

Protein Structure and Function

1.1

Proteins are the most abundant biomolecules. They occur in every part of the cell and constitute about 50% of the dry weight of the cell. The word protein is derived from Greek word “**proteios**” meaning holding the first place. Proteins are the most abundant cellular components. They include enzymes, antibodies, hormones, transport molecules and even component for the cytoskeleton of cell itself. Proteins are also informational macromolecules, the ultimate heirs of genetic information encoded in the sequences of nucleotide bases of DNA. Proteins are predominantly constituted by five major elements in 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.

1.2

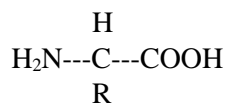
Proteins are polymers of amino acids

Proteins on hydrolysis give amino acids and sometimes non-protein residue (s) in addition to amino acids. Thus amino acids are the building blocks of proteins which are joined to form proteins that have unique three-dimensional structure, making them capable of performing specific biological function.

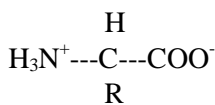
1.3

Structure of Amino acids:

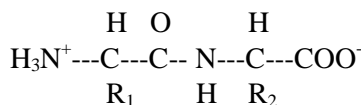
More than three hundred different amino acids have been described in nature, only twenty (20) are commonly found in mammalian proteins. Each amino acid (except Proline) has a **carboxyl group** (-COOH), an **amino group** (-NH₂) and a distinctive **side chain** (-R group) bonded to **α-carbon** atom (fig 1.1) At Physiological pH (approx. 7.4) carboxyl group is dissociated, forming (-COO⁻) ion and amino group is protonated (-NH₃⁺) giving zwitterion form (fig 1.2). The zwitterion form of amino acids have been established by several methods including spectroscopic measurements and X-ray techniques. In proteins -COOH group and -NH₂ are combined in **peptide linkage** (fig1.3) and are not available for chemical reaction. Thus it is the nature of side chains that are important in proteins .



