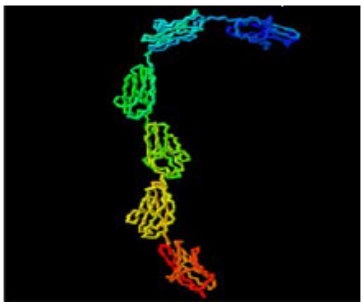
Protein

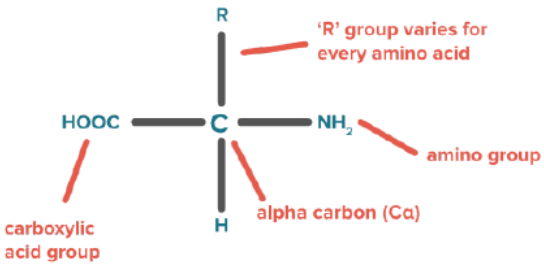
What is protein?

Twenty percent of the human body is made up of proteins. Proteins are the large, complex biomolecules that are critical for normal functioning of cells. They are essential for the structure, function, and regulation of the body's tissues and organs. Proteins are made up of smaller units called amino acids, which are building blocks of proteins. They are attached to one another by peptide bonds forming a long chain of proteins. Therefore, Proteins are chains of amino acids that assemble via peptide linkages.



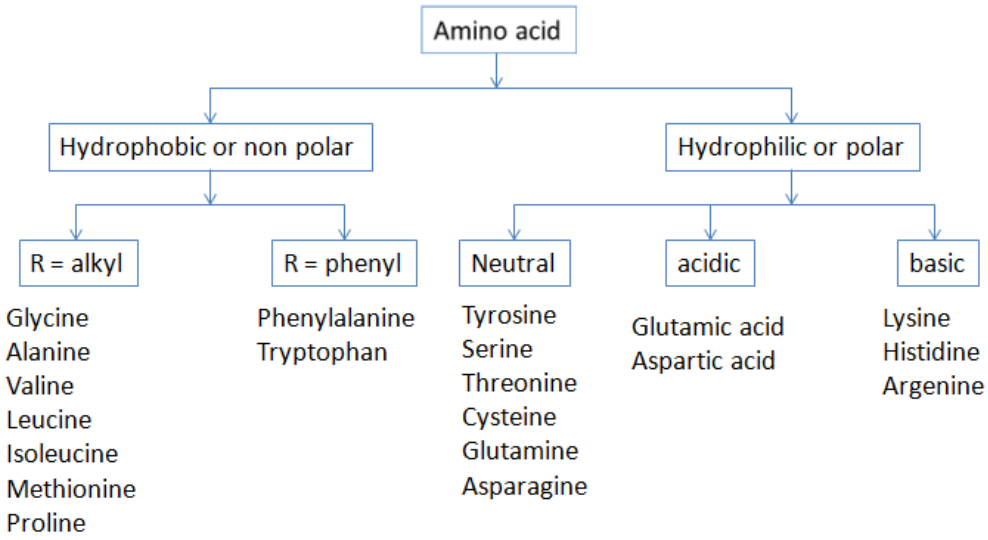
Amino acid

An amino acid is an organic molecule that contains both a carboxylic group and an amino group. Amino acids that have an amino group bonded directly to the alpha-carbon are referred to as alpha amino acids. Every alpha amino acid has a carbon atom, called an alpha carbon, C_α, bonded to a carboxylic acid group (–COOH), an amino group, ($–NH_2$), a hydrogen atom and an R group that is unique for every amino acid.



Although there are hundreds of amino acids found in nature, only about 20 amino acids are needed to make all the proteins found in the human body and most other forms of life. All of them, except for glycine, contain a chiral alpha carbon. These 20 alpha-amino acids are all L-isomer, with an R-absolute configuration except for glycine (no chiral center) and cysteine (S-absolute configuration, because of the sulfur-containing R-group). Along with the 20 standard amino acids there are two amino acids named as selenocysteine and pyrrolysine are considered the 21st and 22nd amino acids, respectively. They may become incorporated into protein chains during ribosomal protein synthesis. Pyrroloysine has functionality in life, however, humans do not use pyrrolysine in protein synthesis.

Based on the nature of their 'R' group, amino acids are classified based on their polarity as:



Glycine = NH_2-CH_2-COOH

Phenylalanine = $H_2N-\underset{\displaystyle CH_2C_6H_5}{\overset{\displaystyle H}{C}}-COOH$

Tyrosine = $H_2N-\underset{\displaystyle CH_2C_6H_5\text{-}OH}{\overset{\displaystyle H}{C}}-COOH$

Glutamicacid = $H_2N-\underset{\displaystyle CH_2CH_2COOH}{\overset{\displaystyle H}{C}}-COOH$

Lysine = $H_2N-\underset{\displaystyle CH_2CH_2CH_2CH_2NH_2}{\overset{\displaystyle H}{C}}-COOH$

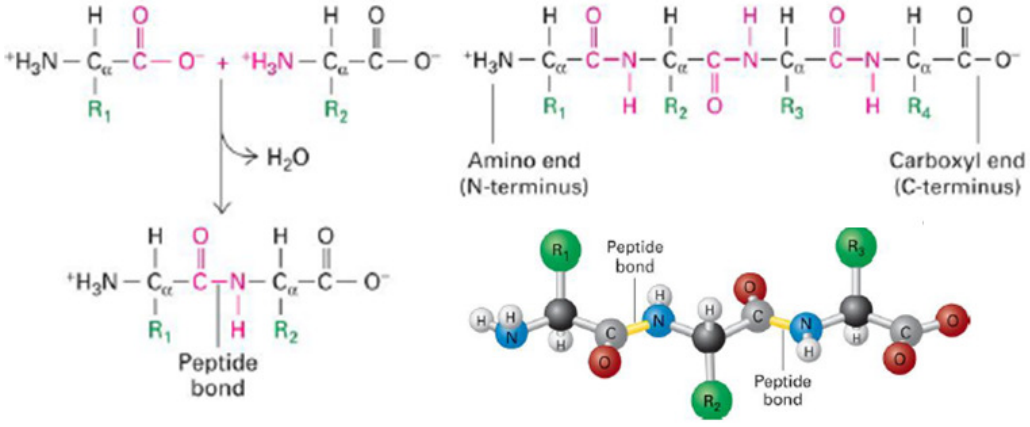
Essential and non-essential amino acid:

In human, there are only 21 proteinogenic amino acids (required for protein synthesis), the 20 of the standard genetic code, plus selenocysteine. Out of 21, humans can synthesize 12 amino acids, from each other or from other molecules of intermediary metabolism. So, these are not an essential part of the diet, called non-essential amino acid.

