

# Structural levels of proteins and stabilizing forces

## Introduction

Alpha-L-amino acids are building blocks of proteins. There are 20 naturally occurring amino acids. Each amino acid has an amino group, carboxyl group, a hydrogen atom and a side chain which can be  $-H$  (glycine) or  $-R$  attached to C alpha carbon atom. There are also many modified amino acids but they occur rarely.

All existing proteins are made up of these 20 amino acids covalently linked by a bond known as peptide bond. Peptide bond is formed upon condensation of carboxyl group of one amino acid with amino group of another with the loss of a water molecule.

There are various levels of organization of protein structure which includes: primary, secondary, tertiary and quaternary.

### 3.1 Primary structure of protein

The primary structure of a protein is defined by amino acid sequence from N-terminal to C-terminal end and disulfide bonds. Amino acids are joined through peptide bonds giving rise to a polypeptide chain. During the formation of polypeptide chain, amino group of first amino acid and carboxyl group of last amino acid remain intact, and thus polypeptide chain is said to be extended from its amino terminus to its carboxy terminus. Formation of polypeptide chain generates "main chain" or "backbone" from which side chain of various amino acids projects. The primary structure basically includes all the covalent bonds in the protein structure.

### 3.2 Secondary structure

Secondary structure of protein arises due to non-covalent interactions between amino acids. Hydrogen bond interaction of backbone NH and CO defines the one of the two type of secondary structure: Alpha helices and beta sheets. Linus Pauling and Robert Corey were first to describe the secondary structure of protein.

#### 3.2.1 Alpha helix

A polypeptide when folded in a right handed helical form along an imaginary axis is called alpha-helix. In alpha-helix, generally side chains are stretched away (or perpendicular to the axis of helix) from the axis of the helix. Alpha helices are found in the protein when a stretch of polypeptide chain possesses phi, psi angle approximately  $-57^\circ$  and  $-47^\circ$  respectively. This corresponds to the allowed region in bottom left or 3<sup>rd</sup> quadrant of the Ramachandran plot.

Right handed alpha helix has 3.6 residues per turn and there is rise of 5.4 Å per turn. In alpha helix, hydrogen bonds are formed between C=O of n residue with the NH of n+4 residue. Thus, all NH and C=O in the alpha helix are hydrogen bonded except

the first NH group and the last CO group, leading to polar at the end of helix. Moreover, all C=O of the helix are in one direction of axis leading to dipole formation.

Some of the amino acids are preferred in alpha helices such as glutamic acid, alanine, histidine, leucine and methionine. Membrane spanning helices generally have hydrophobic amino acid. Cluster of beta branched side chain amino acids hinder helix formation. Proline is known to destabilize helix, as it lacks NH group in backbone & thus cannot participate in form hydrogen-bonding and also cause steric hindrance to the alpha helical conformation. Flexibility of glycine also causes distortion in the helical conformation.

Alpha helix in globular protein structures is commonly found along the outside of the protein, where it faces solution at one side and hydrophobic interior on the other side. Solvent exposed side may experience multiple hydrogen bond between C=O and NH of backbone with water molecule of solvent, which causes distortion / curved structure in  $\alpha$ -helices.

### **3.2.1.1 $\alpha$ -helix in helical/ spiral wheel**

Helical wheel or spiral wheel is a convenient way of representation of amino acid sequences in helices. Since, one turn in helix corresponds to 3.6 residues long, each residue in helical wheel can be plotted at  $360/3.6=100$  around a circle. With this type of representation, location and type of the helix in a protein can be predicted.

### **3.2.2 Beta-sheet**

$\beta$ -sheet is commonly known as extended structure. In  $\beta$ -sheet, two or more polypeptide chain run alongside and forms hydrogen bonds between C=O group of one  $\beta$  strand and NH group of adjacent  $\beta$  strand and vice-versa. The alternate side chains are opposite sides of the strand, similarly alternate C=O groups are opposite sides and C=O groups and side chains are perpendicular to each other.

Beta strands in a  $\beta$ -sheet can interact in two different ways leading to formation of parallel and anti-parallel  $\beta$ -sheet. In parallel  $\beta$ -sheet, amino acids in the all aligned  $\beta$ -strands have same biochemical direction, i.e. amino terminal to carboxy terminal. While, in antiparallel  $\beta$ -sheet, amino acids in adjacent  $\beta$ -strand have alternating direction, amino terminal to carboxy terminal and then followed by carboxy terminal to amino terminal and so on. Beta-sheet can also combine some parallel  $\beta$ -strand and some antiparallel to give rise mixed  $\beta$ -sheet. The  $\beta$ -sheets in protein structures are known to have twisted strands and the twisting of sheets could be from  $25$  to  $30^\circ$ .

