**PROTEINS**

A protein is a naturally occurring, extremely complex substance that consists of [amino acid](https://www.britannica.com/science/amino-acid) residues joined by [peptide bonds](https://www.britannica.com/science/peptide-bond). Proteins are present in all living organisms and include many essential biological compounds such as enzymes, [hormones](https://www.britannica.com/science/hormone), and [antibodies](https://www.britannica.com/science/antibody).

[Proteins](https://en.wikipedia.org/wiki/Protein) are [polymer](https://en.wikipedia.org/wiki/Polymer) chains made of [amino acids](https://en.wikipedia.org/wiki/Amino_acid) linked together by [peptide bonds](https://en.wikipedia.org/wiki/Peptide_bonds). During human [digestion](https://en.wikipedia.org/wiki/Digestion), proteins are broken down in the stomach to smaller [polypeptide chains](https://en.wikipedia.org/wiki/Polypeptide) via [hydrochloric acid](https://en.wikipedia.org/wiki/Hydrochloric_acid) and [protease](https://en.wikipedia.org/wiki/Protease) actions. This is crucial for the [absorption](https://en.wikipedia.org/wiki/Absorption_%28small_intestine%29) of the [essential amino acids](https://en.wikipedia.org/wiki/Essential_amino_acid) that cannot be [biosynthesized](https://en.wikipedia.org/wiki/Biosynthesis) by the body.

**Classification:**

 Proteins have been divided broadly in two classes, namely fibrous proteins and globular proteins on the molecular shapes and functions.

**Fibrous proteins**: Fibrous proteins also known as Schleroprotein are long protein chains shaped liked rodwires. Unlike Globular Protein, they do not denature as easily, and contain many repeats of secondary structures. They are mostly structural proteins that are responsible for organisms in support and protection such as forming connective tissue, muscle fibers, bones, and tendons. The two examples of fibrous proteins are:

1. **α –keratin**: α –keratin (essential in hair, hooves, horn, fingernails, and etc.) is a coiled-coil protein composed of two intertwining α-helices.

2. **Collagen**: Collagen (of tendon, cartilage, blood vessel walls) is the most abundant protein in human’s body. Collagen is a triple helix that is unlike α-helix, it has 3.3 amino acids and 10 Å per turn.

**Globular Protein**: Globular proteins are folded to bury the hydrophobic side chains.

**Types of Structures of Protein:**

Proteins are usually portrayed in 3D structures and categorized into four different characteristics and levels:

**Primary**: The primary structure of a protein is the level of protein structure which refers to the specific sequence of amino acids. When two amino acids are in such a position that the carboxyl groups of each amino acid are adjacent to each other, they can be combined by undergoing a dehydration reaction which results in the formation of a peptide bond. Amino acids in a polypeptide (protein) are linked by peptide bonds that begin with the N-terminal with a free amino group and ends at C-terminal with a free carboxyl group. The peptide bond is planar and cannot rotate freely due to a partial double bond character. While there is a restricted rotation about peptide bond, there are two free rotations on (N-C) bond and (C-C) bond, which are called torsion angles, or more specifically the phi and psi angles. The freedoms of rotation of these two bonds are also limited due to steric hindrance. Genes carry the information to make polypeptides with a defined amino acid sequence. An average polypeptide is about 300 amino acids in length, and some genes encode polypeptides that are a few thousand amino acids long. It's important to know the primary structure of the protein because the primary structure encodes motifs that are of functional importance in their biological function; structure and function are correlated at all levels of biological organization.



A picture of primary structure of protein.

**Secondary**: The amino acid sequence of a polypeptide, together with the laws of chemistry and physics, cause a polypeptide to fold into a more compact structure. Amino acids can rotate around bonds within a protein. This is the reason proteins are flexible and can fold into a variety of shapes. Folding can be irregular or certain regions can have a repeating folding pattern. The coils and folds that result from the hydrogen bonds between the repeating segments of the polypeptide backbone are called secondary structures. Although the individual hydrogen bonds are weak, they are able to support a specific shape for that part of the protein due to the fact that they are repeated many times over a long part of the chain. Secondary structures of a protein are proposed by Pauling and Corey. Its structures are formed by amino acids that are located within short distances of each other. Because of the planar nature of the peptide bonds, only certain types of secondary structure exist. The three important secondary structures are α-helix, β-sheets, and β-turns. Also, the beta sheets can be parallel, antiparallel, or mixed. Antiparallel beta sheets are more stable because the hydrogen bonds are at a ninety degree angles. The a-helix is a coiled structure stabilized by intrachain hydrogen bonds.



