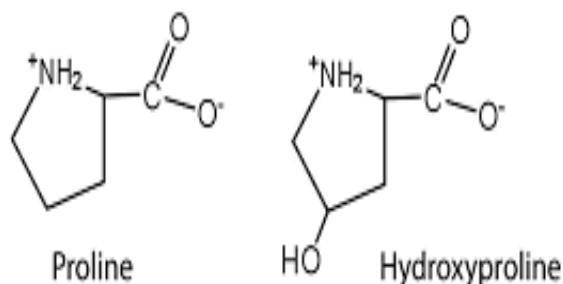
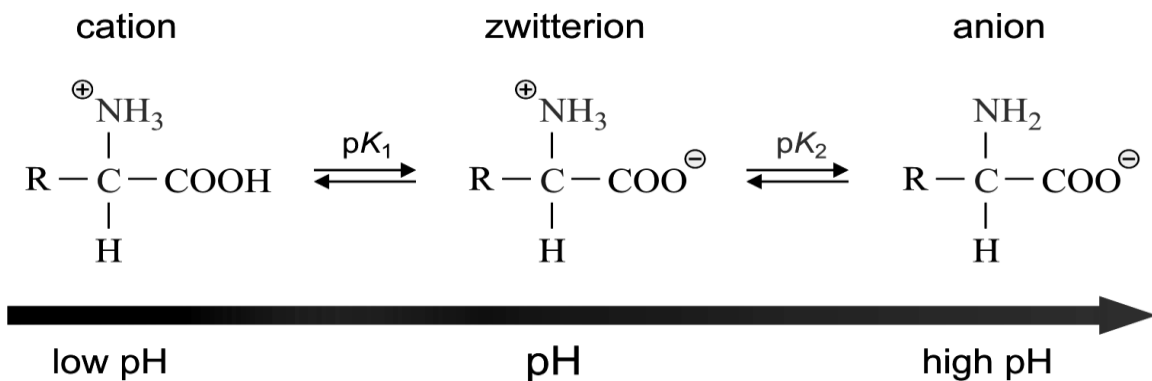
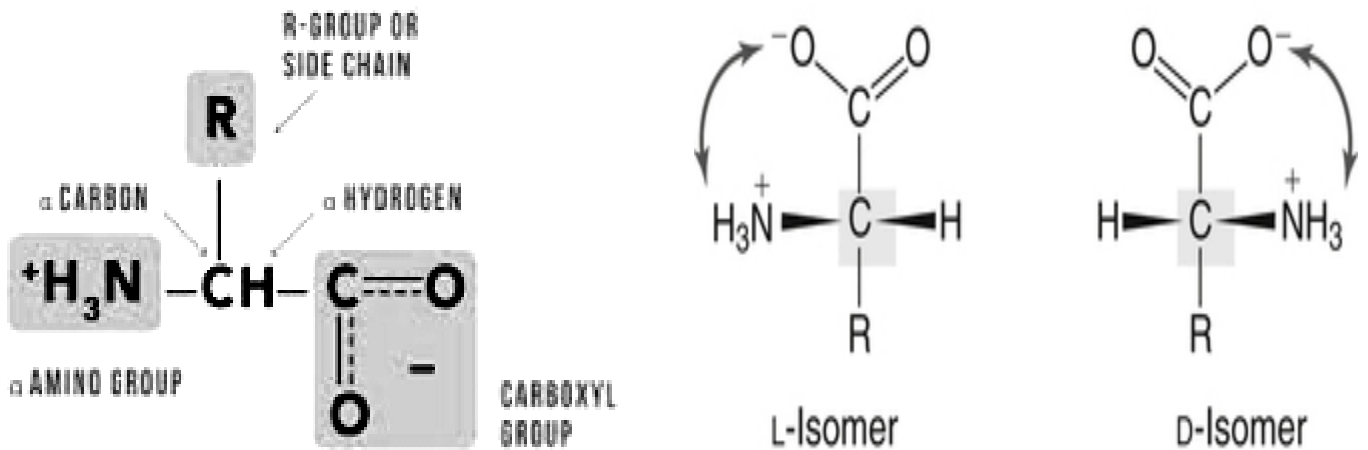
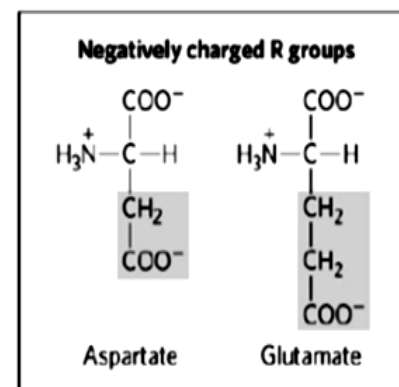
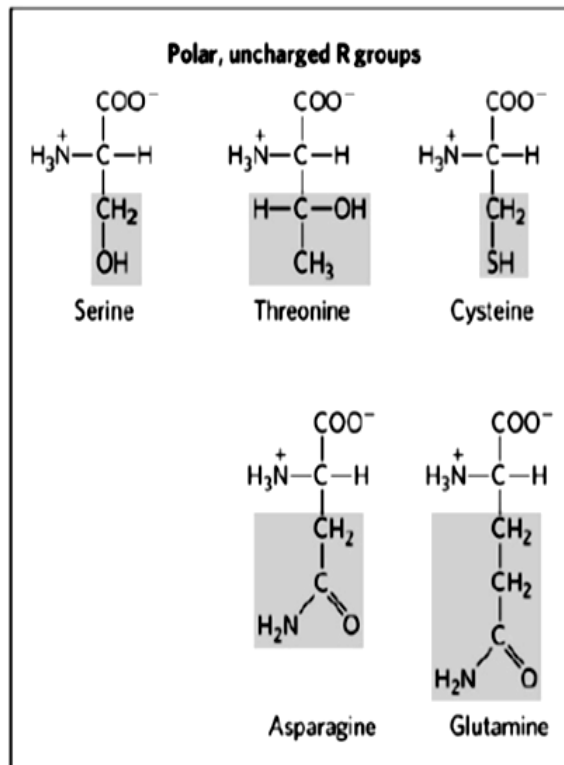
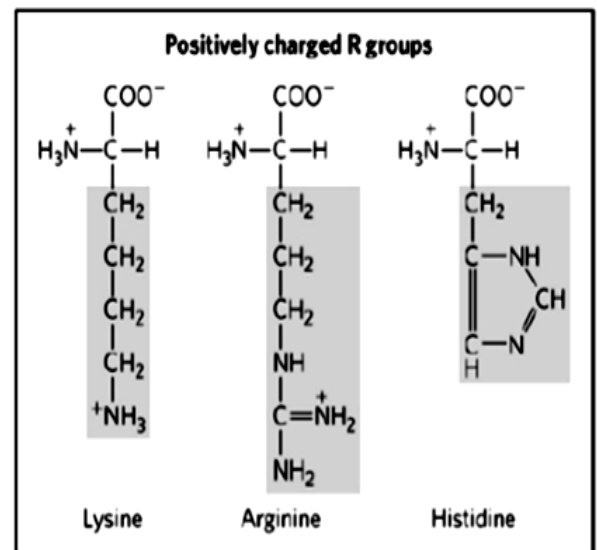
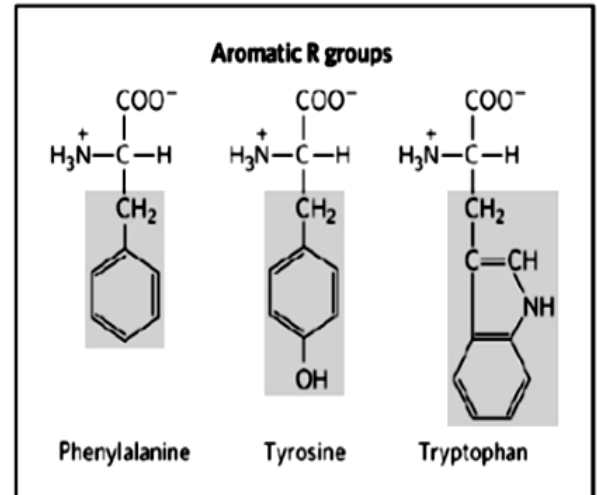
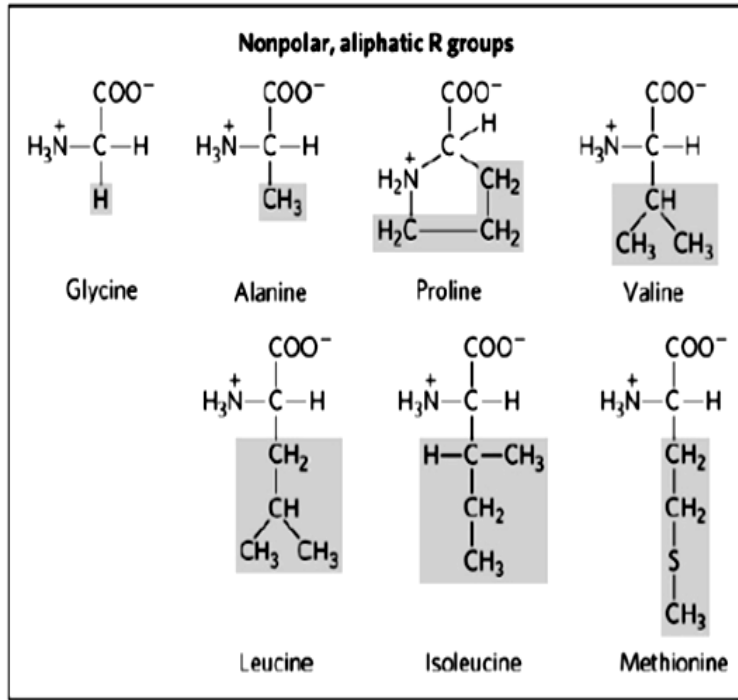


