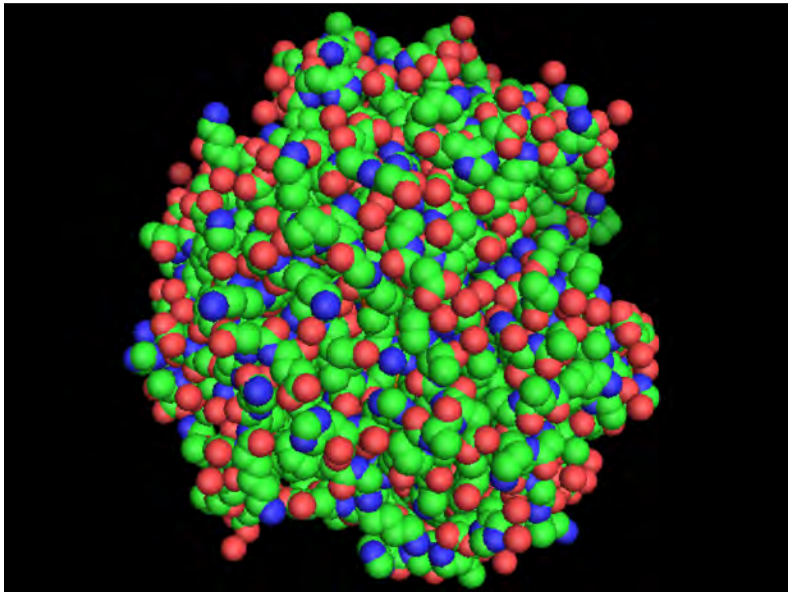
Protein Structure, Function and Disease

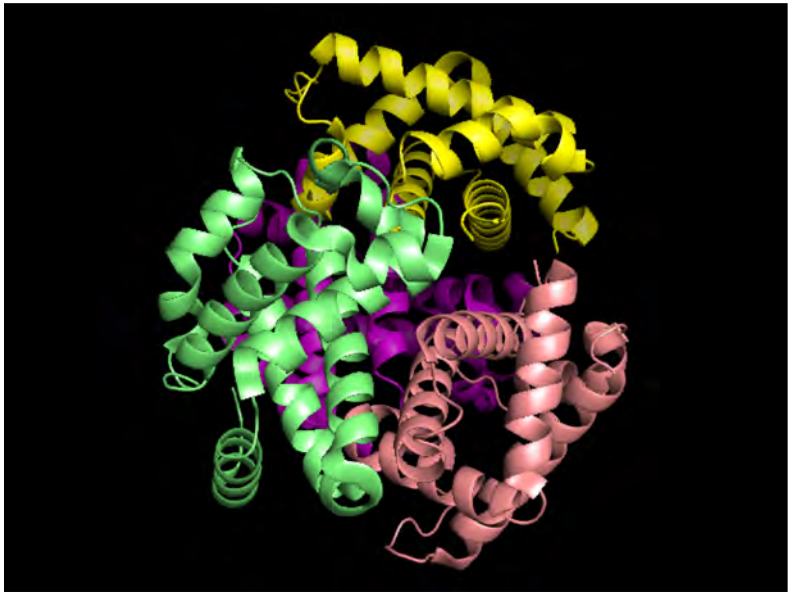
Proteins: from primary to quaternary structures

(Partially adopted from Prof. John Baenziger's former lectures)

Jyh-Yeuan (Eric) Lee, Assistant Professor, BMI

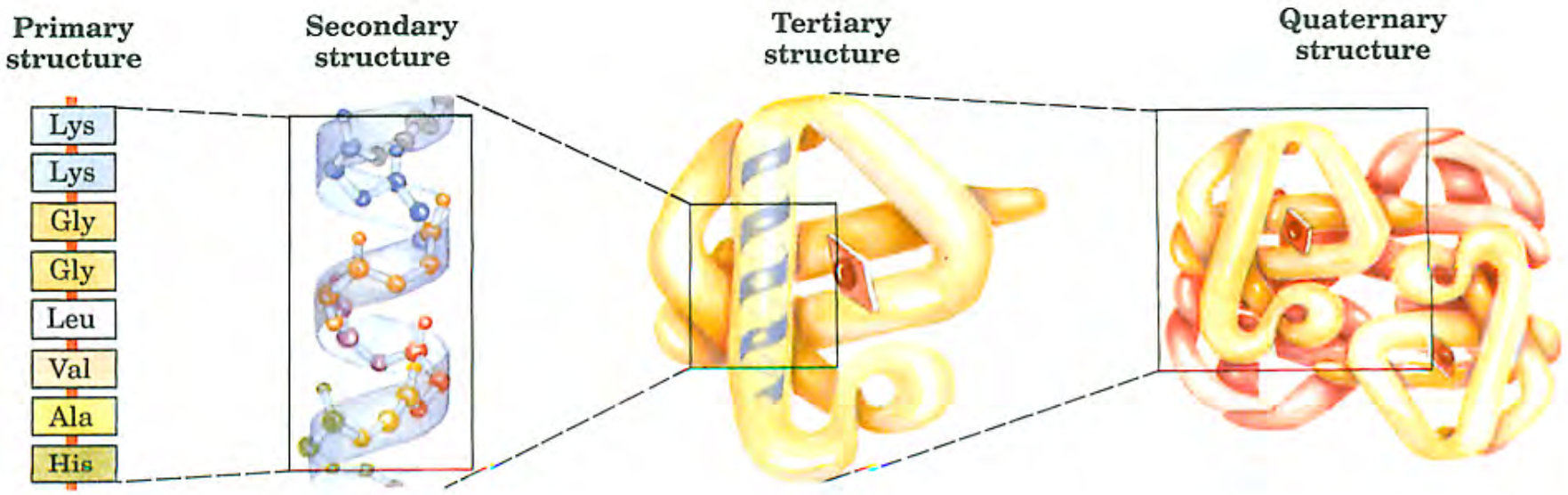
How do we look at structure?





The space filling model of HbA (left) shows van der Waals radii of all atoms, but is too complicated. A ribbon diagram emphasizes well known elements of secondary structure (right), but to understand structure, we must understand the hierarchy of how all these elements fit together. Today we focus on this structural hierarchy to gain insight into the general themes used by nature to create complex protein structures.

The hierarchy of protein structure



Linear sequence of amino acids

α-helix

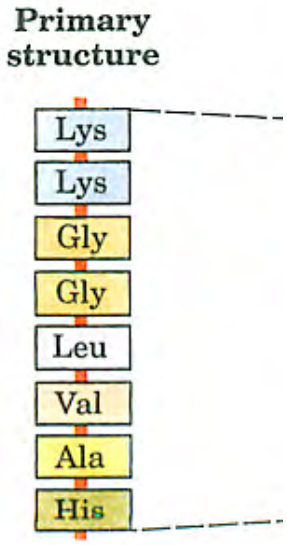
Super-secondary structure/Motif

3D Folded polypeptide chain

Domain

Assembled subunits

Primary Structure



Linear sequence of amino acids

Linear sequence of amino acids in the polypeptide chain

Amino acids are the building blocks of protein function. To understand function, we must understand the chemistry of these amino acid building blocks:

1) Side chain chemistry

2) The peptide bond

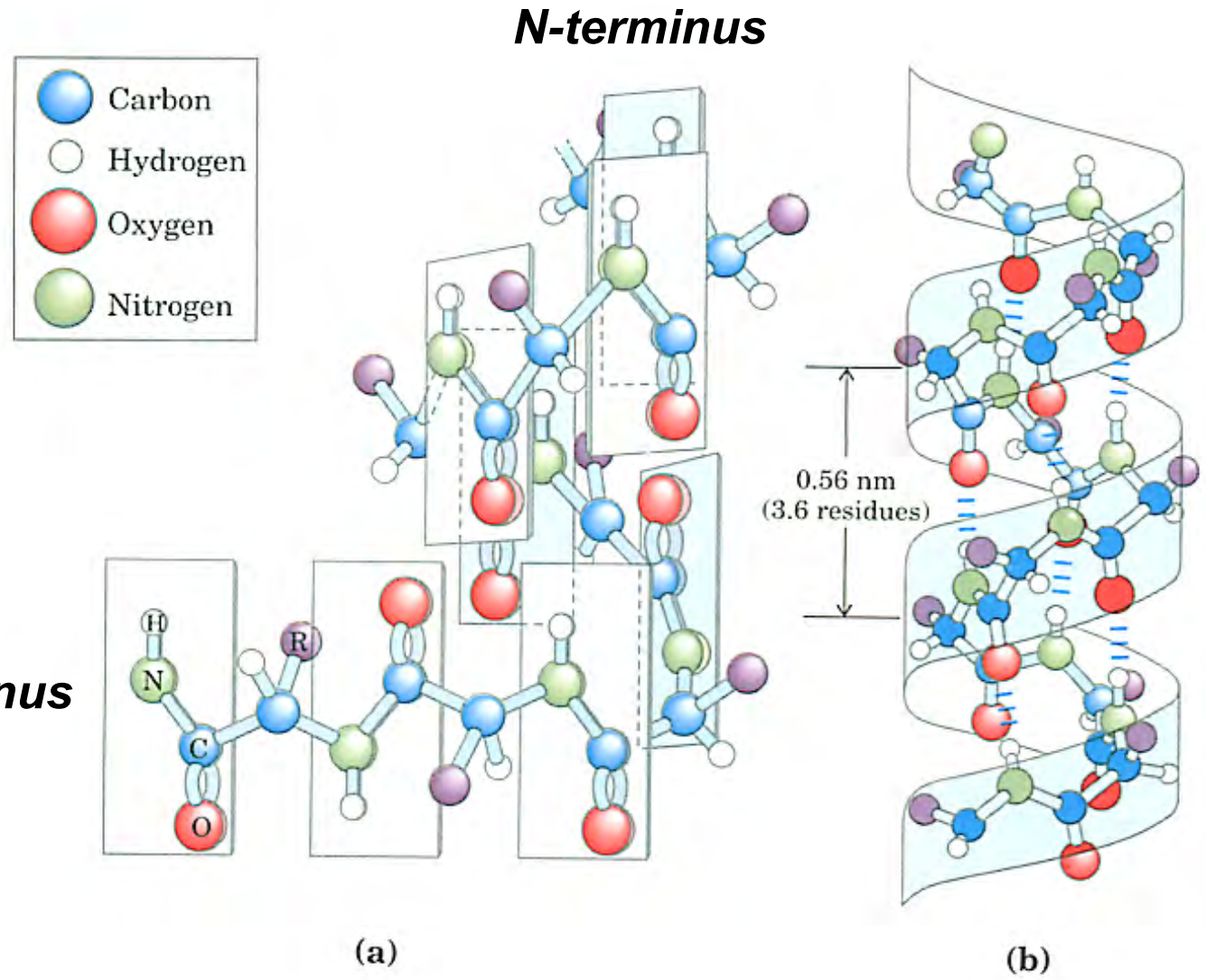
Secondary Structure

α Helix

A regular, recurring arrangement of adjacent amino acids in 3D space defined by main chain hydrogen bonding.

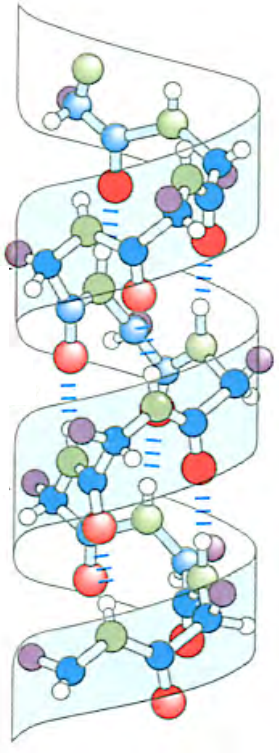
The most prominent forms of secondary structure are the *α-helix*, the *β-sheet*, and the *β-turn*. *These secondary structural features maximize hydrogen bonding of the peptide backbone.*

α-helix


N-terminus
Carbon
Hydrogen
Oxygen
Nitrogen
C-terminus
H
N
R
C
O
0.56 nm
(3.6 residues)
(a)
(b)

α-helix

N-terminus

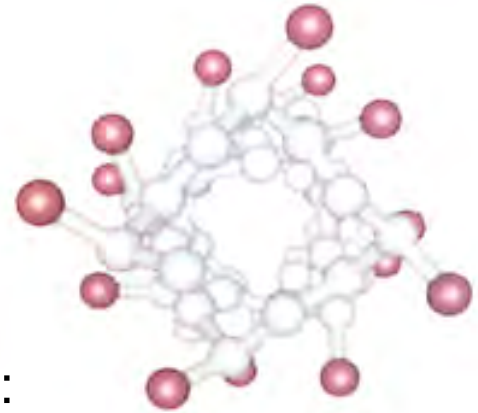


C-terminus

- 3.6 amino acids per turn
- each turn rises 5.41 Å (called the *pitch*) giving 1.5 Å rise per residue
- vast majority are right handed (clockwise rotation looking from N- to C-terminus)
- C=O of residue *i* H-bonds to the N-H of residue *i+4.* H-bonds are straight and of an optimal length (~2.86 Å), but not of optimal geometry (i.e. they are not linear) – still, they are strong! The N-H bond points toward the N-terminus (C=O towards C-terminus)
- planes of the peptide bond are parallel to the helix axis with ϕ (phi) ~ -57^{o} and ψ (psi) ~ -47^{o}.