General structure
of amino acid
fig1.1

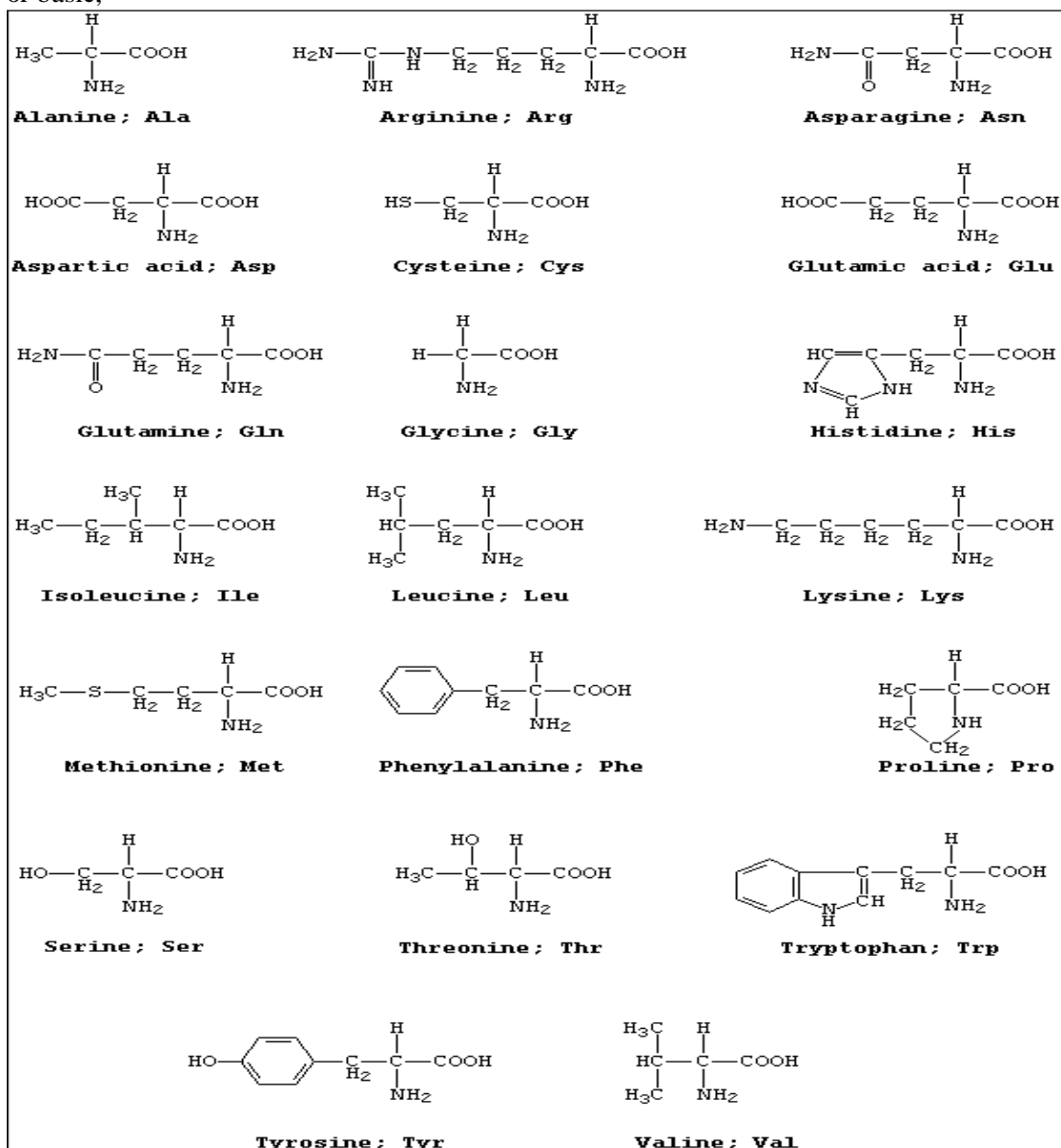


Zwitterionic form
of amino acid
fig1.2



Formation of peptide bond
between two of amino acid
fig1.3

It is, therefore, useful to classify amino acids on the basis of nature of their side chains (-R groups) that is, whether they are non polar or polar (uncharged, acidic or basic,



Structural formulae of the 20 genetically controlled amino acids

a) Amino acids with non polar side chains:-

Each of these has a non polar side chain and does not bind or give off protons or participate in H-bonding. Their side chains are hydrophobic and generally fill up the interior of the folded protein. They are 9 and include *glycine, alanine, valine, leucine, isoleucine, phenylalanine, tryptophan, methionine and proline*

b) Amino acids with Uncharged polar side chains:-

These amino acids have a zero net charge at neutral pH. they include *serine, threonine, asparagine, glutamine, tyrosine and cysteine*. *Serine, threonine and tyrosine* have a polar hydroxyl group that take part in H-bonding, while cysteine and tyrosine can lose proton at alkaline pH. The side chain of asparagine and glutamine contain acid amide group. Side chains of two cysteines covalently cross-link to form a disulfide bond (-S-S-)

c) Amino acids with acidic side chains:-

The amino acids *aspartic acid and glutamic acid* are proton donors. At neutral pH the side chains of these amino acids are fully ionized containing negatively charged (-COO⁻) group. They are therefore, called *aspartate and glutamate*.

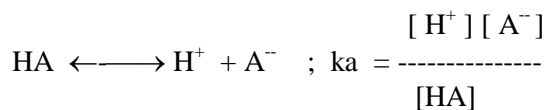
d) Amino acids with basic side chains:-

These amino acids accept protons. At physiological pH the side chain of these amino acids bear positive charge. These include *lysine, arginine and histidine*. Lysine and arginine are fully ionized and positively charged, while histidine is weakly basic at physiological pH.

The 20 amino acids vary considerably in their physicochemical properties such as polarity, acidity, basicity, aromaticity, bulk, conformational flexibility, ability to cross link, ability to hydrogen bond and chemical reactivity.

1.4 Acid-Base Properties of Amino Acids:-

Amino acids in aqueous solutions contains weakly acidic α -carboxyl group and weakly basic α -amino acid. In addition, each of the acidic and basic amino acid contain ionizable group in its chain. Thus both free amino acids as well as peptides and proteins act as buffers. The relation between concentration of a weak acid (HA) and its conjugate base (A⁻) is described by the Henderson-Hasselbach equation:

