Definition

The word protein is derived from Greek word, “***proteios***” which means primary. As the name shows, the proteins are of paramount importance for biological systems. Out of the total dry body weight, 3/4th is made up of proteins.

Proteins contain Carbon, Hydrogen, Oxygen and Nitrogen as the major components while Sulfur and Phosphorus are minor constituents. Nitrogen is characteristic of proteins. **On an average, the nitrogen content of ordinary proteins is 16% by weight.**

Proteins are polymers of amino acids that are covalently joined by a substituted amide linkage named a peptide bond. There are 20 different amino acids that make up food proteins. There are about 500 amino acids in nature, but only 20 are proteogenic. Proteins are synthetized on polysomes in the presence of mRNA, rRNA and tRNA as simple amino acid chains.

All proteins are essentially made up of the same primary 20 amino acids; however, some proteins do not contain all 20 amino acids. The differences in structure and function of these thousands of proteins arise from the sequence in which the amino acids are linked together via amide bonds. Literally, billions of proteins with unique properties can be synthesized by changing the amino acid sequence, the type and ratio of amino acids, and the chain length of polypeptides.

1. Amino acids

1.1. Definition and structure

Amino acids are the basic structural unit and building block of proteins. There are 20 amino acids that are coded genetically that most commonly make up food proteins. Each amino acid has a central carbon, called the α -carbon, to which four different groups are attached:

- a basic amino group (—NH_2)
- an acidic carboxyl group (—COOH)
- a hydrogen atom (—H)
- a distinctive side chain (—R).

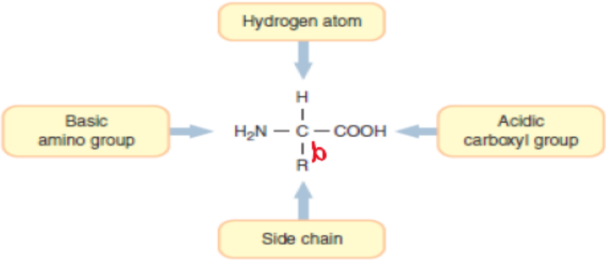


Figure 1: Structure of an amino acid. Excepted for glycine, four different groups are attached to the α-carbon of an amino acid.

Amino acids are asymmetric and can exist as mirror images of each other, both as D and L enantiomers. The only exception to this is glycine, since it contains single hydrogen as its R group. Amino acid residues in proteins including food proteins are exclusively L-amino acids, D-amino acid residues being extremely rare. This nomenclature is derived from D- or L-glyceraldehyde rather than levorotation, the direction of rotation of linearly polarized light.

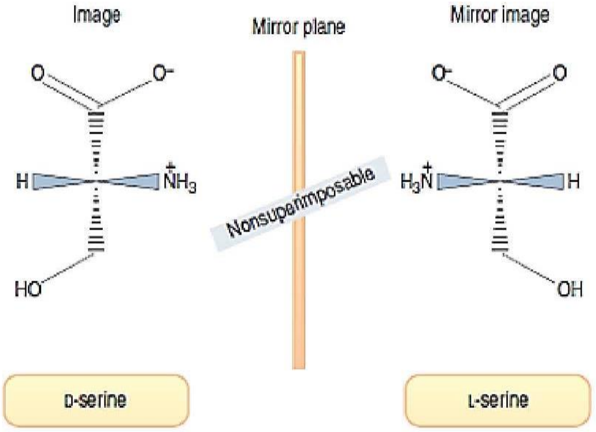


Figure 2: Enantiomers. The mirror-image pair of amino acids.

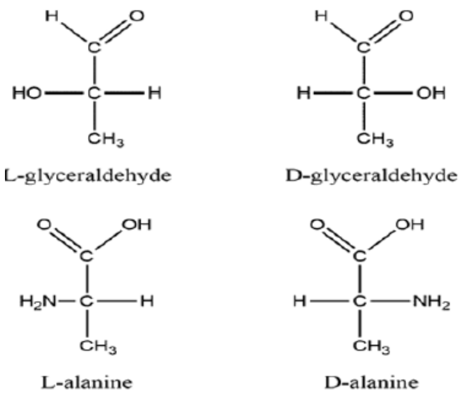


Figure 3: Relationship of stereoisomers of alanine with the absolute configuration of L- and D-glyceraldehyde

Humans cannot synthesize nine of the twenty amino acids that constitute proteins, known as essential amino acids.

