

# SUPER SECONDARY STRUCTURES, MOTIFS AND DOMAIN

## OBJECTIVES

- To learn about super secondary structures in protein structures
- To learn about kinds of domain structures in proteins

Protein structure can be divided into array of structural elements. These include super secondary structures, motifs, domains etc.

## INTRODUCTION

### Super Secondary Structures

Super secondary structures are formed when secondary structure elements arrange or group themselves in specific geometric arrangements. They are formed by a connective link between secondary structures, but they are smaller than a domain or a subunit. Same kind of super secondary structure can exist in proteins of different and unrelated family proteins. These could act as a site of nucleation in the process of protein folding. Super secondary structures are building blocks of protein structure and referred as “structural motif” as well. Examples include Beta hairpin, Helix-turn-helix, Greek key motif etc.

Looking at few of common super secondary structures in detail:

- **Beta hairpin:**

It is formed when two anti parallel beta strands are connected by a relatively tighter loop or a turn. The structure looks similar to that of a hairpin and hence the name. These can form independently and are stable in isolation. The two beta strands are stabilized by hydrogen bonds as shown in the slides. Shown by use of NMR studies, beta hairpins can form in isolation in solution, thus they could act as nucleation sites for protein folding. These are classified into four distinct classes according to the smallest possible number of loop residues and progressively increase the loop size by removing hydrogen bonds in the beta sheet. The primary hairpin of class 1 is a one-residue loop where the bound residues share two hydrogen bonds. One hydrogen bond is then removed to create a three-residue loop, which is the secondary hairpin of class 1. Similarly all the hairpins are defined in this way. In the Class 2, the hairpin begins with two residues. All the classes follow the same pattern.

- **Helix-turn-helix:**

When two alpha helices are connected by a turn or a loop, yields a helix-turn-helix motif. Generally the motif consists of 20 amino acids. The helices are stabilized by non-polar interactions between them. It is the most common motif found in DNA binding proteins, since it is directly involved in contacting DNA. As depicted in the figure, the second helix enters the major groove of the target DNA site and helps in its recognition and binding. In several attempts to classify them, their structure and spatial arrangement of helices have been used. The various classes are:

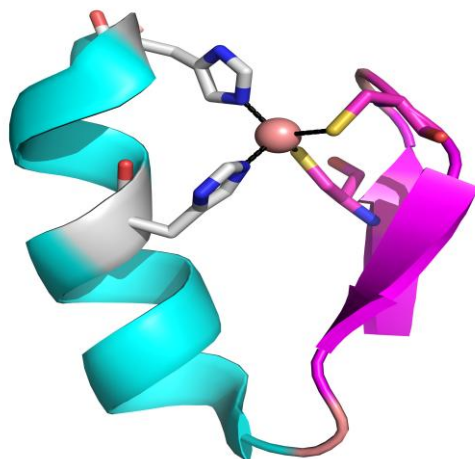
- Di-helical
- Tri-helical
- Tetra-helical
- Winged helix-turn-helix

Example: Cro repressor – The crystal structure of lambda Cro repressor shows a Helix-turn-helix DNA binding motif.

In the figure shown Phage Cro and OR1 complex. PDB ID:3CRO (Mondragon, A et al., 1991)

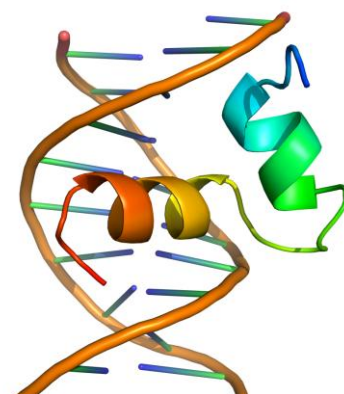
- **Zinc Finger motif:**

As suggested by its name, the motif contains one zinc atom that stabilizes the fold by being centrally placed. Each motif contains 30-50 amino acids; forms one helix and two small  $\beta$ -sheets together make up a single zinc finger motif. A regular string of cysteine/histidine residues coordinate with the zinc atom. Each finger comprises one helix and two small anti-parallel beta sheets. Presence of zinc atom aids in holding the helix and beta sheets together by coordinating cysteine/histidine residues. Generally two coordinating residues come from helix and two residues come from  $\beta$ -sheets. The helix region is involved in binding to major groove of DNA.



