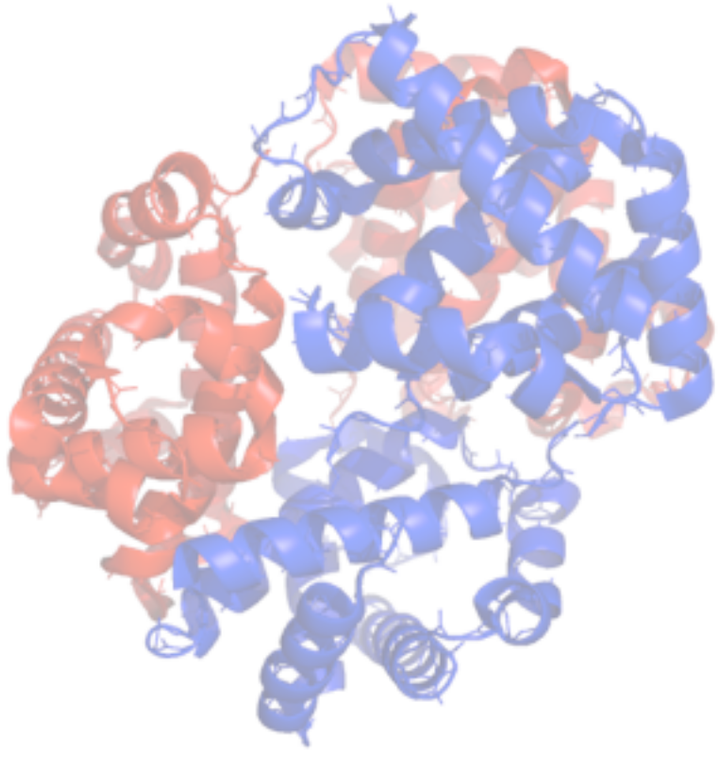
Structural Genomics

Introduction to protein structure



Davide Baù

http://sgu.bioinfo.cipf.es

Structural Genomics Laboratory

Bioinformatics & Genomics Department

Prince Felipe Research Center (CIPF), Valencia, Spain

Outline

- Proteins and protein structure
- Secondary structure elements
- Protein structure-function
- Protein structure-sequence conservation

Proteins

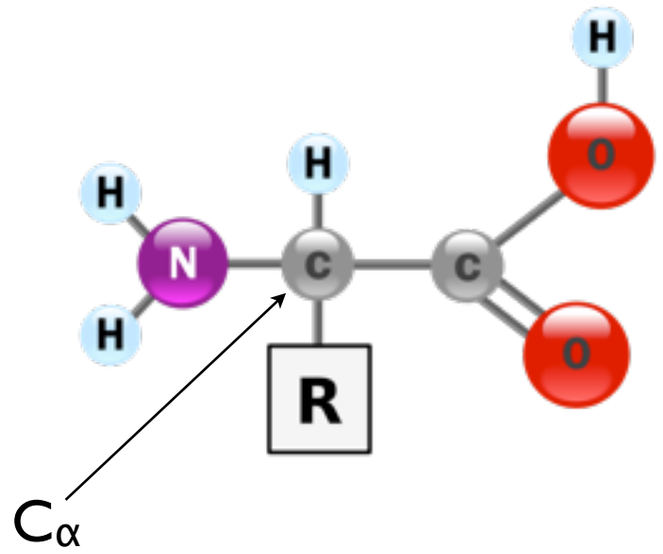
Proteins are biological **macromolecules** and are fundamental constituents of the cell.

Chemically, they are polymers (linear chains) of **amino acids residues** connected via a peptide bond.

Proteins can be found associated with other proteins (complexes), metallic ions and other molecules.

Amino Acids

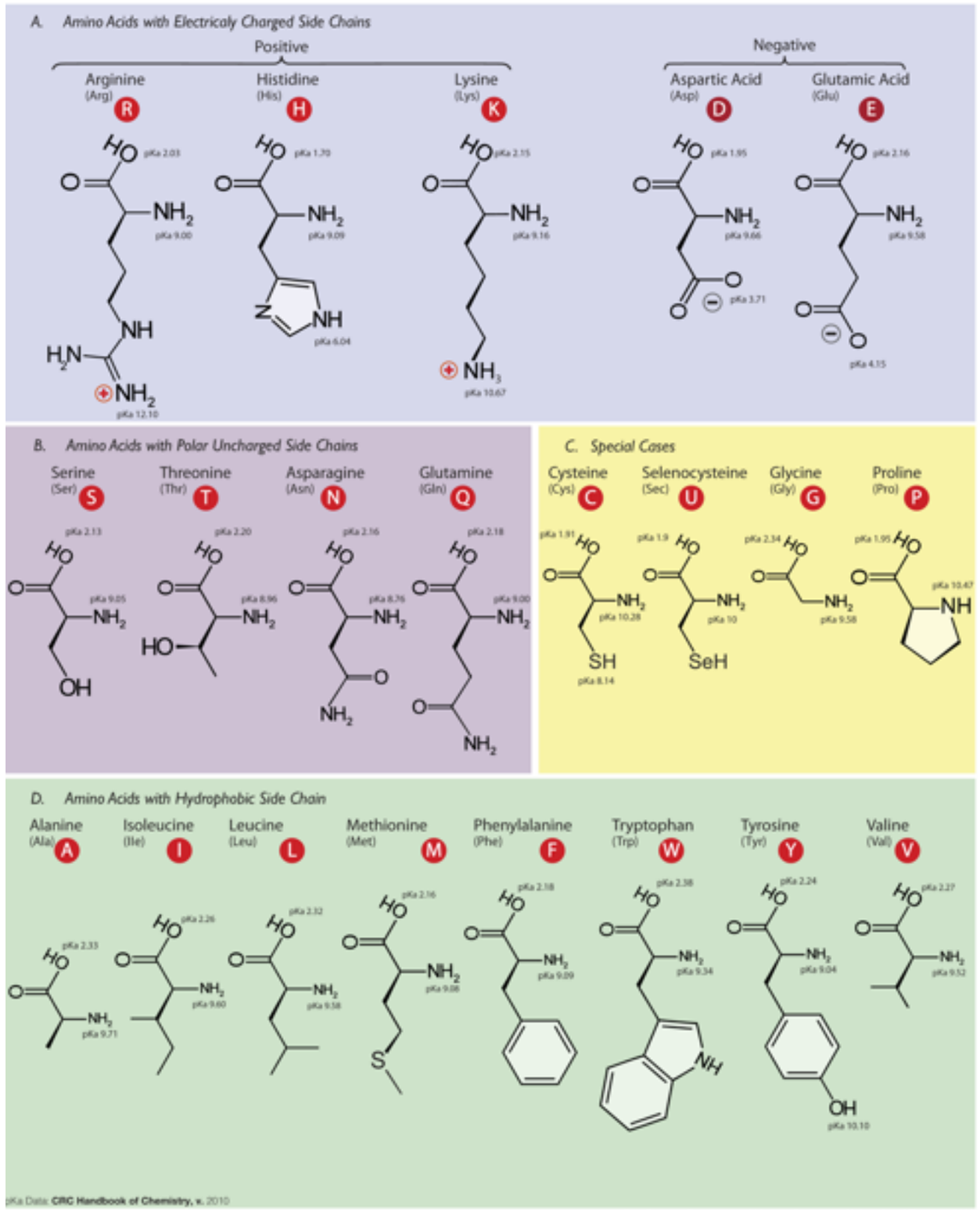
Amino acids are composed by an amine group, a carboxylic acid group and a side-chain that varies between different amino acids:



The carbon atom bound to the side chain (**R**) is called **C_α**.