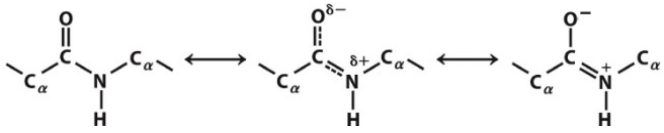
The other nine must be consumed (usually as their protein derivatives) because our body cannot make them and so they are called essential amino acids. The essential amino acids are histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine (i.e. H, I, L, K, M, F, T, W, V).

Peptide bond:

Amino acids are linked together by 'amide groups' called peptide bonds. During protein synthesis, the carboxyl group of amino acid at the end of the growing polypeptide chain reacts with the amino group of an incoming amino acid, releasing a molecule of water. The resulting bond between the amino acids is a peptide bond.

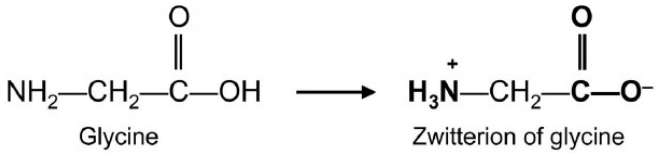


The peptide bond is a resonance hybrid of two canonical structures. The resonance causes the peptide bonds to be less reactive compared to esters, to be quite **rigid** and nearly **planar** and to exhibit a large dipole moment in the favoured trans configuration.



Isoelectric point:

The isoelectric point (IP) is the pH at which the amino acid has an overall zero charge that means at isoelectric point no net migration takes place in an electric field. It is an important characteristic for any amino acid, because every amino acid has at least two acid–base groups. At isoelectric point, amino acids exist as zwitterions. A zwitterion is a molecule that has equal positive and negative regions of charge. Using this property, amino acids and proteins are isolated or separated. The isoelectric points (IP) of amino acids range from 2.8 to 10.8. For example: IP of glycine = 6.0; IP of aspartic acid = 2.8 and IP of lysine = 9.7 etc.



Separation of proteins at the isoelectric point is called isoelectric focusing. In isoelectric focusing a gradient of pH and an electric potential are applied across the gel, making one end more positive than the other. Separation occurs on the basis of the positive or negative groups present on the molecule. If they are positively charged, they will be pulled toward the more negative end of the gel and if they are negatively charged they will be pulled to the more positive end of the gel. The proteins applied in the first dimension will migrate through the gel and will accumulate at their isoelectric point. At this stage the protein net charge is zero and therefore does not move in an electric field.

Classification of Protein:

A) Based on the chemical nature and composition proteins are classified as:

1) Simple proteins: They are composed of only amino acid residue. On hydrolysis these proteins yield only constituent amino acids. For example: albumin(Serum albumin and ovalbumin), Globulins(Serum globulin, fibrinogen, myosin of muscle), Prolamins(Gliadin of wheat and zein of corn), Histones(globin of haemoglobin and nucleoproteins), Protamines etc.

2) Conjugated or compound proteins: Simple proteins combined with some non-protein substances known as prosthetic groups. For example,

Nucleoprotein: Simple basic proteins (protamines or histones) in salt combined with nucleic acids as the prosthetic group. e.g. ribosomes.

Mucoproteins: simple proteins in combination with carbohydrates like mucopolysaccharides, which include hyaluronic acid and chondroitin sulphates.

Chromoproteins: proteins containing coloured prosthetic groups e.g., haemoglobin, flavoprotein and cytochrome.

Phosphoprotein: proteins containing phosphoric acid.

Lipoprotein: proteins conjugated with lipids such as neutral fat, phospholipids and cholesterol.

Metalloprotein: metal-binding proteins. E.g. β-globulin, termed transferrin is capable of combining with iron, copper and zinc etc.

3) Derived proteins: They are derivatives or degraded products derived by partial to complete hydrolysis from simple and conjugated proteins by the action of acids, alkalis or enzymes. They may be:

Primary derived protein: These protein derivatives are formed by processes causing only slight changes in the protein molecule and its properties.

Examples: Proteans (myosan from myosin, fibrin from fibrinogen), Metaproteins, Coagulated proteins (cooked meat and cooked albumin).

Secondary derived proteins: proteins are formed in the progressive hydrolytic cleavage of the peptide bonds of protein molecule. They are roughly grouped into proteoses, peptones and peptides according to average molecular weight. The complete hydrolytic decomposition of the natural protein molecule into amino acids generally progresses through successive stages as follows:

Protein → Protean → Metaprotein

Proteoses → Peptones → Peptides →amino acids

b. Classification of proteins based on function

Proteins are classified based on their functions as:

Catalytic proteins – Enzymes

They enhance the reaction rates a million fold, function as biocatalysts.

Regulatory proteins - Hormones

These are polypeptides and small proteins play highly important regulatory role in maintaining order in complex metabolic reactions e.g., growth hormone, insulin etc.

Protective proteins - Antibodies

These proteins have protective defence function, fight against certain diseases. e.g., immunoglobulin. Fibrin present in the blood is also a protective protein.

Storage proteins

These proteins have the function of storing amino acids as nutrients and as building blocks for the growing embryo. The major storage protein in pulses is globulins and prolamins in cereals. In rice the major storage protein is glutelins.

Transport proteins: Some proteins are capable of binding and transporting specific types of molecules through blood. For example, haemoglobin binds with oxygen and transport through blood to various tissues. Myoglobin, a related protein, transports oxygen in muscle. Lipids bind to serum proteins like albumin and transported as lipoproteins in the blood.

Toxic proteins: Some of the proteins are toxic in nature. Enzyme inhibitors such as trypsin inhibitor bind to digestive enzyme and prevent the availability of the protein. Lectin, a toxic protein present in legumes, agglutinates red blood cells. Snake venom is protein in nature.

Structural proteins: These proteins serve as structural materials or as important components of extra cellular fluid. Examples of structural proteins are myosin of muscles, keratin of skin and hair and collagen of connective tissue.