Essential amino acids are amino acids that the human body cannot synthesize on its own and must be obtained through diet. These amino acids are crucial for various bodily functions, including protein synthesis, tissue repair, and nutrient absorption. There are nine essential amino acids:

1. **Histidine**: Important for growth, the creation of blood cells, and tissue repair.
2. **Isoleucine**: Involved in muscle metabolism, immune function, hemoglobin production, and energy regulation.
3. **Leucine**: Promotes muscle protein synthesis and repair, and plays a role in regulating blood sugar levels.
4. **Lysine**: Necessary for protein synthesis, hormone and enzyme production, and calcium absorption. It also supports immune function.
5. **Methionine**: Essential for tissue growth and repair, and helps in the absorption of zinc and selenium.
6. **Phenylalanine**: Precursor for neurotransmitters like dopamine, norepinephrine, and epinephrine.
7. **Threonine**: Important for collagen and elastin production, which are essential for skin and connective tissues.

8. **Tryptophan**: Precursor for serotonin, a neurotransmitter that regulates mood, sleep, and appetite.
9. **Valine**: Plays a role in muscle growth and tissue repair, and helps maintain proper energy levels.

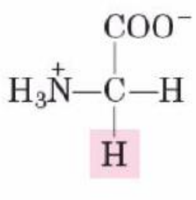
1.2. Classification (based on chemical structure)

The properties of each amino acid are dependent on its side chain (—R); the side chains are the functional groups that are the major determinants of the structure and function of proteins, as well as the electrical charge of the molecule. Knowledge of the properties of these side chains is important for understanding methods of analysis, purification, and identification of proteins.

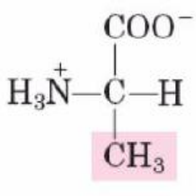
Amino acids with charged, polar or hydrophilic side chains are usually exposed on the surface of proteins. The nonpolar hydrophobic residues are usually buried in the hydrophobic interior or core of a protein and are out of contact with water. The 20 amino acids in proteins encoded by DNA are listed in Table 1 and are classified according to their side chain functional groups.

1.2.1. Aliphatic amino acids

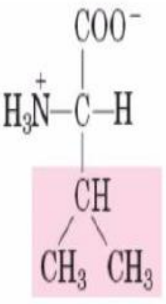
Alanine, **valine**, **leucine**, and **isoleucine**, referred to as aliphatic amino acids, have saturated hydrocarbons as side chains. **Glycine**, which has only a hydrogen side chain, is also included in this group.



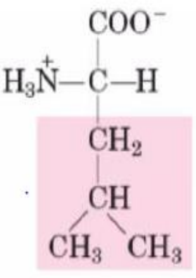
Glycine



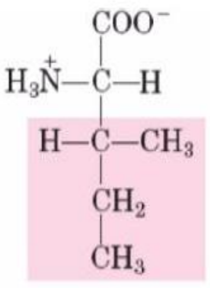
Alanine



Valine



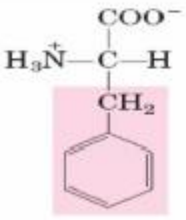
Leucine



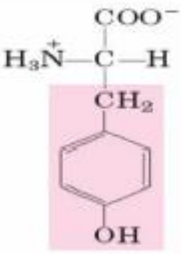
Isoleucine

1.2.2. Aromatic amino acids

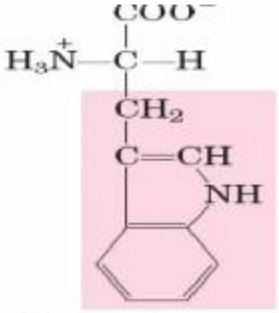
Phenylalanine, **tyrosine**, and **tryptophan** have aromatic side chains. The nonpolar aliphatic and aromatic amino acids are normally buried in the protein core and are involved in hydrophobic interactions with one another.