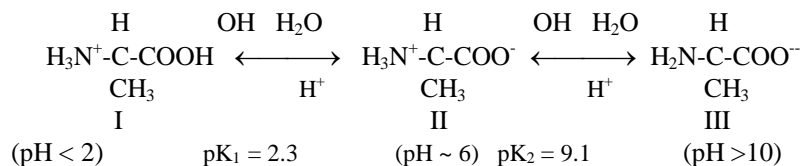


$$\text{pH} = \text{pK} + \log \frac{[\text{A}^-]}{[\text{HA}]}$$

i) Titration of Amino acids:-

Consider an amino acid alanine which contains both a carboxyl and an amino group. At a low pH, both of these groups are protonated. As the pH of the

solution is raised, the -COOH group (I) can dissociate by donating a proton to the medium resulting in the formation of carboxylate group -COO⁻. This form is called *zwitterion* (II)



Henderson-Hasselbalch equation can be used to analyse the dissociation of carboxyl group of alanine

$$\text{pH} = \text{pK}_1 + \log \frac{[\text{III}]}{[\text{I}]}$$

The second titrable group of alanine is amino group. This is a much weaker acid than -COOH group and therefore has much smaller dissociation constant K_2 . Release of proton from protonated group of II results in fully deprotonated form of alanine (III).

By applying Henderson-Hasselbalch equation to each dissociable acidic group, it is possible to calculate the complete titration curve of a weak acid (fig1.3). It follows from the curve:

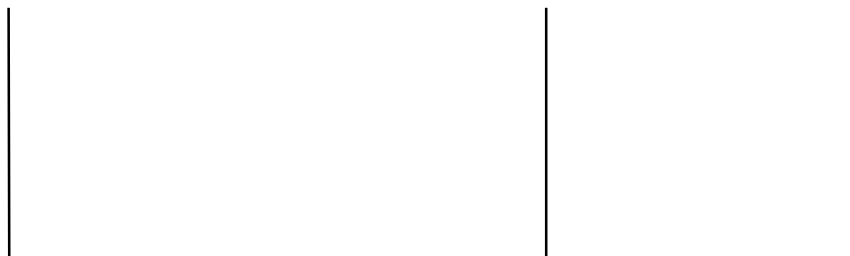
a) The -COOH/-COO⁻ pair can serve as a buffer in the pH region around pK_1 and -NH₃⁺/NH₂ pair can serve buffer in the region around pK_2 ($\text{pK} \pm 1$)

b) When $\text{pH} = \text{pK}_1$ (2.3), equal amounts of form I and II of alanine exist in solution. When pK_2 (9.1), equal amounts form II and III are present in the solution.

c) At neutral pH, alanine exist predominantly as zwitterion with no net charge. This pH is called isoelectric pH (pI), Thus the molecule is electrically neutral. the pI value can be calculated by taking the average of pK_1 and pK_2 i.e.

$$\text{pI} = [\text{pK}_1 + \text{pK}_2]/2 = [2.3 + 9.1]/2 = 5.7$$

At physiological pH, all amino acids exist as zwitterions.



ii) Titration of histidine

Histidine is an example of an amino acid that contains three chemical groups, each of which can reversibly gain or lose a proton: the carboxyl group, the imidazole group of the side chain and the α - amino group.

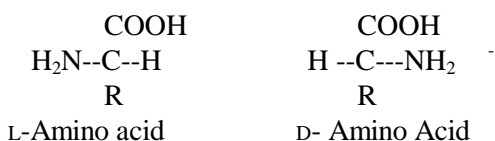
Incremental addition of base to fully protonated histidine results in the sequential removal protons from the carboxyl group ($pK_1 = 1.8$), the imidazole group ($pK_2 = 6.0$) and the amino group ($pK_3 = 9.2$)

The pI for histidine is calculated by first identifying the isoelectric form of the amino acid (form III) then averaging the values of the nearest pKs.

$$pI = [pK_2 + pK_3]/2 = [6.0 + 9.2]/2 = 7.6$$

1.5 Optical properties of amino acids:-

The α - carbon of each amino acid is attached to four different chemical groups and is therefore a chiral, except glycine. All naturally occurring amino acids are of the L- configuration. However D- amino acids also occur and are found in some antibiotics and in bacterial cell walls.



The two forms in each pair are termed as stereoisomers, optical isomers or enantiomers.