## Amino Acids

Amino acids are the structural units or building blocks of proteins. Each AA is a nitrogenous compound having an acidic carboxyl group (-COOH) and a basic amino group (-NH<sub>2</sub>), a hydrogen atom and a distinctive 'R' group or side chain all attached by covalent bonds to a carbon atom which is called an α-carbon. All AA except glycine have an asymmetric carbon resulting in 2 optically active forms—"D" and "L" forms. Only the "L" forms" are constituents of proteins. Amino acids at neutral pH are dipolar or **zwitter** ions. Proline and hydroxyproline have an imino group (NH) with a cyclic structure and are called imino acids.



Side chains (**R**) varying in size, shape, charge, hydrogen bonding capacity and chemical reactivity give rise to 20 different AA which are found in proteins. Based on nature of the attached **R** group these AA may be broadly classified as- **Non-polar and aliphatic, Polar and uncharged, Negatively charged, Positively charged and Aromatic** in nature.



## ESSENTIAL AND NON-ESSENTIAL AMINO ACIDS

Amino acids can be called the “building blocks” of protein and are an important part of every human body. There are 20 different amino acids – nine of which are called “essential” and 11 of which are labeled as “non-essential.” The human body needs all 20 of these amino acids, in varying degrees, to be healthy and fully functional. All 20 have distinct chemical structures and are used for different roles – such as forming neurotransmitters, forming hormones and producing energy. But their primary role is to build proteins. Protein is part of every single cell in the human body and is essential to the body’s functioning. Protein helps build and repair tissues like skin and muscle, and it helps produce antibodies and insulin. From only 20 amino acids, the body is able to generate many thousands of unique proteins with different functions. Each one of these proteins contains between 50 and 2000 amino acids, connected in varying sequences. After all of these amino acids are joined together, they are folded and twisted to make a specific shape. This unique shape is the determining factor for what the protein does for the body.

