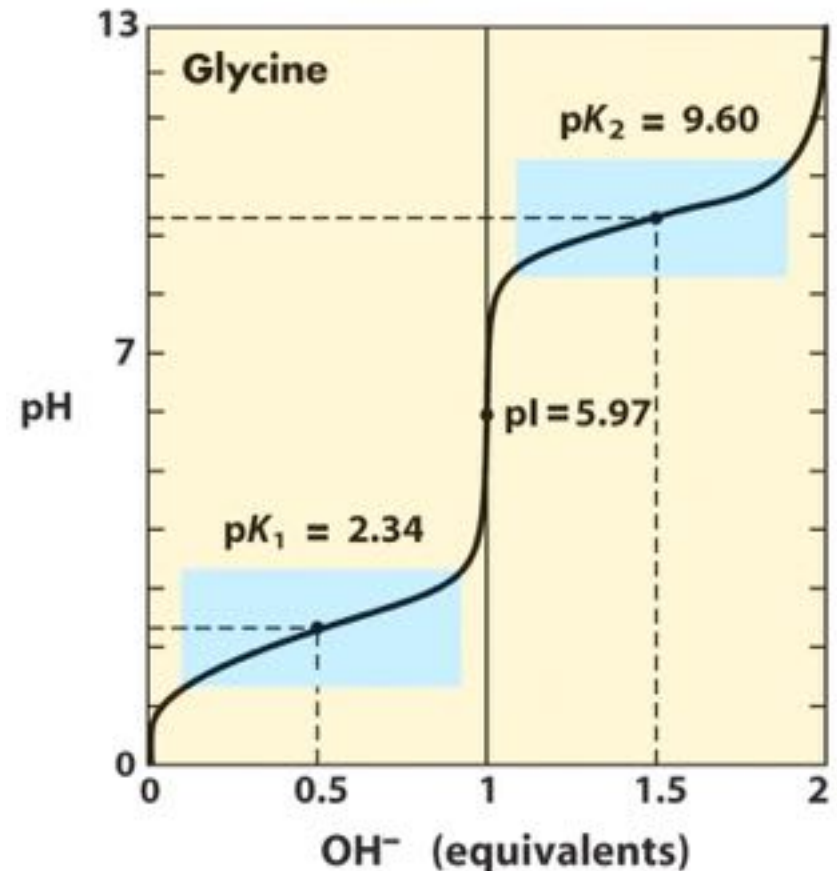
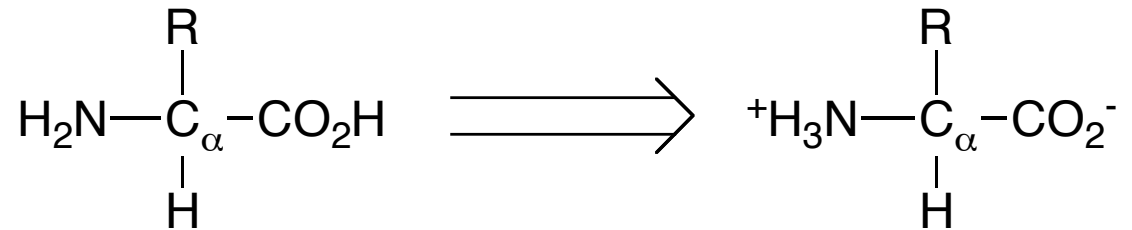
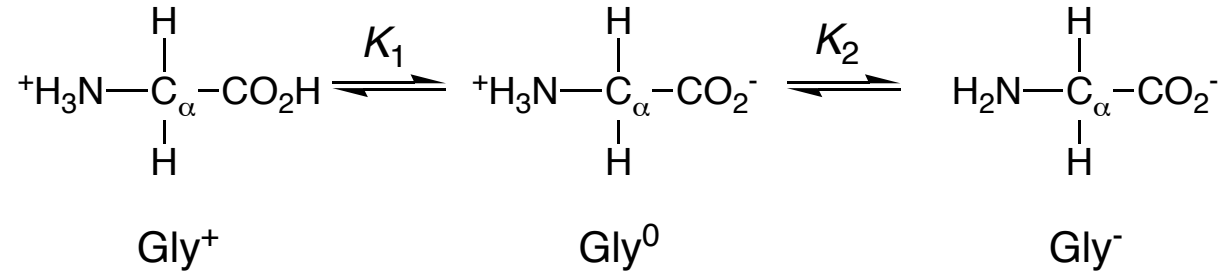


Amino Acids: General Properties

- pK_a of the α -carboxylic acid groups are near 2.2.*
- pK_a of the α -amino groups are near 9.4.*
- At physiological pH, both the carboxylic acid and amino groups are ionized, zwitterions.
- Amino acids can serve in acid or base capacities, they are amphoteric.



