UNIT 5 PROTEINS

Structure

5.1 INTRODUCTION

In Units 1 – 4 of Block 1, we described the cell and its organisation. You were then introduced to some important biomolecules like carbohydrates and lipids and their significance in the functioning of the living system. Besides this, we also discussed nucleic acids and their role in the mechanism of heredity. In this unit you will be introduced to another important and complex class of biomolecules, known as proteins. These molecules are central to all biological processes and constitute half the dry mass of a cell. These molecules are made up of α-amino acids. We shall attempt to give you a broad view of how amino acid molecules join together into polypeptide chains, which in turn form the primary, secondary, tertiary and the quaternary structures of proteins. Thus, in this unit you will study the structural organisation of proteins, as a basis to understand their function as enzymes, which we shall describe in the next unit.

Objectives

After studying this unit, you should be able to:

- describe the chemistry of monomeric structural units of proteins, the α-amino acids and differentiate between different types of amino acids,
- explain how amino acids are covalently linked together to form the linear polypeptide chain, known as the primary structure of protein,
- explain the basis of diversity of structure and function of proteins, due to different amino acids present in the primary structure,
- describe the importance of hydrogen bonding in the formation of the secondary structure of proteins and explain the role of other noncovalent interactions in folding the chain to form tertiary structure of proteins, and
- explain the arrangement of folded polypeptide chains into compact structures known as quaternary structure, using myoglobin and haemoglobin as examples and their function as oxygen carriers.

5.2 BIOLOGICAL SIGNIFICANCE OF PROTEINS

The word protein, which means "of prime importance" was coined by the German chemist G.T. Mulder in 1839, as he recognised their biological importance. They are found in all eukaryotic as well as prokaryotic cells and are the most abundant biochemicals in human beings. Proteins are present in a cell in many forms, performing different tasks that maintain life. For instance, proteins in the form of enzymes act as catalysts which accelerate various chemical reactions in a living cell. We may mention here that life will not be possible without enzymes. Their role as such, is highly significant and will be described in detail in Unit 6 of this block. In addition to their role as enzymes, proteins perform other tasks which are no less essential. Along with lipids, they form the structural components of cell membranes. These molecules are responsible for structural support and also for the movement of the human body. Strong protein fibres, as connective tissue, bind skin and bone. Bones are moved by muscles, which are, in turn, made up of proteins that contract. These molecules also move chromosomes in cells and are involved in other motile and transport functions in the organism. For example, haemoglobin molecules carry oxygen from lungs to the cells. In the blood proteins help in maintaining fluid balance and also transport lipids. They are also a part of the clotting process. As antibodies, they constitute the defence mechanism of the body against infections. Some of them act as chemical messengers in the form of hormones and neurotransmitters. Proteins also act as receptors, which recognise these chemical messengers. By binding to specific regions of DNA, some proteins also regulate gene expression. Further, proteins act as poisons, such as the venom of snakes or as toxins, such as the bacterial toxin which produces botulism in human beings.

5.3 CHEMISTRY OF PROTEINS

In this section we shall briefly discuss the average chemical composition and size of proteins. Besides this we shall mainly describe the chemistry of amino acids, which are the building blocks of proteins.

Proteins are composed of carbon, hydrogen, oxygen, nitrogen and in most cases sulphur also. Although proteins vary a little in their sulphur content, they show a fairly uniform composition with regard to other elements. Table 5.1 outlines the average elemental composition of a protein.

Table 5.1 : Average elemental composition of proteins

Element	Percent
Carbon	53
Hydrogen	7
Oxygen	23
Nitrogen	16
Sulphur	1

Proteins are extremely large molecules with molecular weights in the range of 5000 to a few million. In Fig. 5.1 we have attempted to show you the relative dimensions of various protein molecules in comparison to those of glucose, which is a biomolecule, and to some inorganic ions, such as sodium and chloride ions. This large size of protein molecules gives them colloidal properties. Consequently they cannot pass through membranes of the cell and hence, presence of proteins in the urine warns us of possible damage to the kidney membranes.

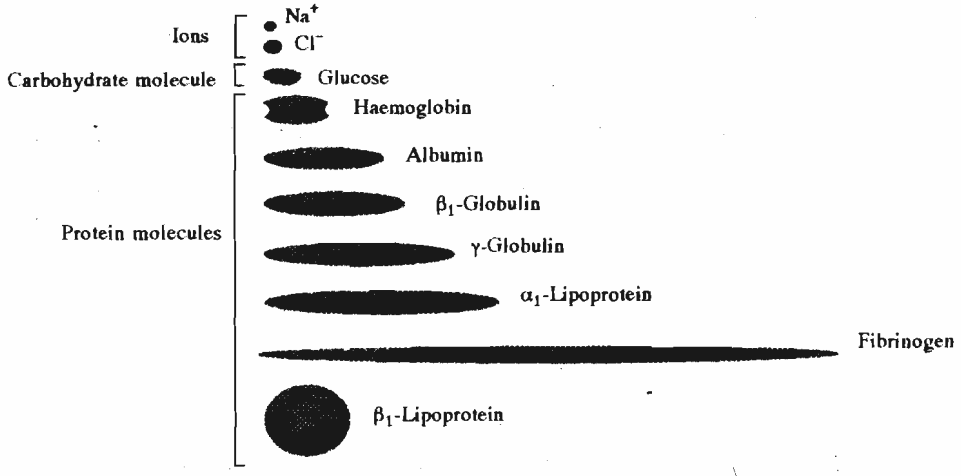


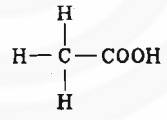
Fig. 5.1 : Comparative sizes of various protein molecules in comparison to sodium and chloride ions and glucose molecule

Let us first describe the basic building blocks of a protein molecule.

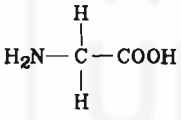
5.3.1 Amino Acids : Building Blocks of Proteins

When we consider that proteins, in the role of enzymes, catalyse thousands of different biochemical reactions and in addition, also perform a multitude of other functions in the living organism, it is obvious that such diversity in their functions must result from an enormous variety of distinct structures. One of the aims of this unit is to explore the basis of this structural diversity and for this purpose, it is necessary to understand the basic building blocks of proteins, the α-amino acids. Let us first try to recollect some basic facts about amino acids.

An amino acid is an organic acid molecule containing an amino ($-NH_2$) group. For example, glycine, a simple amino acid can be generated if one hydrogen atom on the carbon next to the carboxyl group in acetic acid is replaced by an amino group :

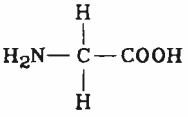


Acetic Acid

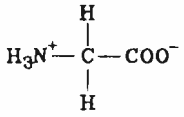


Glycine

An amino acid is usually represented in its unionised form, to show the presence of both amino and the carboxyl groups. However, it is the ionised form which predominates under physiological conditions of pH (around 7). Thus an amino acid like glycine, can be represented in its unionised and ionised form as :



Glycine : unionised form

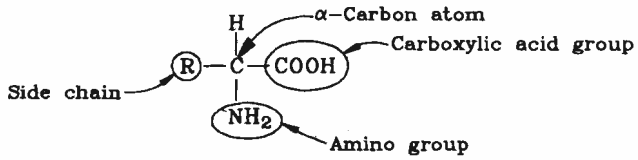


Glycine : ionised form

Further, the carbon atom next to the carboxyl group is known as the α-carbon. Those amino acids which have their amino group attached to this α-carbon, are known as the α-amino acids. The biologically significant amino acids are principally, α-amino acids.

All α-amino acids have a common structural motif. They have an amino group at one end and a carboxyl group at the other, both being attached to the α-carbon atom, which also

carries a hydrogen atom. The fourth group on the α-carbon is known as the side chain and is denoted as R. In the simplest of α-amino acids i.e., glycine, the side chain is a hydrogen atom only. In other α-amino acids, this side chain consists of more complex chemical groupings. All α-amino acids differ from each other in the nature of their side chains only. Thus, the general formula of an α-amino acid can be represented as :



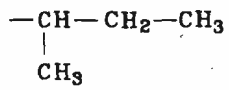
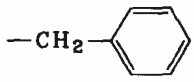
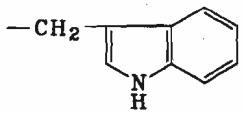
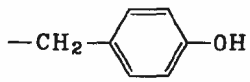
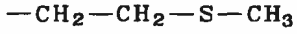
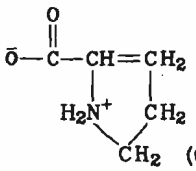
Now let us describe the structure of important amino acids which are normally found in proteins. We shall also group these amino acids according to the nature of their side chains.

5.3.2 Structure of the α-Amino Acids

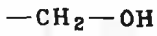
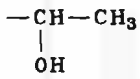
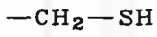
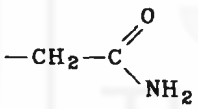
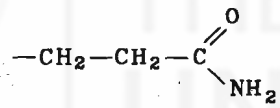
As we have mentioned earlier, amino acids differ from one another with respect to the R groups i.e., the side chains, that are bonded to the α-carbon atom. Most living organisms contain twenty α-amino acids which constitute the vast variety of proteins. In Table 5.2 we have listed these amino acids and illustrated the nature of their side chains. We shall be using the three letter abbreviations for these amino acids in this unit. In addition, one letter abbreviations are also in use. Both these abbreviations are shown in Table 5.2. You would observe from this table that amino acids can be categorised into three main types, on the basis of properties of their side chains. These are amino acids with nonpolar, uncharged polar and charged polar side chains, respectively. The nonpolar group has generally an aliphatic or an aromatic side chain, whereas the uncharged polar types have side chains containing hydroxyl, amide or sulphydryl groups. The charged polar group can be categorised as acidic or basic, depending whether the side chain contains a carboxyl or a basic group. Though tyrosine and tryptophan have nonpolar side chains, they can also be put in the uncharged polar group. Their feeble polarity arises from the presence of – OH group in tyrosine side chain and – NH group in the tryptophan side chain. Polar side chains, whether charged or uncharged, have a strong tendency to inter :t with polar molecules like water, which is the normal medium in which most of the proteins function. Nonpolar side chains on the other hand, shun water molecules and tend to seek a nonpolar environment. This difference in interaction of polar and nonpolar side chains with water, is of great significance for the folding of a linear protein chain into diverse structures, which are characteristic of different functions in the organism.

Table 5.2 : Structures of amino acids normally found in proteins

Amino acid	Abbreviation	Structure of R (side chain) in $H_3N^+-\underset{R}{\underset{\vert}{C}}(H)-C(=O)-O^-$
Nonpolar R group		
Glycine	Gly (G)	$-H$
Alanine	Ala (A)	$-CH_3$
Valine	Val (V)	$-CH(CH_3)_2$
Leucine	Leu (L)	$-CH_2-CH(CH_3)_2$

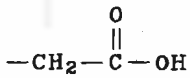
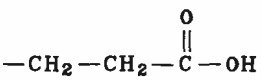
Isoleucine	Ile (I)	
Phenylalanine	Phe (F)	
Tryptophan	Trp (W)	
Tyrosine	Tyr (Y)	
Methionine	Met (M)	
Proline	Pro (P)	 (Complete structure)

Polar, uncharged R group

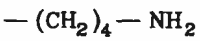
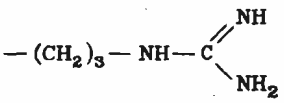
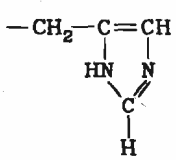
Serine	Ser (S)	
Threonine	Thr (T)	
Cysteine	Cys (C)	
Asparagine	Asn (N)	
Glutamine	Gln (Q)	

Polar, charged R group

Acidic R group

Aspartic acid	Asp (D)	
Glutamic acid	Glu (E)	

Basic R group

Lysine	Lys (K)	
Arginine	Arg (R)	
Histidine	His (H)	

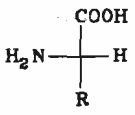
We shall now briefly describe the stereochemistry of amino acids.

