Faculty Science
Everest Shiwach
Department: Botany

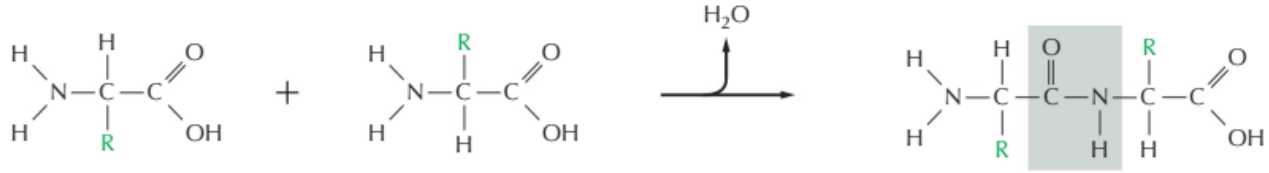
M.Sc. III Sem. Paper- H3002 (Phytochemistry and Metabolism),

Topic- Protein Part-II Structure

A conformation is a spatial arrangement of atoms that depends on the rotation of a bond or bonds. The conformation of a molecule can change without breaking covalent bonds e.g. anti and syn conformations of nucleotides and cis and trans conformations of peptide bond. On the other hand, configurations of a molecule can be changed only by breaking and re-forming covalent bonds. e.g. the L and D forms of amino acids represent different configurations.

Protein structure-The linear sequence of amino acids in a polypeptide chain is called **the primary structure** of a protein. Higher levels of structure are referred to as **secondary, tertiary, and quaternary.**

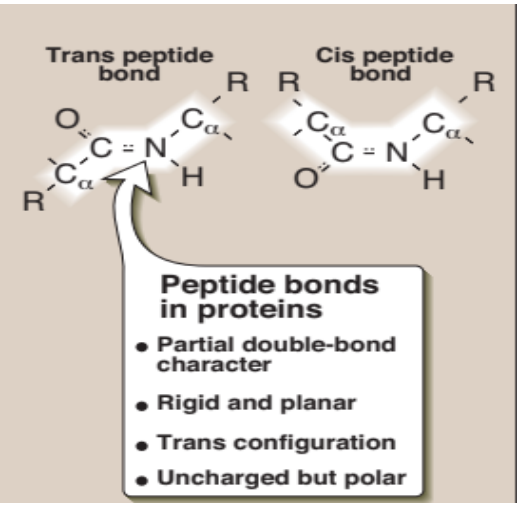
Peptide bond- The linkage formed between amino acids is an amide bond called a peptide bond. This linkage is the product of a simple condensation reaction between the a-carboxyl group of one amino acid and the a-amino group of another. A water molecule is lost from the condensing amino acids in the reaction. Simple condensation reactions are extremely unfavourable in aqueous solutions due to the huge excess of water molecules. The actual pathway of protein synthesis involves reactive intermediates that overcome this limitation. Unlike the carboxyl and amino groups of free amino acids in solution the groups involved in peptide bonds carry no ionic charges.

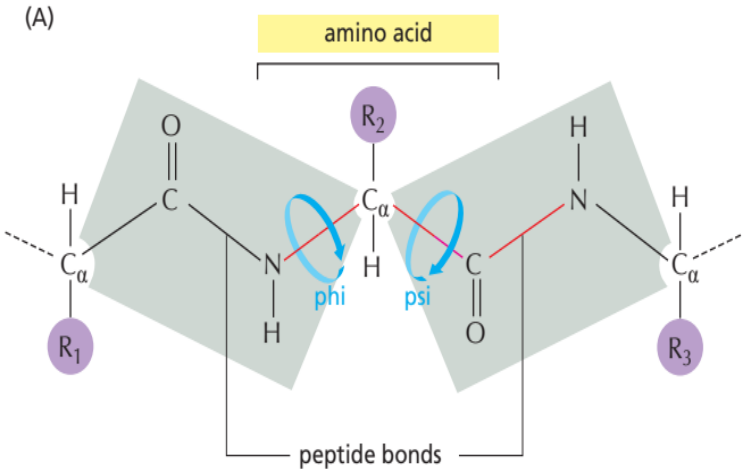


Peptide bond formation

Peptide bonds assume the trans conformation, in which successive C_α atoms are on opposite sides of the peptide bond joining them (see figure-1). The cis conformation, in which successive C_α atoms are on the same side of the peptide bond, is less stable than the trans conformation because of steric interference between neighboring side chains. Consequently, nearly all peptide groups in proteins are in the trans conformation. This steric interference is reduced in peptide bonds to Pro residues, so ~10% of the Pro residues in proteins follow a

Figure-1



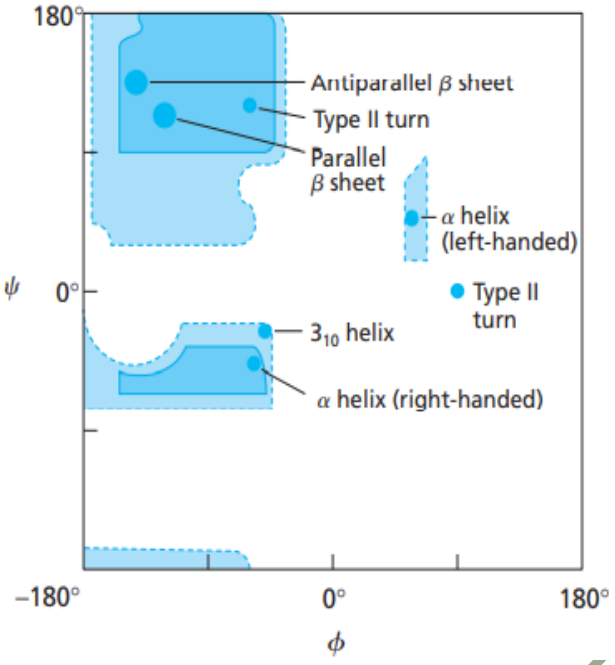


cis peptide bond. The cis and trans conformations arise during protein synthesis when the peptide bond is formed by joining amino acids to the growing polypeptide chain. The two conformations are not easily interconverted by free rotation around the peptide bond once it has formed. The peptide bond exhibits partial double bond character due to the closeness of the carbonyl carbon-oxygen double bond.