*From only 9 essential and 11 non-essential amino acids, the body is able to generate many thousands of unique proteins with different functions.*

### Essential Amino Acids

Essential amino acid, or indispensable amino acid, is an amino acid that cannot be synthesized *de novo* by the organism at a rate as per its demand, and thus must be supplied in its diet. These are the nine amino acids that cannot be create on its own, and which must obtain by eating various foods. Adults need to eat foods that contain the following nine amino acids **phenylalanine, valine, threonine, tryptophan, methionine, leucine, isoleucine, lysine, and histidine.**

Instead of storing up a supply of the essential acids, the body uses them to create new proteins on a regular basis. Therefore, the body needs a continuous supply of these amino acids to stay healthy.

### Non-Essential Amino Acids

The other type is the non-essential amino acid, 11 of which exist and are synthesized by the body- **alanine, arginine, asparagine, aspartic acid, cysteine, glutamic acid, glutamine, glycine, proline, serine, and tyrosine.** Thus, although they are an important part of building proteins, they do not need to be included in an everyday diet. Six of these non-essential acids are also known as “conditional,” meaning that the body may not be capable of producing enough of them when presented with substantial stress or illness These are **arginine, cysteine, glycine, glutamine, proline, and tyrosine..** The **limiting amino acid** is the essential amino acid found in the smallest quantity in the foodstuff, most plant-based foods have a limiting amino acid.

Essential	Conditionally essential	Non-essential
Histidine (H)	Arginine (R)	Alanine (A)
Isoleucine (I)	Cysteine (C)	Aspartic acid (D)
Leucine (L)	Glutamine (Q)	Asparagine (N)
Lysine (K)	Glycine (G)	Glutamic acid (E)
Methionine (M)	Proline (P)	Serine (S)
Phenylalanine (F)	Tyrosine (Y)	
Threonine (T)		
Tryptophan (W)		
Valine (V)		

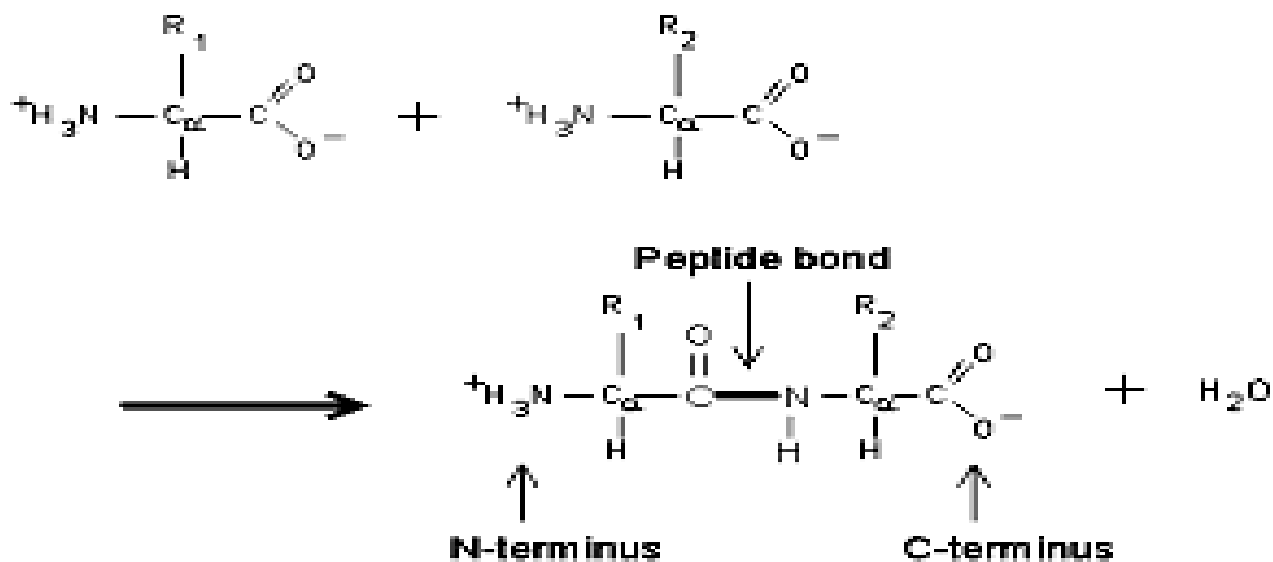
Protein source	Limiting amino acid
Wheat	lysine
Rice	lysine
Legumes	tryptophan
Maize	lysine and tryptophan
Pulses	methionine (or cysteine)
Egg, chicken, milk	none; egg is the reference for complete protein

**Complete proteins** contain a balanced set of essential amino acids for humans. Animal sources such as meat, poultry, eggs, fish, milk, and cheese provide all of the essential amino acids. Near-complete proteins are also found in some plant sources such as **quinoa**. The net protein utilization is affected by the **limiting amino acid** content (the essential amino acid found in the smallest quantity in the foodstuff), and somewhat affected by utilization of essential amino acids in the body. It is therefore important to mix foodstuffs that have different weaknesses in their essential amino acid distributions. This limits the loss of nitrogen through deamination and increases overall net protein utilization. Thus some of the most efficient pairings include sweet corn/tomatoes, apple/coconut and sweet corn/cherry. Certain traditional combinations of foods, such as corn and beans, or beans and rice, contain the essential amino acids necessary for humans in adequate amounts.