R-group extends out from helix

The side view (far right) and the view down the axis (below) of a polyalanine α-helix both show that the side chains (in this case a methyl group) extened out and away from the helix axis.



There are two consequences:

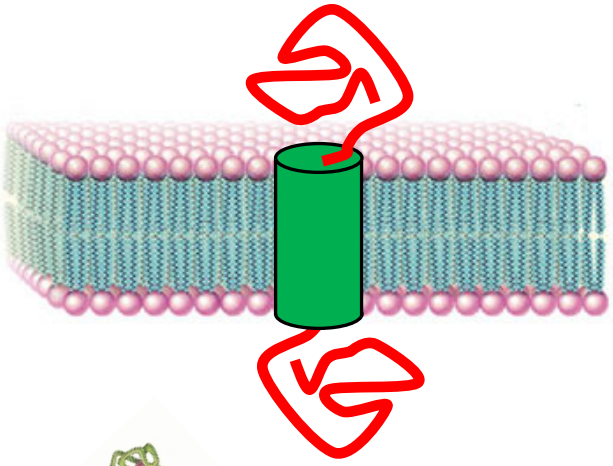
1) Residues 3 – 4 apart in sequence are close together in space

2) The surfaces of α-helices have polarities governed by the side chains

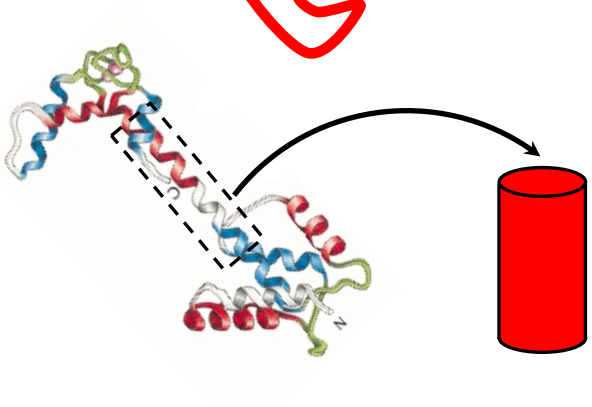
R-groups of residues 3 and 4 apart can interact

- critical interactions occur between side chains on residues that are three or four residues apart in the sequence
- "like" charge at these positions (eg. two E or two K residues) lead to charge repulsion and destabilize an α-helix, opposite charges attract
- bulky residues can sterically interfere with each other
- often find either oppositely charged residues or hydrophobic residues that are 3 or four residues apart in the sequence.

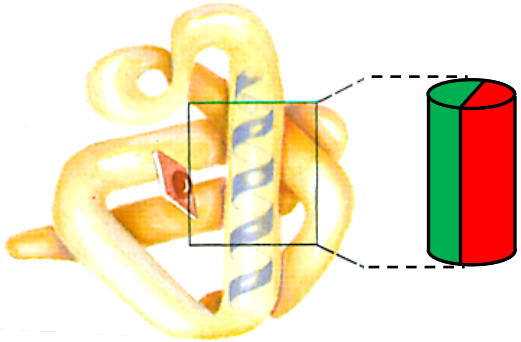
The faces of α-helices have polarity



A single TM α-helix is surrounded by the hydrophobic lipid bilayer and thus its complete surface is typically hydrophobic (green).



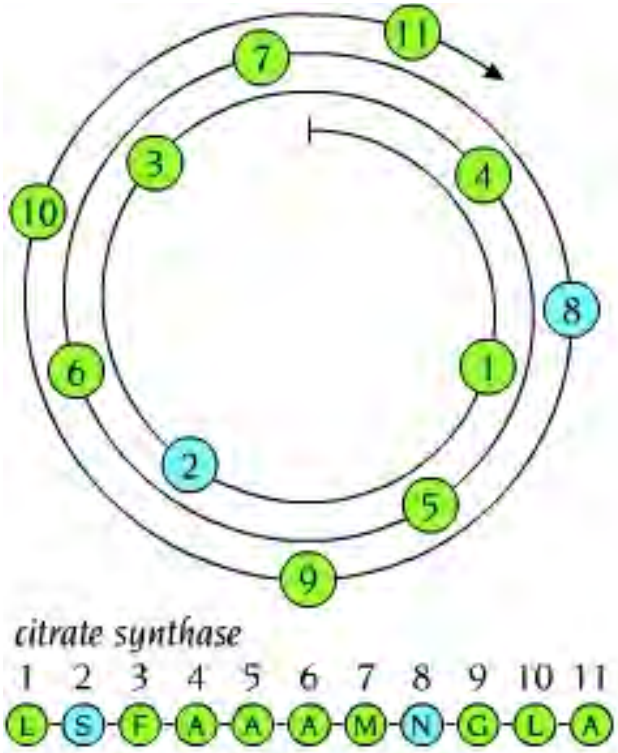
The long α-helix in troponin C is completely exposed to aqueous solvent, and thus its suface is completely hydrophilic (red).



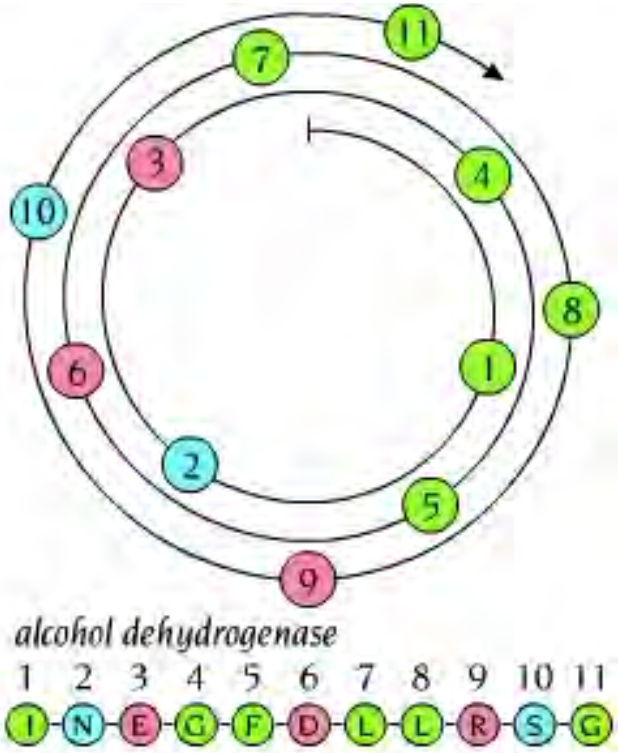
A helix on the surface of myoglobin is exposed to solvent (red) and to the hydrophobic protein core (green). *It is amphipathic!*

α-helices can be polar, non-polar, or amphipathic

Non-polar α-helix

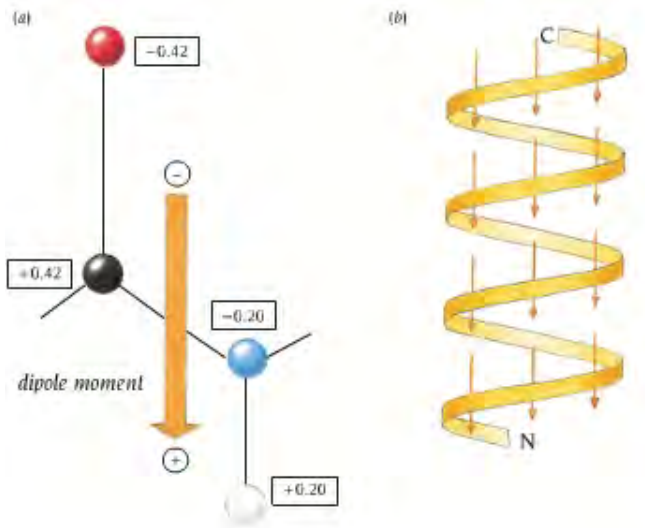


Amphipathic α-helix



The best way to represent the polarity of an α-helix is by a helical wheel representation (like a compressed slinky!), which shows which side chains are on the different faces of the helix. Green=hydrophobic; Red=charged; Blue=polar/neutral

α-helix has a strong dipole moment

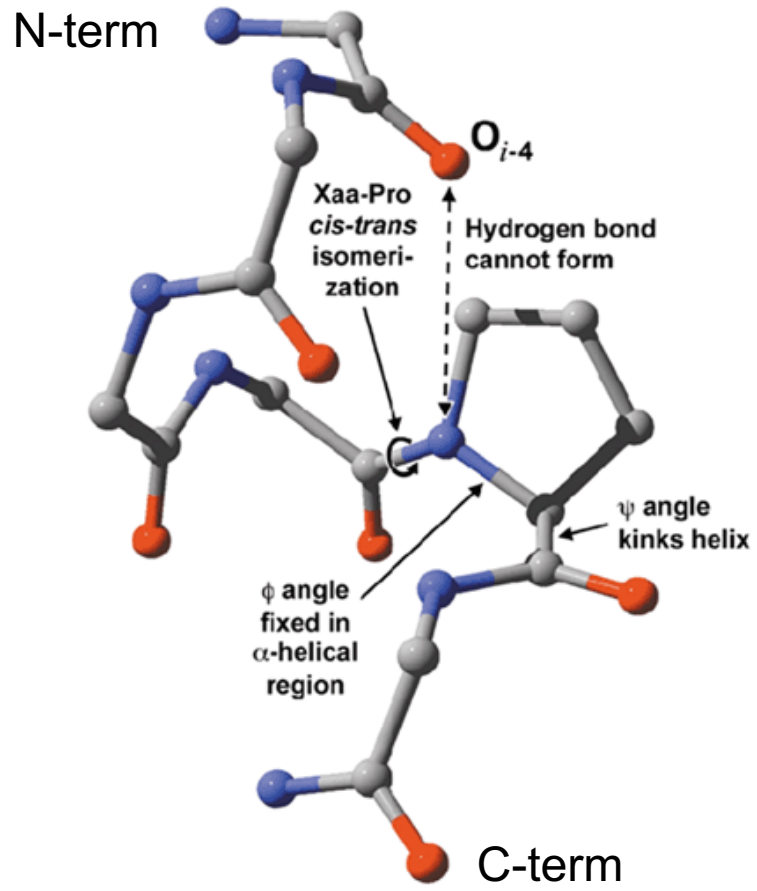


The dipoles of every peptide in an α-helix point in the same direction leading to a large dipole moment roughly equal to the number of residues x 3.5 Debye - equivalent to 0.5-0.7 electron charges at each end of an average α-helix (~8-10 residues). The N-terminus is positive!