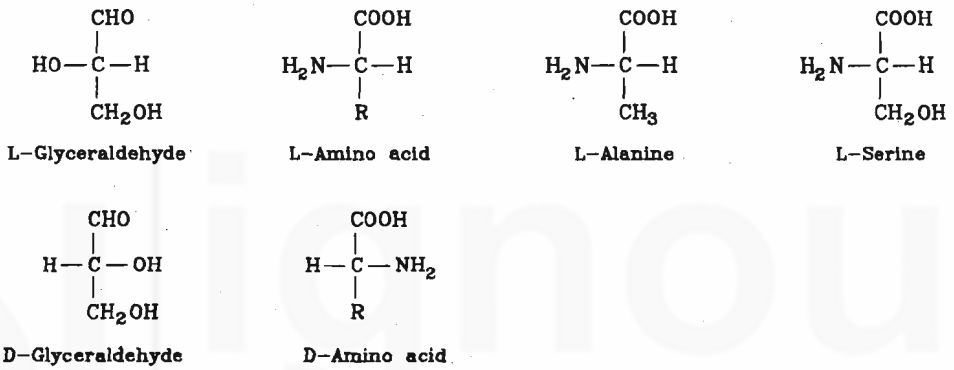
Optical Activity and Configuration of Amino Acids

You would have observed that all the amino acids discussed above, with the exception of glycine, have a chiral or an asymmetric carbon atom, i.e., they have a carbon atom bonded to four different atoms or groups tetrahedrally. In case of amino acids under discussion, it is the α-carbon which is a chiral centre, and you may recall here that molecules containing chiral carbons can exhibit optical activity. Therefore, all amino acids, except glycine, are optically active, i.e., they can rotate plane polarised light either to right or to the left, and are designated as (+) and (–) isomers respectively. Now let us describe the configuration, i.e., the spatial arrangement of atoms or groups, at the α-carbon atom in the amino acids. As in the case of carbohydrates, the configuration at the chiral carbon is designated by the letters D and L. You may recall from Section 2.3 that these notations are used after a comparison with the configuration of glyceraldehyde.

In the case of α-amino acids, the configuration at the α-carbon is compared to the configuration of glyceraldehyde, as shown below:

The Fischer projection formula of L-amino acids can be shown as:





As depicted above, in L-amino acids the —NH_2 group is on the left, and in the D form the —NH_2 group is on the right side. All amino acids found in proteins have the L-configuration and are designated either as L(+) or L(–), depending on whether they rotate the plane polarised light to the right or to the left. Although D-amino acids do occur in nature, they are never present in proteins. In comparison, common carbohydrates generally occur in D-configuration.

Out of the twenty amino acids listed in Table 5.2, some are referred to as "essential amino acids". Let us first discuss which of the amino acids are essential for our body.

Essential Amino Acids

Corn protein is low in lysine and tryptophan. Rice is low in lysine and threonine. Wheat is low in lysine. Vegetarian diet people must eat a combination of vegetables/pulses, so that they can get an adequate supply of essential amino acids and thus avoid protein deficiency diseases. These diseases are commonly found in those places where a single plant, such as corn, is the only or major source of dietary protein.

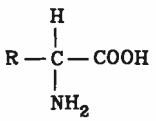
Amino acids are mainly components of protein molecules and free amino acids are not widely distributed in nature. Complete hydrolysis of all proteins produces the twenty L-α-amino acids, described in Table 5.2. Our body can synthesise only ten of these amino acids. The rest which cannot be synthesised by humans, must be supplied by our diet. The latter category of amino acids are known as **essential amino acids**. A balanced diet containing both animal and plant proteins is generally recommended, because it provides all the essential amino acids. The dietary proteins are hydrolysed during digestion to release amino acids which are then reassembled into human proteins. The essential amino acids are luecine, isoleucine, lysine, methionine, phenylalanine, tryptophan, threonine, valine, histidine and arginine. The last two amino acids are essential only during infancy. Proteins that contain all the ten essential amino acids are called as **adequate proteins**. Milk and animal proteins are an excellent source of all amino acids, whereas plant proteins lack one or more of the essential amino acids. However, the protein in soybeans is an exception, and is an adequate protein.

We shall now describe the acid-base properties of amino acids. The amino acid structure, which we have shown up to this stage, is how it would appear in an organic, nonpolar

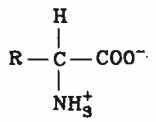
solvent. You would certainly wonder as to how this molecule would appear in an aqueous solution i.e., in a polar medium, since the molecule contains an acidic as well as a basic group. So let us study the behaviour of an amino acid in a polar medium.

Acid-Base Properties of Amino Acids

Since proteins are composed of amino acids (we shall be describing more about this in Section 5.4), a knowledge about the acid-base properties of amino acids will help us in understanding some of the properties of proteins. We have already mentioned that all amino acids contain at least one carboxyl group and one amino group. Both the groups are capable of ionisation in an aqueous solution. Thus in water, amino acids can act as acids and bases. Molecules which exhibit this property are known as **amphoteric** molecules. In other words this means an amphoteric substance can accept as well as donate a hydrogen ion (H^+). At or around neutral pH, an amino acid forms a **dipolar ion** or a **zwitterion**, as the acidic carboxyl group donates a hydrogen ion and the basic amino group accepts one.

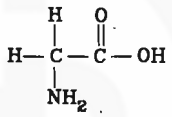


Unionised amino acid

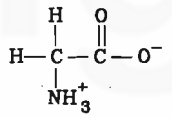


Ionised amino acid
(zwitterion or dipolar ion)

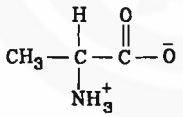
Thus a zwitterion has a positive as well as a negative charge within the same molecule at neutral pH. As the normal biological environment in which amino acids are generally present, is aqueous medium at near neutral pH, they are found as zwitterions only, though we may continue to show these molecules in the unionised form in this unit for the sake of convenience. Now let us show some amino acids in their zwitterion form:



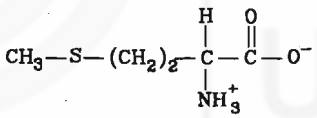
Glycine (unionised)



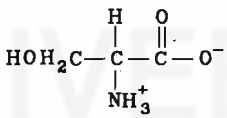
Glycine (ionised, zwitterion)



Alanine



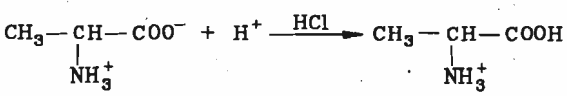
Methionine

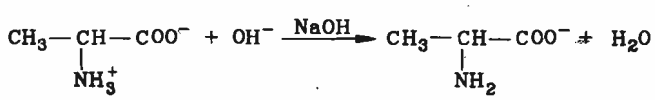


Serine

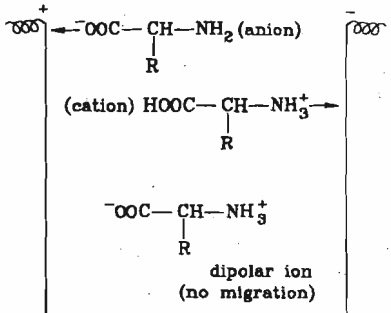
You would have observed from the above that zwitterions have one positive and one negative charge and are, thus, electrically neutral. Therefore, these dipolar ions do not migrate in an electric field. The pH at which amino acids (and also proteins) carry no net charge is called its **isoelectric point (pI)**. For example, the isoelectric point for glycine is pH 6.0, which means that at this pH, glycine will be in its zwitterion form.

Let us now describe briefly how the amino acid molecules (and consequently the proteins) behave in acidic or basic medium. These amphoteric compounds can react with an acid or a base to form a salt. For example, study the reactions for zwitterion of alanine shown below:



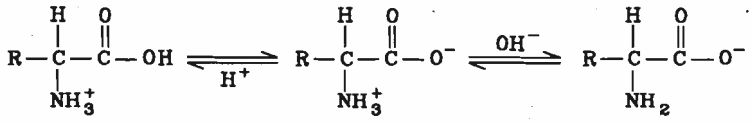


Proteins and amino acids can be separated by a process called **electrophoresis**. It involves placing a mixture of proteins or amino acids in an electric field at constant pH. Molecules with a net charge will move towards the opposite electrodes, whereas the dipolar ions, with a net zero charge will not migrate.



Proteins have minimum solubility at isoelectric point and the electrically neutral molecules can form aggregates, thus rendering their removal from the solution quite easy. This has a practical utility in our day to day life. For example, casein is a protein found in cow's milk. It has an isoelectric point of pH 4.7. The normal pH of cow's milk is 6.3. In the cheese production process lactic acid, produced by bacteria, lowers the pH, which in turn lowers the solubility of casein, thus causing the milk to curdle.

We can observe from these reactions that addition of H^+ or OH^- to a zwitterion results in a net positive or a negative charge on the molecule. **Thus depending on the pH of the aqueous solution, amino acids (as also the proteins) can exist in different ionic forms:**



Acidic pH | pH around neutrality (zwitterion) | Basic pH

This property of amino acids (and hence the proteins) to behave as weak acids (due to — COOH group) or as weak bases (due to — NH_2 group) is employed effectively in nature, as these molecules act as buffers in body fluids. Thus one of the important functions served by proteins in the blood is to maintain its pH in the narrow range of 7.35 to 7.45.

Substances which resist sudden changes in pH of a solution are known as **buffers**. Recall that solutions containing a weak acid (e.g., acetic acid) and its salt (e.g., sodium acetate) function as buffers. Same is true for weak bases and their salts.

We may mention here that only those amino acids with ionisable side chains (such as glutamic acid or lysine) or amino acids at the extreme ends in the proteins with free α-amino and α-carboxyl group, contribute to the acid-base properties of a protein molecule. This is because the amino acids forming the polypeptide chain of proteins (we shall discuss this in subsection 5.3.3) do not have free α-amino and α- carboxyl groups, except at the two terminals of the polypeptide chain. **Some of the side chains (R groups) which are capable of ionisation, will increase the number of ionised forms possible for that particular amino acid.** This is highly important in the overall structure and function of proteins. The tendency of a group to ionise is expressed in terms of pK_a value, which is analogous to that of pH. The lower the pH of a solution, the more acidic it is and larger is its hydrogen ion concentration. In a similar way, lower pK_a value for an ionisable group implies stronger acidic nature of that group and thus more readily it will ionise.

$pK_a = -\log K_a$
$pH = -\log [H^+]$

Before proceeding further attempt the following SAQs.

SAQ 1

Tick [√] mark the appropriate answer.

Amino acids owe their variety to

a) their α-carboxyl group [ ]

b) their differing molecular weights [ ]

c) their α-amino group [ ]

d) the nature of their side chains [ ]

SAQ 2

Tick [√] mark the appropriate answer.