The partial double bond character of the peptide bond that links carbonyl C to amide N requires that the carbonyl C, carbonyl O, and amide N remain coplanar, thus preventing rotation. Free rotation is possible about only two of the three covalent bonds of the polypeptide backbone, the C_α-C_o bond and the C_α -N bond. **The torsion angle about the C_α-N bond is termed the phi (ϕ) angle and that about the C_o-C_α bond the psi (Ψ) angle.** The peptide bond angle is ω (omega). For amino acids other than glycine, most combinations of phi and psi angles are disallowed because of steric hindrance. The conformations of proline are even more restricted due to the absence of free rotation of the N-C_α bond. Most of the conformation of the backbone of a polypeptide can be described by torsion angles ϕ and ψ. Each of these angles is defined by the relative positions of four atoms of the backbone. Clockwise angles are positive, and counter clockwise angles are negative, with each having a 180° sweep. Thus each of the rotation angles can range from -180° to +180°.

Ramachandran plot- The Biophysicist G. N. Ramachandran and his colleagues constructed space-filling models of peptides in 1963 and made calculations to determine which values of ϕ and ψ are sterically permitted in a polypeptide chain. Permissible angles are shown as shaded regions in Ramachandran plots of ϕ verses ψ. The R. plot shows the results of theoretical calculations—the dark blue shaded regions represent permissible angles for most residues and the light blue

areas cover the ϕ and ψ values for smaller amino acid residues where the R groups don't restrict rotation. White areas on a Ramachandran plot are non-permissible areas due largely to steric hindrance. The conformations of several types of ideal secondary structure fall within the shaded areas, as expected.



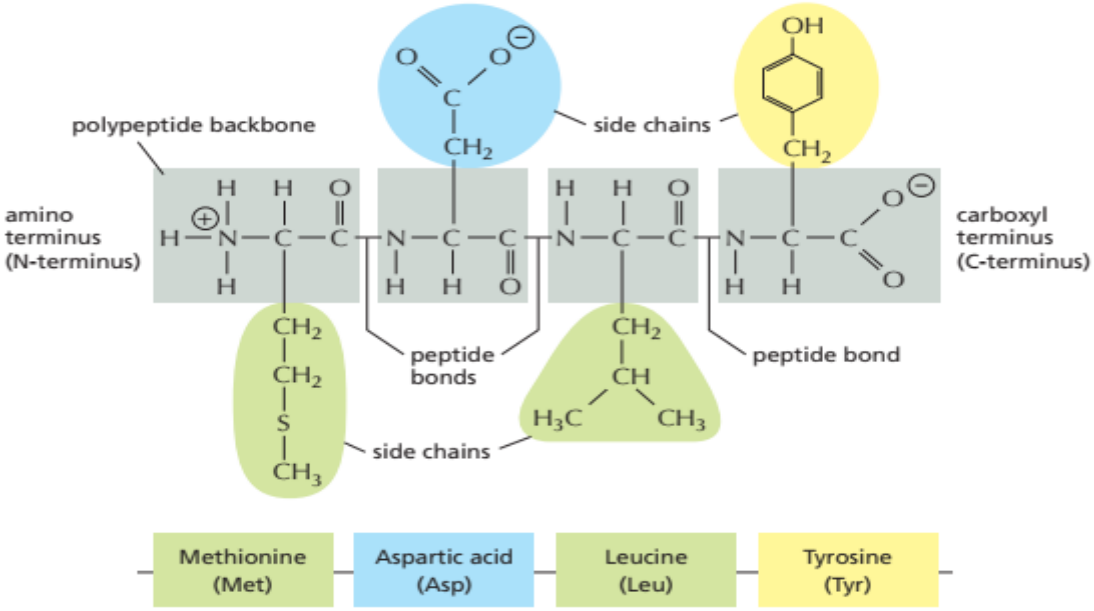
Ramachandran Plot

Idealized ϕ and ψ angles for common secondary structures in protein		
Secondary structure	Torsion angles (in degree)	
	Φ	Ψ
Right handed α- helix	-57	-47
Left handed α- helix	+47	+47
Antiparallel β- sheet	-139	+135
Parallel β- sheet	-119	+113
Right handed 3_{10}- helix	-49	-26

The most important application of Ramachandran plot is the prediction of the quality of various protein structure determined using experimental methods (X-ray crystallography, NMR and Cryo-EM). A good quality structure contains all set of torsional angles in the allowed area whereas, a bad quality (low resolution) protein structure is reflected as a no of torsional angles falling in the forbidden region. Besides experimental methods, protein structure obtained using homology modelling are also routinely checked by plotting Ramachandran plot**.**