Negatively charged amino acids are often found near the amino terminus of the helix to balance the positive pole of the dipole, positive residues are often found at the C-terminus

The α-helix dipole is used to help stabilize protein-ligand interactions

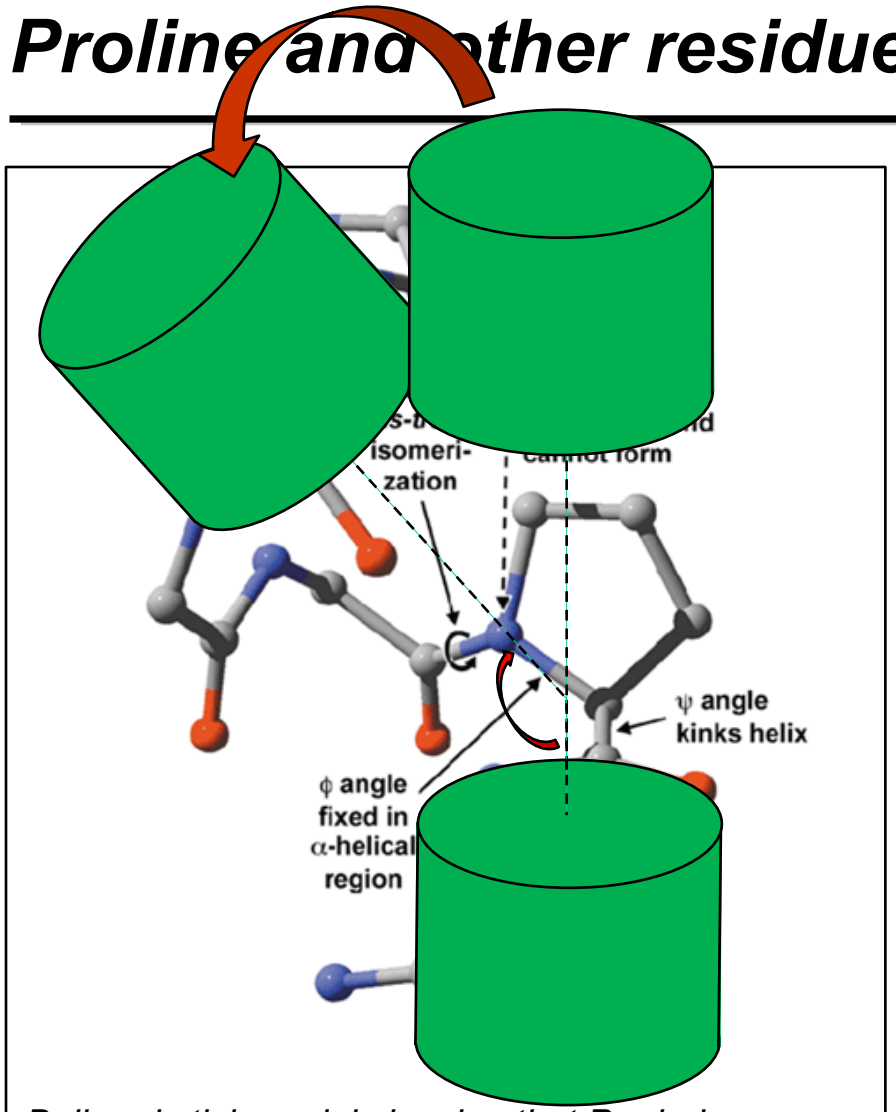
Proline is (usually) an α-helix breaker



Ball and stick model showing that Pro induces a kink and cannot form a hydrogen bond as typically found in an α-helix (Deber & Therien (2002) Nat. Struct. Biol. ***9****, 318.*

Pro residues tend to destabilize α-helices and lead to a kink, as rotation about the N-C_α is not possible and there is no hydrogen on the nitrogen. Pro at position *i* cannot form a hydrogen bond with the N-H of residue *i-4*. The oxygen at position *i-4* thus pulls away from the Pro side chain.

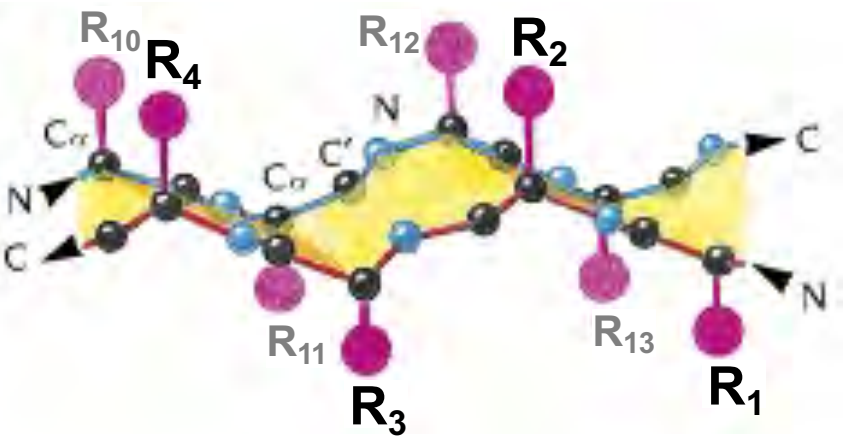
Proline and other residues are α-helix breakers



*Ball and stick model showing that Pro induces a kink and cannot form a hydrogen bond as typically found in an α-helix (Deber & Therien (2002) Nat. Struct. Biol. **9**, 318.*

Pro residues tend to destabilize α-helices and lead to a kink, as rotation about the N-C_α is not possible and there is no hydrogen on the nitrogen. Pro at position *i* cannot form a hydrogen bond with the N-H of residue *i-4*. The oxygen at position *i-4* thus pulls away from the Pro side chain.

β-strand/β-sheet

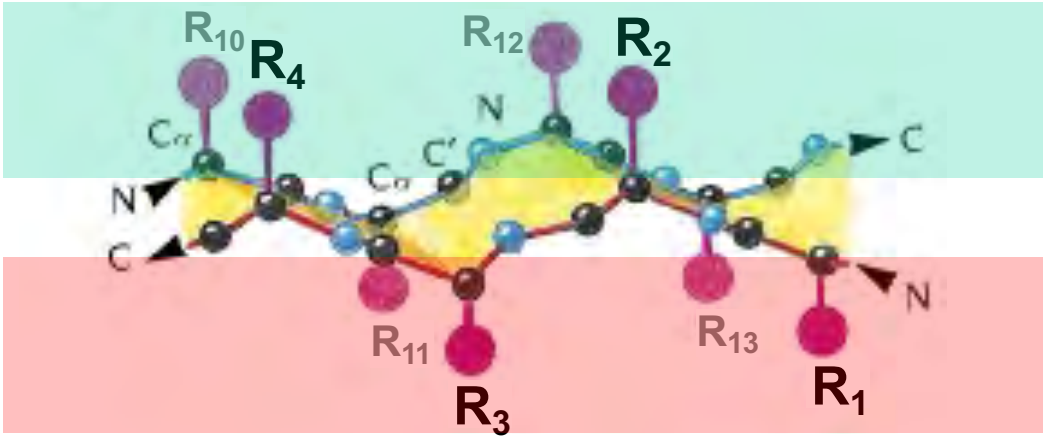


The polypeptide chain is in an extended conformation

Several chains (4-6 or more) typically align to form a *β-sheet* with H-bonding between the peptide C=O and N-H groups of adjacent strands. Adjacent strands can run in parallel or anti-parallel directions.

Strands translate ~3.4 Å per residue

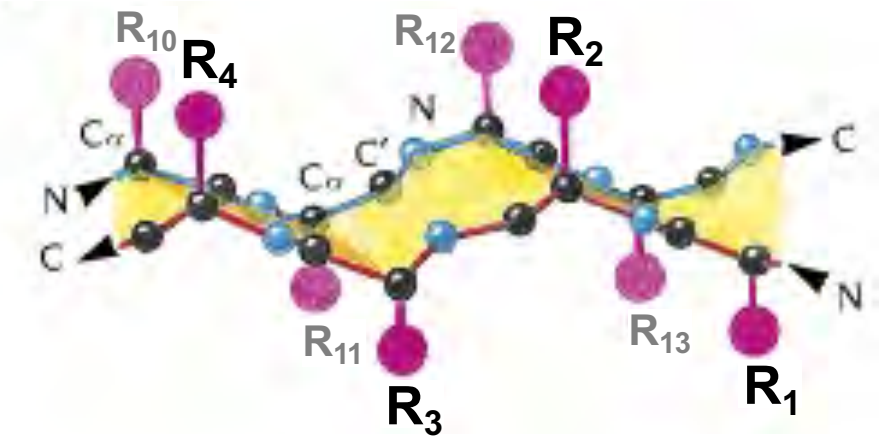
Faces of β-sheet can have distinct polarities

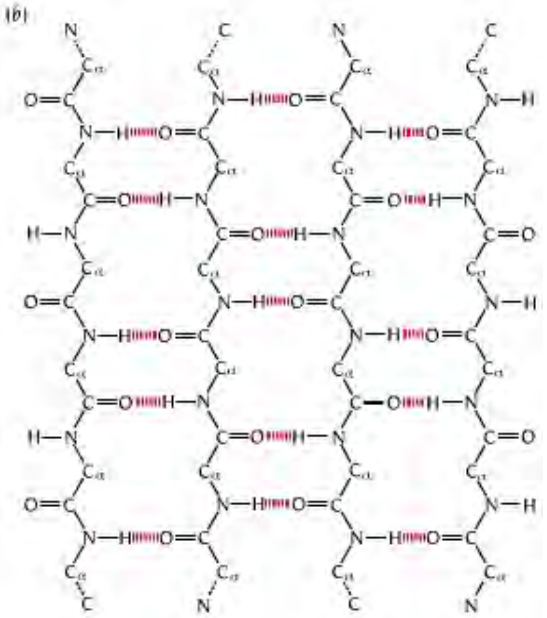


R-groups extend alternatively above and below the plain of the β-sheet. β-sheets can thus be polar, non-polar, or amphipathic (the latter with periodicity of 1polar, 1 non-polar, 1 polar, 1 non-polar, etc.).

For example: If R_2, R_4, R_{10}, and R_{12} are non-polar (green) and R_1, R_3, R_{11}, adn R_{13} are polar (red). The the top surface of the β-sheet is polar and the bottom surface non-polar – the α-helix is amphipathic!

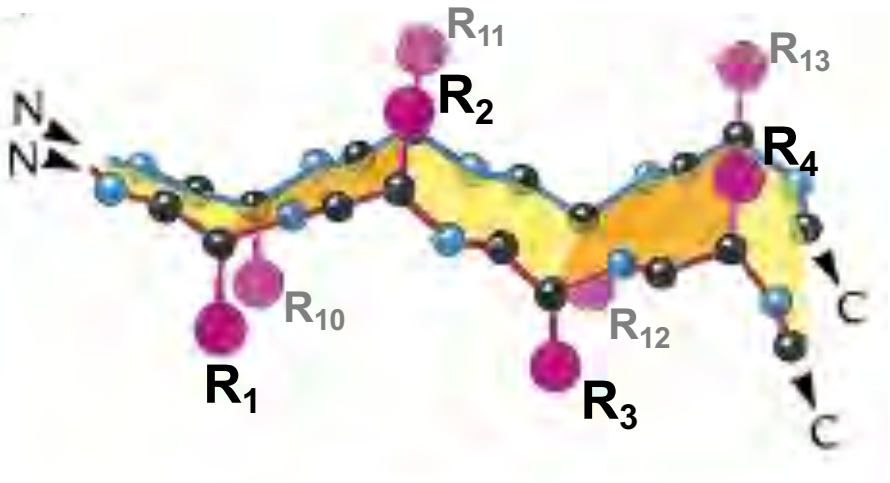
Antiparallel β-sheet

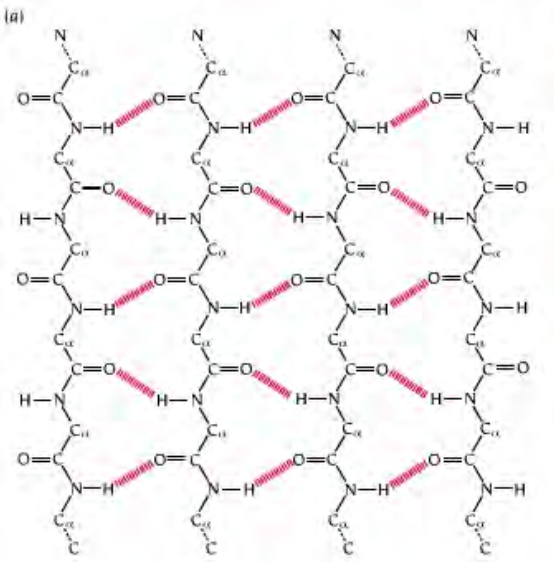




- β-strands run in opposite directions.
- hydrogen bonds are straight and close to optimal length (strong)
- tend to be the most stable
- ϕ (phi) = -139° and ψ (psi) = +135°.