Amino acids are called amphoteric because

a) they have a basic group [ ]

b) they have an acidic group [ ]

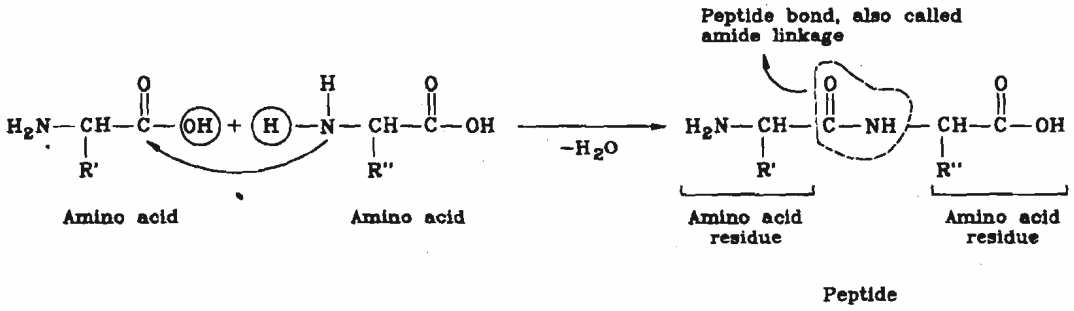
c) they have both acidic and basic groups [ ]

d) none of the above [ ]

Now let us describe an important reaction of amino acids, which is the basis for the formation of proteins, which we shall study in the next sections.

Formation of the Peptide Bond

This is the most important reaction of amino acids and is a condensation reaction between two amino acid molecules, as shown:



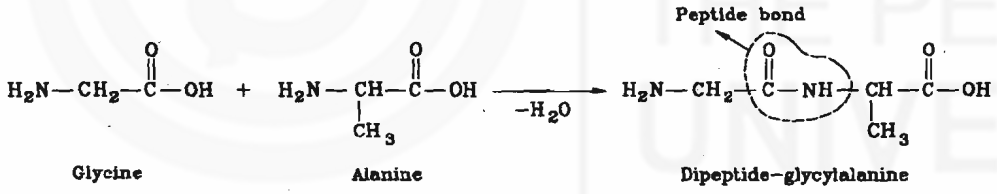
You will recall that amino acids shown here will actually be in ionised form.

You would observe from the above reaction that **peptide bond** formation results in a linkage between the α-carboxyl group of one amino acid and the α-amino group of another amino acid molecule, with the elimination of a water molecule. The resultant molecule is known as a **peptide** and the constituent amino acids in it are called **amino acid residues.** The peptide bond can undergo hydrolysis with an acid or a base or with specific enzymes to give the constituent amino acids.

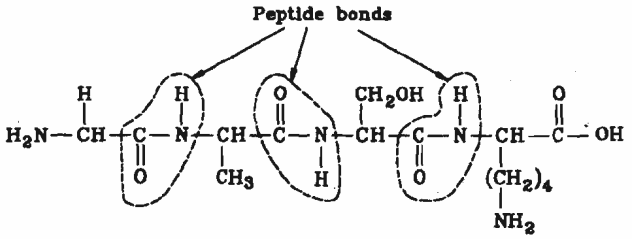
Now let us know more about peptides. This will help you in learning about the larger and complex macromolecules–proteins.

5.3.3 Peptides

From the preceding section you learnt that a peptide is formed by amino acid residues which are held by a peptide bond. Depending upon the number of amino acid residues in a peptide, they are known as dipeptides, tripeptides, etc. For example, a **dipeptide** has two amino acid residues:

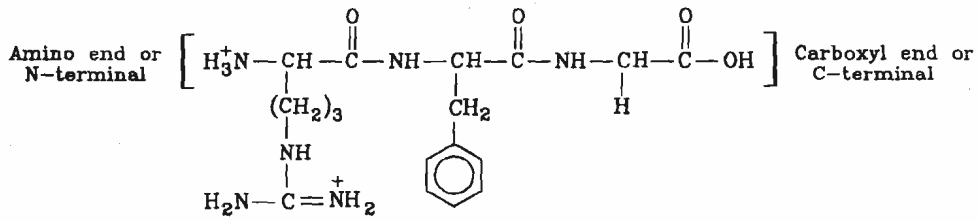


Similarly, a **tripeptide** results from the union of three amino acids and a **tetrapeptide** links together four amino acids. The structure of a typical tetrapeptide is shown below:



Generally, peptides with more than 10 amino acid residues are known as **polypeptides.** We may mention here again that amino acids in a peptide are called residues, since they are residual portions of amino acids left after elimination of water molecules during the formation of peptide bonds. You should also note that a peptide is named from the amino

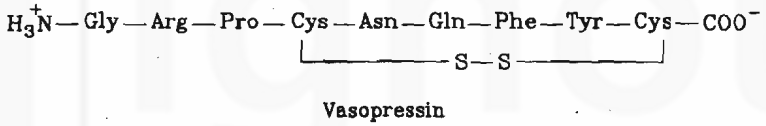
end, which is the end having free α-amino group. For example, in the above tetrapeptide the amino end (also called the N-terminal) is a glycyl residue. Similarly, the other end having a free α-carboxyl group is called the carboxyl end or C- terminal. In the above case, a lysine residue is at the C-terminal. This will be further clear from the tripeptide, arginylphenylalanylglycine, shown below:



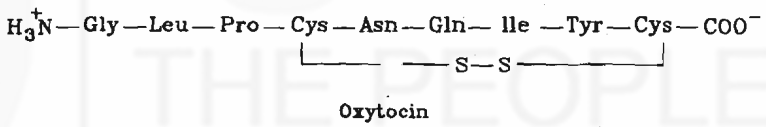
Tripeptide - Arginylphenylalanylglycine

You must have also observed that it is impractical to write the large structure of a peptide, especially of a polypeptide. In such cases, the structures can be easily represented by using the three letter or one letter abbreviations of amino acids, shown in Table 5.2.

Many peptides occur in nature, each performing a specific role in an organism. For example, vasopressin and oxytocin are two peptide hormones secreted by the pituitary gland in humans. Their structures are shown below:

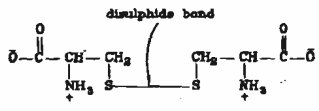


Vasopressin



Oxytocin

Cysteine is often found as a dimer which is bonded through a disulphide bond and is known as **cystine** :



Cystine

We shall be discussing the disulphide bond in Section 5.4.

In addition to the peptide linkages, both of these peptides also contain a disulphide bridge between two cysteine residues. These peptides perform important biological roles. Vasopressin helps in controlling fluid balance in humans by promoting water reabsorption by kidneys, thereby controlling blood pressure by regulating the amount of water in circulation in the body. Oxytocin promotes lactation. It is also used as a drug during child birth to induce labor.

By this stage you have learnt how the building blocks of proteins, the amino acid molecules, are constructed into a polypeptide chain by the peptide bonds. Let us, therefore, describe the structure of proteins in the following sections.

You can check your understanding of peptides by attempting the following SAQ.

SAQ 3

Draw the complete structure for the tripeptide Asp. Leu. Trp.

5.4 STRUCTURE OF PROTEINS

A protein is constructed primarily as a linear polymer of different amino acids, called the polypeptide chain, and we have already described the formation of a polypeptide. The distinction between a polypeptide and a protein is tenuous. But any naturally occurring polypeptide of molecular weight more than 5000 is generally called a **protein**. In other words, it implies that as the polypeptide chain grows into a biopolymer of molecular weight of thousands and more, the term protein becomes more appropriate to use. However, the enormously large size of proteins makes it obvious that they should have extremely complex structures. The structure of a protein molecule will determine its biological role and other properties. It has been found that every protein has a unique three dimensional shape. The complete structure, including specific shape of a protein molecule is discussed under four headings, also called levels of organisation. These are the primary, secondary, tertiary and quaternary structures, respectively. We shall now describe these structures.

If the protein chain would consist of say 100 amino acids and each amino acid would happen to be identical, then only one particular type of protein would be possible. However, if each position in such a chain is occupied by any of the twenty amino acids listed in Table 5.2, then the number of structurally different proteins will be of the order of 20^{100} or almost limitless. The enormity of this structural variation can easily explain the functional versatility of proteins.

5.4.1 Primary Structure

As mentioned above, a protein is assembled as a linear long chain polymer of amino acids, called the polypeptide. The number and sequence of amino acid residues from N-terminal to the C-terminal, which are linked linearly by the peptide bonds, is referred to as the **primary structure** of a protein molecule. In short, the **complete covalent structure of the polypeptide constitutes the primary structure**. For example, the primary structure of glucagon, a protein which stimulates the conversion of glycogen to glucose in the body, is shown below:

Since the amino and carboxyl groups attached to the α- carbon of an amino acid are involved in repetitive amide linkages in the protein chain, except at the extremities of the chain where they are free, it is only the side chains or the R groups which represent variations in the amino acids constituting a protein molecule.

His–Ser–Glu–Gly–Thr–Phe–Thr–Ser–Asp–Tyr–Ser–Lys–Tyr–Leu–Asp–Ser–Arg–Arg–Ala–Glu–Asp–Phe–Val–Glu–Trp–Leu–Met–Asn–Thr
1 10 20 29

It is easy to realise that the sequence of amino acids in the protein chain would determine its physical, chemical and biological properties. Even a minor change in the amino acid sequence results in major effects on the properties of that protein. For example, haemoglobin molecule consists of four polypeptide chains containing in all 574 amino acid residues. This molecule is present in blood and carries oxygen. Changing just one amino acid in one of the chains, results in a defective haemoglobin molecule that is found in patients with sickle cell anemia.

In sickle cell anemia, the haemoglobin molecule carries leucine instead of isoleucine in one of the protein chains. On its return trip to lungs, after delivering oxygen to the tissues, the defective (or mutant) haemoglobin molecules polymerise into large strands and clog up the smaller veins and capillaries, with devastating results.

In essence, the primary structure of a protein corresponds to the structural formula of an organic compound. However, the structural formula of a small organic compound conveys enough information about its chemical function, whereas knowledge of the primary structure of a protein gives no such clue. Linear sequences of amino acids, no matter how different, are one dimensional and hence cannot explain the stupendous variety of protein functions. On the other hand, sequence of amino acids in a polypeptide chain of a protein, determines the complexity of its folding pattern, which holds the secret of its function. We shall examine in the following sections the general principles that govern the folding of the polypeptide chains of proteins into more complex structures.

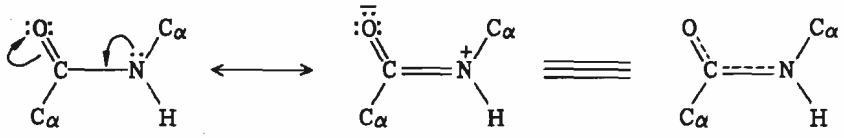
The polypeptide chains forming the primary structure of proteins, fold to form higher order protein structures.

You should note that two strong covalent interactions, namely the peptide and the disulphide bonds, are involved in the integrity of the primary structure of a protein, i.e. of the polypeptide chain. On the other hand, the folding of the polypeptide chain, i.e. the higher order structure of the protein, is stabilised by the collective effect of a large number of weak noncovalent interactions. Further, the folding patterns of the polypeptide chain depend on the geometry of the peptide bonds, which link the individual amino acids. We have already described the peptide bond in subsection 5.3.2. Let us now learn what influence this bond has on the protein folding.

5.4.2 Planarity of the Peptide Bond

The geometrical dimensions of the peptide bond impose certain restrictions on the folding of the polypeptide chain. Pauling and Corey were the first to discover that the

peptide linkage is planar. This is due to resonance between the two canonical forms of peptide group as shown below:



Thus, due to delocalisation of electrons both C═O and C—N bonds of the peptide linkage have a partial double bond character and like any unsaturated system, the peptide linkage is capable of exhibiting geometric isomerism. It has been found to exist in the *trans* configuration (as the bulky α-carbons are farther from each other than they are in *cis* configuration). Another consequence of electron delocalisation is that all the atoms of the peptide linkage i.e., O,C,N,H lie in one plane. Further O and H atoms are in the *trans* position.

The individual bonds constituting the peptide unit are, thus, torsionally rigid (this makes intrabond hydrogen bonding impossible). However, twisting of the peptide plane is still possible around the C_α— N and C_α— C bond axis, as shown in Fig. 5.2.