Twenty standard amino acids are naturally incorporated into proteins and are encoded by the universal genetic code.

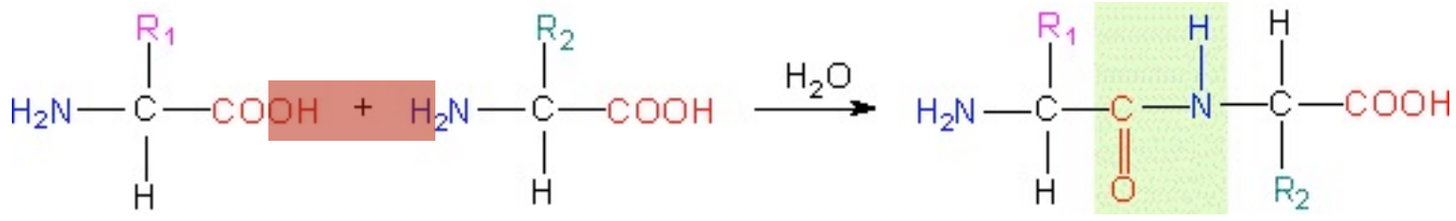
Amino Acids


A. Amino Acids with Electricaly Charged Side Chains
Positive
Arginine (Arg) R
Histidine (His) H
Lysine (Lys) K
Negative
Aspartic Acid (Asp) D
Glutamic Acid (Glu) E
B. Amino Acids with Polar Uncharged Side Chains
Serine (Ser) S
Threonine (Thr) T
Asparagine (Asn) N
Glutamine (Gln) Q
C. Special Cases
Cysteine (Cys) C
Selenocysteine (Sec) U
Glycine (Gly) G
Proline (Pro) P
D. Amino Acids with Hydrophobic Side Chain
Alanine (Ala) A
Isoleucine (Ile) I
Leucine (Leu) L
Methionine (Met) M
Phenylalanine (Phe) F
Tryptophan (Trp) W
Tyrosine (Tyr) Y
Valine (Val) V
pKa Data: CRC Handbook of Chemistry, v. 2010
Dan Cojocari, Department of Medical Biophysics, University of Toronto 2009

The Peptide Bond

Formation

A peptide bond is a **covalent bond** formed between two molecules when the carboxyl group of one molecule reacts with the amino group of the other molecule, causing the release of a molecule of water (H_2O).

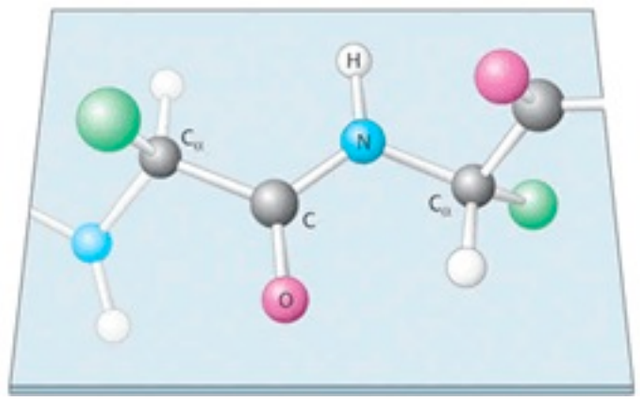


Polypeptides and proteins are chains of amino acids held together by peptide bonds.

The Peptide Bond

Properties

The peptide bond is planar:



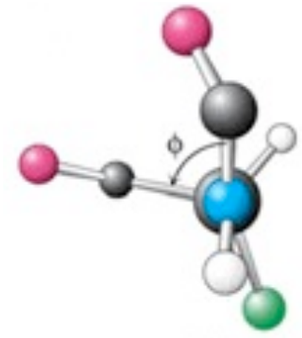
Only 2 bonds can freely rotate, Cα–N bond and CC(O) bond.

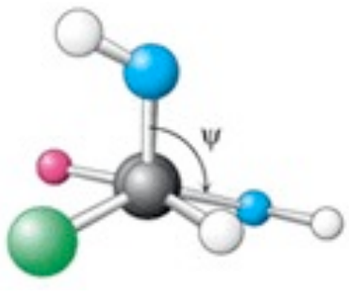
Image credits: http://www.imb-jena.de/~rake

The Peptide Bond

Properties

Limited amount of allowed rotation defined by the **Φ** and **Ψ** torsion angles, which are constrained by the structure of adjacent amino acid residues.





Φ

Ψ

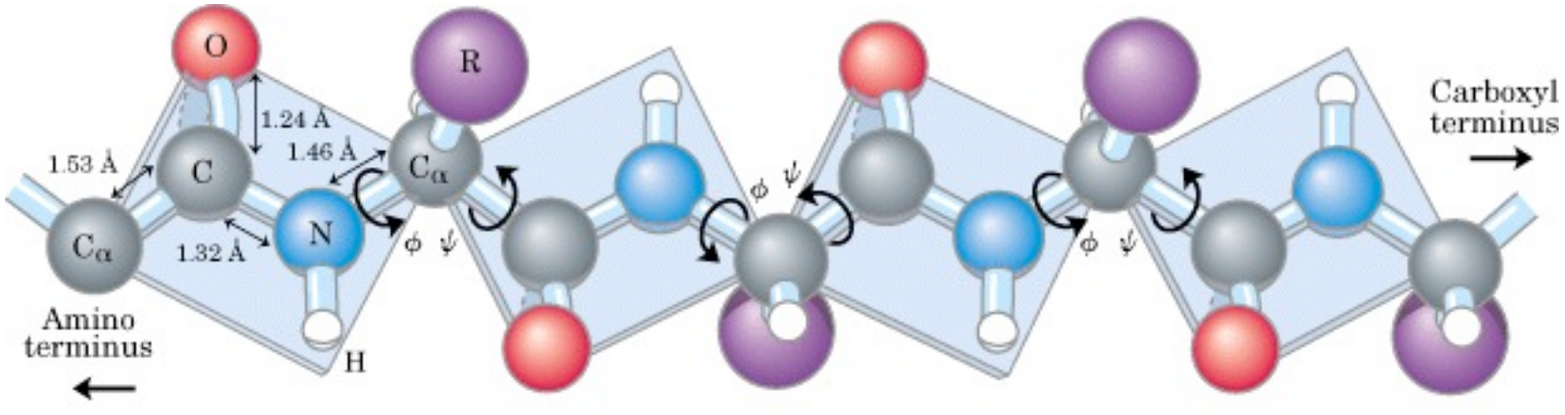
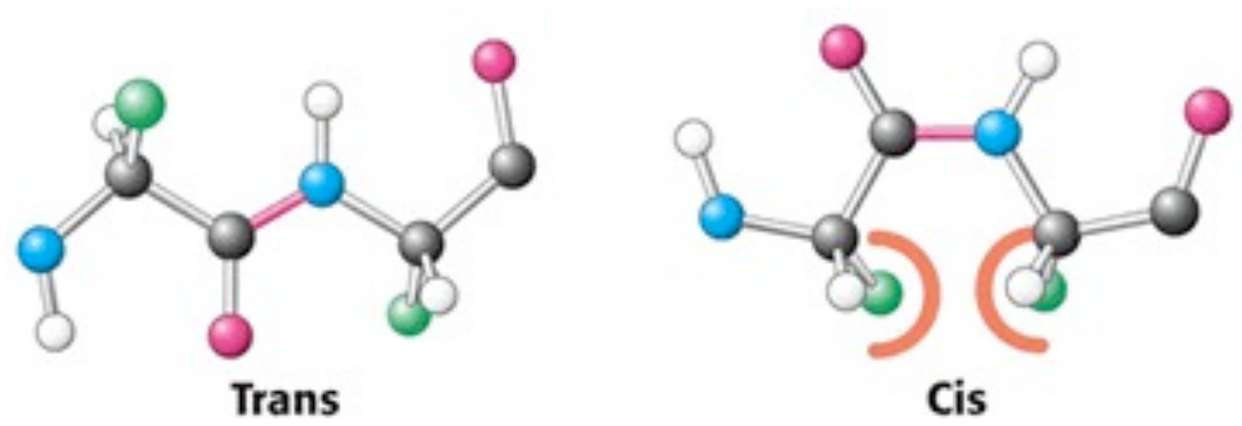


