# IDHAYA COLLEGE FOR WOMEN, KUMBAKONAM

## DEPARTMENT OF MICROBIOLOGY

| | | |
|---|---|---|
| **COURSE** | **:** | **B.Sc., II MICROBIOLOGY** |
| **SEMESTER** | **:** | **IV** |
| **TOPIC** | **:** | **UNIT-V** |
| **SUBJECT NAME** | **:** | **BIOINFORMATICS & COMPUTER APPLICATION IN BIOLOGY** |
| **SUBJECT CODE** | **:** | **16SACBS2** |

**Presented by**

**Dr.M.NITHYA M.Sc., Ph.D., PGDCA**

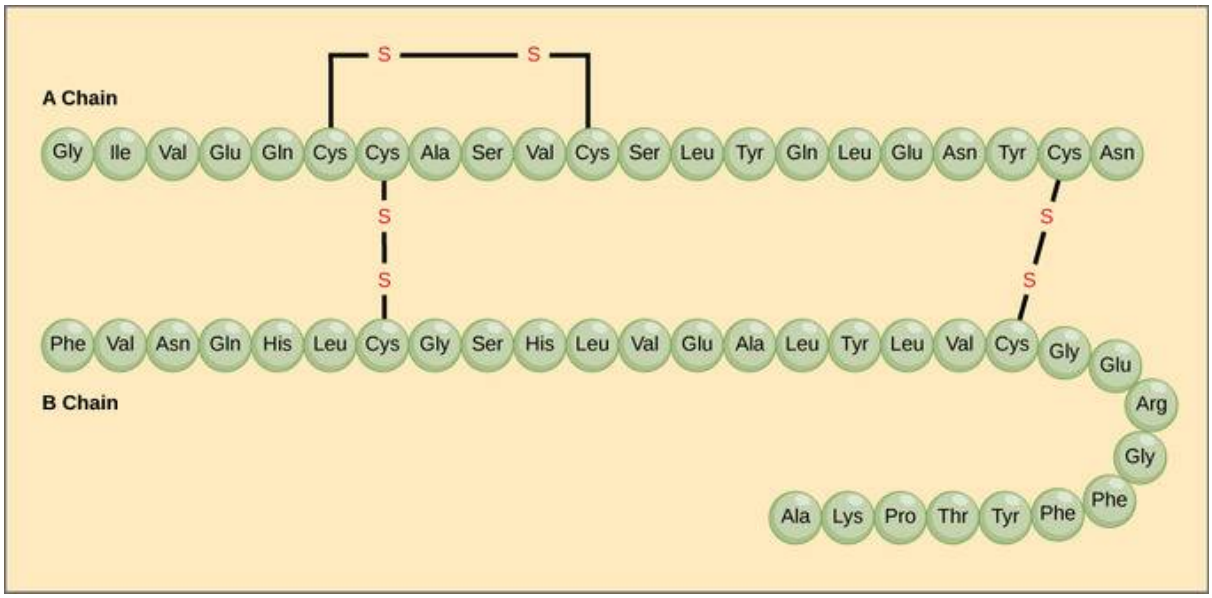
**Assistant Professor,**

**Dept. of Microbiology**

## STRUCTURE OF PROTEIN

- Protein structure depends on its amino acid sequence and local, low-energy chemical bonds between atoms in both the polypeptide backbone and in amino acid side chains.
- Protein structure plays a key role in its function; if a protein loses its shape at any structural level, it may no longer be functional.
- Primary structure is the amino acid sequence.
- Secondary structure is local interactions between stretches of a polypeptide chain and includes α-helix and β-pleated sheet structures.
- Tertiary structure is the overall the three-dimension folding driven largely by interactions between R groups.
- Quarternary structures is the orientation and arrangement of subunits in a multi-subunit protein.

- **Primary Structure**
- A protein's primary structure is the unique sequence of amino acids in each polypeptide chain that makes up the protein. Really, this is just a list of which amino acids appear in which order in a polypeptide chain, not really a structure. But, because the final protein structure ultimately depends on this sequence, this was called the primary structure of the polypeptide chain. For example, the pancreatic hormone insulin has two polypeptide chains, A and B.