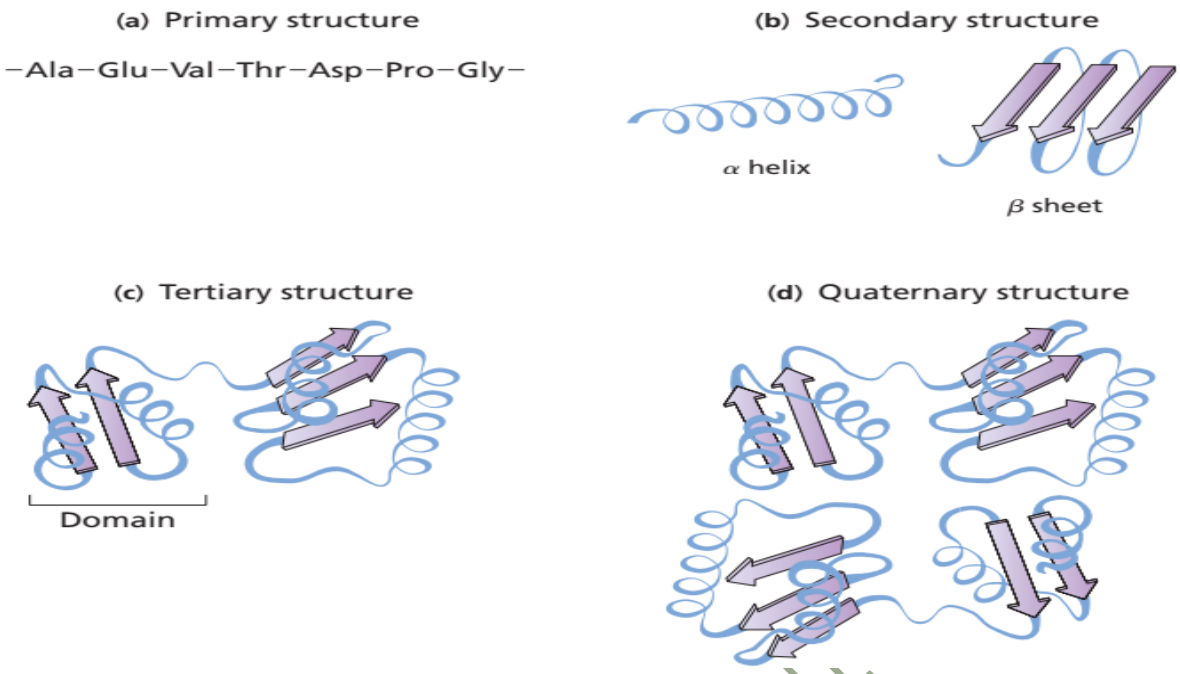
Primary structure- Linked amino acids in a polypeptide chain are called amino acid residues. The primary structure describes the linear sequence of amino acid residues in a protein. The names of residues are formed by replacing the ending -ine or -ate with -yl. For example, a glycine residue in a polypeptide is called glycyl and a glutamate residue is called glutamyl. The free amino group and free carboxyl group at the opposite ends of a peptide chain are called **the N-terminus (amino terminus) and the C-terminus (carboxyl terminus)**, respectively. At neutral pH each terminus carries an ionic charge. By convention, amino acid residues in a peptide chain are numbered from the N-terminus to the C-terminus and are usually written from left to right. This convention corresponds to the direction of protein synthesis. Synthesis begins with the N-terminal amino acid almost always methionine and proceeds sequentially toward the C-terminus by

adding one residue at a time. The terms dipeptide, tripeptide, oligopeptide and polypeptide refer to chains of two, three, several (up to about 20) and many (usually more than 20) amino acid residues, respectively. The mass of a protein is expressed in units of Daltons. One Dalton is equal to one atomic mass unit. A protein with a molecular weight of 50,000 g mol^{-1} has a mass of 50,000 Daltons or 50 kDa (KiloDaltons).



Primary structure

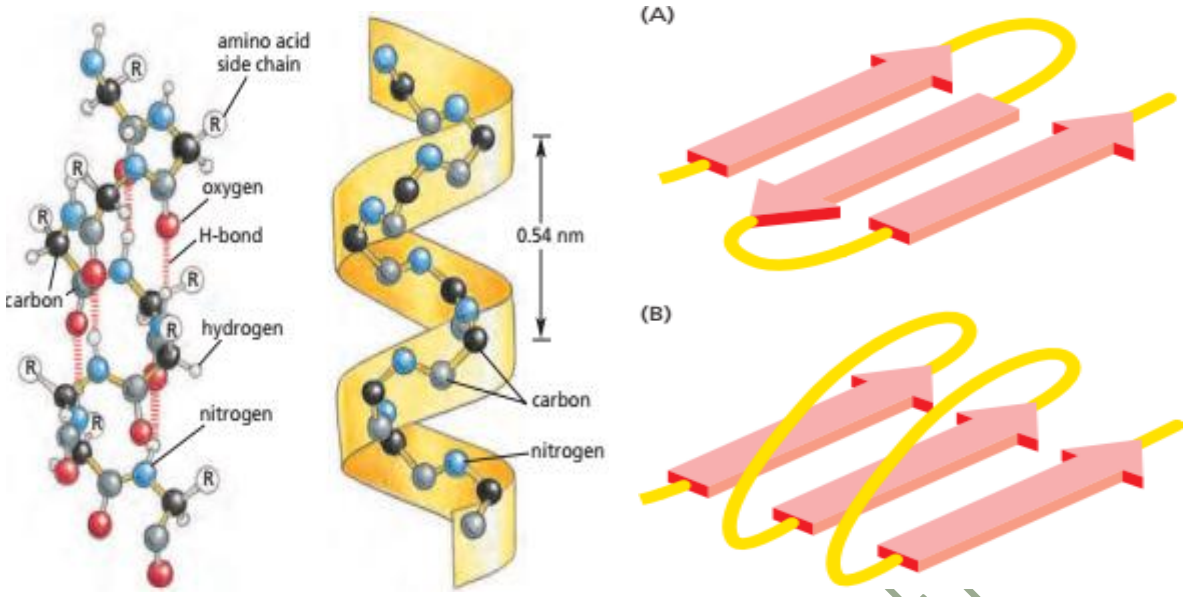
Individual protein molecules have up to four levels of structure. The linear sequence of amino acids is the primary structure of a protein. The three-dimensional structure of a protein is described by three additional levels: secondary structure, tertiary structure, and quaternary structure. The forces responsible for maintaining, or stabilizing, these three levels are primarily noncovalent. **Secondary structure** refers to regularities in local conformations maintained by hydrogen bonds between amide hydrogens and carbonyl oxygens of the peptide backbone. The major secondary structures are α helices, β strands, and turns. **Tertiary structure** describes the completely folded and compacted polypeptide chain. Many folded polypeptides consist of several distinct globular units linked by a short stretch of amino acid residues. Such units are called domains. Tertiary structures are stabilized by the interactions of amino acid side chains in non-neighbouring regions of the polypeptide chain. The formation of tertiary structure brings distant portions of the primary and secondary structures close together. Some proteins possess **quaternary structure**, the association of two or more polypeptide chains into a multi-subunit or oligomeric protein. The polypeptide chains of an oligomeric protein may be identical or different.