The zinc finger proteins can be classified into "fold groups" based on the overall shape of the protein backbone in the folded domain. The most common "fold groups" of zinc fingers are the Cys<sub>2</sub>His<sub>2</sub>-like (the "classic zinc finger"), treble clef, and zinc ribbon.

Originally identified in the transcription factor IIIA, the motif binds DNA through the helix. These fingers have proved out to be ubiquitous in the biological world. They have opened a new window to the area of therapeutics. Zinc finger nucleases and zinc finger transcription factors are two of the



most important applications of this to be realized to date. By engineering these one can direct the specific DNA sequences in the genome in order to rectify congenital diseases at DNA level.

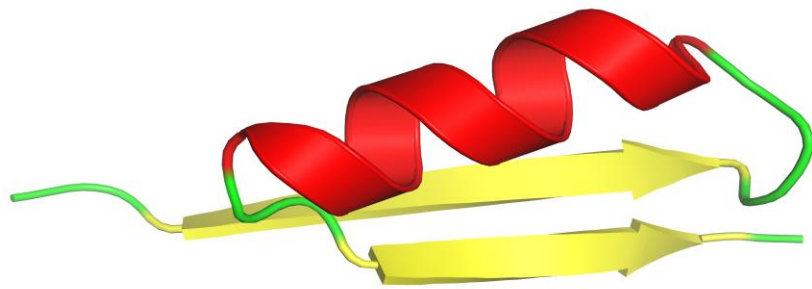
Example: Structure of zinc finger from *Xenopus laevis* transcription factor IIIA as submitted in the Protein Data Bank under its ID 1ZNF (Lee et al., 1989). It shows the coordination of Cys3, Cys9, His19, His23, and one zinc atom.

- **$\beta\alpha\beta$  motif:**

In contrast to beta hairpin,  $\beta\alpha\beta$  motif contains two parallel beta strands connected by longer stretch of amino acids that form an alpha helix. Loop regions connecting sheet and helix can vary in lengths. The helix axis is generally in parallel orientation to the beta sheets. The hydrophobic interactions stabilize the motif and hold the structure. This is a common motif found in most of the proteins that contain parallel beta sheets in their structure. Generally the conformation of the motif is found to be right handed, i.e. plane of the helix is above that of the beta sheets. It is unlikely to find the helix below the plane of beta sheets.

Example:

Ribonuclease inhibitor, which has several  $\beta\alpha\beta$  motifs connected together to form a

