Indira Gandhi National
Open University
School of Sciences

BCHET-149
MOLECULES OF LIFE

Block

2

AMINO ACIDS, PEPTIDES AND PROTEINS

Course Design Committee

Prof. K.K. Arora, Dept. of Chemistry, Zakir Husain College, University of Delhi, Delhi

Dr. Sunita Joshi (Retd.), Dept. of Biochemistry, Daulat Ram College, University of Delhi, Delhi

Dr. Bhupinder Mehta, Dept. of Chemistry, Swami Shraddhanand College, University of Delhi, Delhi

School of Sciences,
IGNOU, New Delhi 110068

Prof. M.S. Nathawat
Prof. Sunita Malhotra
Prof. B.I. Fozdar
Prof. Javed A. Farooqi
Prof. Sanjiv Kumar
Prof. Lalita S. Kumar
Prof. Kamalika Banerjee

Block Preparation Team

Prof. Lalita S. Kumar
School of Sciences, IGNOU

Dr. Swati Omanwar, Ex. Consultant
School of Sciences, IGNOU

Prof. Seemi Bashir Ahmed (Editor)
Department of Chemistry,
Jamia Millia Islamia, New Delhi

Course Coordinator: Prof. Lalita S Kumar

Print Production

Acknowledgements: Sh. Sarabjeet Singh for CRC preparation
Sh. Deepak Kumar and Sh. Mukesh Kumar for word processing
Sh. Ajit Kumar and Sh. Mukesh Kumar for the Art Work
Material partially adapted from the Biochemistry Course (CHE-09).

November, 2021

ISBN:

Further information on Indira Gandhi National Open University courses may be obtained from the University's office at Maidan Garhi, New Delhi-110068 or IGNOU website www.ignou.ac.in

Printed and published on behalf of Indira Gandhi National Open University, New Delhi by the Registrar, MPDD, IGNOU.

Printed at:

BLOCK 2: AMINO ACIDS, PEPTIDES AND PROTEINS

In Block 1 you studied the structure of cell and also the functions of various organelles present in it. You also studied the energy providing biomolecules, namely, carbohydrates, where in the structure, properties and functions of mono-, di- and polysaccharides are dealt in detail. In Block 2 you will study about another very important biomolecules viz., proteins. These molecules are central to all biological processes.

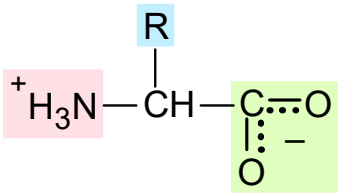
Block 2 comprises three units. In Unit-4, **'Amino Acids'** the amino acids have been explained as the monomeric units of the proteins. The details of the classification, nomenclature, structure and functions of amino acids are given. Unit 5 on **'Peptides'** deals with the molecules that are formed by the combination of amino acids leading to mono-, di- and polypeptides. The nomenclature, synthesis and biological importance of these polymeric molecules have been detailed in the unit. The last unit, Unit 6 on **'Proteins'** discusses the primary structure of these molecules in the form of long chain peptides folding to form the secondary, tertiary and quarternary structure. The variety of functions of proteins makes these very important biomolecules.

Expected Learning Outcomes

After studying this block, you should be able to;

- describe the structure, nomenclature, classification, properties and separation of α-amino acids,
- describe the formation, nomenclature, synthesis and sequence determination of peptides,
- describe the classification, properties and structural organisation of proteins, and
- explain the biological significance of amino acids, peptides and proteins.

ignou
THE PEOPLE'S
UNIVERSITY



UNIT 4

AMINO ACIDS

Structure

4.1 INTRODUCTION

In the units of the first block of this course you were introduced to the important features of a typical cell present in all living organisms. You also studied about carbohydrates, the important energy providing biomolecules and their significance in the functioning of the living system. In the second block you will be introduced to another important and complex class of biomolecules, known as proteins. These molecules are central to all biological process and constitute half the dry mass of a cell. These are made up of α-amino acids as the monomers and are dealt in the present unit of this block.

We would give you a broad view of how amino acid molecules join into polypeptide chains, which in turn form the primary, secondary, tertiary and the quaternary structures of proteins. Thus, you will study the structure, classification and properties of the amino acids. You would also study how these properties help us in the separation of amino acids by paper chromatography and paper electrophoresis. The next unit deals with peptides formed by the combination of amino acids.

Expected Learning Outcomes

After studying this unit, you should be able to:

- describe the structure of amino acids as the monomeric units of proteins,

- ❖ classify amino acids on the basis of their structure,
- ❖ draw and explain the zwitterionic structure of amino acids,
- ❖ define isoelectric point and explain the titration curve for amino acids, and
- ❖ describe the process of separation of amino acids by paper chromatography and paper electrophoresis.

4.2 AMINO ACIDS: THE BUILDING BLOCKS OF PROTEINS

Amino acids are the main components of protein molecules and free amino acids are rarely distributed in nature.

The dietary proteins are hydrolysed during digestion to release amino acids which are then reassembled into proteins.

Proteins are significant biomolecules that are essential for performing a number of functions in the cells. Similar to polysaccharides, proteins are also polymeric in nature. The monomeric units of proteins are amino acids. When proteins are completely hydrolysed, the twenty L- α-amino acids (L notation for amino acids will be explained later) are produced. Let us try to recollect the basic structure of amino acids and learn to represent these in the form of symbols.

4.2.1 Structure and Representation of Amino Acids

The amino acids have a common structural motif. You would note that amino acids have an amino group at one end and a carboxyl group at the other end, both of which are attached to the **α-carbon** atom i.e. the carbon atom next to the carboxyl group. Those amino acids which have their amino group attached to this α-carbon are known as the **α-amino acids**. The α-carbon also carries a hydrogen atom. The fourth group on this carbon is known as the **side chain** and is denoted as **R**. The simplest of α-amino acids i.e., glycine has a hydrogen atom as the side chain. In other α-amino acids, this side chain consists of long carbon chain and more complex chemical groups. All the α-amino acids differ from each other in the nature of their side chains only. Thus, the general structure of an α-amino acid can be represented as given in Fig. 4.1 (a). The structure of glycine as an example is given in Fig. 4.1 (b).

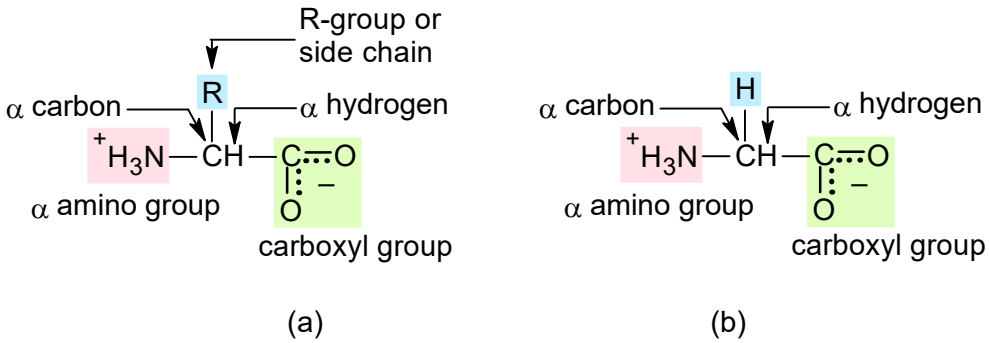
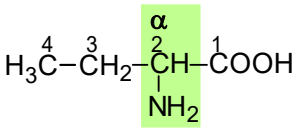


Fig. 4.1: a) General structure of an amino acid; b) Structure of glycine.

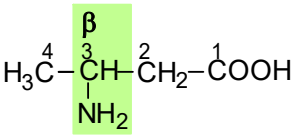
The biologically significant amino acids are principally α-amino acids. However, depending upon the position of amino group on the carbon chain attached to the carboxyl carbon, these are called α, β **and** γ **amino acids**. The amino acids which have amino group attached to the carbon atom adjacent to the carboxyl group are α- amino acids or **2-amino acids**. In amino

acids with longer alkyl chains, the amino group can be placed in different positions with respect to carboxyl groups giving rise to β, γ etc. amino acids. For example, amino group in amino butanoic acid can be placed in three positions viz., α, β and γ which gives rise to three **position isomers** of this amino acid. The structures of the three isomers are given below.

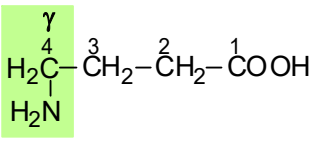
Position isomers: are the isomers that differ in the position of a functional group/multiple bond/substituent in the same carbon chain.



α- Aminobutanoic acid
2- Aminobutanoic acid



β- Aminobutanoic acid
3- Aminobutanoic acid



γ- Aminobutanoic acid
4- Aminobutanoic acid

You can observe from the general structure of amino acids that these differ from one another with respect to the R groups i.e., the side chains bonded to the α-carbon atom. Most living organisms contain twenty α-amino acids which constitute the vast variety of proteins. These amino acids are listed along with their names, structures and representation in symbol forms in Table 4.1. You may note the different groups under which the amino acids are categorised. You will be able to appreciate this when you study about the classification of amino acids in subsection 4.2.3. You would also observe that the symbol representation is using three and one letter. Let us now understand the representation of amino acids to have a better clarity.

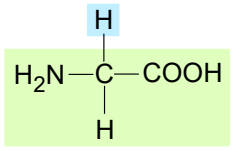
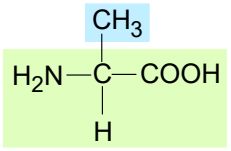
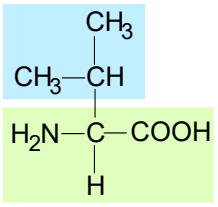
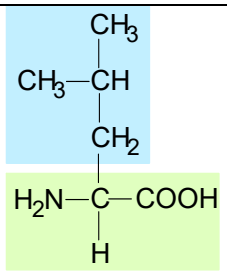
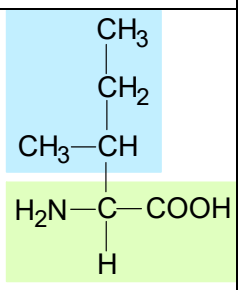
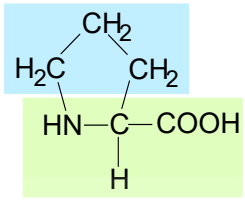
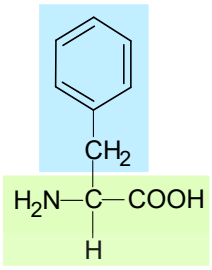
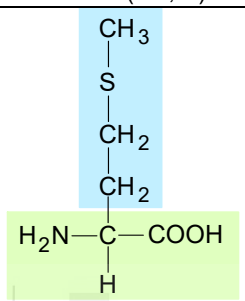
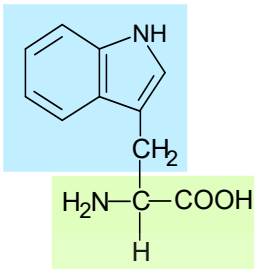
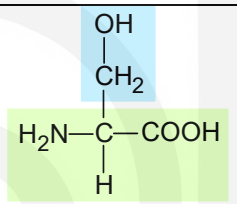
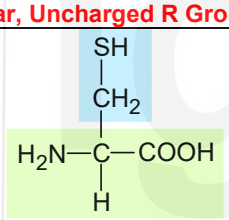
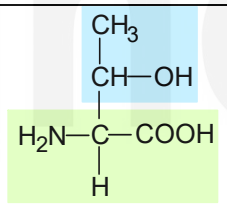
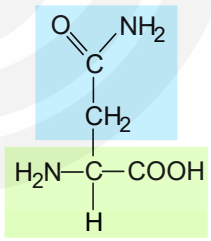
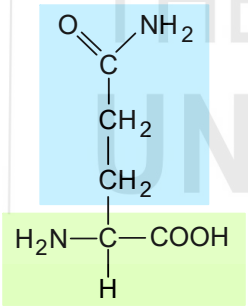
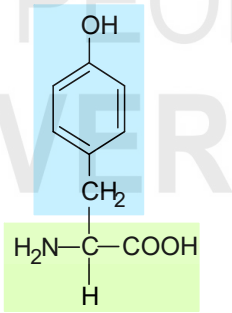
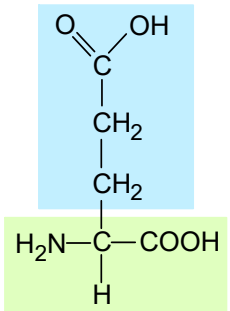
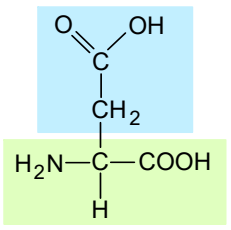
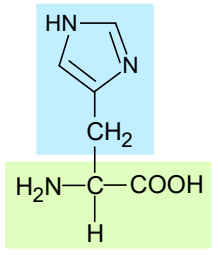
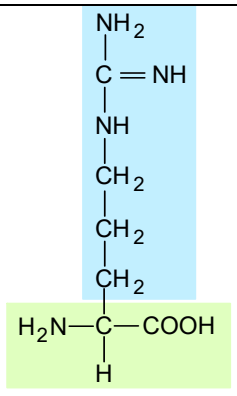
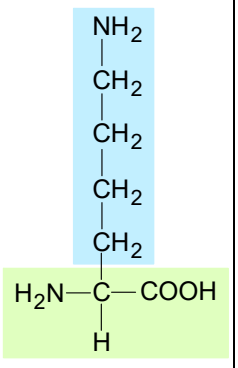
Representation of amino acids

The amino acid is commonly represented by a **three letter symbol** or **one letter symbol**. The three letter symbol is derived from the trivial name of the amino acid and includes the first three letters of this name. The amino acids asparagine, glutamine, isoleucine and tryptophan are the exceptions in this representation. Thus, the three letter symbols are written as one capital letter followed by two lowercase letters. For example, **glycine** is represented as **Gly** and not as GLY or gly. This applies even when we have to write the symbol in the middle of a sentence.

In addition to the above, **one letter symbol** is also in use. The one letter symbol for eleven of the twenty coded amino acids is the first letter of their full name and for the rest nine distinct letters have been assigned so as to avoid any confusion. It is important to know that the symbols or we may say abbreviated forms are commonly used in representing long sequences of peptides or proteins where it would be difficult to write the full sequence of constituent amino acids with their full names.

If we observe the arrangement of atoms or groups around the α-carbon in amino acids as observed in the case of carbohydrates, we would be able to appreciate the stereochemistry of amino acids. You would recall from courses on stereochemistry that the spatial arrangement of atoms or groups in a molecule is called the **configuration** of the molecule. Let us describe the configuration at the α-carbon atom in amino acids after you have attempted SAQ 1.

Table 4.1: Structures of amino acids with three and one letter names

Nonpolar R Group				
				
Glycine (Gly, G)	Alanine (Ala, A)	*Valine (Val, V)	*Leucine (Leu, L)	*Isoleucine (Ile, I)
				
Proline (Pro, P)	*Phenylalanine (Phe, F)	*Methionine (Met, M)	*Tryptophan (Trp, W)	
Polar, Uncharged R Group				
				
	Serine (Ser, S)	Cysteine (Cys, C)	*Threonine (Thr, T)	
				
	Asparagine (Asn, N)	Glutamine (Gln, Q)	Tyrosine (Tyr, Y)	
Polar (Negative Charge) Group		**Polar (Positive Charge) R Group**		
				
Aspartic acid (Asp, D)	Glutamic acid (Glu, E)	*Histidine (his, H)	*Arginine (Arg, R)	*Lysine (Lys, K)
* These are essential amino acids				

SAQ 1

Fill in the blanks with appropriate words in the following sentences:

a) The monomeric units of proteins are _____.

b) The amino acids which have amino group attached to the carbon atom adjacent to the carboxyl group are called_____.

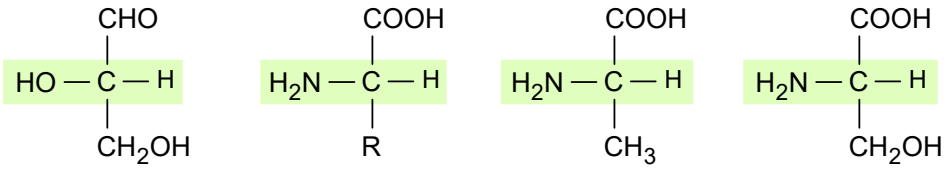
c) Gln is the three letter abbreviation for ---------- .

4.2.2 Configuration of Amino Acids

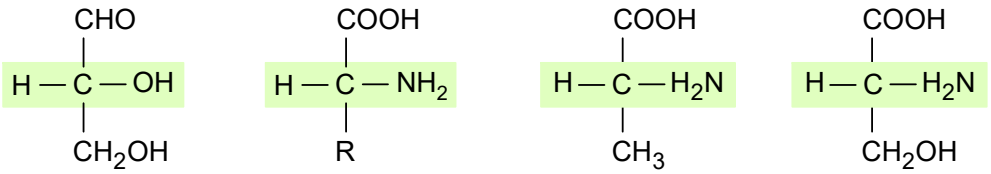
You would have observed that all the amino acids discussed above, with the exception of gycine, have a **chiral** or an asymmetric carbon atom, i.e., they have a carbon atom bonded to four different atoms or groups tetrahedrally. In case of amino acids under discussion, it is the α-carbon which is a chiral centre. Therefore, except glycine, all the amino acids containing chiral carbon exhibit optical activity and are designated as **(+)** and **(-)** isomers respectively. As in the case of carbohydrates, the configuration at the chiral carbon is designated by the letters **D** and **L**. You may recall from Unit 2 that these notations are used after a comparison with the configuration of glyceraldehyde. The D and L glyceraldehydes are the mirror images as shown below.

L-Glyceraldehyde, (-) Rotation **D- Glyceraldehyde, (+) Rotation**

In the case of α-amino acids, the configuration at the α-carbon is compared to the configuration of D and L glyceraldehydes, as shown below:



L-Glyceraldehyde L-Amino acid L-Alanine L-Serine



D-Glyceradehyde D-Amino acid D-Alanine D-Serine

As depicted above, in L-amino acids the $–NH_2$ group is on the left and in the D form the $–NH_2$ group in 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 the left. Although D-amino acids do occur in nature, they are never present in proteins (in comparison to this fact you have read that generally carbohydrates occur in D-configuration).

The amino acids are grouped into different types according to the nature of their side chains. Before studying the way amino acids are classified try to answer the following question.

SAQ 2

Complete the following sentence with appropriate answer.

a) Glycine is not optically active because ……

b) In L amino acids the amino group is on the …….

c) All the amino acids present in proteins have …….configuration.

d) The configuration of amino acids is compared to the configuration of..... to designate D or L.