Contractile proteins: Proteins like actin and myosin function as essential elements in contractile system of skeletal muscle.

Secretary proteins: Fibroin is a protein secreted by spiders and silkworms to form webs and cocoons.

Exotic proteins: Antarctic fishes live in minus temperature waters, well below the temperature at which their blood is expected to freeze. These fishes are prevented from freezing by antifreeze glycoproteins present in their body.

C. Classification based on structure

Globular proteins are mostly water-soluble and fragile in nature e.g., enzymes, hormones and antibodies. For example, Albumin, Globulin, Glutelin, Histones etc.

Fibrous proteins are tough and water-insoluble. They are used to build a variety of materials that support and protect specific tissues, e.g., skin, hair, fingernails. For example, Keratin, Elastin, Collagen.

Biological Function of Protein:

Proteins are vital for the growth and repair, and their functions are endless. They also have enormous diversity of biological function and are the most important final products of the information pathways.

1) Proteins, which are composed of amino acids, serve in many roles in the body (e.g., as enzymes, structural components, hormones, and antibodies).

2) They act as structural components such as keratin of hair and nail, collagen of bone etc.

3) Proteins are the molecular instruments through which genetic information is expressed.

4) They execute their activities in the transport of oxygen and carbon dioxide by haemoglobin and special enzymes in the red cells.

5) They function in the homeostatic control of the volume of the circulating blood and that of the interstitial fluids through the plasma proteins.

6) They are involved in blood clotting through thrombin, fibrinogen and other protein factors.

7) They act as the defence against infections by means of protein antibodies.

8) They perform hereditary transmission by nucleoproteins of the cell nucleus.

9) Ovalbumine, glutelin etc. are storage proteins.

10) Actin, myosin act as contractile protein important for muscle contraction.

Structure of proteins

The sequence of amino acid in a protein is determined by the DNA of the gene. The codon embedded in DNA as sequence of nucleo-bases actually determines the sequence of amino acid in a protein. A change of the gene sequence may lead to a change in the amino acid sequence of the protein that affects the protein's overall structure and function.

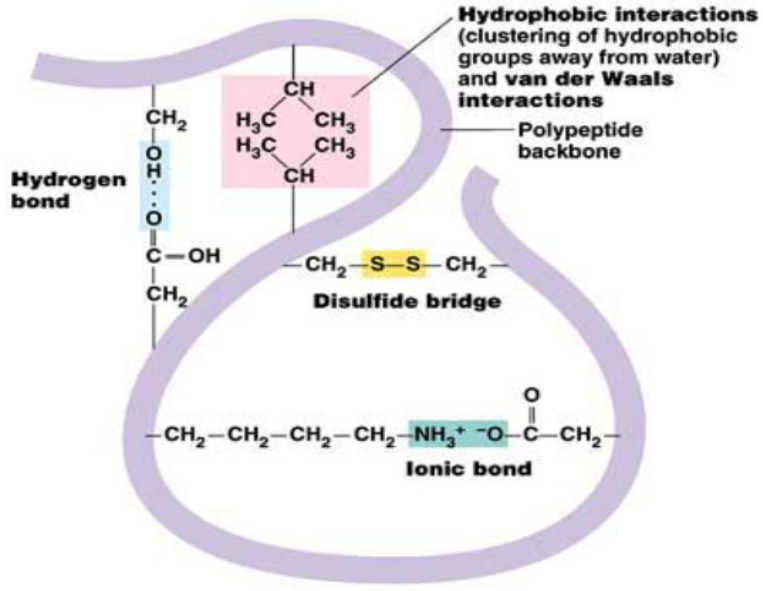
Unlike most organic polymers, protein molecules adopt a specific three-dimensional conformation. This structure is called the native fold responsible to fulfil a specific biological function. There are four levels of protein structure: primary, secondary, tertiary, and quaternary. Different types of interactions exist within these native structures.

Hydrophobic effect: Release of water molecules from the structured solvation layer around the molecule as protein folds increases the net entropy.

Hydrogen bonds: Interaction of N-H and C=O of the peptide bond leads to local regular structures such as α-helices and β-sheets.

London dispersion: Medium-range weak attraction between all atoms contributes significantly to the stability in the interior of the protein.

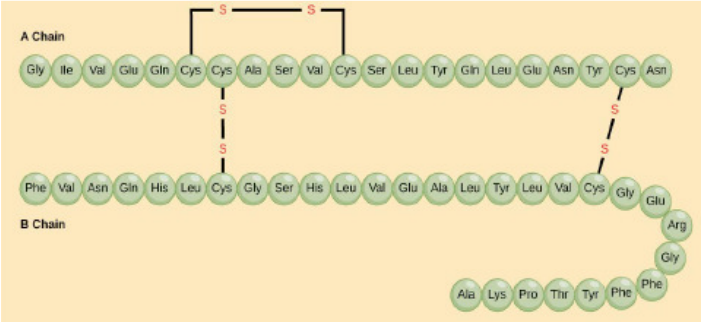
Electrostatic interactions: Long-range strong interactions between permanently charged groups.



Primary Structure

This is simplest level of protein structure. The **primary structure** of a protein is the particular **sequence** of amino acids that form the backbone of a peptide chain or protein.

The hormone insulin has two polypeptide chains A, and B linked by disulfide bonds. The sequence of the A chain with 21 amino acids and a B chain with 30 amino acids can be considered as an example for primary structure.