Four levels of protein structure

Secondary structures: α-helix -The α-helical conformation was proposed in 1950 by Linus Pauling and Robert Corey. A complete turn of the helix contains an average of **3.6 aminoacyl residues**, and the distance it rises per turn **(its pitch) is 0.54 nm.** The R groups of each aminoacyl residue in an α helix face outward. Proteins contain only L-amino acids, for which a right-handed α helix is by far the more stable and only right-handed α helices occur in nature. The stability of an α helix arises primarily from H bonds formed between the oxygen of the peptide bond carbonyl and the H atom of the peptide bond nitrogen of the fourth residue down the polypeptide chain. Each H bond closes a loop containing 13 atoms (the carbonyl oxygen, 11 backbone atoms, and the amide hydrogen). Thus an α-helix is also called as a 3.6_{13} helix. Alanine has a small uncharged side chain and fits well into the α-helical conformation. Alanine residues are prevalent in the helices of all classes of proteins. Proline disrupts an α-helix because its secondary amino group. Instead, it inserts a kink in the chain. Large numbers of charged amino acids also disrupt the helix by forming ionic bonds, or by electrostatically repelling each other. Amino acids with bulky side chains can interfere with formation of the α-helix if they are present in large numbers. A very diverse group of proteins contains α-helices. For example, the **keratins** fibrous and rigid proteins whose structure is nearly entirely α-helical. They are a major component of tissues such as hair and skin. **Myoglobin** is a globular and flexible molecule (details given subsequently).

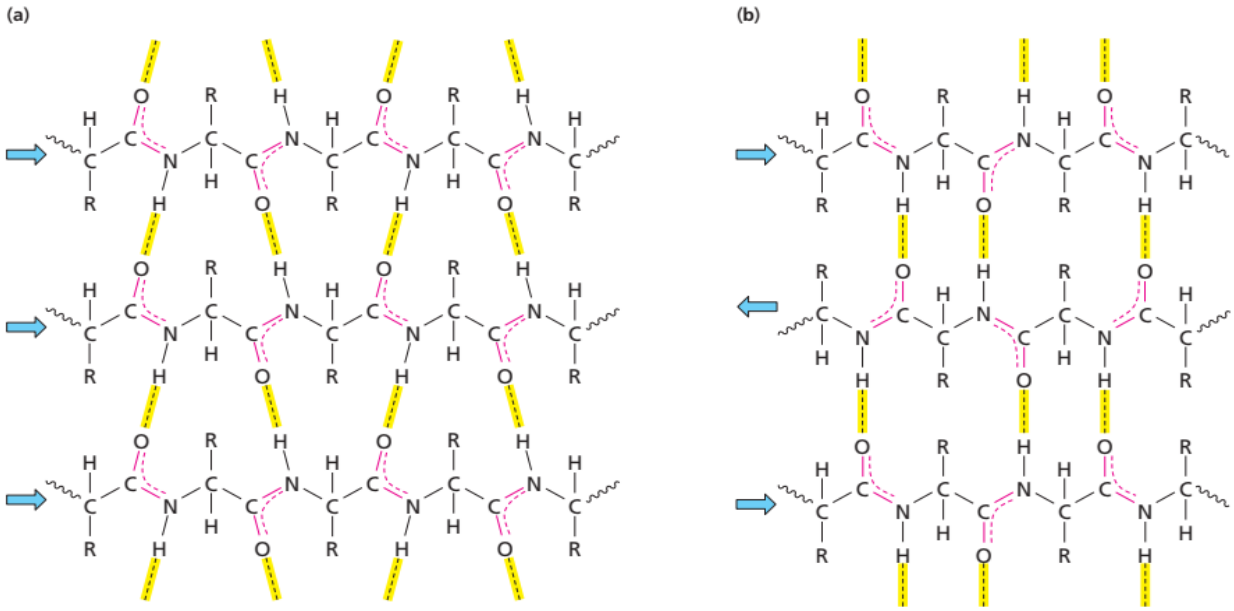
3_{10} helix- Some proteins contain a few short regions of a 3_{10} helix. Like the α-helix, the 3_{10} helix is right-handed. The carbonyl oxygen of a 3_{10} helix forms a



α- helix **(A) Antiparallel β sheet, (B) Parallel β sheet**

hydrogen bond with the amide hydrogen of residue $n + 3$ so the 3_{10} helix has a tighter hydrogen-bonded ring structure than the α-helix and 10 atoms rather than 13 and has fewer residues per turn (3.0) and a longer pitch (0.60 nm).

β- strands and β-sheets- β- strands are portions of the polypeptide chain that are almost fully extended. When multiple β strands are arranged side-by-side they form β sheets, a structure originally proposed by Pauling and Corey. Proteins rarely contain isolated β strands because the structure by itself is not significantly more stable than other conformations. However, β sheets are stabilized by hydrogen bonds between carbonyl oxygens and amide hydrogens on adjacent β strands. Thus, in proteins, the regions of β structure are almost always found in sheets. The hydrogen-bonded β strands can be on separate polypeptide chains or on different segments of the same chain. The β strands in a sheet can be either **parallel** (running in the same N- to C-terminal direction) or **antiparallel** (running in opposite N- to C-terminal directions). When the β strands are antiparallel, the hydrogen bonds are nearly perpendicular to the extended polypeptide chains. In the antiparallel β sheet, the carbonyl oxygen and the amide hydrogen atoms of one residue form hydrogen bonds with the amide hydrogen and carbonyl oxygen of a single residue in the other strand. In the parallel arrangement, the hydrogen bonds are not perpendicular to the extended chains and each residue forms hydrogen bonds with the carbonyl and amide groups of two different residues on the adjacent strand. Parallel sheets are less stable than antiparallel sheets, possibly because the hydrogen bonds are distorted in the parallel arrangement. The β sheet is sometimes called a β pleated sheet since the planar peptide groups meet each

