Self-assembly of Proteins

Dr. K. Uma Maheswari

Professor, School of Chemical & Biotechnology

SASTRA University

Table of Contents

Module Objective

At the end of the module, the learner would be able to understand the factors influencing the self-assembly of proteins. The module also introduces to the reader the different types of self-assembled structures that can be achieved using intelligent design of the peptide sequences.

Preface

The great scientist **Richard Feynman** once said, *"I can appreciate the beauty of a flower. At the same time, I see much more about the flower than an artist sees. I could imagine the cells in there, the complicated actions inside, which also have a beauty. I mean it's not just beauty at this dimension, at one centimeter; there's also beauty at smaller dimensions, the inner structure, also the processes"*. The process of self-assembly of biomolecules is one such process that inspires awe and creates a mystique that adds to the excitement and initiates a quest for understanding and knowledge. This module aims to provide a small glimpse into the fascinating world of protein self-assembly to the learner. The self-assembly of peptides has numerous applications in the fields of medicine and electronics. The following sections focus on several factors that influence self-assembly of peptides and the important structures formed through spontaneous self-assembly by several designer peptides.

This lecture focuses on the basics of self-assembly and the structure of proteins.

1 Self-assembly

1.1 What is self-assembly?

Spontaneous organization of molecules into specific structures without external intervention is known as self-assembly. One of the key requirements for self-assembly process is that the constituent components should be mobile. This will enable them to migrate to form an ordered structure from an initial disordered state. Self-assembly occurs both in the macro scale as well as in the molecular level. The existence of the solar system with its numerous planets, black holes and satellites is itself believed to be a macroscale self-assembly! What will be of more interest to nanotechnologists is the molecular level self-assembly. There are two types of molecular self-assembly – **intramolecular** and **intermolecular**. The term 'self-assembly' commonly refers to the intermolecular self-assembly while the intramolecular self-assembly is referred to as

'folding' which is associated with the structures of biomolecules like proteins and DNA (deoxyribonucleic acid). The molecular level self-assembly is a typical example of the 'bottom-up' approach in fabrication of nano-dimensional structures where molecules in the sub-nano range come together to form assemblies that are in the nano-dimensions.

Another classification based on the dissipation of energy during the process categorizes the self-assembly process as **static** or **dynamic**. A static self-assembly process is one in which the ordered assembly is in equilibrium and does not dissipate energy whereas a dynamic self-assembly process is one in which the self-organized structure dissipates energy. Folding of a globular protein is a typical example of a static self-assembly process. The formation of a galaxy and solar systems are dynamic processes. In the living world, the assembly of the mitotic apparatus during cell division is a dynamic process. However, the dynamic processes are very complex and hence information on the mechanism of self-assembly involved in such systems is rather sketchy.

Why has self-assembly elicited such interest? The main reason is the mono-disperse structures that could be obtained reproducibly with least expenditure of energy by self-assembly. Control of the type of structures by regulating the self-assembling conditions is another additional point in favour of self-assembly. Also, intricate and complicated structures can be achieved using self-assembly which is not possible by other techniques. Generally, self-assembly results in formation of structures that are thermodynamically stable and are defect-free.

Biological system is one of the most investigated systems for understanding the self-assembly process. Bio-inspired engineered systems have been designed based on the mechanistic insights obtained from natural systems for novel self-assembled structures. Among the biomolecules, proteins, deoxyribonucleic acid and the lipid structures have been widely investigated. These biomolecules form the 'molecular trinity' of self-assembly. The attention received by proteins for forming self-assembled ensembles or design of bio-inspired self-assembled structures is due to the huge repertoire of protein structures that can be obtained from the same set of amino acids by changing their position in the sequence. For example, if a peptide containing 18 amino acids has to be synthesized using the twenty commonly available alpha-amino acids, then the theoretical possibility of the number of sequences that can be obtained from these amino acids works out to 20^{18}, which is a colossal number!