## PEPTIDES

**Peptides-** Amino acids are joined covalently to form peptide bonds. Polymerization results in the linkage between two AA which takes place due to the formation of a peptide bond between  $\alpha$ - carboxylic group and one amino acid to the  $\alpha$ - amino group of another. The formation of a peptide (a dipeptide) takes place by loss of water molecule from 2 amino acids. This is a condensation reaction in which  $\alpha$  -amino group of 2<sup>nd</sup> amino acid acts as nucleophile to displace the hydroxyl group of 1<sup>st</sup> amino acid. The OH group is not easily displaced and to make the reaction thermodynamically favorable, the -COOH group has to be activated. The peptide bond formation is an endergonic process and equilibrium favours the reaction. Hydrolysis of peptide is exergonic process but it is a slow because its activation energy is very high so peptide bonds are stable under intracellular conditions. When few amino acids are joined than -oligopeptide and when many join it is - poly peptide.

The formation of peptide bond is due to loss of one positive and one negative charge per peptide bond formed and therefore the peptide bond is uncharged at any physiologic pH. The peptide unit is planar and rigid because of partial double bond between carboxyl carbon and N atom of peptide unit. However there is large degree of rotational freedom on either side of rigid peptide unit.



Many amino acids (more than 100) are joined to form a polypeptide chain. A polypeptide chain has direction because it has 2 different ends  $\alpha$ -amino, and  $\alpha$ -COOH groups. An amino acid unit in a polypeptide chain is written by starting with amino terminal residue and ending with -COOH terminal. One is basic and the other is acidic and these two are the only ionizable groups of the peptide chain except those present in side chain. A polypeptide thus consists of a regularly repeating part called the main chain or backbone and the variable part comprising the side chain. The names are written by adding the suffix -yl except in last amino acid which is written full.

Peptides have different structure-

1. Linear peptides- with a free  $\text{-NH}_2$  and  $\text{+COOH}$  at the ends.
2. Branched peptides- where branching of one of several amino acid on a linear chain takes place either by  $\text{-COOH}$  or  $\text{-NH}_2$  groups of  $\omega$  carboxylic group of dicarboxylic amino acid or amino group of lysine.
3. Cyclic peptides- do not have N and C terminal ends.
4. Semi-cyclic peptides- have only one end if only N terminal end, the  $\alpha\text{-COOH}$  of last amino acid is linked with  $\text{-NH}_2$  of an endopeptide lysine, inversely when there is only one  $\text{-COOH}$  terminal end the  $\alpha\text{-NH}_2$  of first amino acid is linked with  $\omega\text{-COOH}$  of dicarboxylic amino acid.

Peptides may consist of same amino acids - homopeptides like triglycine, polyphenylalanine or may be heteropeptides.

### IMPORTANT PEPTIDES-

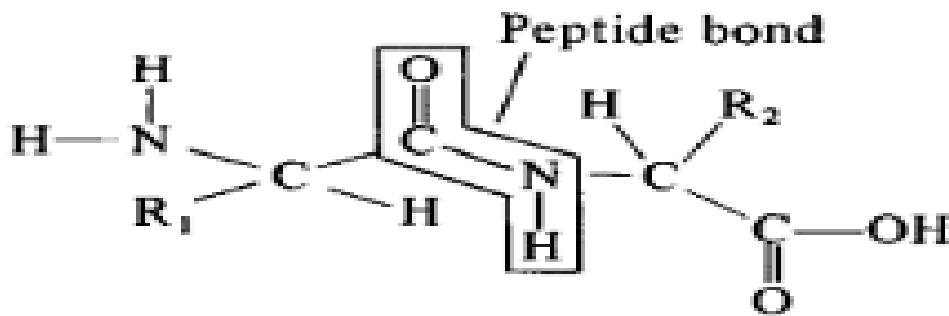
1. **Glutathione**- this is a tripeptide ( $\gamma\text{-L-glutamyl-L-cysteinyl-glycine}$ ) a linear peptide. It exists in reduced form (thiol) and oxidized form in which 2 molecules are linked by disulphide bridge and play a role in redox reactions.
2. **Peptide hormones**- e.g- pituitary hormones, oxytocin, vasopressin. These are cyclic due to a disulphide bridge and side chain.
3. **Insulin**- secreted by pancreas. It has 2 chains- A chain of 21 amino acid (free from basic amino acid) and B chain of 30 amino acid (with basic amino acid) 3 disulphide bridges, 2 inter-chain bridges and 1 intra-chain bridge.
4. **Penicillin**- has valine and cystine but besides peptide linkage other linkage are also present. They possess antibacterial activities present in fungus and bacteria

Peptides like amino acid can be distinguished by their ionization behavior showing characteristic titration curves and isoelectric pH.

## PROTEINS

The basic structural unit of proteins are amino acid however since protein is a macromolecule four basic structural levels of organizations are present:

1. **Primary structure:** this refers to the basic amino acid sequence in a polypeptide chain. Amino acid consists of an amino group, carboxyl group, a H atom and a distinctive R group bonded to 'C' atom which is called  $\alpha$ -carbon. Amino acids are linked by polypeptide chain. A peptide (-CONH) bond links  $\alpha$ -carboxyl group of one amino acid to the  $\alpha$ -amino group of other. A protein consists of one or more polypeptide chain. Each kind of protein has a unique amino acid sequence that is genetically determined. An amino acid unit in a polypeptide chain is called a residue.