Image credits: http://www.imb-jena.de/~rake

The Peptide Bond

Properties

The carbonyl oxygen and and the amide hydrogen are in a **trans** configuration (**energetically more favorable**), because of the steric hindrance (steric clashes) between the functional groups attached to the C_α atom.

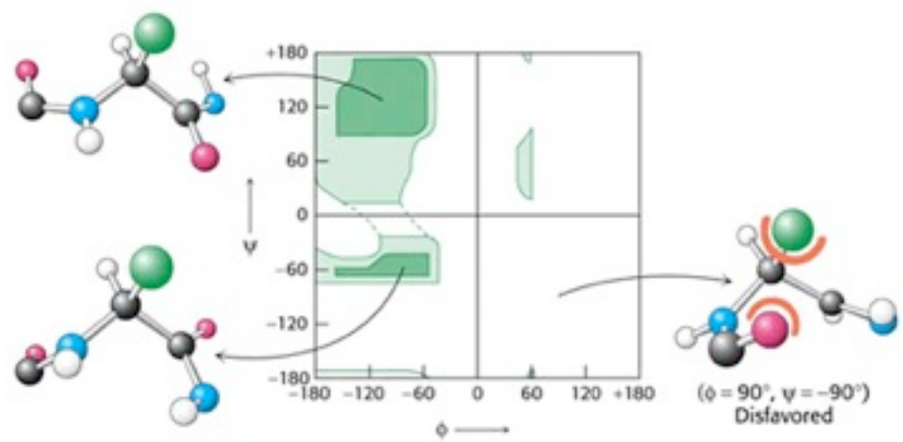


As a consequence, almost all peptide bonds in proteins are in trans configuration.

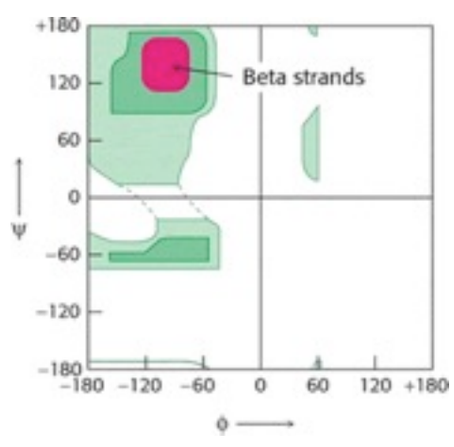
Image credits: http://www.imb-jena.de/~rake

Ramachandran Plots

Protein structures Φ and Ψ angles fall within allowed regions (displayed in green and red).



Secondary structure elements are defined by specific pairs of Φ and Ψ angles:



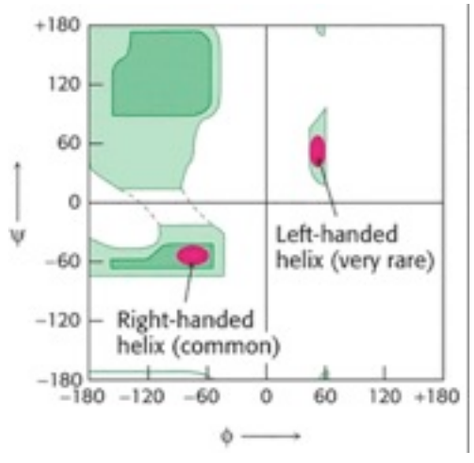


Image credits: http://www.imb-jena.de/~rake

Summary

Proteins are biological **macromolecules** and are fundamental constituents of the cell.

Polypeptides and proteins are **chains of amino acids** held together by peptide bonds.

Protein configurations are defined by specific (and limited) pairs of **Φ** and **Ψ angles**.

Protein Structure

Protein structural levels

- Primary structure
- Secondary structure
- Tertiary structure
- Quaternary structure

Primary Structure

In biochemistry, the primary structure of a molecule is the exact description of its atomic composition and bounds.

The primary structure of a protein is the ordered sequence of its constituents building block (amino acids).

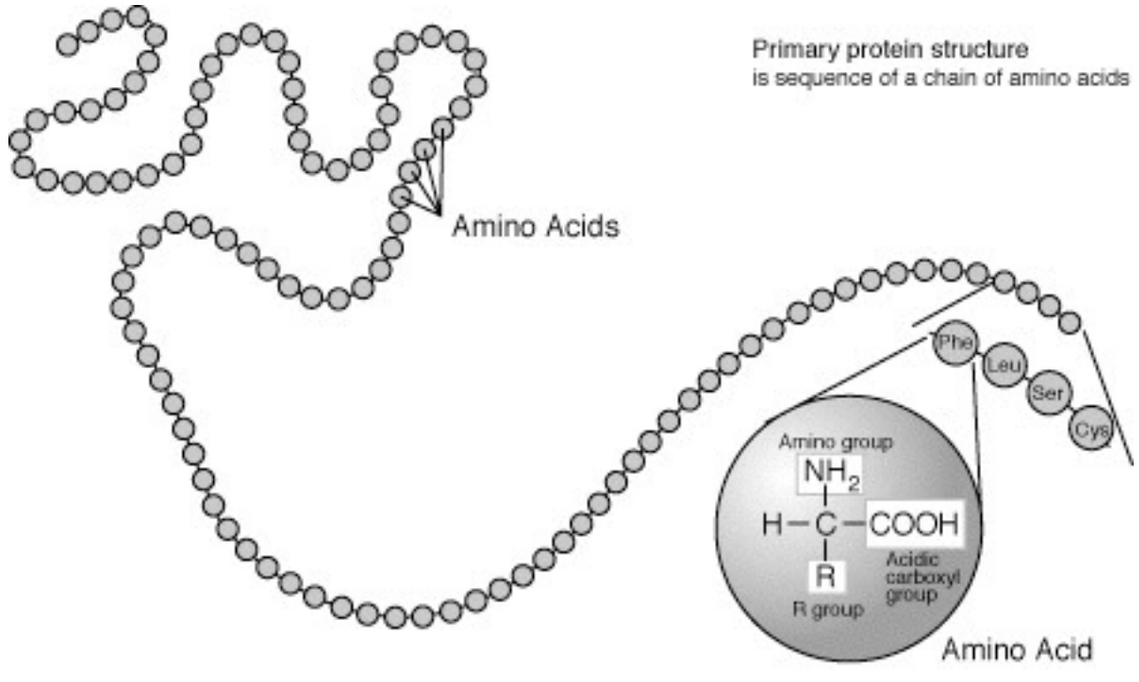


Image credits: Wikipedia

Secondary Structure

The secondary structure of a protein is the ability of a protein of assuming a regular and repetitive spatial arrangement.