4.2.3 Classification of Amino Acids

It was mentioned earlier and from Table 4.1, you can observe that the side chains present in amino acids vary in their nature and polarity. These two factors form the basis of the classification of amino acids. We can understand the classification by either looking at the nature of R group or considering the polarity of R group. Let us see how.

(A) **Classification of amino acids based on the nature of R groups**

- **Aliphatic amino acids:** The aliphatic amino acids have aliphatic chain as R group. Five amino acids; glycine, alanine, valine, leucine and isoleucine are included in this group.
- **Hydroxy amino acids**: The hydroxyl amino acids have hydroxyl group in their R group and include serine, tyrosine and threonine.
- **Sulphur containing amino acids:** These contain sulphur in their R-group. This class includes methionine and cysteine.
- **Acidic amino acids:** These are amino acids containing an additional -COOH group in the side chain e.g., aspartic acid, glutamic acid. Thus these are dicarboxylic acids.
- **Basic amino acids:** The basic amino acids contain additional basic $-NH_2$ groups or a heterocyclic ring. Arginine, lysine and histidine fall into this group.
- **Aromatic amino acids:** In this class aromatic groups are present as the side chain group. The examples are phenylalanine, tyrosine and tryptophan.

- **Imino amino acid:** This amino acid is different from others in that there is an imino group in place of amino group. The only example in this category is proline.

(B) Classification of amino acids based on the polarity of R groups

You may again have a look at Table 4.1. You would observe the amino acids placed into different classes on the basis of polarity. These are:

- **Nonpolar R group**: The nonpolar group has generally a nonpolar aliphatic or a non-polar aromatic side chain. Examples are Glycine, Alanine, Valine, Leucine, Isoleucine, Phenylalanine, Tryptophan, Methionine, Proline
- **Polar uncharged R group:** The uncharged polar groups have side chains containing hydroxyl, amide or sulphydryl groups. Examples are Serine, Threonine, Cysteine, Asparagine, Tyrosine, Glutamine
- **Polar (Negatively charged) group**: The charged polar group can be acidic when the side chain contains a carboxyl group. Examples are Aspartic acid and Glutamic acid.
- **Polar (positively charged) R group**: The charged polar group can be basic when the side chain contains a basic group. Examples are Lysine, Arginine, Histidine

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 interact with polar molecules like water, which is the normal medium in which most of the proteins function. Nonpolar side chains, on the other hand, avoid or repel 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.

(C) **Essential and Nonessential amino acids**

There is yet other way amino acids are classified and that is on the basis of their source in the living system. Our body can synthesise only ten of the amino acids listed in Table 4.1 as is indicated by an asterisk. The rest are not synthesized in humans but are synthesized in either plants or bacteria and must be supplied by our diet. Such amino acids which cannot be synthesised by humans are known as **essential amino acids.** The essential amino acids are leucine, isoleucine, lysine, methionine, phenylalanine, tryptophan, threonine, valine and histidine. Proteins that contain all the ten essential amino acids are called **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.

You should remember that essential or nonessential classes of amino acids do not relate to their significance. In fact all the twenty amino acids are necessary for normal functioning of the body. The amino acids, arginine and histidine belong to the list of **semi-essential amino acids** because the body does not always require dietary sources for it.

The variation in the nature of side chain as explained above is the cause of variation in the physical properties of the amino acids. We will now study their physical properties in the following section. Before proceeding, try to answer the following SAQ.

SAQ 3

Tick [√] mark the appropriate answer.

(a) Amino acids owe their variation to the

i) carboxyl group [ ]

ii) chiral carbon [ ]

iii) α-amino group [ ]

iv) nature of their side chains [ ]

(b) Sulphur containing amino acids are:

i) methionine, serine [ ]

ii) threonine, serine [ ]

iii) methionine, cysteine [ ]

iv) threonine, cysteine [ ]

4.3 PHYSICAL 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 structure of an amino acid in a polar medium.

4.3.1 Zwitterionic Structure

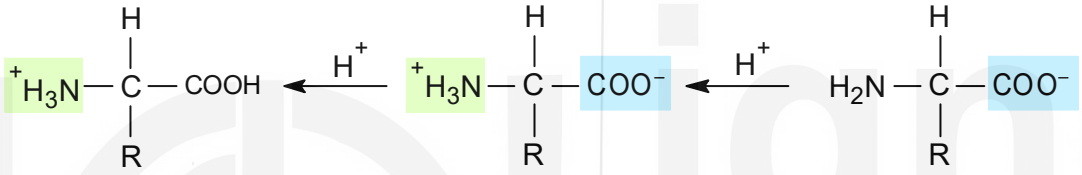
An amino acid is usually represented in its unionised form to show the presence of both amino and the carboxyl groups as you have seen in the previous section. However, these are represented in the ionised form which predominates under physiological conditions of pH around 7. The understanding of the ionised structure of amino acid will help us in understanding some of the properties of proteins.

Thus, an amino acid like glycine can be represented in its unionised and ionised forms as given below.

Unionised amino acid Dipolar amino acid

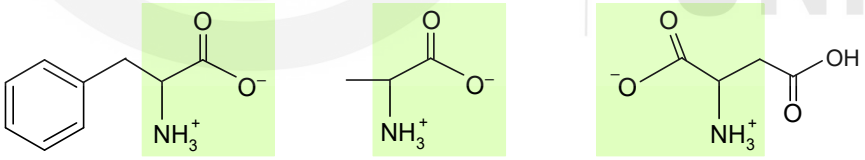
Structure of glycine in ionised and unionised form

We have already mentioned that 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 both acids and bases. Molecules which exhibit this property are known as **amphoteric** molecules. In other words, this means an amphoteric molecule 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 (Z-ion),** as the acidic carboxyl group donates a hydrogen ion and the basic amino group accepts one.



cation (at low pH) zwitterion anion (at high pH)

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. However, generally we show these molecules in the unionised form for the sake of convenience. Some amino acids in their zwitterion form are given below.



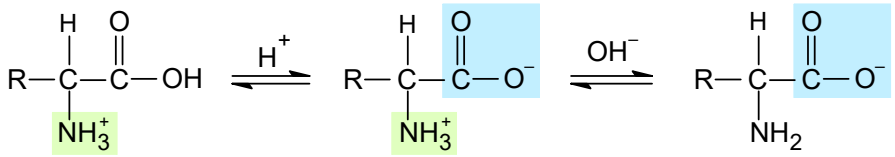
Phenylalanine Alanine Aspartic Acid

The existence of amino acid molecules in the zwitterionic form leads to an important property in these molecules. Let us study it in the following subsection.

4.3.2 Isoelectric Point

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 zwitterionic 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 as shown in the following reaction.



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

Net charge= +1 | **Net charge= 0** **Isoelectric point** | **Net charge= -1**

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.**

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. These molecules act as **buffers** in body fluids. Thus, one of the important functions served by proteins in the blood is to maintain the pH of blood in the narrow range of 7.35 to 7.45.

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 protein 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 (discussed in Unit 6 of this block) 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.

Let us look into the titration curves in amino acids that help in indicating the charged state of the molecules on the basis of the pH of the solution. Before going to the next subsection try to answer the following SAQ.

SAQ 4

Tick [√] mark the appropriate answer.

Amino acids are called amphoteric because they have

i) a basic group [ ]

ii) an acidic group [ ]

iii) both acidic and basic groups [ ]

iv) charged group as a side chain [ ]

4.3.3 Titration Curves in Amino Acids

When the pH of given volume of a sample solution is changed by successive addition of acid or alkali, a titration curve is obtained. The curves are usually plots of pH against the volume of titrant added or more correctly against the number of equivalents added per mole of the sample. This curve empirically defines several characteristics. The precise number of each characteristic depends on the nature of the acid being titrated. The characteristics are:

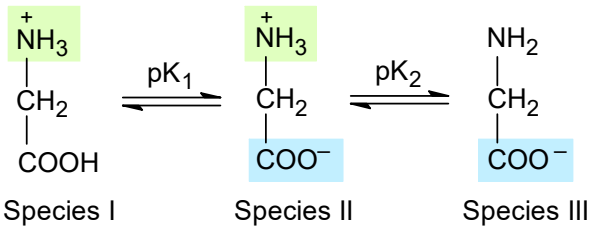
1) the number of ionising groups

2) the pK_a of the ionising group(s)

3) the isoelectric pH value

The ionic form of the amino acid present in an aqueous solution depends on the pH of the solution. Titration curves of amino acids are very useful for their identification. Titration curve of an amino acid is obtained when a solution of an amino acid in a strong acid, such as HCl is titrated against a strong base of the same strength as NaOH. The pH is recorded after each addition, and a graph is plotted between change in pH at y-axis and the amount of base at x-axis as shown in Fig. 4.2. Due to the presence of different ionisable groups, different amino acids have different types of curves. Let us understand the titration curve of amino acids with an example of simplest structured amino acid viz. glycine.

You now know that glycine has one -COOH and one $–NH_3$ group in fully protonated form with a net positive charge. In Fig 4.2 you can see two stages corresponding to the dissociation of these two groups. Initially glycine is fully protonated (species I) with pH approximately 1 and has two ionisable groups capable of giving protons. As the titration proceeds and pH rises glycine undergoes dissociation and becomes a Z-ion (species II). On plotting the titration curve you can see that by addition of NaOH the pH gradually rises and reaches a flat zone at pH 2.3 (pK_1=2.34). This is the stage when –COOH has half dissociated into COO^- and H^+. This flat zone indicates a **buffering zone**. As the titration proceeds a sharp rise in pH can be seen from pH 4.8. This is because at this pH, glycine existing as a Z-ion is unable to provide proton to neutralise the base. With further addition of NaOH and rise of pH, a second flat zone over a pH range of 9-10 can be noted (pK_2=9.6). At this range the NH_3 group deprotonates and generate an anionic form of the amino acid (species III). This is the second buffering zone of glycine. The midpoint between two pK values is pI which is 5.97.



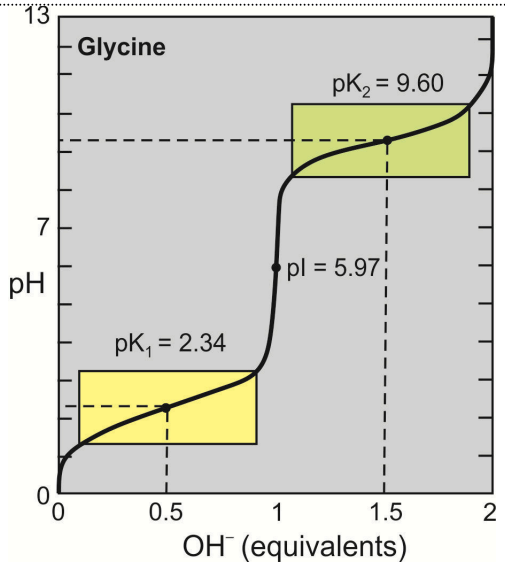


Fig 4.2: Titration curve of glycine (0.1M) at 25°C

The ionic species predominating at key points in the titration are shown above the graph. The shaded boxes, centered at about pK_1 =2.34 and pK_2=9.60 indicate the regions of greatest buffering power. In general, the isoelectric points of the amino acids can be computed by the following formula,

$$pI = 1/2\ (pK_i + pK_j)$$

where, pK_i and pK_j refer to the pK_1 and pK_2 values i.e., the pK_a values for the main chain –COOH and the α- NH_3 group respectively for the neutral amino acids. However, for acidic amino acids these refer to pK_1 and pK_R while for the basic amino acids these represent pK_R and pK_2 values respectively, pK_R being the pK_a value of the side chain functional group.

Now let us understand the technique of separation of amino acids in the next section. Before proceeding further attempt the following SAQ.

SAQ 5

A zwitterion consists of:

i) a positive charge. [ ]

ii) a negative charge. [ ]

iii) both positive and negative charges. [ ]

iv) nil charge. [ ]

4.4 SEPARATION OF AMINO ACIDS

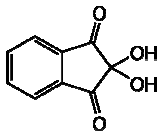
The amino acids earlier separated from the mixture of amino acids are obtained on hydrolysis of plant and animal proteins. The extracted protein is then purified by dialysis and should not contain any ionic impurities. It is then dried and subjected to acidic or alkaline hydrolysis or to enzymatic degradation with proteolytic enzymes. The separation of the amino acids from protein hydrolysate is then carried out with the help of separation procedures

like electrophoresis, ion-exchange chromatography or paper chromatography etc. We will discuss two of the simplest methods of separation of amino acids that is paper chromatography and paper electrophoresis in the following subsections.

4.4.1 Paper Chromatography

In paper chromatography, amino acids are separated as the consequence of differences in their partition coefficients between water and an organic solvent. In this technique the amino acid solution is placed on one side of the sheet of a moist filter paper with the help of a capillary tube, which is then placed in a jar as shown in Fig. 4.3 (a). The organic solvent contained in the jar migrates upward by capillary action across the paper, carrying the amino acids with it along one edge. When the solvent reaches the marked top of the paper, the paper is removed and dried in the air. The spots remain invisible at this stage. These can be made visible by spraying the paper with ninhydrin solution when different amino acids appear at different places depending on their interaction with the spraying agent as shown in Fig. 4.3 (b).

Ninhydrin reacts with amino group of amino acids to form a deep blue coloured solution.



Structure of ninhydrin

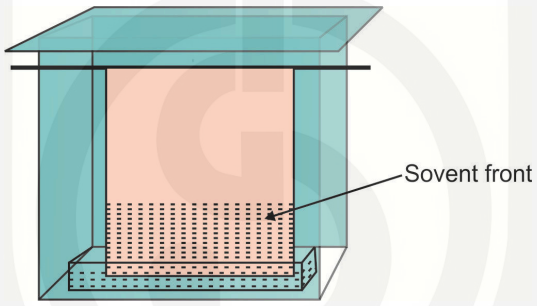
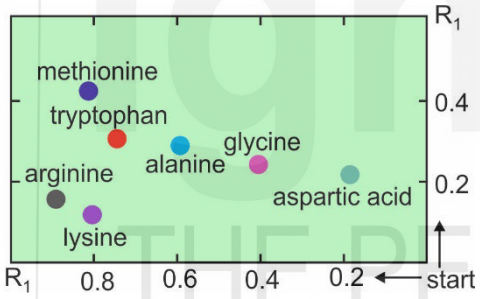


Fig.4.3: (a) Chromatography jar



(b) Amino acids with spraying agent

4.4.2 Paper Electrophoresis

Paper electrophoresis involves placing a mixture of proteins or amino acid 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. Thus, it is the separation of amino acids based on the migration of amino acids in an electric field. In the experimental procedure the amino acid mixture is spotted onto a sheet of filter paper, the paper is wet with a buffered salt solution and placed between two electrodes (cathode and anode) and high voltage applied. At the neutral pH, the acidic amino acids will have a net negative charge and will migrate toward the anode i.e., positive pole while the basic amino acids will migrate towards the cathode i.e., the negative pole. Electrically neutral amino acids will not migrate much, unless the pH is made acidic or basic.

The front and top view of the paper electrophoresis apparatus is given in Fig. 4.4.

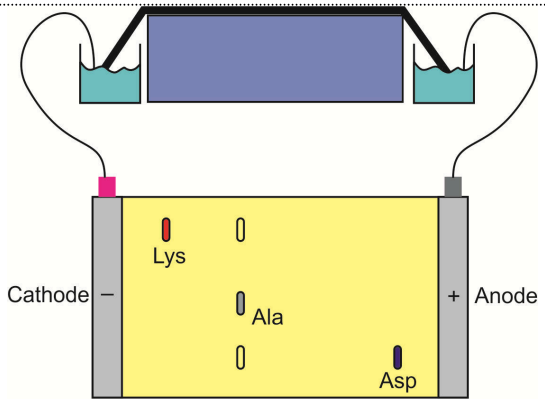


Fig. 4.4: Paper electrophoresis showing migration of amino acid

SAQ 6

Which of the following reagents is used in the identification of amino acids by paper chromatography?

i) Nessler's reagent ii) Benedict's reagent iii) Fehling's solution iv) Ninhydrin

4.5 SUMMARY

- Amino acids are the monomeric units of proteins that are large complex molecules responsible for a multitude of functions essential to life. The amino acids are called α-amino acids because of the presence of amino group at the α- carbon atom i.e. the carbon atom next to the carboxyl group. The amino acids are commonly represented by a three letter symbol or one letter symbol.
- All the 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. The carbon also carries a hydrogen atom and a fourth group represented as the R-group or the side chain. The side chain is characteristic of each amino acid.
- With the exception of glycine, all other amino acids show optical activity and it is the L isomer that is found in the living beings. These are classified based on chemical composition into aliphatic, hydroxy, sulphur containing acidic, basic and aromatic types. On the basis of polarity of side chain, these are classified into nonpolar, polar (uncharged), polar negatively charged) and polar (positively charged) types.
- The carboxyl group and amino group of amino acid are capable of ionisation in an aqueous solution. Therefore, in water, these act as acids and bases. Molecules which exhibit this property are known as amphoteric molecules. As an amphoteric molecule 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 (z-ion). The acidic carboxyl group donates a hydrogen ion and the basic amino group accepts one.
- The ionisable side chains of amino acid residues in proteins contribute to their acid-base properties and also to their buffering capacity. These

properties are depicted by the titration curves of amino acids that are very useful for their identification.

- Two techniques used for separation of amino acids are paper chromatography and paper electrophoresis. The chromatographic technique makes use of ninhydrin as the spraying agent while electrophoresis uses an electric field to separate amino acids.

4.6 TERMINAL QUESTIONS

1. List the ten essential amino acids.
2. Describe the structure of a typical amino acid.
3. Classify amino acids based on the chemical composition.
4. What is Z-ion and how does it affect titration of amino acids.
5. Name and describe the technique used to separate amino acids.
6. Give an example of the amino acid with the following in the side chain.

   i) basic group ii) acidic group iii) aliphatic group iv) aromatic group
7. Write the three letter symbol representation the following amino acids:

   i) Valine ii) Leucine iii) Isoleucine iv) Tyrosine

   v) Alanine vi) Phenylalanine

4.7 ANSWERS

Self Assessment Questions

1. a) amino acids b) α-amino acid c) glutamine
2. a) it does not have a chiral carbon atom

   b) left hand side c) L d) glyceraldehyde
3. a) (iv) b) (iii)
4. iii)

5 iii)

6 iv)

Terminal Questions

1. leucine, isoleucine, lysine, methionine, phenylalanine, tryptophan, threonine, valine, histidine and arginine.
2. The amino acids have a common structural motif. They have an amino group at one end and a carboxyl group at the other end, both of which are attached to the α-carbon atom i.e. the carbon atom next to the carboxyl group. The α-carbon also carries a hydrogen atom. The fourth group on this carbon is known as the side chain and is denoted as R.

