

## Lecture 3

### Protein Structure-I

Proteins are the most abundant organic molecules in living cells. They may be monomeric with one polypeptide chain or multimeric having more than one chain.

In case of a homomultimer the chains are of one kind whereas for a heteromultimer two or more different chains form the protein. (e.g. Hemoglobin is a heterotetramer. It has two alpha chains and two beta chains.)

Proteins may be **simple** or **conjugated** –

Simple – composed only of amino acid residues

Conjugated – in addition to the polypeptide chain these proteins contain other non-amino acid components known as prosthetic groups (e.g. metal ions, cofactors, lipids, carbohydrates)

Example: Hemoglobin – Heme

Each polypeptide chain which is a polymer of amino acids linked by peptide bonds can be classified according to their shape and/or function.

Some common terms:

- Peptide = a short chain of amino acids
- Polypeptide = a longer chain of amino acids
- Protein = a polypeptide that occurs in nature and folds into a defined three-dimensional structure

**This means that ALL proteins are polypeptides but not all polypeptides are proteins.**

The only other covalent bonds present in proteins apart from the peptide bonds (or any post translational modification that may form covalent bonds) are disulfide bonds.

The disulfide bonds are able to link parts of the polypeptide chain that may otherwise be separated by a large number of amino acid residues.

### Protein Structure

- Physical properties of proteins that influence stability
  - Rigidity of backbone
  - Amino acid interactions with water
  - Interactions among amino acids
    - Electrostatic interactions
    - Hydrogen bonds
    - S-S bonds
    - Volume constraints

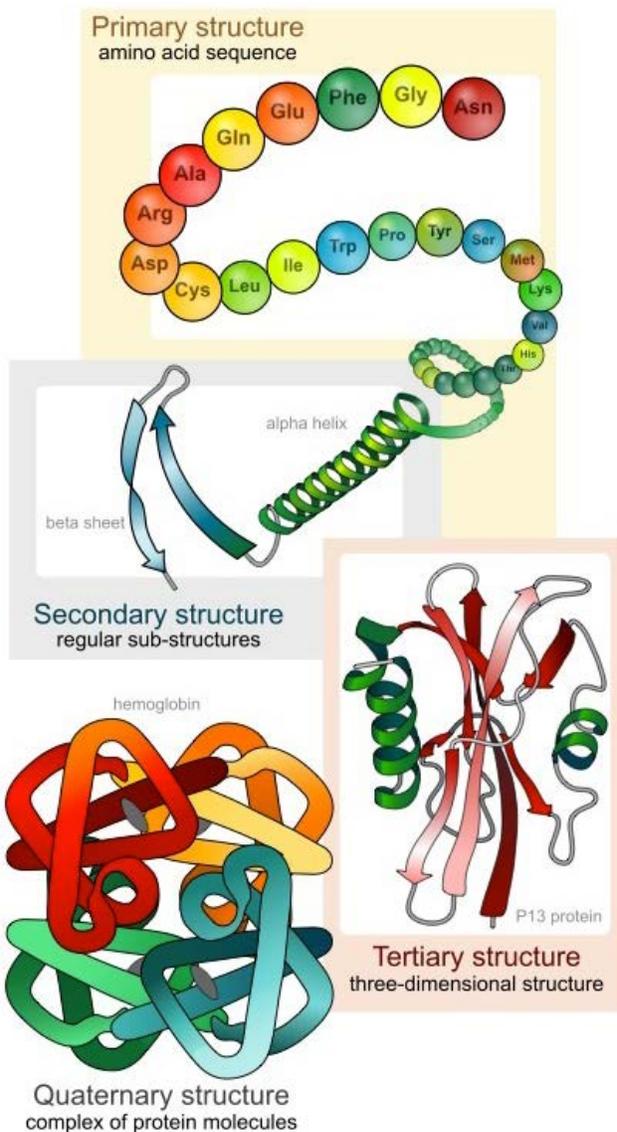
## Protein Hierarchy

Primary structure (1°) : the amino acid sequence.

Secondary structure (2°) : helices, sheets and turns.

Tertiary structure (3°) : side chain packing in the 3-D structure.