1.2 What are proteins?

Proteins are biologically significant molecules that maintain the functional integrity of cells. Chemically, proteins are polymers of amino acids. On the evolution scale, it is believed that proteins are a class of highly efficient and refined biomolecules that have evolved from nucleotides. Proteins have multiple functions in the biological system. The cartoon shown in Figure 1 depicts the diverse functions of proteins in humans.

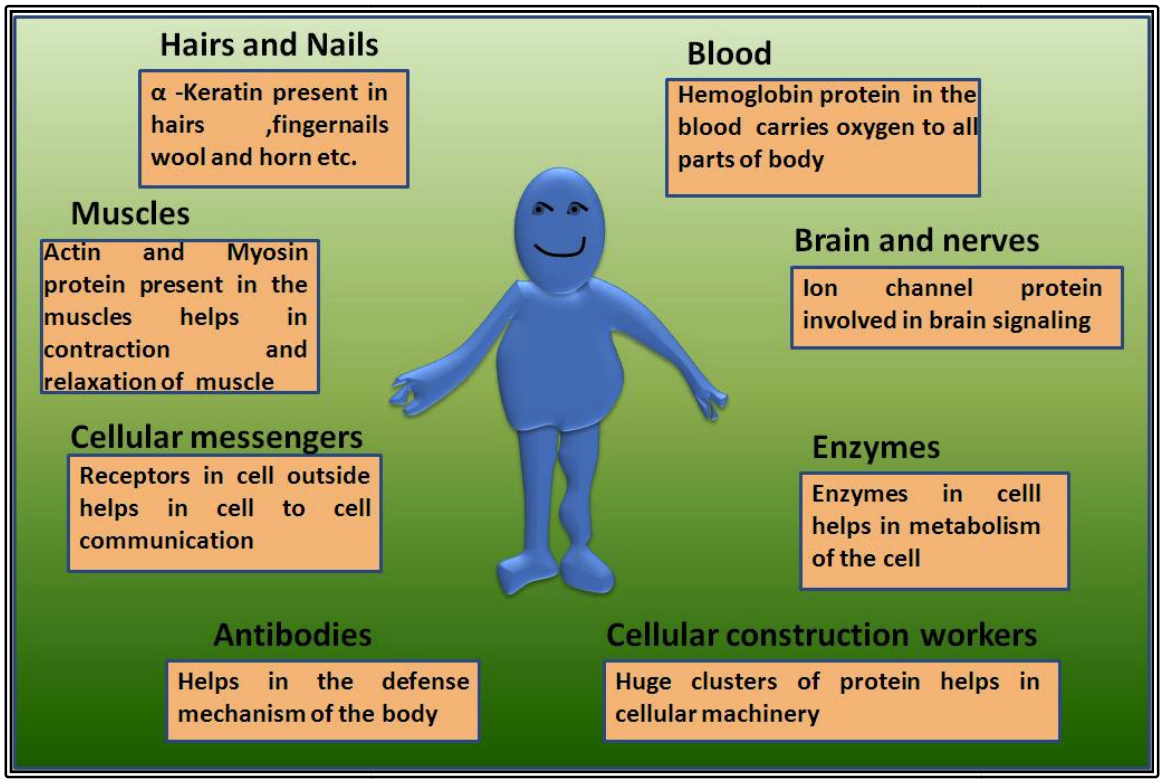

Fig. 1: Cartoon depicting diverse functions of proteins

A major class of proteins is the enzymes, which serve as the biocatalysts. What is unique about these catalysts? – Their high degree of substrate specificity, stereo-specificity, stereo-selectivity and product specificity which is very difficult to achieve with chemical catalysts! For example, the enzyme urease acts only on urea to convert it to carbon dioxide and ammonia. It does not act on any other substrate. Proteins also exhibit highly specific binding affinity, which enables them to function as membrane bound receptors or as antibodies, an important component of the immune system. For instance, streptavidin is a homotetrameric (*consists of four identical sub-units*) protein that is present in the egg white. It has a great affinity for biotin or vitamin B_7. The avidin-biotin link is one among the strongest known non-covalent bond between molecules that has been well exploited in nanobiotechnology. We will discuss more on this aspect later in Modules8 & 9 when we deliberate about drug delivery systems.