3. Amino acids can be classified based on their chemical composition as follows:

   - Aliphatic amino acids: The aliphatic amino acids have aliphatic chain as R group
   - Hydoxy amino acids: The hydroxyl amino acids have hydroxyl group in their R group and include serine, tyrosine and threonine.
   - Sulphur containing amino acids: contain sulphur in their R-group. This group has methionine and cysteine.
   - Acidic amino acids: These are dicarboxylic acids containing an additional-COOH group in the side chain
   - Basic Amino acids: The basic amino acids contain additional basic-NH_2 groups or heterocyclic ring.
   - Aromatic amino acids: Aromatic groups are present as R group.

4. The carboxyl group and amino group in an amino acid can ionise in an aqueous solution. These can accept as well as donate a hydrogen ion and can act as acids and bases. Thus these act as amphoteric molecules. At or around neutral pH, an amino acid forms a dipolar ion or a zwitterion (z-ion). Thus a zwitterion has a positive as well as a negative charge within the same molecule at neutral pH. The slope in titration curve is observed when at a particular pH, amino acid exists in a Z- ion form and therefore is unable to provide proton to neutralise the base.

5. Paper electrophoresis is the technique used for the separation of amino acids. It is based on the migration of amino acids in an electric field. An amino acid mixture is spotted onto a sheet of filter paper, the paper is wet with a buffered salt solution and placed between two electrodes and high voltage applied. At neutral pH, the acidic amino acids will have a net negative charge and will migrate towards the anode while the basic amino acids will migrate towards the cathode. Electrically neutral amino acids will not migrate much unless the pH is made acidic or basic.

6. i) arginine, lysine and histidine

   ii) aspartic acid, glutamic acid

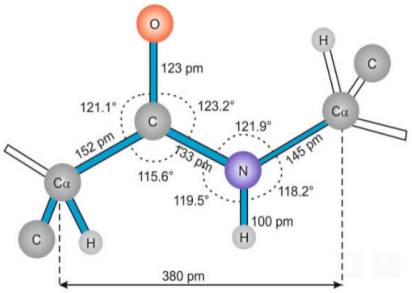
   iii) glycine, alanine, valine, leucine

   iv) phenylalanine, tyrosine and tryptophan

7. i) Val iv) Tyr

   ii) Leu v) Ala

   iii) Ile vi) Phe



UNIT 5

PEPTIDES

Structure

5.1 INTRODUCTION

In the previous unit you studied about α-amino acids which are the monomeric structural building blocks of proteins. A detail account of their structure, classification, physical properties and separation was given there. You might know that amino acids combine together in small to large numbers to form dimeric, trimeric or polymeric molecules called the peptides. Peptides are biologically significant molecules involved in a number of biological functions like, metabolism, regulation of blood pressure, enzyme inhibition, etc.

In this unit you will learn as to how the amino acids combine through an amide bond also called a peptide bond to form peptides. You will also learn about the structure of the peptide formed and how peptides are named. The synthesis of peptides by two methods is discussed in detail. The sequence of amino acids forming a peptide, especially a polypeptide is very important as these are the key biomolecules for the formation of proteins and a slight change in the amino acid sequence would mean changing the physiological state of a cell. You will study the method of determining the amino acid sequence of a peptide. Finally, some of the biologically important peptides have been discussed in the unit. In the next unit you would study how these polypeptide chains form the high molecular weight protein molecules.

Expected Learning Outcomes

After studying this unit, you should be able to:

- ❖ explain how amino acids are covalently linked together to form the linear polypeptide chain;
- ❖ draw the structure of a di- tri- and polypeptide and name these;
- ❖ list and describe the steps involved in the synthesis of peptides;
- ❖ describe the method of amino acid sequencing of peptides; and
- ❖ list some of the biologically important peptides and describe their role in living beings.

5.2 PEPTIDES: FORMATION AND STRUCTURE

Condensation product: a substance obtained by the polymerization of one substance, or by the union of two or more with the elimination of a water molecule.

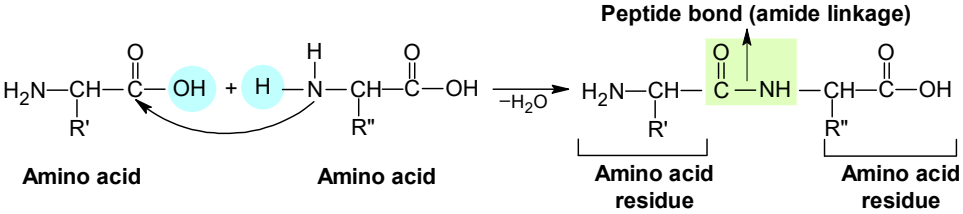
As mentioned earlier, two or more amino acids are linked together to form a condensation product called a **peptide**. This is one of the most important reactions of amino acids about which you will study in the following subsection. You will also learn the structure of peptides and the way these are named.

5.2.1 Formation of Peptides

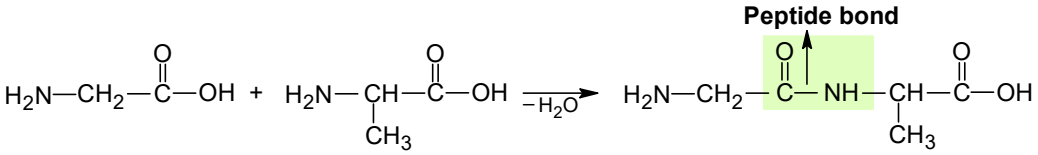
Two or more amino acids react together to form a peptide by the elimination of a water molecule. The amino acids that combine together can be same or different. We may call it a **condensation** reaction as it proceeds with the elimination of a water molecule. The reaction is between the α-carboxy group of one amino acid and the α-amino group of the other amino acid with the formation of an **amide linkage (CONH)**. These amide linkages are also called **peptide bonds**. The peptide bond can undergo hydrolysis with an acid or base or with specific enzymes to give the constituent amino acids.

The amino acids forming the peptide are called **amino acid residues.** In a peptide, the amino acid residue structures lack a hydrogen atom of the amino group (-NH-CHR-COOH), or the hydroxyl moiety of the carboxyl group (NH_2-CHR-CO-), or both (-NH-CHR-CO-). We would elaborate on the structure of the amino acid residues in the next subsection. The general reaction representing a peptide bond is given below:

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



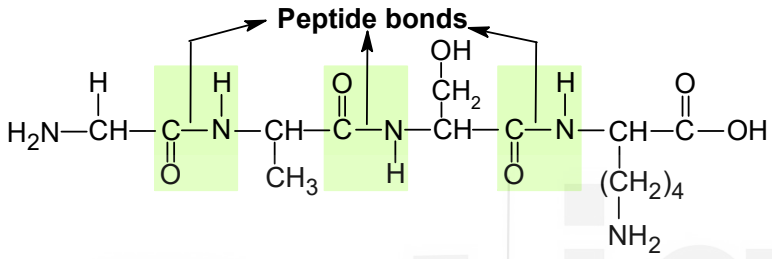
The above reaction can continue by reacting with another amino acid in a similar way and form large peptides. Depending upon the number of amino acid residues in a peptide, they are known as dipeptides, tripeptides, etc. For example, **dipeptide** has two amino acid residues:



Glycine **Alanine** **Dipeptide-glycylalanine**

Similarly, a **tripeptide** results from the union of three amino acids and a **tetrapeptide** links together four amino acids. Generally, peptides with more than 10 amino acid residues are known as **polypeptides.**

The structure of a typical tetrapeptide is given below:



Tetrapeptide-glycylalanylseryllysine

We would look into the structure and stereochemistry of the peptide bond in the following subsection. Before going to the next subsection, define the following.

SAQ 1

Define the following.

(a) Peptide

(b) Peptide bond

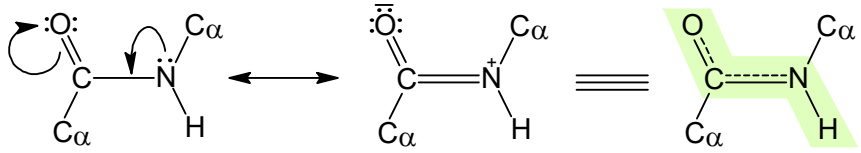
(c) Polypeptide

5.2.2 Structure of a Peptide Bond

The peptide bond chemically is an amide bond i.e., CONH which is planar in nature. In other words, all the atoms involved in bond formation lie on the same plane. The presence of double bond between carbon and oxygen which can move over to carbon and nitrogen makes the overall molecule exhibit **geometrical isomerism**. As a result, the molecule exists in *cis* and *trans* isomeric forms. The *trans* form of the peptide is normally the energetically favourable choice as it is estimated to be about 1000 times more stable than the *cis* form. Therefore, in a peptide bond (–CONH–) the hydrogen of the substituted amino group is almost always *trans* to the oxygen of carboxyl group.

The geometrical dimensions of the peptide bond impose certain restrictions on the folding of the polypeptide chain. The C–N bond length in the peptide bond is 133 pm, which in comparison to the non-peptide C–N bond length (146 pm)

is shorter, but is slightly longer than the normal C=N double bond length of 127 pm. This variation in length of C–N bond is suggestive of a partial double bond character of the peptide bond (explained below). This is due to resonance between the two canonical forms of peptide group as shown below:



Pauling and Corey were the first to discover that the peptide linkage is planar.

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.

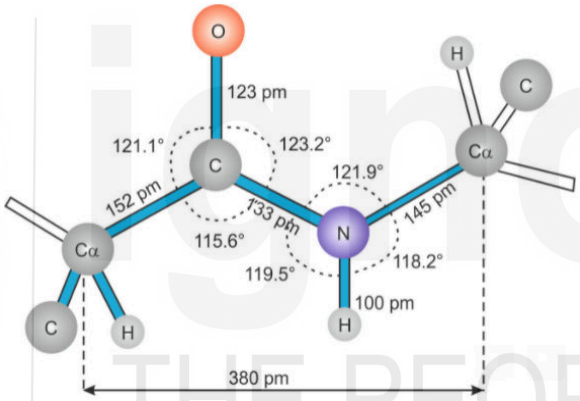


Fig. 5.1: The geometry of the peptide backbone.

In the next subsection we will study as to how the peptides are named. Before that try to answer the following SAQ.

SAQ 2

Match the following:

(a) Trans form of peptide bond	(i) Pauling and Corey
(b) Geometric isomerism	(ii) 1000 times more stable than *cis* form
(c) Planar peptide linkage	(iii) Cis-trans isomersim
(d) Peptide bond	(iv) –CONH–

5.3 NOMENCLATURE OF PEPTIDES

The amino acids, as they exist in the peptides are referred to as

amino acid residues. In a peptide chain, the amino acid residues with R as the side chain can have the following structures.

- lack hydrogen atom of the amino group and represented as –NH–CHR–COOH

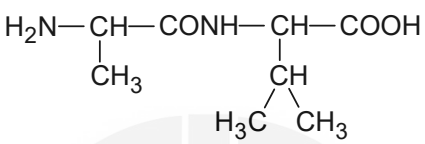
- lack hydroxyl moiety of the carboxyl group

  NH_2–CHR–CO–

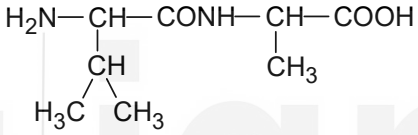
- lack both the H atom and hydroxyl moiety

  –NH–CHR–CO–

You should also note that a peptide is named from the amino end, which is having free or acylated α-amino group. For example, in the tetrapeptide given earlier in subsection 5.2.1, the amino end also called the **N-terminal** is a glycyl residue. Similarly, the other end having a free or derivatised α-carboxyl group is called the carboxyl end or **C-terminal**. In the above case, a lysine residue is at the C-terminal. The name of a peptide is written by replacing 'ine' ending in the names of acylating amino acid residues with 'yl' while that of the C-terminal residue is used as such. Thus, if the amino acids alanine and valine condense in such a way that alanine acylates valine, the dipeptide formed is named alanylvaline and if the condensation is in the reverse order, the product is called valylalanine.

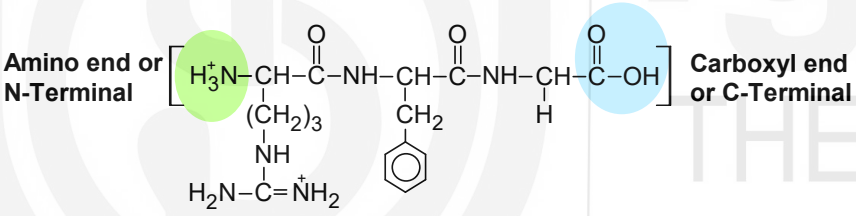


L-Alanyl-L-valine



L-Valyl-L-alanine

This will be clear from the tripeptide, arginylphenylalanylglycine, given below:



Tripeptide - Arginylphenylalanylglycine

Simple peptides containing a few residues of a certain amino acid are named with prefixes as given in Table 5.1 to indicate the number of amino acid residues of the amino acid present, e.g. tetra-L-lysine refers to a peptide containing four residues of L-lysine. 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 abbreviation of amino acids, shown in Table 4.1 of Unit 4.

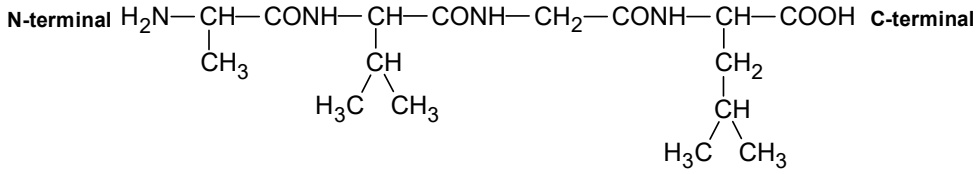
Table 5.1: Prefixes for short peptides

Residues	Prefix
2	di-
3	tri-
4	tetra-
5	penta-
6	hexa-
7	hepta-
8	octa-
9	nona-
10	deca-

Let us learn to represent a peptide or a polypeptide in short manner in the following subsection.

5.3.2 Representation of Peptides

You are now familiar with the terms N-terminal and the C-terminal in peptides. While representing the structural formulae of peptides, the N-terminal is conventionally written to the left and the C-terminal is written to the right.



L-Alanyl-L-valyl-glycyl-L-leucine

This is followed even when we are writing the formulae of the peptides.

$NH_2–CH(CH_2COOH)–CO–NH–CH(CH_3)–CO–NH–CH_2–COOH$

L-Aspartyl-L-alanylglycine

The complete order of amino acids in a protein is called its sequence and is conveniently expressed by using the abbreviated names of the amino acids read from N- to C- terminal. The presence of peptide bonds between different residues is indicated by hyphens. The thirty residue long, amino acid sequence of the B chain of insulin in terms of three letter abbreviations is represented as follows.

Phe-Val-Asn-Gln-His-Leu-Cys-Gly-Ser-His-Leu-Val-Glu-Ala-Leu-Tyr-Leu-Val-Cys-Gly-Glu-Arg-Gly-Phe-Phe-Tyr-Thr-Pro-Lys-Thr

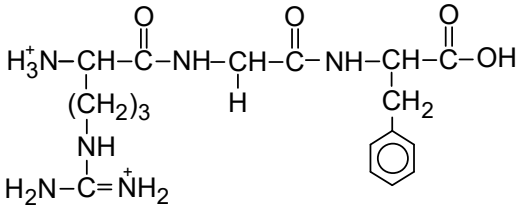
In terms of one letter abbreviation it is given as:

FVNQHLCGSHLVQALYLVCGQRGFFYTPKT

You can check your understanding of peptides by attempting the following SAQ. After answering the SAQ, in the next section you will learn the strategies to synthesise a peptide.

SAQ 3

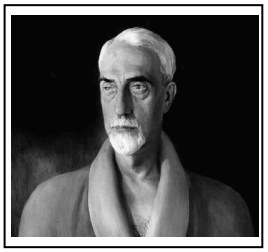
Name the following peptide:



5.4 SYNTHESIS OF PEPTIDES

Synthetic peptides (e.g. dipeptide sugar-substitute aspartame, hormones, such as insulin, oxytocin, adrenocorticotropic hormone, and calcitonin are important commercially or in medicine. Therefore, easy, cost efficient, and

reliable methodology for the chemical synthesis of peptide is important. The concept for stepwise assembly of peptides from amino acid precursors is based on peptide elongation via a coupling reaction between amino acids, followed by removal of a reversible protecting group. The first peptide synthesis, on the basis of analysis of the protein hydrolysis products in the beginning of the twentieth century as well as the creation of the term "peptide," was reported by Emil Fischer and Ernest Fourneau in 1901. They suggested that proteins are made up of a large number of amino acids linked through amide linkages. Fischer supported this fact by actually synthesising a peptide, glycylalanine. The synthesis of peptides can be accomplished by two ways; one is the solution phase synthesis and the other more recent is solid phase synthesis. Let us look into the more elaborate solution phase synthesis in the following subsection.



Emil Fischer

5.4.1 Solution Phase Synthesis of Peptides

Ernest Fourneau

Ernest Fourneau

You have learnt the basic reaction of formation of a peptide. However, there are a number of issues that had to be considered while carrying out the synthesis. These are given below.

- When two amino acids come together to react, there is a probability of reaction in more than one way due to the presence of carboxyl and amino functional groups. We can understand this by taking an example of reaction between glycine and alanine. The reaction can lead to the formation of four dipeptides, Ala-Gly, Ala-Ala, Gly-Gly and the desired Gly-Ala besides a mixture of tri and higher peptides. This can be represented as given below.

  Gly + Ala = Gly-Ala + Ala-Gly + Ala-Ala + Gly-Gly +
  Ala-Ala-Gly + Gly-Gly-Ala……..

- The α-amino and the carboxyl functional groups do not react on their own and may react just by proton transfer to give salts. These groups need to be activated for the reaction to occur.

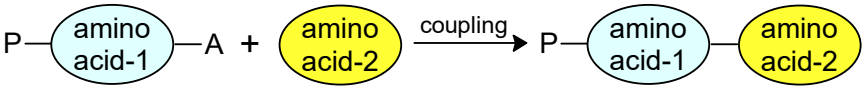
- The side chains of some of the amino acids with functional groups can interfere with the formation of the amide bond and need to be protected.

In the light of the above facts the following three strategies are followed for peptide synthesis.

