

# Proteins- An Overview

Proteins are macromolecules made up of monomers called amino acids. Amino acids are the building block of all proteins. They play crucial functions in different biological processes.

They are present throughout the body of organisms — in muscles, bones, skin, hair, and virtually every other body part or tissue. They are made of 20 amino acids, arranged in different structural forms, forming around 10,000 proteins (or even more than that).

**Some key properties of proteins responsible for their functions are:**

## **i. Proteins Structural Form-**

Proteins are constructed linearly from 20 amino acids, which is a recurring process in a cell. But the linear/primary sequence of the protein is not responsible for their functional roles.

The sequences spontaneously fold into different arrangements, forming three-dimensional structures that are determined by the amino acids present in the sequence. And these folded structures facilitate the functions of these proteins.

## **ii. Proteins different functional groups-**

The most predominant functional groups present in proteins are alcohols, thiols, thioethers, carboxylic acids, carboxamides, and a variety of basic groups.

These functional groups combined in different random patterns in the amino acid sequence accounts for the broad spectrum of protein function. In simple words, the properties of the functional groups of proteins determine their enzymatic and other metabolic functions in the body.

## **iii. Interaction of proteins forming complex assemblies-**

Proteins interact with other proteins or biomolecules to perform tasks that they alone aren't equipped for. Examples of such tasks are DNA replication, the intracellular transmission of signals, and other complex essential biological processes.

## **iv. Rigidity and flexibility of proteins-**

The rigid proteins (have tightly compact structures which restrict any motion or movement) are involved in the structural formation of cells, such as the cytoskeleton which forms internal scaffolding within cells.

Whereas flexible proteins (have loosely bound structures that can easily rearrange or move when required) act as hinges, springs, and levers to perform essential functions and form complexes with other proteins and macromolecules.

## Synthesis of Protein

### 1. Biological Synthesis

The process of protein synthesis is called translation. In this process, the mRNA codes are translated into the respective amino acids that are involved in protein formation. Each amino acid has its unique nucleotide sequence of genes. The genetic code of one amino acid consists of three sets of nucleotides called codons.

The process starts with the transcription of DNA to pre-mRNA using RNA polymerase. The pre-mRNA is then modified through post-transcription to form a mature mRNA. The mature mRNA is used as a template by ribosomes for protein synthesis.

Ribosomes bind to the mRNA and use one codon (consisting of three nucleotides) matching the anticodon present on tRNA (transfer ribonucleic acid is an adaptor molecule that helps in decoding mRNA) for the synthesis of one amino acid. The four nucleotides, A (adenine), G (guanine), C (cytosine), and T (thymine) form a different combination of nucleotides to form different amino acids.

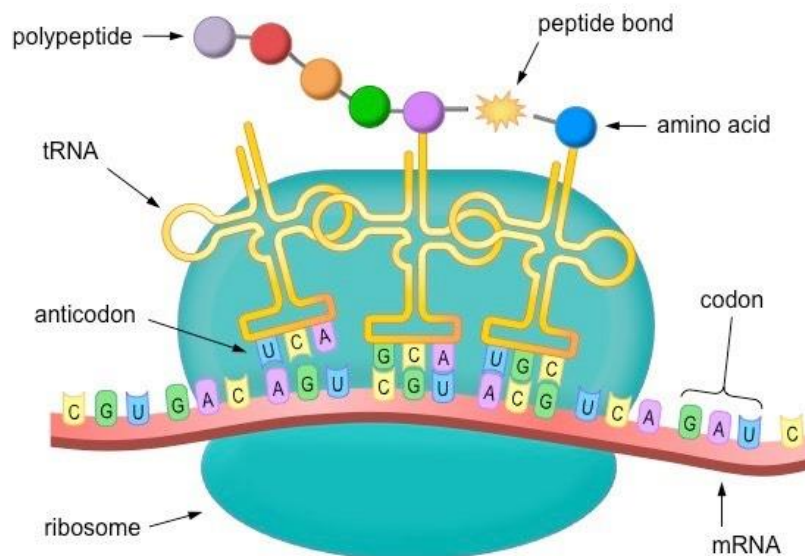


Figure: An illustration of protein synthesis.