Amino Acids: General Properties

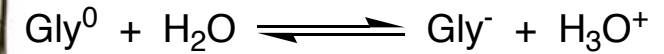
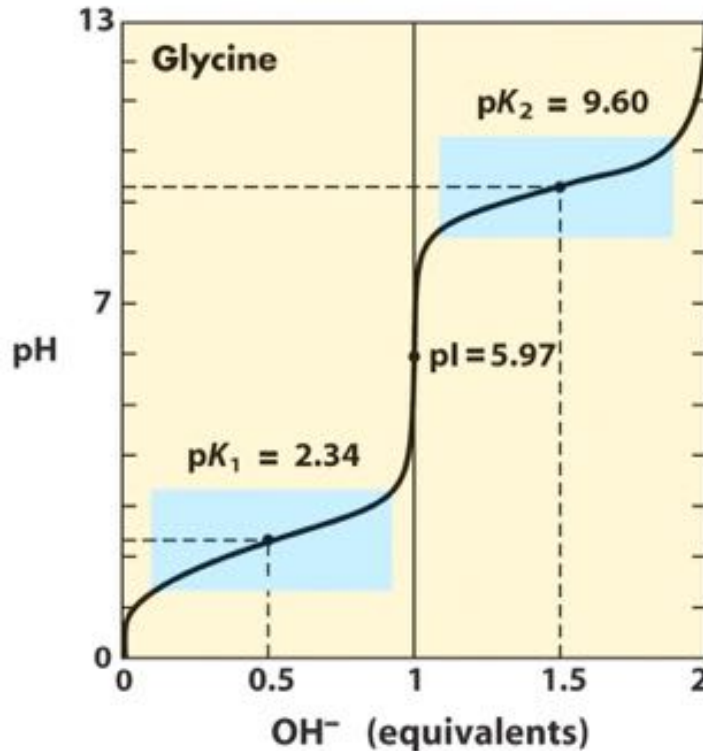


- pK_a of α-carboxylic acid is influenced by α-NH₃⁺. (pK_a of acetic acid = 4.76)

- pK_a of α-NH₃⁺ is influence by α-carboxylate. (pK_a of Gly-methyl ester = 7.75)

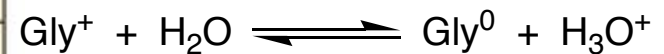
- Side chain functional groups see similar but weaker influence.

- Titration curves of proteins and peptides tend to be complicated, and rarely reflect individual pK_a values.



$$K_2 = \frac{[\text{Gly}^-][\text{H}_3\text{O}^+]}{[\text{Gly}^0]}$$

$$\text{p}K_2 = \sim 9.0-9.8$$



$$K_1 = \frac{[\text{Gly}^0][\text{H}_3\text{O}^+]}{[\text{Gly}^+]}$$

$$\text{p}K_1 = \sim 2.0-2.4$$

Lehninger: Principles of Biochemistry, 4th Ed.,
D. Nelson and M. Cox, 2005, p. 83.