1. ***Protecting the amino group and activating the carboxyl group of one amino acid (aa-1) and reacting it with the other amino acid (aa-2)***

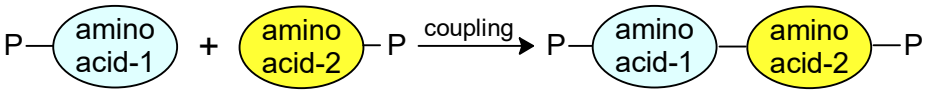
   The protection (**P**) of the amino group is helpful in making this group unreactive so that the amino acid reacts with its carboxyl group only. The resulting N terminal protected dipeptide chain is extended by activating (**A**) its carboxyl group and reacting with another amino acid

   The general reaction for the dipeptide formation can be depicted in the following manner.



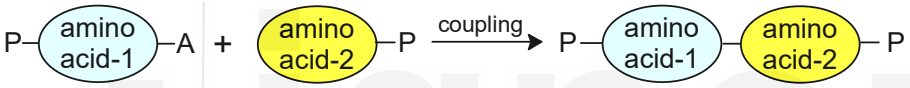
2. ***Protecting the amino group of one amino acid and the carboxyl group of the other amino acid and carrying out the reaction***

The dipeptide formed here is protected on both the ends. Therefore, for extending the amino acid chain one of the ends needs to be deprotected and accordingly the protection has to be done in the amino acid to be added. The final product is then deprotected at both the ends to get the free peptide. The formation of a dipeptide is shown below.



3. ***Protecting the amino group and activating the carboxyl group of one amino acid and reacting with the second amino acid with its protected carboxyl group***

In this strategy also both the ends are protected. The carboxyl group is deprotected and then activated in the dipeptide to combine with another amino acid which is protected at C-terminal to continue the chain.



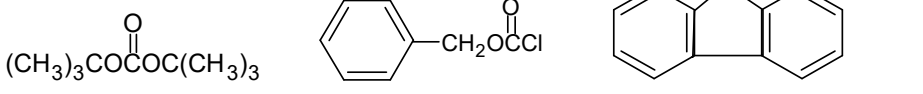
You have to take a note that these strategies follow a process of protection, activation and deprotection of the functional groups so that the multiple formation of peptides is avoided or only a desired peptide is formed. Thus, protection of one of the functional groups is done in one amino acid and the same functional group in the other amino acid is not protected. Besides this, care has to be taken to protect the groups present in the side chains so as to avoid formation of additional compounds. Let us look into the exact steps of peptide synthesis in the next subsection.

5.4.2 Procedural Steps in Peptide Synthesis

Let us look into the compounds and process used for protection, activation and deprotection of the functional groups in amino acids and understand the process of formation of peptides.

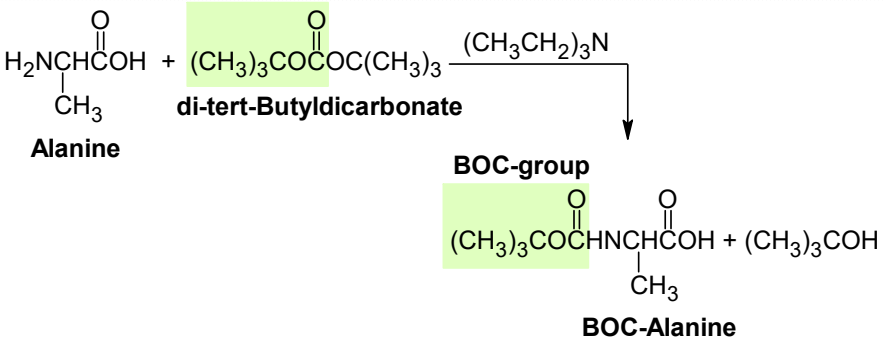
Protection and Deprotection of Amino Group

It is clear that protection of a group in the constituent amino acids is essential for getting a desired peptide. The protection is done by the acylation of amino group by treatment with acyl chlorides or anhydrides. The acyl (amide) formation reduces the basicity and nucleophilicity of amines as a result their reactivity gets reduced. There are a number of reagents used for protection of amino group, the structures of a few are listed below.

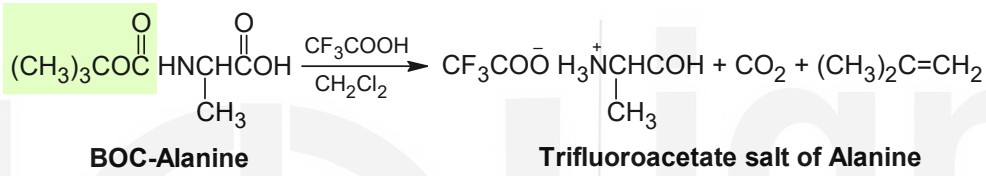


di-tert-Butyldicarbonate (t-BOC) **Benzylchoroformate (Cbz or Z)** **Fluoren-9-ylmethylchloroformate (Fmoc)**

As an example, the reaction between an amino acid and di-tert-butyl dicarbonate (gives a t-BOC group) in presence of a base like, triethylamine (TEA) can be represented as given below:

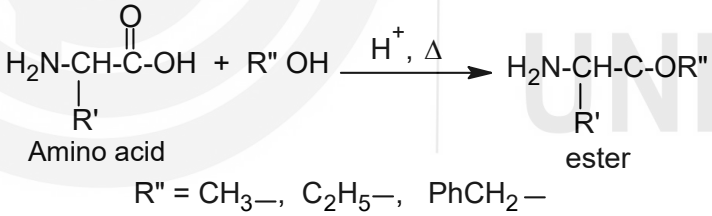


The removal of the protecting group is required after the reaction of peptide formation is complete. The removal of the protecting group is called **deprotection**. In case of N-protection with BOC, deprotection can be done by hydrogenolysis or in presence of mild acids. Hydrogenolysis is done by reaction with H_2 in the presence of a transition metal catalyst. The reaction under mild acidic conditions: trifluoroacetic acid diluted with dichloro- methane is written as following.

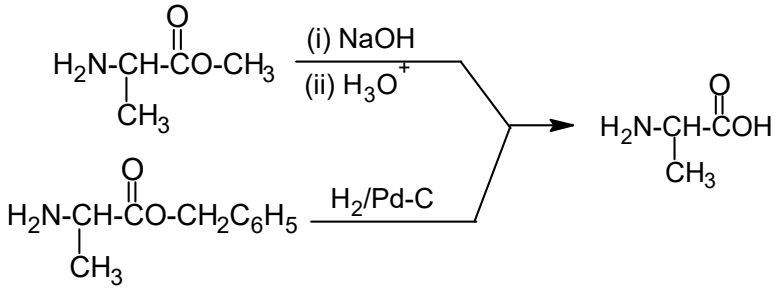


Protection and Deprotection of Carboxyl Group

The carboxyl group of amino acids is generally protected as methyl, ethyl or benzyl esters, prepared by the standard methods of ester formation. Fischer esterification is used for preparing the methyl and ethyl esters as per the reaction given below.



As for amino group, deprotection of protected carboxyl group can be done by hydrolysis using aqueous base. The benzyl esters can be deprotected by using HBr in acetic acid.



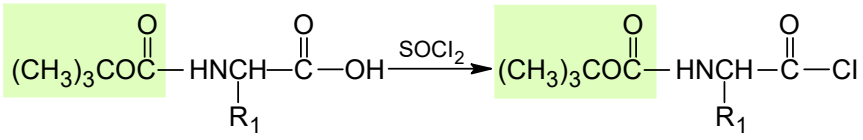
Activation of Carboxyl Group and Amino acid Coupling

Coupling is the generation of peptide bond between the two suitably protected amino acids.

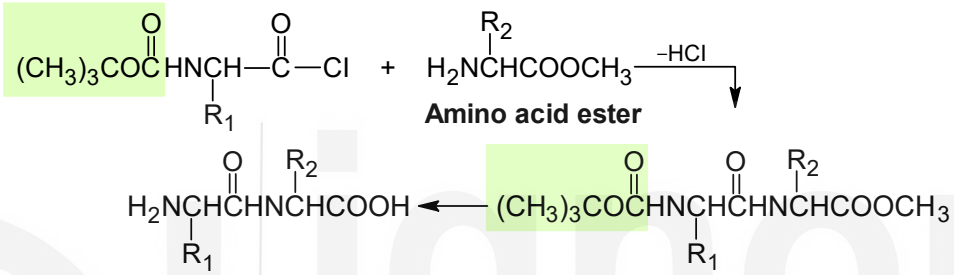
The final step in peptide synthesis is coupling *i.e.* combining of the amino acids protected at appropriate terminals. This requires first the activation of the

carboxyl group. There are number of methods available for this purpose. We will be discussing only a couple of these.

In the simplest way the N-protected amino acid is activated at C-terminal by converting it into its **acid chloride** by treating with thionyl chloride ($SOCl_2$). The amino acid which is C-terminal activated is made to react with an amino acid ester to give the protected peptide (Schotten-Baumen reaction). The desired peptide with free amino and carboxyl groups is obtained thereafter with mild hydrolytic reactions. The reactions are given below.



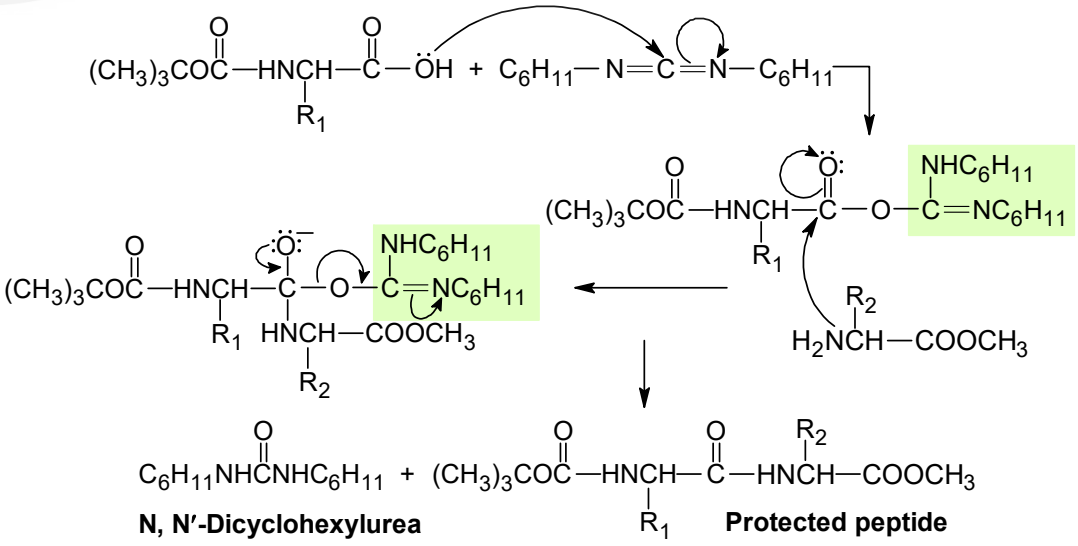
BOC-Amino acid **BOC-Amino acylchloride**



Amino acid ester

Free Dipeptide **Protected Dipeptide**

The other type of coupling reagents are mainly carbodiimides such as **dicyclohexylcarbodiimide** (DCC) or diisopropylcarbodiimide (DIC) required for the activation of the C-terminal carboxylic acid on the incoming amino acid. These coupling reagents react with the carboxyl group of the amino acid to form a highly reactive O-acylisourea intermediate. The O-acylisourea intermediate is rapidly displaced by nucleophilic attack from the deprotected primary amino group on the N-terminal of the growing peptide chain to form the nascent peptide bond. This method was introduced by Sheehan and Hess in 1955. The reactions are given below.



N, N′-Dicyclohexylurea **Protected peptide**

You can revise in general the whole process of peptide synthesis by the schematic diagram given in Fig. 5.3.

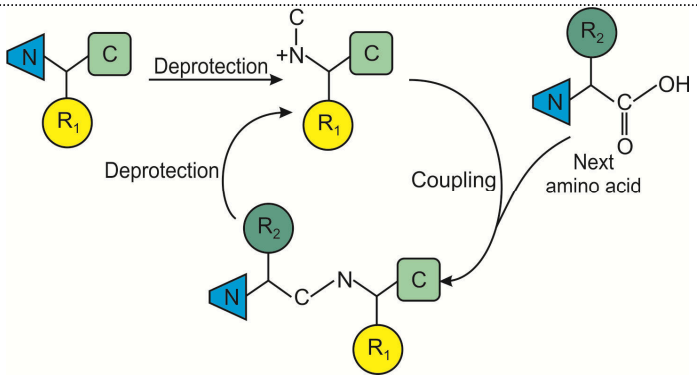


Fig. 5.3: Schematic diagram indicating the process of peptide synthesis.

Before learning about the second process of peptide synthesis answer the following questions.

SAQ 4

Match the following action with chemicals used in peptide synthesis

(a)	Amino acid coupling	(i)	mild acid
(b)	Protection of amino group	(ii)	benzyl ester
(c)	Deprotection of amino group	(iii)	carbodiimides
(d)	Protection of carboxyl group	(iv)	t BOC

5.4.2 Merrifield Solid Phase Synthesis of Peptides

R. Bruce Merrifield was awarded Nobel prize in 1984

The synthesis of a peptide of significant length (e.g., ten residues) by solution phase approach requires many steps, and the product must be carefully purified after each step to prevent unwanted cross reactions. To facilitate the tedious and time-consuming purifications, and reduce the material losses that occur in handling, a modification of this strategy has been developed. This procedure, known as the **Merrifield synthesis** after its inventor R. Bruce Merrifield, involves attachment of the C-terminus of the peptide chain to an insoluble polymeric resin as a solid support, usually having the form of very small beads. Separation and purification is simply done by filteration and washing the beads with appropriate solvents. The reagents for the next peptide bond addition are then added, and the purification steps repeated. The entire process can be automated and peptide synthesis machines based on the Merrifield approach are commercially available. In general terms we can say that solid phase peptide synthesis is done by the following three sets of operations (Fig. 5.4).

1. Chain assembly on the resin: The N-terminal protected amino acid is covalently attached to the polymer. This is called **anchoring**.

2. Cleavage and deprotection of the peptide chain: Here the assembled peptide is cleaved from the resin and deprotected simultaneously or sequentially.

3. Purification and characterisation of the target peptide: The peptide is purified by advance chromatographic methods and characterised by spectroscopic methods.

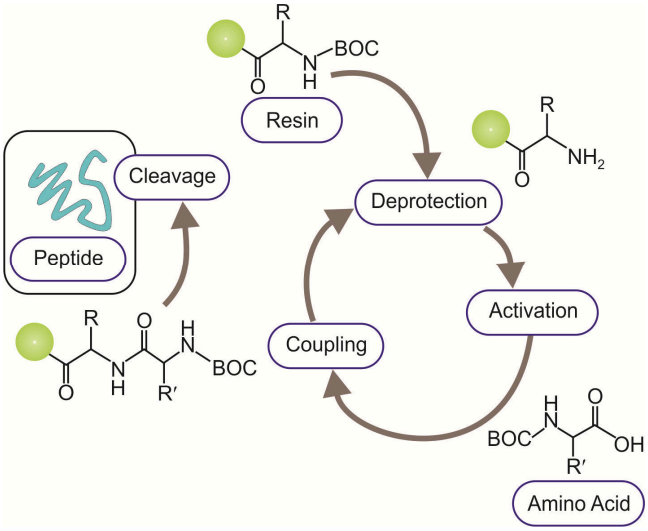


Fig. 5.4: Schematic representation of Merrifield solid phase synthesis of peptides.

Two or more moderately sized peptides can be joined together by selective peptide bond formation, provided side-chain functions are protected and do not interfere. In this manner peptides and small proteins can be synthesized in the laboratory. However, even if chemists assemble the primary structure of a natural protein in this or any other fashion, it may not immediately adopt its native secondary, tertiary and quaternary structure. Many factors, such as pH, temperature and inorganic ion concentration influence the conformational coiling of peptide chains. Scientists are still trying to understand how and why these higher structures are established in living organisms.

SAQ 5

Fill in the blanks with appropriate words.

(a) The process where the N-terminal protected amino acid is covalently attached to the resin is called ……

(b) Solid phase synthesis of peptides was invented by… ……..

(c) In the solid phase synthesis a polymer support is used at the……

5.5 DETERMINATION OF AMINO ACID SEQUENCE

The determination of sequence of amino acids constituting a protein is very significant as each type of protein has different function to perform in a cell. A number of methods are available for this purpose. These require that the N-terminal amino acid has a free amino group and the C-terminal amino acid has a free carboxyl group.

There are three common methods by which the entire sequence of a protein can be determined. In two of these methods, the actual protein is sequenced; in the third method, the DNA which encodes for the protein is sequenced, and

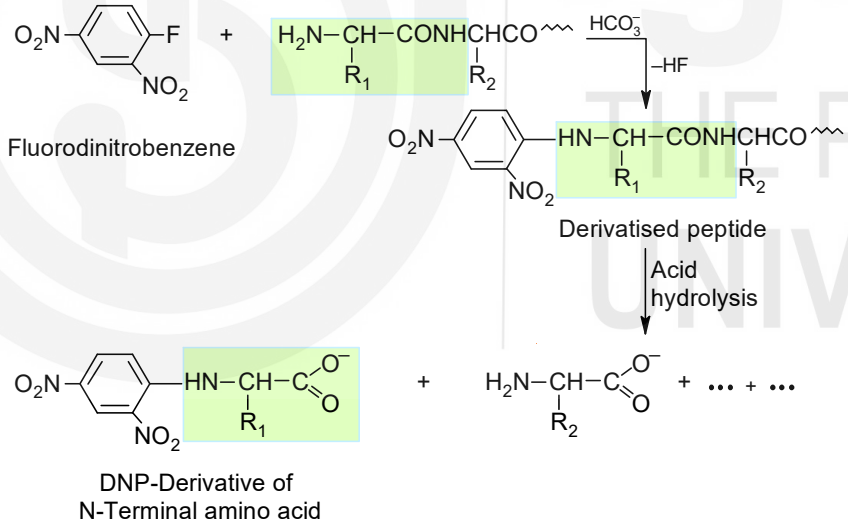
then from the DNA sequence, the amino acid sequence is determined. The first peptide hormone to be sequenced was insulin by Sanger's degradation method. Now a days the sequence of amino acid in actual protein is done by automated, sequential Edman degradation. Let us study both the methods in detail.

Sanger's Method

In 1945, Sanger developed a three stage method for sequencing of amino acids in insulin. He was awarded the Noble prize for his contribution. Sanger's method has three steps:

1. The amino terminal of amino acid is labeled or derivatised with flurodinitrobenzene (FDNB), commonly called Sanger's reagent.
2. The labeled or the derivatised peptide is subjected to acid (6N HCl) hydrolysis.
3. The labeled/derivatised amino acid is identified by its characteristic migration rate on thin layer chromatography or paper electrophoresis.