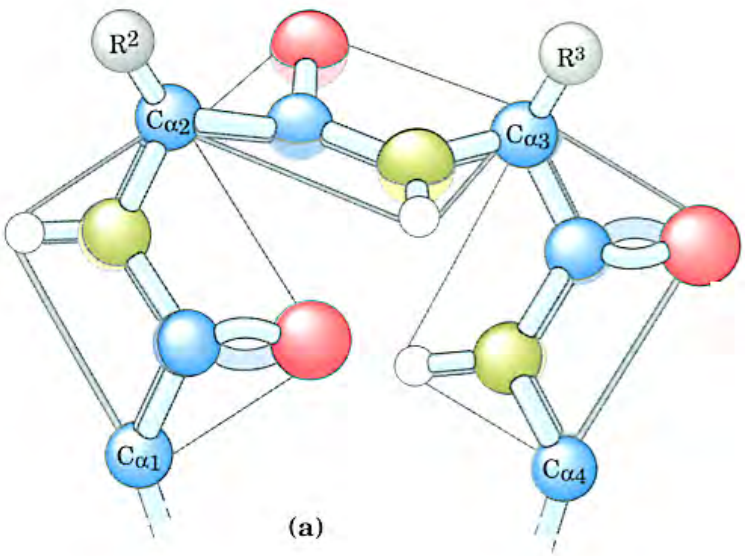
Parallel β-sheet

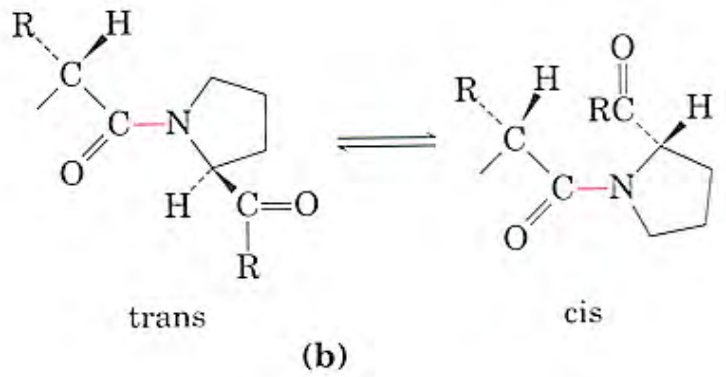




- β-strands run in the same direction
- hydrogen bonds are not optimal
- usually slightly less stabile than anti-parallel β-sheets
- ϕ (phi) = -119^o and ψ (psi) = +133^o.

β-bend/β-turn

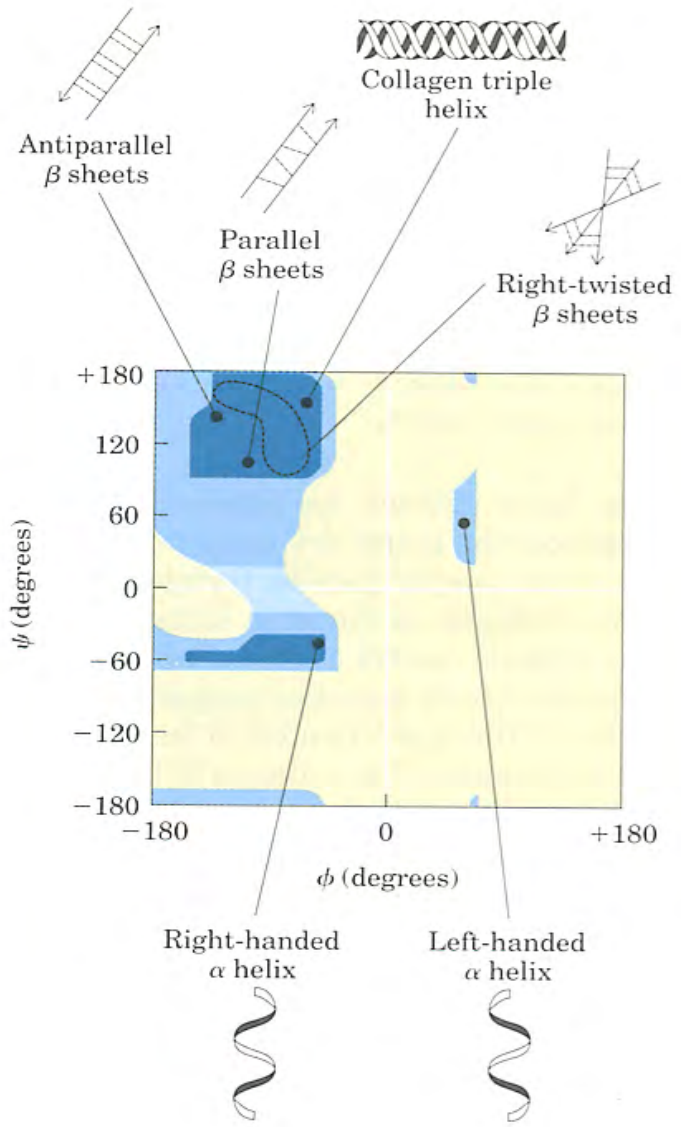




In a β-turn, the polypeptide backbone makes an abrupt turn in direction using 4 amino acids with the C=O of residue #1 H-bonding to the N-H of residue #4.

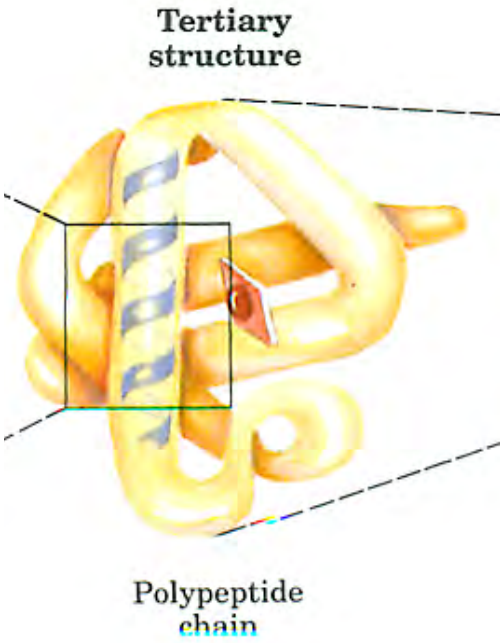
Both Gly and Pro are common. Gly is common because it is small, so it can adopt the strained ϕ and ψ angles required in a β-turn. Pro is common because it more easily adopts the *cis* conformation, which fits better in a β-turn.

Ramachandran plot



α-helices and β-sheet all fit into allowed regions of the Ramachandran plot. The tight β-turn requires unique ϕ and φ angles that can most easily be accomplished by Gly and *cis*-Pro.

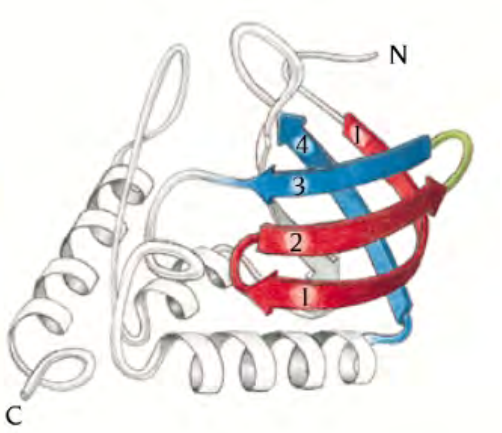
Tertiary Structure



The 3D arrangement of all secondary structural features as well as those amino acid residues that are not found in defined secondary structures.

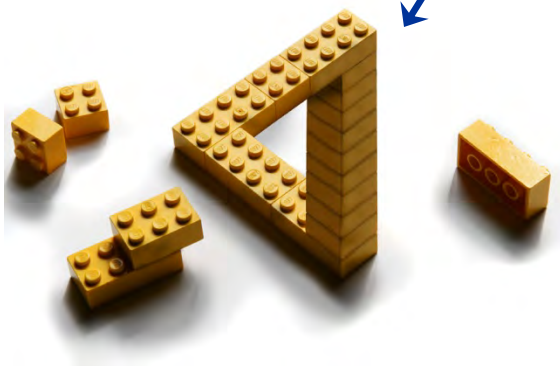
Protein tertiary structures can be grouped into 3 general categories: *α-structures*, *α/β-structures*, and *β-structures*.

Super-secondary structure/motifs

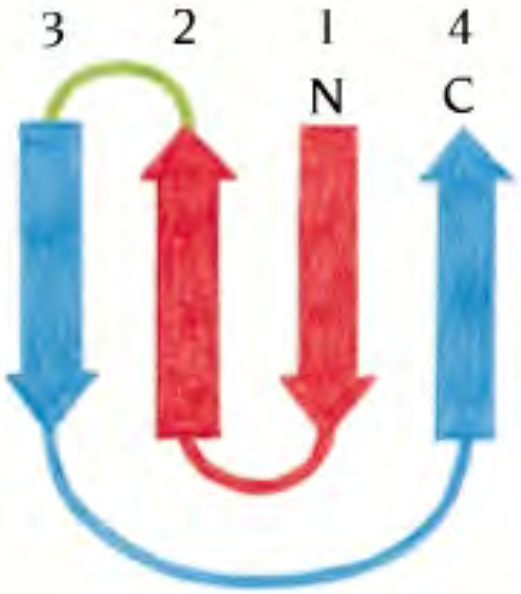


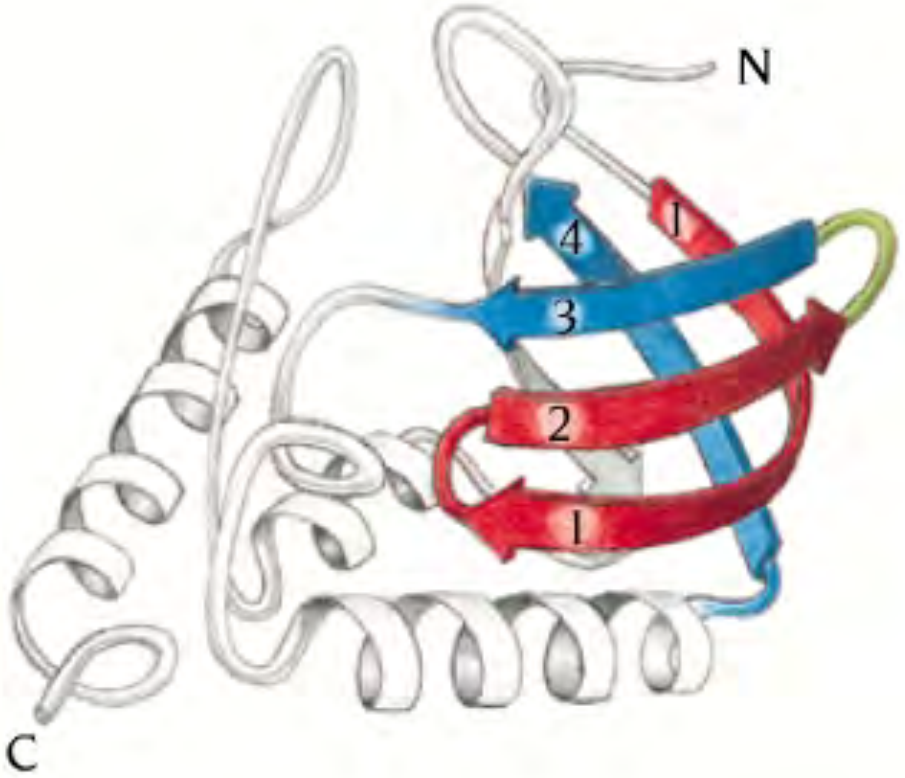
A regular arrangement of two or three secondary structures that is repeated in many different proteins and/or many times within a protein.

A super-secondary structure can have a either a *structural* or a *functional* role in proteins.