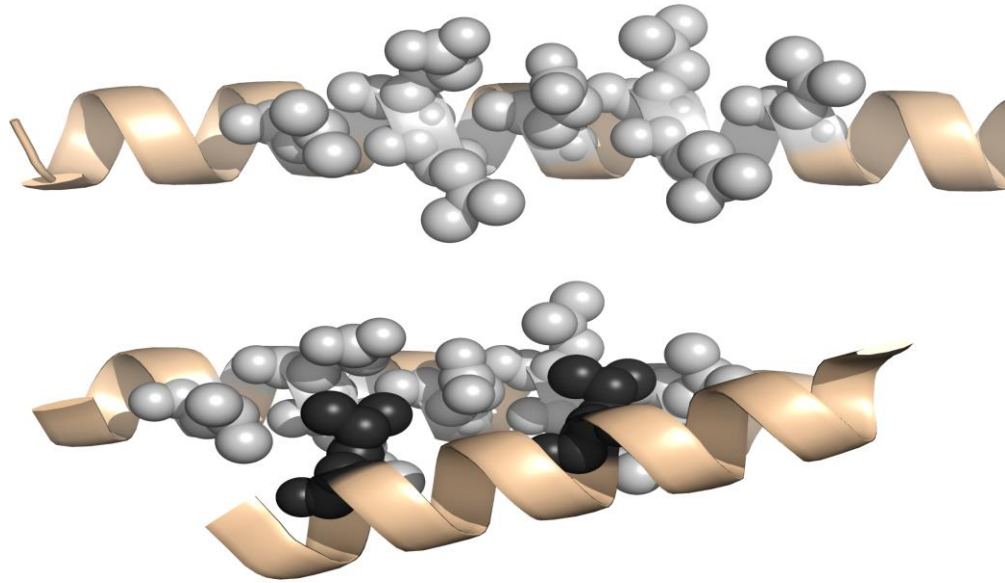


concave shaped structure. (Kobe, B. & Dicesenhofer, J. 1993 ). In the above image has one  $\beta\alpha\beta$  motif is shown.

- **$\alpha\alpha$  motif / Coiled coils:**

Two alpha helices are packed together by stably winding around each other. The two helices are stabilized by hydrophobic interactions and are also found to be intact in solutions as well. The motif is generally found in transcription factors and intermediate filaments. A heptad (7 amino acid) repeats sequence is present in this motif. About 7 amino acids are present per two turns in the helix. The arrangement is such a way that for every turn, a hydrophobic residue would come on the same side of the helix. This helps in the hydrophobic interactions when two helices wind around. The seven amino acids are represented as a,b,c,d,e,f,g. Where “a” and “d” are generally small hydrophobic residues, preferably Leu and Met. The residue “e” and “g” are generally charged residues forming salt bridges between helices, hiding the hydrophobic interactions from water. This motif can also be referred as leucine zipper, due to presence of leucine as the 7<sup>th</sup> residue in each heptad. On one helix there is gap

between four residues and hydrophobic residue, generally Leu from another helix enters this gap like knob in the hole as seen in the figure.

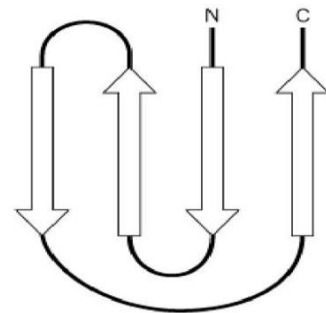


Example: GCN4 transcription factor displays this motif as part of its structural set up that aids in its functioning to bind DNA. PDB id:2ZTA (O'Shea, EK et al., 1996).

- **Greek key motif:**

Another motif made up of beta sheets in an artsy arrangement and hence, deriving its name from a common pattern of the Greek ornamental work.

This is one of the most common motifs found in  $\beta$ -sheet rich proteins. Four anti-parallel beta sheets are present in which three strands are connected by loops and the first and fourth strand is connected via longer loop (refer to figure in the slides).



Example: Gamma-crystallin structure deposited under the ID 1ELP, shows this motif in its structure.

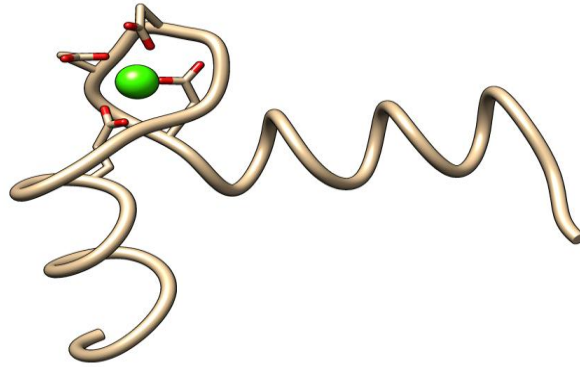
- **EF hand motif:**

EF hand motif was first time observed in Parvalbumin, which contains three of such motifs. It was named so because in the protein, a 12 amino acid loop connected the 5<sup>th</sup> and the 6<sup>th</sup> helix, which binds calcium (E is the 5<sup>th</sup> letter and F is the 6<sup>th</sup>), thus named as EF hand motif.