This method is useful for proteins with less than 50-70 amino acids. Larger proteins can be hydrolysed by protease or by chemicals into shorter fragments and then labeled with FDNB. The reactions can be depicted as follows.

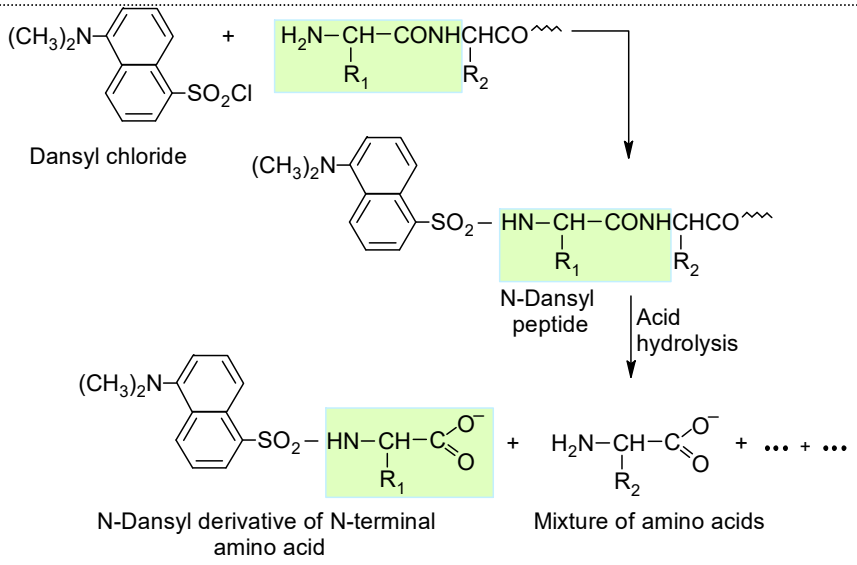


Dansyl Method

This is a much more sensitive method than the Sanger's method. It involves the reaction of the terminal amino group of the protein with dansyl chloride (5-dimethylaminonapnthalene-1-sulphonyl chloride) to give N-dansyl derivative.

The derivatised protein is then hydrolysed to give a highly fluorescent N-dansyl amino acid and a mixture of unprotected amino acids. The dansylated amino acid is detected chromatographically or with the help of its characteristic fluorescence in the UV region.

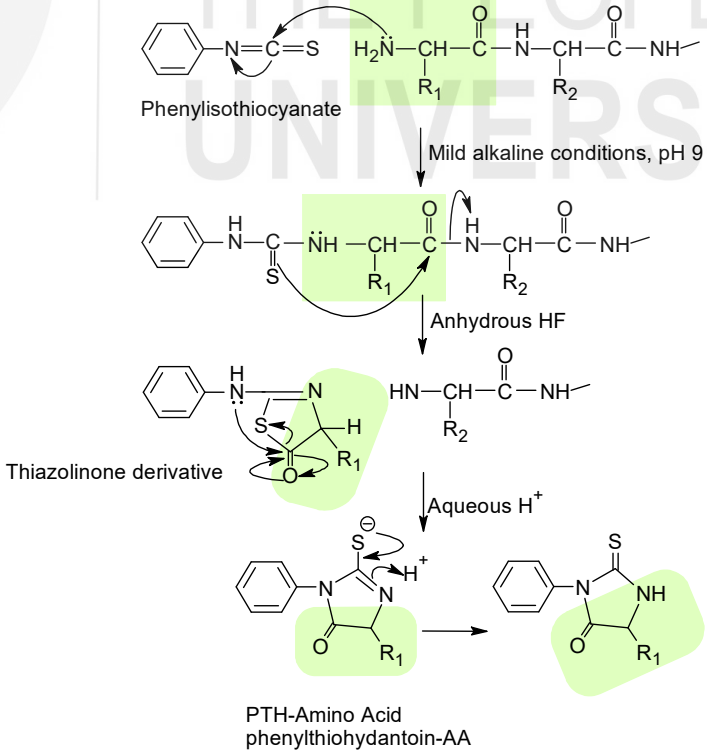
The reactions involved in the method are given below.



Edman Degradation– Sequential determination of amino terminal residue

HPLC provides a sensitive means of distinguishing the various amino acids.

Pehr Edman formulated a method where the amino-terminal residue is labeled and cleaved from the peptide in such a way that the peptide bonds between the other amino acid residues are not disturbed. In the Edman degradation, a phenylthiocarbamoyl derivative is formed when the uncharged terminal amino group of the peptide reacts with phenyl isothiocyanate. Then, under mild acidic conditions, a cyclic derivative of the terminal amino acid i.e. phenylthiohydantoin (PTH)-amino acid is liberated, which leaves an intact peptide shortened by one amino acid. The phenylthiohydantoin (PTH)-amino acid, can be identified by chromatographic procedures like high-pressure liquid chromatography (HPLC). The reactions can be given as follows.



Edman degradation of a polypeptide with Ala-Gly-Asp-Phe-Arg-Gly amino acids is given in Fig. 5.5. You can see in the figure that the labeled amino-terminal residue (PTH-alanine in the first round) can be released without

hydrolyzing the rest of the peptide. Hence, the amino-terminal residue of the shortened peptide (Gly-Asp-Phe-Arg-Gly) can be determined in the second round. The Edman procedure can then be repeated on the shortened peptide, yielding another PTH-amino acid, which can again be identified by chromatography. Three more rounds of the Edman degradations will reveal the complete sequence of the original peptide pentapeptide.

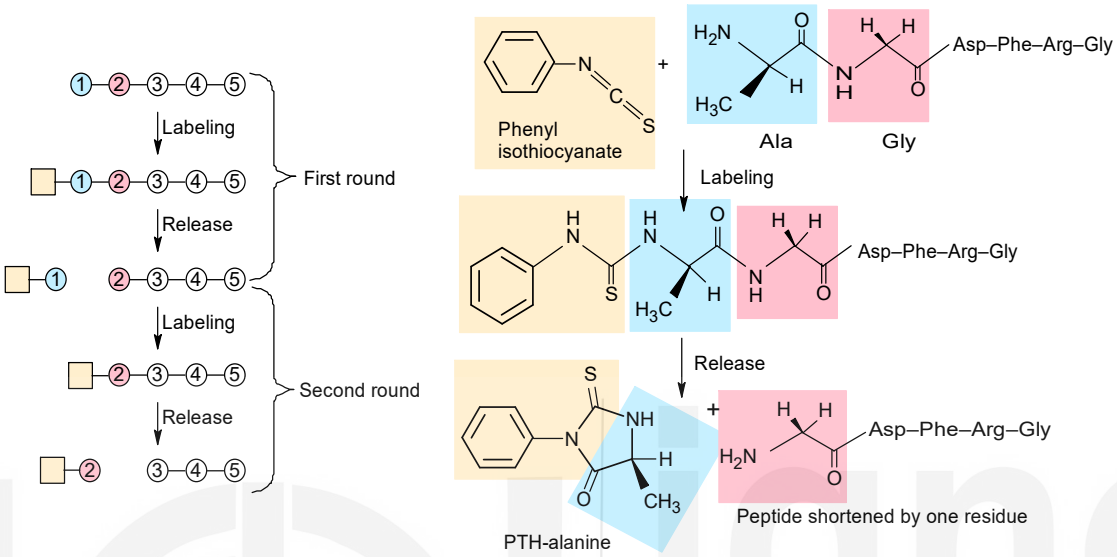


Fig. 5.5: Edman degradation of polypeptide with Ala-Gly-Asp-Phe-Arg-Gly amino acids.

The development of automated sequencers has markedly decreased the time required to determine protein sequences. One cycle of the Edman degradation, the cleavage of an amino acid from a peptide and its identification is carried out in less than 1 hour. By repeated degradations, the amino acid sequence of some 50 residues in a protein can be determined.

You will be able to appreciate the role of some biologically important peptides in the next section. Try to answer the following SAQ first.

SAQ 6

Fill in the blanks with appropriate words.

(a) In amino acid sequencing by Edman degradation a phenylthiocarbamoyl derivative is formed when the uncharged terminal amino group of the peptide reacts with _____.

(b) In amino acid sequencing by Edman degradation under mildly acidic conditions, a cyclic derivative of the terminal amino acid ----------is liberated, which leaves an intact peptide shortened by one amino acid.

(c) _____ provides a sensitive technique of distinguishing the various amino acids

5.6 BIOLOGICALLY IMPORTANT PEPTIDES

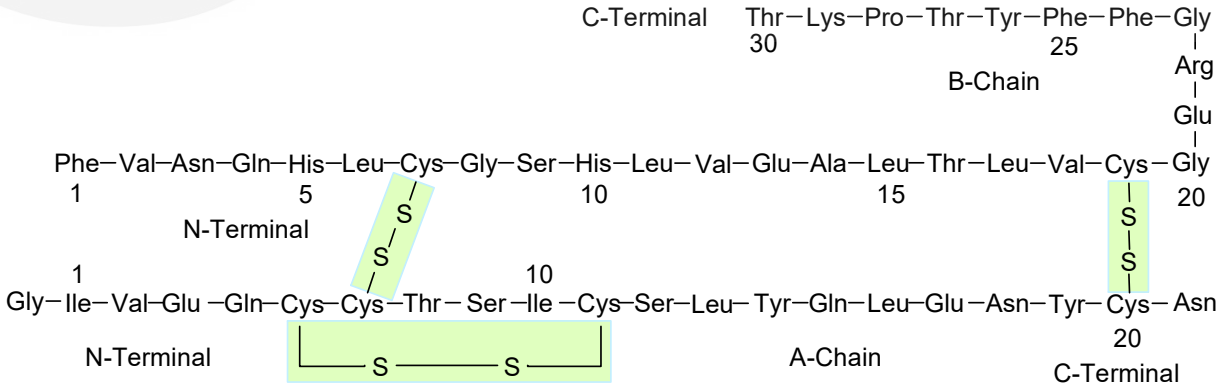
You have seen that peptides vary in their structure due to the type and number of amino acids forming these polymeric molecules. These are the key

constituents of a large number of proteins too. The variation in structure is responsible for their functional diversity and their role in physiological and biochemical processes. Let us look into some interesting and important examples of biologically important peptides given below.

Oxytocin: Oxytocin is a peptide hormone, containing nine amino acids. The peptide has the following sequence of amino acids. As you can see it contains a disulphide linkage between cysteine residues at position 1 and 6 and the C-terminal carboxyl is in the form of an amide.

Oxytocin has an important role in birth and lactation. It helps in stimulation of milk ejection, uterine smooth muscle contraction at the time of parturition (child-birth) and development of maternal behaviour towards the child soon after giving birth. It gives the feeling of satiation (fullness) when the food is absorbed and is involved in facilitating sperm transport within the male reproductive system. Due to the small size of the peptide, researchers have been putting efforts towards synthesising analogues of this hormone for application in the biological systems.

Insulin: Insulin is a peptide hormone, secreted by *islets of Langerhans* which is an endocrine gland in the pancreas. Most of us are aware that this hormone controls the blood glucose level and that insulin is a very significant hormone used in the treatment of diabetes. It was the first peptide hormone sequenced by Sangers. It is one of the smallest protein molecules having just 51 amino acid residues on two different chains of 21(A-chain) and 30(B-chain) amino acid residues linked through disulphide bonds. The side chains of Cys7 and Cys20 residues of the A-chain form the disulphide linkage with the Cys7 and Cys19 residues of the B-chain. In addition, A-chain has an intra-chain S-S loop between Cys6 and Cys11 residues.



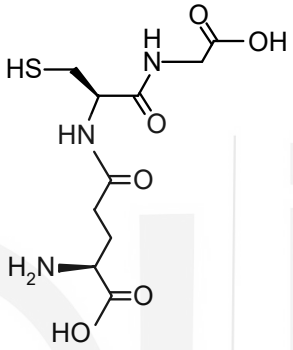
Structure of Insulin

Glutathione: Glutathione (γ-glutamylcysteinylglycine) is a tripeptide found in nearly all cells of plants, animals and microorganisms. The presence of glutathione in the body is required to maintain the normal functioning of the immune system. Glutathione is also used to prevent oxidative stress in most

cells and helps to trap free radicals that can damage DNA and RNA. There is a direct correlation with the speed of aging and the reduction of glutathione concentrations in intracellular fluids. As individuals grow older glutathione levels drop and the ability to detoxify free radicals decreases. It is known to play a critical role in the multiplication of lymphocytes (the cells that mediate specific immunity) which occurs in the development of an effective immune response.

Glutathione is also involved in the transport of amino acid and trace minerals across cell membranes and may be involved in the repair of damaged cells.

Structurally, glutathione is an unusual peptide in the sense that the N-terminal glutamyl residue forms a peptide bond through its γ-carboxyl group (in the side chain) instead of the usual carboxyl group.



Structure of glutathione

Bradykinin: Bradykinin is a nonapeptide (peptide containing nine residues) which is released by the blood plasma in response to a wasp sting. Physiologically bradykinin is a pain causing agent having the following amino acid sequence.

Arg–Pro–Pro–Gly–Phe–Ser–Pro–Phe–Arg

Bradykinin is also known as **Kalladin I** which has a closely related peptide **Kalladin II** (a decapeptide) with the following sequence.

Lys–Arg–Pro–Pro–Gly–Phe–Ser–Pro–Phe–Arg

Both the Kalladins have excellent muscle relaxing properties.

Gramicidin S: Gramicidin S is a peptide antibiotic isolated in 1944. This cyclic decapeptide contains two identical pentapeptides viz., D-Phe-L-Pro-L-Val-L-Orn-L-Leu. The two pentapeptides are condensed head to tail with each other. Ornithine is an amino acid similar to lysine with one methylene group lesser. The structure of the peptide is given below.

Val-Orn-Leu-Phe-Pro
| |
Pro-Phe-Leu-Orn-Val

Gramicidin S exhibits a broad spectrum antibiotic activity, which is attributed to its ability to bind cell membranes. However, its clinical applications are quite limited due to its low bioactivity and cytotoxicity.

Some other important biological functions of peptides are listed in Table 5.2.

Table 5.2: Structures and functions of some biologically active peptides

Name of the peptide	Primary Structure	Biological functions
Angiotensin II (Horse)	1 8 Asp-Arg-Val-Tyr-Ile-His-Pro-Phe	Hypertensive peptide; controls blood pressure by constriction of arteries. Also involved in the release of aldosterone
Glucagon (Bovine)	10 His-Ser-Gln-Gly-Thr-Phe-Thr-Ser-Asp-Tyr- 11 20 Ser-Lys-Tyr-Leu-Asp-Ser-Arg-Arg-Ala-Gln- 21 29 Asp-Phe-Val-Gln-Trp-Leu-Met-Asn-Thr	Pancreatic hormone involved in regulation of glucose metabolism
Enkephalin (Metenkaphalin)	1 5 Tyr-Gly-Gly-Phe-Met	A member of family of opiate-like peptides found in brain that inhibit sense of pain by binding to receptors in brain
Substance P	1 10 Arg-Pro-Lys-Pro-Gln-Phe-Phe-Gly-Leu-Met	A neurotransmitter
Thyrotropin Releasing Factor(TRF)	1 3 Glu-His-Pro	Secreted by hypothalamus and causes pituitary gland to release thyrotropin (or Thyroid Stimulating Hormone, TSH)
Vasopressin (Antidiuretic hormone)	1 9 Cys-Tyr-Phe-Gln-Asn-Cys-Pro-Arg-Gly I I S ——————— S	Regulates the tonicity of body fluids, maintains the volume of water in the extracellular fluid

SAQ 7

Match the following correctly.

(a)	Glutathione	(i)	hypoglycemia
(b)	Insulin	(ii)	oxidative stress
(c)	Glucagon	(iii)	hyperglycemia
(d)	Oxytocin	(iv)	blood pressure
(e)	Angiotensin	(v)	posterior pituitary

5.7 SUMMARY

- Two or more amino acids react together to form a peptide by the elimination of a water molecule by condensation reaction. The reaction is between the α-carboxyl group of one amino acid and the α-amino group of the other amino acid with the formation of an amide linkage (CONH). These amide linkages are also called peptide bonds. All the atoms involved in peptide bond formation lie on the same plane.
- A peptide is named from the amino end, which is having free or acylated α-amino group. The name of a peptide is written by replacing ‘ine’ ending in the names of acylating amino acid residues with ‘yl’ while that of the C-terminal residue is used as such. Depending upon the number of amino acid residues in a peptide, they are known as dipeptides, tripeptides, etc.

- While representing the structural formulae of peptides, the N-terminal is conventionally written to the left and the C-terminal is written to the right.
- The peptides can be synthesised by solution phase synthesis and solid phase synthesis. Peptide synthesis follows the strategy of protection, activation and deprotection of the functional groups so that the multiple formation of peptides is avoided or only a desired peptide is formed.
- Peptide sequencing is generally done by three methods- Sanger's, Dansyl and Edman degradation methods. In Sangers degradation the amino terminal of amino acid is labeled with flurodinitrobenzene, then the labeled peptide is hydrolysed by acid and then separated by its chromatographic properties. The Edman degradation sequentially removes one amino acid residue at a time from the amino end of a peptide. The uncharged terminal amino group of the peptide is labeled with phenyl isothiocyanate. Then, under mild acidic conditions, a cyclic derivative of the terminal amino acid *i.e*. phenylthiohydantoin (PTH)-amino acid. The phenylthiohydantoin (PTH) - amino acid, can be identified by chromatographic procedures like high-pressure liquid chromatography (HPLC).
- Peptides play an essential role in fundamental physiological processes and are necessary for many biochemical processes. The variation in the peptide structure is responsible for their functional diversity and their role in physiological and biochemical processes. Oxytocin, insulin, glutathione, bradykinin etc. are some examples of biologically active peptides.

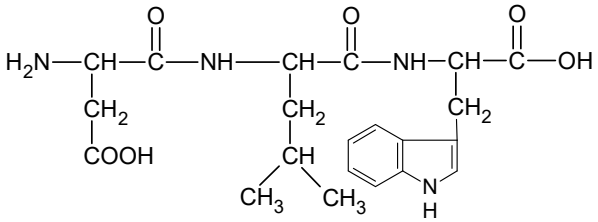
5.8 TERMINAL QUESTIONS

1. Differentiate between the liquid phase and solid phase protein synthesis.
2. List the steps involved in the Sanger's method of protein synthesis.
3. Write short notes on the following:

   (a) Oxytocin

   (b) Insulin

   (c) Glutathione
4. Explain why the peptide bond is planar.
5. Write the structure of the peptide Phe-Val-Asn.
6. The structure of a tripeptide is given below. Write its one letter and three letter representations.