**Primary structure**

- The A chain of insulin is 21 amino acids long and the B chain is 30 amino acids long, and each sequence is unique to the insulin protein.
- The gene, or sequence of DNA, ultimately determines the unique sequence of amino acids in each peptide chain. A change in nucleotide sequence of the gene's coding region may lead to a different amino acid being added to the growing polypeptide chain, causing a change in protein structure and therefore function.
- The oxygen-transport protein hemoglobin consists of four polypeptide chains, two identical α chains and two identical β chains. In sickle cell anemia, a single amino substitution in the hemoglobin β chain causes a change the structure of the entire protein. When the amino acid glutamic acid is replaced by valine in the β chain, the polypeptide folds into an slightly-different shape that creates a dysfunctional hemoglobin protein. So, just one amino acid substitution can cause dramatic changes. These dysfunctional hemoglobin proteins, under low-oxygen conditions, start associating with one another, forming long fibers made from millions of aggregated hemoglobins that distort the red blood cells into crescent or "sickle" shapes, which clog arteries . People affected by the disease often experience breathlessness, dizziness, headaches, and abdominal pain.

**Secondary Structure**

- A protein's secondary structure is whatever regular structures arise from interactions between neighboring or near-by amino acids as the polypeptide starts to fold into its functional three-dimensional form. Secondary structures arise as H bonds form between local groups of amino acids in a region of the polypeptide chain. Rarely does a single secondary structure extend throughout the polypeptide chain. It is usually just in a section of the chain. The most common forms of secondary structure are the α-helix and β-pleated sheet structures and they play an important structural role in most globular and fibrous proteins.

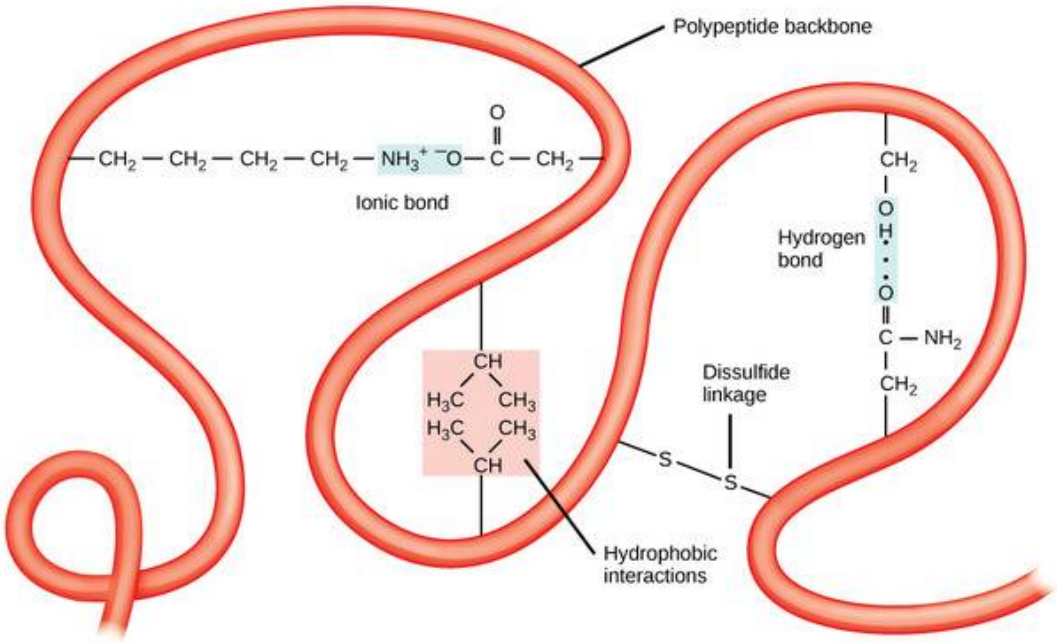
  The α-helix and β-pleated sheet form because of hydrogen bonding between carbonyl and amino groups in the peptide backbone. Certain amino acids have a propensity to form an α-helix, while others have a propensity to form a β-pleated sheet.
- In the α-helix chain, the hydrogen bond forms between the oxygen atom in the polypeptide backbone carbonyl group in one amino acid and the hydrogen atom in the polypeptide

backbone amino group of another amino acid that is four amino acids farther along the chain. This holds the stretch of amino acids in a right-handed coil. Every helical turn in an alpha helix has 3.6 amino acid residues. The R groups (the side chains) of the polypeptide protrude out from the α-helix chain and are not involved in the H bonds that maintain the α-helix structure.