## 2. Chemical Synthesis

The chemical synthesis of proteins involves peptide synthesis, which uses strategies like chemical ligation, Staudinger ligation, or other orthogonal chemical reactions to couple synthetic peptides. Here, peptides (a chain of 30-50 amino acids) are produced and linked together via amide or peptide bonds to form specific proteins.

But it's an inefficient technique when it comes to producing a polypeptide chain of more than 300 amino acids. It's important to note that the chemical synthesis of protein occurs from C-terminus to N-terminus, whereas in biological processes, the synthesis occurs in the opposite direction, that is from N-terminus to C-terminus.

### *Structures of Proteins*

Proteins are built by linking two or more amino acid residues together in different orientations or configurations. The amino acids in the proteins are joined together by peptide bonds formed through a condensation reaction between two amino acids, releasing water molecules in the process.

The description of protein structure may sound similar to peptides, but there is a fine line between proteins and peptides as both differ in size, structure, and functions, as explained below:

- Peptides are composed of 2 to 50 amino acids and can be classified further into oligopeptides and polypeptides. The polypeptide is a chain of 20 or more amino acids, joined together by peptide bonds, while oligopeptides are composed of 2-20 amino acid residues. However, proteins are made up of a long chain of 50 or more amino acids linked through peptide bonds.
- Peptides mainly have a linear structure, whereas proteins have four structural levels: primary (linear), secondary, tertiary, and quaternary.
- Peptides are involved in regulating the activities of other molecules while proteins have a wider spectrum of functions which include structural determination, enzymatic reactions, and hormonal functions.

### **Four Levels of Protein Structure**

The different levels of protein structure are possible as a result of different chemical interactions within its structure. For example, the folding of the linear sequence of amino acids into three-dimensional structures is driven by several non-covalent interactions such as hydrogen bonding, ionic interactions, Van der Waals forces, and hydrophobic interactions.

Given below is a short description of the **four levels of protein structure**.

## 1. Primary Structure

The primary structure is the arrangement of amino acids into a linear polypeptide chain. Here, the amino acids are only joined by peptide bonds and disulfide bonds. The sequence of amino acids in a polypeptide chain is responsible for the functions of the proteins, which is decided by decoding the genetic codes or genes corresponding to the proteins.

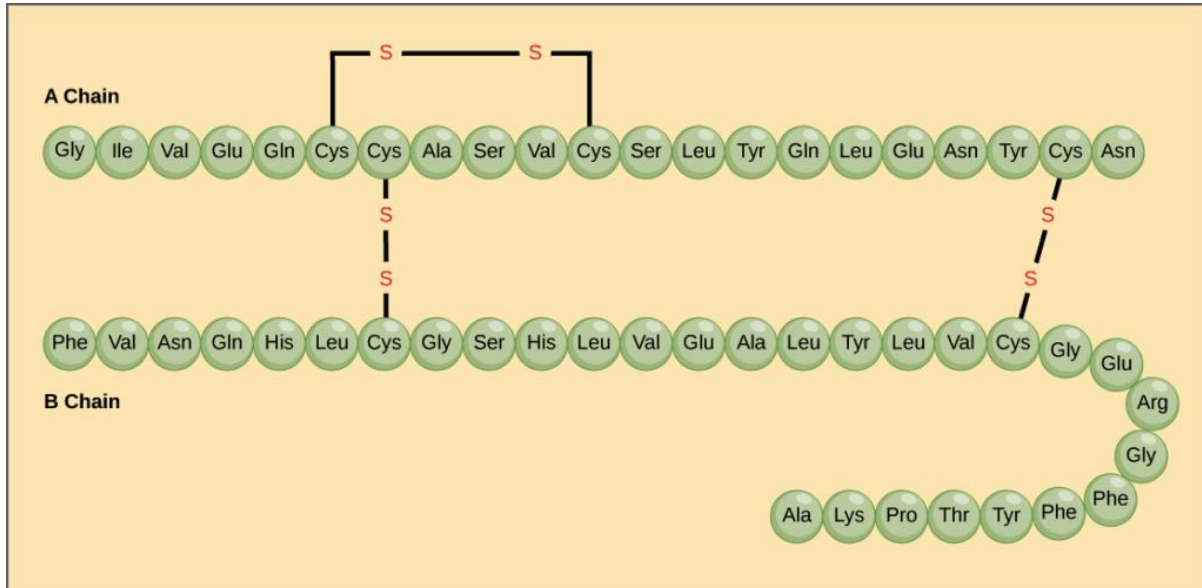


Figure: The representation of the primary structure of the protein.