So basically the EF hand motif is two alpha helices connected via 12 amino acid loop that usually binds calcium. "DXD/NXDGXIXXXE" is the sequence that is generally found in the loop region, where "X" is any amino acid. Carboxyl groups of aspartic acid, glutamic acid and main

chain carbonyl groups coordinate with calcium to form pentagonal bipyramid coordination. This motif is inevitably part of almost all calcium binding and signaling proteins. The most common example widely studied includes Calmodulin and Troponin C.

In the native, i.e. unbound form, the helices are close and almost parallel to each other. However, when it is in calcium bound state, the helices attain a perpendicular orientation. In the figure it is shown the calcium coordinated EF hand motif (PDB ID: 1CLL, Chattopadhyaya et al., 1992). Two EF hand motifs together become one functional unit.



The helices have hydrophobic residues, which are hidden in the native state become exposed when there is a conformation change due to binding of calcium by the EF hand motifs. In this open state, the protein can then interact with their targets and alter their function. In this manner they sense the calcium signal and alter the function of other proteins.

- **Omega loop:**

Another motif, whose name has been derived from the Greek alphabet, gets its name due to its similar shape. It is a compact globular structure with no regular repeat pattern of hydrogen bonds, where beginning and end of the loops come together. The side chains of the loop come in the center of the loop and stabilize. Omega loops of two different proteins interact with each other to form dimers, with their side chains packing in the center of the loop. These can be found in the phosphate-binding loop of nucleotide binding proteins.

Example: Interleukin-1b

## **Domains**

Varied combinations of secondary structures and motifs, connected by loops and turns, form a functional domain of a protein. It is a conserved part of a given protein sequence and structure. Wetlaufer first proposed domains in 1973 after X-ray crystallographic studies of hen lysozyme, papain and limited proteolysis of immunoglobulins.

A domain is a compact three-dimensional structure, which often folds independently in a stable form such that it has its own hydrophobic core surrounded by hydrophilic residues. Domains form the fundamental unit of tertiary structure. It can evolve, function and exist independently. Size of a protein domain can vary in length from 30 to 500 amino acids. Though about

90% of domains consist of less than 200 residues with an average of 100 amino acids with a single hydrophobic core.

The structural domains aids in protein folding, as they are able to fold themselves independently; thus speeding up the process of protein folding. This also decreases a large number of residue interactions that can occur during the course of formation of mature protein.

During molecular evolution, domains are used as building blocks. These are recombined and assorted to form proteins with varied functions. A great variety of multidomain proteins can also be found which have arisen due to gene duplications. Independent proteins in prokaryotes can be seen as combination of domains in eukaryotic multidomain proteins.

Large proteins can have multiple functional domains. Small changes in the linker regions can lead to domain swapping.

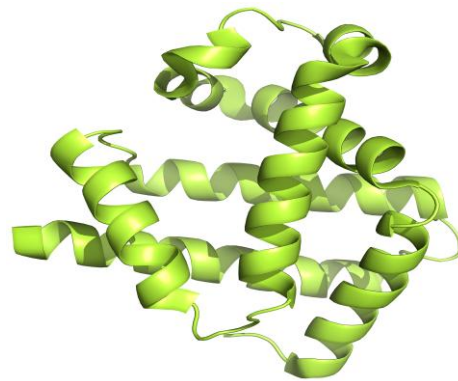
Domain swapping is a mechanism to form oligomeric structures. A secondary or tertiary element of a monomeric protein is replaced by the same element of another protein. An example to quote would be oligomeric enzymes with active sites at subunit interfaces.

Domain folds are classified into approximately 800 fold families. The simplest classification is on the basis of secondary structural content as discussed below.