1.3 Intramolecular self-assembly in proteins

There are twenty amino acids that serve as the building blocks for the naturally occurring proteins. Biologically functional proteins are made from L-amino acids and replacement with the D-amino acids might result in disruption of its biological activity. Why? The three-dimensional structure of the protein is altered when the stereo-isomers of its constituent amino acids are replaced!

The structure of proteins comprises of four levels – primary, secondary, tertiary and quaternary. The primary structure of a protein is determined by the sequence of amino acids from which it is formed. Figure 2 depicts the representation of the primary structure of a protein.

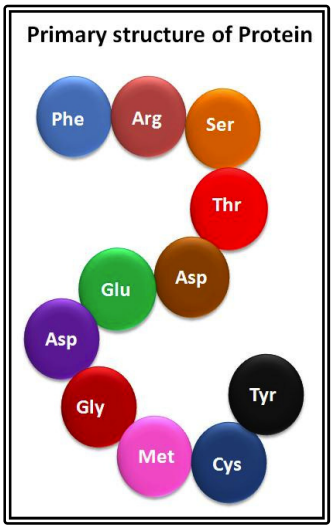

Fig. 2: Primary structure of a protein

(Phe: Phenyl alanine, Arg: Arginine, Ser: Serine, Thr: Threonine, Asp: Asparate, Glu: Glutamate, Gly: Glycine, Met: Methionine, Cys: Cysteine, Tyr: Tyrosine)

Is the primary structure very significant very significant? Yes, the sequence of amino acids is very critical in determining the structural and functional properties of the protein. Let us look at an example of a protein sequence where replacement of a single amino acid causes a huge transformation in the morphology and hence the property of the protein. Hemoglobin, which is responsible for the oxygen carrying function of the blood, contains a protein apart from the heme prosthetic group. The replacement of a single amino acid glutamic acid, a hydrophilic amino acid, by the hydrophobic amino acid valine at the 6^{th} position in the beta globin chain of hemoglobin, results in a drastic change in the morphology of the red blood cell from a puff-shaped biconcave spherical structure to an elongated sickle shape. This is because the presence of a hydrophobic amino acid in the place of a hydrophilic amino acid causes aggregation and loss in elasticity of the red blood cell. As a result, the red blood cell loses its ability to squeeze through capillaries, suffers a reduction in the oxygen carrying ability and also has a reduced life-time. This condition is known as ‘sickle cell anemia’ and is an inherited disorder. Thus the sequence of amino acids in the protein plays an important role in determining the structure and function of a protein. The following animation (Figure 3) describes the phenomenon of sickle cell anemia due to a single point mutation in the primary structure of the globin protein.