In β-pleated sheets, stretches of amino acids are held in an almost fully-extended conformation that "pleats" or zig-zags due to the non-linear nature of single C-C and C-N covalent bonds. β-pleated sheets never occur alone. They have to held in place by other β-pleated sheets. The stretches of amino acids in β-pleated sheets are held in their pleated sheet structure because hydrogen bonds form between the oxygen atom in a polypeptide backbone carbonyl group of one β-pleated sheet and the hydrogen atom in a polypeptide backbone amino group of another β-pleated sheet. The β-pleated sheets which hold each other together align parallel or antiparallel to each other. The R groups of the amino acids in a β-pleated sheet point out perpendicular to the hydrogen bonds holding the β-pleated sheets together, and are not involved in maintaining the β-pleated sheet structure.

**Tertiary Structure**

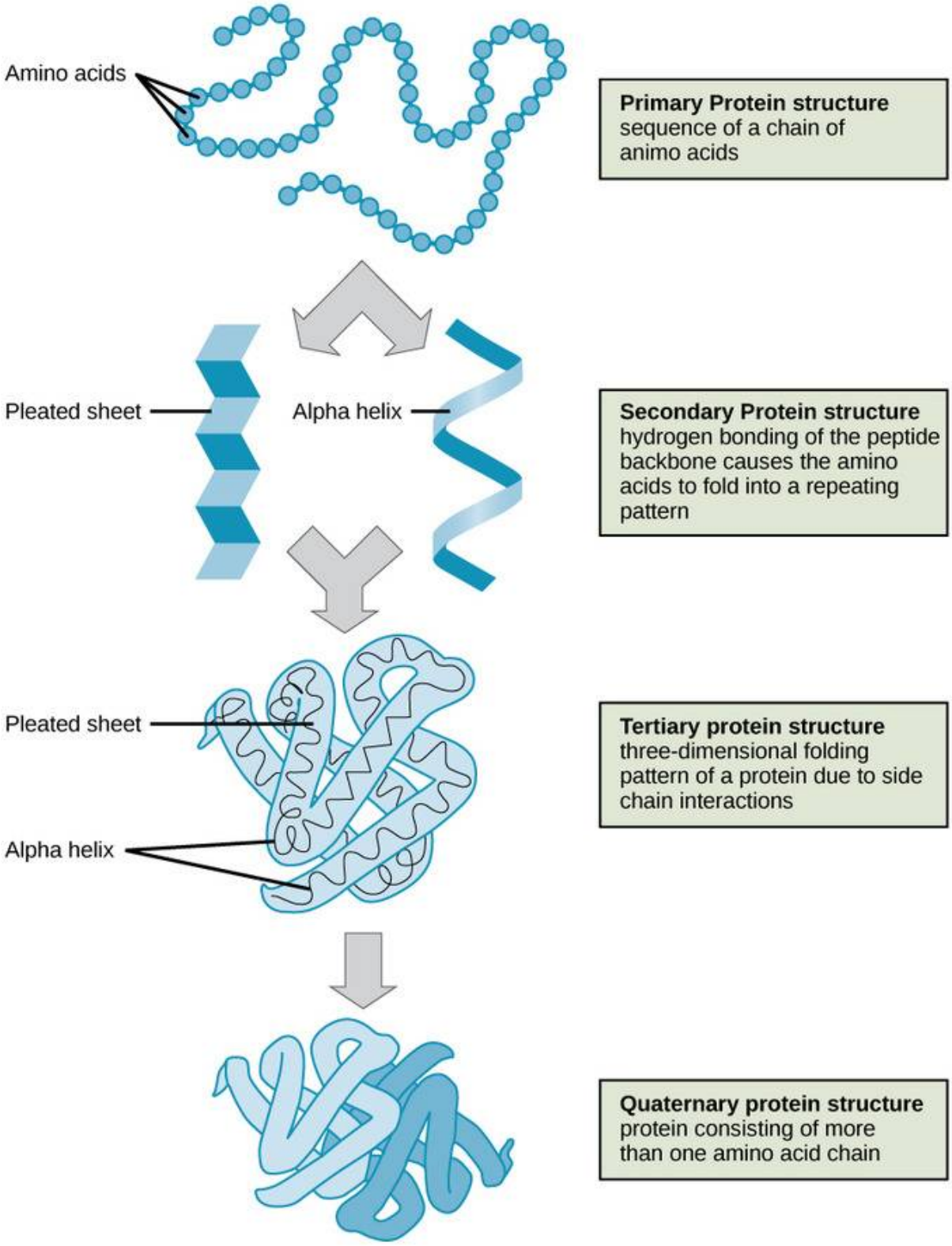
- The tertiary structure of a polypeptide chain is its overall three-dimensional shape, once all the secondary structure elements have folded together among each other. Interactions between polar, nonpolar, acidic, and basic R group within the polypeptide chain create the complex three-dimensional tertiary structure of a protein. When protein folding takes place in the aqueous environment of the body, the hydrophobic R groups of nonpolar amino acids mostly lie in the interior of the protein, while the hydrophilic R groups lie mostly on the outside. Cysteine side chains form disulfide linkages in the presence of oxygen, the only covalent bond forming during protein folding. All of these interactions, weak and strong, determine the final three-dimensional shape of the protein. When a protein loses its three-dimensional shape, it will no longer be functional.