Phenylalanine



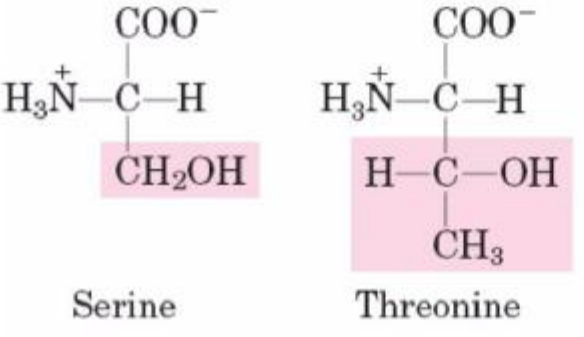
Tyrosine

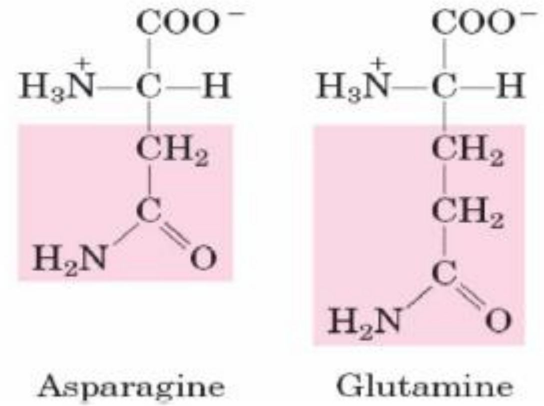


Tryptophan

1.2.3. Neutral polar amino acids

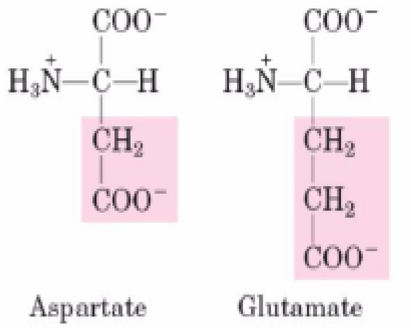
Neutral polar amino acids contain hydroxyl or amide side chain groups. **Serine** and **threonine** contain hydroxyl groups. **Asparagine** and **glutamine** have amide-bearing side chains. These are polar but uncharged under physiological conditions. Serine, threonine and asparagine are the primary sites of linkage of sugars to proteins, forming glycoproteins.





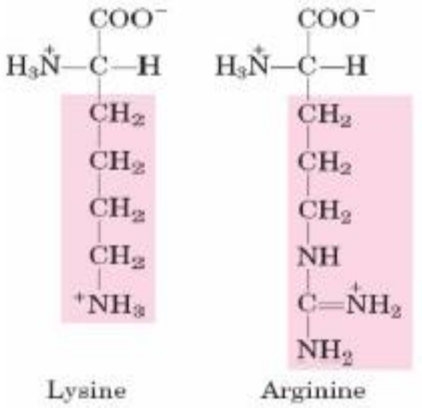
1.2.4. Acidic amino acids

Aspartic and **glutamic acids** contain carboxylic acids on their side chains and are ionized at pH 7.0 and, as a result, carry negative charges on their β - and γ -carboxyl groups, respectively. In the ionized state, these amino acids are referred to as aspartate and glutamate, respectively.



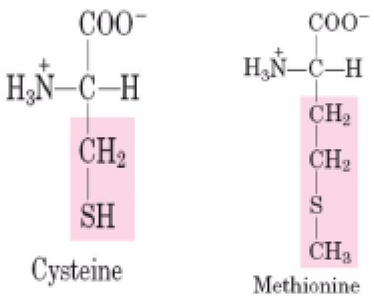
1.2.5. Basic amino acids

The side chains of **lysine** and **arginine** are fully protonated at neutral pH and, therefore, positively charged.



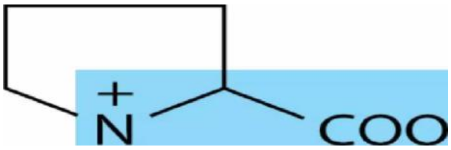
1.2.6. Sulfur-containing amino acids

Cysteine is a sulfur-containing amino acids characterized by low polarity. Cysteine plays an important role in stabilization of protein structure, since it can participate in formation of a disulfide bond with other cysteine residues to form cystine residues, crosslinking protein chains and stabilizing protein structure. **Methionine** is the third sulfur-containing amino acid and contains a nonpolar methyl thioether group in its side chain.



1.2.7. Cyclic imino acid

Proline is different from other amino acids in that its side chain pyrrolidine ring includes both the α -amino group and the α -carbon. This imino acid forces a "bend" in a polypeptide chain, sometimes causing abrupt changes in the direction of the chain.



1.3. Physicochemical properties of amino acids

Physical properties of amino acids

- All the natural acids are colorless crystalline solids.
- **Solubility:** Amino acids vary in their solubility in water depending on their side chains. Polar and charged amino acids are more soluble in water, while nonpolar amino acids are less soluble.
- **Optical Activity:** Because of this asymmetric center, amino acids exhibit optical activity, that is, they rotate the plane of linearly polarized light. All proteins found in nature contain only L-amino acids.