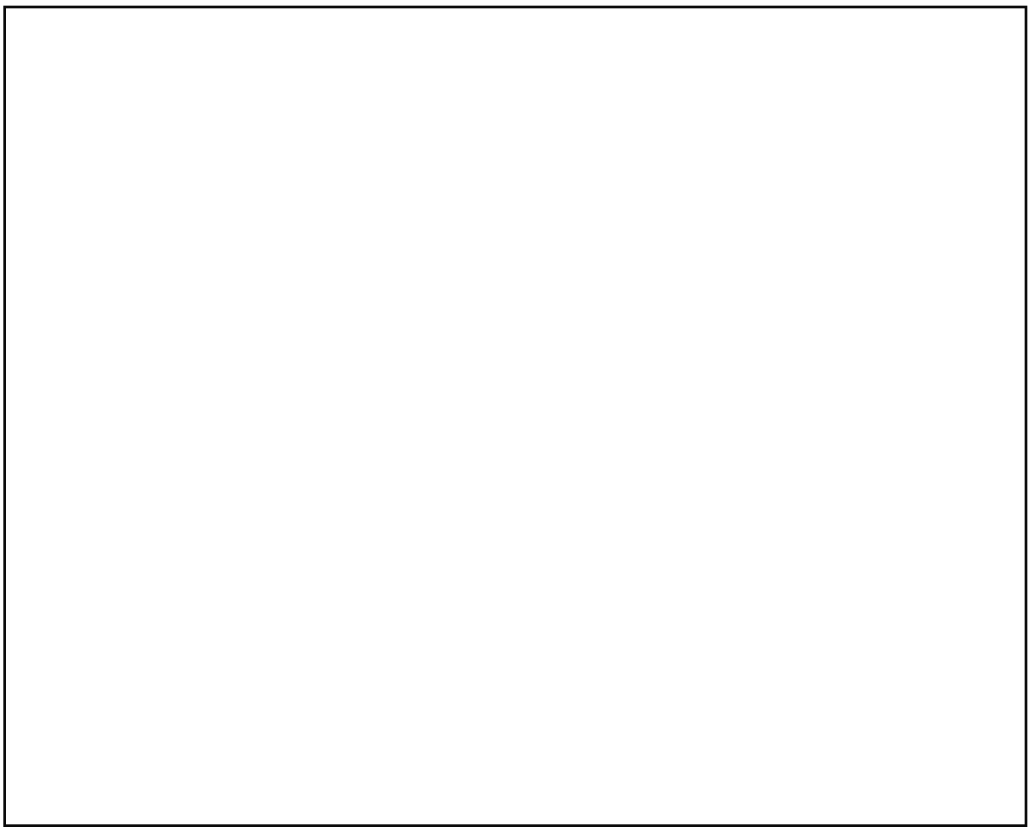

Fig.3: Animation on Sickle cell anemia

Note: Can be viewed only in Acrobat Reader 9.0 and above

The next level of protein structure is the secondary structure, which involves small, localized domains of the proteins that interact to form alpha helices or beta sheets or random coils. Figure 4 represents two commonly encountered secondary structures in proteins – the alpha helix and beta sheets.

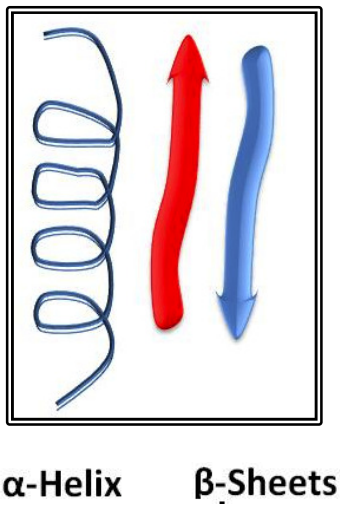

Fig. 4: Secondary structures of proteins

The secondary structure in a domain is dependent upon the nature of side groups of the amino acids present in that region. Chou-Fasman rule gives an empirical set of conditions that can help in predicting the propensity of a particular sequence of amino acids to form a specific secondary structure.

Chou-Fasman rules

- *Chou-Fasman predictions classify amino acid residues as alpha helix formers or breakers or as indifferent.*
- *Glutamate, alanine and leucine are classified as strong alpha helix formers*
- *Histidine, methionine, valine, glutamine, tryptophan and phenyl alanine are also alpha helix formers but weaker than the first category*
- *Lysine and isoleucine are weak alpha helix formers*
- *Aspartate, threonine, serine, arginine and cysteine are classed as indifferent*
- *Asparagine and tyrosine are alpha helix breakers*
- *Proline and glycine are strong alpha helix breakers*

Does this secondary structure contribute to the structure and function of a protein? Well, the answer is YES! In the case of Alzheimer's disease, which is characterized by formation of amyloid plaque deposits in the brain, resulting in progressive memory loss and dementia, a sequence of events leads to the release of a fragment (1-42) of the amyloid precursor protein, which transforms from an alpha helical structure to a beta sheet resulting in the insoluble deposits causing neuronal death and the debilitating consequences in the brain function of an individual. Such types of disorders are termed as 'prion' diseases.