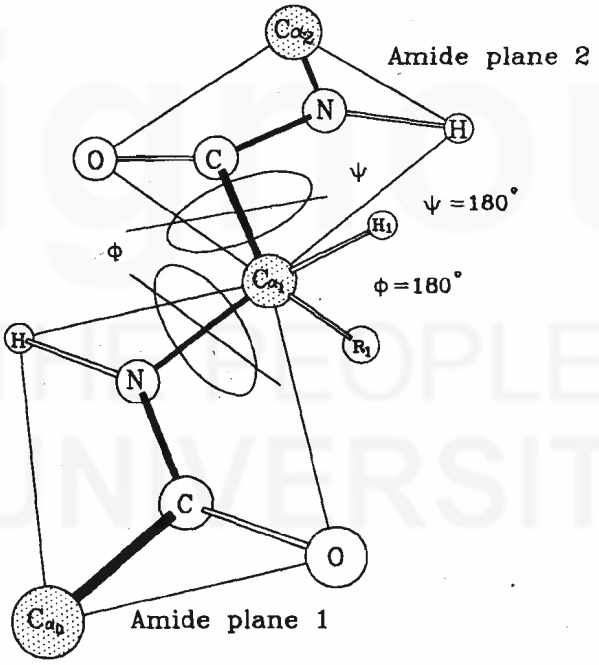
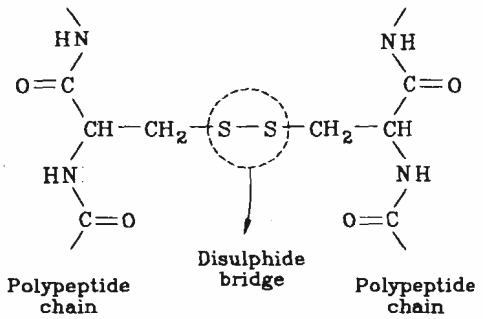


Fig. 5.2 : A portion of a polypeptide chain showing the rotation of adjacent peptide planes and the corresponding angles of rotation about the bonds to the central α-carbon atom. $\phi = \psi = 180^\circ$ corresponds to the fully extended polypeptide chain.

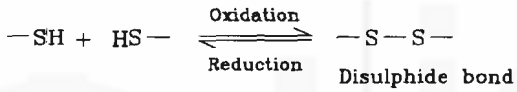
The angles ϕ and ψ indicate the rotational movement of the peptide planes around N—C_α and C_α— C bonds respectively. When the chain is fully extended, the ϕ and ψ values are considered to be equal to 180°. As viewed from the central α-carbon, any clockwise rotation of the peptide plane through 180° is assigned a positive value and any anticlockwise rotation is considered negative. Thus each peptide plane can twist and rotate around the α-carbon bonds. The relative angles of rotation will define the direction and type of folding of polypeptide chain. Some values for angles ϕ and ψ bring unbonded atoms too close to each other, causing steric repulsions, and are consequently "not allowed". It is easy to realise that this would substantially restrict the number of conformations a polypeptide chain could have.

We shall now describe the other covalent interaction, which is involved in stabilising the protein structures. This is known as the **disulphide bond.**

Disulphide bonds or bridges are extremely important in stabilising higher order protein structures. This bond may connect two polypeptide chains through cysteine residues as shown below :



You will recall that the amino acid, cysteine contains a thiol group (—SH). This group easily undergoes an oxidative reaction with another thiol group and results in a disulphide bond:



As you can see this is a reversible reaction under reducing conditions. Disulphide bridges are relatively stable and can also connect different parts of the same polypeptide chain, as in vasopressin or oxytocin (subsection 5.3.3). In such cases, they impose some restrictions on the folding pattern of the polypeptide chains. At the same time, they impart stability to the finally folded conformation.

Now let us describe the noncovalent interactions which are also important in stabilising the protein structures.

5.4.3 Importance of Weak Noncovalent Interactions in Protein Folding

As we have already discussed, the primary structure of a polypeptide chain is maintained by covalent amide bonds, capable of withstanding random thermal motion. Another, covalent linkage that maintains a higher order protein structure is the disulphide bond. However, the folding of a polypeptide chain to form secondary, tertiary and quaternary structures (we shall be describing them in next section) is mainly due to the cooperative action of a multitude of weak noncovalent interactions. Singly, these interactions are weak and easily disrupted by thermal motion. However, the cumulative effect of a large number of these interactions is considerable. These, acting in concert not only provide reasonable stability to higher order protein structures, but also confer on them a degree of flexibility, so essential for biological function.

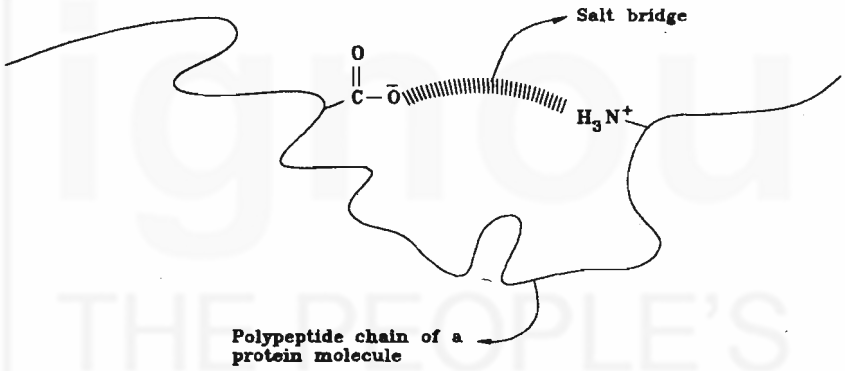
The weak noncovalent interactions that contribute to the stability of protein structures have been identified as electrostatic forces, van der Waals forces, hydrogen bonds and hydrophobic interactions. We shall describe these forces briefly. You will find an estimate of the energies involved in these interactions and the covalent disulphide bonds in Table 5.3.

Table 5.3 : Energies of noncovalent interactions and of a disulphide bond

Type of interaction or bond	Interaction or bond energy (kJ mol^{-1})
Electrostatic interactions	12.5 – 20.9
Hydrogen bond	12.5 – 20.9
van der Waals dispersion interactions	1 – 5
Hydrophobic interactions	2.8 per CH_2 group
Disulphide bond	200

You will now briefly learn what electrostatic interactions are:

The amino acid side chains of many proteins, such as those of lysine or glutamic acid, carry charged groups like $-N^+H_3$ or $-\overset{\overset{O}{\|}}{C}-O^-$ at neutral pH. The attractive forces generated by the proximity of such charges on a polypeptide chain constitute the **electrostatic bonds**, which are also known as **ionic bonds** or **salt bridges**.

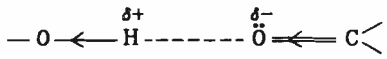


This bond would thus contribute to the folding process of a polypeptide chain. As would be expected from an ionic bond, the interaction of a salt bridge is negligible at the interface of a protein molecule with water, where the dielectric constant would be very high. However, its contribution to stability is significant in the protein interior, which is relatively less accessible to water.

Another weak interaction involved in protein folding is **van der Waals dispersion forces**. This is due to the attractive force between uncharged molecules and comes into effect when two atoms approach each other so closely that their electron clouds penetrate each other. Though repulsive forces also develop, at van der Waals contact distance, the attractive forces predominate. Individually these dispersion interactions are of very low energy, but their cumulative effect, when summed up over interacting protein surfaces, is considerable and would be of great significance in the maintenance of the folded state of proteins.

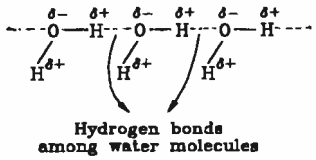
Now let us describe hydrogen bond, an important linkage in biological structure and function. It is a weak noncovalent interaction.

Hydrogen bond results from an electrostatic interaction between a hydrogen atom bound covalently to an electronegative atom like O, N or S and a second electronegative atom which has a lone pair of electrons available.



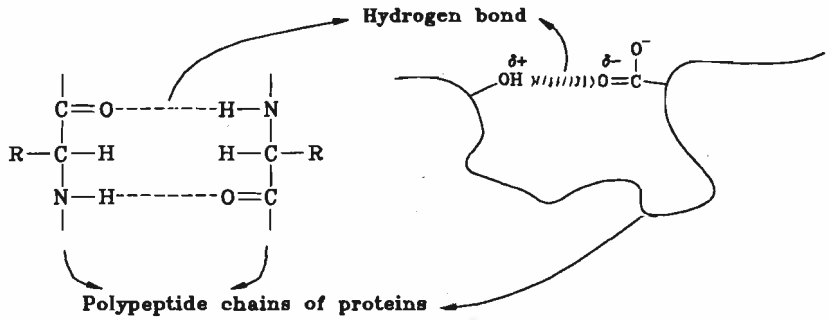
The H atom of —OH group acquires a partial positive character since the electron cloud is attracted more towards the electronegative oxygen. The positively charged H atom thus interacts electrostatically with the unshared electron pair on the oxygen of the C=O bond. This interaction is called the **hydrogen bond**. You would observe that the net effect of this linkage is sharing of a hydrogen atom between two electronegative atoms. The strength of the hydrogen bond depends on the electronegativities of the two sharing atoms. The more electronegative these are, the stronger is the hydrogen bond.

In water hydrogen bonding arises between hydrogen atom of one molecule and the oxygen atom of another molecule:



Hydrogen bonds among water molecules

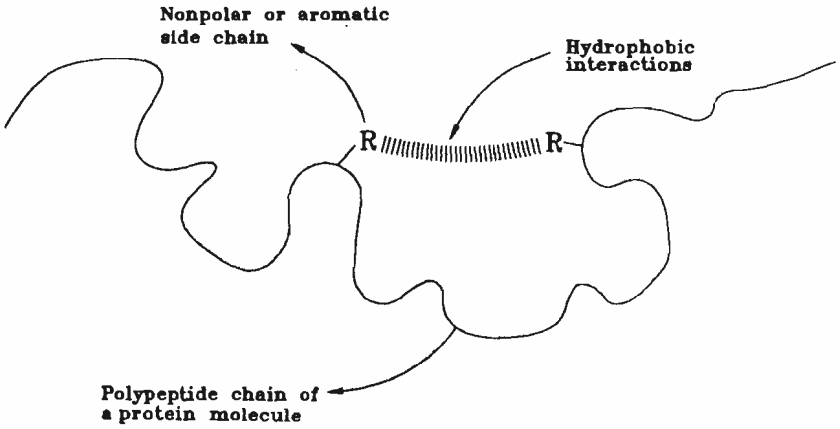
In proteins hydrogen bonds result when hydrogen atoms are shared between the electronegative nitrogen and the carbonyl oxygen of peptide bonds, within the same or different polypeptide chains. They are also formed between the side chain —OH groups and the carbonyl group of another amino acid residue.



The importance of hydrogen bonds in the stability of protein structures is, however, debatable. This is because water, which is the universal biological medium, forms hydrogen bonds with other water molecules, as well as with the appropriate groups on the polypeptide chain. Thus before the folding process starts, the amino acid side chains of proteins which are capable of forming hydrogen bonds, would form these linkages with water molecules first. During the folding of protein chains these hydrogen bonds with water molecules would have to be broken and new ones formed between protein side chains. Thus the stabilisation provided by hydrogen bonds in higher order protein structures, is the difference between the energy required to break the hydrogen bonds with water and the energy obtained in forming inter-residue hydrogen bonds, which is marginal. Nonetheless, their importance in protein folding arises due to their extremely large numbers.

Let us now learn what a **hydrophobic interaction** means and what is its contribution towards stabilising proteins.