- **Tertiary structure**The tertiary structure of proteins is determined by hydrophobic interactions, ionic bonding, hydrogen bonding, and disulfide linkages.

## Quaternary Structure

- The quaternary structure of a protein is how its subunits are oriented and arranged with respect to one another. As a result, quaternary structure only applies to multi-subunit proteins; that is, proteins made from one than one polypeptide chain. Proteins made from a single polypeptide will not have a quaternary structure.
- In proteins with more than one subunit, weak interactions between the subunits help to stabilize the overall structure. Enzymes often play key roles in bonding subunits to form the final, functioning protein.
- For example, insulin is a ball-shaped, globular protein that contains both hydrogen bonds and disulfide bonds that hold its two polypeptide chains together. Silk is a fibrous protein that results from hydrogen bonding between different β-pleated chains.




- **Four levels of protein structure**The four levels of protein structure can be observed in these illustrations.

- **CLASSIFICATION-PDB, Swiss-PROT, SCOP, CATH**

- The **Protein Data Bank** (**PDB**) is a database for the three-dimensional structural data of large biological molecules, such as proteins and nucleic acids. The data, typically obtained by X-ray crystallography, NMR spectroscopy, or, increasingly, cryo-electron microscopy, and submitted by biologists and biochemists from around the world, are freely accessible on the Internet via the websites of its member organisations (PDBe, PDBj, RCSB, and BMRB). The PDB is overseen by an organization called the Worldwide Protein Data Bank, wwPDB.

- The PDB is a key in areas of structural biology, such as structural genomics. Most major scientific journals, and some funding agencies, now require scientists to submit their structure data to the PDB. Many other databases use protein structures deposited in the PDB. For example, SCOP and CATH classify protein structures, while PDBsum provides a graphic overview of PDB entries using information from other sources, such as Gene ontology.

  The file format initially used by the PDB was called the PDB file format. The original format was restricted by the width of computer punch cards to 80 characters per line. Around 1996, the "macromolecular Crystallographic Information file" format, mmCIF, which is an extension of the CIF format was phased in. mmCIF became the standard format for the PDB archive in 2014. In 2019, the wwPDB announced that depositions for crystallographic methods would only be accepted in mmCIF format

  An XML version of PDB, called PDBML, was described in 2005. The structure files can be downloaded in any of these three formats, though an increasing number of structures do not fit the legacy PDB format. Individual files are easily downloaded into graphics packages from Internet URLs:

  The "4hhb" is the PDB identifier. Each structure published in PDB receives a four-character alphanumeric identifier, its PDB ID. (This is not a unique identifier for biomolecules, because several structures for the same molecule—in different environments or conformations—may be contained in PDB with different PDB IDs.

  The structure files may be viewed using one of several free and open source computer programs, including Jmol, Pymol, VMD, and Rasmol. Other non-free, shareware programs include ICM-Browser, MDL Chime, UCSF Chimera, Swiss-PDB Viewer, StarBiochem (a Java-based interactive molecular viewer with integrated search of protein databank), Sirius, and

VisProt3DS (a tool for Protein Visualization in 3D stereoscopic view in anaglyth and other modes), and Discovery Studio. The RCSB PDB website contains an extensive list of both free and commercial molecule visualization programs and web browser plugins.