## 2. Secondary Structure

Secondary structure is the regular, recurring arrangements of adjacent amino acids in a polypeptide chain. This occurs by forming hydrogen bonds between molecules and it represents the initial stage of polypeptides folding into three-dimensional structural form, that is, before tertiary and quaternary structures.

The most common secondary structures are alpha-helix and beta-sheet.

- **Alpha helix:** In this structural form, the carbonyl group (C=O) of one amino acid is hydrogen-bonded to the amino hydrogen (N-H) of the fourth amino acid in the sequence. This pulls the linear chain into a helical ribbon form, with each helical turn containing 3.6 amino acids.
- **Beta sheet or strand:** Here, two or more segments of a polypeptide chain are lined up next to each other. The strands are hydrogen-bonded together in either parallel or antiparallel forms.

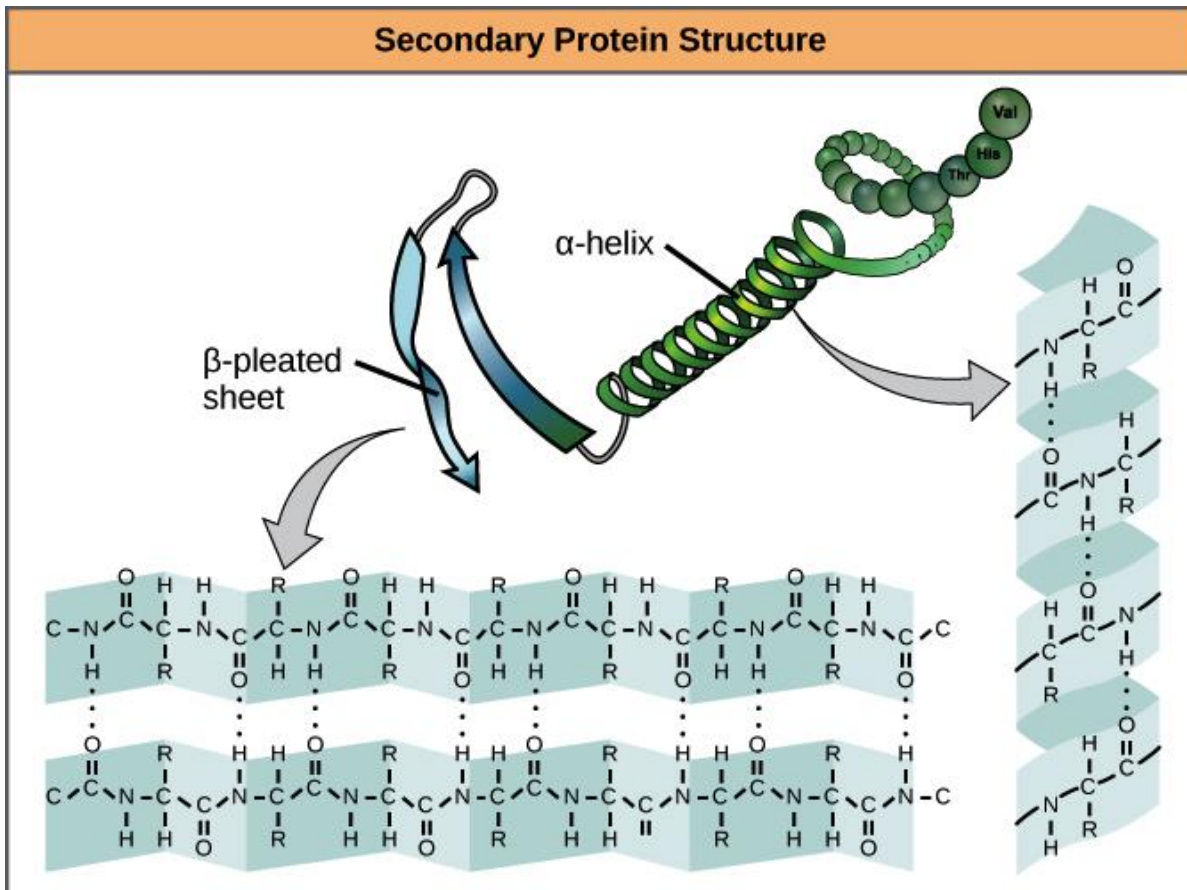


Figure: Structural forms of alpha-helix and beta-pleated sheets.[8]

### 3. Tertiary Structure

Here, the polypeptide chain further folds into a three-dimensional space, involving one or several domains. The alpha-helix and beta sheets are folded into a compact globular structure that's driven by non-specific hydrophobic interactions.