There are three types of secondary structure: **helices**, **β-sheets** and **turns**.

The secondary structure is formally stabilized by the hydrogen bonds.

Secondary Structure

α-helix and 3_{10}-helix

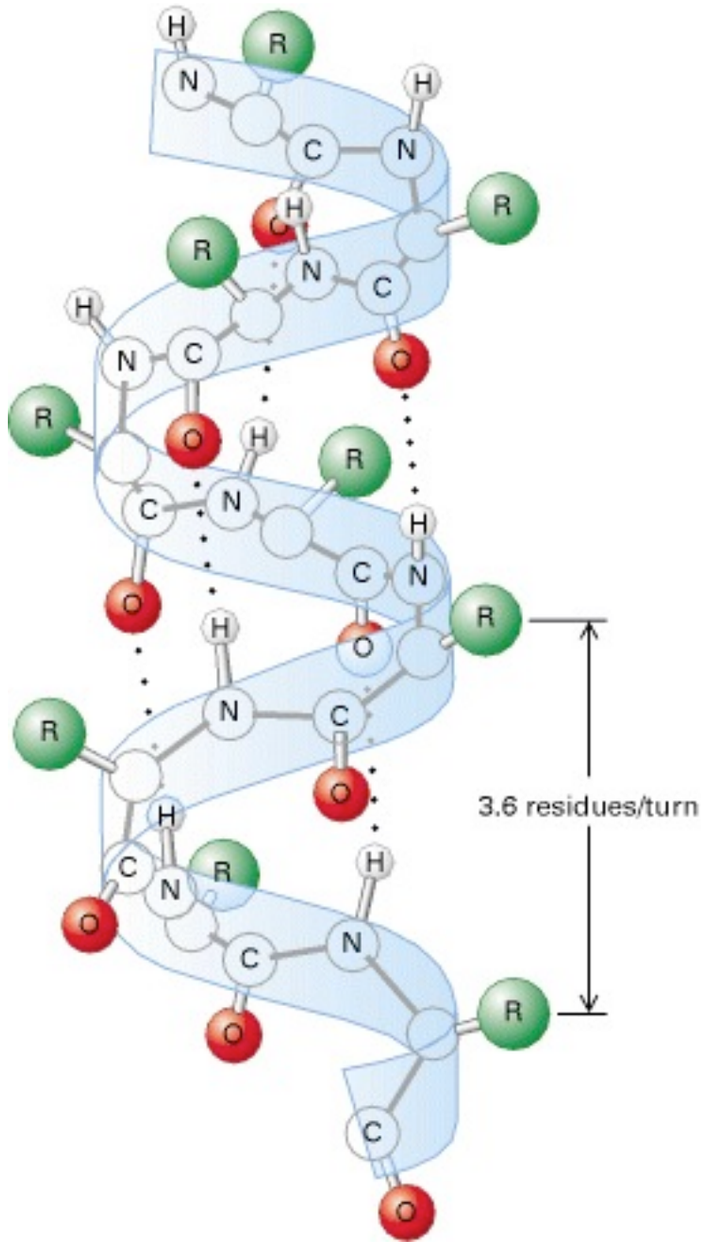
α-helices form when consecutive residues adopt specific values of the (Φ, Ψ) angles.

The structure is stabilized by hydrogen bonds between the C=O of residue *i* and the N-H of residue (*i+4*).

The side chains (-**R**) point outwards minimizing steric interference.

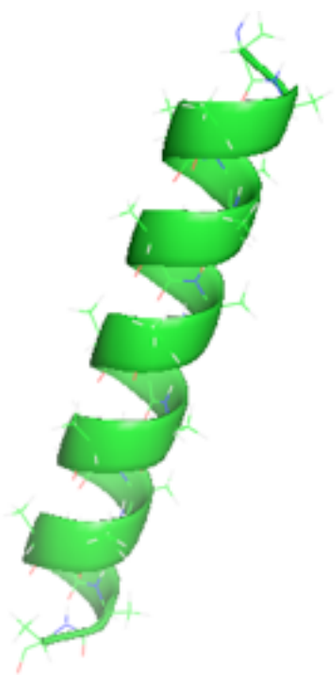
α-helix: 3.6 residues/turn, 12 backbone atoms/turn and a distance of 5.4 Å.

3_{10} helix: 3 residues/turn, 10 backbone atoms/turn and a distance of 6 Å. H-bonds between residue *i* and (*i+3*).