Secondary structure

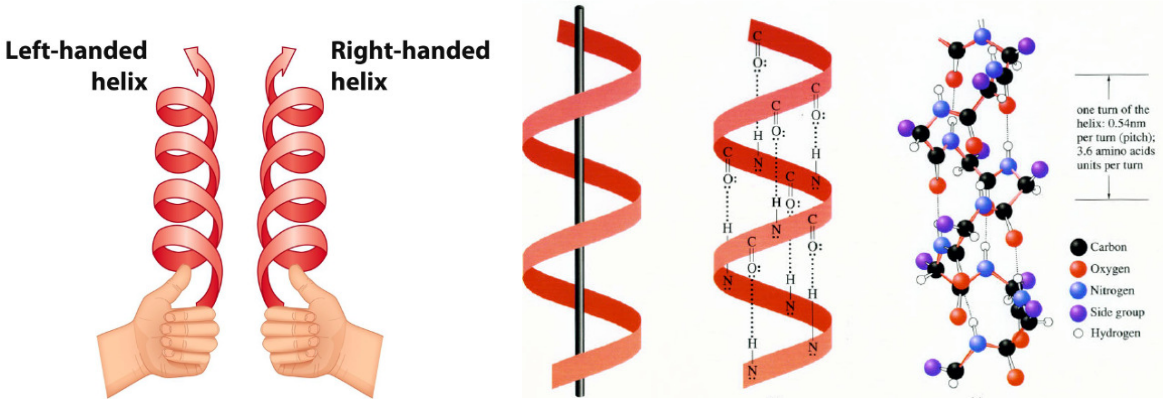
The secondary structures of proteins indicate the three-dimensional spatial arrangements of the polypeptide chains. It has local folded structures that form within a polypeptide due to H-bonding interaction. The two most common types of secondary structures are the α-helix and the β-pleated sheet. Both structures are held in shape by hydrogen bonds, which form between the carbonyl O of one amino acid and the amino H of another. The Irregular arrangement of the polypeptide chain is called the **random coil or extended chain.**

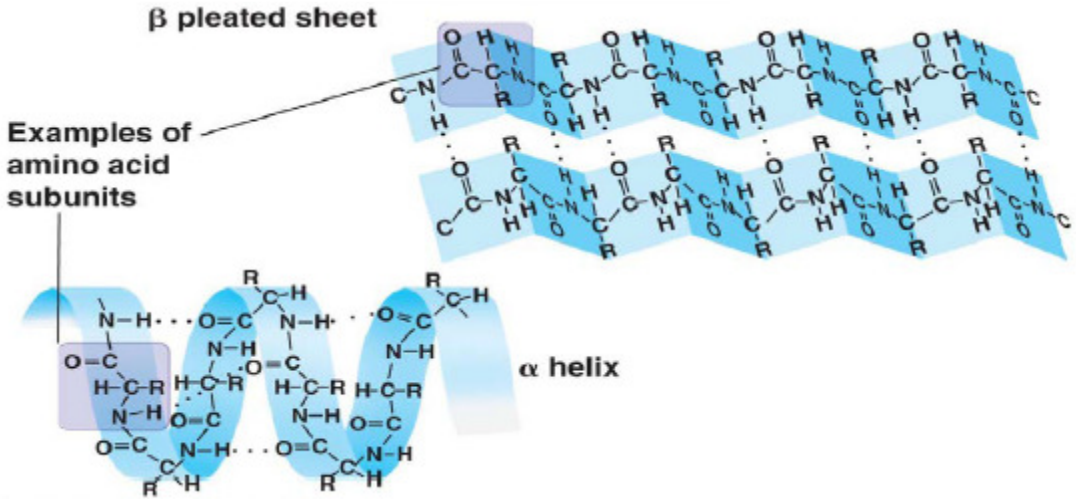
Characteristics of α-Helix

A ***α*-helix** has a coiled shape one polypeptide chain. All peptide bonds have a similar orientation and are aligned roughly parallel with the helical axis. The atoms of main chain remain on the inside and R-group of the side chains are projected outwards. The inner diameter of the helix (no side chains) is about 4–5 Å whether the outer diameter of the helix (with side chains) is 10–12 Å. The helix is stabilized by hydrogen bonds. The Carbonyl (C = O) of each amino acid is H-bonded to the amide (N-H) of the amino acid and it is in between

n^{th} amino acids with $(n+4)^{th}$ toward the C-terminus e.g. amino acid 1 is H-bonded to amino acid 5. The α-helices have sidedness i.e. n+4 are on the same side of the helix. There are 3.6 amino acid residues per turn. Each turn has length 5.4 Å called pitch of the helix. The α-helix has a large macroscopic dipole moment and is projected from positively charged N-terminus to negatively charged C-terminus. Helices can be right or left handed. The proteins are right handed.

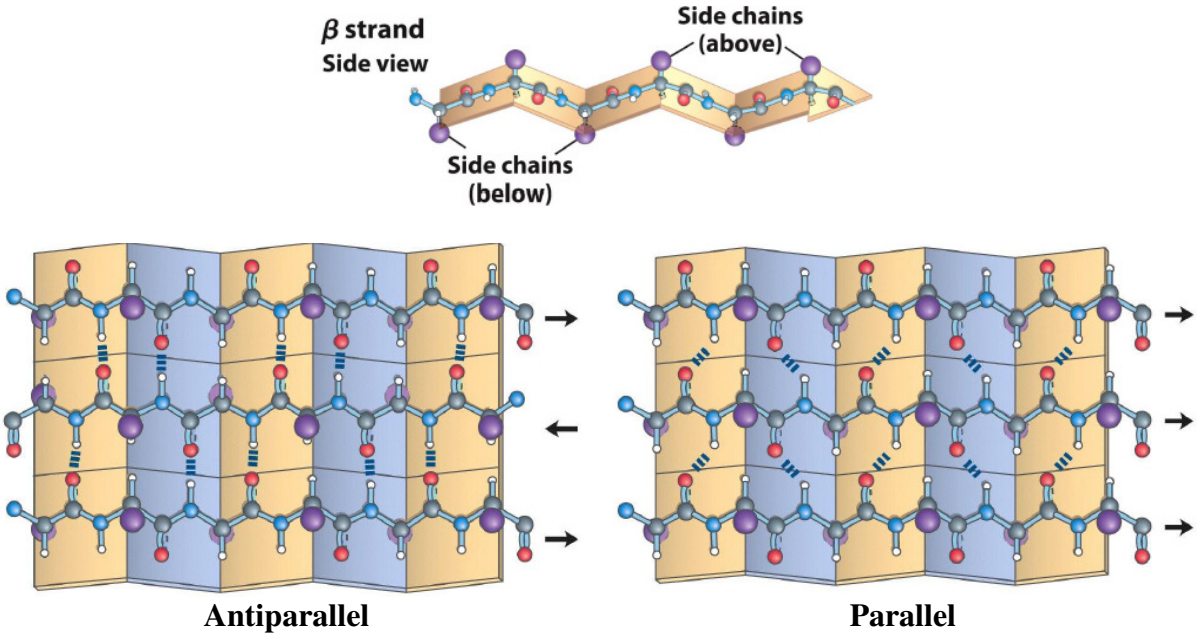
Not all polypeptide sequences adopt α-helical structures. Small hydrophobic residues such as alanine and leucine are strong helix formers where as proline acts as a helix breaker because the rotation around the N-C_α bond is impossible and glycine also acts as a helix breaker because the tiny R-group supports other conformations. Too many + or – charged groups near each other in space are unfavourable for helical structure because of electrostatic repulsion or attraction. α-helix observed in α-keratin of hair, feathers and nails.





