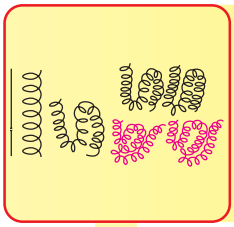
Chapter

4 Proteins and Amino Acids



The proteins speak :

"We are the basis of structure and function of life;
Composed of twenty amino acids, the building blocks;
Organized into primary, secondary, tertiary and quaternary structure;
Classified as simple, conjugated and derived proteins."

Proteins are the ***most abundant organic molecules of the living system***. They occur in every part of the cell and constitute about 50% of the cellular dry weight. Proteins form the fundamental basis of structure and function of life.

Origin of the word 'protein'

The term **protein** is derived from a *Greek* word ***proteios***, meaning *holding the first place*. Berzelius (Swedish chemist) suggested the name proteins to the group of organic compounds that are utmost important to life. Mulder (Dutch chemist) in 1838 used the term **proteins** for the high molecular weight nitrogen-rich and most abundant substances present in animals and plants.

Functions of proteins

Proteins perform a great variety of specialized and essential functions in the living cells. These functions may be broadly grouped as ***static*** (***structural***) and ***dynamic***.

Structural functions : Certain proteins perform *brick and mortar* roles and are primarily responsible for structure and strength of body. These include ***collagen*** and ***elastin*** found in bone matrix, vascular system and other organs and ***α-keratin*** present in epidermal tissues.

Dynamic functions : The dynamic functions of proteins are more diversified in nature. These include proteins acting as ***enzymes, hormones, blood clotting factors, immunoglobulins***, membrane receptors, storage proteins, besides their function in genetic control, muscle contraction, respiration etc. Proteins performing dynamic functions are appropriately regarded as ***the working horses*** of cell.

Elemental composition of proteins

Proteins are predominantly constituted by five major elements in the following proportion.

Carbon	:	50 – 55%
Hydrogen	:	6 – 7.3%
Oxygen	:	19 – 24%
Nitrogen	:	13 – 19%
Sulfur	:	0 – 4%

Besides the above, proteins may also contain other elements such as P, Fe, Cu, I, Mg, Mn, Zn etc.

The content of ***nitrogen,*** an essential component of proteins, on an average is ***16%.*** Estimation of nitrogen in the laboratory (mostly by ***Kjeldahl's method***) is also used to find out the amount of protein in biological fluids and foods.

Proteins are polymers of amino acids

Proteins on complete hydrolysis (with concentrated HCl for several hours) yield L-α-amino acids. This is a common property of all the proteins. Therefore, ***proteins are the polymers of L-α-amino acids.***

STANDARD AMINO ACIDS

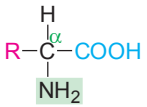
As many as 300 amino acids occur in nature—Of these, ***only 20***—known as standard amino acids are repeatedly ***found in the structure of proteins***, isolated from different forms of life—animal, plant and microbial. This is because of the universal nature of the genetic code available for the incorporation of only 20 amino acids when the proteins are synthesized in the cells. The process in turn is controlled by DNA, the genetic material of the cell. After the synthesis of proteins, some of the incorporated amino acids undergo modifications to form their derivatives.

AMINO ACIDS

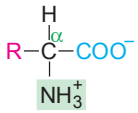
Amino acids are a group of organic compounds containing two ***functional groups—amino*** and ***carboxyl***. The amino group (—NH_2) is basic while the carboxyl group (—COOH) is acidic in nature.

General structure of amino acids

The amino acids are termed as α-amino acids, if both the carboxyl and amino groups are attached to the same carbon atom, as depicted below



General structure



Exists as ion

The α-carbon atom binds to a side chain represented by R which is different for each of the 20 amino acids found in proteins. The amino acids mostly exist in the ionized form in the biological system (shown above).

Optical isomers of amino acids

If a carbon atom is attached to four different groups, it is ***asymmetric*** and therefore exhibits ***optical isomerism***. The amino acids (except glycine) possess four distinct groups (R, H, COO^-, NH_3^+) held by α-carbon. Thus all the amino acids (except glycine where R = H) have optical isomers.

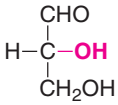
The structures of L- and D-amino acids are written based on the configuration of L- and D-glyceraldehyde as shown in ***Fig.4.1***. The proteins are composed of L-α-amino acids.

Classification of amino acids

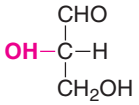
There are different ways of classifying the amino acids based on the structure and chemical nature, nutritional requirement, metabolic fate etc.

A. Amino acid classification based on the structure : A comprehensive classification of amino acids is based on their structure and chemical nature. Each amino acid is assigned a 3 letter or 1 letter symbol. These symbols are commonly used to represent the amino acids in protein structure. The 20 amino acids found in proteins are divided into seven distinct groups.

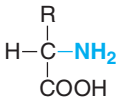
In ***Table 4.1***, the different groups of amino acids, their symbols and structures are given. The salient features of different groups are described next



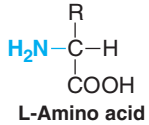
D-Glyceraldehyde



L-Glyceraldehyde



D-Amino acid



L-Amino acid

Fig. 4.1 : D- and L-forms of amino acid based on the structure of glyceraldehyde.

TABLE 4.1 Structural classification of L-α-amino acids found in proteins

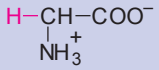
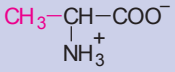
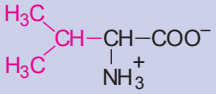
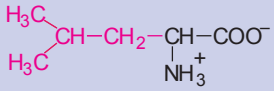
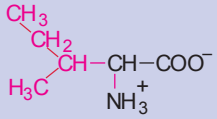
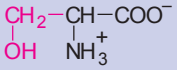
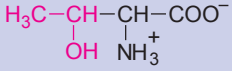
Name	Symbol		Structure	Special group present
	3 letters	*1 letter*		
I. Amino acids with aliphatic side chains				
1. Glycine	Gly	G		
2. Alanine	Ala	A		
3. Valine	Val	V		Branched chain
4. Leucine	Leu	L		Branched chain
5. Isoleucine	Ile	I		Branched chain
II. Amino acids containing hydroxyl (—OH) groups				
6. Serine	Ser	S		Hydroxyl
7. Threonine	Thr	T		Hydroxyl
Tyrosine	Tyr	Y	See under aromatic	Hydroxyl

Table 4.1 contd. next page

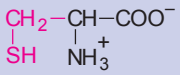
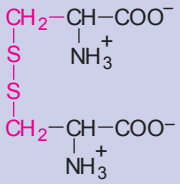
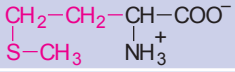
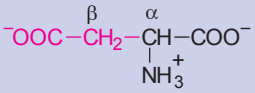
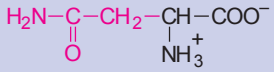
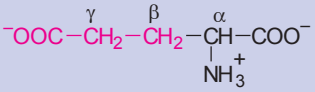
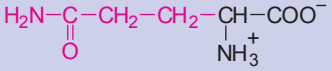
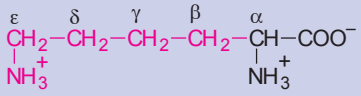
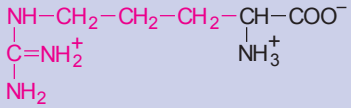
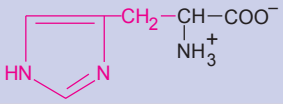
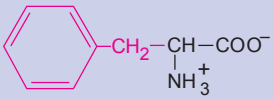
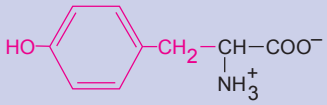
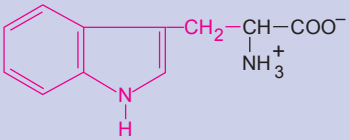
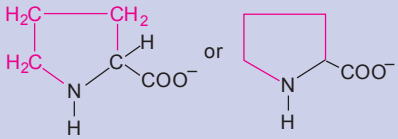
Name	Symbol		Structure	Special group present
	3 letters	1 letter		
III. Sulfur containing amino acids				
8. Cysteine	Cys	C		Sulfhydryl
Cystine	—	—		Disulfide
9. Methionine	Met	M		Thioether
IV. Acidic amino acids and their amides				
10. Aspartic acid	Asp	D		β-Carboxyl
11. Asparagine	Asn	N		Amide
12. Glutamic acid	Glu	E		γ-Carboxyl
13. Glutamine	Gln	Q		Amide
V. Basic amino acids				
14. Lysine	Lys	K		ε-Amino
15. Arginine	Arg	R		Guanidino
16. Histidine	His	H		Imidazole

Table 4.1 contd. next page

Name	Symbol		Structure	Special group present
	3 letters	1 letter		
VI. Aromatic amino acids				
17. Phenylalanine	Phe	F		Benzene or phenyl
18. Tyrosine	Tyr	Y		Phenol
19. Tryptophan	Trp	W		Indole
VII. Imino acid				
20. Proline	Pro	P		Pyrrolidine

*(**Note** : R group is shown in red)*

1. **Amino acids with aliphatic side chains :** These are monoamino monocarboxylic acids. This group consists of the most simple amino acids—glycine, alanine, valine, leucine and isoleucine. The last three amino acids (Leu, Ile, Val) contain branched aliphatic side chains, hence they are referred to as ***branched chain amino acids.***
2. **Hydroxyl group containing amino acids :** Serine, threonine and tyrosine are hydroxyl group containing amino acids. Tyrosine—being aromatic in nature—is usually considered under aromatic amino acids.
3. **Sulfur containing amino acids :** Cysteine with sulfhydryl group and methionine with thioether group are the two amino acids incorporated during the course of protein synthesis. Cystine, another importa nt sulfur containing amino acid, is formed by condensation of two molecules of cysteine.
4. **Acidic amino acids and their amides :** Aspartic acid and glutamic acids are ***dicarboxylic monoamino acids*** while asparagine and glutamine are their respective amide derivatives. All these four amino acids possess distinct codons for their incorporation into proteins.
5. **Basic amino acids :** The three amino acids lysine, arginine (with guanidino group) and histidine (with imidazole ring) are dibasic monocarboxylic acids. They are highly basic in character.
6. **Aromatic amino acids :** Phenylalanine, tyrosine and tryptophan (with indole ring)

are aromatic amino acids. Besides these, histidine may also be considered under this category.