One type of secondary structure, an alpha helix.



Another type of secondary structure, a beta sheet.

**Characteristics of the Secondary Structures**:

1. **α-helix**: In an α-helix, the polypeptide backbone forms a repeating helical structure that is stabilized by hydrogen bonds between a carbonyl oxygen and an amine hydrogen. These hydrogen bonds occur at regular intervals of one hydrogen bond every fourth amino acid and cause the polypeptide backbone to form a helix . The most common helical structure is a right-handed helix with its hydrogen bonds parallel to its axis. The hydrogen bonds are formed between carbonyl oxygen and amine hydrogen groups of four amino acid residues away. Each amino acid advances the helix, along its axis, by 1.5 Å. Each turn of the helix is composed of 3.6 amino acids; therefore the pitch of the helix is 5.4 Å. There is an average of ten amino acid residues per helix with its side chains orientated outside of the helix. Different amino acids have different propensities for forming x-helix, however proline is a helix breaker because proline does not have a free amino group. Amino acids that prefer to adopt helical conformations in proteins include methionine, alanine, leucine, glutamate and lysine (malek).

2. **β-sheet**: ß-sheets are stabilized by hydrogen bonding between peptide strands. In a β-sheet, regions of the polypeptide backbone come to lie parallel to each other and are connected by hydrogen bonds [[1]](https://en.wikibooks.org/wiki/Structural_Biochemistry/Proteins#cite_note-Campbell-1). The hydrogen bonds are formed between the carbonyl oxygen and the amine hydrogen of amino acid in adjacent strands in a polypeptide, which means that the hydrogen bonds are inter-stand. β-sheet regions are more extended than an α-helix, and the distance between adjacent amino acids is 3.5 Å. Hydrogen bonding in β-strand can occur as parallel, anti- parallel, or a mixture. Amino acid residues in β- parallel configuration runs in the same orientation. Pleated sheets makes up the core of many globular proteins and also are dominant in some fibrous proteins such as a spiders web [[1]](https://en.wikibooks.org/wiki/Structural_Biochemistry/Proteins#cite_note-Campbell-1). The large aromatics such as: tryptophan, tyrosine and phenylalanine, and beta-branched amino acids like: isoleucine, valine, and threonine prefer to adopt β-strand conformations.This orientation is energetically less favorable because of its slanted, non-vertical hydrogen bonds. Trytophan, tyrosine, and phenylalanine are hydrophobics while the other amino acids are hydrophilics.

3. **β-turns**: Poly peptide chains can change direction by making reverse turns and loops. Loop regions that connect two anti-parallel β-strands are known as reverse turns or β-turns. These loop regions have irregular lengths and shapes and are usually found on the surface of the protein. The turn is stabilized by hydrogen bond between the backbone of carbonyl oxygen and amine hydrogen. The CO group of the residue, in many reverse turns, which is bonded to the NH group of residue i + 3 . The interaction stabilizes abrupt changes in direction of the polypeptide chain. Unlike the alpha-helices and ß-strands, loops do not have regular periodic structures. However, they are usually rigid and well defined. Since they loops lie on the surface of the proteins, they are able to participate in interactions between proteins and other molecules. Ramachandran plot is a plot that shows the available torsion angles of where proteins can be found. However, in the plot, if there are many dots that locate all over the place, it means that there exists a loop.

**Tertiary**: As the secondary structure becomes established due to the primary structure, a polypeptide folds and refolds upon itself to assume a complex three-dimensional shape called the protein tertiary structure. Tertiary structure is the overall shape of a polypeptide.[[1]](https://en.wikibooks.org/wiki/Structural_Biochemistry/Proteins#cite_note-Campbell-1) Tertiary structure results from the interactions between the side chains (R groups) of the various amino acids [[1]](https://en.wikibooks.org/wiki/Structural_Biochemistry/Proteins#cite_note-Campbell-1). This three dimensional structure is due to intramolecular interactions between the side groups along the polypeptide chain. Its domain typically contains 300 – 400 amino acids, and it adopts a stable tertiary structure when it is isolated from their parent protein. As a polypeptide folds into its functional shape, amino acids that have hydrophobic side chains tend to end up clustered at the core of the protein so that they are out of contact with water [[2]](https://en.wikibooks.org/wiki/Structural_Biochemistry/Proteins#cite_note-Viadiu-2). Covalent bonds called disulfide bridges can also affect the shape of a protein [[1]](https://en.wikibooks.org/wiki/Structural_Biochemistry/Proteins#cite_note-Campbell-1). Disulfide Bridges form where two amino acids containing sulfhydryl groups on their side chains are brought close together by how the protein is folding [[1]](https://en.wikibooks.org/wiki/Structural_Biochemistry/Proteins#cite_note-Campbell-1). For some proteins, such as ribonuclease, the tertiary structure is the final structure of a functional protein. Other proteins are composed of two or more polypeptides and adopt a quaternary structure.