1.6 a) Essential amino acids

The amino acids which can not 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 essential amino acids are:

Arginine, Valine, Histidine, Isoleucine, Leucine, Lysine, Methionine, Phenylalanine, Tyrosine, Tryptophan. (A.V HILL, MP,T.T)

Of the ten listed above two amino acid namely arginine and histidine can be partly synthesized by adult human, hence are considered as semi-essential (Ah). Thus 8 are absolutely essential while 2 semi essential.

b) Non-essential amino acids:-

The body can synthesize about 10 amino acids to meet the biological needs, hence they need not be consumed in diet. These *Glycine, alanine, serine, cystein, aspartate, asparagine, glutamate, glutamine, tyrosine, and proline.*

Structure of Proteins

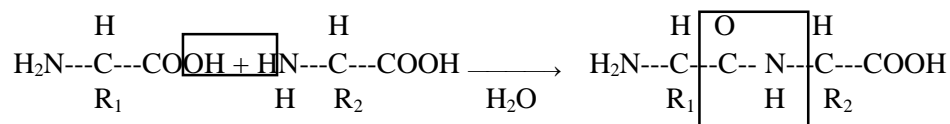
The 20 amino acids commonly found in proteins are linked together by peptide bonds. The linear sequence of linked amino acids contain the information necessary to generate a protein molecule with unique three dimensional shape. the complexity of protein structure is best analysed in terms of four organizational levels, namely, primary, secondary, tertiary and quarternary level.

2.1.Primary structure of proteins:-

The sequence of amino acids in a protein is called the primary structure of protein. determining the order of amino acids in a polypeptide chain requires the application of several experimental techniques.

i)Peptide bond:-

The amino acids are joined covalently by peptide bond which are amide linkages between the α - carboxyl group of amino acid and α - amino group of another.



ii) Characteristics of peptide bond:-

The peptide bond has a partial double bond character and is rigid and planar. This prevents the free rotation about this bond. However, the bonds between the α -carbons and α -amino or α -carboxyl groups can be freely rotated and thus allows the polypeptide to assume variety of possible configuration. The peptide bond is generally trans bond. Peptide bond neither accepts nor gives off proton over the pH of 2 to 12

iii) Order of amino acid:-

By convention free amino end of the polypeptide chain (N-terminus) is written on left and free carboxyl end (C-terminus) to the right. Therefore amino acid sequence is read from N terminal end to C terminal end.

iv)Naming of polypeptide:-

Linkage of many amino acid through peptide bond result in the formation of unbranched chain called polypeptide chain. Each component amino acid in the polypeptide chain is called residue or moiety. In naming a polypeptide, the names of amino acid end in “yl” with exception of C- terminal amino acid e.g. in case of tripeptide *H₂N-Val-Gly-Leu-COOH* the chain can be named as “valylglycylleucine”.

2.2 Secondary structure of protein:-

The polypeptide backbone does not assume a random three dimensional structure , but instead generally forms arrangement of amino acids that are located near to

each other in linear sequence. These arrangements are termed the secondary structure. The α -helix, β -sheets and β -bends are examples of secondary structure.

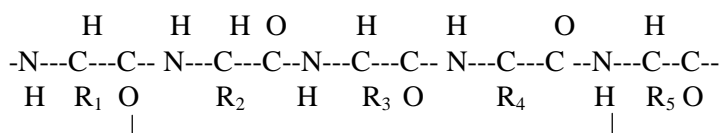
a) α -helix :-

There are several different polypeptides helices found in nature, but the α -helix is the most common. It is a spiral structure, consisting of a tightly packed, coiled polypeptide backbone core with the side chain of the component amino acids extending outward from the central axis in order to avoid interfering sterically with each other.

A very diverse group of proteins contain α -helices. e.g. keratin are a family of closely related, fibrous proteins, whose structure is nearly entirely α -helical. they are a major component of tissues such as hair and skin. Their rigidity is determined by the number of disulfide bond between the constituent polypeptide chains.

i) Hydrogen Bonds:-

An α -helix is stabilised by extensive hydrogen bonding between the peptide bond carbonyl oxygen and amide hydrogen that are part polypeptide backbone. The H-bond extends up the spiral from the carbonyl oxygen of one peptide bond to the -NH- group of a peptide linkage four residues ahead.



ii) Amino acids per turn:-

Each turn of an α -helix contains 3.6 amino acids per turn covering a distance of 0.54nm and each amino acid representing an advance of 0.15 nm along the axis of the helix.