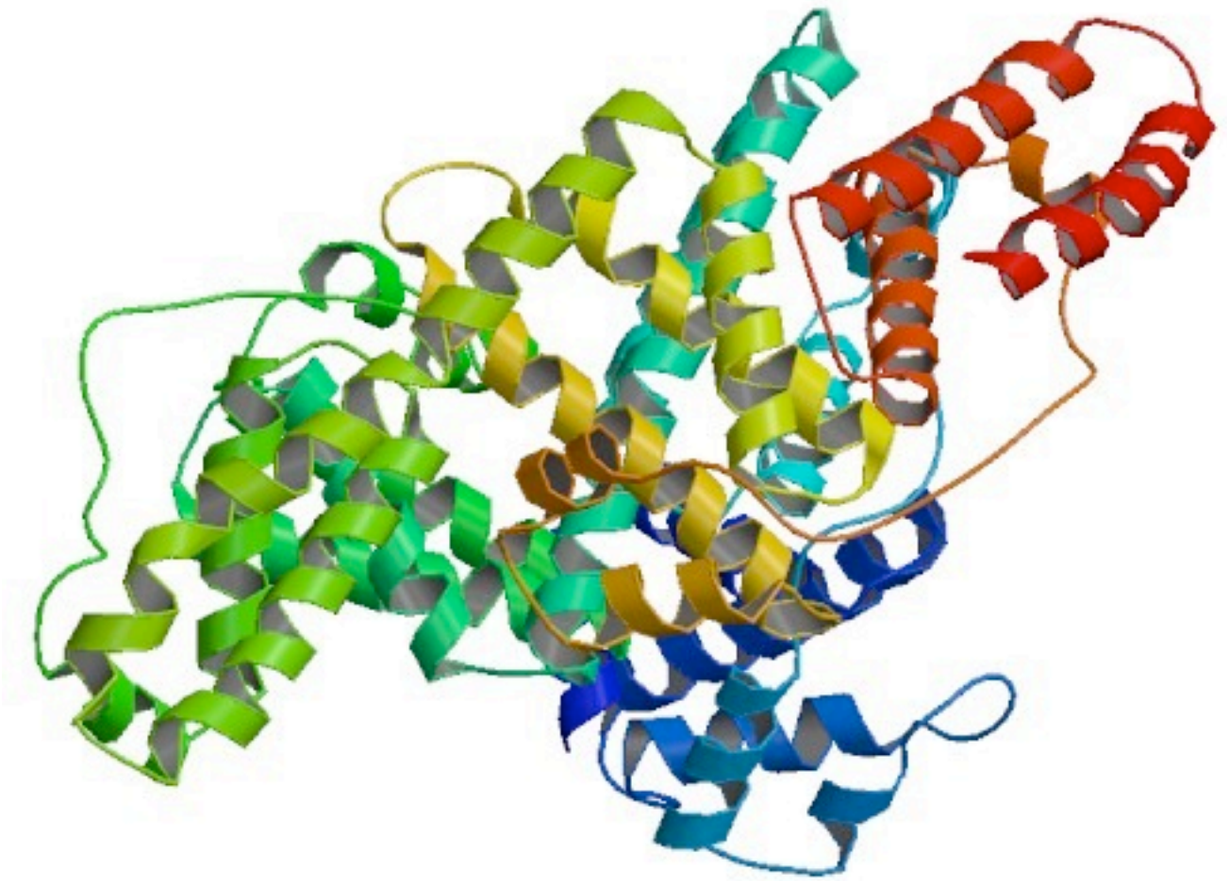


α-helix Example

Human serum albumin (PDB: 1ao6)



Ideal α-helix



Real α-helices

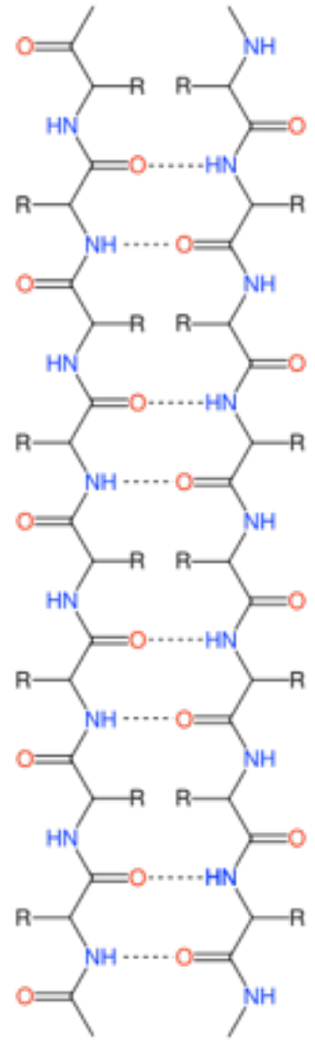
Secondary Structure

β-sheets

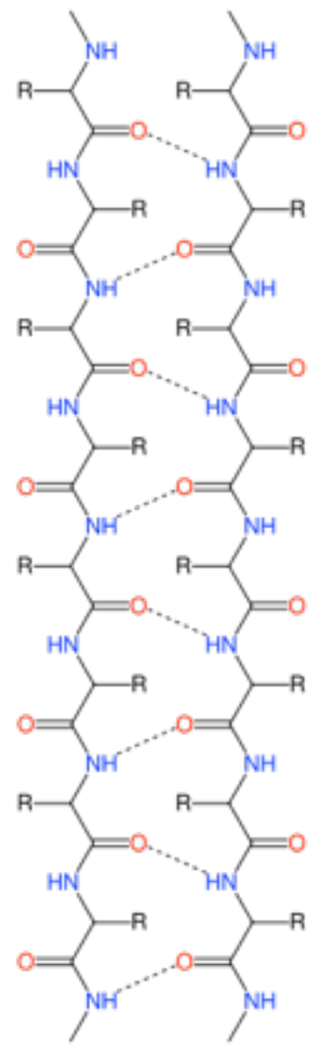
β-sheets consist of β-strands connected laterally by at least two or three backbone hydrogen bonds in a anti-parallel or parallel orientation.

In an antiparallel arrangement, the successive β-strands alternate directions of the N and C-terminus. This is the most stable β-sheet arrangement.

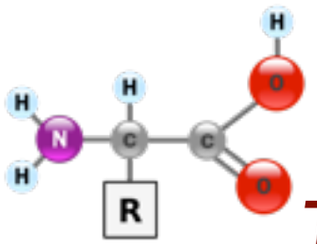
In a parallel arrangement, the N-termini of successive strands are oriented in the same direction, generating a less stable β-sheet due to the non-planarity of the inter-strand H-bonds.