Hydrogen atoms of aliphatic side chains of proteins, such as those of valine or isoleucine are attached to carbon and the electron pair forming the bond is equally shared between the two. Therefore, these hydrogen atoms do not carry any fractional positive charge. Because of this aliphatic side chains do not interact with water molecules via hydrogen bonds and are, therefore, called as nonpolar. When such nonpolar side chains are exposed to water, the hydrogen bonded network of water molecules gets disturbed. Water consists of a loose fluctuating cluster of hydrogen bonded molecules which constantly change their partners in hydrogen bonding. Since water molecules cannot interact with a nonpolar side chain immersed in it, they are forced to form more hydrogen bonds with each other, resulting in a highly ordered structure around the nonpolar side chains. This increase in local order results in a decrease in entropy of water locally, which is thermodynamically unfavourable. The nonpolar side chain is thus forced to avoid exposure to water and to seek interaction with its own kind in the interior of the protein molecule, resulting in the liberation of water molecules and consequent increase in the entropy of the system. The folding of the polypeptide chains thus proceeds in such a way that all the nonpolar side chains are squeezed into the interior close to each other and the polar and charged side chains are brought on to the surface of the molecule, where they can have favourable interaction with water molecules. The free energy gained by the transfer of a $—CH_2$ group of a nonpolar side chain from aqueous surroundings to the protein interior is of the order of 2.8 kJ mol^{-1}.



Aromatic amino acid residues, such as those belonging to phenylalanine, tyrosine or tryptophan also tend to associate with one another in a stack like arrangement under the influence of above hydrophobic interactions. Once the nonpolar residues, either aliphatic or aromatic, come close together van der Waals dispersion forces also come into play, providing further stabilisation to the folded structure. **The hydrophobic interactions play a significant role in maintaining the folded protein structures, although no true bonds exist in these.**

We have so far described the primary structure of proteins and the various interactions which help to stabilise the folding of polypeptide chains into complex higher order protein structures. Having gained this knowledge, let us describe the folding of the proteins into secondary, tertiary and quaternary structures. While discussing these higher order protein structures you will be able to understand more clearly how formation of these structures is dependent on various noncovalent interactions and also on the nature of the peptide bond.

SAQ 4

Tick [√] mark the correct statement.

Electrostatic interaction between charges in proteins is highest

a) just below the protein surface [ ]

b) at the protein surface [ ]

c) in the protein interior [ ]

d) none of the above [ ]

5.4.4 Secondary Structure

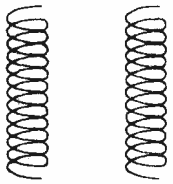
Extensive investigations by Pauling and Corey using X-ray diffraction studies led them to identify the most ordered and stable elements of protein folding called **secondary structures**. The secondary structure is basically the specific geometric arrangement of the amino acids. that results from amide linkages that are close to each other in the polypeptide chain of a protein. You may recall from subsection 5.4.2, the planarity and *trans* configuration of the peptide linkage. Subject to the structural restrictions imposed by these properties of the peptide unit and the need to maximise dispersion interactions between nonbonded atoms by close packing, Pauling and Corey built models of protein structures which would have maximum stability. These turned out to be the ones in which carbonyl oxygen and amide nitrogen atoms of the polypeptide backbone were involved in the formation of linear hydrogen bonds. Two regularly repeating structures which satisfied these criteria of maximum stability and minimum distortion were a helical structure called the **α-helix** and the **β-pleated sheet**. Although these two arrangements in the secondary structure were proposed on the basis of X-ray diffraction studies on small synthetic polypeptides and model building,

they were later on found to be present in various protein molecules. We shall now explain what an α-helix and a β-pleated sheet signify.

The α-helix

The α-helical structure results from a particular pattern of rotation around the N — C_α and C_α — C bonds of the polypeptide chain, which means the polypeptide chain can turn back under itself in a spiral or a helix. The α-helix is thus a tightly coiled structure that resembles a right handed screw or a circular staircase. This arrangement allows hydrogen bonds to form between the carbonyl group of one amino acid and the — NH — group, four amino acids ahead in the chain as shown in Fig. 5.3.

A coiled spring is helical in structure.



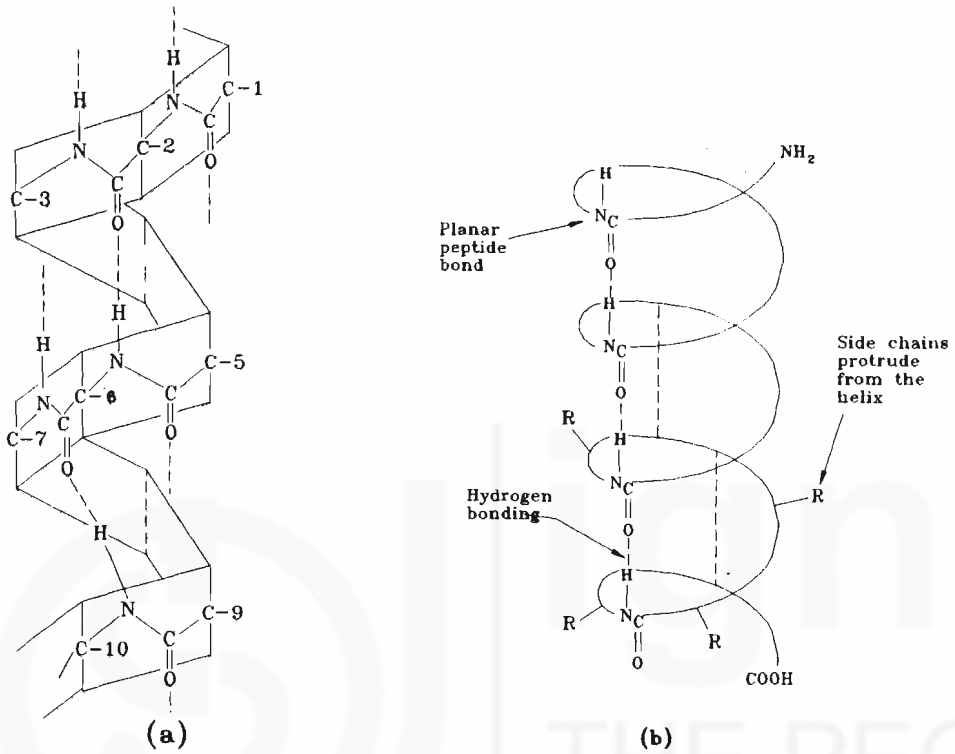
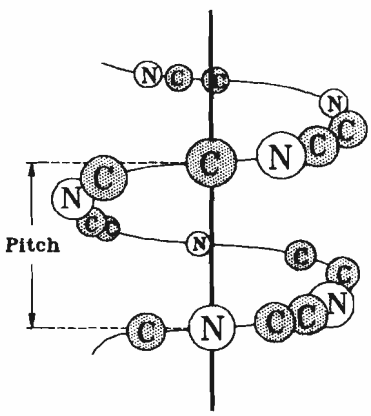


Fig. 5.3 : The right handed α-helix. a) shows the peptide groups as planes, with α-carbon atoms occupying junctions of successive planes,b) the helix, as generally represented in this unit.

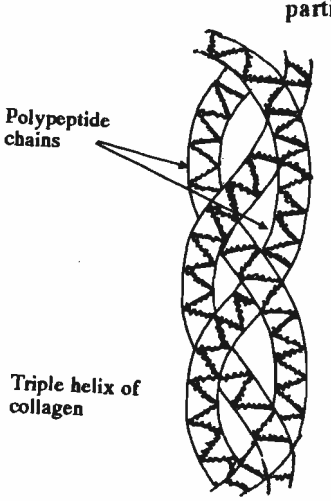
You would observe that the hydrogen bonds are parallel to the axis of the helix and are aligned linearly for maximum strength. Besides, the side chains of the amino acid residues protrude away from the axis. Each complete turn of the helix contains about 3.6 amino acids. In polypeptide chains composed of L-amino acids, the right handed helix is more stable than the left handed helix. The α-helical secondary structure is found in keratins, which are proteins that make up hair/fur, wool, claws, hooves and feathers. In keratins three α-helices are wound together like the fibres in a rope and these helices are held together by disulphide bridges. The overall number of these disulphide bridges imparts hardness and lesser flexibility to the keratins.

Collagens, which are the most abundant proteins in the body are found in skin, bones, teeth, cartilage, tendons, blood vessels and connective tissue. These proteins do not form a true α-helix but are present as a triple helix and three polypeptide chains in left handed helical conformation are twisted together in a right handed fashion. This makes these collagen fibres quite strong. You may ask as to why collagens do not form a true α-helix. This is because of a large number of proline residues in the polypeptide chain. Because of its cyclic structure



The **pitch** of the helix or the distance between two consecutive turns is 540 pm.

(Table 5.2) proline interrupts the helix and tends to bend or **kink** the polypeptide backbone. Besides this, the imino group involved in peptide link, for lack of hydrogen does not participate in hydrogen bonding with the carbonyl group.



The pitch in the helix of collagen chains is about 900 pm. Recall that in a true α-helix it is 540 pm. Lack of hydrogen bonding explains why the successive turns are not closely held.

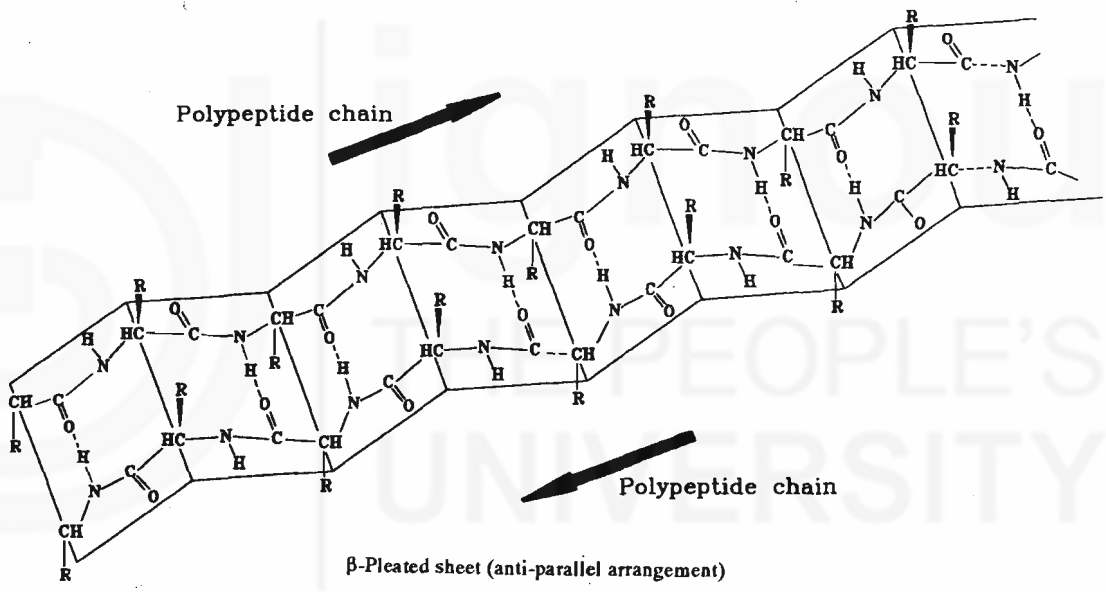
The α-helix was discovered first and the β-pleated sheet structure was the second secondary structure to be identified. That is why it is known as the β- structure.

Let us now explain the configuration of protein chains in the β-pleated sheets.

β-pleated sheets

In this arrangement the polypeptide chains are almost fully extended and hydrogen bonding occurs between close zigzagging parallel chains. The direction of hydrogen bonding is perpendicular to the direction of the polypeptide chain, unlike in α-helix where it is in the same direction. In the β-pleated structure, the side chain groups extend above and below the pleated sheet.