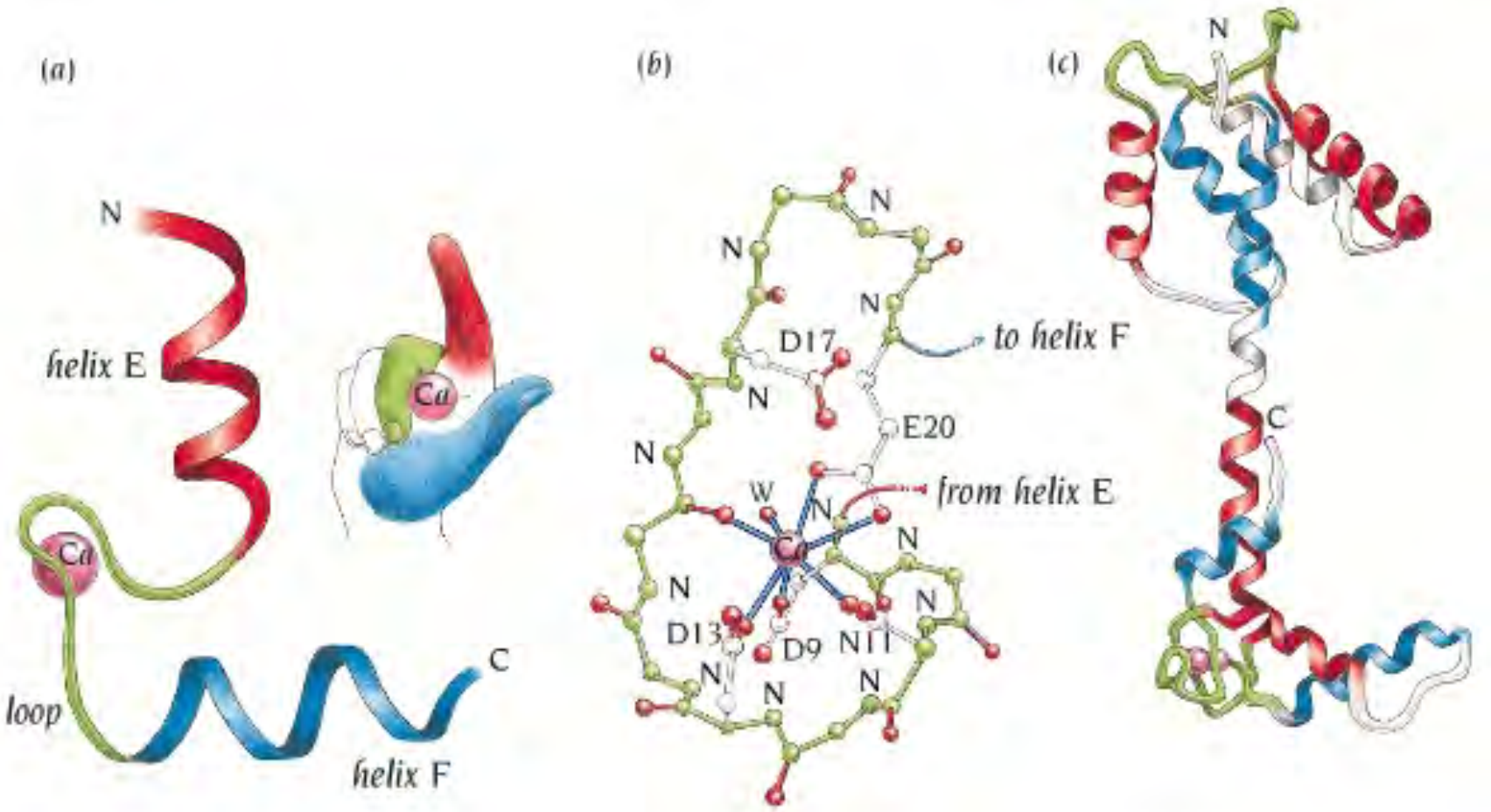
Structural motifs: Greek key





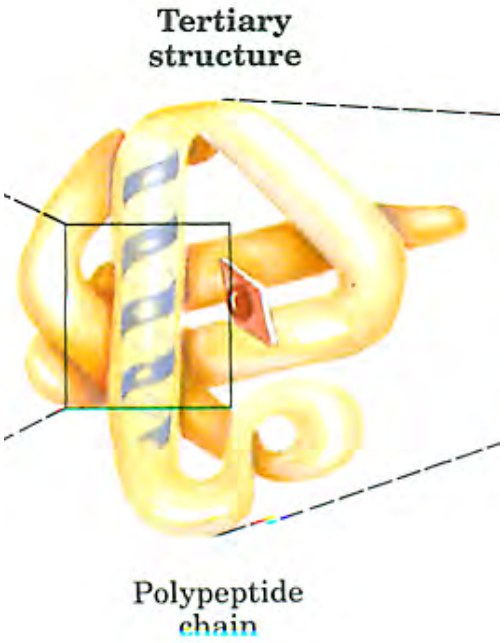
The Greek key motif is a super-secondary structural element found in many proteins. It has a structural as opposed to a functional role. It is a structural building block – analogous to a LEGO building block.

Functional motifs: EF hand



The EF hand is a functional super-secondary structural element found in many proteins, such as troponin C (far right), that binds Ca^{2+}. It has a functional as opposed to a structural building block role.

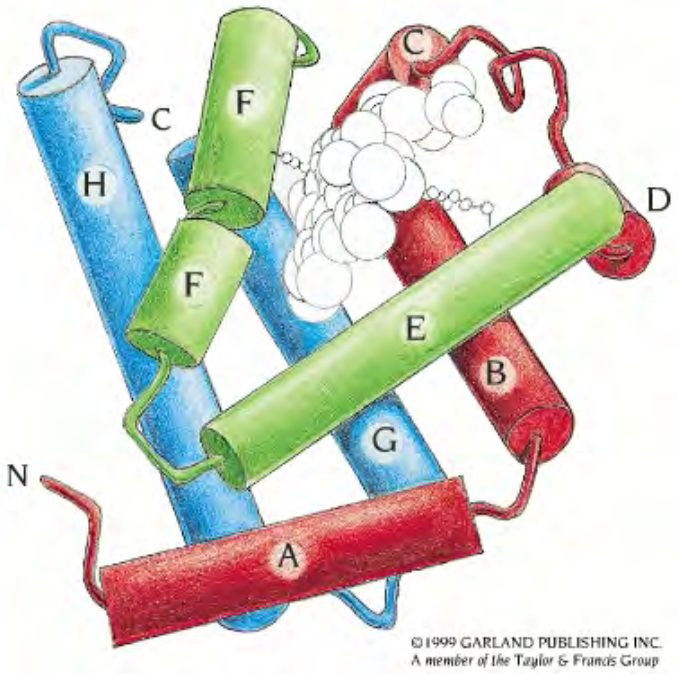
Tertiary Structure



The 3D arrangement of all secondary structural features as well as those amino acid residues that are not found in defined secondary structures.

Protein tertiary structures can be grouped into 3 general categories: *α-structures*, *α/β-structures*, and *β-structures*.

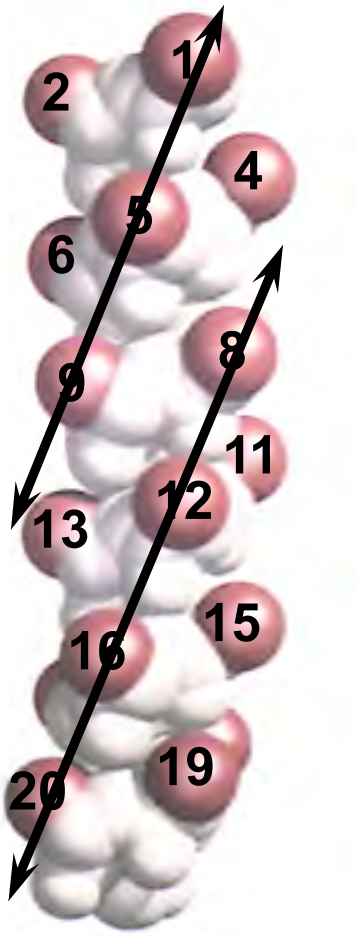
Alpha-structure proteins: the globin fold



α-structure proteins are proteins formed exclusively from α-helices, the most common member are the oxygen binding proteins, myoglobin and hemoglobin. Both adopt what is referred to as the globin fold. This fold consists of 8 α-helices (A-H) connected by short loop regions arranged so that they enclose a pocket for the active site oxygen-binding heme group. *The globin fold illustrates the "ridges in grooves" model of α-helix/α-helix packing.*

Helix ridges and grooves

The side view of a space filling model of a polyalanine α-helix is shown on the right

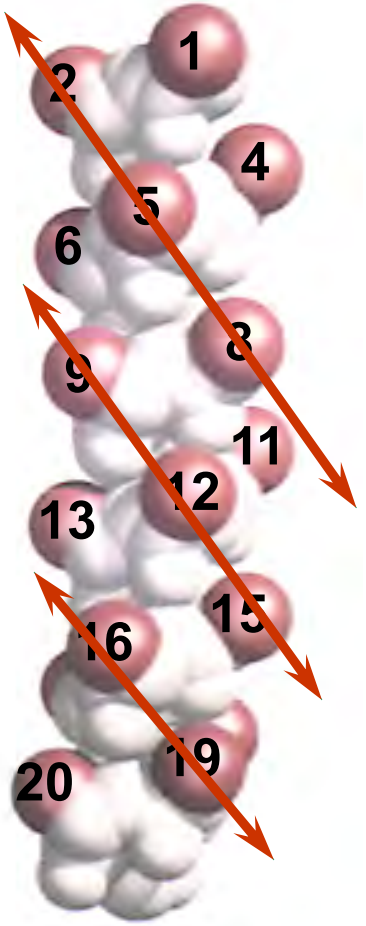


The methyl R groups are arranged on the surface in helical rows that form ridges separated by shallow grooves (think of furrows in a ploughed field)

There are two “sets” of ridges and grooves – those formed from amino acids that are 3 apart in sequence (red lines), and those that are formed from residues 4 apart (black lines)

Helix ridges and grooves

The side view of a space filling model of a polyalanine α-helix is shown on the right



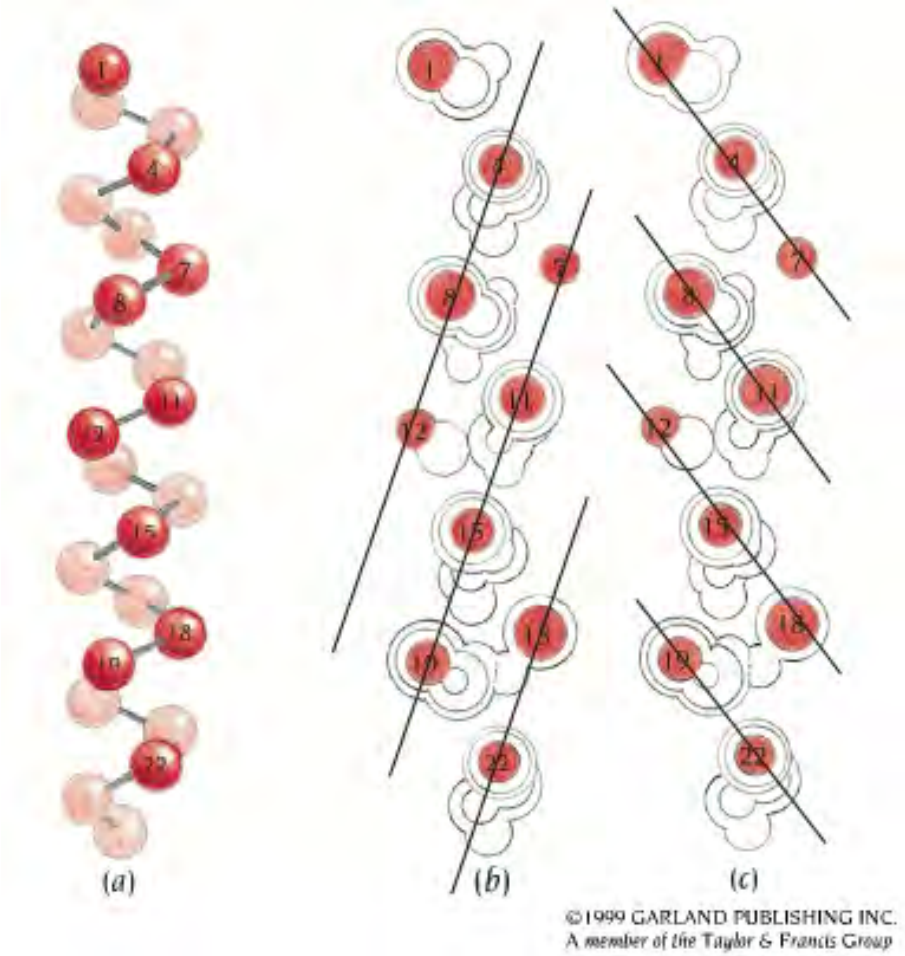
The methyl R groups are arranged on the surface in helical rows that form ridges separated by shallow grooves (think of furrows in a ploughed field)