## SWISS-PROT file format

| General information about the entry | |
|---|---|
| Entry name | **FA12_HUMAN** |
| Primary accession number | **P00748** |
| Secondary accession number(s) | None |
| Entered in SWISS-PROT in | Release 01, July 1986 |
| Sequence was last modified in | Release 12, October 1989 |
| Annotations were last modified in | Release 35, November 1997 |
| **Name and origin of the protein** | |
| Protein name | COAGULATION FACTOR XII [Precursor] |
| Synonym(s) | EC 3.4.21.38<br>HAGEMAN FACTOR<br>HAF |
| Gene name(s) | F12 |
| From | Homo sapiens (Human) |
| Taxonomy | Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Mammalia, Eutheria, Primates, Catarrhini, Hominidae, Homo |

## SWISS-PROT file format

**Comments**

- ***FUNCTION***: FACTOR XII IS A SERUM GLYCOPROTEIN THAT PARTICIPATES IN THE INITIATION OF BLOOD COAGULATION, FIBRINOLYSIS, AND THE GENERATION OF BRADYKININ AND ANGIOTENSIN.
- ***CATALYTIC ACTIVITY***: CLEAVES SELECTIVELY ARG-|-ILE BONDS AND ACTIVATES COAGULATION FACTORS VII AND XI.
- ***PTM***: O- AND N-GLYCOSYLATED.
- ***DISEASE***: DEFECTS IN F12 DO NOT CAUSE ANY CLINICAL SYMPTOMS. THE SOLE EFFECT IS THAT WHOLE-BLOOD CLOTTING TIME IS PROLONGED.
- ***MISCELLANEOUS***: FACTOR XII, PREKALLIKREIN, AND HMW KININOGEN FORM A COMPLEX BOUND TO AN ANIONIC SURFACE. PREKALLIKREIN IS CLEAVED BY FACTOR XII TO FORM KALLIKREIN, WHICH THEN CLEAVES FACTOR XII FIRST TO ALPHA-FACTOR XIIA AND THEN TO BETA-FACTOR XIIA. ALPHA-FACTOR XIIA ACTIVATES FACTOR XI TO FACTOR XIA.
- ***SIMILARITY***: CONTAINS 2 EGF-LIKE DOMAINS.
- ***SIMILARITY***: CONTAINS 1 FIBRONECTIN TYPE-I DOMAIN.
- ***SIMILARITY***: CONTAINS 1 FIBRONECTIN TYPE-II DOMAIN.
- ***SIMILARITY***: CONTAINS 1 KRINGLE REGION.
- ***SIMILARITY***: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE TRYPSIN FAMILY.

## PROTEIN VISUALIZATION TOOLS-RASMOL, Swiss PDB Viewer

### RasMol & PDB VIEWER

RasMol is a computer program written for molecular graphics visualization intended and used mainly to depict and explore biological macromolecule structures, such as those found in the Protein Data Bank. It was originally developed by Roger Sayle in the early 1990s.

Historically, it was an important tool for molecular biologists since the extremely optimized program allowed the software to run on (then) modestly powerful personal computers. Before RasMol, visualization software ran on graphics workstations that, due to their cost, were less accessible to scholars. RasMol continues to be important for research in structural biology, and has become important in education.

RasMol has a complex licensing version history. Starting with the version 2.7 series, RasMol source code is dual-licensed under a GNU General Public License (GPL), or custom license RASLIC.[3] Starting with version 2.7.5, a GPL is the only license valid for binary distributions.

RasMol includes a scripting language, to perform many functions such as selecting certain protein chains, changing colors, etc. Jmol and Sirius software have incorporated this language into their commands.

Protein Data Bank (PDB) files can be downloaded for visualization from members of the Worldwide Protein Data Bank (wwPDB). These have been uploaded by researchers who have characterized the structure of molecules usually by X-ray crystallography, protein NMR spectroscopy, or cryo-electron microscopy.

### Swiss-PdbViewer

Swiss-PdbViewer (aka DeepView) is an application that provides a user friendly interface allowing to analyze several proteins at the same time. The proteins can be superimposed in order to

deduce structural alignments and compare their active sites or any other relevant parts. Amino acid mutations, H-bonds, angles and distances between atoms are easy to obtain thanks to the intuitive graphic and menu interface.

Swiss-PdbViewer (aka DeepView) has been developped since 1994 by Nicolas Guex. Swiss-PdbViewer is tightly linked to SWISS-MODEL, an automated homology modeling server developed within the Swiss Institute of Bioinformatics (SIB) at the Structural Bioinformatics Group at the Biozentrum in Basel.

Working with these two programs greatly reduces the amount of work necessary to generate models, as it is possible to thread a protein primary sequence onto a 3D template and get an immediate feedback of how well the threaded protein will be accepted by the reference structure before submitting a request to build missing loops and refine sidechain packing.

Swiss-PdbViewer can also read electron density maps, and provides various tools to build into the density. In addition, various modeling tools are integrated and residues can be mutated.