7. **Imino acids :** Proline containing pyrrolidine ring is a unique amino acid. It has an imino group (=NH), instead of an amino group ($-NH_2$) found in other amino acids. Therefore, proline is an α-imino acid.

Heterocyclic amino acids : Histidine, tryptophan and proline.

B. Classification of amino acids based on polarity : Amino acids are classified into 4 groups based on their polarity. Polarity is important for protein structure.

1. **Non-polar amino acids :** These amino acids are also referred to as hydrophobic (water hating). They have no charge on the 'R' group. The amino acids included in this group are — alanine, leucine, isoleucine, valine, methionine, phenyl-alanine, tryptophan and proline.
2. **Polar amino acids with no charge on 'R' group :** These amino acids, as such, carry no charge on the 'R' group. They however possess groups such as hydroxyl, sulfhydryl and amide and participate in hydrogen bonding of protein structure. The simple amino acid glycine (where R = H) is also considered in this category. The amino acids in this group are—glycine, serine, threonine, cysteine, glutamine, asparagine and tyrosine.
3. **Polar amino acids with positive 'R' group :** The three amino acids lysine, arginine and histidine are included in this group.
4. **Polar amino acids with negative 'R' group :** The dicarboxylic monoamino acids—aspartic acid and glutamic acid are considered in this group.

C. Nutritional classification of amino acids : The 20 amino acids (***Table 4.1***) are required for the synthesis of variety proteins, besides other biological functions. However, all these 20 amino acids need not be taken in the diet. Based on the nutritional requirements, amino acids are grouped into two classes—essential and non-essential.

1. **Essential or indispensable amino acids :** The amino acids which ***cannot be synthesized by the body*** and, therefore, need to be supplied through the diet are called essential amino acids. They are required for proper growth and maintenance of the individual. The ten amino acids listed below are essential for humans (and also rats) :

   Arginine, **V**aline, **H**istidine, **I**soleucine, **L**eucine, **L**ysine, **M**ethionine, **P**henyla-lanine, **T**hreonine, **T**ryptophan.

   [The code ***A.V. HILL, MP., T. T.*** (first letter of each amino acid) may be memorized to recall essential amino acids. Other useful codes are H. VITTAL, LMP; PH. VILLMA, TT, PVT TIM HALL and MATTVILPhLy.]

   The two amino acids namely ***a***rginine and ***h***istidine can be synthesized by adults and not by growing children, hence these are considered as ***semi–essential amino acids*** (remember ***Ah***, to recall). Thus, 8 amino acids are absolutely essential while 2 are semi-essential.
2. **Non-essential or dispensable amino acids :** The body can synthesize about 10 amino acids to meet the biological needs, hence they need not be consumed in the diet. These are—glycine, alanine, serine, cysteine, aspartate, asparagine, glutamate, glutamine, tyrosine and proline.

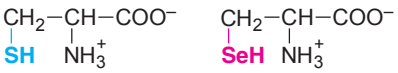
D. Amino acid classification based on their metabolic fate : The carbon skeleton of amino acids can serve as a precursor for the synthesis of glucose (glycogenic) or fat (ketogenic) or both. From metabolic view point, amino acids are divided into three groups (for details, ***Refer Chapter 15***).

1. **Glycogenic amino acids :** These amino acids can serve as precursors for the formation of glucose or glycogen. e.g. alanine, aspartate, glycine, methionine etc.
2. **Ketogenic amino acids :** Fat can be synthesized from these amino acids. Two amino acids leucine and lysine are exclusively ketogenic.

3. **Glycogenic and ketogenic amino acids :** The four amino acids isoleucine, phenylalanine, tryptophan, tyrosine are precursors for synthesis of glucose as well as fat.

Selenocysteine – the 21st amino acid

As already stated, 20 amino acids are commonly found in proteins. In recent years, a 21st amino acid namely selenocysteine has been added. It is found at the active sites of certain enzymes/proteins (***selenoproteins***). e.g. ***glutathione peroxidase***, glycine reductase, 5′-deiodinase, thioredoxin reductase. Selenocysteine is an unusual amino acid containing the trace element selenium in place of the sulfur atom of cysteine.



Incorporation of selenocysteine into the proteins during translation is carried out by the codon namely UGA. It is interesting to note that UGA is normally a stop codon that terminates protein biosynthesis. Another unique feature is that selenocysteine is enzymatically generated from serine directly on the tRNA (selenocysteine-tRNA), and then incorporated into proteins.

Pyrrolysine – the 22nd amino acid? : In the year 2002, some researchers have described yet another amino acid namely pyrrolysine as the 22nd amino acid present in protein. The stop codon UAG can code for pyrrolysine.

Properties of amino acids

The amino acids differ in their physico–chemical properties which ultimately determine the characteristics of proteins.

A. Physical properties

1. **Solubility :** Most of the amino acids are ***usually soluble in water*** and insoluble in organic solvents.

2. **Melting points :** Amino acids generally melt at higher temperatures, often above 200°C.

3. **Taste :** Amino acids may be sweet (Gly, Ala, Val), tasteless (Leu) or bitter (Arg, Ile). ***Monosodium glutamate*** (MSG; ***ajinomoto***) is used as a flavoring agent in food industry, and Chinese foods to increase taste and flavor. In some individuals intolerant to MSG, ***Chinese restaurant syndrome*** (brief and reversible flu-like symptoms) is observed.

4. **Optical properties :** All the amino acids ***except glycine*** possess optical isomers due to the presence of asymmetric carbon atom. Some amino acids also have a second asymmetric carbon e.g. isoleucine, threonine. The structure of L- and D-amino acids in comparison with glyceraldehyde has been given (***See Fig.4.1***).

5. **Amino acids as ampholytes :** Amino acids contain both acidic (−COOH) and basic ($-NH_2$) groups. They can donate a proton or accept a proton, hence amino acids are regarded as ampholytes.

Zwitterion or dipolar ion : The name *zwitter* is derived from the German word which means ***hybrid***. Zwitter ion (or dipolar ion) is a hybrid molecule containing ***positive and negative ionic groups***.

The amino acids rarely exist in a neutral form with free carboxylic (−COOH) and free amino ($-NH_2$) groups. In strongly acidic pH (low pH), the amino acid is positively charged (cation) while in strongly alkaline pH (high pH), it is negatively charged (anion). Each amino acid has a characteristic pH (e.g. leucine, pH 6.0) at which it carries both positive and negative charges and exists as zwitterion (***Fig.4.2***).

Isoelectric pH (symbol pI) is defined as the pH at which a ***molecule exists as a zwitterion*** or ***dipolar ion*** and carries no net charge. Thus, the molecule is electrically neutral.

The pI value can be calculated by taking the average pKa values corresponding to the ionizable groups. For instance, leucine has two ionizable groups, and its pI can be calculated as follows.

$$pH = \frac{pK_1\left(COO^-\right) + pK_2\left(NH_3^+\right)}{2}$$

$$pI = \frac{2.4 + 9.6}{2} = 6.0$$

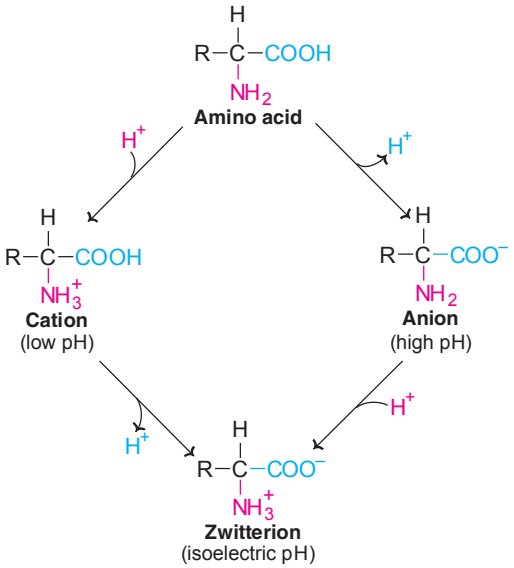


Fig. 4.2 : *Existence of an amino acid as cation, anion and zwitterion.*

Leucine exists as cation at pH below 6 and anion at pH above 6. At the isoelectric pH (pI = 6.0), leucine is found as zwitterion. Thus the pH of the medium determines the ionic nature of amino acids.

For the calculation of pI of amino acids with more than two ionizable groups, the pKas for all the groups have to be taken into account.

Titration of amino acids : The existence of different ionic forms of amino acids can be more easily understood by the titration curves. The graphic representation of leucine titration is depicted in ***Fig.4.3***. At low pH, leucine exists in a fully protonated form as cation. As the titration proceeds with NaOH, leucine loses its protons and at isoelectric pH (pI), it becomes a zwitterion. Further titration results in the formation of anionic form of leucine.

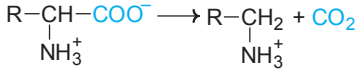
Some more details on isoelectric pH are discussed under the properties of proteins (p. 60).