iii) Amino acids that disrupt an α -helix

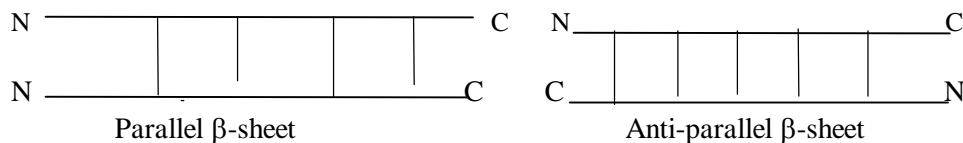
Certain amino acids are more often found in α -helix than other. In particular proline is rarely found as it can not form the correct pattern of H-bonds due to the lack of H atoms. It inserts a kink in the chain that disrupts the smooth helical structure. Large numbers of charged amino acids (Glu, Asp, His, Lys or Arg) also disrupt α -helix by forming ionic bonds with bulky side chains, such as Trp, Val or Ile.

b) β - Sheets

β -sheet is another form of secondary structure in which all of the peptide bond components are involved in H-bonding. The structure appears pleated, hence called β -pleated sheets.

i) Parallel and anti parallel sheets:-

A β -sheet can be formed from two or more separate polypeptide chains or segments of polypeptide chains that are arranged either parallel or anti parallel to each other. A β -sheet can also be formed by a single polypeptide chain folding back on itself. In globular proteins, β -sheet always have a right-handed curl or twist when viewed along polypeptide backbone.



c) β -Bends:-

β -Bends reverse the direction of a polypeptide chain, helping it to form a compact, globular shape. They are usually found on the surface of protein molecule and often include charged residues. β -Bends are generally composed of four amino acid one of which may be proline- the amino acid that imparts a kink in the chain. Glycine, the amino acid with smallest R-group is also found. β -Bends are stabilised by H- bonds.

2.3 Tertiary structure :-

Tertiary structure refers to spatial arrangement of amino acids that are far apart in the linear sequence as well as those residues that are adjacent. The structure of globular protein in aqueous solution is compact, with a high density of atoms in the core of the molecule.

The polypeptide chain folds spontaneously so that the majority of hydrophobic side chains are buried in the interior, whereas hydrophilic groups are generally found on the surface of the molecule. Once folded the three dimensional, biologically active (native) conformation of protein is maintained not only by hydrophobic interactions but also by electrostatic forces, H-bonds and if present by covalent disulfide bonds. The electrostatic forces include salt bridges between oppositely charged groups and multiple weak van der Waal's interactions between tightly packed aliphatic side chains. In the interior of the protein.

It is generally accepted that the information needed for correct folding is contained in the primary structure of the polypeptide itself. In addition a specialized group of proteins, named chaperons (also called polypeptide chain binding proteins, PCB) are required for the proper folding of many species of proteins. They act as catalysts by increasing the rate of final stages in the folding process.

2.4 Quaternary structure:-

Many proteins consist of single polypeptide chain; these are monomeric proteins. But many others consist of two or more polypeptide chains that may be structurally identical or totally unrelated. The arrangement of these polypeptide subunits is called quaternary structure of protein. Such proteins are called oligomeric proteins. The individual polypeptides are called as monomers, protomers or subunits.

Subunits are held by noncovalent interactions (e.g. hydrophobic interactions, H-bonds ionic bonds). Sub units may function independently of each other or may work together cooperatively, as in hemoglobin(tetra-meric protein) where binding of oxygen to one polypeptide facilitates the binding of other oxygen to the other polypeptide.

2.5 Determination of primary structure proteins:-

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

- A) *Determination of amino acid composition.*
- B) *Degradation of protein or polypeptide into smaller fragments*
- C) *Determination of amino acid sequencing*

A)

Determination of amino acid composition:-

The protein or polypeptide is completely hydrolysed to liberate the amino acids which are quantitatively estimated. The hydrolysis may be carried out by acids or by alkali.

a) Acid hydrolysis:-

The proteins or polypeptide is dissolved in 6N HCl and heated at 110°C in a sealed evacuated tube for 20-70 hours. By this treatment, peptide bonds are cleaved to release the amino acids. Acid hydrolysis however, destroys tryptophan and converts asparagine and glutamine to aspartic and glutamic acid.

b) Alkaline hydrolysis:-

This is done by using 2 to 4N NaOH at 100°C for 5 to 8 hours. It however causes decomposition of serine, threonine and cysteine.