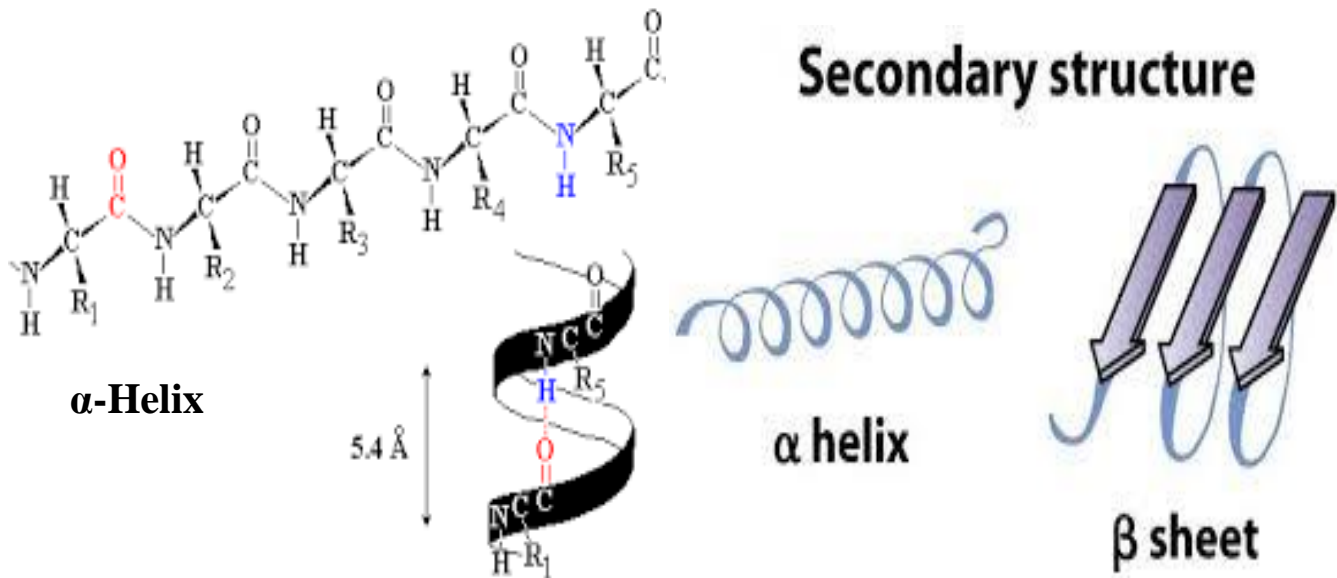


2. **Secondary structure:** many structural proteins eg. Silk fibroin are linear unfolded type having gly-alan – ser units. A polypeptide chain has a spatial organization. This refers to the regular folding of the chain due to presence of intramolecular H bonds. Folding and H bonding between amino acids results in a rigid stable structure called 1. Helix and 2. Sheets.

The secondary structure includes the polypeptide amino acid side chain because secondary structure is stabilized by H bonding between peptide imide and carboxyl groups of polypeptide and not the side chain. The peptide bond has a planar structure and the electrons are delocalized in amide linkage giving a double bond nature to C=N bonds. Since the C-N bond is double the  $\alpha$ -C to carboxyl C and imide N to  $\alpha$ -carbon bond result in rotation. The chain thus assumes either a helical configuration or two degenerate forms of a circle (very rare) or a zigzag structure. Of these, the most abundant are  $\alpha$ -helix, parallel  $\beta$  pleated sheets and anti- parallel  $\beta$ -pleated sheets.

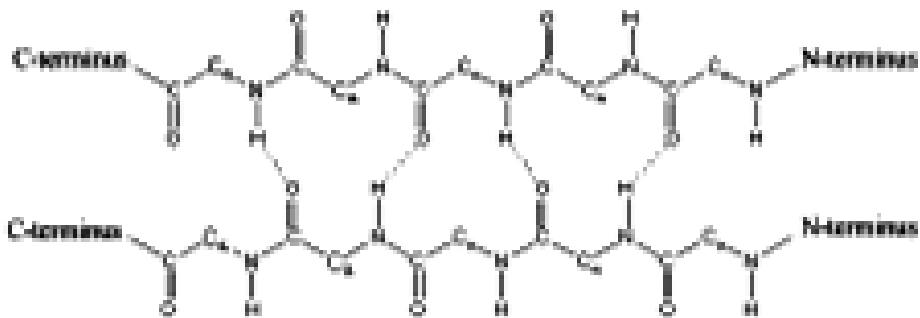
**$\alpha$ -helix** is a right handed helix i.e, the chain rotates clockwise. The  $\alpha$ -helix has 3.6 amino acids residue per turn and is stabilized by a nearly straight H bonding between an imide (NH) group and a carboxyl (COO) group at a position 4 residues away in the same chain.  $\alpha$ -keratin is example of fibrous protein.

Haemoglobin and myoglobin are example of globular protein.

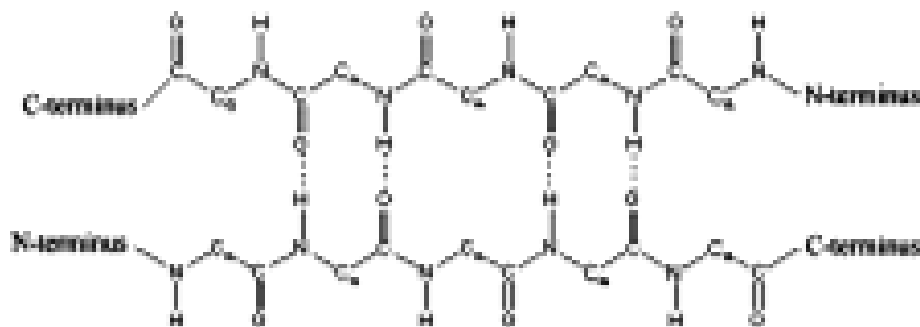


**$\beta$ -pleated sheets:** The atoms lie in folded plane and the side chains of succeeding residues along the polypeptide chain protrude alternatively above and below the general plain of the structure. There is a maximum H bonding between -NH and -CO group of different polypeptide chain and the H bonds are straight enough to be stable. They are not only pleated but slightly twisted. The polypeptide chain lies in a extended or  $\beta$ - form with the C=O and N-H group. Hydrogen atom are bonded to those of neighboring chain which maybe the same chain folded back or separate chain. It is named  $\beta$ -form because it was first founded in  $\beta$ -keratin. It is pleated because successive  $\alpha$ -carbon atoms of amino acids residue lie slightly above and below the plane of  $\beta$ -sheet. The adjacent polypeptide chain in same direction is parallel or opposite direction is anti-parallel.