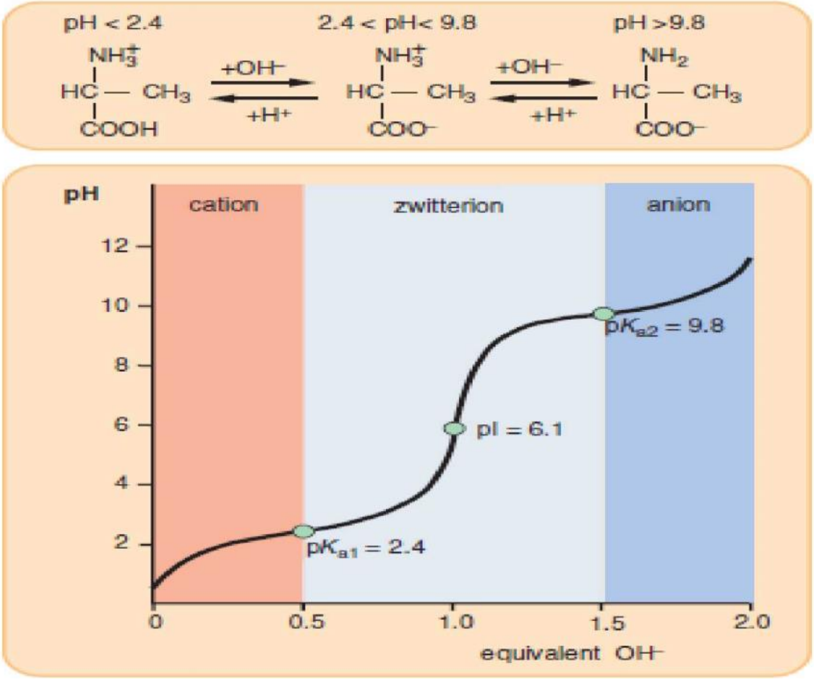
Chemical properties of amino acids

- **Acid-Base Properties**

Amphoteric Nature: Amino acids can act as both acids and bases. The amino group can accept a proton (acting as a base), and the carboxyl group can donate a proton (acting as an acid).

Zwitterions: In aqueous solutions, amino acids exist as zwitterions at certain pH levels, where the amino group is protonated ($-NH_3^+$) and the carboxyl group is deprotonated ($-COO^-$). This makes them neutral overall but charged internally.

Isoelectric Point (pI): The pH at which an amino acid exists predominantly as a zwitterion and has no net charge. Each amino acid has a specific pI, which depends on the nature of its side chain.



➢ ***Properties due to the carboxylic group COOH***:

- **Esterification of amino acids** involves the reaction of the carboxylic group (-COOH) in the amino acid with an alcohol to form an ester.

  Amino Acid (R-COOH) + Alcohol (R'-OH)→Ester (R-COOR')+H2O

- **Amidification of amino acids** refers to the formation of an amide bond by reacting the amino acid's carboxyl group (-COOH) with an amine group ($-NH_2$ or $-NH_3^+$). This process is central to the formation of peptide bonds between amino acids, which are the building blocks of proteins.

  Amino Acid 1 (R-COOH) + Amino Acid 2 (R'-NH_2)→Peptide (R-CONH-R')+H_2O

- **Decarboxylation of amino acids** refers to the removal of the carboxyl group (-COOH) from an amino acid, resulting in the formation of an amine and the release of carbon dioxide (CO_2).

  This reaction plays a significant role in various biological processes and the synthesis of bioactive amines.

  $R-CH(NH_2)-COOH \rightarrow R-CH_2-NH_2+CO_2$

- **Formation of Salts with Bases:** When an amino acid reacts with a strong base, the carboxylic group (-COOH) loses a proton to form -COO⁻. This leads to the formation of a salt where the amino acid acts as an acid.

Example: Reaction with Sodium Hydroxide (NaOH)

$R\text{-}CH(NH_2)\text{-}COOH + NaOH \rightarrow R\text{-}CH(NH_2)\text{-}COO^{-}\cdot Na^{+} + H_2O$ (This produces a carboxylate salt).

➢ **Properties due to the NH_2 group:**

- **Carbonyl Addition Reaction:** The α-amino functions of amino acids react reversibly with aldehydes to give a highly fluorescent product. It is one of the very sensitive methods for detecting amino acids. The amino group ($-NH_2$) attacks the carbonyl carbon of an aldehyde or ketone. (Schiff base formation)

  $R\text{-}CH(NH_2)\text{-}COOH + R'\text{-}CHO$ (Aldehyde)$\rightarrow R\text{-}CH(N{=}CHR')\text{-}COOH + H_2O$

- **Deamination** is the process by which an amino group ($-NH_2$) is removed from an amino acid, resulting in the formation of a corresponding keto acid and the release of ammonia (NH_3). This reaction plays a crucial role in nitrogen metabolism, particularly in the catabolism of amino acids.
- To maintain the intracellular reserve of the 20 amino acids used for protein synthesis, metabolism will undergo deaminations with oxidation that will produce α-keto acids, the main source, if not the only one, from which amino acids are synthesized.