Anti-parallel β-sheets



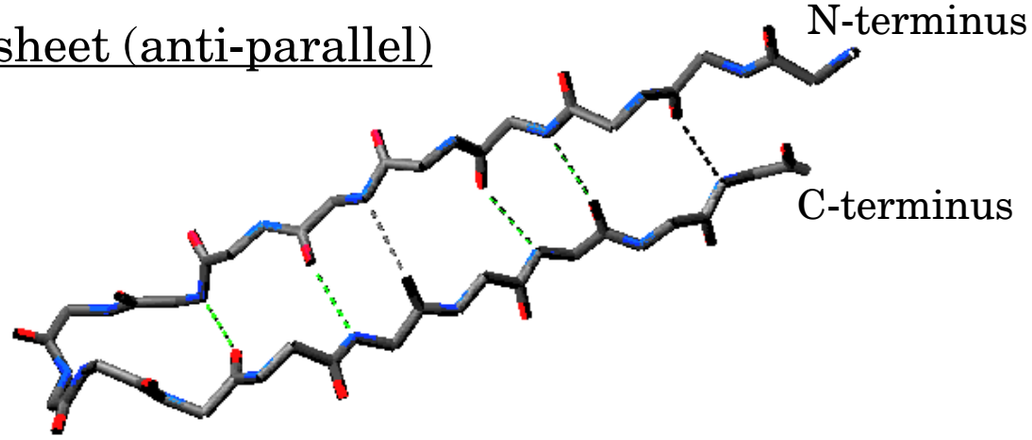
Parallel β-sheets

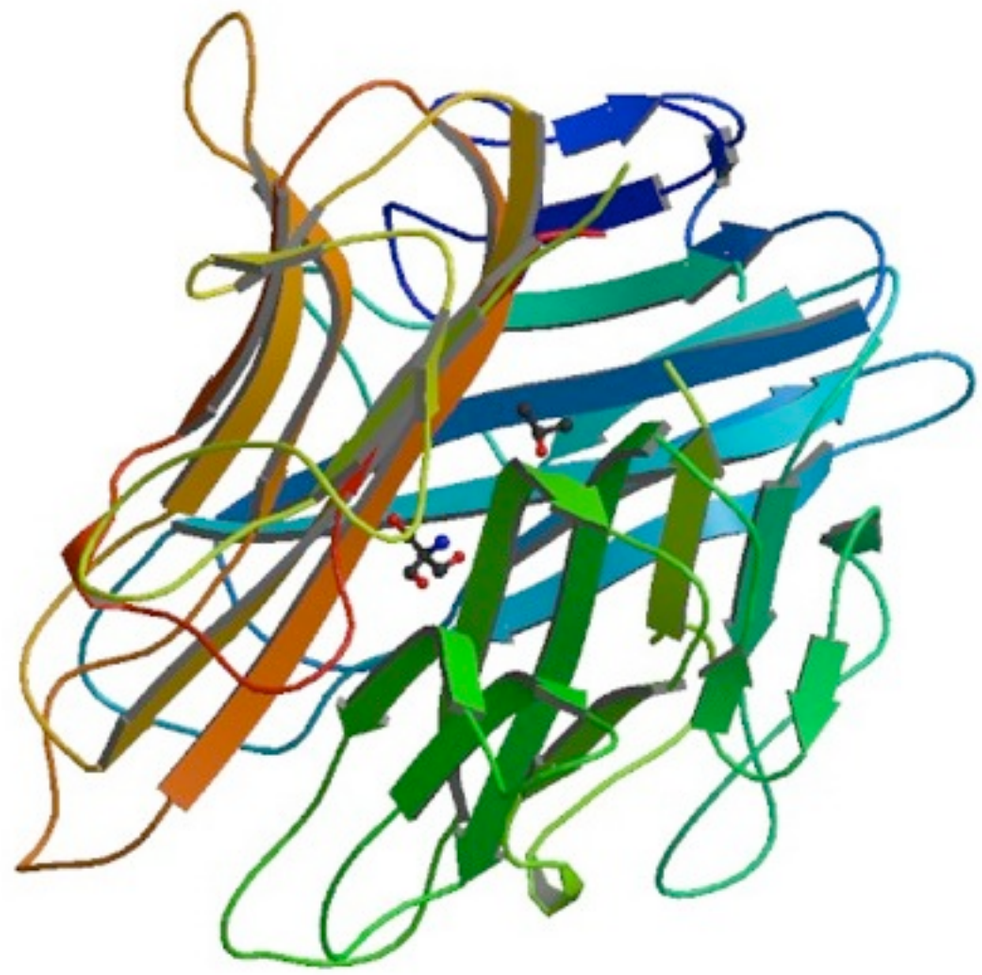


β-sheets Example

Tumor necrosis factor (TNF) from Mouse (PDB: 2tnf)

β-sheet (anti-parallel)





Ideal β-sheets

Real β-sheets

Image credits: Mark Brandt

Secondary Structure

Turns

A turn is **non-regular structure** that connects secondary structure elements and reverses the overall chain direction.

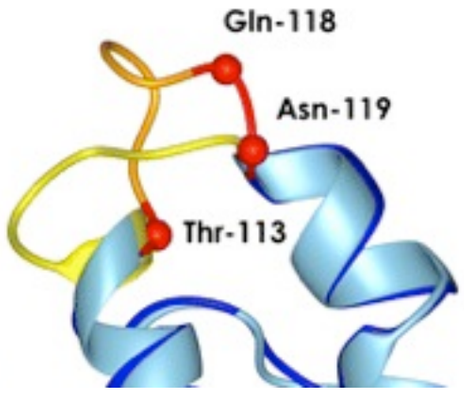
A turn is a structural motif where the C_α atoms of two residues (anchor points) separated by few others (usually 1 to 5) are close in space (< 7 Å).

Turns are classified depending on the number of peptide bonds between the anchor points.

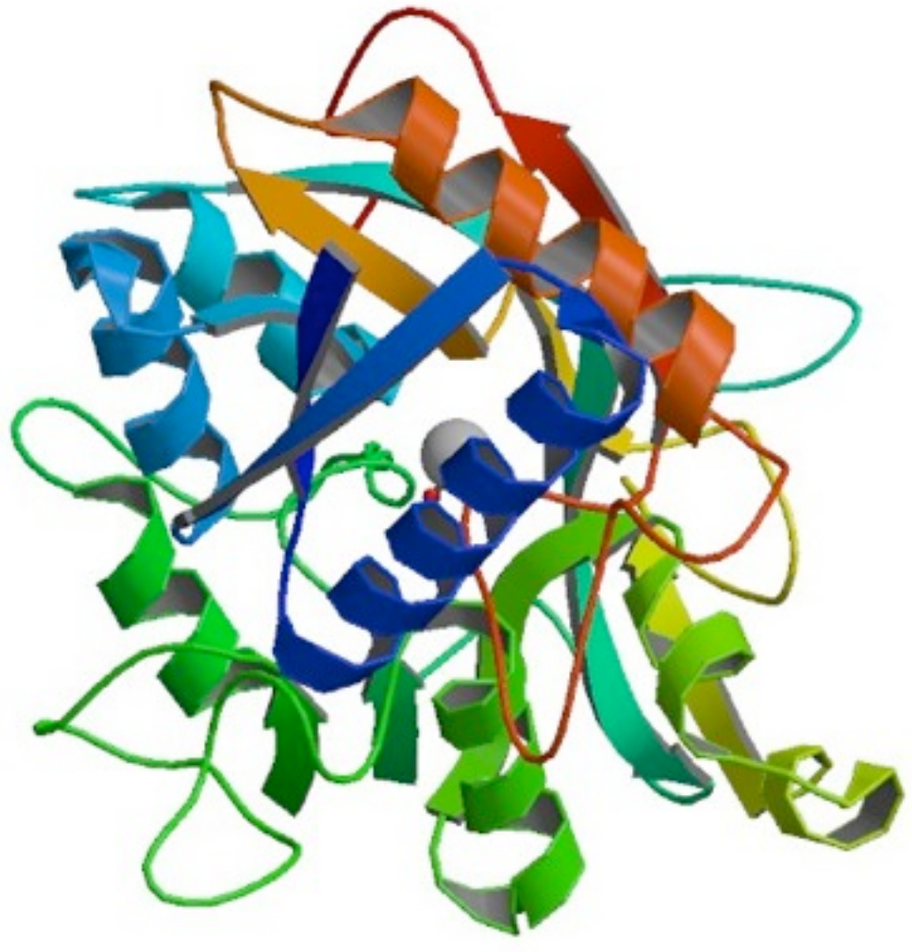
Loops defines longer, extended or disordered turns without fixed internal hydrogen bonding.