5.9 ANSWERS

Self-Assessment Questions

1. (a) Two or more amino acids are linked together to form a condensation product called a peptide.

   (b) The α-carboxyl group of one amino acid binds with the α-amino group of the other amino acid to form an amide linkage (CONH) which is called peptide bond.

   (c) Peptides with more than 10 amino acid residues are known as polypeptides.

2. (a): ii), (b): iii), (c): i), (d): iv)

3. Tripeptide – Arginylglycylphenylalanine

4. (a): iii), (b): iv), (c): i), (d): ii)

5. (a) anchoring (b) R-Bruce Merrifield (c) C-terminal

6. (a) phenyl isothiocyanate

   (b) phenylthiohydantoin (PTH)-amino acid

   (c) HPLC

7. (a): ii), (b): i), (c): iii), (d): v) (e): vi)

Terminal Questions

1. There are three steps involved in solid phase protein synthesis. First step is chain assembly on the resin where the N-terminal protected amino acid is covalently attached to the polymer. This is called anchoring. The second step involves cleavage and deprotection of the peptide chain in which the assembled peptide is cleaved from the resin and deprotected simultaneously or sequentially. The third and final step is purification by advance chromatographic methods and characterisation by spectroscopic methods.

2. In Sanger's method firstly the amino terminal of amino acid is labeled with flurodinitrobenzene (FDNB; commonly called Sanger's reagent) and then labeled peptide is subjected to acid (6N HCL) hydrolysis and separated by its chromatographic properties.

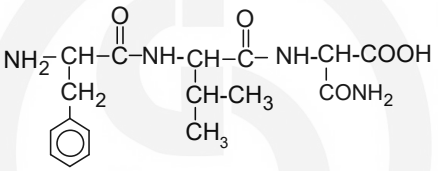
3. (a) Oxytocin- is a peptide hormone. It has an important role in birth and lactation as it helps in stimulation of milk ejection. It also causes uterine smooth muscle contraction at the time of parturition and helps in development of maternal behaviour towards the child soon after giving birth.

   (b) Insulin- is a peptide hormone, secreted by *islets of Langerhans* in the pancreas. Insulin controls the blood glucose level and that insulin is a very significant hormone used in the treatment of diabetes. It is one of the smallest protein molecule having just 51 amino acid residues on

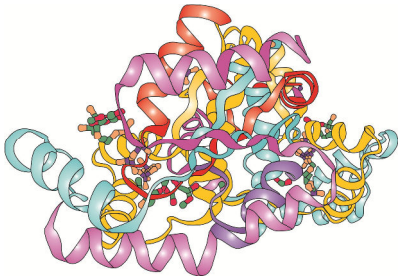
two different chains of 21(A-chain) and 30 (B-chain) residues linked through disulphide bonds.

(c) Glutathione- γ-Glutamylcysteinylglycine is a tripeptide found in nearly all cells of plants, animals and microorganisms. Glutathione is used to prevent oxidative stress in most cells and helps to trap free radicals that can damage DNA and RNA. There is a direct correlation with the speed of aging and the reduction of glutathione concentrations in intracellular fluids. It is known to play a critical role in the multiplication of lymphocytes (the cells that mediate specific immunity) which occurs in the development of an effective immune response.

4. Structure of peptide bond-The peptide bond chemically is an amide bond i.e., CONH which is planar in nature. All the atoms in bond formation lie on the same plane. The presence of double bond between carbon and oxygen which can move over to carbon and nitrogen makes the bond between C and N rigid. As a result, the atoms/groups attached to these take fixed position. Thus, the molecule exists in *cis* and *trans* isomeric forms. In other words, the molecule shows geometrical isomerism. Therefore, in a peptide bond (–CONH–) the hydrogen of the substituted amino group is almost always *trans* to the oxygen of carboxyl group.

5. 

6. Asp-Leu-Trp; DLW



UNIT 6

PROTEINS

Structure

6.1 INTRODUCTION

Proteins are the most abundant biomolecules in the human body and make up for 75% dry weight of our body. The term protein originates from the Greek word '$\pi\rho\sigma\tau\varepsilon\iota\sigma\sigma$' (*proteios*, meaning 'of first importance') and was coined by Dutch chemist G.T. Mulder in 1839. A number of constituents of our body are made up of proteins like skin, bone, tendon, muscle, tooth, nail, etc. Besides the structural components, these biomolecules are significant in performing a number of functions like providing structure to the body, transporting oxygen and other substances within an organism, regulating body chemistry etc. in the living systems. Proteins work as hormones, enzymes, storage proteins, transport proteins, structural proteins, immunoglobulins etc. Improper protein

functions cause many diseases such as lactosuria, thalassemia, sickle cell anemia etc. Therefore, it is important to understand their structural and functional relationships.

In the first two units of Block 2, you studied about the monomeric unit of proteins i.e., the amino acids and how two or more amino acids join together by peptide bonds to form a peptide. A linear polymer of different amino acids is called the polypeptide chain. Proteins contain one or more polypeptides. Thus, proteins are long chains of amino acids (50-2000) held together by peptide bonds. Naturally occurring polypeptides with molecular weight of more than 5000 are generally called **proteins**. 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.

In this unit you will first learn how proteins are classified based on their properties. You will also learn that the enormously large sized proteins have extremely complex structures and the structure of a protein molecule determines its biological role and other properties. In the later part of the unit you will learn and be able to appreciate the unique three-dimensional shape of proteins. The complete structure, including specific shape of a protein molecule is discussed under four types viz., primary, secondary, tertiary and quaternary structures. The function of the protein depends on its structure and any change in structure cause improper protein function and eventually a disease. So last but not the least the functional properties of proteins will be discussed.

Expected Learning Outcomes

After studying this unit, you should be able to:

- classify proteins on the basis of solubility, composition, shape and function;
- describe different properties of proteins;
- explain the structural organizational levels of proteins viz. primary, secondary, tertiary and quaternary; and
- list the biological functions of proteins.

6.2 CLASSIFICATION OF PROTEINS

You know that different proteins have different types and numbers of amino acids. In other words we can say that the structures of different proteins differ only in the composition and sequence of constituent amino acids. The functions and properties of proteins are related to their diversified structures. It is quite difficult to classify these macromolecules on the basis of one single property or characteristic. The classification of proteins is dependent on the following.

- Solubility
- Shape and structure

- Chemical composition
- Biological function

Let us study the types of proteins under each of the above factors.

6.2.1 Based on Solubility

The types of proteins can be classified based on their solubility in different solvents, such as water, salt solution and alcohol as follows. Some proteins are:

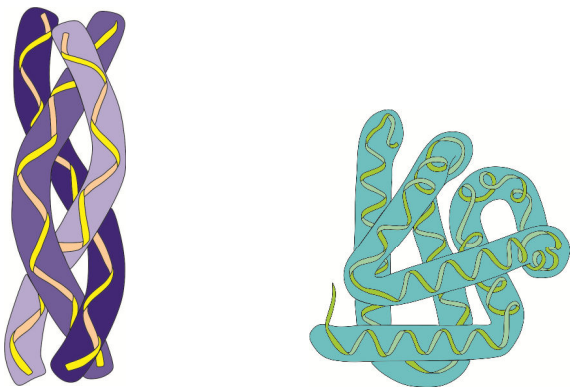
- soluble in water and salt solution e.g., albumin
- soluble in salt solution but sparingly soluble in water e.g., globulin
- soluble in 70-80% ethanol, are usually rich in proline amino acid e.g., protamines.
- nucleoproteins soluble in salt solution e.g., histones

6.2.2 Based on Shape

On the basis of their shape proteins can be divided into two classes:

- **Fibrous proteins**- Fibrous proteins are formed by polypeptide chains and appear as long thin filaments as can be seen in Fig. 6.1 (a). Their primary function is to provide support to the cells as well as the whole organism. Examples of fibrous proteins are keratin (in nails, claws), collagen (in bone, teeth, connective tissues of tendons and ligaments) and elastin, etc.
- **Globular proteins**- They appear as spherical globular structure therefore, also called as **spheroproteins** or **corpuscular** proteins and their structure is more complex than fibrous proteins as is given in Fig. 6.1(b). Most of the proteins belong to this class; all enzymes are globular proteins in nature. Examples of globular proteins are insulin, albumin, globulin, lysozymes, histones and protamines (the proteins of animal origin) and prolamines and glutelins found in the food grains.

In globular proteins the polypeptide chain can also fold back on itself, resulting in a change by 180^0 in the direction of the chain. These are known as **β-turns.**



(a) **(b)**

Fig. 6.1: Structures of: (a) fibrous and (b) globular proteins.

6.2.3 Based on Chemical Composition

On the basis of their chemical composition or the product of hydrolysis, proteins may be divided into following classes.

- **Simple proteins** or **homoproteins** are made up of only amino acids and do not contain any non-protein part. Examples are plasma albumins, collagens, protamines, prolamines, keratins, etc.
- **Conjugated proteins** or **heteroproteins** contain in their structure a non-protein part or group. The protein part of conjugated proteins is called **apoprotein** and the nonprotein part is referred to as the **prosthetic group**. The entire molecule is called a **haloprotein**. The proteins and their prosthetic groups are listed below in Table 6.1.

Table 6.1: Some proteins and their prosthetic groups

S. No.	Name of the protein	Prosthetic group
1.	Glycoproteins	carbohydrates
2.	Metalloproteins	metal ions
3.	Phosphoproteins	phosphoric acid
4.	Lipoproteins	lipids
5.	Nucleoproteins	nucleic acids
6.	Chromoproteins	coloured

- **Derived proteins** are the proteins derived by the action of different physical or chemical agents like, heat, acid, alkali or enzyme on the **native protein**. The derived proteins may be **primary** or **secondary** derived types. In the case of primary type, the protein gets denatured without getting hydrolysed. The secondary type undergoes a progressive hydrolysis by enzymes, dilute acids and alkalies.

Native protein is the protein in its folded state i.e. the 3D shape and is functional but remains unaltered by denaturating agents like, heat, chemicals or enzymes.

6.2.4 Based on Biological Functions

The multitude of functions that proteins perform is the consequence of both the folding of the polypeptide chain, and the presence of many different functional groups in the amino acid side chains. From the functional point of view, they may be divided into several groups.

The human beings have thousands of enzymes taking part in various biological reactions.

- **Structural proteins**- Proteins have a central role in the mechanical support and stabilization of many structures. Examples are α-keratins, myosin, collagen and elastin.
- **Enzymes** (biochemical catalysts)- In living organisms, almost all reactions are catalyzed by specific proteins called enzymes. Almost all the known enzymes are proteins except some catalytic RNA molecules called ribozymes, i.e., ribonucleic acid enzymes. Examples are pepsin, amylase, lipase, etc. You will study more about enzymes in Unit 7.
- **Regulatory proteins**- Many hormones are like insulin, glucagon, and thyroid-stimulating hormone (TSH) are proteins. These are regulatory molecules involved in the control of many cellular functions, from metabolism to reproduction.

- **Transport proteins**- Specific proteins classified as transport proteins, transport many organic and inorganic small molecules, in the bloodstream and extracellular fluids, across the cell membranes, and inside the cells from one compartment to another, Examples are: haemoglobin that carries oxygen, transferrin which carries iron in the blood, etc.
- **Storage proteins**- are proteins which help to store organic and inorganic molecules. For example, ferritin stores iron intracellularly in a non-toxic form in the liver, myoglobin stores oxygen in the skeletal muscles ovalbumin.
- **Protective or defense proteins**- Antibodies or immunoglobulins are glycoproteins that recognize antigens expressed on the surface of viruses, bacteria and other infectious agents. Thus these proteins defend the body against infection.

SAQ 1

Write true or false for the following statements.

(a) Albumin is soluble in water and salt solution.

(b) Collagen is an example of a fibrous protein.

(c) Non protein part of protein is called apoprotein.

(d) Immunoglobulins are storage proteins.

6.3 PROPERTIES OF PROTEINS

Being very large in size and colloidal in nature, protein molecules cannot pass through membranes of the cell and hence, presence of proteins in the urine warns us of possible damage to the kidney membranes.

We have mentioned that proteins are extremely large molecules with molecular weights in the range of 5000 to a few million Daltons (Da). In Fig. 6.2 we have attempted to show you the relative dimensions of various protein molecules in comparison to those of glucose molecule and to some inorganic ions, such as sodium and chloride ions. This large size of protein molecules is responsible for their colloidal properties.

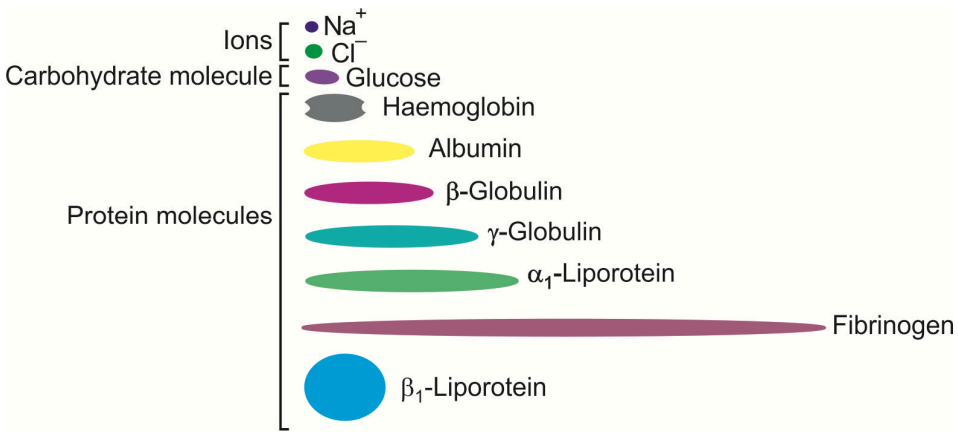


Fig. 6.2: Comparative sizes of various protein molecules in comparison to sodium and chloride ions and glucose molecule.

The properties of proteins are mainly due to the nature of constituent amino acids. However, the large polymeric protein molecules show some typical

physical and chemical behaviour. You will study some such properties in the following subsections. We would start with a brief on the molecular weight of these macromolecules as the high molecular weight is responsible for few of their properties.

6.3.1 Molecular Weight

You know that proteins are polypeptides usually containing between 50-25000 or more amino acids. The molecular weights of a large number of proteins are known. Average molecular weight of proteins ranges between 5 kDa (kilodalton) and 220 kDa. Proteins of very high molecular weights are found among hemocyanins, the copper-containing respiratory proteins of invertebrates. The molecular weight of proteins can be determined by molecular sieve chromatography (gel filtration), the sedimentation analysis, osmotic pressure, light scattering, x-ray diffraction, etc. However, the method most frequently used to determine the molecular weight of proteins is ultracentrifugation.

The molecular weight of proteins is expressed in the units of Daltons (Da)-a unit of mass defined to be equal to 1 a.m.u which is almost equal to the mass of a hydrogen atom.

6.3.2 Amphoteric Nature

The proteins have ionizable $-COO^-$ and $-NH_3^+$ groups on the surface because of the constitution of amino acids. Therefore, the proteins are able to donate or accept a proton and are amphoteric in nature. Besides, these can also adsorb anions and cations depending upon the pH of the solution. At alkaline pH, the negative charge predominates and the proteins move to the anode. At acidic pH, the positive charge predominates and the proteins move to the cathode. It is because of the zwitterionic structure that in a crystal lattice the protein molecules are held together tightly. As a result, proteins have high melting points. You may note here that the amphoteric nature of proteins facilitates their separation by gel electrophoresis and affects their solubility.

6.3.3 Solubility

The solubility of proteins depends on the amino acid residues present on the surfaces of proteins as these interact with solvents. Charged and polar surface residues interact readily with ionic groups in the polar solvents like water, ethanol etc. and increase the solubility of a protein whereas these remain insoluble in nonpolar solvents like benzene, hexane, ether etc. If there is a charge at the protein surface, the protein prefers to interact with water by the formation of hydrogen bonds, rather than with other protein molecules. This charge makes it more soluble. Without a net charge, protein-protein interactions and precipitation are more likely occur.

The solubility of proteins in aqueous buffers depends on the distribution of hydrophilic and hydrophobic amino acid residues on the protein surface. Proteins that have high hydrophobic amino acid content on the surface have low solubility in an aqueous solvent. At its isoelectric point, the solubility of protein is the lowest. The solubility of proteins in blood requires pH in the range of 7.35 to 7.45. It can be said that protein solubility is influenced by a number of extrinsic factors like pH, ionic strength, temperature, and the presence of various solvent additives and intrinsic factors like amino acids on the protein surface.

6.3.4 Precipitation

You might know that precipitation is the reverse of solubility so the factors responsible for low solubility or nonsolubility will lead to precipitation. Thus, if the charges on the protein surface are somehow neutralised, the protein molecules would come close, aggregate and result into precipitation. Proteins form colloidal solutions and proteins can be precipitated out from its solution by a variety of substances. Such reactions which precipitate out protein from their solution are called as **precipitation reactions**. Proteins can be precipitated either by removal of water layer (dehydration), denaturation, adjusting the isoelectric pH or by neutralization of charge present on protein molecule. Precipitation of proteins occurs because the change in pH or hydrophobicity alters interactions between the protein and the aqueous environment. Precipitation of proteins may also occur through binding of salts or metals to protein functional groups so that intramolecular interactions are disrupted causing the proteins to denature, aggregate, and fall out of solution. This property of precipitation is used in processing of biological products in order to concentrate proteins and purify them from various contaminants.

6.3.5 Denaturation

Denaturation refers to the unfolding of the polypeptide chain i.e. disruption or change in structure of protein. The process of denaturation does not change the primary structure of proteins i.e. the amino acid sequence remains the same Denaturation occurs due to break in the hydrogen and disulfide bonds of proteins and leads to an increase in the ionizable groups. It brings about change in the physical, chemical as well as biological properties of proteins. Heat, X-rays and UV radiation are well known physical denaturing agents. Other chemical agents include acids, alkalis, heavy metals (lead and mercury), urea and salicylate. Denaturation is generally an irreversible process, however, the native protein can be regenerated by gradually removing the denaturing agent and the process is called **renaturation**.

6.3.6 Colour Reactions

Proteins are made of amino acids and as all proteins are not made from the same set of amino acids therefore they respond differently to all colour reactions. These reactions act as the basis for qualitative detection of proteins. The most common colour reactions of proteins indicating the reagent, colour and the site of the reaction are summarised in Table 6.2.