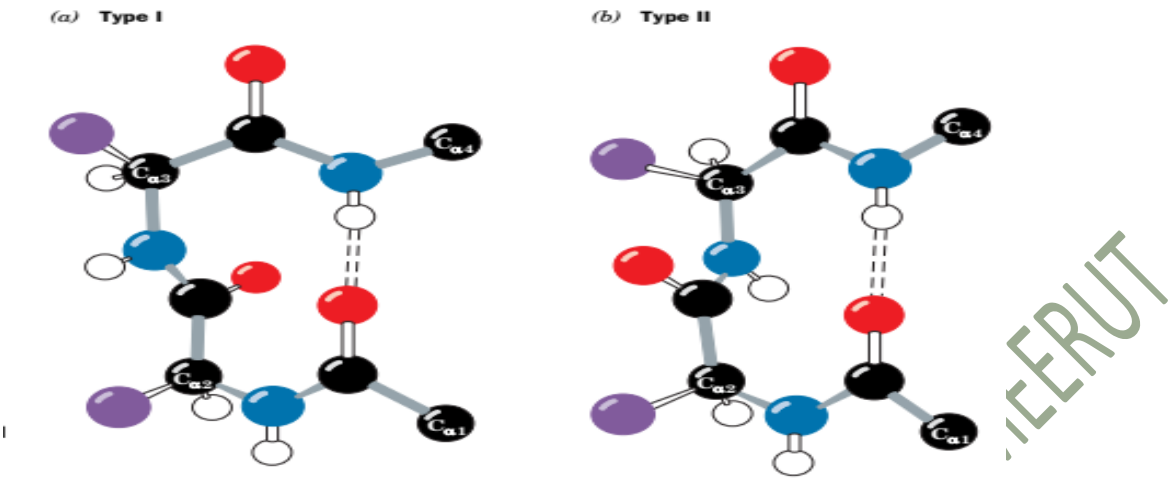


(a) Parallel β sheet **(b) Antiparallel β sheet**

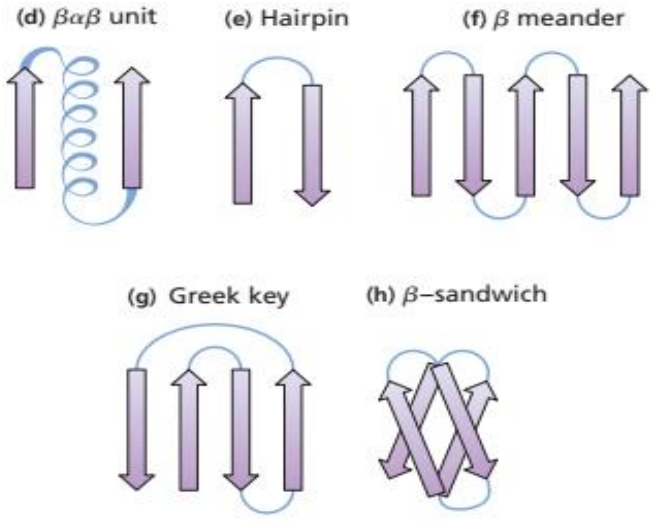
other at angles, like the folds of an accordion. As a result of the bond angles between peptide groups, the amino acid - side chains point alternately above and below the plane of the sheet. A typical β sheet contains from 2 to as many as 15 individual β strands. Each strand has an average of 6 amino acid residues.

Loops or turns- Proteins also contain stretches of nonrepeating 3D structure. Most of these are known as loops or turns because they cause directional changes in the polypeptide backbone. Loops and turns connect α helices and β strands and allow the polypeptide chain to fold back on itself producing the compact three-dimensional shape seen in the native structure. As much as one-third of the amino acid residues in a typical protein are found in such nonrepetitive structures. Loops often contain hydrophilic residues and are usually found on the surfaces of proteins where they are exposed to solvent and form hydrogen bonds with water. Some loops consist of many residues of extended nonrepetitive structure. About 10% of the residues can be found in such regions. Loops containing only a few (up to 5) residues are referred to as turns if they cause an abrupt change in the direction of a polypeptide chain. The most common types of tight turns are called **reverse turns or β turns** because they often connect different antiparallel β strands. This terminology is misleading since β turns can also connect α helices or a α helix and a β strand. There are two common types of β turn, designated type I and type II. Both types of turn contain 4 amino acid residues and are stabilized by hydrogen bonding between the carbonyl O of the 1st residue and the amide H of the 4th residue. Both type I and type II turns produce an abrupt (usually about 180°) change in the direction of the polypeptide chain. In type II turns, the

3rd residue is glycine about 60% of the time. Proline is often the 2nd residue in both types of turns.



Supersecondary structures or motifs- Certain combinations of secondary structure form motifs. Groupings of secondary structural elements, called supersecondary structures or motifs, occur in many unrelated globular proteins:
1. The βαβ motif- In which an α helix connects two parallel strands of a β sheet.
2. The β hairpin motif- It consists of antiparallel strands connected by relatively tight reverse turns.
3. The β meander motif- It is an antiparallel β sheet composed of sequential β strands connected by loops or turns. The order of strands in the β sheet is the same as their order in the sequence of the polypeptide chain. The β meander sheet may contain one or more hairpins but usually the strands are joined by larger loops.



Different types of supersecondary structures

4. The Greek key motif- The named after an ornamental design commonly used in ancient Greece. A β hairpin is folded over to form a 4-stranded antiparallel β sheets.

5. The β sandwich motif- It is formed when β strands or sheets stack on top of one another. The figure shows an example of a β sandwich where the β strands are connected by short loops and turns, but β sandwiches can also be formed by the interaction of two β sheets in different regions of the polypeptide chain.

Tertiary structure- The primary structure of a polypeptide chain determines its tertiary structure. Tertiary refers both to the folding of domains and to the final arrangement of domains in the polypeptide. The structure of globular proteins in aqueous solution is compact, with a high-density of the atoms in the core of the molecule. Hydrophobic side chains are buried in the interior, whereas hydrophilic groups are generally found on the surface of the molecule.

Domains- Domains are the fundamental functional and three-dimensional structural units of polypeptides. Polypeptide chains that are greater than 200 amino acids in length generally consist of two or more domains. The core of a domain is built from combinations of supersecondary structural elements (motifs). Folding of the peptide chain within a domain usually occurs independently of folding in other domains. Therefore, each domain has the characteristics of a small, compact globular protein that is structurally independent of the other domains in the polypeptide chain.