### Parallel $\beta$ Sheet



### Antiparallel $\beta$ Sheet





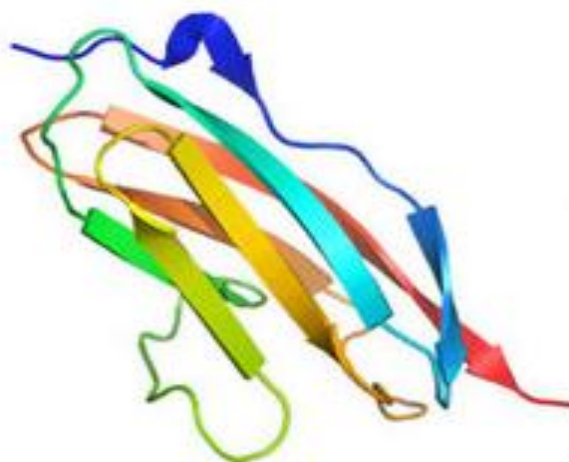
$\beta$ -pleated sheet differ from  $\alpha$ -helix in having a polypeptide arranged side by side resembling a sheet rather than a rod. The chain is fully extended in a zigzag fashion rather than being lightly coiled as in a helix. In **fibrous proteins parallel  $\beta$  sheet is  $\beta$ -keratin and antiparallel is silk.** In **globular proteins antiparallel –  $\beta$  sheet** are 6 stranded bovine pancreatic **Chymotrypsin** and 4 parallel stranded globular protein 4 stranded parallel  $\beta$  sheet in **Glutathione Reductase**.

**Super secondary structure-** In fibrous protein collagen 3 left handed  $\alpha$  helix coil in parallel around each other to form a right handed super secondary structure which is stabilized by inter chain H bond. This is mainly due to presence of glycine and no  $\alpha$  helix is formed. The special structure of fibrous protein is super secondary structure. Super secondary structures are also motif or folds are stable arrangement of several elements of secondary structure and their connections. Motifs are the basis of protein structural classification (SCOP) in which proteins are divided into 4 classes **all  $\alpha$** , **all  $\beta$** ,  **$\alpha/\beta$**  in which  $\alpha+\beta$  segments alternate and  **$\alpha+\beta$**  in which  $\alpha+\beta$  regions are segregated. Within each class are 10 to 100s of different folding arrangements. The top two level **class** and **folds** are purely structural. Proteins with significant primary sequence similarity or similar structure and function form **family**. eg., globin family. Two or more families having same major structural motif and functional similarities are grouped as **super families**.

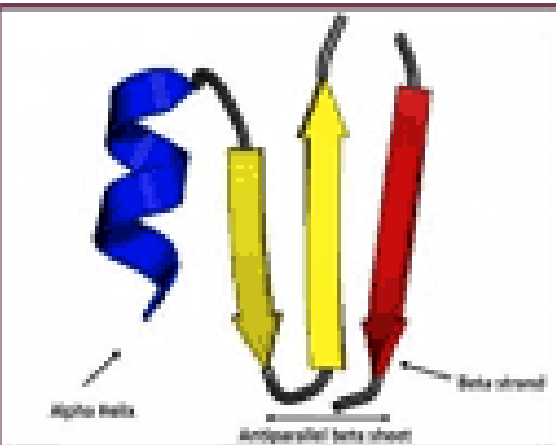
Polypeptides with more than a few hundred amino acid residues after fold into two or more stable globular units are called **domains**. In many cases a domain from a large protein will retain its correct 3-D structure even when it is separated from the remainder of the polypeptide chain. A protein with multiple domains may appear to have distinct globular lobe for each domain. Different domains have distinct functions. Two simple motifs the  **$\beta$ - $\alpha$ - $\beta$**  and  **$\alpha$ - $\alpha$**  corner are involved.



**All  $\alpha$ -helix**




**All  $\beta$ -sheets**




Folding patterns involving  $\alpha$ -helices and  $\beta$ -pleated sheets.

**SUPER SECONDARY STRUCTURES (MOTIFS)**


Certain groupings of secondary structural elements are called motifs.