Table 6.2: Some common colour reactions of proteins

Reaction/ test	Reagent	Colour	Reaction site	Remark
Biuret	Dil. $CuSO_4$ + NaOH	Blue-violet	Two or more than two peptide bonds	Red or blue-violet colour not given by amino acids or dipeptides
Millon's	Mercuric sulphate in H_2SO_4 and sodium nitrite	Red precipitate	Phenolic groups, tyrosine	Red colour given by phenols

Ninhydrin	Ninhydrin	Blue-violet absorbing at 540 nm	Amino group	Proline and hydroxyproline give yellow colour
Nitroprusside	Sodium nitroprusside in dil. NH_4OH	Red	Cysteine	Done to detect the presence of sulphur in proteins
Sakaguchi	Sodium hypochlorite and α-naphthol	Red	Arginine residue	Interference by compounds containing guanidine
Xanthroproteic	Boiling with HNO_3	Yellow or orange	Phenolic or indolic amino acids, e.g., phenylalanine, tyrosine	This reaction is responsible for yellowing of skin when HNO_3 falls on it.

SAQ 2

Fill in the blanks with appropriate words.

(a) The method most frequently used to determine the molecular weight of proteins is _____.

(b) Proteins are soluble in _____solvents.

(c) Denaturation of protein occurs due to break in the hydrogen and _____ bonds in the proteins

(d) Biuret test identifies _____ in the molecule of a protein.

6.4 STRUCTURAL ORGANISATION OF PROTEINS: PRIMARY STRUCTURE

A protein is constructed primarily as a linear polymer of different amino acids, called the polypeptide chain. We have already described the formation of a polypeptide in the previous unit. If the protein chain consists of say 100 amino acids and each amino acid happens 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 4.1 (Unit 4), 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 you have read in the previous section. 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 levels of organisation. These are the primary, secondary, tertiary and quaternary structures. We start with primary structure in this section.

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, Fig. 6.3. In short, the complete covalent structure of the polypeptide constitutes the primary structure.

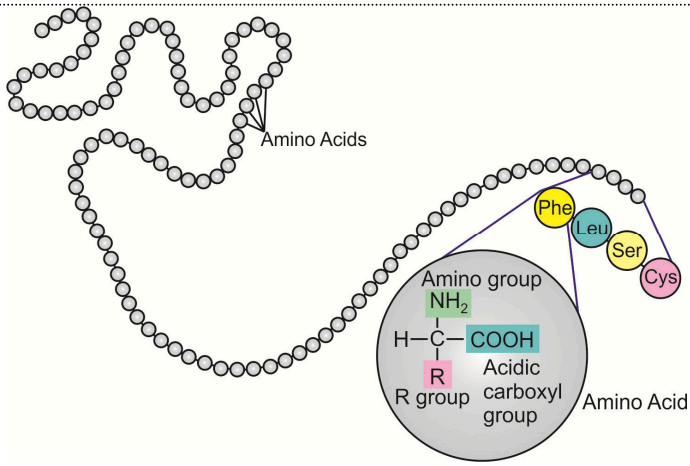
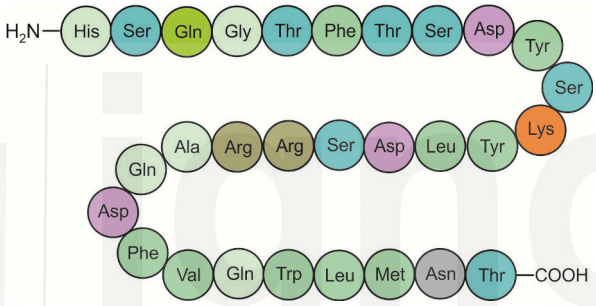


Fig. 6.3: Primary structure of a protein

For example, the primary structure of glucagon, a protein which stimulates the conversion of glycogen to glucose in the body, is shown in Fig. 6.4.



(HSQGTFTSDYSKYLDSRRAQDFVQWLMNT)

Fig. 6.4: Primary structure of glucagon

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 red blood cells 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.

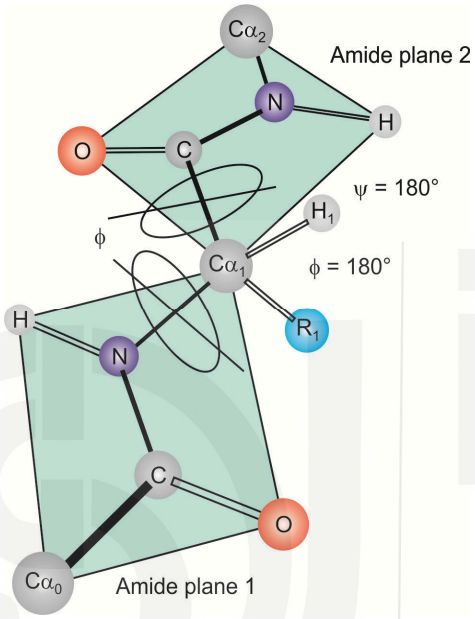
In essence, the primary structure of a protein corresponds to the structural formula of an organic compound. However, the structural formula of a small organic compound conveys enough information about its chemical function, whereas knowledge of the primary structure of a protein gives no such clue. Linear sequences of amino acids, no matter how different, are one dimensional and hence cannot explain the extraordinary 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.

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 and also disulphide bonds. We shall examine in the following subsections the general principles that govern the folding of the polypeptide chains of proteins into more complex structures.

6.4.1 Peptide Bond and Protein Folding

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 and its planar nature in subsection 5.2.2 of Unit 5. Let us now learn what influence this bond has on protein folding.

The individual bonds constituting the peptide unit are torsionally rigid that make intrabond hydrogen bonding impossible. However, twisting of the peptide plane is still possible around the Cα - N and Cα - C bond axis, as shown in Fig. 6.5.



ϕ = φ = 180° corresponds to the fully extended polypeptide chain.

Fig. 6.5: 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.

The angles ϕ and φ 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 ϕand φ 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 φ bring unbounded atoms too close to each other, causing steric repulsions, and are consequently "not allowed". It is easy to appreciate that this would substantially restrict the number of conformations in a polypeptide chain.

6.4.2 Disulphide Bonds

We shall now describe the other type of 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 in Fig. 6.6.

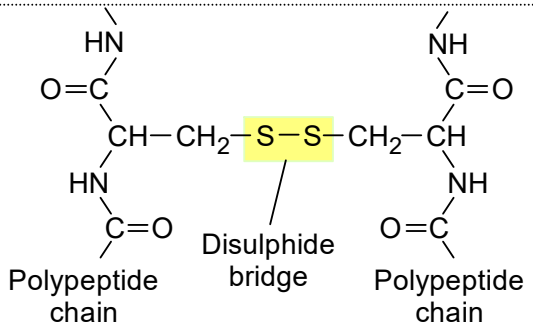


Fig. 6.6: Cysteine residues holding two polypeptide chains

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 depicted in Fig. 6.7.

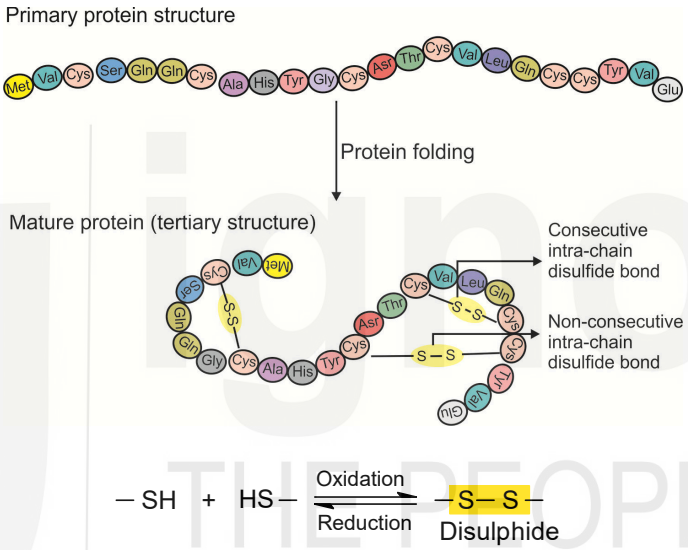


Fig. 6.7: A disulphide bond formed by two cysteine residues in a protein molecule.

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. 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 that are also important in stabilising the protein structures.

6.4.3 Weak Noncovalent Interactions in Protein Folding

As we have already discussed, the primary structure of a polypeptide chain is maintained by covalent amide bonds. 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 the next subsection) 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 not only provide reasonable stability to higher order protein structures,

but also confer on these a degree of flexibility, which is very essential for biological function.

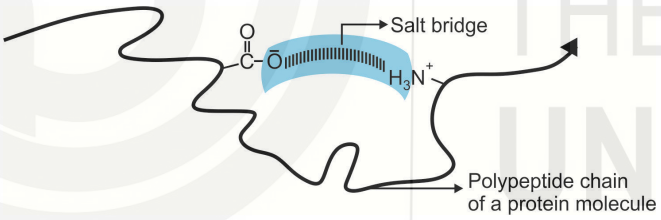
The weak noncovalent interactions that contribute to the stability of protein structures have been identified as **electrostatic forces**, **van der Waals forces**, **hydrogen bonds** and **hydrophobic interactions**. We shall describe these forces briefly. You will find an estimate of the energies involved in these interactions and the covalent disulphide bonds in Table 6.3.

Table 6.3: Energies of noncovalent Interactions and 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 $–NH_3^+$ or $-COO^-$ 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.

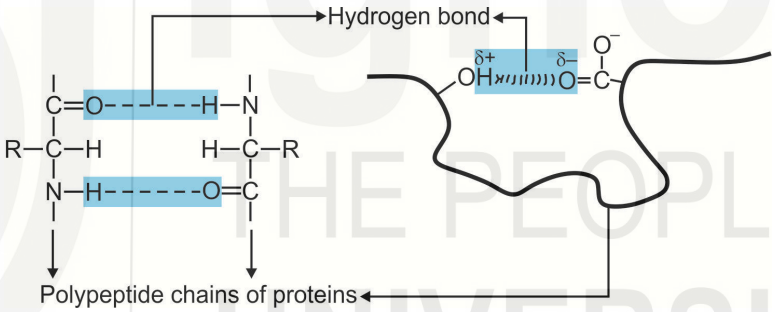
Another weak interaction involved in protein folding is **van der Waals dispersion forces**. This is due to the attractive force between uncharged molecules and comes into effect when two atoms approach each other so closely that their electron clouds penetrate each other. Though repulsive forces also 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 surface, is considerable and would be of great significance in the maintenance of the folded state of proteins.

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

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

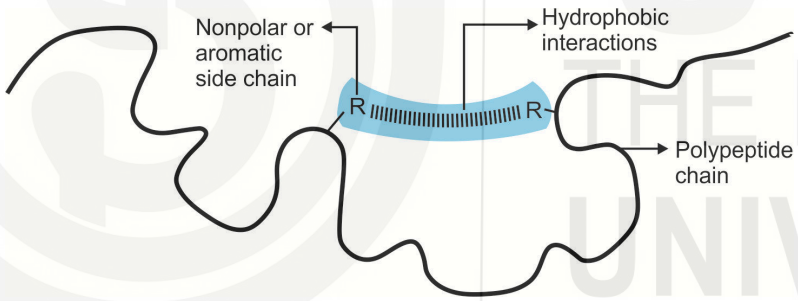


The importance of hydrogen bonds in the stability of protein structures is, however, debatable. This is because water, which is the universal biological medium, forms hydrogen bonds with other water molecules, as well as with the appropriate groups on the polypeptide chain. Thus before the folding process 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.

You might very well know that hydrophobic means repelling or not interacting water. You will study how and what does it lead to. 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 3

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 [ ]

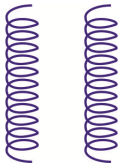
6.5 SECONDARY STRUCTURE

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 5.2.2 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 carboxyl 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**. These two arrangements in the secondary structure were proposed on the basis of X-ray diffraction studies on small synthetic polypeptides and model building. These were later on found to be present in various protein molecules.

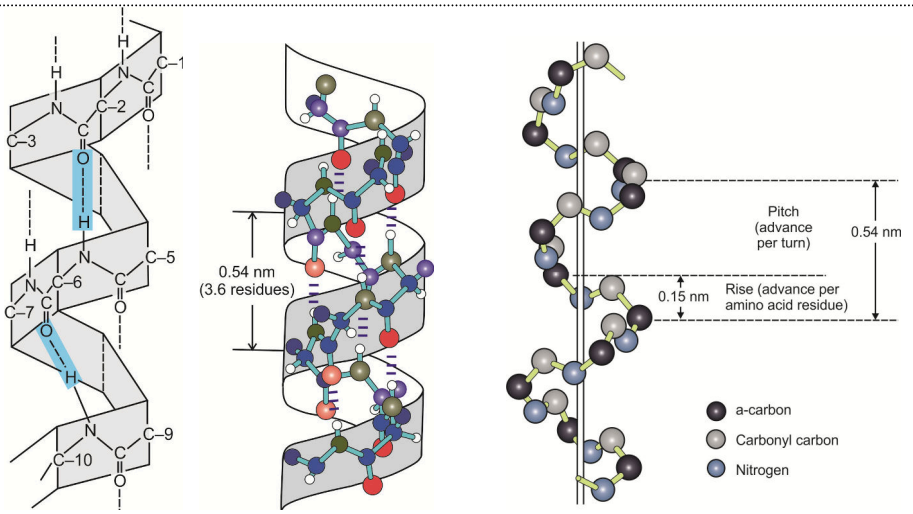
We shall now explain what an α-helix and a β-pleated sheet signify.

6.5.1 The α-Helix

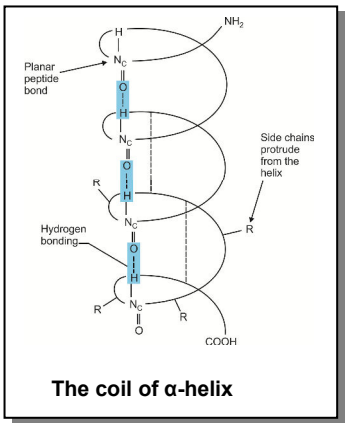
A coiled spring is helical in structure.



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 an α–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 carboxyl group of one amino acid and the -NH group, four amino acids ahead in the chain as shown in Fig. 6.8 (a, b and c).



(a) (b) (c)



The coil of α-helix

Fig 6.8: The right-handed α-helix a) the peptide groups as planes with α-carbon atoms occupying junctions of successive planes; b) ball and stick model of the helix indicating H-bonds; c) the pitch of the α-helix

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. The **pitch** of the helix or the distance between two consecutive turns is 540 pm (or 0.54 nm) as shown in Fig. 6.5(c). 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. Three polypeptide chains in left-handed helical conformation are twisted together in a right-handed fashion (Fig 6.9). 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 (Table 4.1) 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 carboxyl group.

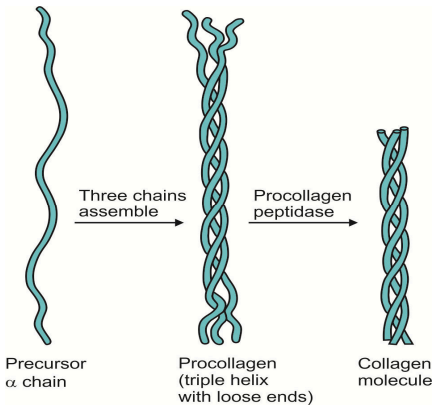


Fig 6.9: Triple helix of collagen

The pitch in the helix of collagen chains is about 900 pm. Lack of hydrogen bonding explains why the successive turns are not closely held.

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

6.5.2 β-Pleated Sheets

The α-helix was the first secondary structure discovered. That is why β-pleated sheet structure is known as β.

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 as shown in Fig.6.10.

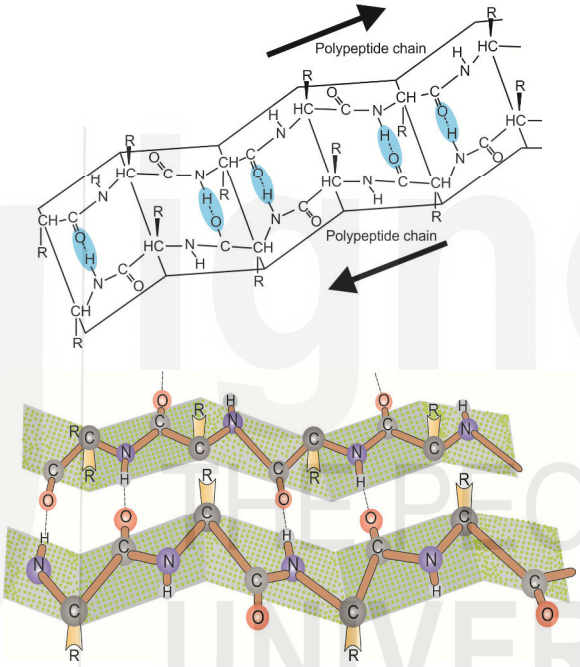
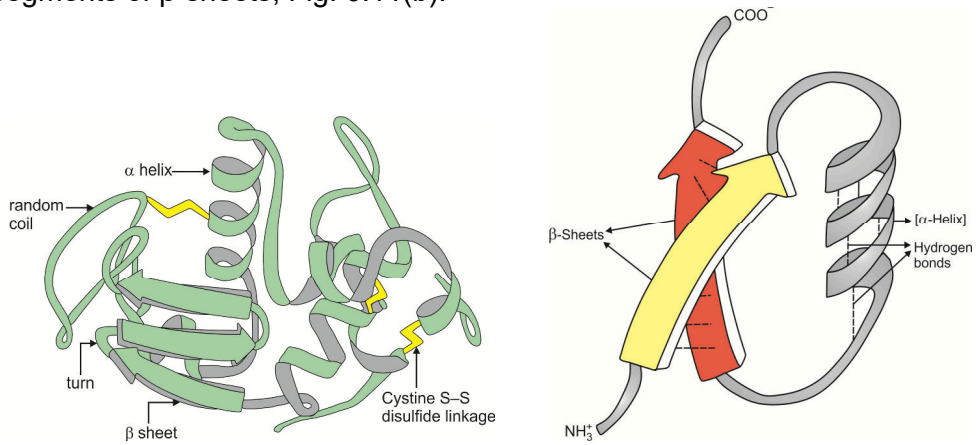


Fig. 6.10: The antiparallel chains of β-pleated sheet. Note that the R-groups are directed perpendicularly to the plane of the sheet.

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

The antiparallel sheet is more stable because hydrogen bonds in it are more linearly oriented. In the β-structures several polypeptide chains participate in the formation of sheets, which in turn are 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, Fig. 6.11 (a). 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 φ angles of the amino acid residues in this region are also limited by the same considerations as described earlier. The term "random coil" denotes that the conformation of the polypeptide chain in this region does not fit in with any of the regular and repetitive patterns described above. 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, Fig. 6.11(b).