B. Chemical properties

The general reactions of amino acids are mostly due to the presence of two functional groups namely carboxyl (−COOH) group and amino ($-NH_2$) group.

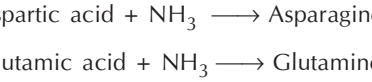
Reactions due to −COOH group

1. Amino acids form salts (−COONa) with bases and esters (−COOR′) with alcohols.
2. **Decarboxylation :** Amino acids undergo decarboxylation to produce corresponding amines.



This reaction assumes significance in the living cells due to the formation of many ***biologically important amines***. These include histamine, tyramine and γ-amino butyric acid (GABA) from the amino acids histidine, tyrosine and glutamate, respectively.

3. **Reaction with ammonia :** The carboxyl group of dicarboxylic amino acids reacts with NH_3 to form amide



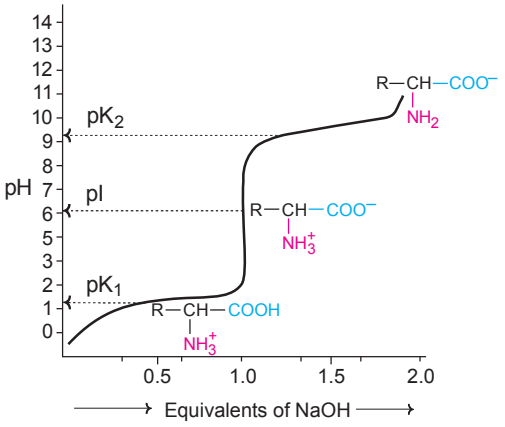


Fig. 4.3 : *Titration curve of an amino acid-leucine ($R = (CH_3)_2—CH—CH_2—$; pK_1 = Dissociation constant for COOH; pI = Isoelectric pH; pK_2 = Dissociation constant for NH_3^+).*

Reactions due to $-NH_2$ group

4. The amino groups behave as bases and combine with acids (e.g. HCl) to form salts ($-NH_3^+Cl^-$).

5. **Reaction with ninhydrin :** The α-amino acids react with ninhydrin to form a purple, blue or pink colour complex (***Ruhemann's purple***).

   Amino acid + Ninhydrin ⟶ Keto acid + NH_3 + CO_2 + Hydrindantin

   Hydrindantin + NH_3 + Ninhydrin ⟶ Ruhemann's purple

   Ninhydrin reaction is effectively used for the quantitative determination of amino acids and proteins. (***Note :*** Proline and hydroxyproline give yellow colour with ninhydrin).

6. **Colour reactions of amino acids :** Amino acids can be identified by specific colour reactions (***See Table 4.3***).

7. **Transamination :** Transfer of an amino group from an amino acid to a keto acid to form a new amino acid is a very important reaction in amino acid metabolism (details given in ***Chapter 15***).

8. **Oxidative deamination :** The amino acids undergo oxidative deamination to liberate free ammonia (***Refer Chapter 15***).

NON-STANDARD AMINO ACIDS

Besides the 20 standard amino acids (described above) present in the protein structure, there are several other amino acids which are biologically important. These include the amino acid derivatives found in proteins, non-protein amino acids performing specialized functions and the D-amino acids.

A. Amino acid derivatives in proteins : The 20 standard amino acids can be incorporated into proteins due to the presence of universal genetic code. Some of these amino acids undergo specific modification after the protein synthesis occurs. These derivatives of amino acids are very important for protein structure and functions. Selected examples are given hereunder.

- Collagen—the most abundant protein in mammals—contains ***4-hydroxyproline*** and ***5-hydroxylysine***.
- Histones—the proteins found in association with DNA—contain many methylated, phosphorylated or acetylated amino acids.
- ***γ-Carboxyglutamic acid*** is found in certain plasma proteins involved in blood clotting.
- Cystine is formed by combination of two cysteines. Cystine is also considered as derived amino acid.

B. Non-protein amino acids : These amino acids, although never found in proteins, perform several biologically important functions. They may be either α-or non-α-amino acids. A selected list of these amino acids along with their functions is given in ***Table 4.2***.

C. D-Amino acids : The vast majority of amino acids isolated from animals and plants are of L-category. Certain D-amino acids are also found in the ***antibiotics*** (actinomycin-D, valinomycin, gramicidin-S). D-serine and D-aspartate are found in brain tissue. D-Glutamic acid and D-alanine are present in bacterial cell walls.

Amino acids useful as drugs

There a certain non-standard amino acids that are used as drugs.

- ***D-Penicillamine*** (D-dimethylglycine), a metabolite of penicillin, is employed in the chelation therapy of Wilson's disease. This is possible since D-penicillamine can effectively chelate copper.
- ***N-Acetylcysteine*** is used in cystic fibrosis, and chronic renal insufficiency, as it can function as an antioxidant.
- ***Gabapentin*** (γ-aminobutyrate linked to cyclohexane) is used as an anticonvulsant.

TABLE 4.2 A selected list of important non-protein amino acids along with their functions

Amino acids	Function(s)
I. α-Amino acids	
Ornithine, Citrulline, Arginosuccinic acid	Intermediates in the biosynthesis of urea.
Thyroxine, Triiodothyronine	Thyroid hormones derived from tyrosine.
S-Adenosylmethionine	Methyl donor in biological system.
Homocysteine	Intermediate in methionine metabolism. A risk factor for coronary heart diseases
Homoserine	Intermediate in threonine, aspartate and methionine metabolisms.
3, 4-Dihydroxy phenylalanine (DOPA)	A neurotransmitter, serves as a precursor for melanin pigment.
Creatinine	Derived from muscle and excreted in urine
Ovothiol	Sulfur containing amino acid found in fertilized eggs, and acts as an antioxidant
Azaserine	Anticancer drug
Cycloserine	Antituberculosis drug
II. Non-α-amino acids	
β-Alanine	Component of vitamin pantothenic acid and coenzyme A
β-Aminoisobutyric acid	End product of pyrimidine metabolism.
γ-Aminobutyric acid (GABA)	A neurotransmitter produced from glutamic acid
δ-Aminolevulinic acid (ALA)	Intermediate in the synthesis of porphyrin (finally heme)
Taurine	Found in association with bile acids.

STRUCTURE OF PROTEINS

Proteins are the polymers of L-α-amino acids. The structure of proteins is rather complex which can be divided into 4 levels of organization (***Fig.4.4***) :

1. **Primary structure :** The linear sequence of amino acids forming the backbone of proteins (polypeptides).

2. **Secondary structure :** The spatial arrangement of protein by twisting of the polypeptide chain.

3. **Tertiary structure :** The three dimensional structure of a functional protein.

4. **Quaternary structure :** Some of the proteins are composed of two or more polypeptide chains referred to as subunits. The spatial arrangement of these subunits is known as quaternary structure.

[The structural hierarchy of proteins is comparable with the structure of a building. The amino acids may be considered as the bricks, the wall as the primary structure, the twists in a wall as the secondary structure, a full-fledged self-contained room as the tertiary structure. A building with similar and dissimilar rooms will be the quaternary structure].

The term ***protein*** is generally used for a polypeptide containing ***more than 50 amino acids.*** In recent years, however, some authors have been using '***polypeptide***' even if the number of amino acids is a few hundreds. They prefer to use protein to an assembly of polypeptide chains with quaternary structure.

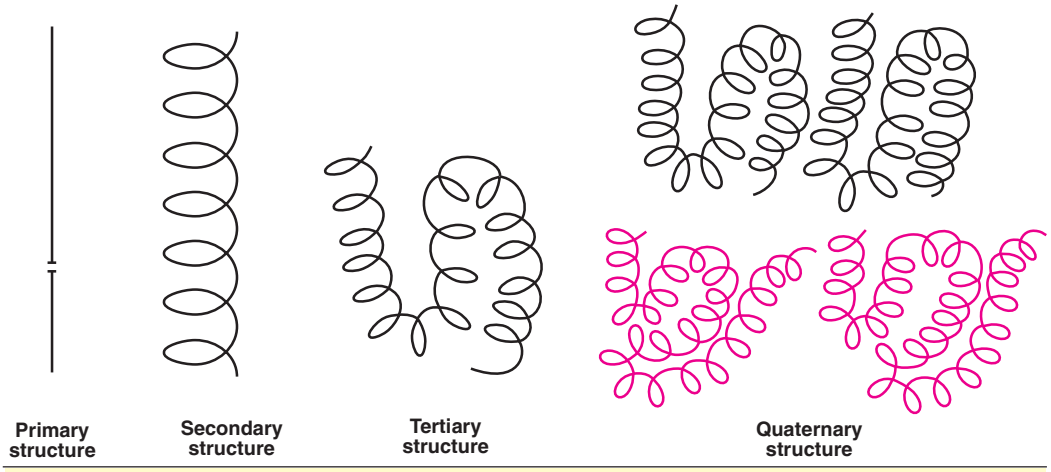


Fig. 4.4 : *Diagrammatic representation of protein structure*
*(**Note :** The four subunits of two types in quaternary structure).*

PRIMARY STRUCTURE OF PROTEIN

Each protein has a unique sequence of amino acids which is determined by the genes contained in DNA. The primary structure of a protein is largely responsible for its function. A vast majority of genetic diseases are due to abnormalities in the amino acid sequences of proteins i.e. changes associated with primary structure of protein.

The amino acid composition of a protein determines its physical and chemical properties.

Peptide bond

The amino acids are held together in a protein by covalent peptide bonds or linkages. These bonds are rather strong and serve as the cementing material between the individual amino acids (considered as bricks).

Formation of a peptide bond : When the ***amino group*** of an amino acid combines with the ***carboxyl group*** of another amino acid, a peptide bond is formed (***Fig.4.5***). Note that a dipeptide will have two amino acids and one peptide (not two) bond. Peptides containing more than 10 amino acids (decapeptide) are referred to as polypeptides.