There are two “sets” of ridges and grooves – those formed from amino acids that are 3 apart in sequence (red lines), and those that are formed from residues 4 apart (black lines)

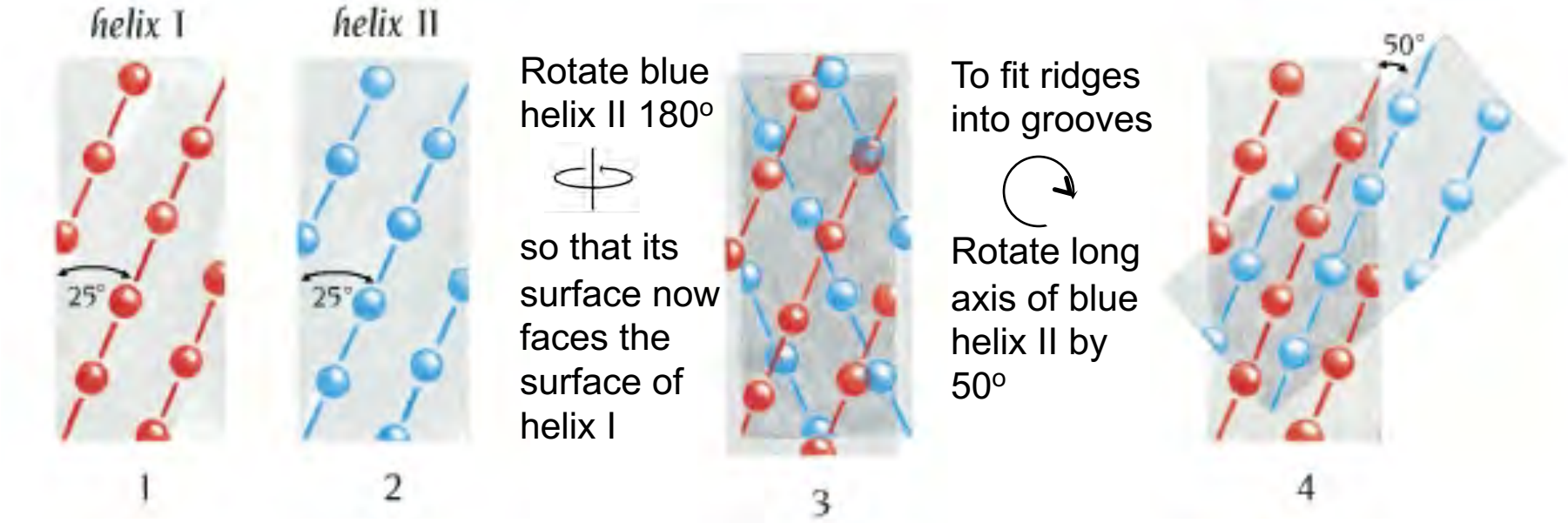
Helix-helix packing

Ridges and grooves formed by residues 4 apart in sequence make a 25^{o} angle relative to the long axis of the α-helix.

Ridges and grooves from residues 3 apart in sequence make an angle of 45^{o} relative to the long axis of the α-helix.

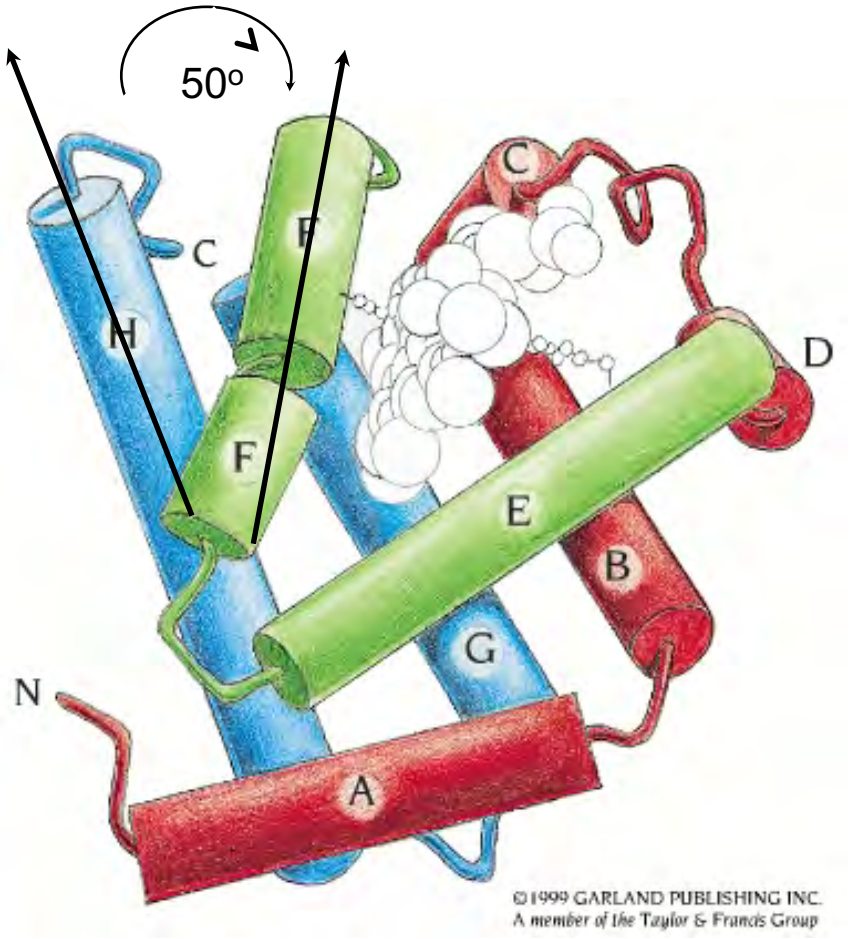


Helix-helix packing – method I



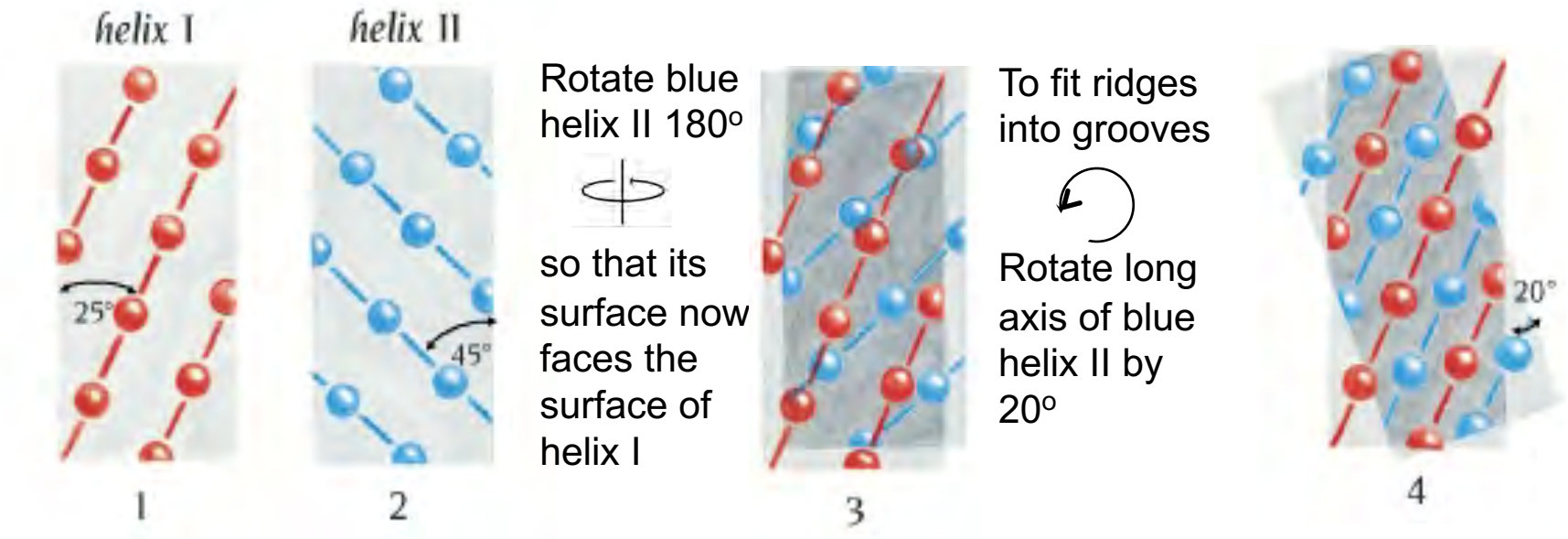
You can fit the ridges formed by residues 4 apart on one helix into the grooves formed by residues 4 apart on the adjacent helix. These ridges form a 25° angle relative to the helix axis (*1&2*), so if you face the two helices together, these ridges and grooves are 2x25° apart (*3*). One helix axis must therefore be rotated by 50° relative to the other to fit the ridges into the appropriate grooves (*4*). Inter-helical angles of 50° are found in the globin fold!

Ridges into grooves packing in the globin fold



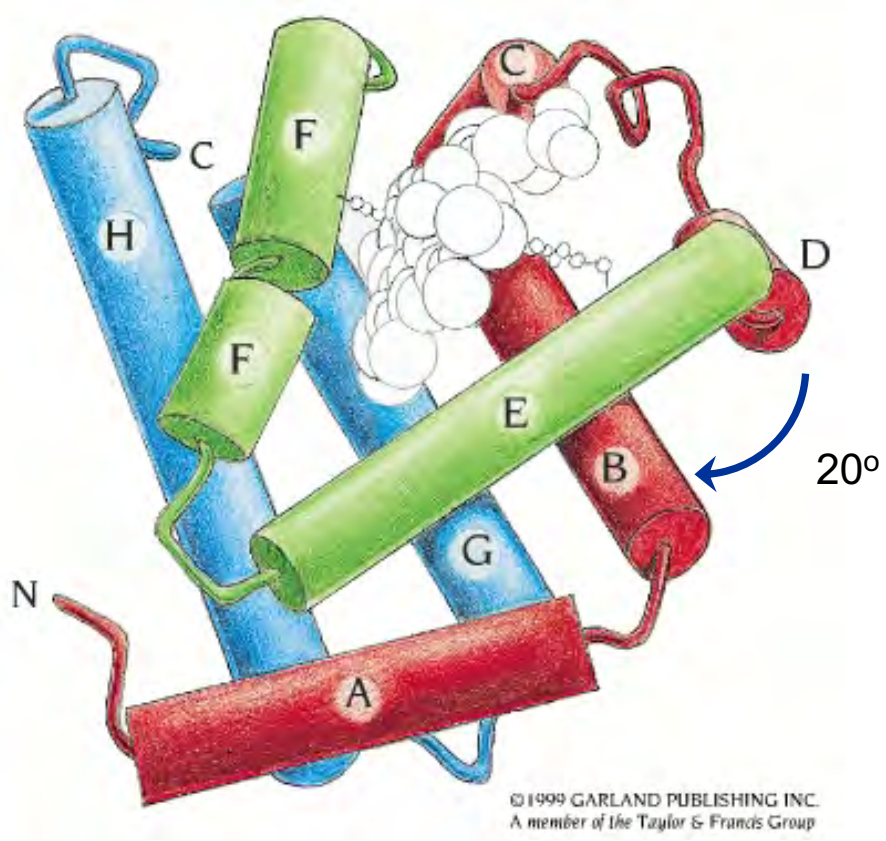
Many α-helices pack the ridges (4 residues apart) of one α-helix into grooves (4 residues apart) of the adjacent α-helix leading to an angle between the two α-helices of 50° (e.g. helix F vs H; helix E vs B; etc.).

Helix-helix packing – method II



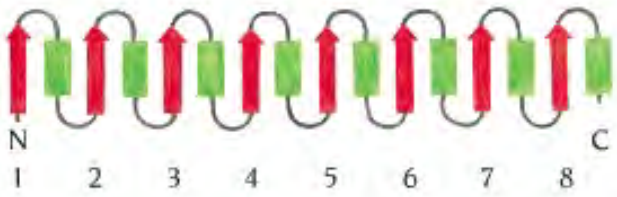
You can fit the ridges formed by residues 4 apart on one helix into the grooves formed from residues 3 apart on the other helix (or vice versa). The ridges and grooves formed from residues 3 apart in the sequence are at a 45° angle to the helix axis (those 4 apart at a 25° angle)(*4*&*3*, respectively). The two helices packed together in this manner give an inter-helical angle of 45°-25° = 20° (*4*).