(a) **(b)**

Fig. 6.11: (a) Random coil of a protein molecule; (b) The β-α-β folding pattern of a protein.

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 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 stack 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 that are the 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, 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 6.4.2-3. You can learn about the tertiary structure of the protein molecules after attempting the following SAQ.

SAQ 4

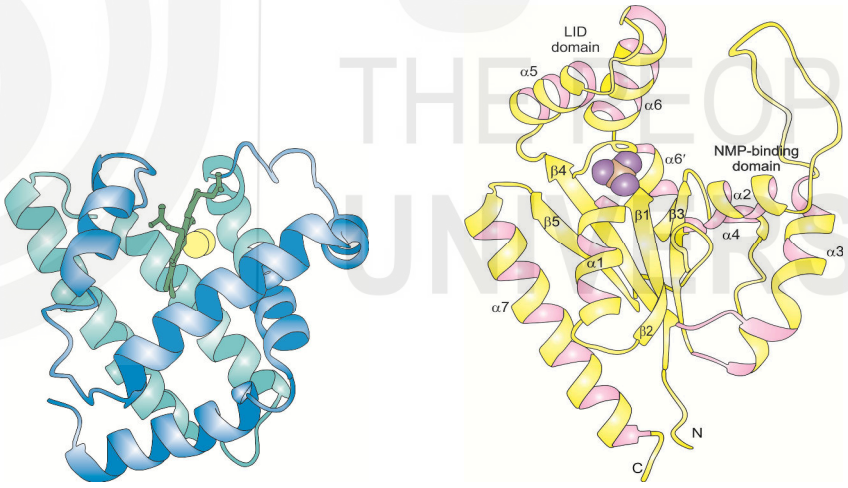
Tick [√] mark the correct answer.

β-sheets are present in pure form in:

a) silk fibroin [ ]
b) α-keratin [ ]
c) collagen [ ]
d) none of the above [ ]

6.6 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. However 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. The three-dimensional tertiary structure of two typical proteins – myoglobin and adenylate kinase is given in Fig. 6.12.



(a) (b)

Fig. 6.12: (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.

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 amino acid residue and – NH^{+}_{3} group of another. Disulphide bonds also

provide further stability, as in some enzymes. In Fig. 6.13 a schematic representation of a globular protein is given representing the various types of forces which are involved in stabilising its tertiary structure. 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.

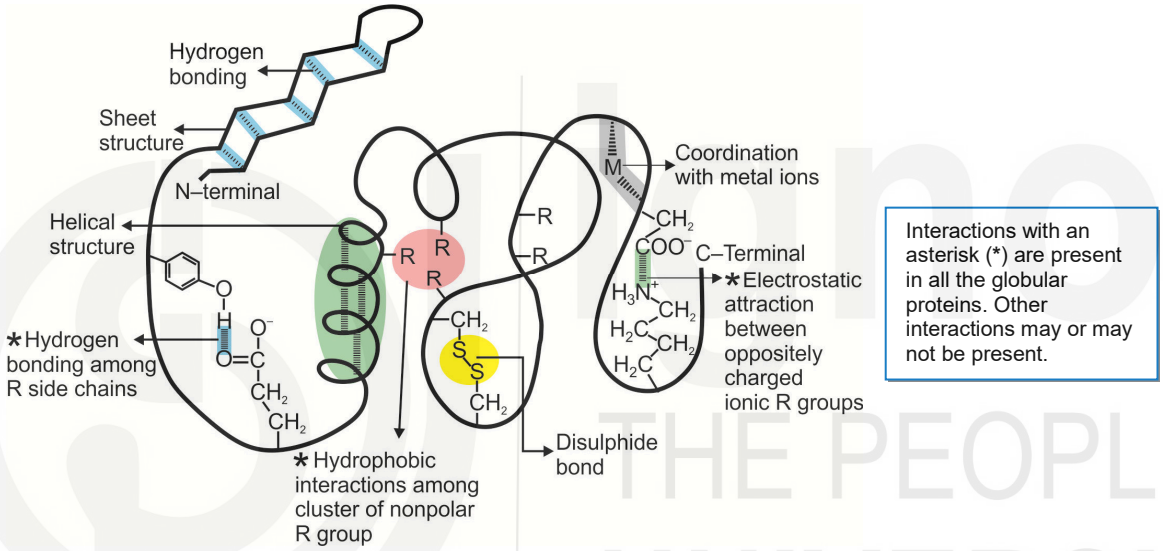


Fig. 6.13: Schematic representation of bonds and forces that stabilise globular structure in proteins.

We may add here that in some proteins which require metal ions for their function, coordination with negativity 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. 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 (GAPDH) 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. 6.14.

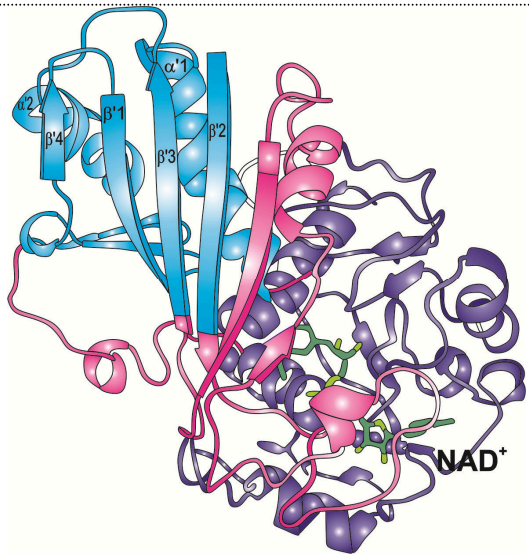


Fig. 6.14: Domain structures in a subunit of the enzyme glyceraldehyde 3-phosphate dehydrogenase. The NAD^+-binding domain of GAPDH is shown in white, the catalytic domain is shown in yellow and orange.

Quite often polypeptide chains of the same type or different types, each of those 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 study this highest level of protein structure but after answering the following SAQ.

SAQ 5

Tick [√] mark the correct statement.

Globular proteins always contain,

a) some α-helices [ ]

b) some β-helices [ ]

c) some random coil regions [ ]

d) α-helices and β-sheets [ ]

6.7 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 the 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. 6.12 (a). 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.

The oxygen binding ability of haemoglobin is also due to the haem group. 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. The haemoglobin molecule is represented in Fig. 6.15.

Haemoglobin is present in red blood cells where it binds oxygen and then transports it to tissues throughout the body.

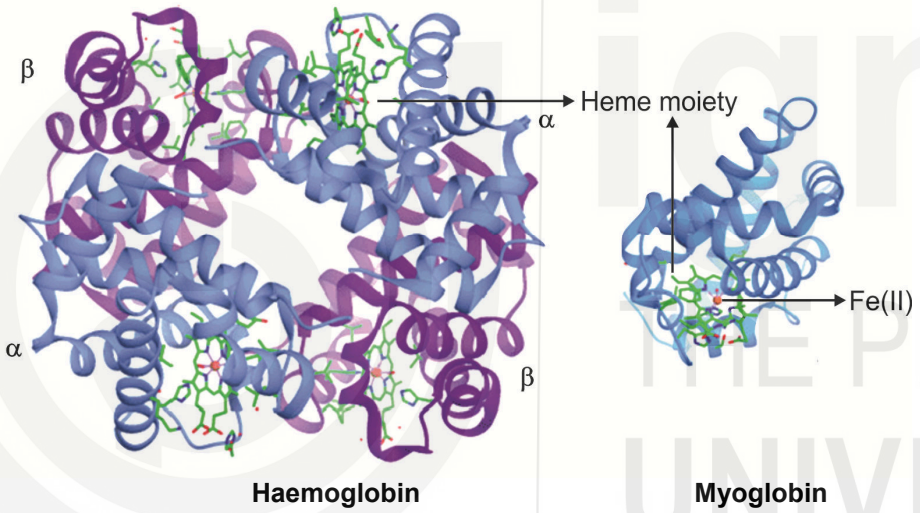


Fig. 6.15: Comparison of the structures of myoglobin and haemoglobin. Observe the positions of the haem (Fe) groups and the close similarity in the folding pattern of the myoglobin chain and that of the β chain of haemoglobin

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 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 so 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.

6.7.1 Haemoglobin as a Transport Protein

You will recall from subsection 6.4.6, 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. 6.16.

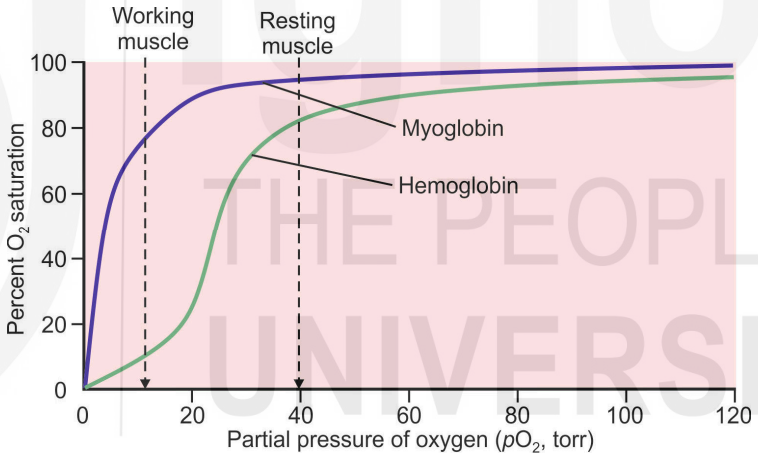


Fig. 6.16: Diagram showing oxygen binding curves of myoglobin and haemoglobin.

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 low, it can easily release its bound oxygen, thus fulfilling its role as an oxygen transporter. From the curve in Fig. 6.16, 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 group is 400 times stronger than binding to the first haem group.

After attempting the following SAQ, you will study the biological functions of proteins in the next section.

SAQ 6

Tick [√] mark the correct option.

The oxygen binding curve for haemoglobin is:

a) linear [ ]

b) simple [ ]

c) sigmoidal [ ]

d) none of the above [ ]

6.8 BIOLOGICAL FUNCTIONS OF PROTEINS

You have already learnt about the general biological functions of proteins in the section that dealt with the classification of proteins based on biological functions (Subsec. 6.2.4). You studied that proteins have diverse functions viz., providing structural support, metabolism of nutrients, transport of nutrients, storage of nutrients, providing protection from pathogens, acting as enzymes, carriers, or hormones. Further you know that shapes and functions of proteins are intricately linked. So, any change in shape caused by changes in temperature or pH leads to protein denaturation and a loss in its function. The types and functions of proteins are listed in Table 6.4.

Table 6.4: Types and functions of proteins

Types	Function	Examples
Enzymes	Catalysts in biochemical reactions (like digestion); is specific for the substrate; help in breakdown, rearrangement or synthesis reactions.	Catalase, NADPH oxidase, digestive enzymes-Amylase, lipase, pepsin, trypsin
Transport proteins	Carry substances in the blood or lymph throughout the body	Hemoglobin, albumin
Hormones	Chemical-signaling molecules of the body and are secreted by endocrine cells that act to control or regulate specific physiological processes including growth, development, metabolism, and reproduction. Hormones are either small proteins or steroids.	Insulin, glucagon, thyroxine
Contractile protein	Effect muscle contraction	Actin, myosin
Storage protein	Provide nourishment in early development of the embryo and the seedling	Legume storage proteins, egg white (albumin)
Defense proteins	Protect the body from foreign pathogens	Immunoglobulins
Structural proteins	Construct different structures and give shape to cell like the cytoskeleton	Actin, tubulin, keratin

SAQ 7

Match the following two columns:

(a) Immunoglobulin	(i) hormone
(b) Haemoglobin	(ii) protection
(c) Insulin	(iii) shape
(d) Actin	(iv) transport
(e) Cytoskeleton	(v) contraction

6.9 SUMMARY

- Polypeptides with molecular weight of more than 5000 are generally called **proteins**. The functions and properties of proteins are related to their diversified structure. Proteins are classified on the basis of their solubility, shape and structure, chemical composition and biological function.
- Proteins classified based on solubility are Albumin, globulin, protamines and histones. On the basis of their shape proteins can be divided into fibrous proteins and globular proteins. On the basis of their chemical composition proteins may be divided into simple proteins, derived proteins and conjugated proteins.
- The multitude of functions that proteins perform is the consequence of both the folding of the polypeptide chain, and the presence of many different functional groups in the amino acid side chains. They can be classified on the basis of function into structural protein, enzymes, hormones, storage proteins, defence proteins, transport proteins.
- The properties of proteins are mainly due to the nature of constituent amino acids. They have between 50-2000 amino acid. Average molecular weight of proteins ranges between 5KDa and 220KDa. The protein is amphoteric and has $-COO^-$ and $-NH3^+$ groups on its surface. Proteins are easily soluble in polar solvents like water, ethanol etc. but insoluble in polar solvents like benzene, hexane, ether etc.
- Proteins form colloidal solutions can be precipitated out from solution by a variety of substances. This property of precipation is used in processing of biological products in order to concentrate proteins and purify them from various contaminants. Denaturation of proteins occurs due to break in the hydrogen and disulfide bonds. Biuret and Xanthoproteic tests are used for detection of proteins based on its colour reactions.
- Proteins are organized at four levels: primary, secondary, tertiary, and quaternary. The primary structure is the unique sequence of amino acids. The local folding of the polypeptide to form structures such as the α helix and β-pleated sheet constitutes the secondary structure. The overall three-dimensional structure is the tertiary structure. When two or more

polypeptides combine to form the complete protein structure, the configuration is known as the quaternary structure of a protein.

- Noncovalent Interactions (Electrostatic interactions, Hydrogen bond, van der Waals dispersion interactions, Hydrophobic interactions and disulphide bond help in determining the structure of protein.
- Haemoglobin binds and transports oxygen in the whole body. This transport function of haemoglobin molecule is strongly dependent on its quaternary structure.
- Protein have diverse functions: providing structural support, metabolism of nutrients, transport of nutrients, storage of nutrients, providing protection from pathogens; and by acting as enzymes, carriers, or hormones.

6.10 TERMINAL QUESTIONS

1. Classify proteins on the basis of their chemical composition.
2. Discuss the role of non covalent interactions in protein folding.
3. What is meant by denaturation of proteins?
4. Describe in brief the structure and function of hemoglobin:
5. Describe in brief the quaternary structure of proteins.
6. Name the types of secondary structures in proteins and describe in brief their constitutions.
7. List the biological functions of proteins.

6.11 ANSWERS

Self-Assessment Questions

1. (a) True (b) True (c) False (d) False
2. (a) ultracentrifugation (b) polar (c) disulphide (d) peptide bond
3. (c) 4. (a) 5. (c) 6. (c)
7. (a) ii) (b) iv) (c) i) (d) v) (e) iii)

Terminal Questions

1. Proteins are classified on basis of their chemical composition into simple, conjugated and derived proteins. Simple proteins or homoproteins are made up of only amino acids and do not contain any non-protein part. Examples are plasma albumins, globulins, collagens, protamines, prolamines, keratins, etc. Conjugated proteins or heteroproteins contain a non-protein part or group in their structure. The protein part of conjugated proteins is called apoprotein and the nonprotein part is referred to as the prosthetic group. The entire molecule is called a haloprotein.

Derived proteins are the proteins derived by the action of different physical or chemical agents like, heat, acid, alkali or enzyme on the native protein. The derived proteins may be primary or secondary derived types. In the case of primary type, the protein gets denatured without getting hydrolysed. The secondary type undergoes a progressive hydrolysis by enzymes, dilute acids and alkalies.

2. The folding of a polypeptide chain to form secondary, tertiary and quaternary structures is due to the weak noncovalent interactions. 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 (Refer 6.4.4).

3. Denaturation of protein refers to the unfolding of the polypeptide chain *i.e.* disruption or change in t6he native structure of protein. Denaturation occurs due to break in the hydrogen and disulphide bonds of proteins and leads to an increase in the ionizable groups. Heat, X-rays and UV radiation and chemical agents like acids, alkalis, heavy metals (lead and mercury), urea and salicylate are well known denaturing agents. Denaturation is generally an irreversible process, however, the native protein can be regenerated by gradually removing the denaturing agent and the process is called renaturation.

4. Haemoglobin is present in red blood cells where it binds oxygen and then transports it to tissues throughout the body. It releases the bound oxygen to tissues where oxygen is low. This molecule has four polypeptide α-chains, with 141 residues each, and the β-chains with 146 residues. Each chain has a haem group. 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.

5. The quaternary structure of a protein is the association of several protein chains or subunits into a closely packed arrangement. Each of the subunits has its own primary, secondary, and tertiary structure. The subunits are held together by hydrogen bonds and van der Waals forces between nonpolar side chains. The subunits in a quaternary structure is arranged for the entire protein to function properly. Any alteration in the structure of the subunits or how they are associated causes marked changes in biological activity.

6. Secondary structure of proteins is arrangements of adjacent amino acid residues in a polypeptide chain. It is maintained by hydrogen bonds between amide hydrogens and carbonyl oxygens of the peptide backbone. The major secondary structures are α-helices and β-structures.

   In the α-helical structure, the polypeptide backbone is tightly wound around the long axis of the molecule, and R groups of the amino acid residues protrude outward from the helical backbone. The hydrogen

bonds, which stabilize the helix, are nearly parallel to the long axis of the helix. The β-sheets are stabilized by hydrogen bonds between carbonyl oxygens and amide hydrogens on adjacent β-strands. In β-sheets, the hydrogen bonds are nearly perpendicular to the extended polypeptide chains. The β-strands may be either parallel (running in the same N- to C-terminal) or antiparrallel (running in opposite N- to C- terminal directions).

7. Proteins perform various functions which depends on the presence of the different functional groups in the amino acid side chains. Structural proteins have a central role in the mechanical support and stabilization of many structures. Examples are α-keratins, myosin, collagen and elastin. Enzymes (biochemical catalysts) catalyse almost all reactions. Examples are pepsin, amylase, lipase, etc. Many hormones are like insulin, glucagon, and thyroid-stimulating hormone (TSH) are proteins. These are regulatory molecules involved in the control of many cellular functions, from metabolism to reproduction. Specific proteins classified as transport proteins, transport many organic and inorganic small molecules, in the bloodstream and extracellular fluids, across the cell membranes, and inside the cells from one compartment to another, Examples are: haemoglobin that carries oxygen, transferrin which carries iron in the blood, etc. Storage proteins to store organic and inorganic molecules. For example, ferritin stores iron intracellularly in a non-toxic form in the liver, myoglobin stores oxygen in the skeletal muscles. Antibodies or immunoglobulins are glycoproteins that recognize antigens expressed on the surface of viruses, bacteria and other infectious agents. Thus these proteins defend the body against infection.

ignou
THE PEOPLE'S
UNIVERSITY