$$\text{pH} = \text{p}K + \log\left(\frac{[\text{A}^-]}{[\text{HA}]}\right)$$

$$\text{pI} = 1/2(\text{p}K_1 + \text{p}K_2)$$

Protein Structure

Levels of Organization

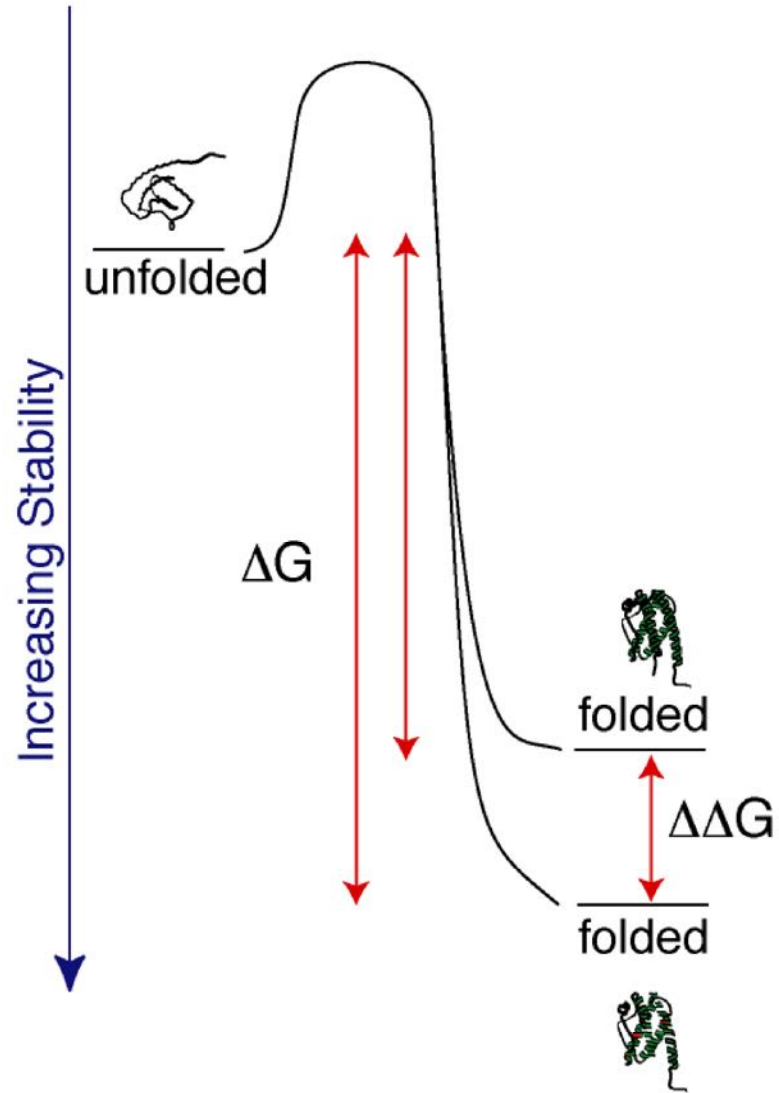
- **Primary structure (1° structure):** the amino acid sequence of polypeptide chain.
- **Secondary structure (2° structure):** local spatial organization and arrangement of the peptide backbone. Generally refers to easily localized structural elements (i.e. helices and sheets).
- **Tertiary structure (3° structure):** the comprehensive three-dimensional structure of a protein (single polypeptide chain).
- **Quaternary structure (4° structure):** assembly through noncovalent interactions) of a larger protein structure from 2 or more polypeptide chains (subunits), and the organization of these subunits.

Physical Interactions

- The unique properties of proteins is inextricably linked to the complex three-dimensional folded conformations they assume.
- The three-dimensional folded conformation is the result of many simultaneous noncovalent interactions between different parts of the protein and with the environment.
- These interactions are the result of a limited set of fundamental noncovalent forces.
- The complexity of water and an aqueous environment limits our understanding of proteins.

Protein Stability

- Native proteins are only marginally stable under physiological conditions.
- The stability of most folded proteins under physiological conditions is $\sim 20\text{-}60 \text{ kJ mol}^{-1}$.
- The three-dimensional folded conformation arises through a delicate balance of stabilizing and destabilizing forces.
- The observed stability of the folded protein is the result of a very small difference between very large but compensating factors.
- The enthalpic and entropic contributions vary similarly and compensate each other. This results in the free energy being relatively small difference between the two.