## **3.3 Tertiary structure**

A compact, asymmetric structure arises due to individual polypeptide chain is called tertiary structure. Tertiary structure arises due to interaction of side chains of amino acids bring all the secondary structures together. Each individual residue in the protein tertiary structure exists in most stable conformation. Hydrophobic amino acids are preferably found in the interior of protein, while charged residues are located on the surface. In tertiary structure, almost all hydrogen bond acceptors and donors are generally satisfied. Globular proteins are tightly packed in the centre &

loosely packed near the surface. Low packing density indicates towards flexibility, movement and functional role.

Sometimes, two cysteine residues present distantly from each other in polypeptide chain may come adjacent in three-dimensional structure. Adjacent placed cysteine can be oxidized to form a disulfide bridge.

### 3.4 Quaternary structure

In quaternary structure, different polypeptide chains or tertiary structures come together. Subunit in quaternary structure can be of same kind (homotropic) or of different type (heterotropic). However, quaternary structures generally have internal symmetry (e.g. Lactate dehydrogenase). There is cooperativity in substrate binding in quaternary structure. Mostly show allosteric regulation, can have activators or suppressors, which are different from substrates or products. Quaternary structure leads to genetic economy and efficiency of the system. The allosteric activators or suppressors could be related to the whole pathway, where these allosteric enzymes are part of it.

### 3.5 Stabilizing forces in protein structure.

#### 3.5.1 Covalent bond

1. Disulfide bond- disulfide bond is formed between two sulfur atoms of cysteine residues when they are located adjacent to each other in the tertiary structure. Typical S-S bond length is approximately 2.2Å. When there is a reducing physiological environment, S-S disulfide bonds exist in reduced form i.e. as two SH.

If, more than two cysteines are present in the protein sequence then, there are a number of ways S-S bond can form. An equation predicts possible number of ways in which disulfide bonds can be formed.

$(N-1)(N-3)N-5)...$  where N is number of cysteine present in the protein sequence.

For example protein with 8 cysteine can have  $7 \times 5 \times 3 \times 1 = 105$  possible ways the disulfide bridges can be formed.

Ribonuclease A, is a classical example to exhibit the above stated fact, it has 8 cysteine amino acid residues which ultimately forms 4 disulfide bonds. When the protein was treated with urea (8M) and DTT, it unfolded and all its disulfide bonds get reduced. Protein would again refold upon removal of urea and DTT from the solution. This removal can be done in two different ways, which yields two different results. When we removed DTT first followed by removal of urea, protein retains only 1% activity. This is due to formation of random disulfide bond formation between 8 cysteines present in the protein, which leads to about 105 conformations and out of which only one conformation is the active conformation. However, removal of urea followed by removal of DTT retains 99% of protein's activity after refolding. Thus from this experiment we can conclude that disulfide bridges do not play an important role during protein folding.

Insulin as an example doesn't refold similar to Ribonuclease A. A mature insulin has two polypeptide chains that are connected by two disulphide bonds, whereas pre-pro-insulin exists as a single polypeptide with three disulphide bonds. When we try to refold mature insulin it doesn't yield a fruitful result, but when we try refolding pre-pro-insulin, we can recover it fairly well. Nature has designed the proteins such that their native form and functional forms are different. Pre-pro insulin can be folded to native state, while mature insulin can't be folded to native form.

### 3.5.2 Non-covalent interactions

**3.5.2.1 Electrostatic interactions:** It is also known as salt bridges, these are formed between the charged interacting residues. In proteins, salt bridges are formed between carboxyl group of aspartic acid or glutamic acid with guanidinium group of arginine or amino group of lysine. Salt bridges in protein are about 2.8Å. Energy contributed by formation of salt bridge in protein structure is approximately 12-17 KJ/mol in water (on the surface) while its formation in the buried hydrophobic environment contributes about 30KJ/mol. In water their strength is less as water can also form ionic interactions with the residues. Thus, ionic interactions forming on the surface of the protein do not contribute much to the protein structure, while its formation in the buried leads to thermally stable protein.

**3.5.2.2 Hydrogen bonds-** Hydrogen bond is weak electrostatic interaction with a directional property. Hydrogen bond involves at least one hydrogen atom. In hydrogen bond, there is one hydrogen donor group and a hydrogen acceptor group. In proteins, hydrogen bonds are generally formed between atoms: N-H---O. Hydrogen bond distance is about 3Å (range from 2.7- 3.5 Å). Energy contributed by hydrogen bond in protein structure is approximately 2 to 6 KJ/mol in water and 12- 30 KJ/mol in buried hydrophobic environment.

#### **Importance of hydrogen bonds in structural stability For support of the argument**

- Melting point & boiling point of water is high as good as hydrogen bond but increase in temperature disrupt the hydrogen bond of protein that denatures it.
- Proteins have internal architecture with various helices and sheets. These secondary structures are stabilized by hydrogen bonds.