Ridges into grooves packing in the globin fold



Many α-helices pack the ridges (4 residues apart) of one α-helix into grooves (3 residues apart) of the adjacent α-helix leading to an angle between the two α-helices of 20° (e.g. helix G vs H; helix B vs G).

Alpha/Beta Structures: α/β or TIM barrel



α/β-structured proteins are abundant. The α/β barrel (or TIM barrel) is a common fold found in enzymes, and is composed of 8 alternating β-strands and α-helices (the TIM barrel is named after the enzyme, Triosphosphate IsoMerase).

The β-sheet folds back on itself to form a β-barrel, with hydrophobic residues located inside and α-helices on the outside of the barrel.

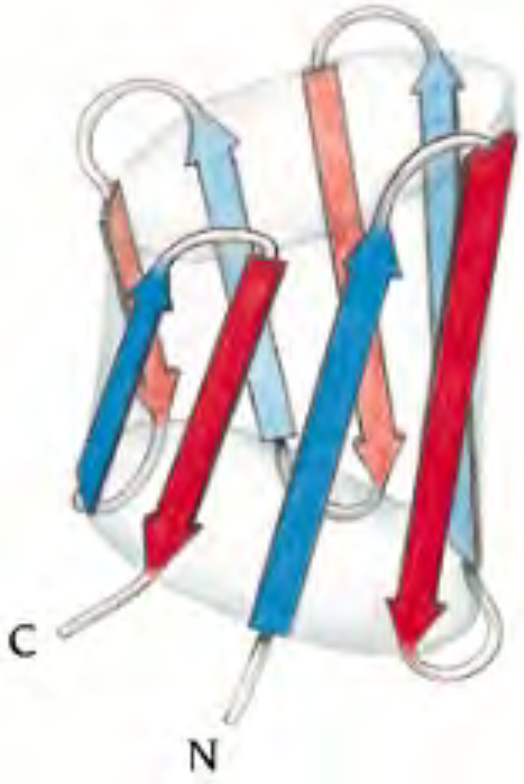
Alpha/Beta Structures:

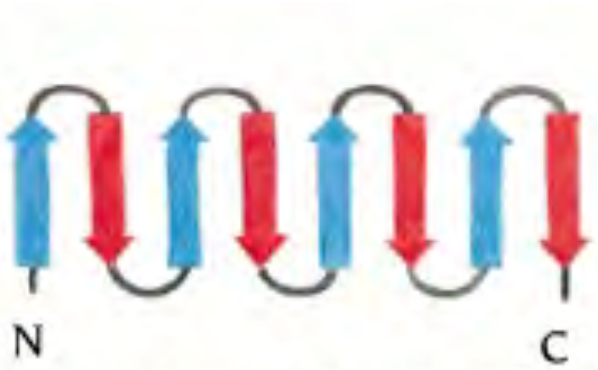
TIM barrels are built from overlapping β-α-β super-secondary structural motifs.

Alpha/Beta Structures: α/β barrel

The loop regions that connect the C-termini of the β-strands with the N-termini of the α-helices are variable and often form the enzyme active sites

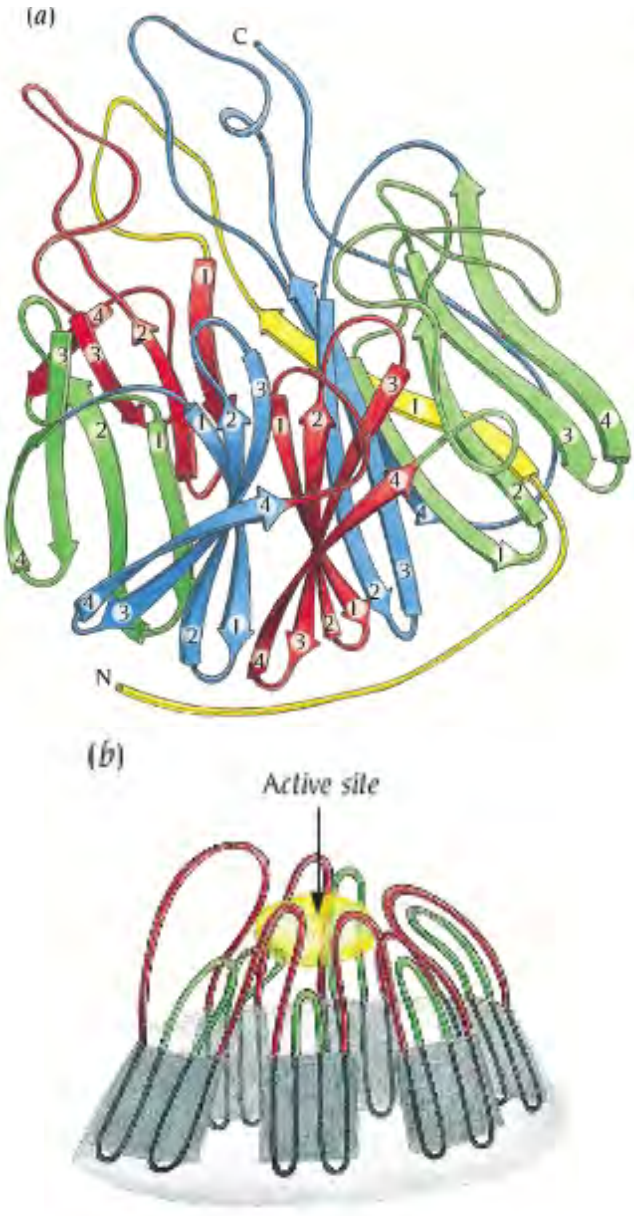
Beta Structures: up and down β-barrel





The up and down β-barrel is formed using the Greek key super-secondary structural element

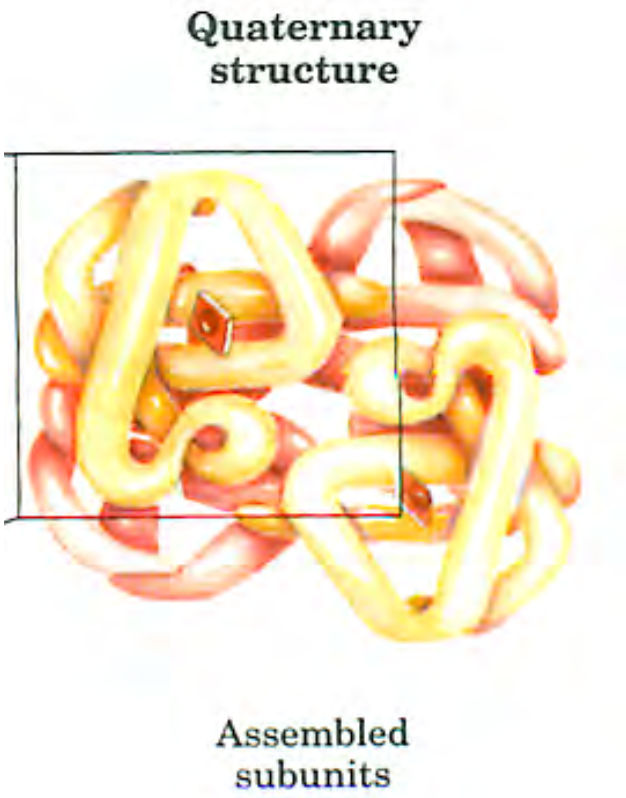
Beta Structures: up and down β-barrel



β-structured proteins are also common. The up and down β-barrel family is structurally and functionally diverse with cores comprised of 4 to 10 (or more) β-strands arranged in an antiparallel fashion.

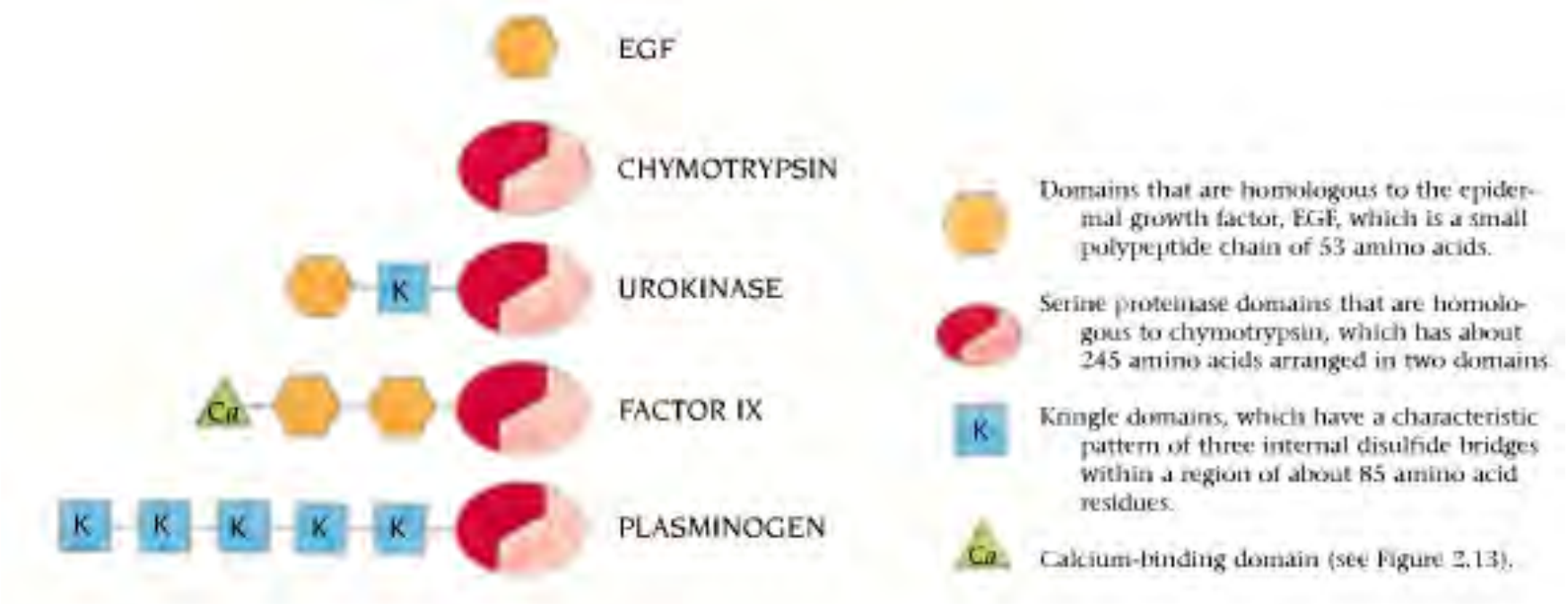
Usually two sheets pack against each other sometimes forming a barrel.The inside of the barrel is usually filled with hydrophobic residues. The loops connecting the β-strands are often variable in length and tend to form the active site of the molecule.