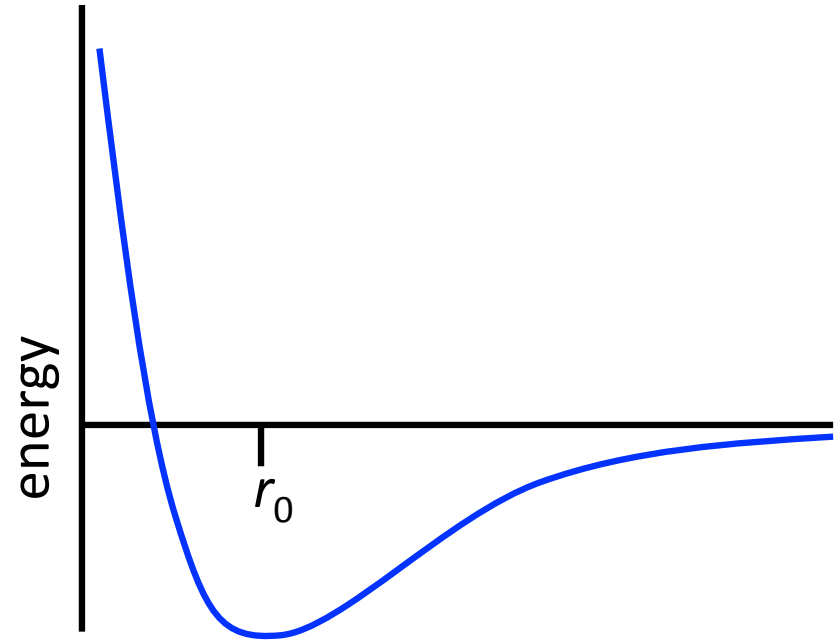


Forces Stabilizing Macromolecular Structure

- Noncovalent interactions are key biological forces:
 - ★ Electrostatic Forces:
 - ▶ Ionic interactions
 - ▶ Van der Waals
 - ▶ Hydrogen bonding
 - ★ Hydrophobic interactions
- These forces are transient in nature.
- Individually all are weak (C-C bond ~ 80 Kcal/mol), but they add up and collectively can be very strong.

Short-Range Repulsions

- Repulsion arises as molecules/atoms approach near enough for their respective electron orbitals to begin to overlap.
- Repulsion increases because electrons on different molecules/atoms cannot occupy the same space at the same time (increases exponentially with the inverse of distance).
- Because repulsion rises so steeply, it is possible to consider molecules/atoms to have definite dimensions with defined volumes (**van der Waals radius**).
- van der Waals radius is based on smallest distance that can exist between two nonbonded atoms in the crystalline state.



Electrostatic Forces: Point Charges

- All intermolecular forces are thought to be essentially electrostatic in origin.
- The most fundamental noncovalent interaction would therefore be the interaction between electrostatic charges.
- Coulomb's law describes the interaction between two point charges in a vacuum.
- It describes an interaction that is effective over relatively long distances.
- For other environments (such as in solution) the electrostatic interaction is modulated by other interactions.
- In homogenous environments, the electrostatic interaction is diminished by the dielectric constant of the medium.
- At short distances, molecules and atoms cannot be treated as point charges.
- Interactions between very close, oppositely charged groups in proteins usually involve not only electrostatic interactions, but also some degree of hydrogen bonding (**salt bridges**).

Coulomb's Law

$$F = \frac{k q_1 q_2}{\epsilon r^2}$$

F = force between two charges (1 & 2).

q_1 and q_2 = charges on 1 and 2.

k = proportionality constant ($8.99 \times 10^9 \text{ J}\cdot\text{m}\cdot\text{C}^{-2}$).

ϵ = Dielectric constant of environment.

r = Distance between interacting charges.

Representative Dielectric Constants

Vacuum $\epsilon = 1$

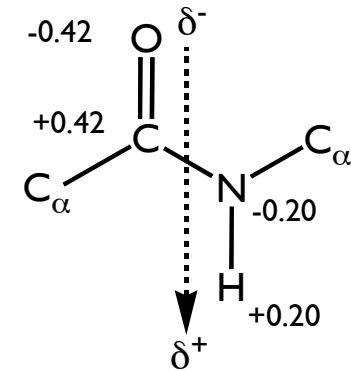
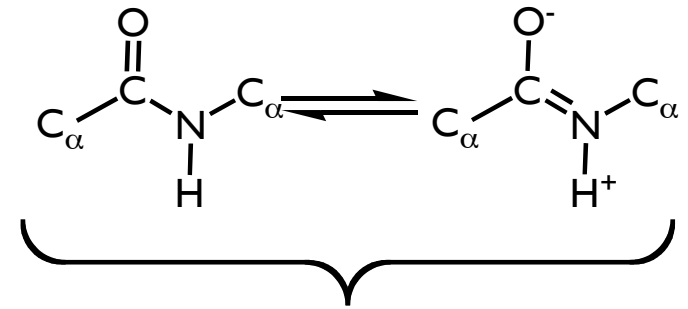
Water $\epsilon = 80$