**Quaternary**: While all proteins contain primary, secondary and tertiary structures, quaternary structures are reserved for proteins composed of two or more polypeptide chains [[1]](https://en.wikibooks.org/wiki/Structural_Biochemistry/Proteins#cite_note-Campbell-1). Proteins that have quaternary structures contain more than one polypeptide and each adopt a tertiary structure and then assemble with each other via intermolecular interactions. The quaternary structure of a protein is the overall structure that is the result of the addition of these polypeptide subunits [[1]](https://en.wikibooks.org/wiki/Structural_Biochemistry/Proteins#cite_note-Campbell-1). The individual polypeptides are called protein subunits, which means different polypeptides folded separately. Subunits may be identical polypeptides or they may be different. When proteins consist of more than one polypeptide chain, they are said to have quaternary structure and are also known as multimeric proteins, meaning proteins consisting of many parts. Quaternary structures can also defined as when more than one protein come together to create either a dimer, trimer, tetramer, etc... [[2]](https://en.wikibooks.org/wiki/Structural_Biochemistry/Proteins#cite_note-Viadiu-2). Hemoglobin is an example of a quaternary structure that is composed of two alpha subunits and two beta subunits.

**Determination of Structure of Proteins or Polypeptides**

The various steps involved in the determination of a protein or polypeptides are

1. I R Spectrum
2. U. V. Spectrum
3. Amino acid analysis
4. Terminal residue analysis

The amino acids residues at the two ends of a peptide chain are different from another residue and from each other. One N terminal residue contains α-amino group and other, C-termnal residue contains a free –COOH group α- to the peptide linkage.

**A. N-Terminal Residue Analysis**

There are many different reagents which can be used to label terminal amino acids. They all react with amine groups and will therefore also bind to amine groups in the side chains of amino acids such as lysine - for this reason it is necessary to be careful in interpreting chromatograms to ensure that the right spot is chosen. Two of the more common reagents are **Sanger's reagent** ([1-fluoro-2,4-dinitrobenzene](https://en.wikipedia.org/wiki/1-fluoro-2%2C4-dinitrobenzene)) and dansyl derivatives such as [dansyl chloride](https://en.wikipedia.org/wiki/Dansyl_chloride%22%20%5Co%20%22Dansyl%20chloride). [Phenylisothiocyanate](https://en.wikipedia.org/wiki/Phenylisothiocyanate%22%20%5Co%20%22Phenylisothiocyanate), the reagent for the Edman degradation, can also be used.

**Sanger’s method:** Sanger's reagent ([1-fluoro-2,4-dinitrobenzene](https://en.wikipedia.org/wiki/1-fluoro-2%2C4-dinitrobenzene)) and dansyl derivatives such as [dansyl chloride](https://en.wikipedia.org/wiki/Dansyl_chloride%22%20%5Co%20%22Dansyl%20chloride) readily undergoes nucleophilic substitution by the free amino group of the N-terminal amino acid residue in mildly basic solution. The resulting N-dintrophenyl derivative then hydrolysed to a mixture of labelled N-terminal amino acid and other component amino acids.



Sanger's method of peptide end-group analysis: **A** derivatization of *N*-terminal end with [Sanger's reagent](https://en.wikipedia.org/wiki/1-fluoro-2%2C4-dinitrobenzene) (DNFB), **B** total acid hydrolysis of the dinitrophenyl peptide

**B. C-Terminal Residue Analysis**

The number of methods available for [C-terminal](https://en.wikipedia.org/wiki/C-terminal) amino acid analysis is much smaller than the number of available methods of N-terminal analysis. The most common method is to add [carboxypeptidases](https://en.wikipedia.org/wiki/Carboxypeptidase%22%20%5Co%20%22Carboxypeptidase) to a solution of the protein, take samples at regular intervals, and determine the terminal amino acid by analysing a plot of amino acid concentrations against time. This method will be very useful in the case of polypeptides and protein-blocked N termini. C-terminal sequencing would greatly help in verifying the primary structures of proteins predicted from DNA sequences and to detect any postranslational processing of gene products from known codon sequences.