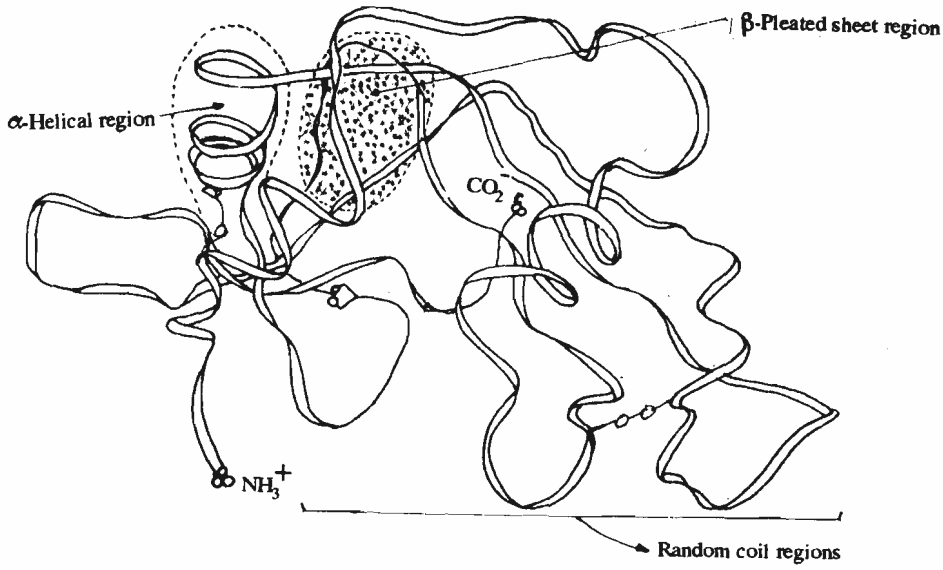
The β-sheets can have a parallel or antiparallel arrangement of the protein chains. In the **parallel β- sheets** the polypeptide chains run in the same direction from amino to carboxyl terminals whereas in the **antiparallel β-sheets** the chains run in opposite direction. The antiparallel sheet is more stable because hydrogen bonds in it are more linearly oriented. In the β-structures several polypeptide chains



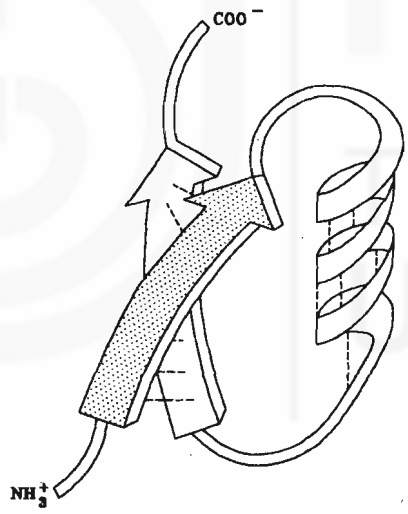
β-Pleated sheet (anti-parallel arrangement)

participate in the formation of sheets, which are in turn stacked on each other. However, these stacked sheet structures are not favoured by amino acids with bulky side chains. This is because of steric hindrance, as the side chains extend above and below the plane of the sheet. The β-sheet structures are present in silk. We may point out here that although α-helix and β-sheet structures are present in pure form in many proteins, the secondary structures of most proteins are mixtures of α-helix, β-sheets and regions of protein chain known as **random coil**. The term "random coil" does not mean that the various atoms and groups may occupy any arbitrary position. Their positions are fixed. The ϕ and ψ angles of the amino acid residues in this region are also limited by the same considerations as described in subsection 5.4.2. The term "random coil" denotes that the conformation of polypeptide chain in this region does not fit in with any of the regular and repetitive patterns described above.

In globular proteins the polypeptide chain can also fold back on itself, resulting in a change by 180° in the direction of the chain. These are known as **β-turns.**



Another pattern present frequently in proteins, such as globular proteins, is β-α-β arrangement, where an α-helical segment is flanked on both ends by segments of β-sheets:



Silk is composed of the protein **fibroin**, which has a β- pleated sheet structure.

Fig.5.4 : Schematic representation of the β-α- β folding pattern of a protein. The α-helix is represented by the coiled structure and β-sheets by the large arrows. The dashed lines are indicative of hydrogen bonds.

We may emphasise here that these secondary configurations impart structure dependent properties to the proteins. For example, fibrous proteins exemplified by α-keratin of hair or wool, silk fibroin and collagen are best examples of structure dependent function in proteins. Thus wool fibre is flexible, extensible and elastic, properties which are obviously due to the α-helical structure of wool keratin. Silk fibre on the other hand is

Proteins have been classified in several different ways. One of these classifications is based on the physical characteristic of the proteins and recognises them as globular proteins and fibrous proteins. **Globular proteins** are soluble in water i.e., form colloidal dispersions, are fragile and are involved in active functional work, such as catalysing biological reactions or transporting other substances. Examples are enzymes and haemoglobin. **Fibrous proteins** on the other hand are insoluble in water, are tough and are involved in structural or protective functions. Examples are keratins, collagen and silk.

unusually strong, but not extensible. These properties are derived from its β-sheet structures built from polypeptide chains which are already fully extended and, therefore, not extensible any further. Since the silk fibre is made up of β-sheets stacked on each other and held together by van der Waals forces, it is very flexible. Similarly, the triple helical structure of collagen gives it rigidity, mechanical strength and resistance to stretching – properties ideal for a connective tissue protein.

We have so far explained that the particular amino acid sequence constituting the polypeptide chain of a protein molecule is known as its primary structure. The particular conformation which this primary structure, i.e. the polypeptide backbone, assumes constitutes the secondary structure. This structure is influenced by the planarity of the peptide bond and is stabilised by hydrogen bonding. You will now learn that the way a protein molecule, as constructed by its primary and varying degrees of secondary structures, folds into a particular specific shape constitutes its **tertiary structure**. You will also learn that the forces which stabilise this structure are many in number. These forces have already been described in subsection 5.4.3. Let us now discuss the tertiary structure of the protein molecules.

SAQ 5

Tick [√] mark the correct answer.

β-sheets are present in pure form in

a) silk fibroin [ ]

b) α-keratin [ ]

c) collagen [ ]

d) none of the above [ ]

5.4.5 The Tertiary Structure

We have identified regularly repeating structural features such as α-helix and β-sheets as being characteristic of fibrous proteins such as keratin, silk fibroin or collagen. In most of the other large number of proteins, including enzymes, the polypeptide chain is twisted, folded and packed into a compact almost globular three dimensional shape, which is known as the **tertiary structure**. In the tertiary structure stretches of secondary structure, such as α-helices and β-sheets, along with randomly coiled regions i.e., those without any particular order of the polypeptide chain, fold back on each other in three dimensions forming compact globular shaped structures. In these globular proteins the polypeptide chains are closely packed, leaving very few cavities, which are filled with water molecules. The formation of globular tertiary structure does not necessarily imply the presence of secondary structural elements in these proteins, since folding to a compact globular shape can occur even in the absence of any significant secondary structure. The varying degree of secondary structural elements, whenever they are present in the tertiary structures, have generally the same geometrical specifications as those in the fibrous proteins. Though some slight deviations from their normal dimensions have been occasionally noticed. A feature of the tertiary structure is that segments of the protein chain far removed in sequence come very close together in its folded three dimensional structure. We have shown the three dimensional tertiary structure of two typical proteins – myoglobin and adenylate kinase in Fig. 5.5.

All enzymes are globular proteins.

We have already described hydrogen bonding and other interactions in subsection 5.4.3.

The folding of the entire protein molecule, i.e., the tertiary structure is stabilised by hydrogen bonding and hydrophobic interactions, which are principally responsible for the specificity of the folding. However, other interactions, such as van der Waals dispersion forces do make a significant contribution, because amino acid side chains are closely packed in globular proteins. A small contribution to the stability of the folded structure also comes from electrostatic forces, like the interaction between $-COO^-$ group of one

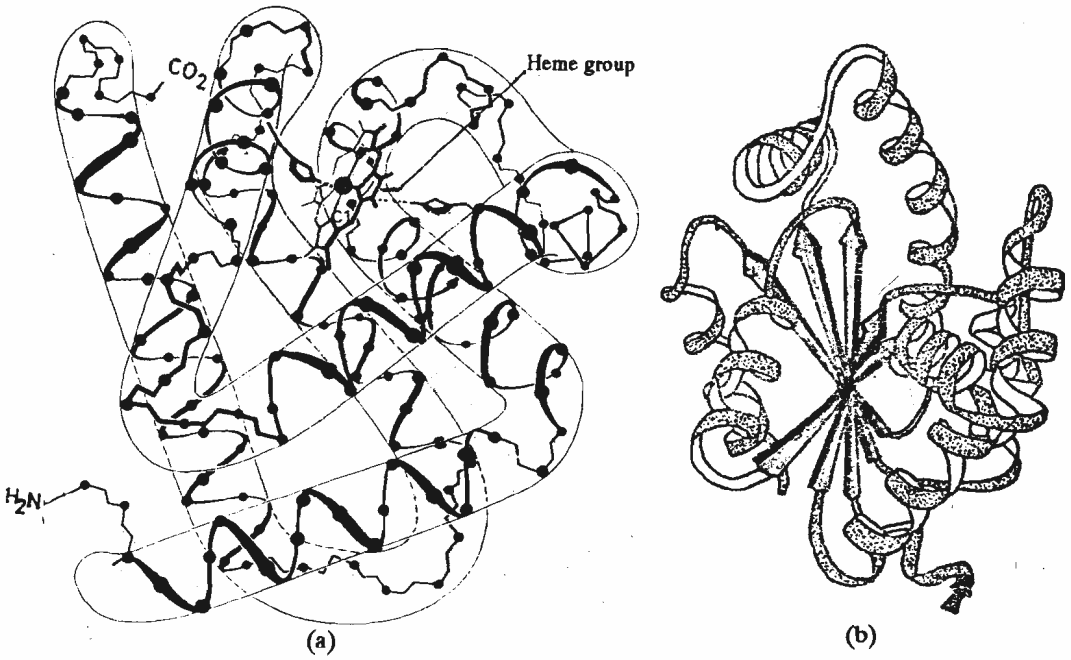


Fig.5.5 : a) Myoglobin; you will observe that the molecule is predominantly made up of helical regions. b) Adenylate kinase; the molecule comprises α-helices and β-sheets. Helices have been shown as ribbons and sheets as arrows.

amino acid residue and $-NH_3^+$ group of another. Disulphide bonds also provide further stability, as in some enzymes. In Fig. 5.6 we have given a schematic representation of a globular protein and represented the various types of forces which are involved in stabilising its tertiary structure.

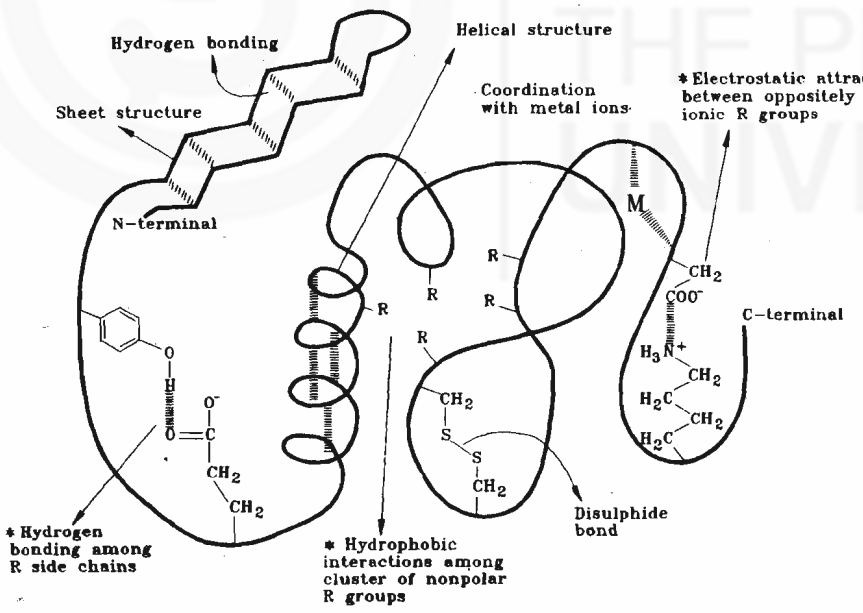


Fig.5.6 : Schematic representation of bonds and forces that stabilise globular structure in proteins. Interactions identified with an asterisk are present in all the globular proteins. Other interactions may or may not be present.