2.Peptides

2.1. Generalities

- Peptides are short chains of amino acids linked by peptide (or amide) bonds (**Alpha carboxyl group of one amino acid reacts with alpha amino group of another amino acid to form a peptide bond or CO-NH bridge**).
- **They are classified as:** di, tri, tetra and poly peptides.
- A polypeptide is a long, continuous, unbranched peptide chain. Peptides are different from proteins on the basis of their size and they contain up to 50 amino acids in their chain.
- Two amino acids combined to form a dipeptide; three amino acids form a tripeptide; four will make a tetrapeptide; a few amino acids together will make an oligopeptide; and combination of 10-50 amino acids is a polypeptide. By convention, long polypeptide chains containing more than 50 amino acids are called proteins.
- In a tripeptide, there are 3 amino acids, but these 3 can be any of the total 20 amino acids. Thus 20^3 = 8000 different permutations and combinations are possible in a tripeptide. An ordinary protein

having about 100 amino acids, will have 20100 different possibilities. This number is more than the total number of atoms present in the whole universe. Thus, even though there are only 20 amino acids; by changing the sequence of combination of these amino acids, nature produces enormous number of markedly different proteins.

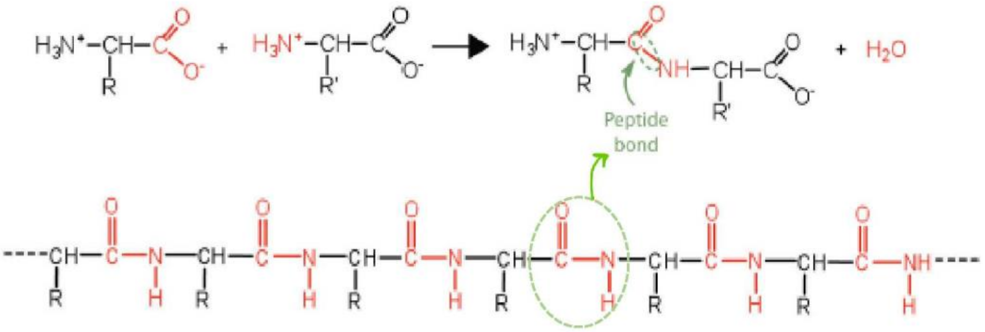


Figure 4: Peptide general structure

2.2.Polypeptide chain

Amino acids that are incorporated into peptides are termed as “residues”. All peptides except cyclic peptides have a N-terminal (amino group) and a C-terminal (Carboxyl group) residue at the end of the peptide.

Peptides are named based on the sequence and number of amino acids they contain, starting from the **N-terminus** (amino end) to the **C-terminus** (carboxyl end). Here’s how peptides are named:

a) **Amino Acid Prefixes**: Each amino acid in the peptide chain is represented by its abbreviated name (using either a three-letter or one-letter code). For example:

- ✓ Glycine → Gly or G
- ✓ Alanine → Ala or A
- ✓ Serine → Ser or S

b) **Naming Convention**: Peptides are named by listing the names of amino acids from the N-terminus to the C-terminus. If a peptide has three amino acids (tripeptide), such as glycine, alanine, and serine, it would be named **Gly-Ala-Ser**.

c) **Suffix “-yl”**: For longer peptides, the amino acids are often named using the “-yl” suffix for all residues except the last one. The last amino acid keeps its full name.

For example:

- ✓ A dipeptide consisting of glycine and alanine would be named **glycylalanine**.
- ✓ A tripeptide consisting of glycine, alanine, and serine would be named **glycylalanylserine**.

d) **Peptide Bond Directionality**: The naming always starts from the amino acid at the **N-terminus** and ends at the **C-terminus**.

e) **Example**: A peptide with the sequence serine, valine, and leucine would be named **Ser-Val-Leu** or

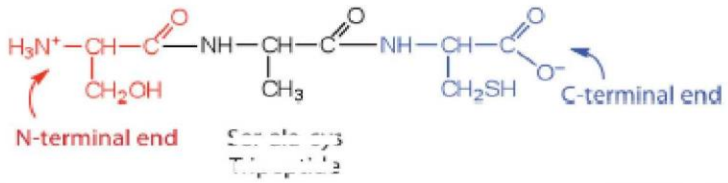


Figure 5: Structure of a tripeptide formed with serine, alanine, and cysteine.

2.3.Example of some interesting peptides:

Oxytocin

▪ **Composition:** Oxytocin is a nonapeptide (nine amino acids long) with the sequence: Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly.

✓ **Structure:** It is cyclic due to a disulfide bond between the two cysteine residues, giving it a stable structure.