Several types of domains on the basis of secondary structures are:

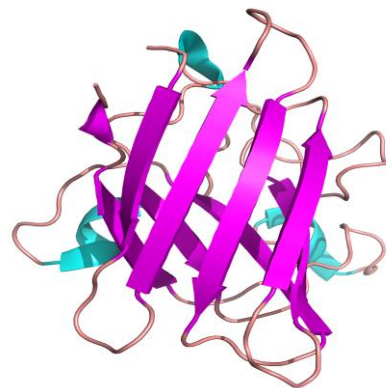
- All  $\alpha$  domains:

All  $\alpha$ -domains have only  $\alpha$  helices. They could be coiled-coils as discussed above, four helix bundles, which have four helices held together by hydrophobic core. For details see module 5. The  $\alpha$ -domains can have large complex structures. The simple example  $\alpha$ -domain structure is – Myoglobin (PDB id: 3RGK, *Hubbard, SR et al., 1990*), which is shown in the figure.

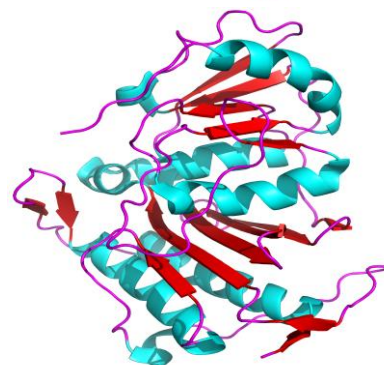


- All  $\beta$  domains:

All  $\beta$  domains have only  $\beta$  sheets. They could be divided into many groups. Some of them are discussed in module 5. Many of these structures have Greek key motifs and some of them have  $\beta$ -helices. The simplest example of all  $\beta$  structure is superoxide dismutase (SOD), which is shown in the figure (PDB id: 2V0A (*Strange, RW et al., 2007*)). The front 4 sheets are arranged in Greek key motif pattern.



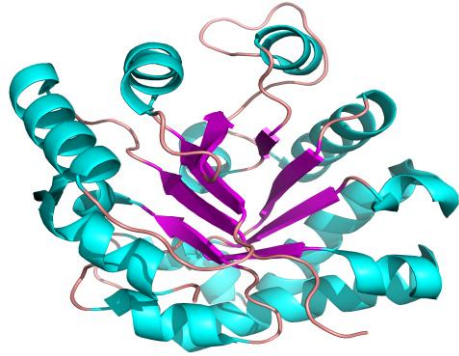
- $\alpha + \beta$  domains:





This domain contains mixture of  $\alpha$  and  $\beta$  secondary structures. Here generally the  $\beta$ -sheets are arranged in different mode or in combinations and helices are sheets are partially segregated. For example N-terminal domain of PAP phosphatase (3'- phosphoadenosine 5' - phosphate Phosphatase) shown in the figure. PDB id: 4QXD (Tarique, KF et al., 2014).

- $\alpha/\beta$  domains:  
These domains contains combination of  $\beta$ - $\alpha$ - $\beta$  motifs, predominantly form a parallel  $\beta$  sheets surrounded by  $\alpha$  helices. For example Triosephosphate isomerase, shown in the figure. (PDB id: 1TPF (Kishan, KV et al., 1994).



**Summary:** Proteins are made up of either single domain or multiple domains. Several motifs pack together to form compact, local, semi-independent units called domains. They help in protein folding by forming organized regions independently and they can act as nucleation in protein folding thus accelerating the process. Protein domain dynamics plays key roles in a multitude of molecular recognition and signaling processes. Domains are important area for molecular evolution. Due to gene duplication many multidomain proteins can be found especially in eukaryotes. Domains can be classified into more than 800 groups but by simple classification can be done into 4 classes. Each domain is combination of motifs / super secondary structures. When looking at evolution by functional adaptation, domain swapping gives an insight into this through oligomerisation. Like domains, many super secondary structures can exists independently in solution.