We may add here that in some proteins which require metal ions for their function, coordination with negatively charged side chain may be an important element in their

stability. Similarly a nonpolar side chain may be found occasionally on the protein surface, exposed to the solvent water. The function of such residues may lie in binding to other polypeptide chains to form quaternary structures or in binding to the hydrophobic interiors of various membranes or in binding to the nonpolar substrates of enzymes. In general, most of the polar groups in globular proteins lie on the surface, and most of the hydrophobic side chains lie inside the molecule.

The folded state i.e., the tertiary structure of a protein that it assumes under normal conditions of temperature and pH, has minimum energy and hence most stable. This state is known as **native configuration**. Proteins in native state can be easily unfolded or **denatured** by extremes of pH, heat or organic solvents. Denaturation results in uncoiling of the proteins into a random state with loss of biological activity. These denaturing agents have a disruptive effect on the hydrogen bonds, salt bridges and other interactions. The energy of stabilisation of native state is thus marginally higher in comparison to that of the unfolded state. This is because the conversion of a random, highly disordered unfolded polypeptide chain into a highly ordered folded state of the native protein is disfavoured on entropic considerations.

You can now attempt the following SAQ and then proceed further.

SAQ 6

Tick [√] mark the appropriate statement.

Globular proteins always contain

a) some α-helices [ ]

b) some β-sheets [ ]

c) some random coil regions [ ]

d) α-helix and β-sheets [ ]

Many globular proteins are organised into structural **domains** or lobes. These are themselves globular in nature and are linked by a strand of the polypeptide chain. Each domain or lobe has a specialised function. For example, in the enzyme glyceraldehyde 3-phosphate dehydrogenase each of the four chains that constitute it are made up of two distinct domains. One of these binds NAD which is a cofactor for the enzyme. The other lobe functions as the catalytic domain and binds the substrate glyceraldehyde 3-phosphate. Their structures are shown in Fig. 5.7.

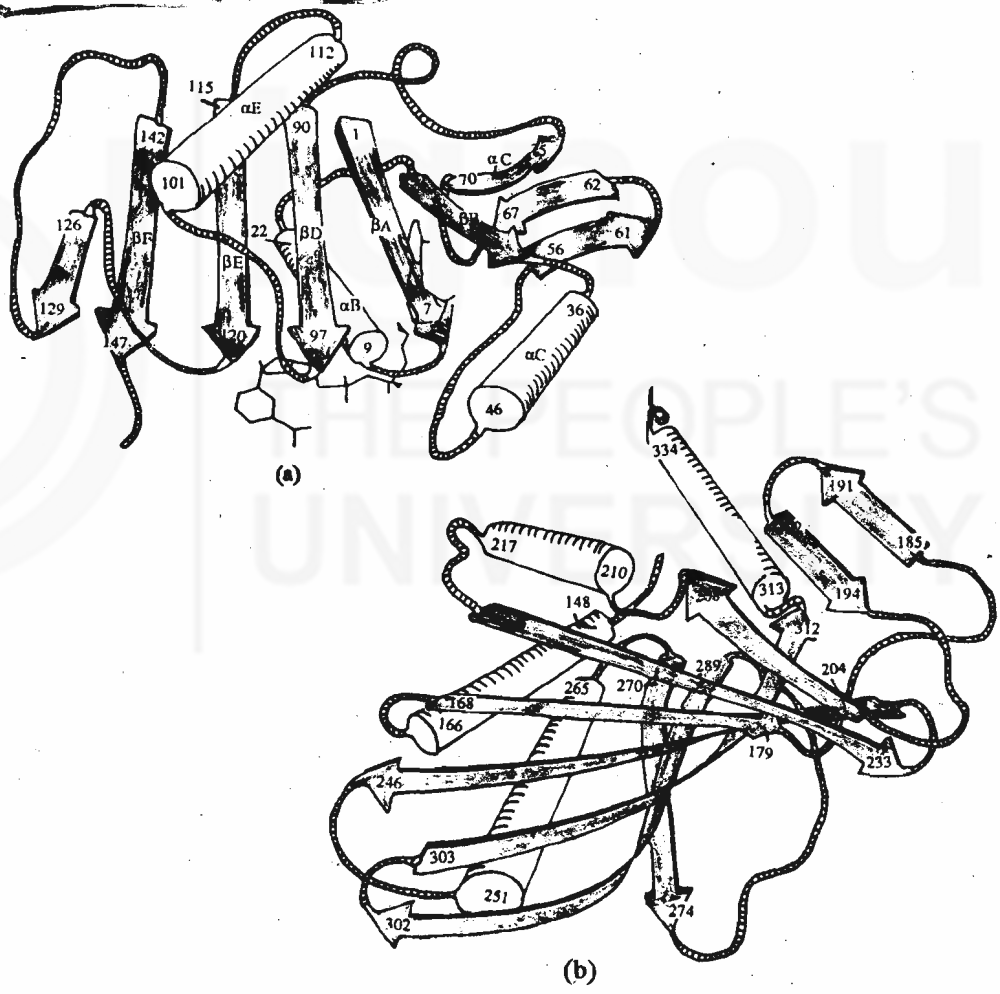


Fig. 5.7 : Domain structures in a subunit of the enzyme glyceraldehyde 3-phosphate dehydrogenase a) The NAD binding domain with bound NAD b) The catalytic domain

Quite often polypeptide chains of the same type or different types, each of them generally folded into a compact unit i.e., a tertiary structure, coalesce together to form a structural aggregate. This level of protein organisation is called as its **quaternary structure**. Let us now describe this highest level of protein structure.

5.4.6 Quaternary Structure

As we mentioned above, this level of protein organisation involves the association of two or more individual protein units, each with its own tertiary structure, into a complex and functional unit. This association of various protein subunits is known as the quaternary structure. With the exception of disulphide bond, the forces which hold these subunits together are the same as those present in the tertiary structure.

The relationship between tertiary and quaternary structures can be best illustrated by myoglobin and haemoglobin. Myoglobin functions as an oxygen storage protein in the muscles whereas haemoglobin transports oxygen throughout the body. Myoglobin binds oxygen when O_2 content in the cells is high and releases it when its level in the cell falls. You will recall that myoglobin has tertiary level of protein organisation and we have already represented its structure in Fig. 5.5. It consists of a single polypeptide chain of 153 amino acid residues organised into eight α-helical portions, which are connected by seven non-helical strands. These nonhelical strands help the α-helical portions to fold back on one another to form a compact globular structure. A nonprotein molecule called **haem**, with an iron atom at its centre, is the oxygen binding site in the myoglobin molecule. It is located in a nonpolar cavity of myoglobin molecule and bonded to a histidine molecule in the chain.

Cooking besides denaturing proteins in our food also denatures protein toxins and bacterial proteins, thus preventing food poisoning.

The oxygen binding ability of haemoglobin is also due to the haem group. Haemoglobin is present in red blood cells where it binds oxygen and then transports it to tissues throughout the body. This molecule has four polypeptide chains, each carrying a haem group. Two of them are called α-chains, with 141 residues each, and the other two are called β-chains with 146 residues each. The four polypeptide chains or subunits are held together by noncovalent interactions principally of hydrophobic type, and by van der Waals dispersion forces. About one third of the interchain contacts consist of hydrogen bonds and electrostatic interactions, which provide specificity for the folding process. We have represented the haemoglobin molecule in Fig. 5.8.

Fibrous proteins as well as globular proteins may be composed of only one polypeptide chain or multiple chains. When the proteins consist of more than one chain, they are called **oligomeric** proteins and the individual chains are known as **protomers** or **subunits**.

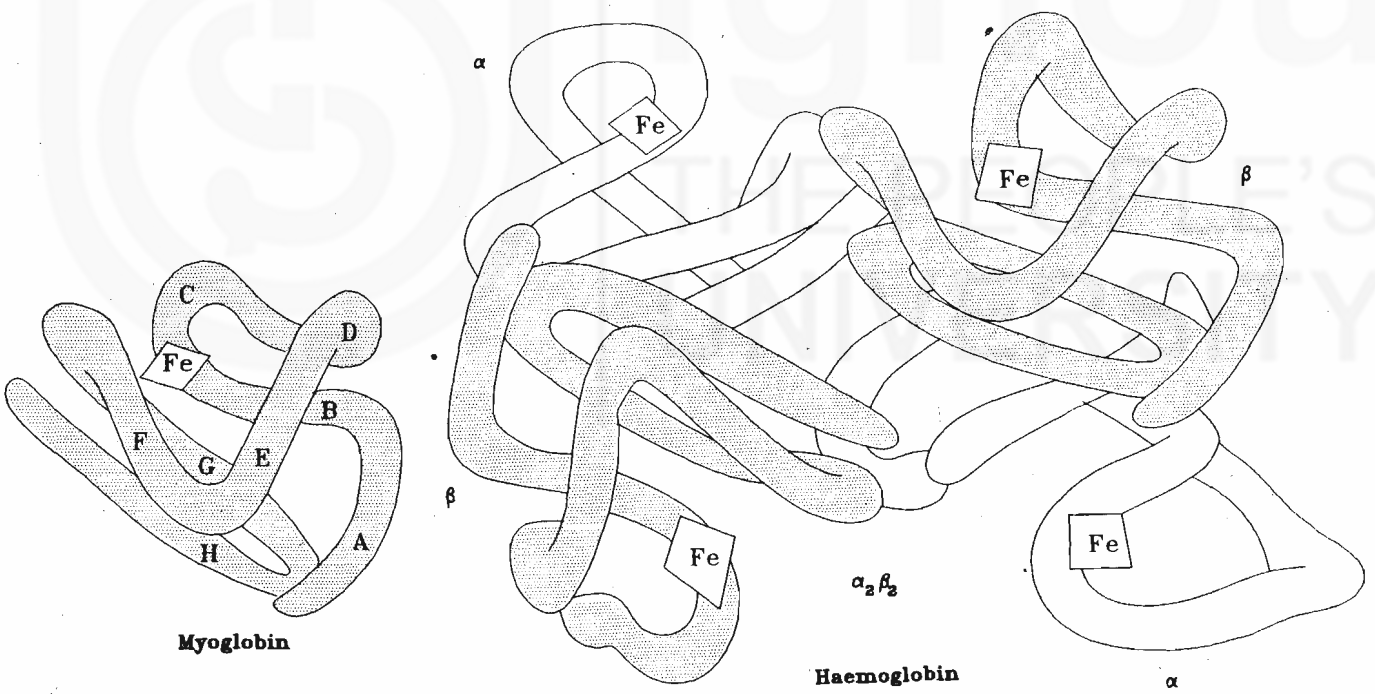


Fig.5.8 : Comparison of the structures of myoglobin and haemoglobin. Observe the positions of the haem (Fe) groups and the close similarity in the folding pattern of the myoglobin chain and that of the β chain of haemoglobin.