Secondary Structure

Turns



Loop example



Loop in a protein

Image credits: Liebau et al, FALC loop server

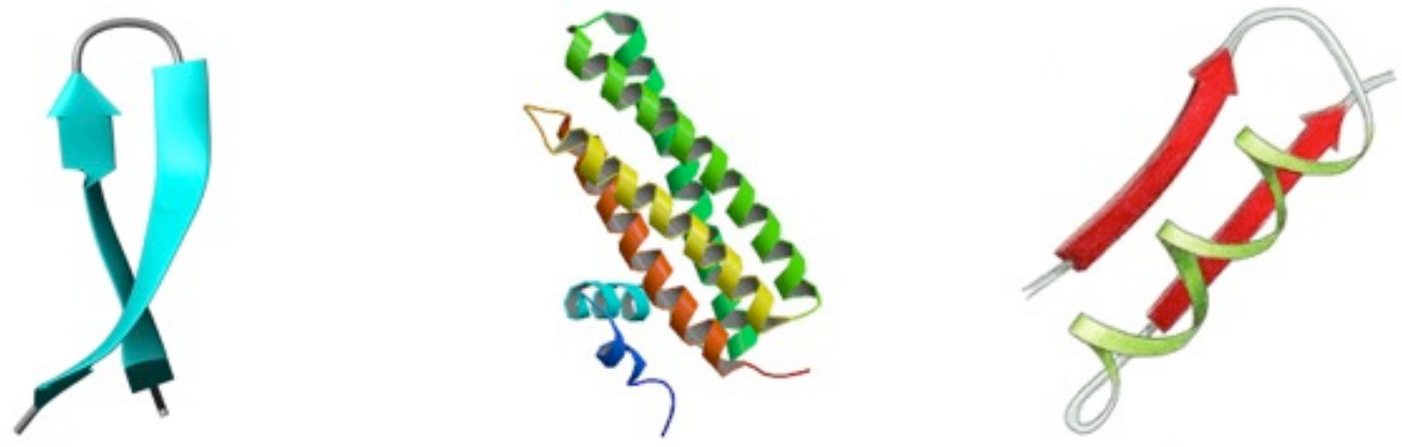
Super Secondary Structure

Structural motifs

A super secondary structure is a compact three-dimensional structure composed of several adjacent elements of secondary structure.

Super secondary structures are smaller than protein domains or subunits .

Examples: β (*a*) and α-helix (*b*) hairpins, and β-α-β motifs (*c*).



a *b* *c*

Tertiary Structure

The tertiary structure is the overall three-dimensional structure of a single protein.

The alpha-helices and beta-sheets are folded into a compact globule.

The folding is driven by the non-specific hydrophobic interactions (the burial of hydrophobic residues from water).

The structure is stabilized by nonlocal interactions (salt bridges, hydrogen bonds, and disulfide bonds).

Quaternary Structure

The quaternary structure is an **assembly** of several protein molecules which form a multimer.

The quaternary structure is stabilized by the same non-covalent interactions and disulfide bonds as the tertiary structure.

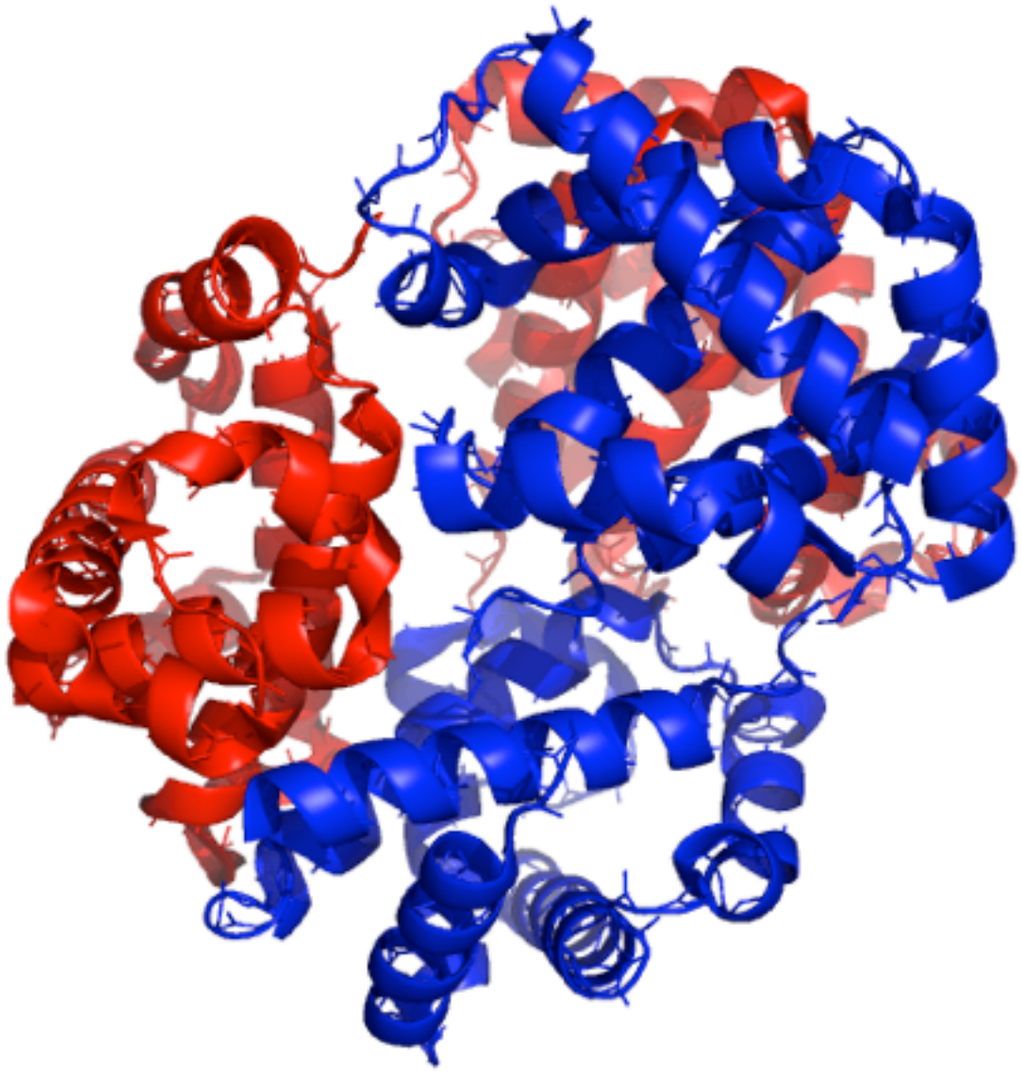
Multimer can be made up of identical subunits ("homo-mer" (e.g. a homotetramer) or of different subunits "hetero-" (e.g. a heterotetramer).

Many proteins do not have the quaternary structure and function as monomers.