Characteristics of peptide bonds : The peptide bond is rigid and planar with partial double bond in character. It generally exists in *trans* configuration. Both −C=O and −NH groups of peptide bonds are polar and are involved in hydrogen bond formation.

Writing of peptide structures : Conventionally, the peptide chains are written with the free amino end (N-terminal residue) at the left, and the free carboxyl end (C-terminal residue) at the right. The amino acid sequence is read from N-terminal end to C-terminal end. Incidentally, the protein biosynthesis also starts from the N-terminal amino acid.

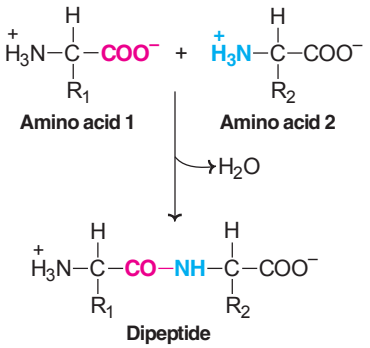


Fig. 4.5 : *Formation of a peptide bond.*

Shorthand to read peptides : The amino acids in a peptide or protein are represented by the 3-letter or one letter abbreviation. This is the ***chemical shorthand*** to write proteins.

Naming of peptides : For naming peptides, the amino acid suffixes ***-ine*** (glycine), ***-an*** (tryptophan), ***-ate*** (glutamate) are changed to ***-yl*** with the exception of C-terminal amino acid. Thus a tripeptide composed of an N-terminal glutamate, a cysteine and a C-terminal glycine is called glutamyl-cysteinyl-glycine.

In the ***Fig.4.6***, the naming and representation of a tripeptide are shown.

$\overset{+}{H_3}N$—glutamate—cysteine—glycine—COO^-	Amino acids in a peptide
E — C — G	One letter symbols
Glu — Cys — Gly	Three letter symbols
Glutamyl — cysteinyl — glycine	Peptide name

Fig. 4.6 : *Use of symbols in representing a peptide (**Note :** A tripeptide with 3 amino acids and two peptide bonds is shown; Free —NH_3^+ is on the left while free —COO^- is on the right).*

Dimensions of a peptide chain : The dimensions of a fully extended polypeptide chain are depicted in ***Fig.4.7***. The two adjacent α-carbon atoms are placed at a distance of 0.36 nm. The interatomic distances and bond angles are also shown in this figure.

Determination of primary structure

The primary structure comprises the identification of constituent amino acids with regard to their quality, quantity and sequence in a protein structure. A pure sample of a protein or a polypeptide is essential for the determination of primary structure which involves 3 stages :

1. Determination of amino acid composition.
2. Degradation of protein or polypeptide into smaller fragments.
3. Determination of the amino acid sequence.

1. Determination of amino acid composition in a protein : The protein or polypeptide is completely hydrolysed to liberate the amino acids which are quantitatively estimated. The hydrolysis may be carried out either by acid or alkali treatment or by enzyme hydrolysis. Treatment with enzymes, however results in smaller peptides rather than amino acids.

Pronase is a mixture of non-specific proteolytic enzymes that causes complete hydrolysis of proteins.

Separation and estimation of amino acids : The mixture of amino acids liberated by protein hydrolysis can be determined by chromatographic techniques. The reader must refer ***Chapter 41*** for the separation and quantitative determination of amino acids. Knowledge on

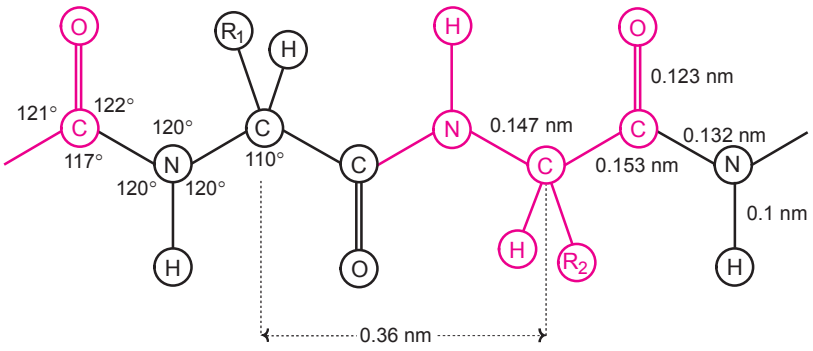


Fig. 4.7 : *Dimensions of a fully extended polypeptide chain. (The distance between two adjacent α-carbon atoms is 0.36 nm).*

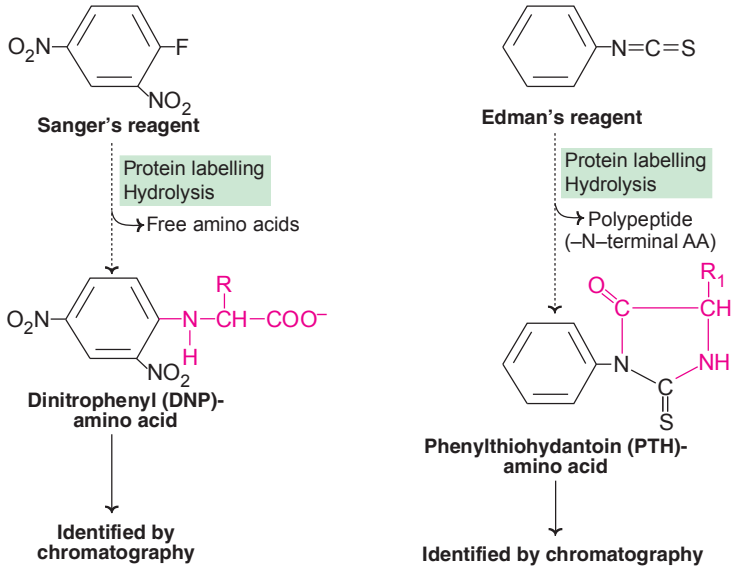


Fig. 4.8 : *Sanger's reagent (1-fluoro 2,4-dinitrobenzene) and Edman's reagent (Phenyl isothiocyanate) in the determination of amino acid sequence of a protein (AA-Amino acid).*

primary structure of proteins will be incomplete without a thorough understanding of chromatography.

2. Degradation of protein into smaller fragments : Protein is a large molecule which is sometimes composed of individual polypeptide chains. Separation of polypeptides is essential before degradation.

(a) **Liberation of polypeptides :** Treatment with urea or guanidine hydrochloride disrupts the non-covalent bonds and dissociates the protein into polypeptide units. For cleaving the disulfide linkages between the polypeptide units, treatment with performic acid is necessary.

(b) **Number of polypeptides :** The number of polypeptide chains can be identified by treatment of protein with ***dansyl chloride***. It specifically binds with N-terminal amino acids to form dansyl polypeptides which on hydrolysis yield N-terminal dansyl amino acid. The number of dansyl amino acids produced is equal to the number of polypeptide chains in a protein.

(c) **Breakdown of polypeptides into fragments :** Polypeptides are degraded into smaller peptides by enzymatic or chemical methods.

Enzymatic cleavage : The proteolytic enzymes such as trypsin, chymotrypsin, pepsin and elastase exhibit specificity in cleaving the peptide bonds (***Refer Fig.8.7***). Among these enzymes, trypsin is most commonly used. It hydrolyses the peptide bonds containing lysine or arginine on the carbonyl (−C=O) side of peptide linkage.

Chemical cleavage : Cyanogen bromide (CNBr) is commonly used to split polypeptides into smaller fragments. CNBr specifically splits peptide bonds, the carbonyl side of which is contributed by the amino acid methionine.

3. Determination of amino acid sequence : The polypeptides or their smaller fragments are conveniently utilized for the determination of sequence of amino acids. This is done in a stepwise manner to finally build up the order of amino acids in a protein. Certain reagents are employed for sequence determination (***Fig.4.8***).

Sanger's reagent : Sanger used 1-fluoro 2, 4-dinitrobenzene (FDNB) to determine insulin structure. ***FDNB*** specifically binds with N-terminal amino acid to form a dinitrophenyl (DNP) derivative of peptide. This on hydrolysis yields DNP-amino acid (N-terminal) and free amino acids from the rest of the peptide chain. DNP-amino acid can be identified by chromatography.

Sanger's reagent has limited use since the peptide chain is hydrolysed to amino acids.

Edman's reagent : ***Phenyl isothiocyanate*** is the Edman's reagent. It reacts with the N-terminal amino acid of peptide to form a phenyl thiocarbamyl derivative. On treatment with mild acid, phenyl thiohydantoin (PTH)–amino acid, a cyclic compound is liberated. This can be identified by chromatography (***Fig.4.8***).

Edman's reagent has an advantage since a peptide can be sequentially degraded liberating N-terminal amino acids one after another which can be identified. This is due to the fact that the peptide as a whole is not hydrolysed but only releases PTH-amino acid.

Sequenator : This is an ***automatic machine*** to determine the amino acid sequence in a polypeptide (with around 100 residues). It is based on the principle of Edman's degradation (described above). Amino acids are determined sequentially from N-terminal end. The PTH-amino acid liberated is identified by high-performance liquid chromatography (HPLC). Sequenator takes about 2 hours to determine each amino acid.

Overlapping peptides

In the determination of primary structure of protein, several methods (enzymatic or chemical) are simultaneously employed. This results in the formation of overlapping peptides. This is due to the specific action of different agents on different sites in the polypeptide. Overlapping peptides are very useful in determining the amino acid sequence.