▪ **Function:**

- ✓ ***Hormone and Neurotransmitter:*** Oxytocin acts both as a hormone and a neurotransmitter in the brain.
- ✓ ***Role in Childbirth and Lactation:*** It is well-known for its critical role in childbirth, where it stimulates uterine contractions to facilitate labor. After childbirth, it promotes the ejection of milk during breastfeeding by stimulating the muscles around the mammary glands.
- ✓ ***Social Bonding and Behavior:*** Oxytocin is often called the "love hormone" or "cuddle hormone" because it enhances social bonding, maternal behavior, and trust. It contributes to the emotional connections between partners, parent-infant bonding, and social interactions.
- ✓ ***Stress and Emotional Regulation:*** Oxytocin can reduce stress and anxiety by counteracting the effects of cortisol (the stress hormone). It fosters feelings of relaxation and psychological stability.

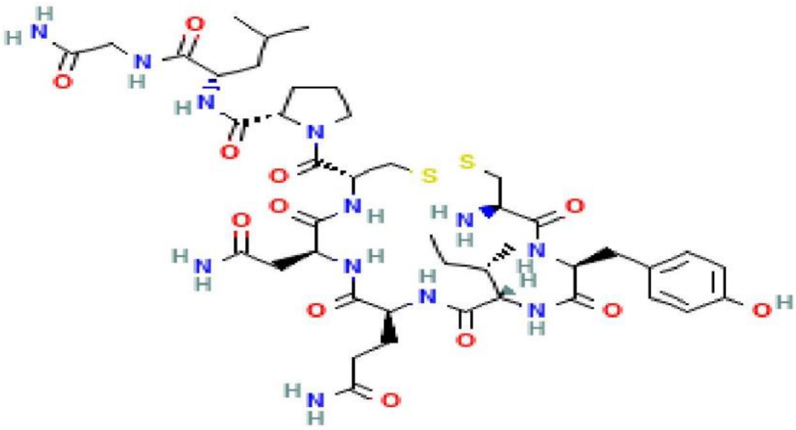


Figure 6: Oxytocin structure

3.Protein

3.1. Definition

Protein is common, important and most abundant biological macromolecules present in all cells. It is composed of amino acids which are commonly called building blocks of protein. It is synthesized by polymerization of amino acids through peptide bonds.

3.2. Classification

The classification of proteins is dependent on the following:

- ✓ Solubility
- ✓ Shape and structure
- ✓ Chemical composition
- ✓ Biological function

3.2.1. Based on Solubility

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

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

3.2.2.3Based on Shape

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

- ✓ **Fibrous proteins**: are formed by polypeptide chains and appear as long thin filaments. Their primary function is to provide support to the cells as well as the whole organism. Examples of fibrous proteins are keratin (in nails, claws), collagen (in bone, teeth, connective tissues of tendons and ligaments) and elastin, etc.

- ✓ **Globular proteins**: appear as spherical globular structure therefore, also called as spheroproteins or corpuscular proteins and their structure is more complex than fibrous proteins. Most of the proteins belong to this class; all enzymes are globular proteins in nature. Examples of globular proteins are insulin, albumin, globulin, lysozymes, histones and protamines (the proteins of animal origin) and prolamines and glutelins found in the food grains.

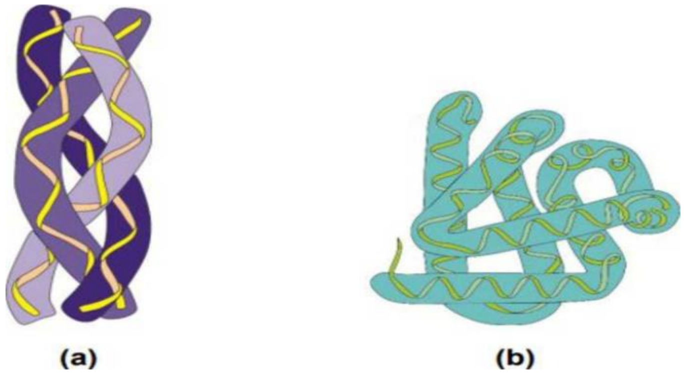


Figure 7: Structures of: (a) Fibrous and (b) Globular proteins.

3.2.3. Based on Chemical Composition

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

- ✓ **Simple proteins or homoproteins** are made up of only amino acids and do not contain any non-protein part. Examples are plasma albumins, collagens, protamines, prolamines, keratins, etc.
- ✓ **Conjugated proteins or heteroproteins** contain in their structure a nonprotein part or group. The protein part of conjugated proteins is called apoprotein and the nonprotein part is referred to as the prosthetic group. The entire mol ecule is called a haloprotein.
- ✓ **Derived proteins** are the proteins derived by the action of different physical or chemical agents like, heat, acid, alkali or enzyme on the native protein. The derived proteins may be primary or secondary derived types. In the case of primary type, the protein gets denatured without getting hydrolysed. The secondary type undergoes a progressive hydrolysis by enzymes, dilute acids and alkalies.