Figure-2 (A)

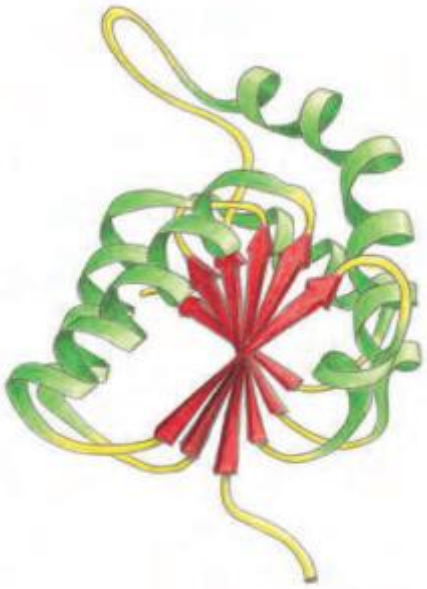


Figure-2 (B)

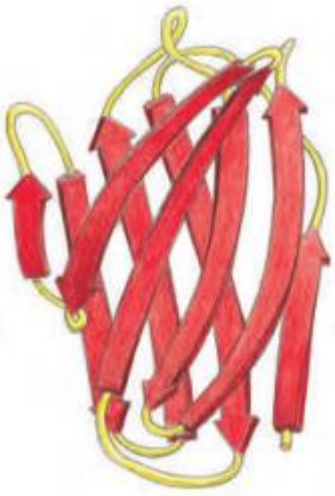


Figure-2 (C)

(A) Cytochrome *b*562, a single-domain protein involved in electron transport in mitochondria. This protein is composed almost entirely of α helices. (B) The NAD-binding domain of the enzyme lactic dehydrogenase, which is composed of a mixture of α helices and parallel β sheets. (C) The variable domain of an immunoglobulin (antibody) light chain, composed of a sandwich of two antiparallel β sheets (see figure-2).

Interactions stabilizing tertiary structure- The following types of interactions cooperate in stabilizing the tertiary structures of globular proteins.

1.Disulfide bonds: A disulfide bond is a covalent linkage formed from the sulfhydryl group (–SH) of each of two cysteine residues, to produce a cystine residue. The folding of the polypeptide chain(s) brings the cysteine residues into proximity and permits covalent bonding of their side chains. e.g. many disulfide bonds are found in immunoglobulins that are secreted by cells.

2. Hydrophobic interactions: Amino acids with nonpolar side chains tend to be located in the interior of the polypeptide molecule, where they associate with other hydrophobic amino acids. In contrast, amino acids with polar or charged side chains tend to be located on the surface of the molecule in contact with the polar solvent.

3. Hydrogen bonds: Amino acid side chains containing oxygen- or nitrogen-bound hydrogen, such as in the -OH groups of serine and threonine, can form H bonds with electron-rich atoms, such as the oxygen of a carboxyl group or carbonyl group of a peptide bond. Formation of H bonds between polar groups on the surface of proteins and the aqueous solvent enhances the solubility of the protein.

4.Van der Waals Interactions: Van der Waals forces arise from electrostatic interactions between permanent and or induced dipoles, such as the carbonyl group in peptide bonds. The cumulative effect of many van der Waals interactions probably makes a significant contribution to stability because nonpolar side chains in the interior of a protein are densely packed.

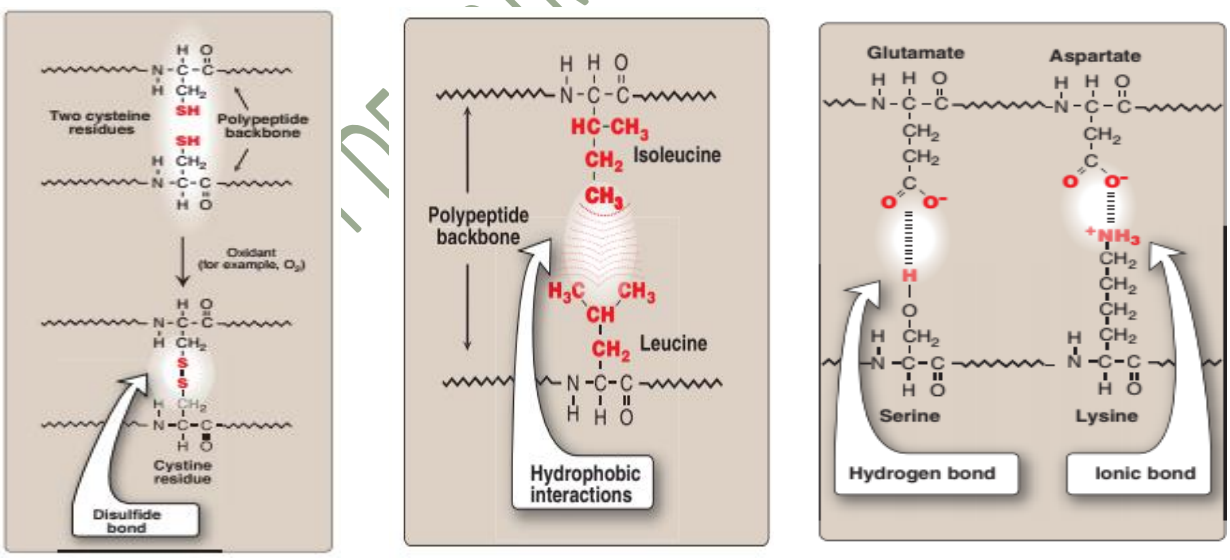


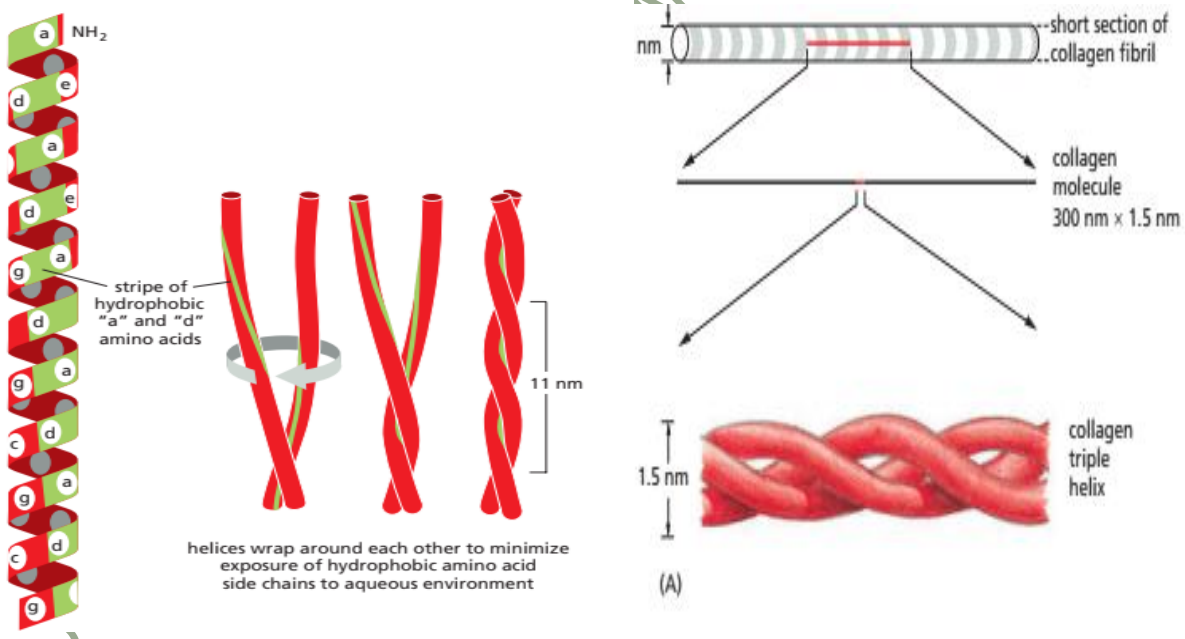
Diagram showing various interactions stabilizing tertiary structure