In nature, there are quite a formidable range of peptides that predominantly possess a specific type of secondary structure. For example, the **human beta defensin-2** is an antimicrobial peptide that possesses a beta pleated sheet structure while the cationic peptide **magainin** exclusively possesses an alpha-helical structure. **Bactenecin** and **indolicidin** from bovine sources, both possessing anti-microbial activities completely lack either alpha helix or beta pleated sheet structures. Instead, bactenecin possesses a loop structure while indolicidin functions as an extended chain. Interestingly, all four peptides exhibit anti-microbial activity. Does this structure show any correlation with their function? YES! The beta sheet structures tend to disrupt the lipid bilayer membranes of the bacterial membrane forming pores or channels while the alpha helical structures tend to incorporate into the lipid bilayer thereby altering the membrane permeability. The cationic charge of the peptide also is crucial to their function. The bacterial membrane

generally contains anionic groups and hence cationic peptides are able to interact more effectively with them.

Now, we move on to the next higher level of protein structure, namely the tertiary structure. Here interactions between the various domains lead to formation of a supramolecular structure that is stabilized by both primary and secondary forces. The disruption of these attractive forces alters the protein conformation, structure and its function. One example is the influence of 2,3, bisphosphoglycerate (2,3-BPG) on hemoglobin. The charged BPG provides additional electrostatic forces that convert the hemoglobin from a relaxed form to a taut form thereby facilitating release of oxygen to the tissues. This effect is known as 'Bohr effect'. Figure 5 represents the tertiary structure of a protein.

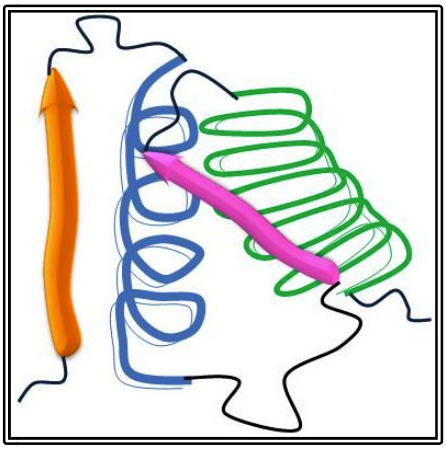

Fig. 5: Tertiary structure of a protein

The highest level of protein structure is the quaternary structure. This form is distinct from the other levels of structure in that it is found only in proteins that contain sub-units and involves intermolecular self-assembly. The different polypeptide chains forming the sub-units of a multimeric protein associate via secondary forces of attraction. These forces when altered can dramatically influence the protein function. Figure 6 represents the quaternary structure of a protein.

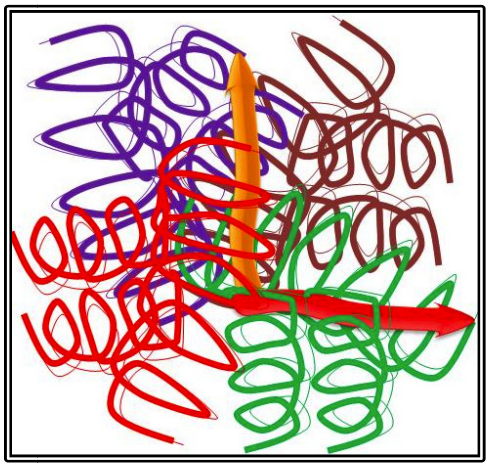

Fig. 6: Quaternary structure of a protein

The secondary, tertiary and quaternary structures of a protein can be 'denatured' by disrupting the attractive forces between them while the primary structure cannot be denatured but the peptide bond can only be hydrolysed.

2 Reference

"Biochemistry", Donald Voet and Judith Voet, John Wiley & Sons, 2011