3.2.4. Based on Biological Functions

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

- **Structural proteins:** They have a central role in the mechanical support and stabilization of many structures. Examples are α-keratins, myosin, collagen and elastin.
- **Enzymes** (biochemical catalysts): In living organisms, almost all reactions are catalyzed by specific proteins called enzymes. Almost all the known enzymes are proteins except some catalytic RNA molecules called ribozymes, i.e., ribonucleic acid enzymes. Examples are pepsin, amylase, lipase, etc.
- **Regulatory proteins**: Many hormones are like insulin, glucagon, and thyroid-stimulating hormone (TSH) are proteins. These are regulatory molecules involved in the control of many cellular functions, from metabolism to reproduction.
- **Transport proteins**: Specific proteins classified as transport proteins, transport many organic and inorganic small molecules, in the bloodstream and extracellular fluids, across the cell membranes, and inside the cells from one compartment to another, Examples are: hemoglobin that carries oxygen, transferrin which carries iron in the blood, etc.
- **Storage proteins- ar**e proteins which help to store organic and inorganic molecules. For example, ferritin stores iron intracellularly in a non-toxic form in the liver, myoglobin stores oxygen in the skeletal muscle's ovalbumin.

- **Protective or defense proteins**- Antibodies or immunoglobulins are glycoproteins that recognize antigens expressed on the surface of viruses, bacteria and other infectious agents. Thus, these proteins defend the body against infection.

3.3. Protein structure

The levels of protein structural organization have been divided into four orders, one leading to the other.

3.3.1. Primary structure:

- The sequence of amino acids in its peptide chain is known as primary structure of a protein.
- Primary structure mostly formed by covalent peptide bonds linking the α- carboxyl carbon of each amino acids with α-nitrogen of the next amino acids.

3.3.2. Secondary structure

- Secondary structure consists of the three-dimensional conformation of the protein. Depending on hydrogen bond, amino acid residues close-up, hold and stabilize this three-dimensional form of the protein. Detailed studies in chemistry confirmed three different types of secondary structure of protein viz. α-helix, β-pleated and triple helix.

- ***The α-helix*** is right-handed, compact and rigid coiling-like structure. Each turn of the helix contains about 3,6 amino acid residues. The alpha helix is stabilized by a hydrogen bond between the H-atom attached to electronegative nitrogen atom of a peptide linkage and the electronegative carbonyl oxygen atom of the fourth amino acid on the amino-terminal side of that peptide bond. The amount of helix content differs in various proteins.

- ***The β-pleated sheet*** is a less common, nonhelical, fully extended, slightly zig-zag, sheetlike structure. It is stabilized by regular hydrogen bonds between α-amide N of a peptide bond of one peptide strand and the α- carbonyl O of another peptide linkage of an adjacent peptide strand.
- ***The triple helix*** structure is formed by three congruent geometrical helices of polypeptide chains with same axis winding around one another to form a kind of stiff cablelike structure. It is also known as super helical structure. Examples of triple helices include collagen like proteins, triplex DNA, triplex RNA. This type of structure is stronger and relatively rigid compared to others.

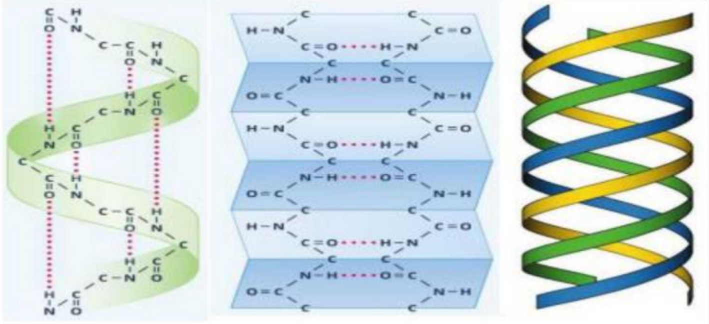


Fig 5a: α- helix Fig. 5b: β-pleated Fig 5c: Triple helix

Figure 8: Secondary structures of a protein.

3.3.3. Tertiary structure

- A protein needs to adopt a final and stable 3-dimentional shape in order to function properly. The tertiary structure refers to this configuration of a protein subunit in three-dimensional space.
- The forces that stabilize the tertiary structure are intermolecular disulfide bonds between cystein residues, electrostatic interactions between ionic groups of opposite charge and also hydrogen bonding.

- Tertiary structure of a protein plays a role in catalytic activity of proteins like enzymes and in how the hormone regulates receptor activation. Physiologic alterations of mature tertiary protein structure occur much more rapidly than does primary sequence.