Reverse sequencing technique

It is the genetic material (chemically DNA) which ultimately determines the sequence of amino acids in a polypeptide chain. By analysing the nucleotide sequence of DNA that codes for protein, it is possible to translate the nucleotide sequence into amino acid sequence. This technique, however, fails to identify the disulfide bonds and changes that occur in the amino acids after the protein is synthesized (post-translational modifications).

SECONDARY STRUCTURE OF PROTEIN

The conformation of polypeptide chain by twisting or folding is referred to as secondary structure. The amino acids are located close to each other in their sequence. Two types of secondary structures, ***α-helix*** and ***β-sheet***, are mainly identified.

Indian scientist Ramachandran made a significant contribution in understanding the spatial arrangement of polypeptide chains.

α-Helix

α-Helix is the ***most common*** spiral structure of protein. It has a rigid arrangement of polypeptide chain. α-Helical structure was proposed by Pauling and Corey (1951) which is regarded as one of the milestones in the biochemistry research. The salient features of α-helix (***Fig.4.9***) are given below

1. The α-helix is a tightly packed coiled structure with amino acid side chains extending outward from the central axis.

2. The α-helix is ***stabilized by*** extensive ***hydrogen bonding***. It is formed between H atom attached to peptide N, and O atom attached to peptide C. The hydrogen bonds are individually weak but collectively, they are strong enough to stabilize the helix.

3. All the ***peptide bonds***, except the first and last in a polypeptide chain, participate in ***hydrogen bonding***.

4. Each turn of α-helix contains 3.6 amino acids and travels a distance of 0.54 nm. The spacing of each amino acid is 0.15 nm.

5. α-Helix is a stable conformation formed spontaneously with the lowest energy.

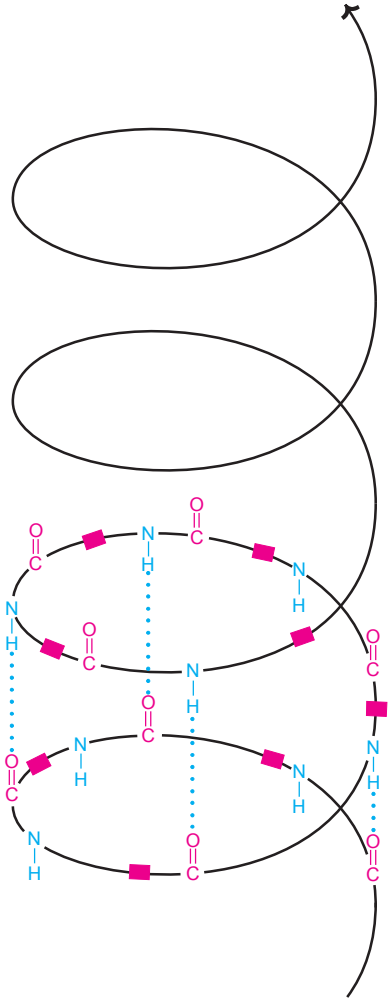


Fig. 4.9 : Diagrammatic representation of secondary structure of protein—a right handed α-helix

(■ -Indicate $-\overset{H}{\underset{|}{C}}-R$ *groups of amino acids; dotted blue lines are hydrogen bonds; Note that only a few hydrogen bonds shown for clarity).*

6. The ***right handed α-helix*** is more stable than left handed helix (a right handed helix turns in the direction that the fingers of right hand curl when its thumb points in the direction the helix rises).

7. Certain amino acids (particularly proline) disrupt the α-helix. Large number of acidic (Asp, Glu) or basic (Lys, Arg, His) amino acids also interfere with α-helix structure.

β-Pleated sheet

This is the second type of structure (hence β after α) proposed by Pauling and Corey. β-Pleated sheets (or simply β-sheets) are composed of two or more segments of ***fully extended peptide chains*** (***Fig.4.10***). In the β-sheets, the hydrogen bonds are formed between the neighbouring segments of polypeptide chain(s).

Parallel and anti-parallel β-sheets

The polypeptide chains in the β-sheets may be arranged either in parallel (the same direction) or anti-parallel (opposite direction). This is illustrated in ***Fig.4.10***.

β-Pleated sheet may be formed either by separate polypeptide chains (H-bonds are interchain) or a single polypeptide chain folding back on to itself (H-bonds are intrachain).

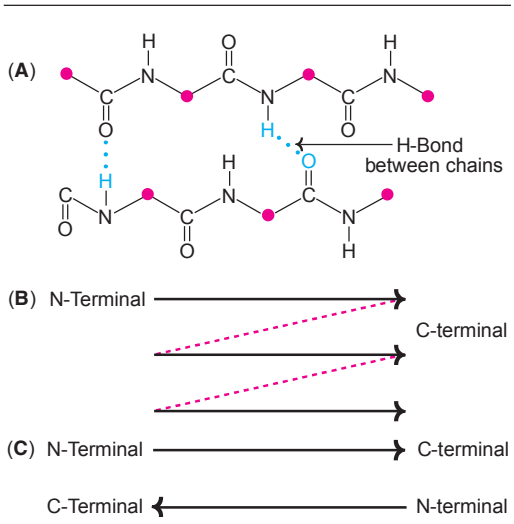


*Fig. 4.10 : Structure of β-pleated sheet (**A**) Hydrogen bonds between polypeptide chains (**B**) Parallel β-sheet (**C**) Antiparallel β-sheet. (**Note :** Red circles in A represent amino acid skeleton* $-\overset{H}{\underset{|}{C}}-R$ *).*

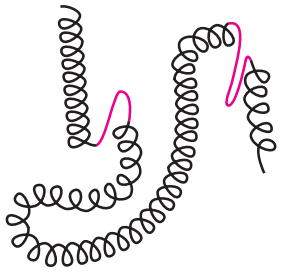


Fig. 4.11 : Diagrammatic representation of a protein containing α-helix and β-pleated sheet (blue).

Occurrence of β-sheets : Many proteins contain β-pleated sheets. As such, the α-helix and β-sheet are commonly found in the same protein structure (***Fig.4.11***). In the globular proteins, β-sheets form the core structure.

Other types of secondary structures : Besides the α-and β-structures described above, the β-bends and nonrepetitive (less organised structures) secondary structures are also found in proteins.

TERTIARY STRUCTURE OF PROTEIN

The ***three-dimensional arrangement of protein*** structure is referred to as tertiary structure. It is a compact structure with hydrophobic side chains held interior while the hydrophilic groups are on the surface of the protein molecule. This type of arrangement ensures stability of the molecule.

Bonds of tertiary structure : Besides the hydrogen bonds, disulfide bonds (−S−S), ionic interactions (electrostatic bonds), hydrophobic interactions and van der Waals forces also contribute to the tertiary structure of proteins.

Domains : The term domain is used to represent the ***basic units of protein structure*** (tertiary) ***and function***. A polypeptide with 200 amino acids normally consists of two or more domains.

QUATERNARY STRUCTURE OF PROTEIN

A great majority of the proteins are composed of single polypeptide chains. Some of the proteins, however, consist of two or more polypeptides which may be identical or unrelated. Such proteins are termed as ***oligomers*** and possess quaternary structure. The individual polypeptide chains are known as ***monomers***, ***protomers*** or ***subunits***. A ***dimer*** consits of ***two*** polypeptides while a ***tetramer*** has four.

Bonds in quaternary structure : The monomeric subunits are held together by non-convalent bonds namely hydrogen bonds, hydrophobic interactions and ionic bonds.

Importance of oligomeric proteins : These proteins play a significant role in the regulation of metabolism and cellular function.

Examples of oligomeric proteins : Hemoglobin, aspartate transcarbomylase, lactate dehydrogenase.

Bonds responsible for protein structure

Protein structure is stabilized by two types of bonds—covalent and non-covalent.

1. **Covalent bonds :** The peptide and disulfide bonds are the strong bonds in protein structure. The formation of ***peptide bond*** and its chracteristics have been described.

Disulfide bonds : A disulfide bond (−***S***−***S***) is formed by the sulfhydryl groups (−SH) of two cysteine residues, to produce cystine (***Fig.4.12A***). The disulfide bonds may be formed in a single polypeptide chain or between different polypeptides. These bonds contribute to the structural conformation and stability of proteins.

2. **Non-covalent bonds :** There are, mainly, four types of non-covalent bonds.

(a) **Hydrogen bonds :** The hydrogen bonds are formed by sharing of hydrogen atoms between the nitrogen and carbonyl oxygen of different peptide bonds (***Fig.4.12B***). Each hydrogen bond is weak but collectively they are strong. A large number of hydrogen bonds significantly contribute to the protein structure.

(b) **Hydrophobic bonds :** The non-polar side chains of neutral amino acids tend to be

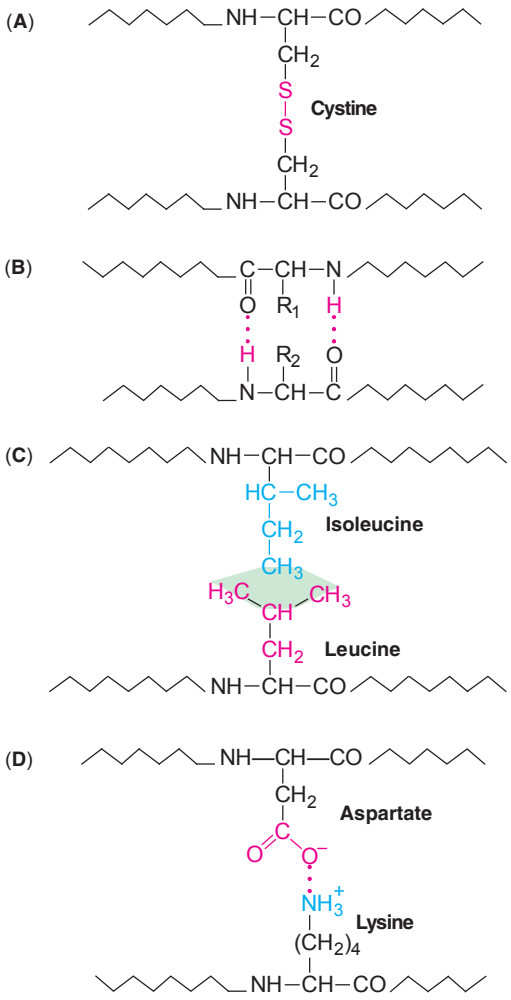


Fig. 4.12 : *Major bonds in protein structure (**A**) Disulfide bond (**B**) Hydrogen bonds (**C**) Hydrophic bonds (**D**) Electrostatic bond.*
*(**Note : See Fig. 4.5** for peptide bond).*

closely associated with each other in proteins (***Fig.4.12C***). As such, these are not true bonds. The occurrence of hydrophobic forces is observed in aqueous environment wherein the molecules are forced to stay together.