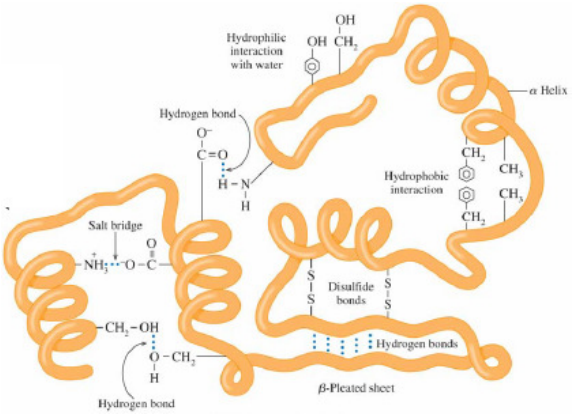
Characteristics of β-Sheets

A ***β*-pleated sheet** is a secondary structure that consists of polypeptide chains arranged side by side. The planarity of the peptide bond and tetrahedral geometry of the α-carbon create this pleated sheet-like structure. The backbone of the sheet is held together by hydrogen bonds between the amides in different strands of one or more polypeptide chains from inter-chain or intra-chain. These adjacent strands can come from completely different section of polypeptide chain or even different polypeptides i.e. β-sheets are composed of long range interaction. In β-sheets, peptide backbone is almost completely extended with R-groups stick up and down from β-sheets. It is stabilized by hydrogen bonds between the carbonyl of each amino acid with the NH of another amino acid and H-bond is nearly perpendicular to direction of peptide backbone. The adjacent chains can be parallel or anti-parallel. In parallel β–sheet, H-bonded strands run in the same direction where as in anti-parallel β–sheet, H-bonded strands run in opposite directions. Each stand is usually 6-12 residues long. Silk is made from a β-pleated sheet.



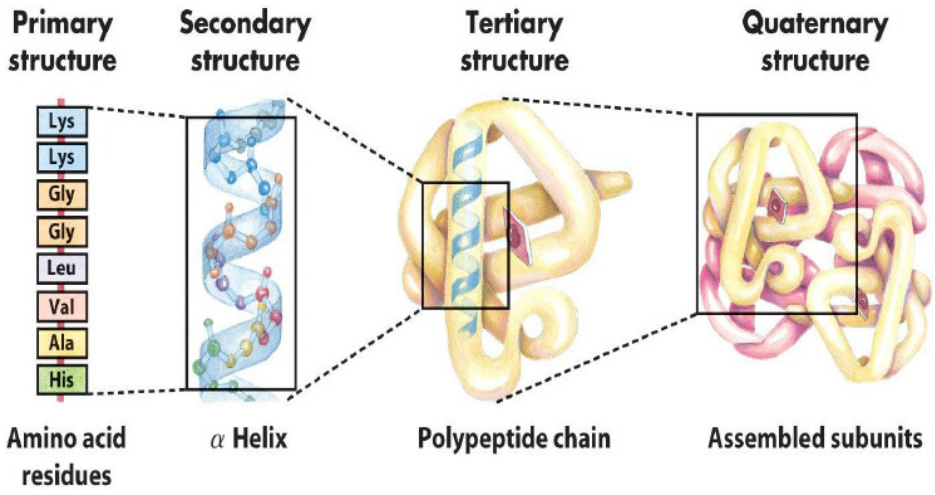
Tertiary Structure

The overall three-dimensional structure of a polypeptide is called its tertiary structure stabilized by numerous weak interactions and cross links between amino acid side chains. The tertiary structure is primarily due to hydrophobic interactions between the R groups of the amino acids that make up the protein. Generally, the nonpolar, hydrophobic R groups cluster together on the inside of the protein, leaving hydrophilic amino acids on the outside to interact with surrounding water molecules. Disulfide bonds between the sulfur-containing side chains also stabilized the tertiary structure.



Quaternary Structure

The quaternary structure is the combination of two or more tertiary units. An example for quaternary structure is haemoglobin. The haemoglobin carries oxygen in the blood and is made up of four subunits, two each of the α and β types.



Isolation: The process by which the proteins from the cells or tissues are recovered for the analysis purpose is known as protein isolation or protein extraction.

Necessity of extraction:

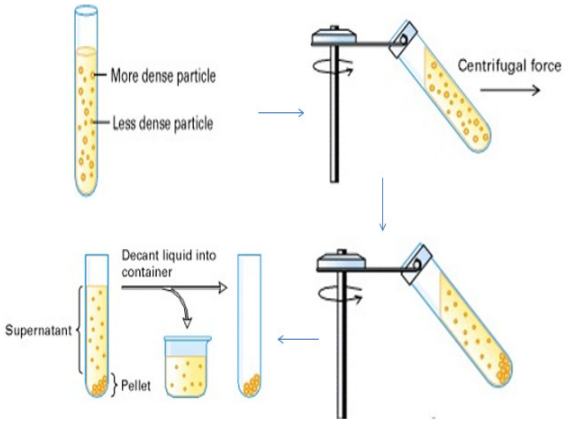
1) To compare the structure of proteins by different organism. 2) To purify protein in order to indentify the gene that encodes it. 3) To examine an enzyme in a crude extract for physiological studies. 4) To study the structure and mechanism of action of an enzyme. 5) To diagnose parasitic disease.

Isolation and Separation of protein

To isolate extracellular proteins, cell disruption is not necessary. Secreted soluble proteins can be collected in the cell supernatant after centrifugation and the membrane-bound proteins might be released from the cell simply using detergents. But, for isolation of intracellular protein, cell disruption is necessary. It is effected by the different process as (a) Detergents lysis, (b) Enzymatic lysis, (c) Osmotic lysis, (d) Freeze-thaw cycles, (f) Ultra-sonication, (g) Homogenization.

Centrifugation:

Centrifugation separates the proteins from their solutions depending on their mass or density. The heavier and/or denser molecule will settle to the bottom of a tube more quickly than lighter or less dense molecules. Generally, the separation of proteins is carried out in a solution containing a layer of increasing or decreasing concentration of sucrose or some other media, like percoll.