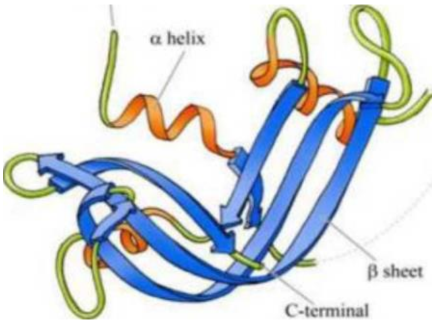


Figure 8: Secondary structures of a protein.

3.3.4. Quaternary structure

- Many proteins consist of more than one polypeptide chain which is known as subunit. Quaternary structure refers spatial arrangement of these subunits of a naïve protein.
- Due to presence of two or more subunits, it is called as oligomers. The subunits are then held together mainly by hydrogen bonds and ionic bonds between polar amino acid sidechains of surface residues, and some hydrophobic interactions between nonpolar amino acid sidechains of the contact regions.
- Hemoglobin is the simple example of quaternary structure. It contains two α and two β chains. The four chains are held together to give a globular shape. Quaternary structure allows a protein to have multiple functions and also to undergo complicated conformational changes.

3.4. Physicochemical properties of proteins

The physicochemical properties of a protein are determined by the analogous properties of the amino acids in it.

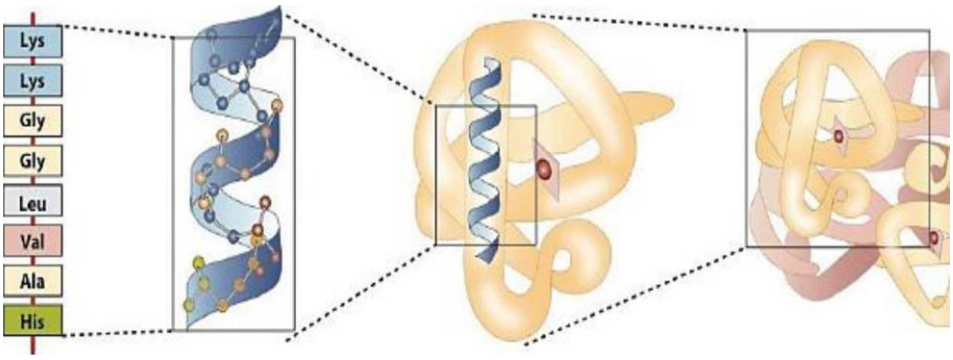


Figure 9: Levels of structure in proteins.

- **Molecular Weight:** The sum of the masses of the amino acids in a protein determines its molecular weight, which can be measured in daltons (Da) or kilodaltons (kDa). proteins are extremely large molecules with molecular weights in the range of 5000 to a few million Daltons (Da). proteins are polypeptides usually containing between 50-25000 or more amino acids. The molecular weights of a large number of proteins are known. Average molecular weight of proteins ranges between 5 kDa (kilodalton) and 220 kDa.
- **Amphoteric Nature:** The proteins have ionizable $-COO^-$ and $-NH_3^+$ groups on the surface because of the constitution of amino acids. Therefore, the proteins are able to donate or accept a proton and are amphoteric in nature. Besides, these can also adsorb anions and cations depending upon the pH of the solution. At alkaline pH, the negative charge predominates and the proteins move to the anode. At acidic pH, the positive charge predominates and the proteins move to the cathode. It is because of the zwitterionic structure that in a crystal lattice the protein molecules are held together tightly.
- **Solubility:** The solubility of proteins depends on the amino acid residues present on the surfaces of proteins as these interact with solvents. It varies greatly from one protein to another; it is influenced by:
  - ***pH:*** The solubility of a protein is minimal at the isoelectric point (the pH at which the net electric charge of the molecule is zero).
  - ***Ionic strength:*** When the concentration of salt in the medium increases, the ionic strength increases and the solubility of proteins decreases, leading to protein precipitation. This technique is used in protein separation.
  - ***Temperature:*** The increase in temperature leads to the denaturation of proteins, making them insoluble.
- **Denaturation** refers to the unfolding of the polypeptide chain i.e. disruption or change in structure of protein. The process of denaturation does not change the primary structure of proteins i.e. the amino acid sequence remains the same. Denaturation occurs due to break in the hydrogen and disulfide bonds of proteins and leads to an increase in the ionizable groups. It brings about change in the physical, chemical as well as biological properties of proteins. Heat, X-rays and UV radiation are well known physical denaturing agents. Other chemical agents include acids, alkalis, heavy metals (lead and mercury), urea and salicylate.