$\beta$ - $\alpha$ - $\beta$  motif



Greek key motif



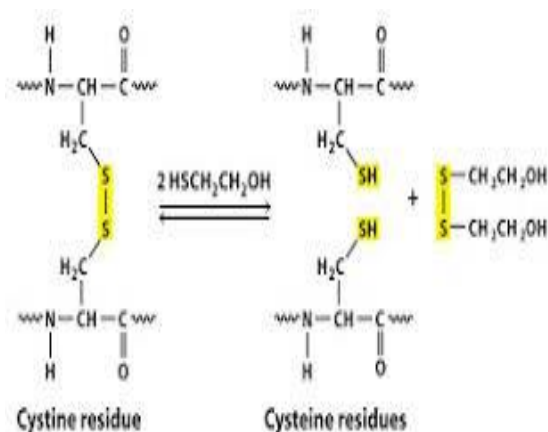
$\beta$  barrel motif

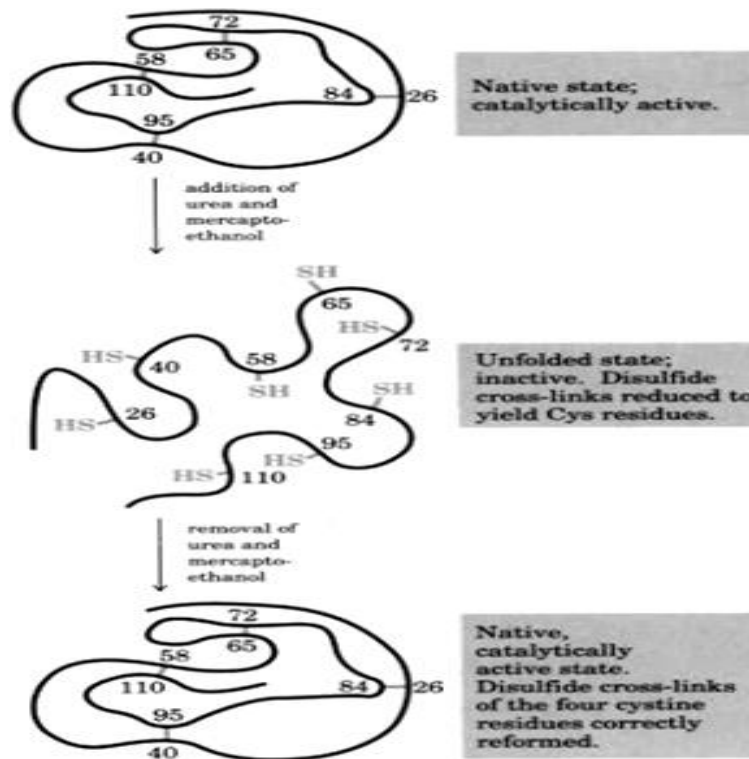


Beta sheet

$\beta$ - $\alpha$ - $\beta$  : an  $\alpha$ -helix connects two parallel strands of  $\beta$ -pleated sheets.★

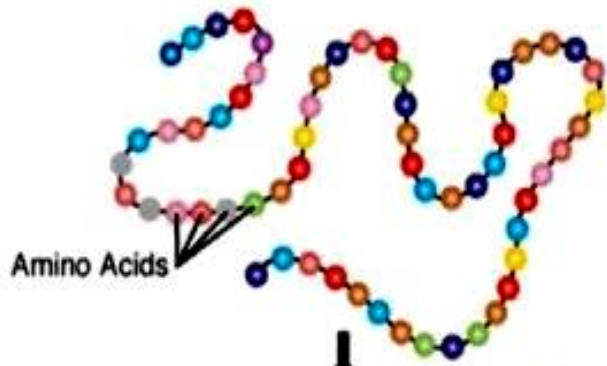
3. **Tertiary structure:** refers to the three dimensional structure especially the bonds between amino acids residue that are distant from each other in a polypeptide chain and the arrangement of secondary elements relative to one another. The term conformation refers to secondary and tertiary structure i.e, the folding pattern of polypeptide chain and all contacts between amino acid side chain of that polypeptide chain. The biological activity of protein molecule is due to its well defined 3-D structure. The steric relationship between amino acid residue that lie far part in a linear sequence and relative arrangements of the secondary structure elements to another in a polypeptide chain is the tertiary structure. The tertiary structure is stabilized by side chains of amino acids. e.g., Ribonuclease is a single polypeptide chain of 124 AA residues. It has 4 disulphide bonds which can be reversibly cleaved by reducing them with reagent like  $\beta$ -mercaptoethanol so that disulphide (cystine) all converts to sulfhydryl (cysteine). If urea is added the protein is totally denatured by unfolding. The reduced unfolded inactive RNase regains its enzyme activity when oxidised with air after removal of urea + mercaptoethanol to give S-S. Thus the four disulphide bond pairing gives the 3-D structure to RNase and determines its activity.





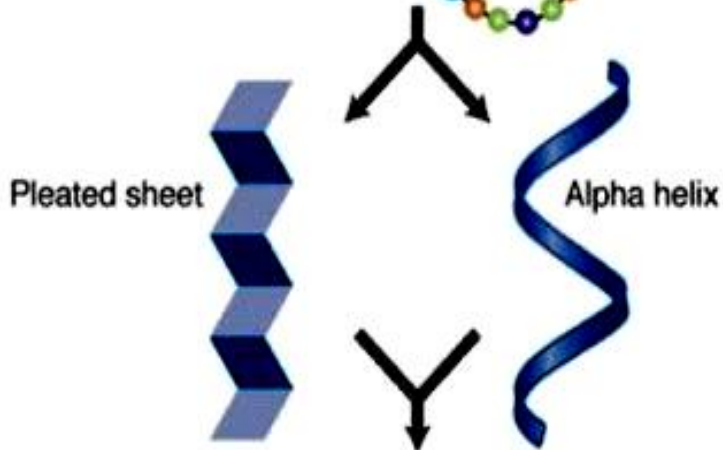
4. **Quarternary structure:** Oligomeric proteins that contains more than one polypeptide chains have an additional structure which refers to the way in which the chain are folded together in relation to each other. Each polypeptide subunit or protomer (repeating subunit or a group of subunit) has its own primary, secondary and tertiary structure (a multi subunit protein is also referred to as **multimer**.). Thus quarternary structure defines the structure resulting from interaction between separate polypeptide units of protein containing more than one subunit. It maybe **homogeneous** when subunits have similar structure example phosphorylase which has 2 sub identical subunits that are inactive separately but when they join as dimers they become active. **Heterogeneous** quaternary structure when subunits are dissimilar example hemoglobin has two identical  $\alpha$ -chains and  $\beta$ -chains i.e. protomers are  $\alpha_2 \beta_2$ ; Ribulose bisphosphate carboxylase has 8 small and 8 large subunits.

**Random coil** or loop region- of a polypeptide chain is neither random nor a coil. The term indicates a section of polypeptide in a protein whose conformation is not responsible as one of defined structure above. A more common term is **connecting loop region**. The structure of random coil in a protein is mainly determined by side chain interactions and it is fixed within a protein. The structure maybe along conformation as determined by various group interactions in protein structure despite named coil. Since structure of random coil does not satisfy the H bonding potentials of C=O and N-H group of the backbone or those of side groups, such section are found at the exterior of proteins, in contact with water.