5. Ionic interactions: Negatively charged groups, such as the carboxylate group ($– COO^-$) in the side chain of aspartate or glutamate, can interact with positively charged groups, such as the amino group ($– NH_3^+$) in the side chain of lysine.

Quaternary structure-Many proteins consist of a single polypeptide chain, and

are defined as monomeric proteins. However, others may consist of two or more polypeptide chains that may be structurally identical or totally unrelated. The arrangement of these polypeptide subunits is called the quaternary structure of the protein. Subunits are held together by noncovalent interactions (e.g. hydrogen bonds, ionic bonds, and hydrophobic interactions). A multisubunit protein is referred to as an oligomer (proteins with only one polypeptide chain are monomers). When the subunits are identical, dimers and tetramers predominate. Subunits may either function independently of each other, or may work cooperatively, as in hemoglobin, in which the binding of oxygen to one subunit of the tetramer increases the affinity of the other subunits for oxygen.

Structure of fibrous proteins- α-Keratin and collagen both form higher order structures that are insoluble in water. (a) α Keratin consists of two α-helical chains that wrap around each other in a left-handed coiled coil. Each polypeptide chain has a 7-residue pseudorepeat, *a-b-c-d-e-f-g*, such that hydrophobic residues at positions *a* and *d* form the contacts between the helices. This dimer is assembled further into higher order structures that are less well characterized.



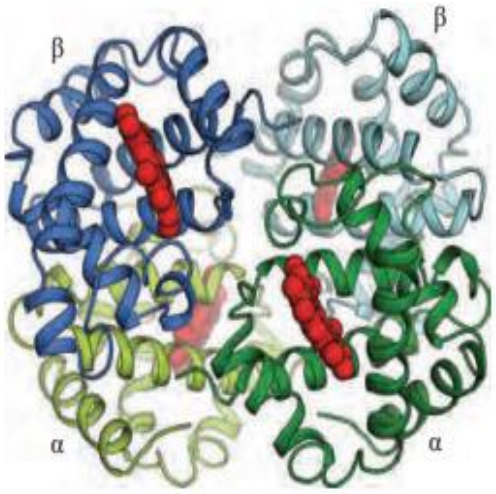
Coiled coil of α-Keratin **Collagen**

(b) Collagen contains long stretches of the three-residue repeating unit Gly-X-Y, in which X is often Pro, and Y is often 4-hydroxyPro (Hyp; hydroxylation of Pro residues requires ascorbic acid). Each polypeptide strand forms a left-handed helix with around three residues per turn. Three such parallel chains then wrap around each other in gentle right-handed coil. Interchain hydrogen bonds connect all three strands, which are staggered so that a Gly residue occurs at every position along the triple helix. Covalent cross-links between modified Lys residues (allysine) and His residues bind together aggregates of collagen fibers.

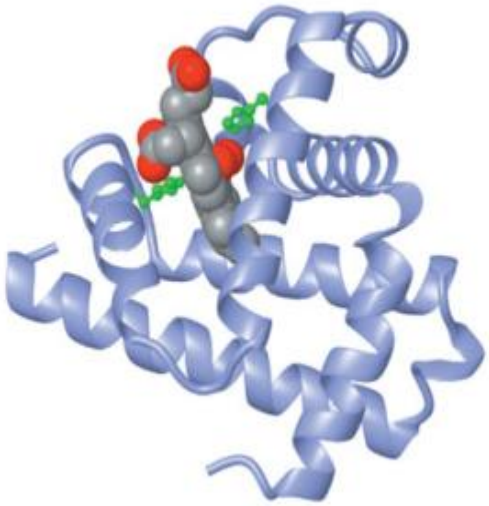
Properties of Some Fibrous Proteins

Structure	Characteristics	Examples of occurrence
α Helix, cross-linked by disulfide bonds	Tough, insoluble protective structures of varying hardness and flexibility	α-Keratin of hair, feathers, nails
β Conformation	Soft, flexible filaments	Silk fibroin
Collagen triple helix	High tensile strength, without stretch	Collagen of tendons, bone matrix

Structure of globular hemeproteins: Hemoglobin and myoglobin



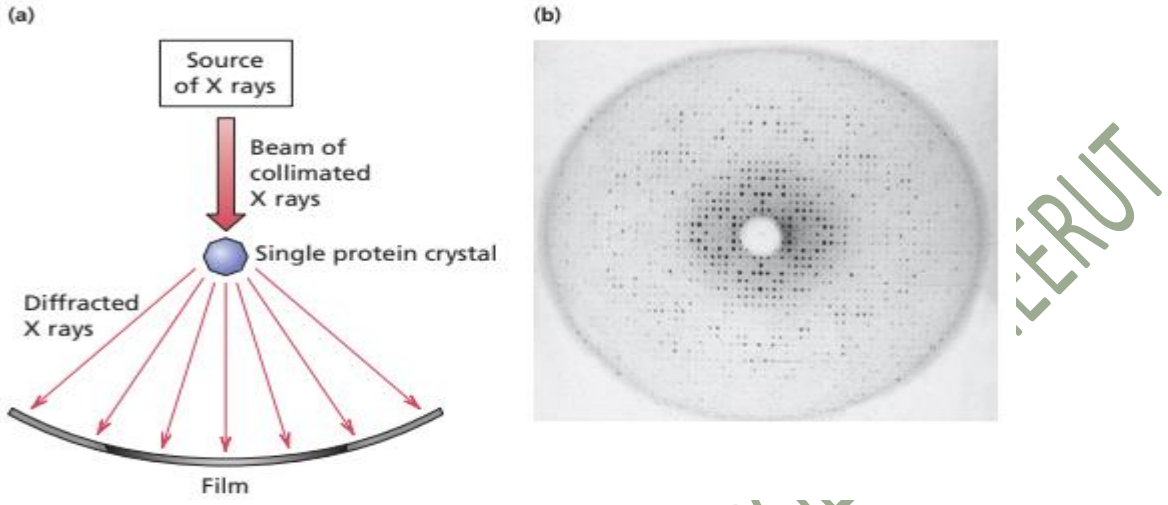
Hemoglobin is an abundant tetrameric globular protein in red blood cells. It has a quaternary structure and contains two copies of α-globin **(green)** and two copies of β-globin **(blue).** Each of these four polypeptide chains contains a heme molecule **(red)**, which is the site that binds oxygen (O2). Thus, each molecule of hemoglobin in the blood carries four molecules of oxygen.