Quaternary structure (4°) : association of subunits.



from wikipedia

The **primary structure** is the sequence of amino acid side chains along the backbone of peptide bonds in a polypeptide. Considering the 20 common amino acids only, with no limitation as to the length or the type of amino acid to be linked, there are a huge number of possibilities for the primary structure.

For example if we would want to determine how many primary structures with three amino acids are possible, we would see that there are 8000 possibilities.

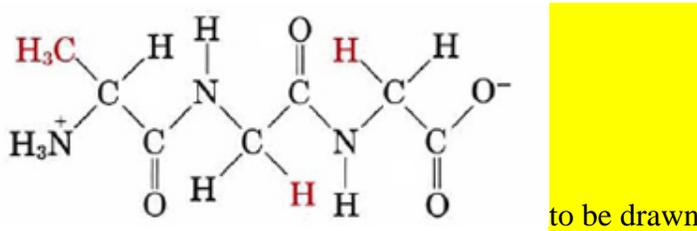
There are 20 possibilities for the first place in the chain. Similarly there are 20 possibilities for the second place in the chain, making a total of  $20 \times 20 = 400$  possible combinations for a dipeptide. Extending this to a tripeptide means there are  $400 \times 20 = 8000$  possibilities!

With the polypeptide chains ranging from 20-1000+ residues the possibilities are immense.

The primary structure of a protein is conventionally specified by writing the three-letter abbreviations for each amino acid. The amino terminal is the  $-\text{NH}_3^+$  end of the polypeptide chain. This is the convention because protein synthesis starts from the N-terminal end

In some cases, this is even shortened to the 1-letter abbreviation sequence.

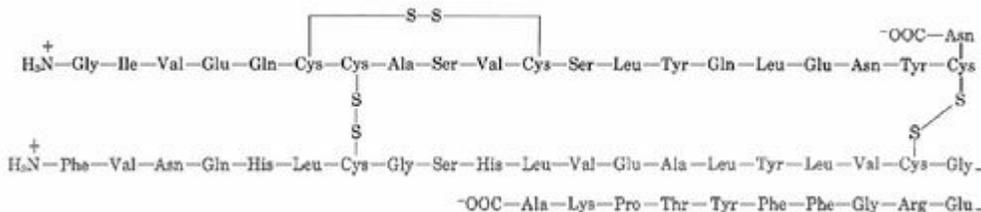
For example, the structure



which reading from the  $-\text{NH}_3^+$  end is alanine, glycine, glycine would be specified as **Ala-Gly-Gly** in the three-letter code format and simply as **AGG** in the one letter code format

The primary sequence of a protein determines the three-dimensional shape that ultimately relates to the function of the protein. Therefore a lot of effort has gone into determining the primary sequence of proteins. The common method for amino acid sequence determination is called [Edman degradation](#).

The first protein whose amino acid sequence was determined is insulin. This was determined by Frederick Sanger in 1953 after many years of effort. There are two chains in this structure that are linked together by disulfide bonds between Cys residues of the different chains.

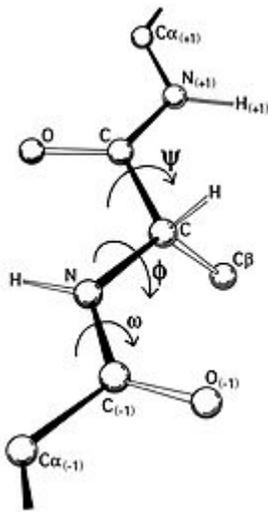


The **secondary structure** involves a specific geometry of the polypeptide backbone where the backbone atoms are linked by hydrogen bonds. The defined backbone angles depicted by the  $\phi$  and  $\psi$  angles result in a regular pattern of the polypeptide chain.

We have seen what the  $\phi$  and  $\psi$  angles are in the previous lecture and realize that rotation about these bonds can change the corresponding dihedral angles of the backbone. This brings about conformational flexibility to the molecule and allows distant regions of the polypeptide chain to come close to one another.

As we have seen earlier the other possibility that may bring distant part of the polypeptide close to one another is the disulfide bond.