The interactions occur between polar, nonpolar, acidic, and basic R groups within the polypeptide chain. When proteins lose the tertiary structural form, they can't perform any of their functions.

In an aqueous environment, the hydrophobic or nonpolar groups move into the interior of the proteins while the hydrophilic R groups lie on the surface or outside the structure.

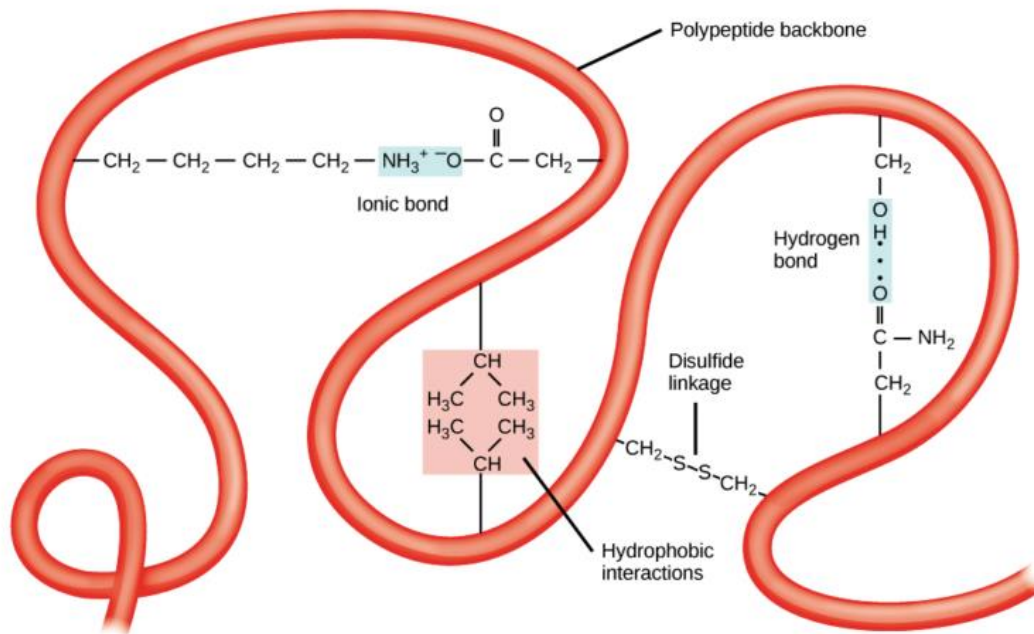


Figure: The representation of the tertiary structure of the protein.

#### 4. Quaternary Structure

More than one polypeptide chain is involved in the formation of a quaternary structure. The polypeptide chains can be identical multiple copies or different in amino acid sequences. It's the weak interactions such as hydrogen bonding and London dispersion forces, that keep the multiple polypeptides bound together in forming a quaternary structure.

A well-known example of a quaternary structure is hemoglobin that carries oxygen in the blood and consists of four peptide chains (two alpha and two beta chains) to form a tetramer.