(c) **Electrostatic bonds :** These bonds are formed by interactions between negatively charged groups (e.g. COO^-) of acidic amino acids with positively charged groups (e.g. $-NH_3^+$) of basic amino acids (***Fig.4.12D***).

(d) **Van der Waals forces :** These are the non-covalent associations between electrically neutral molecules. They are formed by the electrostatic interactions due to permanent or induced dipoles.

Examples of protein structure

Structure of human insulin : Insulin consists of two polypeptide chains, A and B (***Fig.4.13***). The A chain has glycine at the N-terminal end and asparagine at the C-terminal end. The B chain has phenylalanine and alanine at the N- and C-terminal ends, respectively. Originally, insulin is synthesized as a single polypeptide *preproinsulin* which undergoes proteolytic processing to give *proinsulin* and finally *insulin*.

The structural aspects of hemoglobin and collagen are respectively given in ***Chapters 10*** and ***22***.

Methods to determine protein structure

For the determination of secondary and tertiary protein structures, X-ray crystallography is most commonly used. Nuclear magnetic resonance (NMR) spectra of proteins provides structural and functional information on the atoms and groups present in the proteins.

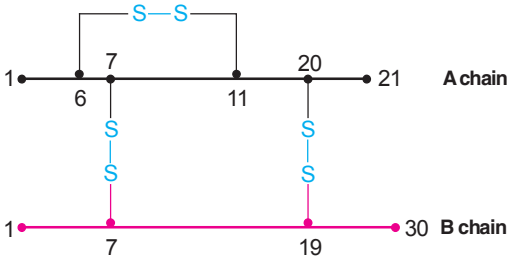


Fig. 4.13 : *Diagrammatic representation of human insulin structure*
*(**Note :** A and B polypeptide chains are held by two disulfide linkages).*

Methods for the isolation and purification of proteins

Several methods are employed to isolate and purify proteins. Initially, proteins are fractionated by using different concentrations of ammonium sulfate or sodium sulfate. Protein fractionation may also be carried out by ultracentrifugation.

Protein separation is achieved by utilizing electrophoresis, isoelectric focussing, immuno-electrophoresis, ion-exchange chromatography, gel-filtration, high performance liquid chromatography (HPLC) etc. The details of these techniques are described in ***Chapter 41***.

PROPERTIES OF PROTEINS

1. **Solubility :** Proteins form ***colloidal solutions*** instead of true solutions in water. This is due to huge size of protein molecules.

2. **Molecular weight :** The proteins vary in their molecular weights, which, in turn, is dependent on the number of amino acid residues. Each amino acid on an average contributes to a molecular weight of about 110. Majority of proteins/polypeptides may be composed of 40 to 4,000 amino acids with a molecular weight ranging from 4,000 to 440,000. A few proteins with their molecular weights are listed below :

Insulin-5,700; Myoglobin-17,000; Hemoglobin-64,450; Serum albumin-69,000.

3. **Shape :** There is a wide variation in the protein shape. It may be globular (insulin), oval (albumin) fibrous or elongated (fibrinogen).

4. **Isoelectric pH :** Isoelectric pH (pI) as a property of amino acids has been described. The nature of the amino acids (particularly their ionizable groups) determines the pI of a protein. The acidic amino acids (Asp, Glu) and basic amino acids (His, Lys, Arg) strongly influence the pI. At isoelectric pH, the proteins exist as ***zwitterions*** or ***dipolar ions***. They are electrically neutral (do not migrate in the electric field) with minimum solubility, maximum precipitability and least buffering capacity. The isoelectric pH(pI) for some proteins are given here

Pepsin-1.1; Casein-4.6; Human albumin-4.7; Urease-5.0; Hemoglobin-6.7; Lysozyme-11.0.

5. **Acidic and basic proteins :** Proteins in which the ratio (ε Lys + ε Arg)/(ε Glu + ε Asp) is greater than 1 are referred to as basic proteins. For acidic proteins, the ratio is less than 1.

6. **Precipitation of proteins :** Proteins exist in colloidal solution due to hydration of polar groups ($-COO^-$, $-NH_3^+$, $-OH$). Proteins can be precipitated by dehydration or neutralization of polar groups.

Precipitation at pI : The proteins in general are least soluble at isoelectric pH. Certain proteins (e.g. casein) get easily precipitated when the pH is adjusted to pI (4.6 for casein). Formation of curd from milk is a marvellous example of slow precipitation of milk protein, casein at pI. This occurs due to the lactic acid produced by fermentation of bacteria which lowers the pH to the pI of casein.

Precipitation by salting out : The process of protein precipitation by the additional of neutral salts such as ***ammonium sulfate*** or ***sodium sulfate*** is known as salting out. This phenomenon is explained on the basis of ***dehydration of protein*** molecules by salts. This causes increased protein-protein interaction, resulting in molecular aggregation and precipitation.

The amount of salt required for protein precipitation depends on the size (molecular weight) of the protein molecule. In general, the higher is the protein molecular weight, the lower is the salt required for precipitation. Thus, serum ***globulins*** are ***precipitated by half saturation*** with ammonium sulfate while ***albumin*** is precipitated ***by full saturation***. Salting out procedure is conveniently used for separating serum albumins from globulins.

The addition of small quantities of neutral salts increases the solubility of proteins. This process called as ***salting in*** is due to the diminished protein–protein interaction at low salt concentration.

Precipitation by salts of heavy metals : Heavy metal ions like Pb^{2+}, Hg^{2+}, Fe^{2+}, Zn^{2+}, Cd^{2+} cause precipitation of proteins. These metals

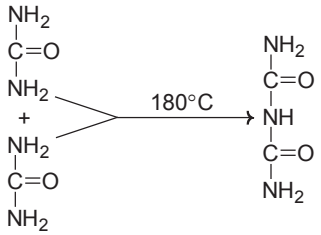
being positively charged, when added to protein solution (negatively charged) in alkaline medium results in precipitate formation. Based on the principle of precipitation, raw egg-white (protein-albumin) is sometimes used to overcome the toxicity of mercury.

Precipitation by anionic or alkaloid reagents : Proteins can be precipitated by trichloroacetic acid, sulphosalicylic acid, phosphotungstic acid, picric acid, tannic acid, phosphomolybdic acid etc. By the addition of these acids, the proteins existing as cations are precipitated by the anionic form of acids to produce protein-sulphosalicylate, protein-tungstate, protein-picrate etc. Industrial ***tanning of leather*** is based on the principle of protein precipitation by tannic acid.

Precipitation by organic solvents : Organic solvents such as alcohol are good protein precipitating agents. They dehydrate the protein molecule by removing the water envelope and cause precipitation. The use of ***surgical spirit*** (about 20% alcohol) as a disinfectant is based on the ***precipitation of proteins*** and the death of bacteria.

7. **Colour reactions of proteins :** The proteins give several colour reactions which are often useful to identify the nature of the amino acids present in them (***Table 4.3***).

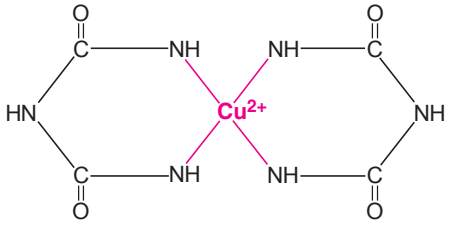
Biuret reaction : Biuret is a compound formed by heating urea to 180°C.



2 Moles of **urea** **Biuret**

When biuret is treated with dilute copper sulfate in alkaline medium, a purple colour is obtained. This is the basis of biuret test widely used for identification of proteins and peptides.

Biuret test is answered by compounds containing two or more CO−NH groups i.e., peptide bonds. All proteins and peptides possessing at least two peptide linkages i.e., tripeptides (with 3 amino acids) give positive biuret test. Histidine is the only amino acid that answers biuret test. The principle of biuret test is conveniently used to detect the presence of proteins in biological fluids. The mechanism of biuret test is not clearly known. It is believed that the colour is due to the formation of a ***copper co-ordinated complex***, as shown below.



The presence of magnesium and ammonium ions interfere in the biuret test. This can be overcome by using excess alkali.

DENATURATION

The phenomenon of ***disorganization of native protein structure*** is known as denaturation. Denaturation results in the loss of secondary, tertiary and quaternary structure of proteins. This involves a change in physical, chemical and biological properties of protein molecules.