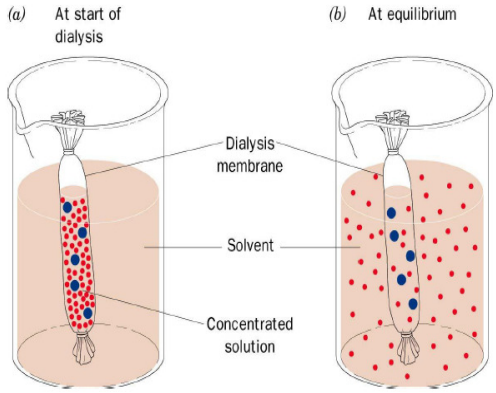
Depending on the centrifugal force and duration of centrifugation the insoluble materials are sediment into a pellet. After centrifugation of a mixture of cell homogenate, cell organelles such as nuclei collect into a pellet, but the soluble proteins remain in the supernatant. The supernatant fraction still contains a large mixture of proteins, which can be collected by decanting the supernatant and then subjecting it to further purification methods.

Salt-induced precipitation or salt fractionation:

It is a purification technique that utilizes the reduced solubility of certain molecules such as proteins or DNA in a solution of very high ionic strength. The salt concentration needed to precipitate out differs from protein to protein, a specific salt concentration can be used to precipitate a target protein. This process is also used to concentrate dilute solutions of proteins. The proteins are solvated in water. But, with increasing salt concentration, these water molecules are attracted by the ions produced from the salt. As a result, only fewer water molecules are available for interaction with the protein. This leads to proteins are interact to each other more strongly than the protein-water interaction resulting the precipitation of protein.

Dialysis:

Dialysis is used to remove lower-molecular components from protein solutions, or to exchange the medium. Dialysis is based on the fact that due to their size, protein molecules are unable to pass through the pores of a semi-permeable membrane, while lower-molecular substances distribute themselves evenly between the inner and outer spaces over time.

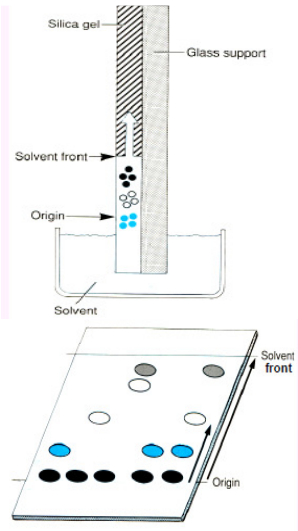


Separation and purification of proteins:

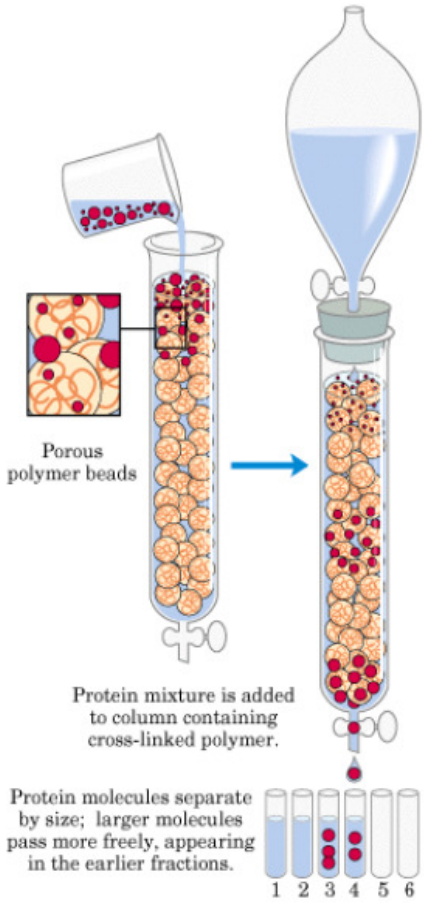
Proteins are separated on the basis of (a) Molecular size (b) Solubility (c) Charge (d) Specific binding-affinity. The techniques used for separation or purifications are (a) Chromatography which is based on differential interaction of molecules with stationary phase and solvent (mobile phase) (b) Electrophoresis which separates charged molecules according to size or charge when moving in an electricfield through a fluid phase.

Thin Layer Chromatography:

Hydrophobic and hydrophilic molecules can crudely separated by partitioning in biphasic solvent systems of chloroform, methanol, and water. The molecule those interacts strongly with the solvent than the stationary phase move faster through the thin layer of stationay phase and others move slowly. These type of differential interaction of proetein molecule with mobile and statinary phase separates the proteins.



Gel Filtration Chromatography:



Gel filtration chromatography separates proteins according to size and shape as they pass through a hydrophilic polymer. Polymer beads composed of cross-linked dextran (dextrose) which is highly and uniformly porous (like Swiss cheese). Large proteins come out first (can't fit in pores), small proteins come out last (get stuck in the pores).

Resolution of gel filtration is determined by three factors:

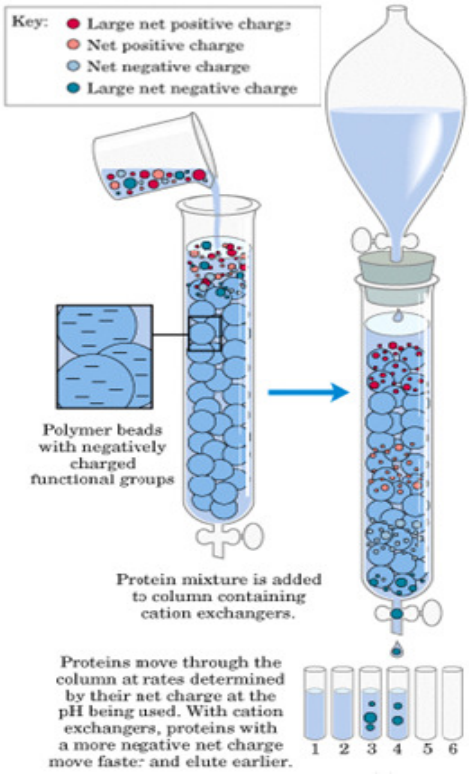
(a) Flow rate (solvent) in relation to column size.

(b) Sample volume in relation to column size.

(c) Length/diameter ratio of the column.