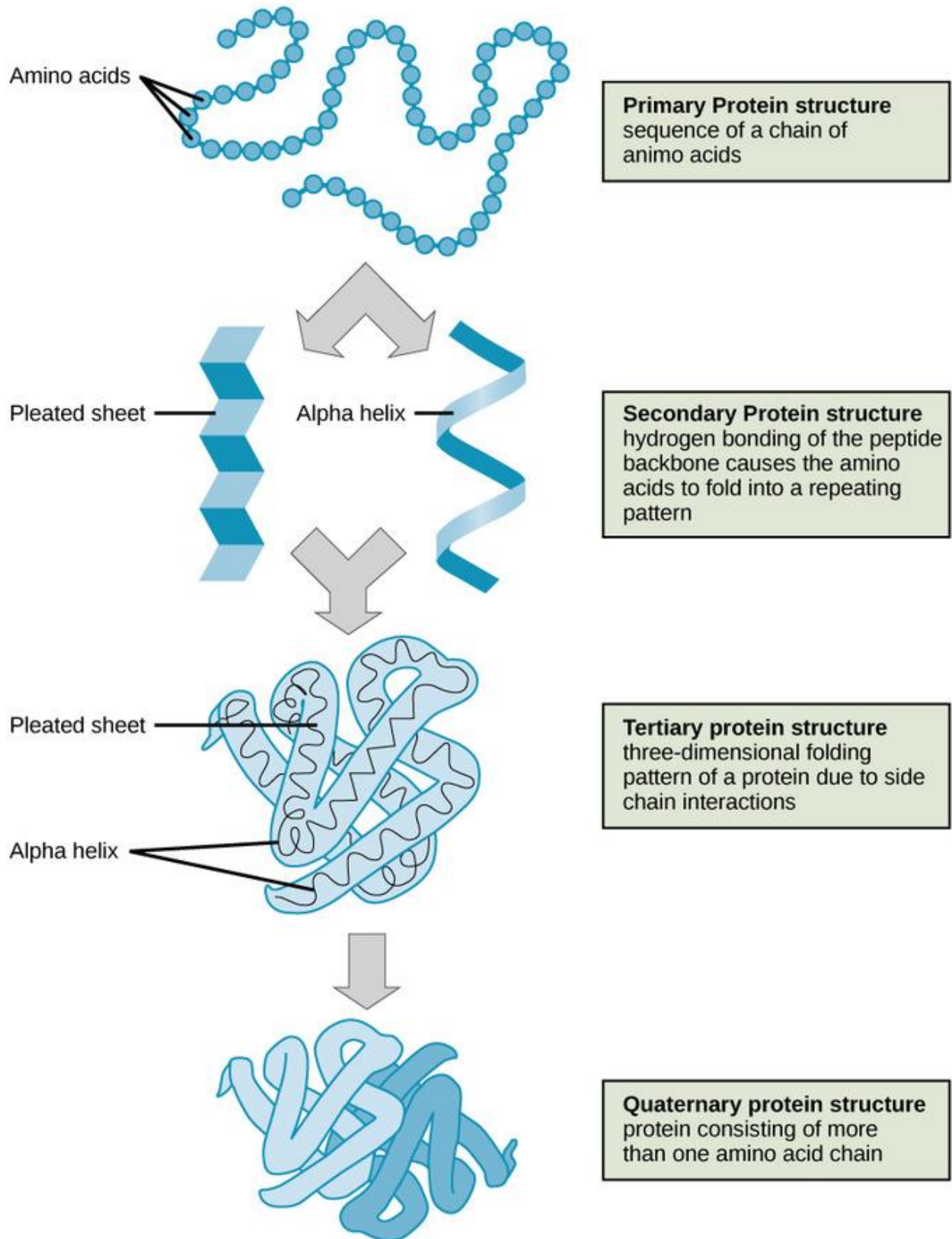


Figure: An illustration of the four levels of protein structure.

## ***Physicochemical Properties of Protein***

1. **Color and taste:** Proteins are colorless and tasteless. They are homogenous and crystalline.
2. **Molecular weights of proteins:** This can be calculated if the molecular weight of amino acids in a linear chain is known. So, the total number of amino acids multiplied by 110 (subtracting the number of water molecules released during condensation reaction) gives the exact molecular weight of proteins. Note: The average molecular weight of amino acids is 110 Daltons.
3. **Coagulation:** This is the heat denaturation of proteins like albumins and globulins to form insoluble coagulates or aggregates, known as coagulum.
4. **Solubility:** Increase in the acidity or alkalinity of the solution increase the solubility of the protein. That is, the solubility of proteins depends on the pH of the solution, and at the isoelectric pH, they are the least soluble.
5. **Optical activity:** The optical activity of proteins is due to the asymmetry of the polypeptides chains and at the  $\alpha$ -carbon atoms in the amino acid residues. The proteins rotate the plane for polarized light to the left (they are levorotatory).
6. **Denaturation:** A protein is said to be denatured when heat, acids, alkalis, alcohol, acetone, urea, beta-mercaptoethanol cause partial or complete unfolding of the native conformation of polypeptide chains.
7. **Isoelectric pH (pI):** The pH at which the protein structure contains an equal number of positive and negative charges on their functional groups is called the isoelectric pH. This property is applied in labs to isolate proteins from different samples. When proteins are subjected to an electric field at isoelectric pH, they neither move to anode or cathode and become less soluble – this causes precipitation of the proteins.
8. **Electrochemistry of proteins:** The electrochemistry of proteins is defined as the redox potential of proteins, due to the presence of free positive and negative charges on the functional groups of the amino acids. This electrochemical character in proteins is mainly decided by the charged ammonium groups ( $-\text{NH}_3^+$ ) of lysine and arginine, and negatively charged carboxyl groups ( $-\text{COO}^-$ ) of aspartic acid and glutamic acid. The charges on these side groups of amino acid of proteins make them move towards respective (cathode or anode) electrodes in the presence of an electric field and participate in any chemical reaction.
9. **Condensation Reaction:** Condensation reaction occurs between the carboxyl group of one amino acid and the amino group of another amino acid. It leads to the formation of peptide bonds and the release of water molecules.