Myoglobin was the **first protein** to have 3D structure solved by X-ray crystallography by John Kendrew in 1959. It is a monomeric globular protein in vertebrate muscle to store the oxygen and made up of single polypeptide chain of 153 amino acid residues that is folded into eight α-helices **(blue).** The heme prosthetic group is located within a hydrophobic cleft of polypeptide chain. The heme prosthetic group binds oxygen **(red).** His-64 **(green)** forms a hydrogen bond with oxygen and His-93 **(green)** is complexed to the iron atom of the heme.

Protein structure determination- The presence of α-helices and β-pleated sheets in proteins can often be predicted from the primary amino acid sequence, it is not possible to predict the precise three-dimensional structure of a protein

from its amino acid sequence, unless its sequence is very similar to that of a protein whose three-dimensional structure is already known. The three-dimensional structure of a protein can be determined to the atomic level by the techniques of X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and cryoelectron microscopy.



(a) Diagram of X rays diffracted by a protein crystal. (b) X-ray diffraction pattern of a crystal of deoxyhemoglobin adult human.

X-ray crystallography- In X-ray crystallography the first requirement are crystals of highly purified protein. In the crystal millions of protein molecules are precisely aligned with one another in a rigid array characteristic of that particular protein. Beams of X-rays are then passed through the crystal. The wavelengths of X-rays are 0.1—0.2 nm, short enough to resolve the atoms in the protein crystal. The atoms in the crystal scatter the X-rays, producing a diffraction pattern of discrete spots on photographic film. The intensities of the diffraction maxima (the darkness of the spots on the film) are then used mathematically to construct the three dimensional image of the protein crystal.

Nuclear magnetic resonance (NMR) spectroscopy- This technique can be used to determine the three-dimensional structures of small (up to approximately 30 kDa) proteins in aqueous solution. It does not require the crystallization of the protein. In this technique, a concentrated protein solution is placed in a magnetic field and the effects of different radio frequencies on the spin of different atoms in the protein measured. The behaviour of any particular atom is influenced by neighbouring atoms in adjacent residues, with closer residues causing more perturbation than distant ones. From the magnitude of the effect, the distances between residues can be calculated and then used to generate the three-dimensional structure of the protein.

Cryoelectron microscopy- This is often used to determine the three-dimensional structures of multisubunit proteins that are difficult to crystallize. In this technique, the protein sample is rapidly frozen in liquid helium to preserve its

structure. The frozen, hydrated protein is then examined in a cryoelectron microscope using a low dose of electrons to prevent radiation induced damage to the structure. The resulting images are analyzed by complex computer programs and the three-dimensional structure of the protein reconstructed.

Protein folding- Under appropriate physiological conditions proteins spontaneously fold into their native conformation. The primary structure of the protein dictates its three-dimensional structure. It has also been observed that protein folding is driven primarily by hydrophobic forces and proceeds through an ordered set of pathways. In vivo (in the cell) this process requires only a few minutes because the cells contain accessory proteins which assist the polypeptides to fold to their native conformation. Accessory proteins: **Protein disulfide isomerases** catalyze disulfide interchange reactions, thereby facilitating the shuffling of the disulfide bonds in a protein until they achieve their correct pairing. **Peptidyl prolyl cis–trans isomerases** catalyze the interconversion of the cis and trans isomers of peptide bonds with the amino acid Pro, thereby accelerating the folding of Pro-containing polypeptides. One of the classes of peptidyl prolyl cis–trans isomerases is inhibited by the **immunosuppressive drug cyclosporin A**.

Molecular chaperones, which include proteins such as the **heat shock proteins 70 (Hsp 70)**, the **chaperonins**, and the **lectins calnexin and calreticulin**. These prevent the improper folding and aggregation of proteins that may otherwise occur as internal hydrophobic regions are exposed to one another.

Specificities of various proteases (enzymes cleave the proteins)

Proteases with source	**Cleavage point**
Trypsin (bovine pancreas)	**Lys, Arg (C)**
Chymotrypsin (bovine pancreas)	**Phe, Trp, Tyr (C)**
Pepsin (bovine gastric mucosa)	**Leu, Phe, Trp, Tyr (N)**
peptide bond cleavage occurs on either the carbonyl (C) or the amino (N) side of the indicated amino acid residues.	

Role of chemicals in protein sequencing

A. Dansyl chloride- A reagent used in sequencing polypeptides from the N-terminus.

B. Phenylisothiocyanate- An Edman's reagent used to identify N-terminal residues.

C. Iodoacetate- An alkylating agent that reacts with Cys residues.

D. 2-mercaptoethanol- A reducing agent that carries out reductive cleavage of disulfide bonds.

E. Cyanogen bromide A reagent that cleaves polypeptides at Met (C) into smaller fragments.

Reference

1. Alberts B et al. (2015) in "The Molecular biology of the cell", 6th edition Garland Science, New York.
2.Voet D, Voet J (2018) in "Biochemistry", 5th edition, J. Wiley & Sons, USA
3. Richard A. Harvey, Denise R Ferrier (2014) in "Lippincott's Illustrated Reviews: Biochemistry", 6th edition, Lippincott Williams &Wilkins, USA.
4. Robert Horton H. et al. (2012) in "Principles of Biochemistry", 5th edition Pearson, New York.