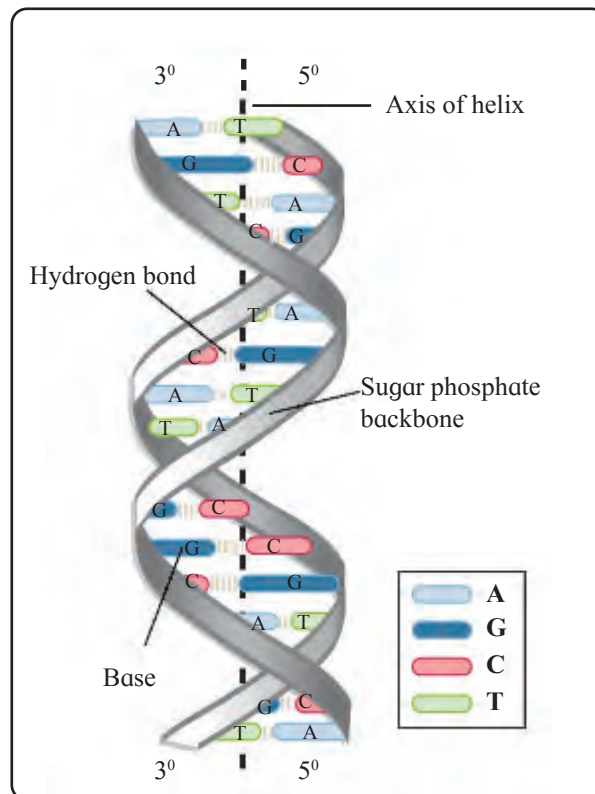


**14.4.3 DNA double helix :** James Watson and Francis Crick put forth in 1953 a double helix model for DNA structure, which was later verified by electron microscopy. Salient features of the Watson and Crick model of DNA are :

- DNA consists of two polynucleotide strands that wind into a **right-handed double helix**.
- The two strands run in opposite directions; one from the 5' end to the 3' end, while the other from the 3' end to the 5' end.
- The sugar- phosphate backbone lies on the outside of the helix and the bases lie on the inside, perpendicular to the axis of the helix.
- The double helix is stabilized by hydrogen bonding between the bases of the two DNA strands. This gives rise to a ladder-like structure of DNA double helix.
- Adenine always forms two hydrogen bonds with thymine, and guanine forms three hydrogen bonds with cytosine. Thus **A - T and C - G are complementary**

**base pairs** and the two strands of the double helix are complementary to each other.

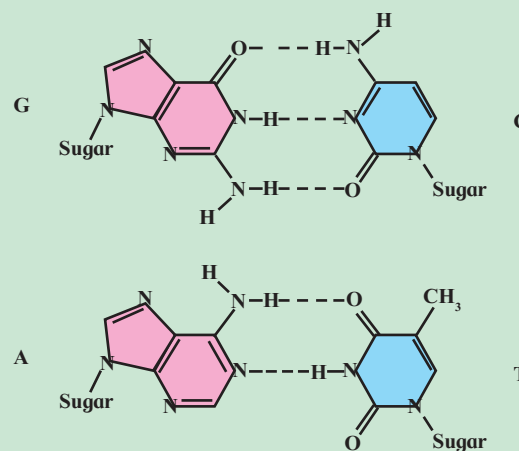
It may be noted that RNA exists as single stranded structure.



**Fig. 14.30 : DNA double helix**

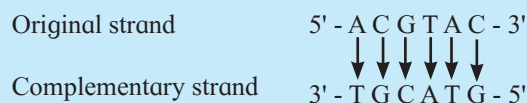
### Do you know ?

Hydrogen bonding between complementary base pairs.



**Problem 14.6 :** Write the sequence of the complementary strand of the following portion of a DNA molecule : 5' - ACGTAC-3'

**Solution :** The complementary strand runs in opposite direction from the 3' end to the 5' end. It has the base sequence decided by complementary base pairs A - T and C - G.



## Exercises

### 1. Select the most correct choice.

- $\text{CH}_2\text{OH}-\text{CO}-(\text{CHOH})_4-\text{CH}_2\text{OH}$  is an example of
  - Aldohexose
  - Aldoheptose
  - Ketotetrose
  - Ketoheptose
- Open chain formula of glucose does not contain
  - Formyl group
  - Anomeric hydroxyl group
  - Primary hydroxyl group
  - Secondary hydroxyl group
- Which of the following does not apply to  $\text{CH}_2\text{NH}_2 - \text{COOH}$ 
  - Neutral amino acid
  - L - amino acid
  - Exists as zwitter ion
  - Natural amino acid
- Tryptophan is called essential amino acid because
  - It contains aromatic nucleus.
  - It is present in all the human proteins.
  - It cannot be synthesised by human body.
  - It is essential constituent of enzymes.
- A disulfide link gives rise to the following structure of protein.
  - Primary
  - Secondary
  - Tertiary
  - Quaternary
- RNA has
  - A - U base pairing
  - P-S-P-S backbone
  - double helix
  - G - C base pairing

### 2. Give scientific reasons :

- The disaccharide sucrose gives negative Tollens test while the disaccharide maltose gives positive Tollens test.
- On complete hydrolysis DNA gives equimolar quantities of adenine and thymine.
- $\alpha$  - Amino acids have high melting points compared to the corresponding amines or carboxylic acids of comparable molecular mass.
- Hydrolysis of sucrose is called inversion.
- On boiling egg albumin becomes opaque white.

### 3. Answer the following

- i. Some of the following statements apply to DNA only, some to RNA only and some to both. Label them accordingly.
  - a. The polynucleotide is double stranded. (.....)
  - b. The polynucleotide contains uracil. (.....)
  - c. The polynucleotide contains D-ribose (.....).
  - d. The polynucleotide contains Guanine (.....).
- ii. Write the sequence of the complementary strand for the following segments of a DNA molecule.
  - a. 5' - CGTTTAAG - 3'
  - b. 5' - CCGGTTAATACGGC - 3'
- iii. Write the names and schematic representations of all the possible dipeptides formed from alanine, glycine and tyrosine.
- iv. Give two evidences for presence of formyl group in glucose.

### 4. Draw a neat diagram for the following:

- i. Haworth formula of glucopyranose
- ii. Zwitter ion
- iii. Haworth formula of maltose
- iv. Secondary structure of protein
- v. AMP
- vi. dAMP
- vii. One purine base from nucleic acid
- viii. Enzyme catalysis

#### Activity :



- Draw structure of a segment of DNA comprising at least ten nucleotides on a big chart paper.
- Make a model of DNA double stranded structure as group activity.