Quaternary Structure Example

The two α (blue) and two β (red) chains of hemoglobin

Side view



Front view

Summary

Protein structural levels

Primary Secondary Tertiary Quaternary

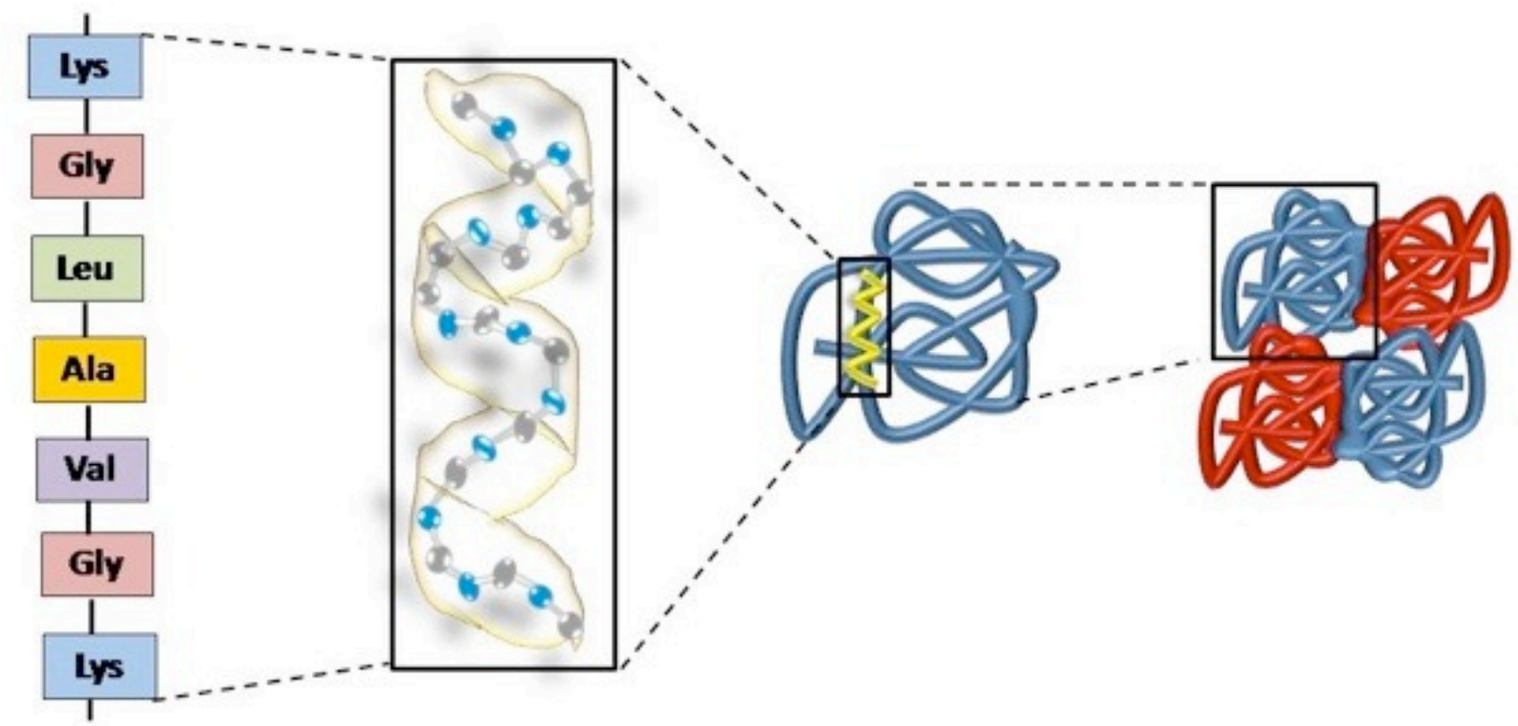


Image credits: http://iitb.vlab.co.in/

Domains

A protein domain is a part of protein that exist **independently** of the rest of the protein chain.

Each domain forms a compact three-dimensional structure and can be independently stable and folded (~25 up to 500 AA).

Many proteins consist of several structural domains.

One domain may appear in a variety of different proteins.

Domains often form functional units.

Protein types

Fibrous, membrane, and globular

Fibrous proteins are long narrow molecules, mostly involved in forming macroscopic structural elements (e.g. keratin or collagen).

Membrane proteins typically have a hydrophobic region (frequently α-helical) that interacts with the non-polar interior of membranes.

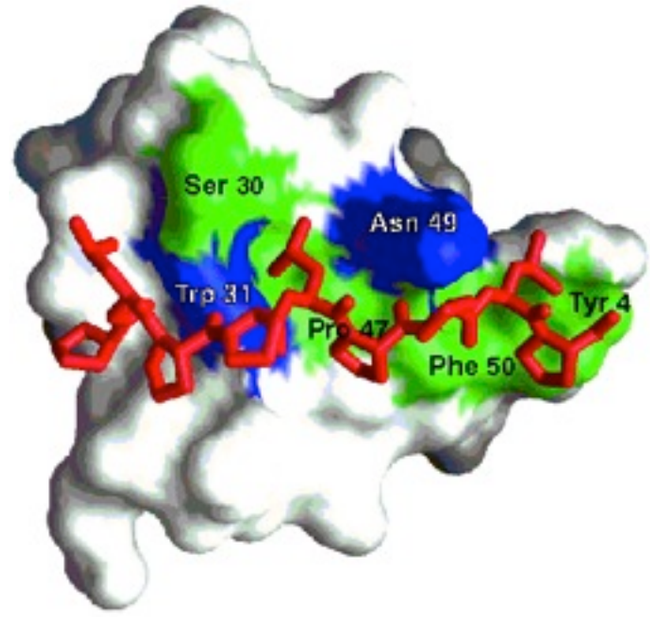
Globular proteins are a diverse class of soluble proteins. Many of the most heavily studied proteins are members of this class of proteins.

Protein Structure Relevance

The biochemical function (activity) of a protein is defined by its interactions with other molecules.

The biological function is in large part a consequence of these interactions.

The 3D structure is more informative than sequence because interactions are determined by residues that are close in space but are frequently distant in sequence.



Relation Structure-Function

Evolution tends to conserve function.

Function depends directly on structure rather then on sequence.

Consequently, structure is more conserved in evolution than sequence.

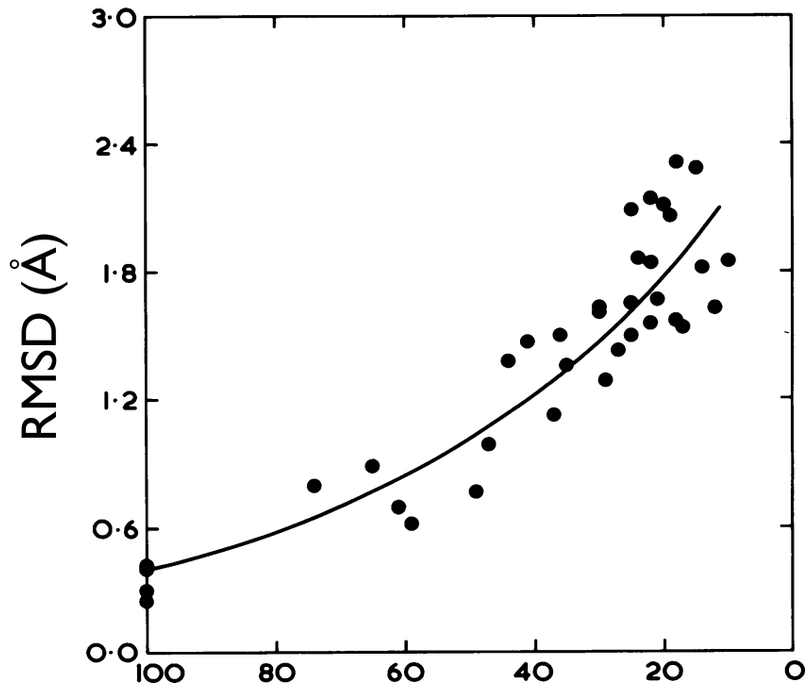
Patterns in space are more recognizable than patterns in sequence.

Sequence-Structure Conservation

Highly conserved proteins are often required for basic cellular function, stability or reproduction.

Conservation of protein sequences is indicated by the presence of identical amino acid residues at analogous parts of proteins.

Conservation of protein structures is indicated by the presence of functionally equivalent, though not necessarily identical, amino acid residues and structures between analogous parts of proteins.



Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. EMBO J. 1986 Apr;5(4): 823-6.

Summary

There are three main types of proteins: **Fibrous**, **Membrane** and **Globular**.

The biochemical **function** (activity) of a protein depends on its **three-dimensional structure**.

Since evolution tends to conserve function, structure is more conserved in evolution than sequence.

Questions?
(and coffee break!)