The question now is – do the backbone angles have any regularity or are they completely random? To answer this question, we need to understand a bit more of the geometry or conformation of protein structures.



## The $\alpha$ -helix

The  $\alpha$ -helix is one of two secondary structures (the other being the  $\beta$  -sheet) predicted and discovered by Linus Pauling in 1951.

### Helix nomenclature:

- 1 Hydrogen bond (dotted line) between



(residue i)            (residue i+3/4/5)

Repeating unit:

(i)  $3_{10}$  helix

- 10 atoms
- 3 residues per turn
- i to i+3 hydrogen bonding

(ii)  $3.6_{13}$  helix

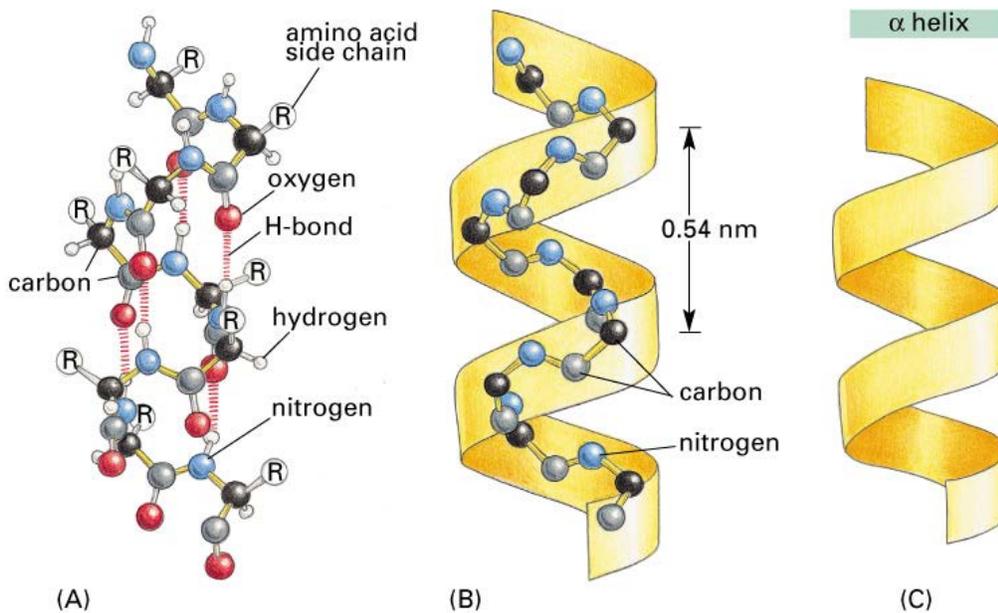
- 13 atoms
- 3.6 residues per turn
- i to i+4 hydrogen bonding
- 

(iii)  $5_{18}$  helix

- 18 atoms
- 5 residues per turn
- i to i+5 hydrogen bonding

A regular right-handed  $\alpha$ -helix has the following spatial parameters:

$\phi$	=	-57°
$\psi$	=	-47°
n = 3.6	(number of residues per turn)	
pitch 0.54nm (or 5.4Å)		

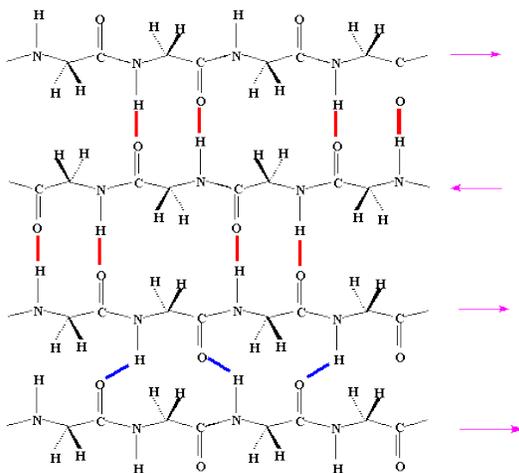


## The $\beta$ -sheet

The  $\beta$ -sheet forms a result of H-bonding between polypeptide chains that are referred to as strands in this case.