Table 4.3 Colour reactions of proteins/amino acids

Reaction	*Specific group or amino acid*
1. Biuret reaction	Two peptide linkages
2. Ninhydrin reaction	α-Amino acids
3. Xanthoproteic reaction	Benzene ring of aromatic amino acids (Phe, Tyr, Trp)
4. Milllons reaction	Phenolic group (Tyr)
5. Hopkins-Cole reaction	Indole ring (Trp)
6. Sakaguchi reaction	Guanidino group (Arg)
7. Nitroprusside reaction	Sulfhydryl groups (Cys)
8. Sulfur test	Sulfhydryl groups (Cys)
9. Pauly's test	Imidazole ring (His)
10. Folin–Coicalteau's test	Phenolic groups (Tyr)

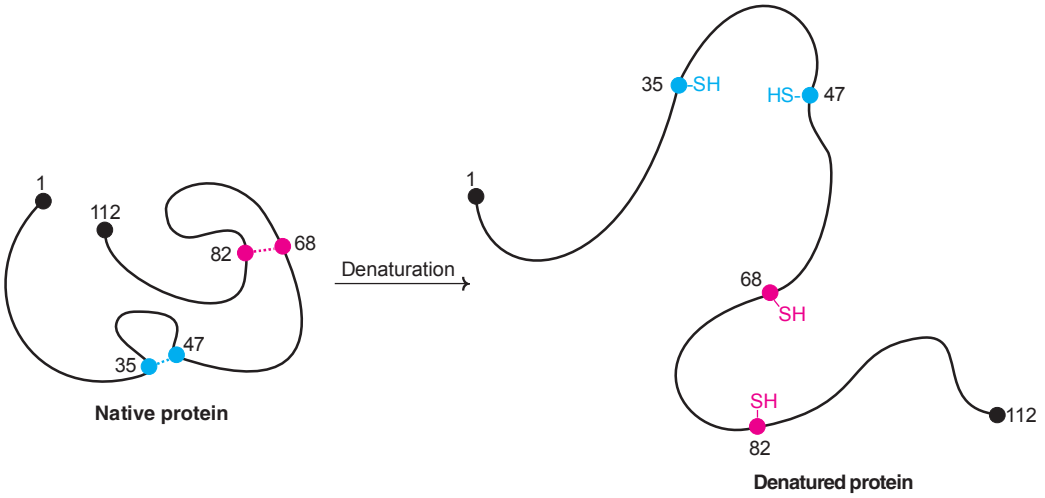


Fig. 4.14 : Denaturation of a protein.

Agents of denaturation

Physical agents : ***Heat***, violent shaking, X-rays, UV radiation.

Chemical agents : ***Acids***, alkalies, organic solvents (ether, alcohol), salts of heavy metals (Pb, Hg), urea, salicylate, detergents (e.g. sodium dodecyl sulfate).

Characteristics of denaturation

1. The native helical structure of protein is lost (***Fig.4.14***).

2. The ***primary structure*** of a protein with peptide linkages remains ***intact*** i.e., peptide bonds are not hydrolysed.

3. The protein ***loses*** its ***biological activity***.

4. Denatured protein becomes insoluble in the solvent in which it was originally soluble.

5 The viscosity of denatured protein (solution) increases while its surface tension decreases.

6. Denaturation is associated with increase in ionizable and sulfhydryl groups of protein. This is due to loss of hydrogen and disulfide bonds.

7. Denatured protein is more easily digested. This is due to increased exposure of peptide bonds to enzymes. Cooking causes protein denaturation and, therefore, cooked food (protein) is more easily digested. Further, denaturation of dietary protein by gastric HCl enchances protein digestion by pepsin.

8. Denaturation is ***usually irreversible***. For instance, omelet can be prepared from an egg (protein-albumin) but the reversal is not possible.

9. Careful denaturation is sometimes reversible (known as ***renaturation***). Hemoglobin undergoes denaturation in the presence of salicylate. By removal of salicylate, hemoglobin is renatured.

10. Denatured protein cannot be crystallized.

Coagulation : The term 'coagulum' refers to a semi-solid viscous precipitate of protein. Irreversible denaturation results in coagulation. Coagulation is optimum and requires lowest temperature at isoelectric pH. Albumins and globulins (to a lesser extent) are coagulable proteins. ***Heat coagulation test is commonly used to detect the presence of albumin in urine***.

Flocculation : It is the process of protein precipitation at isoelectric pH. The precipitate is referred to as flocculum. Casein (milk protein) can be easily precipitated when adjusted to isoelectric pH (4.6) by dilute acetic acid.

Flocculation is reversible. On application of heat, flocculum can be converted into an irreversible mass, coagulum.

CLASSIFICATION OF PROTEINS

Proteins are classified in several ways. Three major types of classifying proteins based on their function, chemical nature and solubility properties and nutritional importance are discussed here.

A. Functional classification of proteins

Based on the functions they perform, proteins are classified into the following groups (with examples)

1. **Structural proteins :** Keratin of hair and nails, collagen of bone.
2. **Enzymes or catalytic proteins :** Hexokinase, pepsin.
3. **Transport proteins :** Hemoglobin, serum albumin.
4. **Hormonal proteins :** Insulin, growth hormone.
5. **Contractile proteins :** Actin, myosin.
6. **Storage proteins :** Ovalbumin, glutelin.
7. **Genetic proteins :** Nucleoproteins.
8. **Defense proteins :** Snake venoms, Immunoglobulins.
9. **Receptor proteins** for hormones, viruses.

B. Protein classification based on chemical nature and solubility

This is a more comprehensive and popular classification of proteins. It is based on the amino acid composition, structure, shape and solubility properties. Proteins are broadly classified into 3 major groups

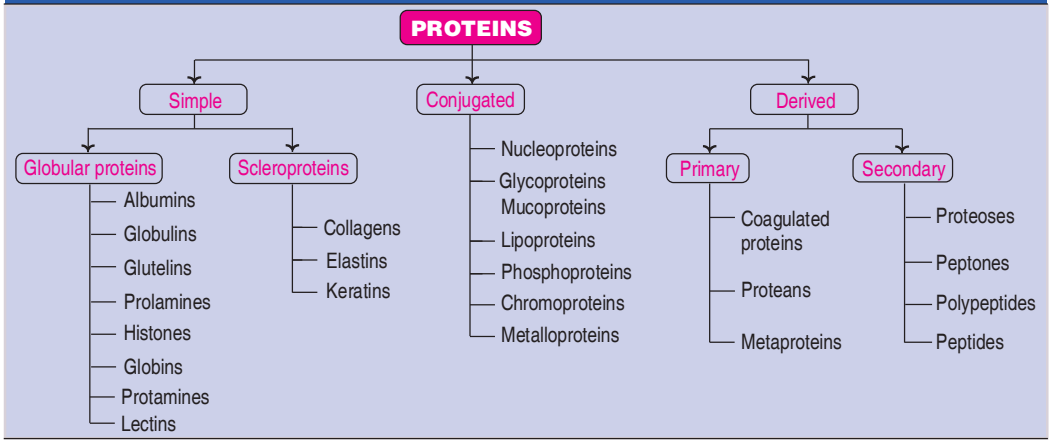
1. **Simple proteins :** They are composed of ***only amino acid*** residues.
2. **Conjugated proteins :** Besides the amino acids, these proteins contain a non-protein moiety known as ***prosthetic group*** or conjugating group.
3. **Derived proteins :** These are the denatured or degraded products of simple and conjugated proteins.

The above three classes are further subdivided into different groups. The summary of protein classification is given in the ***Table 4.4***.

BIOMEDICAL / CLINICAL CONCEPTS

- *Proteins are the most abundant organic molecules of life. They perform static (structural) and dynamic functions in the living cells.*
- *The dynamic functions of proteins are highly diversified such as enzymes, hormones, clotting factors, immunoglobulins, storage proteins and membrane receptors.*
- *Half of the amino acids (about 10) that occur in proteins have to be consumed by humans in the diet, hence they are essential.*
- *A protein is said to be complete (or first class) protein if all the essential amino acids are present in the required proportion by the human body e.g. egg albumin.*
- *Cooking results in protein denaturation exposing more peptide bonds for easy digestion.*
- *Monosodium glutamate (MSG) is used as a flavoring agent in foods to increase taste and flavour. In some individuals intolerant to MSG, Chinese restaurant syndrome (brief and reversible flu-like symptoms) is observed.*

TABLE 4.4 Summary of classification of proteins



1. Simple proteins

(a) **Globular proteins :** These are spherical or oval in shape, soluble in water or other solvents and digestible.

(i) **Albumins :** Soluble in water and dilute salt solutions and coagulated by heat. e.g. serum albumin, ovalbumin (egg), lactalbumin (milk).

(ii) **Globulins :** Soluble in neutral and dilute salt solutions e.g. serum globulins, vitelline (egg yolk).

(iii) **Glutelins :** Soluble in dilute acids and alkalies and mostly found in plants e.g. glutelin (wheat), oryzenin (rice).

(iv) **Prolamines :** Soluble in 70% alcohol e.g. gliadin (wheat), zein (maize).

(v) **Histones :** Strongly basic proteins, soluble in water and dilute acids but insoluble in dilute ammonium hydroxide e.g. thymus histones.

(vi) **Globins :** These are generally considered along with histones. However, globins are not basic proteins and are not precipitated by NH_4OH.

(vii) **Protamines :** They are strongly basic and resemble histones but smaller in size and soluble in NH_4OH. Protamines are also found in association with nucleic acids e.g. sperm proteins.

(viii) **Lectins** are carbohydrate-binding proteins, and are involved in the interaction between cells and proteins. They help to maintain tissue and organ structures. In the laboratory, lectins are useful for the purification of carbohydrates by affinity chromatography e.g. concanavalin A, agglutinin.

(b) **Fibrous proteins :** These are fiber like in shape, insoluble in water and resistant to digestion. Albuminoids or ***scleroproteins*** are predominant group of fibrous proteins.

(i) **Collagens** are connective tissue proteins lacking tryptophan. Collagens, on boiling with water or dilute acids, yield gelatin which is soluble and digestible (**Chapter 22**).