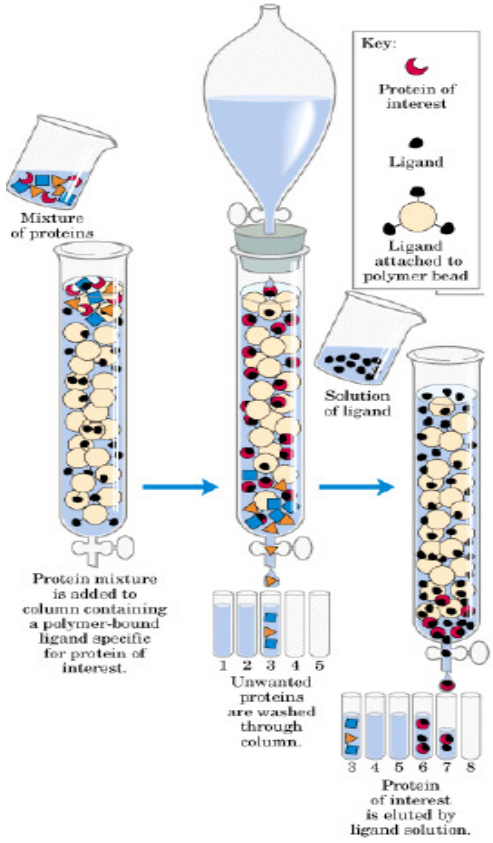
The resolving power of this method is less than that of electrophoretic methods. This method is considered as preparative rather than analytical.

Ion Exchange Chromatography



Ion exchange chromatography separates protein based on their surface ionic charge. Resolusion depends on the net charge of molecules under given solvent conditions and of their retardation on a column. The positively charged proteins bind to negatively charged column bed and so negatively charged protein elute first and followed by positively charged protein. Their charge properties depend upon the ionic composition and pH of the solvent The bounded proteins molecules are eluted by increasing the ionic strength of the mobile phase or by a change in pH.

Affinity Chromatography:



Affinity Chromatography separates molecules based on highly specific interaction between the molecule to be purified and a complementary substance (a ligand) bound to the stationary phase. This inteeaction is specific and reversible.

For example, interaction between antigen –antibody or enzyme – substrate or receptor protein-its ligand or glycosylated protein -specific lectin etc.

In immunoaffinity, dissociation of antibody –antigen complex can be achieved by gradual dropping pH ofthe mobile phase down to 2.7.

In receptor –ligand binding, elution is accomplished byaddition of large excess of free ligand.

Electrophoretic Methods:

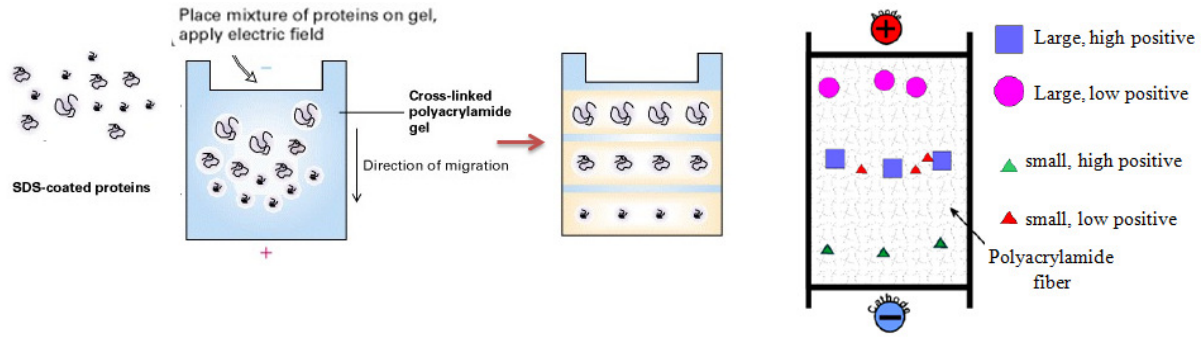
In electrophoresis, molecules in a mixture is subjected under the influence of an applied electric field. The mobility of the dissolved molecules in an electric field is proportional to the net number of charges on them, inversely proportional to the particle radius and the viscosity of the medium. For example, if two molecules have the same mass and shape, the one with the greater net charge will move faster toward an electrode.

The molecule having equal size are sepearated by their charge differences and molecules having equal charge are separated by their size differences. For example: Nucleic acids are repeated units of equally charge/size ratio are effectively separated according to their size.

A small drop of sample is deposited on a strip of filter paper or other porous substrate, which is then soaked with a conducting solution. When an electric field is applied at the ends of the strip, the molecules dissolved in the conducting solution move along the strip at a rate corresponding to the magnitude of their charge. But for better resolution, agarose (plant polysaccharide) or polyacrylamide (synthetic polymer) gel is used.

1-D Protein Electrophoresis

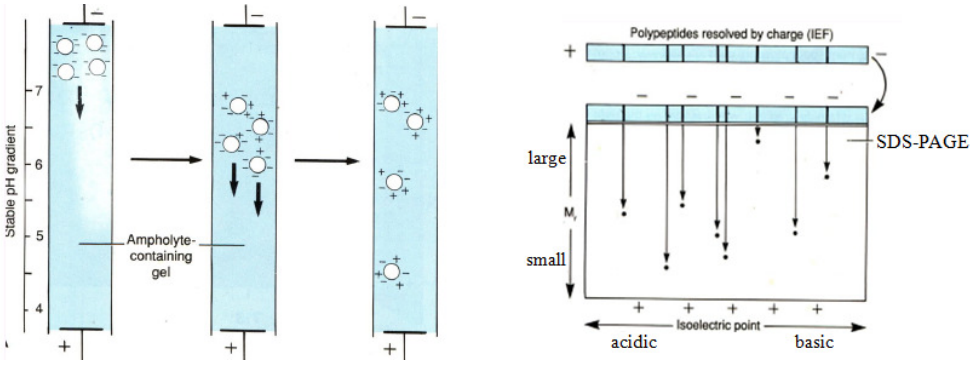
This technique is used to separate proteins according to their molecular size. Before placing the protein mixture, SDS (sodium dodecylsulfate) is added to denaturates the proteins, causing multimeric proteins dissociate into their subunits, and all polypeptide chains are forced into extended conformations with similar charge:mass ratios. SDS treatment thus eliminates the effect of differences in shape, so that chain length, which reflects mass, is the sole determinant of the migration rate of proteins. Thus, SDS which is negatively charged detergent unfolding the proteins, binds uniformly to protein. During electrophoresis, the SDS-protein complexes migrate through the gel. Small proteins are able to move through the pores more easily, and faster, than larger proteins. Thus, the proteins separate into bands according to their size. The separated protein bands are visualized by staining with a dye.