You would have observed that the α and β chains of haemoglobin resemble myoglobin a great deal and the four subunits, resembling myoglobin, associate to form the haemoglobin molecule. The evolution of myoglobin into haemoglobin appears to have been accompanied

α and β chains of haemoglobin have similar but distinct amino acid sequence. The β-chain or subunit contains 8 α-helices whereas the α-chain or subunit contains only 7 α-helices.

by replacement of the polar surface residues of myoglobin by nonpolar residues in the related α and β chains of haemoglobin. Since the exposure of nonpolar residues to water would have lead to instability, α and β chains are associated with each other in such a way as to drive the nonpolar residues into the interior and to form the quaternary structure of haemoglobin, which now acquires a new function, not present in the myoglobin molecule. We shall discuss more about this function in the next subsection.

The presence of a quaternary or oligomeric structure in a protein confers several advantages on it. For example, regulation of a protein function or acquisition of a new function is much better achieved when several subunits associate together to form the quaternary structure. Also the presence of dissimilar subunits in a quaternary structure may permit variation in the specificity of a function, such as catalysis. An important reason for the presence of quaternary structures may be the survival of the species. There is a finite possibility of the biosynthetic machinery of the cell committing a mistake, leading to the formation of a faulty protein molecule, with a wrong amino acid in its sequence. In a single unit protein such a mistake in the sequence may lead to a total loss of function with disastrous consequences for the species. However, in a multi subunit protein the malfunction caused by one faulty subunit would be nonsignificant, if other subunits are normal.

Thus quaternary structures are significant not only due to their complex structure related functions, but also in correcting the mistakes caused during biosynthetic process. Let us now briefly discuss the role of haemoglobin in the body.

5.4.7 Haemoglobin as a Transport Protein

You will recall from subsection 5.4.6, that haemoglobin binds and transports oxygen in the whole body. This transport function of haemoglobin molecule is strongly dependent on its quaternary structure. You will find a comparison between the oxygen binding ability of a single unit myoglobin molecule and multi subunit haemoglobin in Fig. 5.9.

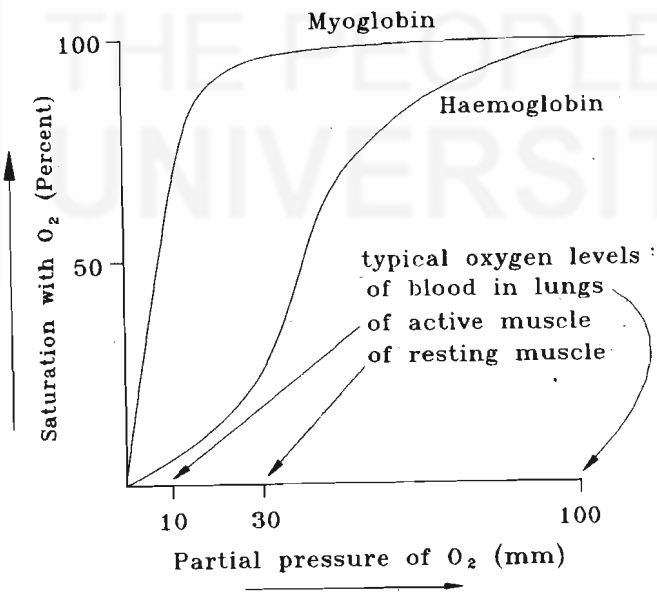


Fig.5.9 : Diagram showing oxygen binding curves of myoglobin and haemoglobin

You would observe that the oxygen binding curve of myoglobin is simple, befitting its storage function. It is easily saturated with oxygen and releases it only when the concentration of oxygen falls very low. The oxygen binding curve for haemoglobin on the other hand is sigmoidal, indicating that its binding ability increases as more and more

oxygen molecules bind to it. Further, haemoglobin has low affinity for oxygen and gets saturated with it only in the lungs where oxygen pressure is high. But in the tissues where oxygen pressure is low, it would easily release its bound oxygen, thus fulfilling its role as an oxygen transporter. From the curve in Fig. 5.9 it would be clear that under the same conditions, myoglobin would be still bound with oxygen. The behaviour of haemoglobin molecule can be explained on the basis that successive binding of O_2 molecules to the haem groups triggers changes in the inter subunit alignments, increasing the oxygen affinity of remaining haem groups considerably. You will recall that there are four haem groups in a haemoglobin molecule. It has been estimated that binding of oxygen to the fourth haem group is 400 times stronger than binding to the first haem group.

SAQ 7

Tick [√] mark the correct statement.

Haemoglobin is evolutionarily related to

a) haem [ ]

b) adenylate kinase [ ]

c) myoglobin [ ]

d) none of the above [ ]

5.5 SUMMARY

- Proteins are large complex molecules with molecular weights ranging from thousands to a few millions. They perform a multitude of functions essential to life.
- Basic structural units of proteins are the twenty α-amino acids, all with L-configuration.
- All these amino acids, with the exception of proline, have a common structural feature. They have the amino and carboxyl groups attached to the same carbon atom and are, therefore, called as α-amino acids. The α-carbon atom also carries a hydrogen atom and a fourth group known as the R-group or the side chain. The side chain is characteristic of each amino acid.
- All amino acids, except glycine (in which R is a H atom), display optical activity.
- A protein is constructed as a linear polymer of different amino acids which are linked together by peptide bonds. These chains are known as polypeptide chains.
- Since each protein chain consists of hundreds of amino acids and each position in such a chain is occupied by any of the twenty α-amino acids with different physico-chemical properties, the total number of structurally different proteins that can exist is large, explaining in part the enormous variety of functions that proteins are capable of.
- The ionisable side chains of amino acid residues in proteins contribute to their acid-base properties and also to their buffering capacity.
- The complexity of the protein structure is best illustrated by resolving it into hierarchical levels of organisation, known as primary, secondary, tertiary and quaternary structures, each of which contributes to the formation of the next higher level.
- The number and sequence of amino acid residues, linked linearly by peptide bonds, is called the primary structure of proteins.

- Protein chains do not normally exist as random coils but assume precise three dimensional folding patterns which are dictated by amino acid sequence and influenced by various noncovalent interactions.
- Constrained by the planarity of the peptide linkage and hydrogen bonds between $>C=O$ and $-NH-$ groups of the polypeptide chain, contiguous regions of a protein organise themselves into repetitive ordered structural elements, such as α-helices or β-sheets, forming secondary structures in proteins.
- In many proteins polypeptide segments, with secondary structural elements and connected by random coils of polypeptide chain, fold back on each other to assume a compact globular shape called the tertiary level of protein structure. In this structure distant segments of the polypeptide chain come close to each other.
- The quaternary level of protein structure is illustrated by haemoglobin, an oxygen transport protein, in which four polypeptide chains of two different types associate together into an aggregate or oligomeric structure.
- The transport function of haemoglobin is strongly dependent on its quaternary structure.
- The folding process involved in the formation of tertiary and quaternary structures are triggered principally by hydrophobic forces generated by the tendency of some amino acid side chains to avoid water and seek the interior of the protein molecule. The folded state is stabilised by van der Waals dispersion interactions and electrostatic forces. The specificity of the folding process is achieved by hydrogen bonds.

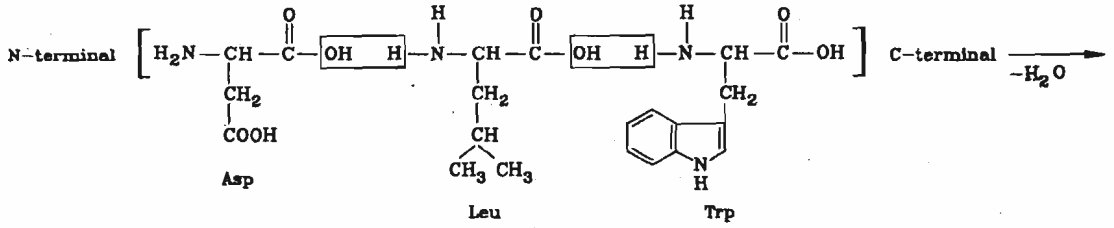
5.6 TERMINAL QUESTIONS

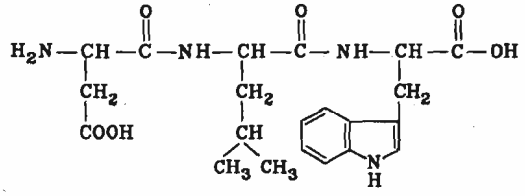
1. Describe some important functions of proteins.
2. List ten amino acids, the side chains of which can participate in hydrogen bonding.
3. What type of amino acids are responsible for the acid base behaviour of proteins?
4. Explain why the peptide bond is planar?
5. Describe the origin of hydrophobic forces.
6. Describe briefly two repetitive arrangements which constitute the secondary structures in proteins.
7. Explain the advantages of quaternary structure in proteins.

5.7 ANSWERS

Self Assessment Questions

1. d
2. c
3. You will recall that abbreviated forms for peptides are written from N-terminal to the C-terminal. Write the structures for these amino acids in their unionised forms and remove water molecules from the carboxyl and amino groups of adjacent amino acids, joining these molecules by peptide bonds:



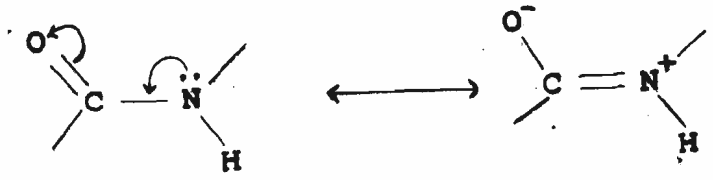


tripeptide Asp. Leu. Trp.

4. c
5. a
6. c
7. c

Terminal Questions

1. Protein molecules as enzymes, catalyse a large number of biochemical reactions. Besides this they form structural elements in cells, are involved in various motile, transport and storage functions in the organism, and also act as defence mechanism against infections. Proteins also regulate gene expression. Some proteins function as hormones.
2. The side chains of serine, threonine, asparagine, glutamine, aspartic acid, glutamic acid, cysteine, histidine, lysine and arginine can participate in hydrogen bonding as their side chains contain hydrogens attached to O, N or S atoms or have a >C=O group in their side chains.
3. Amino acids with charged side chains, such as glutamic acid, aspartic acid, lysine, arginine, histidine, and amino acids at the N and C-terminals with free α-amino and α- carboxyl groups respectively, are responsible for the acid-base behaviour of proteins.
4. Peptide group resonates between its two canonical structures as shown below:



Thus both > C= O group and the C —N bond have a double bond character. Since rotation about the axis of a double bond is not possible, all the atoms of the peptide linkage are forced to be coplanar.

5. Hydrophobic forces arise because of the inability of nonpolar side chains of amino acids to interact with water via hydrogen bonds. Water is thus forced to increase its own hydrogen bonded network, when exposed to these side chains. This results in a highly ordered structure in water, leading to decreased entropy and, therefore, thermodynamic instability. Therefore, proteins fold spontaneously to transfer these nonpolar side chains to their interior, so that water molecules are liberated from the hydrogen bonded network, entropy is increased and thermodynamic stability thus restored.

6. α-Helix and β-pleated sheets are two repetitive arrangements present in secondary structure of proteins. In the α-helix, the polypeptide chain turns back under itself in a spiral which resembles a screw or a coiled spring. The side chains protrude away from the axis of the helix. In the β-pleated sheet arrangement the polypeptide chains run parallel or antiparallel to each other and are almost fully stretched. The side chain in the β-sheets extend above and below the plane of the chains or sheets. The α-helix is stabilised by hydrogen bonds which are parallel to the axis of the helix, whereas in β-sheets the direction of hydrogen bonding is perpendicular to the direction of the polypeptide chains.

7. Quaternary structure permits the acquisition of new regulatory or catalytic functions not present in the original subunits forming the quaternary structure. It also leads to increased protein stability and has the ability to correct a malfunction caused by the presence of a faulty subunit.