## ***Classification of Proteins***

### ***A. Based on solubility***

1. **Fibrous proteins:** These are insoluble in water and are mainly involved in supportive and protective functions in organisms. They are tough, strong, and linear

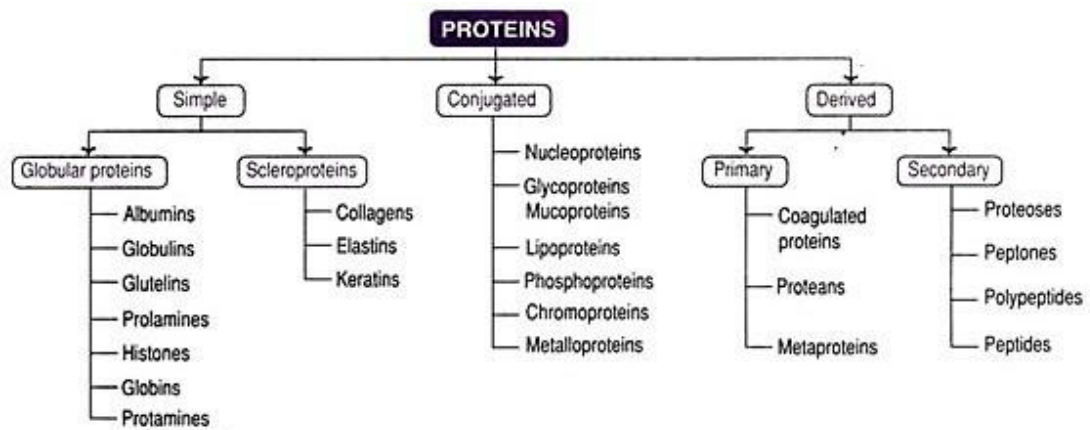


in shape. Here, long, parallel polypeptide chains are cross-linked together, forming the protein. Collagen, keratin, silk, and fibrin are common examples of fibrous proteins.

2. **Globular proteins:** These are soluble in water and have metabolic functional roles, such as forming enzymes, hormones, and antibodies. Here, the polypeptide chains are tightly folded into a sphere. The majority of proteins in the cells belong to this category. Examples include DNA polymerase, RNA polymerase, and hemoglobin.

## B. Based on structural complexity

1. **Simple proteins:** They have a simple structural organization and are only composed of amino acid residues. They are also known as homoproteins. They can be globular or fibrous proteins. Examples include keratin, elastin, albumin, collagen, and histones.
2. **Conjugated proteins:** These are complex proteins and here, proteins are loosely bound with one or more non-protein moieties. The non-protein groups are called prosthetic groups and they can be carbohydrates, lipids, metal ions, nucleic acid, phosphoric acid, and FAD. These proteins are usually globular in shape and are soluble in water. Examples include nucleoprotein, metalloprotein, and lipoproteins.
3. **Derived Proteins:** They are low molecular weight derivatives of proteins — derived from the partial hydrolysis of simple or conjugated proteins by acid, enzyme, or alkalis. Examples are coagulated proteins, proteans, peptones, and peptides.



## Biological Functions of Protein

Proteins perform several essential life-sustaining metabolic functions in organisms. Based on their functions, they are classified into:

1. **Structural proteins:** Most of the structural proteins are fibrous proteins and are insoluble in water. They form the components of bone, tendons, cartilage, skin, connective tissue, hair, and horn. Examples include collagen, keratin, and elastin.