#### **Against the Argument**

- Solvents that bind to peptide through hydrogen bond should favor random coil but polyethylene glycol (PEG), glycerol stabilize the protein.
- N-methyl acetamide (NMA), which mimics the peptide backbone, forms dimers at low concentrations in chloroform, it needs to be present in high concentration in water.

If the Hydrogen bond is dominant force for protein structure determination, then proteins should have structure independent of amino acid sequence and should be regular and periodic, but it is not true. Generally In proteins, C=O groups form 9% beta sheet 25% helix, 5% turns, 19% loops and irregular remaining with water & ligands.

### **Metallic bonds**

It is a form of bond that exists in proteins formed due to electrostatic attractive forces. This bond formation is due to inherent property of metals to give up electrons. Many metals present in proteins form coordination complexes, to provide stability to the protein structure. Metals generally serve as co factors for many enzymes.

### **3.5.2.3 Hydrophobic interactions**

Hydrophobic interactions also known as Van der Waals interactions, can be seen when non-polar atoms gather together to minimize their contact with water. Oils and lipids cluster themselves to form micelle and reduce the surface area of hydrophobic groups and minimize their contact with water. Soluble, globular proteins tend to hide hydrophobic amino acids in the core of the protein, preventing their contact with water molecule. Attractive and repulsive forces in Van der Waals interaction can be described by “Lennard-Jones” potential.

### **Importance of hydrophobic interaction in structural stability**

#### **For support of the argument**

- Organic solvents denature the proteins
- Ionic strength & higher charged ions stabilize the proteins
- Proteins have hydrophobic residues in the core

#### **Against the Argument**

- Proteins have internal architecture
- None polar components associate more strongly as temp increases but Proteins denature as temperature is increased.

### **3.6 Total energy of the protein:**

Total energy of protein includes energy contributed by bonded and non-bonded interactions. Energy contributed by bonded interactions includes bond-stretch, bond bend and dihydral angle. While, energy contributed by non-bonded includes electrostatic, van-der waals and hydrogen bonding.

$$E_{\text{total}} = E_{\text{bonded}} + E_{\text{non-bonded}}$$

$$E_{\text{bonded}} = E_{\text{bond-stretch}} + E_{\text{bond-bend}} + E_{\text{dihydral angle}}$$

$$E_{\text{non-bonded}} = E_{\text{electrostatic}} + E_{\text{van-der_waals}} + E_{\text{H-bond}}$$

**3.6.1 Energy: Bond-stretch-** when two atoms are connected by a covalent bond, it maintain an equilibrium distance between them. If there is displacement of atoms from the equilibrium position leads to bond stretch energy, which contributes to the total potential energy.

**3.6.2 Energy: bond-bend/ angle** – When three atoms are connected by two covalent bonds, then those bonds tend to form a fixed angle between them. Any variation from this fixed equilibrium angle contributes toward the potential energy.

**3.6.3 Energy: dihedral angle-** Dihedral angle is defined by angle formed by two planes or it can be defined as the angle formed at middle bond of the three bonds connecting four atoms in the space. Deviation from equilibrium bond angle due to rotation about middle bond also contributes towards potential energy of the protein

**3.6.4 Non-bonded interactions:** The energy potential of non-bonded interactions is due to electrostatic, H-bonds and Van der Waals forces. These interactions are crucial for the stability of the tertiary structure. Energy due to electrostatic forces can be calculated using Coulomb's law, as equation given in slides. The energy contribution by Van der Waals forces can be calculated using the Lennard-Jones potential, as the equation given in the slides. The energy contribution by Van der Waals interactions / hydrophobic interactions is well explained by London dispersion forces. When two non-polar atoms come together, nucleus of one atom attracts the electron cloud of the other atom. Due to this, a dipole moment is induced in both atoms and these dipoles attract each other. But, as these two atoms come very close, electrons of both atoms and nucleus of both atoms repel each other and in best possible case equilibrium is achieved, where the atoms stay at certain distance. At this distance, best attractive force and least repulsive force are maintained. Hydrogen bond is another strong energy potential, which plays important role in protein structure, which explained in more detail in module 1.

Total energy of the protein helps in prediction of stability and lowest energy state of the molecule during energy minimization, refinement of NMR and crystal structures. Moreover, it is also helpful in determining binding energy upon interaction of protein-inhibitor, protein-ligand and protein-protein interactions. This kind of energy parameter is used for structure prediction of proteins and simulations of the protein structure.

### **3.7 Summary**

This module explains the structural levels of protein and different interactions, which stabilize the protein structure. Secondary structures ( $\alpha$ -helix and  $\beta$ -sheet) are explained in detail. Spiral wheel arrangement of  $\alpha$ -helix and how it can be used to predict the location of helix in the tertiary structure is discussed. Occurrence of Gly or Pro in the helix induces distortion or bend in the helix. The importance of covalent and non-covalent interactions in the protein structure is discussed. The total energy of the protein can be calculated by summation of individual potentials and how this parameter is used in different theoretical and experimental calculations is discussed.