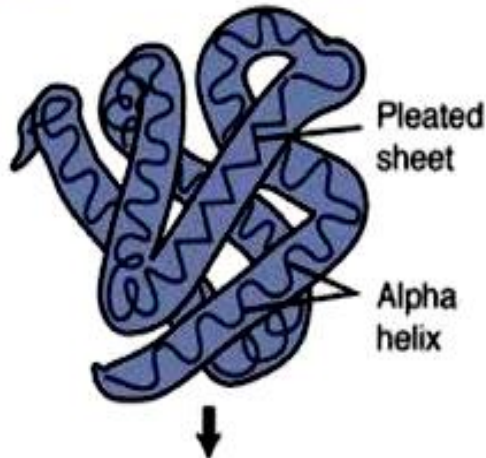


## Levels of protein organization

**Primary protein structure**  
is sequence of a chain of amino acids



**Secondary protein structure**  
occurs when the sequence of amino acids are linked by hydrogen bonds



**Tertiary protein structure**  
occurs when certain attractions are present between alpha helices and pleated sheets.



**Quaternary protein structure**  
is a protein consisting of more than one amino acid chain.

## CLASSIFICATION OF PROTEIN

### A. Fibrous Protein

**Scleroproteins-** insoluble in water, dilute acids, bases and salts. e.g animal fibres like collagen, keratin, elastin, fibroin.

### B. Globular Protein

1. **Simple proteins-** this group includes proteins containing only amino acids as structural components.

On decomposition with acids they liberate constituent acid. On the basis of solubility these may be

a) **Protamines and histones-** possess a simple structure, low molecular weight (approx 5000), water soluble, heat resistant, strongly basic due to high content of basic amino acids (lysine, arginine) form salts with mineral acids and nucleic proteins. Mainly found in animal sperm cells etc.

b) **Albumins-** abundant in seeds, water soluble, soluble in dilute solution of acids and bases or salts, heat labile (coagulation) precipitated by saturated solution of acid salt  $(\text{NH}_4)_2\text{SO}_4$  or neutral salt  $\text{Na}_2\text{SO}_4$  e.g., **leucosine in cereals, legumelline in legumes.**

c) **Globulin- soluble in water are** pseudoglobulin (milk whey) and insoluble are euglobulin (more common), coagulated by heat, precipitated by solution of  $(\text{NH}_4)\text{SO}_4$ . **Euglobulin e.g., glycine (soybean) legumine (peas).**

d) **Glutelins-** isolated from plant seeds, insoluble in water, dilute salt and alcohol solutions, soluble in dilute acids and alkaline, coagulated by heat e.g., **glutenin (wheat), glutelin (maize, oryzin (rice).**

e) **Prolamines-** insoluble in water and dilute salt solution but soluble in dilute acid and alkaline and alcohol (60-80%), not coagulated by heat e.g **gliadin (wheat), zein (corn), hordein (oat).**

2. **Conjugated or Complex proteins:-** these are linked with a separable non protein part or prosthetic group. This prosthetic group may be metal or an organic compound. On decomposition they liberate amino acids and prosthetic group.

a) **Metalloproteins-** proteins linked with metals. Maybe stable or labile.

**Metals strongly bound by proteins-** some heavy metals like Hg, Ag, Cu, Zn are strongly bound to protein by  $-\text{SH}$  group of side chain. Some have strong affinity for Fe (siderophilin) and Cu (ceruloplasmin).

**Metals loosely bound-** e.g Ca.

**Metals which do not bind with proteins-** Na and K form compounds with nucleic acids by electrostatic bonds.

- b) **Chromoproteins**- these are proteins coupled with colored pigments. Chlorophyll is present in leaf cells as chloroplastin which readily dissolves in water and is readily denatured. eg haem protein in haemoglobin, cytochromes, flavoproteins- flavin nucleic acid (FAD).
- c) **Glycoprotein**- egg albumin.
- d) **Mucoprotein**- large amount of carbohydrates e.g mucin
- e) **Phospholipids**- casein, vitellin
- f) **Lipoprotein**- lipovitellin.
- g) **Nucleoprotein**- nucleohistones from nuclei.

### **BIOLOGICAL FUNCTIONS OF PROTEINS**

1. **Enzymes**- most varied and highly specialized proteins with catalytic activity are the enzymes.
2. **Transport proteins**- bind and carry specific molecule or ions from one organ to another e.g lipoproteins, ceruloplasmin (Cu), siderophilin (Fe).
3. **Nutrient storage**- seeds of many plants store nutrient proteins. Ovalbumin, egg white, casein milk, ferritin- Fe protein plant, haemoglobin.
4. **Contractile or motile protein**- Actin and myosin in contractile system. Tubulin is protein from which micro tubule is built.
5. **Structural proteins**- Supporting filaments gives structural strength or protection. Fibrous protein e.g.- collagen, cartilage, elastin in ligaments, hair, finger nails- keratin. Silk fibers and spider web- fibroid, resilin in wings of insect.
6. **Defense proteins**- The immunoglobulins or antibodies which can recognize and precipitate invading viruses , bacteria and foreign protein e.g. fibrinogen, thrombin, snake venom, ricin- plant toxic protein have defensive function.
7. **Regulatory protein**- Help to regulate cellular or physiological activity. eg.- hormones in cellular response to many hormonal signals is mediated by a loss of ATP binding protein called G- proteins. Other regulatory proteins bind to amino acid and regulate the biosynthesis of enzymes. t- RNA molecule involved in cell division.
8. **Other proteins**- Function cannot be classified. Monellin- African plant sweet protein. Also antifreeze protein in Antarctic fish.