Domain structure



Domains are large units of tertiary structure (i.e. more complex than a motif) that are repeated either within a protein or in different proteins. They either have a structural or functional role – usually the latter

Domains

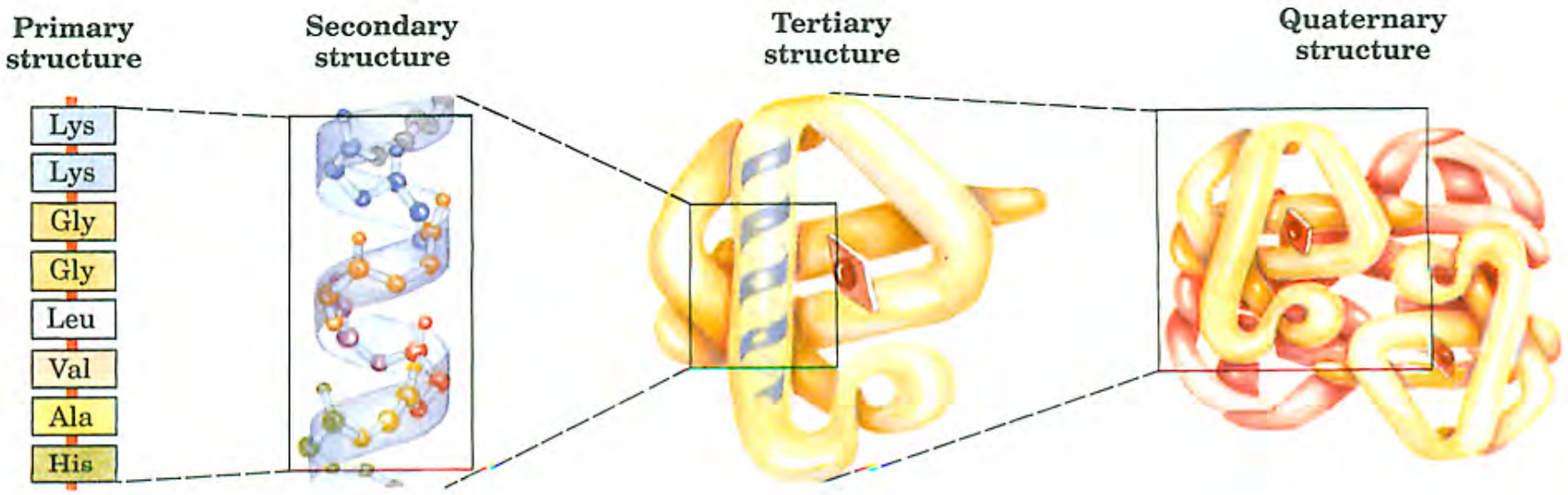


A conceptual diagram showing that proteins can be made by combining different functional motifs

Domains can often be isolated as separate functional units upon mild proteolytic treatment.

For example, the enzyme urokinase has three domains: a serine protease domain, a Kringle domain, and an epidermal growth factor domain.

The hierarchy of protein structure



Amino acids

α-helix

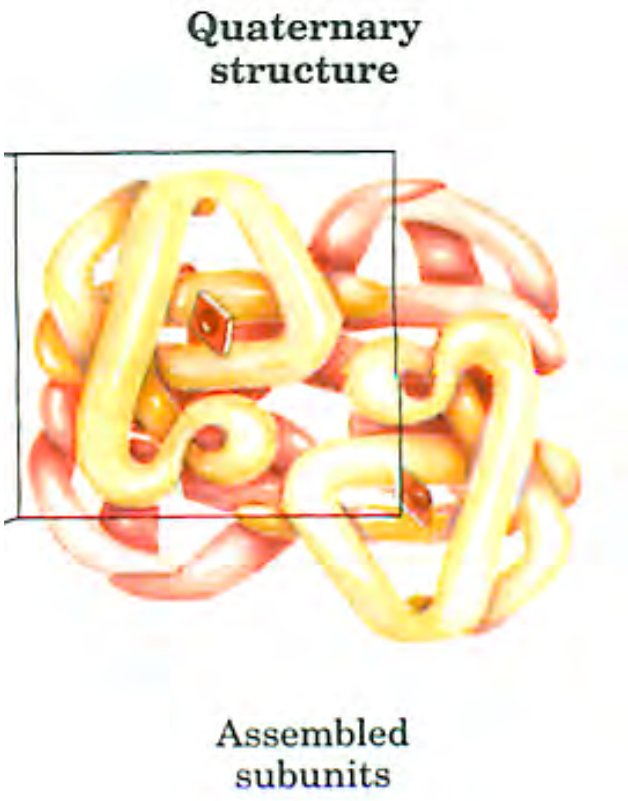
Super-secondary Structure/motif

Folded polypeptide chain

Domain

Assembled subunits

Quaternary structure

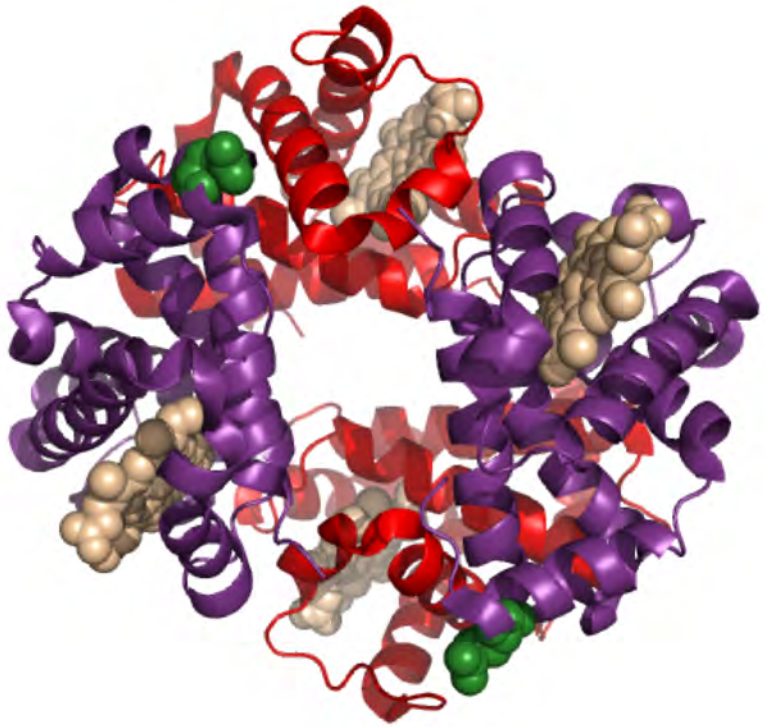


Proteins are often formed from more than one polypeptide chain, each forming a distinct tertiary structure.

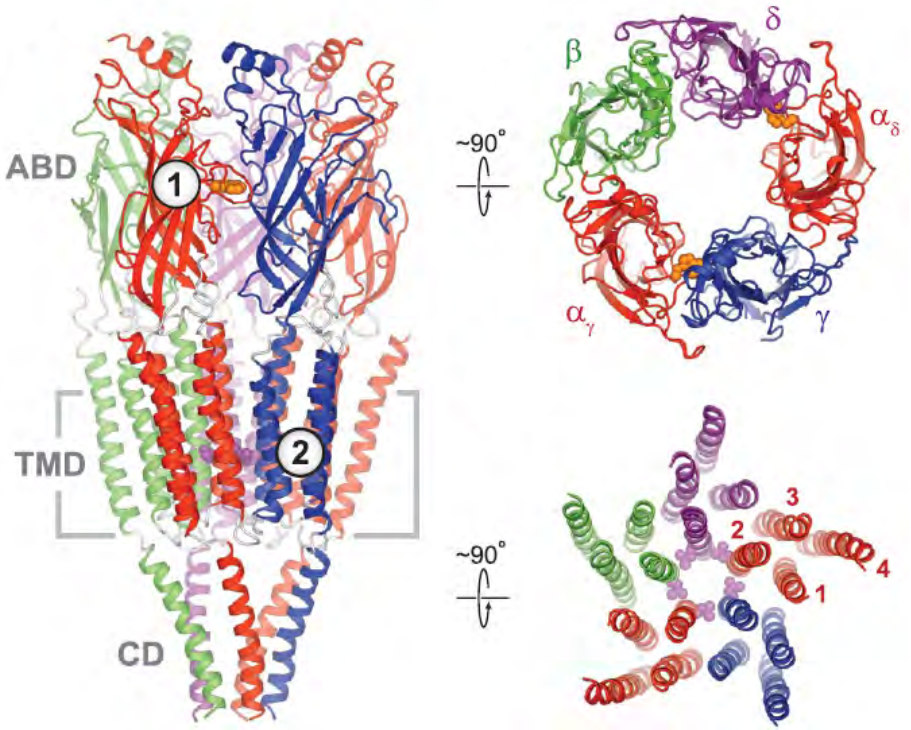
Quaternary structure refers to how distinct polypeptide chains or independent 3D structures pack together to give a protein complex.

The diagram to the left shows how 4 individual globin-folds pack together to form hemoglobin

Quaternary structure

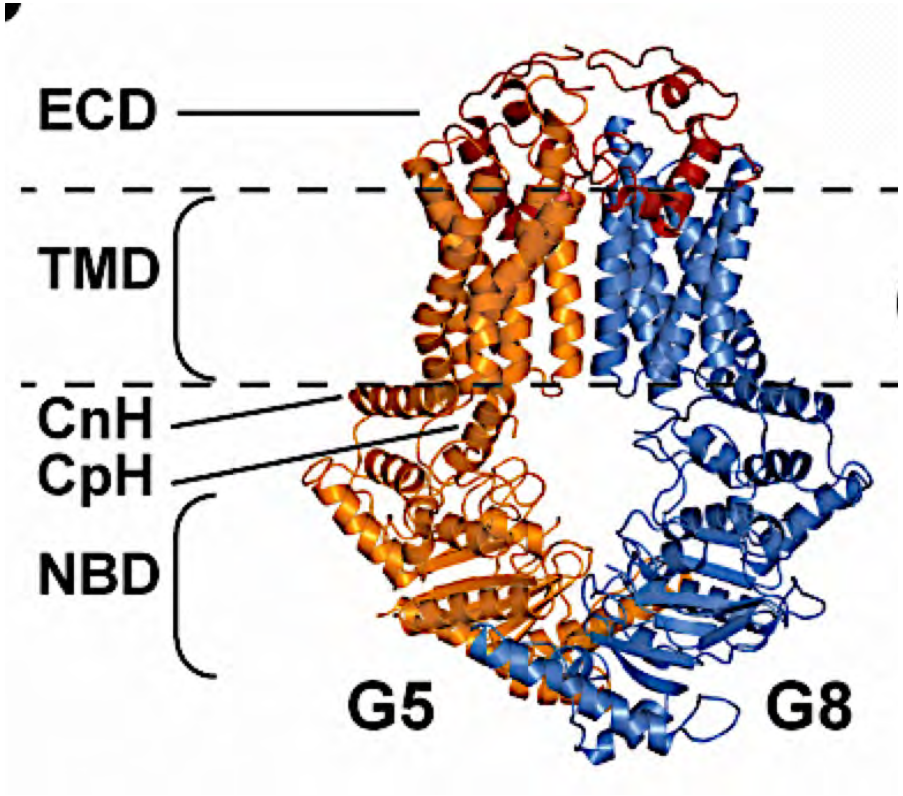


Hemoglobin, has a quaternary structure composed of two α (red) and two β (purple) subunits organized as a dimer of dimers.

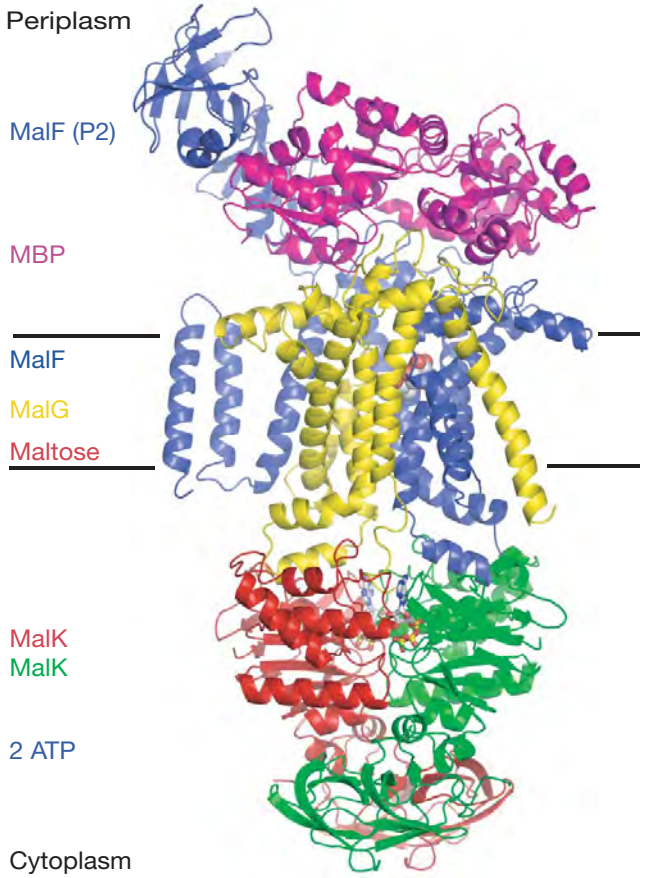


The muscle-type nicotinic acetylcholine receptor (nAChR), is composed of five subunits ($\alpha 2\beta\gamma\delta$) organized pseudo-symmetrically around a central ion channel pore.

Quaternary structure



A sterol/cholesterol transporter:
Human ABCG5/ABCG8 (a heterodimer)



A bacterial maltose transporter