2. **Enzymes:** They are [biological catalysts](#) and work by reducing the activation energy of reactants. During this process, they speed up the metabolic reactions of cells. Most of the enzymes are globular conjugated proteins. Examples include nitrogenase, DNA polymerase, and lipase.
3. **Hormones:** The protein hormones in the cells include glucagon, insulin, and adrenocorticotrophic hormone.
4. **Respiratory pigments:** These are colored proteins that are conjugated and contain pigments (chrome) as their prosthetic group. Examples are hemoglobin and myoglobin.<sup>[12]</sup>
5. **Contractile proteins:** These proteins are involved in muscle contraction at the expense of energy from ATP molecules. Examples are actin and myosin.
6. **Storage proteins:** These proteins store metal or amino acids in the cells and are found in seeds, eggs, milk, and pulses. Examples are casein, gluten, and ferritin.
7. **Transport proteins:** These proteins are responsible for transporting molecules or materials to their target destination. They form channels in the plasma membrane and are also involved in the formation of blood and lymph in animals. Examples include serum albumin.
8. **Defense proteins:** These are proteins that are involved in protecting the organism from foreign microbes or materials. Examples include immunoglobulin (antibodies) and fibrinogen.
9. **Toxins:** Snake venoms are toxic proteins.

**Methods to Study Proteins-** Different characteristics of proteins are studied using different techniques which include: *in vivo*, *in vitro*, and *in silico*.

***In vivo* techniques** are used to study the functional roles of proteins inside cells; ***in vitro* techniques** are preferred when it comes to understanding the working mechanism of proteins in a certain environment; while ***in silico*** is a computation method of studying proteins, like understanding protein complex formation and structural determination.

Given below are techniques based on the studies they are used for:

1. **Protein purification:** The process starts with lysing cells; here, the cell membrane is disrupted to release all cellular components which are then purified using ultracentrifugation. It fractionates the cellular components in different fractions that contain soluble proteins. The proteins in the lysates are precipitated by salting out and isolated using different chromatography techniques based on molecular weight, net charge, and binding affinity. The protein is then purified by using either gel electrophoresis, spectroscopy, or electrofocusing, depending on the experiment requirements.
2. **Presence of protein in samples:** Two tests are used to study the presence of proteins in the given sample: Biuret Test and Ninhydrin Test. In the **Biuret test**, 2 ml of the sample is added to 2 ml 10% NaOH and one drop of 10% CuSO<sub>4</sub> solution. If the color of the solution changes to violet, it indicates the presence of peptide linkage. In the **ninhydrin test**, the formation of violet color after 1 ml of Ninhydrin

solution is added to 1 ml protein solution and gently heated indicates the presence of  $\alpha$ -amino acids in the solution.

3. **Structure prediction and determination:** Homology modeling is a computational tool used to predict the protein structure. It's useful for protein engineering, such as in designing novel folds for drug applications. X-ray crystallography, NMR spectroscopy, dual-polarization interferometry, and circular dichroism are other popular techniques used for the determination of protein structure.
4. **Proteomics:** Total proteins present in a cell is called proteomes. And the study of isolation, separation, identification, characterization, and purification of these proteins is known as proteomics. The most common techniques used for this purpose are 2D electrophoresis, mass spectrometry, protein microarrays, and two-hybrid screening.

## Conclusion

Proteins are one of the essential biomolecules required to sustain life. They consist of amino acids arranged in four structural levels: primary, secondary, tertiary, and quaternary.

They differ from peptides at both structural and functional levels. The functions of proteins range from transportation of molecules, structural formation, to storage and enzymatic roles, whereas peptides only influence the activities of other molecules.

The spectrum of functions of proteins in organisms has pulled the scientist's brain towards the complexity of their actions, hence scientists are curious to understand the structure and working mechanism of the molecule. Some common tools used to study proteins are mass spectrometry, chromatography, circular dichroism, and spectrometry.

Working with proteins requires a high level of precaution and expertise. And considering its importance in organisms, scientists are trying to develop, synthesize, and purify the molecule at an individual level for application in medical areas. The area of proteomics also opens the doors for young scientists to make a breakthrough and contribute to enhancing the quality of human life.