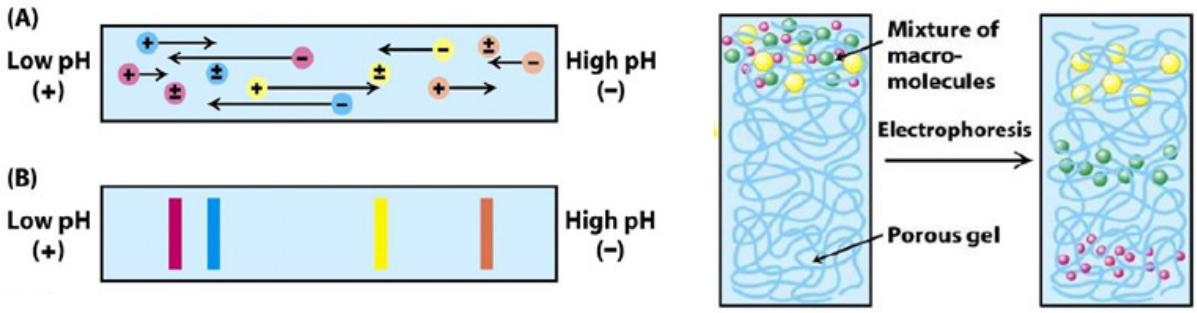


2-D Protein Electrophoresis

This technique can separate proteins according to both their isoelectricpoint (Ip), and their size. When protein mixture is subjected to an electric field with a pH gradient, charged molecules migrate until they reach their isoelectric point, at which they becomes neutral.

Now, these separated proteins on a strip are subjected to a vertical slap of SDS-PAGE (sodium dodecylsulfate–polyacrylamid gel) where proteins are further separated according to their size.





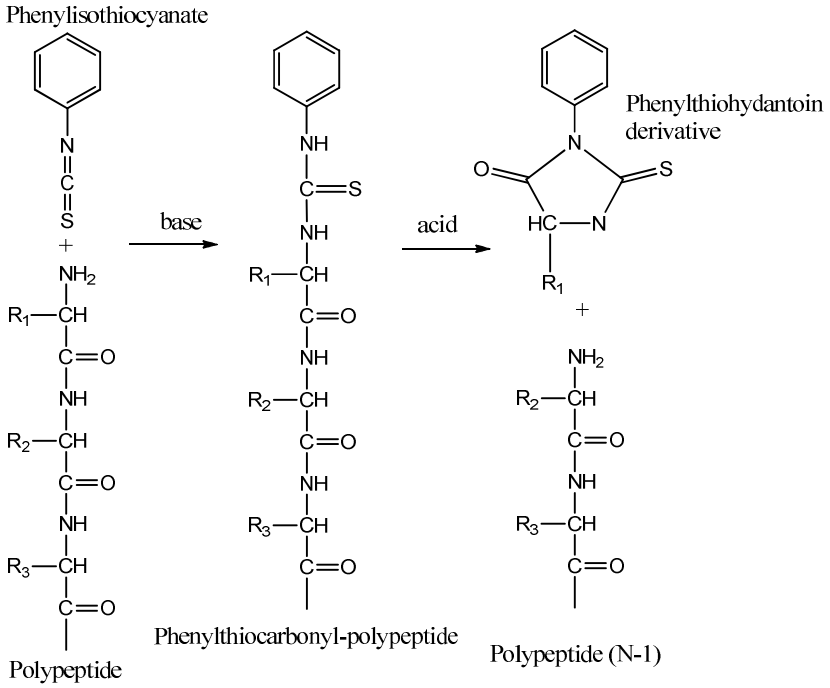
Characterization:

Characterization of proteins includes (a) determination of size (b) determination of charge (c) analysis of co-factor (d) analysis of three dimensional structure (e) activity test etc.

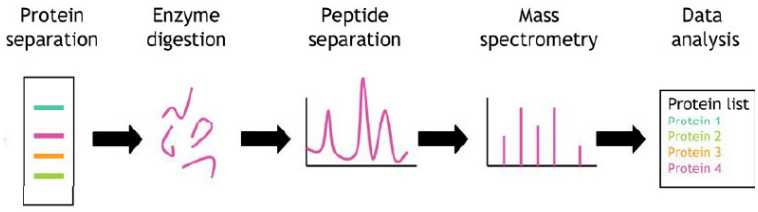
(a) Idea about the size of the proteins are obtained from Gel electrophoresis and gel filtration chromatography. (b) Isoelectric focusing (2-D Protein Electrophoresis) gives the idea about the charge of protein molecules. (c) UV-vis absorption and fluorescence spectroscopy, atomic absorption spectroscopy, inductively coupled plasma-mass-spectrometry are used to analysis the reactivity and co-factor present in it. (d) Circular dichroism spectroscopy, X-ray crystallography are used to analyze the structure of a protein molecule.

Identification of protein:

(a) The N-terminal amino acid sequence of a protein is identified by Edman degradation procedure. Phenylisothiocyanate bind the N-terminal amino acid and then cleaved from the main polypeptide and form phenylthiohydantoin derivative, a stable cyclic compound which is identified by reverse phase HPLC and compare with HPLC of standard mixture.



Mass spectroscopy is also used to identify a protein molecule. The process involves fragmentation of protein, determination of fragment sizes, and subsequent comparison to the known fragmentation pattern.



Denaturation:

Each protein has its own unique shape and 3D-structure which is responsible for its specific function. Loss of this structural integrity with accompanying loss of activity is called denaturation. This is happen when temperature or pH of a protein's environment is changed, or if it is exposed to chemicals. Under such condition, structure maintaining interaction like H-bonding, electrostatic, hydrophobic interaction may be disrupted that causes protein to lose its three-dimensional structure and turn back into an unstructured string of amino acids called **denatured protein**. Denatured proteins are usually non-functional.

Example: Denaturation of protein occurs when an egg is cooked, the skin is wiped with alcohol, and heat is used to cauterize blood vessels.