(ii) **Elastins :** These proteins are found in elastic tissues such as tendons and arteries.

(iii) **Keratins :** These are present in exoskeletal structures e.g. hair, nails, horns. Human hair keratin contains as much as 14% cysteine (**Chapter 22**).

2. Conjugated proteins

(a) **Nucleoproteins :** Nucleic acid (DNA or RNA) is the prosthetic group e.g. nucleohistones, nucleoprotamines.

(b) **Glycoproteins :** The prosthetic group is carbohydrate, which is less than 4% of

protein. The term ***mucoprotein*** is used if the carbohydrate content is more than 4%. e.g. mucin (saliva), ovomucoid (egg white).

(c) **Lipoproteins** : Protein found in combination with lipids as the prosthetic group e.g. serum lipoproteins.

(d) **Phosphoproteins :** Phosphoric acid is the prosthetic group e.g. casein (milk), vitelline (egg yolk).

(e) **Chromoproteins :** The prosthetic group is coloured in nature e.g. hemoglobins, cytochromes.

(f) **Metalloproteins :** These proteins contain metal ions such as Fe, Co, Zn, Cu, Mg etc., e.g. ceruloplasmin (Cu), carbonic anhydrase (Zn).

3. Derived proteins : The derived proteins are of two types. The primary derived are the denatured or coagulated or first hydrolysed products of proteins. The secondary derived are the degraded (due to breakdown of peptide bonds) products of proteins.

(a) **Primary derived proteins**

(i) **Coagulated proteins :** These are the denatured proteins produced by agents such as heat, acids, alkalies etc. e.g. cooked proteins, coagulated albumin (egg white).

(ii) **Proteans :** These are the earliest products of protein hydrolysis by enzymes, dilute acids, alkalies etc. which are insoluble in water. e.g. fibrin formed from fibrinogen.

(iii) **Metaproteins :** These are the second stage products of protein hydrolysis obtained by treatment with slightly stronger acids and alkalies e.g. acid and alkali metaproteins.

(b) **Secondary derived proteins :** These are the progressive hydrolytic products of protein hydrolysis. These include proteoses, peptones, polypeptides and peptides.

C. Nutritional classification of proteins

The nutritive value of proteins is determined by the composition of essential amino acids (described already). From the nutritional point of view, proteins are classified into 3 categories.

1. **Complete proteins :** These proteins have all the ten essential amino acids in the required proportion by the human body to promote good growth. e.g. ***egg albumin***, milk casein.

2. **Partially incomplete proteins :** These proteins partially lack one or more essential amino acids, and can promote moderate growth. e.g. wheat and rice proteins (limiting Lys, Thr).

3. **Incomplete proteins :** These proteins completely lack one or more essential amino acids. Hence they do not promote growth at all e.g. ***gelatin*** (lacks Trp), zein (lacks Trp, Lys).

BIOLOGICALLY IMPORTANT PEPTIDES

Several peptides occur in the living organisms that display a wide spectrum of biological functions. Generally, the term 'peptide' is applied when the number of constituent amino acids is less than 10. Some examples of biologically active peptides and their functions are described here.

1. **Glutathione :** It is a tripeptide composed of 3 amino acids. Chemically, glutathione is γ-glutamyl-cysteinyl-glycine. It is widely distributed in nature and exists in reduced or oxidized states.

$$\underset{\text{Reduced}}{2G-SH} \rightleftharpoons \underset{\text{Oxidized}}{G=S-S-G}$$

Functions : In a steady state, the cells generally maintain a ratio of about 100/1 of GSH to G−S−S−G. The reversible oxidation-reduction of glutathione is important for many of its biological functions.

- Glutathione serves as a coenzyme for certain enzymes e.g. prostaglandin PGE_2 synthetase, glyoxylase.
- It prevents the oxidation of sulfhydryl (−SH) groups of several proteins to disulfide (−S−S−) groups. This is essential for the protein function, including that of enzymes.

- It is believed that glutathione in association with glutathione reductase participates in the formation of correct disulfide bonds in several proteins.
- Glutathione (reduced) performs specialized functions in erythrocytes
  - (i) It maintains RBC membrane structure and integrity.
  - (ii) It protects hemoglobin from getting oxidized by agents such as H_2O_2.
- Glutathione is involved in the transport of amino acids in the intestine and kidney tubules via ***γ-glutamyl cycle*** or ***Meister cycle*** (***Refer Chapter 8***).
- Glutathione is involved in the detoxication process. The toxic substances (organo-phosphates, nitro compounds) are converted to mercapturic acids.
- Toxic amounts of peroxides and free radicals produced in the cells are scavanged by glutathione peroxidase (a selenium containing enzyme).

$$2\,GSH + H_2O_2 \xrightarrow{\text{Peroxidase}} G - S - S - G + 2\,H_2O$$

2. **Thyrotropin releasing hormone** (*TRH*) **:** It is a tripeptide secreted by hypothalamus. TRH stimulates pituitary gland to release thyrotropic hormone.

3. **Oxytocin :** It is a hormone secreted by posterior pituitary gland and contains 9 amino acids (nonapeptide). Oxytocin causes ***contraction of uterus***.

4. **Vasopressin** (***antidiuretic hormone, ADH***) **:** ADH is also a nonapeptide produced by posterior pituitary gland. It stimulates kidneys to retain water and thus ***increases the blood pressure***.

5. **Angiotensins :** Angiotensin I is a decapeptide (10 amino acids) which is converted to angiotensin II (8 amino acids). The later has more hypertensive effect. Angiotensin II also stimulates the release of aldosterone from adrenal gland.

BIOMEDICAL / CLINICAL CONCEPTS

- *Collagen is the most abundant protein in mammals. It is rich in hydroxyproline and hydroxylysine.*
- *Several biologically important peptides are known in the living organism. These include glutathione for the maintenance of RBC structure and integrity; oxytocin that causes uterus contraction; vasopressin that stimulates retention of water by kidneys; enkephalins that inhibit the sense of pain in the brain.*
- *Antibiotics such as actinomycin, gramicidin, bacitracin and tyrocidin are peptide in nature.*
- *γ-Carboxyglutamic acid is an amino acid derivative found in certain plasma proteins involved in blood clotting.*
- *Homocysteine has been implicated as a risk factor in the onset of coronary heart diseases.*
- *Several non-protein amino acids of biological importance are known. These include ornithine, citrulline and arginosuccinic acid (intermediates of urea synthesis), thyroxine and triiodothyronine (hormones), and β-alanine (of coenzyme A).*
- *The protein-free filtrate of blood, required for biochemical investigations (e.g. urea, sugar) can be obtained by using protein precipitating agents such as phosphotungstic acid and trichloroacetic acid.*
- *Heat coagulation test is most commonly employed to detect the presence of albumin in urine.*

6. **Methionine enkephalin :** It is a penta-peptide found in the brain and has opiate like function. It ***inhibits*** the sense of a ***pain***.

7. **Bradykinin and kallidin :** They are nona- and decapeptides, respectively. Both of them act as powerful vasodilators. They are produced from plasma proteins by snake venom enzymes.

8. **Peptide antibiotics :** Antibiotics such as ***gramicidin***, ***bacitracin***, tyrocidin and actinomycin are peptide in nature.

9. **Aspartame :** It is a ***dipeptide*** (aspartyl-phenylalanine methyl ester), produced by a combination of aspartic acid and phenylalanine. Aspartame is about ***200 times sweeter than sucrose***, and is used as a low-calorie artificial sweetner in softdrink industry.

10. **Gastrointestinal hormones :** Gastrin, secretin etc. are the gastrointestinal peptides which serve as hormones.

SUMMARY

1. *Proteins are nitrogen containing, most abundant organic macromolecules widely distributed in animals and plants. They perform structural and dynamic functions in the organisms.*
2. *Proteins are polymers composed of L-α-amino acids. They are 20 in number and classified into different groups based on their structure, chemical nature, nutritional requirement and metabolic fate. Selenocysteine has been recently identified as the 21st amino acid, and is found in certain proteins.*
3. *Amino acids possess two functional groups namely carboxyl (−COOH) and amino ($-NH_2$). In the physiological system, they exist as dipolar ions commonly referred to as zwitterions.*
4. *Besides the 20 standard amino acids present in proteins, there are several non-standard amino acids. These include the amino acid derivatives found in proteins (e.g. hydroxy-proline, hydroxylysine) and, non-protein amino acids (e.g. ornithine, citrulline).*
5. *The structure of protein is divided into four levels of organization. The primary structure represents the linear sequence of amino acids. The twisting and spatial arrangement of polypeptide chain is the secondary structure. Tertiary structure constitutes the three dimensional structure of a functional protein. The assembly of similar or dissimilar polypeptide subunits comprises quaternary structure.*
6. *The determination of primary structure of a protein involves the knowledge of quality, quantity and the sequence of amino acids in the polypeptide. Chemical and enzymatic methods are employed for the determination of primary structure.*
7. *The secondary structure of protein mainly consists of α-helix and/or β-sheet. α-Helix is stabilized by extensive hydrogen bonding. β-Pleated sheet is composed of two or more segments of fully extended polypeptide chains.*
8. *The tertiary and quaternary structures of protein are stabilized by non-covalent bonds such as hydrogen bonds, hydrophobic interactions, ionic bonds etc.*
9. *Proteins are classified into three major groups. Simple proteins contain only amino acid residues (e.g. albumin). Conjugated proteins contain a non-protein moiety known as prosthetic group, besides the amino acids (e.g. glycoproteins). Derived proteins are obtained by degradation of simple or conjugated proteins.*
10. *In addition to proteins, several peptides perform biologically important functions. These include glutathione, oxytocin and vasopressin.*