Enzyme hydrolysis can also be employed which results in smaller peptides.

c) Separation and estimation of amino acids:

The mixture of amino acids liberated by any of the above hydrolysis process can be separated by cation-exchange chromatography. In this method a mixture of amino acid is applied to a column that contains a resin to which a negatively charged group is tightly attached .

Each amino acid is sequentially released from the chromatography column by eluting with solution of increasing ionic strength and pH. As the pH increases , the amino acids lose hydrogen ion, first from α -carboxyl group and then from side chain and α -amino group, first become neutral, then negatively charged and are released from resin. Each amino acid emerges at specific pH and ionic strength.

d)Quantitative analysis:-

The separated amino acids contained in the elute from the column are quantitated by heating them with ninhydrin, a reagent that forms purple compound with most amino acids , ammonia, and amines (yellow with proline) The amount of each amino acid is determined spectrophotometrically. The analysis can be done using amino acid analyser.

B) Degradation of protein or polypeptide into smaller fragments

If the protein contains many polypeptides then each polypeptide chain is liberated By treating with urea or guanidine hydrochloride and number of polypeptides determined by treating with dansyl chloride. Polypeptides produced are degraded into smaller fragments enzymatically or chemically.

a) Enzymic cleavage :-

Trypsin is a digestive enzyme and is commonly used in laboratory to cleave a polypeptide into fragments that are of a size amenable to sequence analysis.

Trypsin cleaves peptide bonds on the carbonyl side of either lysine or arginine.

The fragments produced are called tryptic peptides

b) Chemical cleavage:-

Treatment with cyanogen bromide (CNBr) is also commonly used to split polypeptide on the carbonyl side of methionine residues. This reagent like trypsin is highly specific.

C) Determination of amino acid sequencing:-

The polypeptide or their smaller fragments are conveniently utilized for determination of amino acid sequence.

This is done in a stepwise manner to finally build up the order of amino acids in a protein. Several methods are used for this purpose.

a)Sanger's reagent:-

Sanger used 1-Fluoro-2,4-dinitrobenzene (FDNB) to determine insulin structure.

It binds with the N-terminal amino acid to form a dinitrophenyl (DNP) derivative of peptide. This on

hydrolysis yields DNP-amino acid (N-terminal) and free amino acid from the rest of the peptide chain. DNP-amino acid can be identified by chromatography.

b)Edman's reagent:- Phenyl isothiocyanate, known as Edman's reagent is used to label the amino terminal residue under mildly alkaline conditions.

The resulting phenylthiohydantion (PTH) derivative introduces an instability in the N-terminal peptide bond that can be selectively hydrolysed without cleaving the other peptide bonds. The PTH-amino acid can be identified by chromatography. This process has been automated in sequenator to determine the sequence of over 100 amino acids, starting from the N terminal end of polypeptide.

D)

Overlapping peptides:-

The peptides produced by cleavage of a protein with proteolytic enzymes of CNBr are usually small enough to be sequenced by Edman's degradation. However, even after the sequence of peptides in the original polypeptide is not yet known. To determine the position of these fragments it is necessary to prepare overlapping peptides. These are formed by treating separate samples of the original protein or polypeptide with reagents that cleave the chain at different sites. Overlapping peptides act as bridges in ordering the peptide fragments, because a peptide produced by one reagent overlaps the sequence of fragments.

2.6

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 as :

Functional classification of proteins:-

Based on functions that proteins perform, they are classified into:

i) Structural protein:-

Many proteins serve as supporting filaments, cables or sheets to give biological structures strength or protection. The major components of tendons and cartilage is the fibrous protein collagen, which has very high tensile strength. Leather is almost pure collagen. Ligaments contain elastin, a structural protein capable of stretching in two dimensions. Hair, finger nails and feathers consist largely of the tough, insoluble protein keratin. Silk fibres, spider web, wings and hinges of some insects are proteins.

ii) Enzymes:-

The most varied and most highly specialized proteins are those with catalytic activity- the enzymes. Virtually all the chemical biochemical reaction are catalyzed by enzymes. Many thousand enzymes purified are proteins.

iii) Transport proteins:-

Transport proteins in blood plasma bind and carry specific molecules or ions from one organ to another. Hemoglobin carries oxygen from lungs to peripheral tissues. Albumin carries a number of ions.

iv) Defense proteins:-

Many proteins help in defending organisms against invasion by other species. The immunoglobulins are specialized proteins made by lymphocytes of vertebrates. Fibrinogen and thrombin are blood clotting proteins that prevent the blood loss. Snake venoms, bacterial toxins are proteins.

v) Contractile proteins:-

Some proteins endow cells and organisms with the ability to contract, to change shape, or to move about. Actin and myosin are examples of such proteins

vi) Storage proteins :-

The seeds of many plants store nutrient proteins required for growth of germinating seedling. Similarly ovalbumin of egg, casein of milk are examples of nutrient proteins.

vii) Regulatory proteins:-

Some proteins help regulate cellular or physiological activities. Among them are many hormones, G-proteins etc.