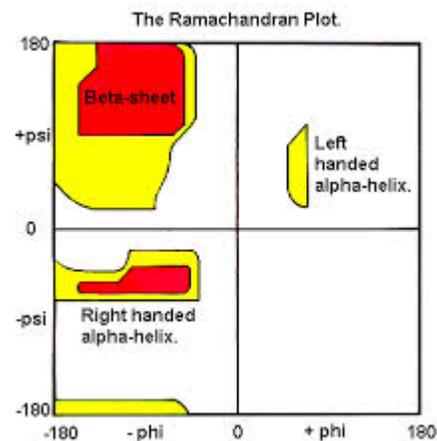
$\beta$ -sheet: parallel ( $\phi = -119$  and  $\psi = 113$ ) or anti-parallel pleated sheet structures

A four-stranded beta sheet that contains three antiparallel strands and one parallel strand is shown. Hydrogen bonds are indicated with red lines (antiparallel strands) and blue lines (parallel strands). Note the difference in the hydrogen bonding pattern.



A Ramachandran plot, often referred to as a  $\phi, \psi$  plot was developed in 1963 by G. N. Ramachandran and colleagues to be able to understand the geometrical features of the backbone. It was demonstrated that the backbone dihedral angles  $\psi$  against  $\phi$  of amino acid residues in the protein structure are unique for specific secondary structural elements as shown.

The plot is usually used to determine which conformations are possible for a protein but its usage is very important in structure validation of proteins. Since some regions of the map are forbidden due to geometrical restraints in the backbone, the calculations of the dihedral angles serve as a measure of the structural integrity of the protein structure.



Ramachandran, G.N.; Ramakrishnan, C.; Sasisekharan, V. (1963). "Stereochemistry of polypeptide chain configurations". *Journal of Molecular Biology* 7: 95–9. [doi:10.1016/S0022-2836\(63\)80023-6](https://doi.org/10.1016/S0022-2836(63)80023-6)

Secondary structure elements are connected by regions of *turns* and *loops*

Turns – short regions of non- $\alpha$ , non- $\beta$  conformation. These are also referred to the coil regions of the protein and they act as linkers of the secondary structural elements the  $\alpha$ -helix and  $\beta$ -sheet.

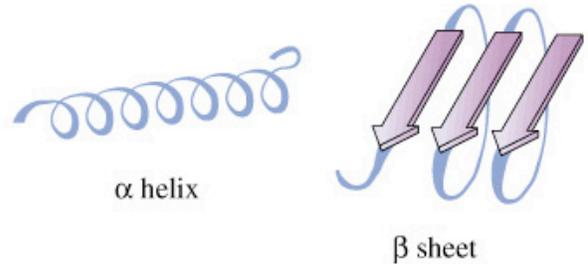
This leads to the formation of a tertiary structure that is held together primarily by noncovalent interactions which we will see later.

The different levels of structure are given below:

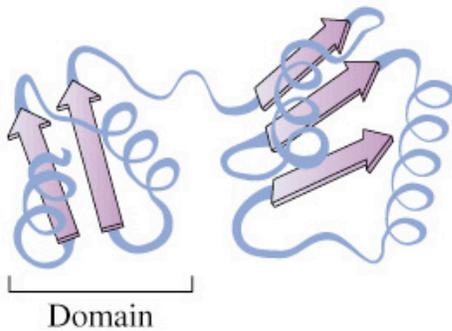
(a) Primary structure

–Ala–Glu–Val–Thr–Asp–Pro–Gly–

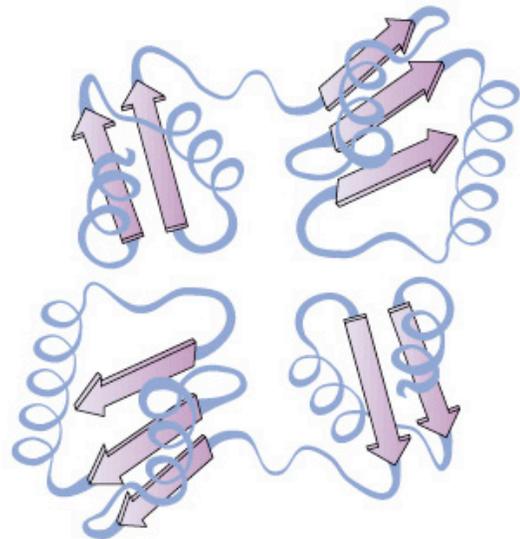
(b) Secondary structure



(c) Tertiary structure



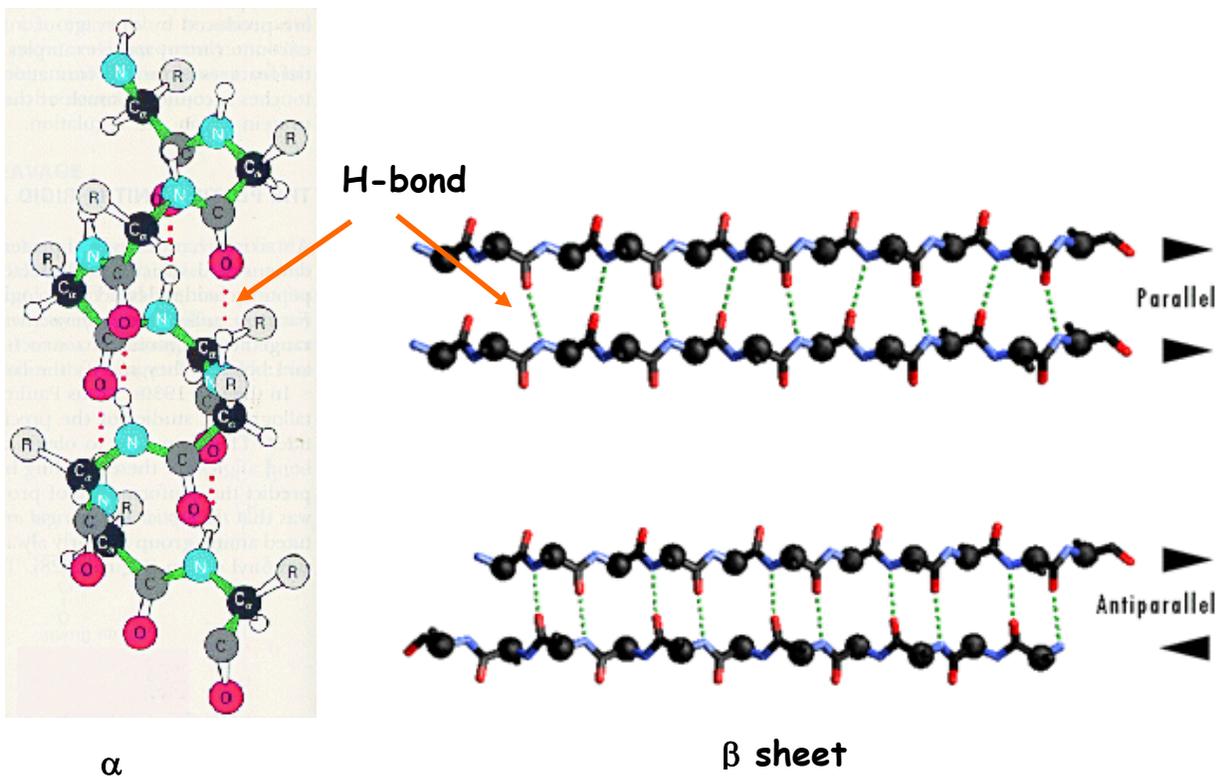
(d) Quaternary structure



### Principle forces in protein folding:

- 1) van der Waals interactions along with hydrophobic forces that occur favors packing of the amino acid interactions and are responsible for the formation of a hydrophobic core of the protein. This exploits the hydrophobic tendency of non-polar amino-acid side-chains.
- 2) Polar side-chain interactions occur either through hydrogen bonds or electrostatic interactions.
- 3) Disulfide bond formation: provides stability

Structure Stabilizing Interactions in secondary structure of proteins are primarily hydrogen bonding as seen earlier. This has a higher interaction energy than a simple van der Waals interaction



Overall interactions that comprise the tertiary structure of proteins