A) Classification on the basis structural composition

This is a more comprehensive and popular classification of proteins . It is based on the amino acid composition, shape and solubility of proteins. Proteins are divided into 3 types:

1. **Simple proteins**
2. **Conjugated proteins**
3. **Derived proteins**

1.Simple proteins:-

Simple proteins are those which are composed of only amino acid residues. They are further divided into different groups as follows:

a) Globular proteins:-

These are spherical or oval in shape , soluble in water or other solvents and are digestible. Globular proteins include:

i) Albumins:-

They are soluble in water and dilute salt solutions and coagulated by heat e.g. serum albumin, ovalbumin.

ii) Globulins:-

They are soluble in neutral and dilute salt solutions e.g. viteline tuberin, legumin etc.

iii) Glutelins:-

They are soluble in dilute acid and alkalis and are mostly found in plants. e.g. glutelin, oryzenin etc.

iv) Prolamines:-

They are soluble in 70% alcohol e.g. gliadin, zein

v) Histones:-

They are strongly basic, soluble in water and dilute acids but insoluble in dilute ammonium hydroxide e.g. thymus histone

vi) Globins:-

These are generally considered along with histones but they are not basic 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.

b) Fibrous Proteins :-

These are fiber like in shape, insoluble in water and resistant to digestion. Albuminoids or scleroproteins constitute the most predominant group of fibrous proteins.

i) Collagens:-

They are connective tissue proteins lacking tryptophan. Collagens on boiling in water or dilute acids, yield gelatin which is soluble and digestible.

ii) Elatins:-

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 etc.

c) Conjugated proteins:-

These proteins in addition to amino acid residue contain other non-proteinic moiety called prosthetic group. These include:

i) Nucleoproteins:-

These proteins contain nucleic acids as prosthetic group e.g. nucleohistones, nucleoprotamines

ii) Glycoproteins:-

These proteins contain carbohydrates, which is less than 4% as prosthetic group. If carbohydrate content is more than 45 they are called mucoproteins e.g. mucin, ovomucoid.

iii) Lipoproteins:-

These proteins contain lipids as prosthetic group e.g. membrane lipoproteins.

iv) Phosphoproteins:-

These proteins contain phosphoric acids as prosthetic group e.g. casein.

v) Chromoproteins:-

These proteins contain nucleic acids a coloured prosthetic group e.g. hemoglobin, cytochromes

vi) Metalloproteins:-

These proteins contain metal ions such as Fe, Co, Zn, Mg, etc. e.g. ceruloplasmin(Cu) carbonic acid (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 proteins of proteins.

a) Primary derived proteins

i) Coagulated proteins:-

These are the denatured proteins produced by agents such as heat, acids, alkalis etc. e.g. cooked proteins, coagulated albumin

ii) Protean:

These are the earliest products of protein hydrolysis by enzymes, dilute acids alkalis etc. which are insoluble in water e.g. fibrin formed from fibrinogen

iii) Metaproteins:

These are second stage products of protein hydrolysis obtained by treatment with slightly stronger acids and alkalis

b) Secondary derived proteins:

These are the progressive hydrolytic products of protein hydrolysis and include proteoses, peptones polypeptides

C) Nutritional Classification of proteins

The nutritive value of protein is determined by the composition of essential amino acids. From nutritional point of view proteins are divided into 3 types:

i) Complete proteins:-

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

ii) Partially incomplete proteins:-

These proteins are partially lacking one or more essential amino acids and hence can promote moderate growth e.g. wheat and rice proteins (limiting lys, Thr)

iii) Incomplete protein:

These proteins completely lack one or more essential amino acids and do not promote growth at all e.g. gelatin (lacks Trp) Zein (lacks Trp and lys)