Hexanes $\epsilon = 2$

Protein interior $\epsilon = 4$

Electrostatic Forces: van der Waals

- A molecule or functional group does not need to have a net charge to participate in electrostatic interactions.
- Electron densities can be localized if covalently linked atoms have different electronegativities.
- The separation of charge in a molecule determines its dipole moment (μ_D), corresponding to the magnitude of the separated charge (Z) and the distance (d) by which it is separated.
- The dipole moment has directionality as well as magnitude.
- The peptide bond which has partial double bond character exemplifies this polarization. The oxygen has a partial negative charge and the -NH- group a partial positive.
- Dipoles interact with point charges, other dipoles and more complex interactions.



$$\mu_D = Zd$$

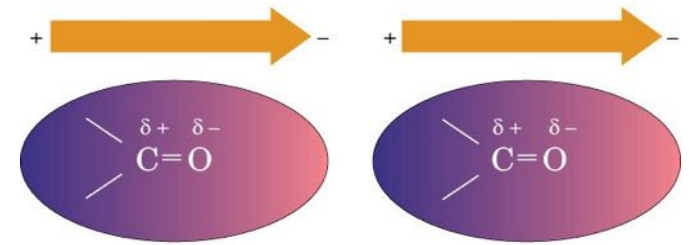
Dipole Moment

Electronegativities
of common atoms
in proteins:
O = 3.45
N = 2.98
C = 2.55
S = 2.53
H = 2.13

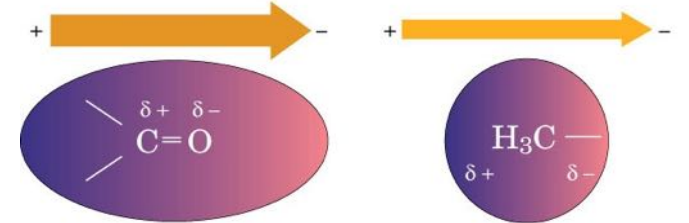
Electrostatic Forces: van der Waals

- Dipolar interactions are weaker than those between ionic groups. This is due to the fact that both attraction and repulsion occur between the two separated charges.
- The strength of dipolar interactions drops much more abruptly with distance than is the case with the interaction between ions (inversely with $\sim d^3$).
- Interactions involving dipoles also effects the dipole charge distribution within the interacting molecules. **Polarizability** describes the disposition of a molecule to have its electron (charge) distribution influenced by an applied electronic field.
- An induced dipole always interacts favorably with the inducing field. However this interaction is only half of that which would have occurred had the dipole already existed.

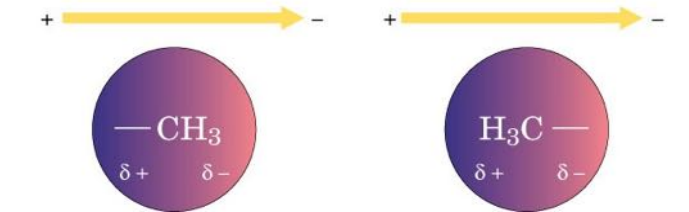
(a) Interactions between permanent dipoles



(b) Dipole-induced dipole interactions



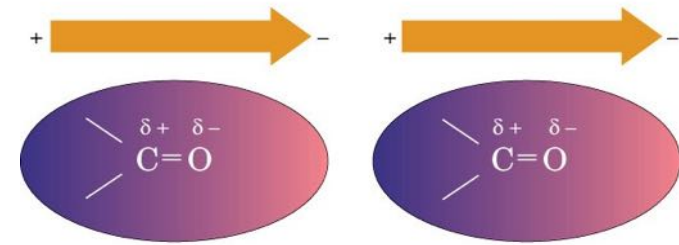
(c) London dispersion forces



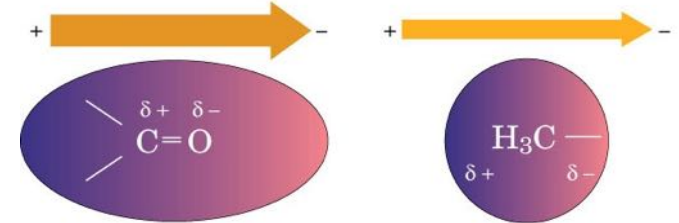
Electrostatic Forces: van der Waals

- All atoms and molecules attract each other even in the absence of charged groups as a result of mutual interactions and induced polarization.
- These are weak interactions only effective at short distances (varying with d^{-6}).
- Can arise from interaction between:
 - Two permanent dipoles.
 - A permanent dipole and an induced dipole.
 - Two induced dipoles (London dispersion forces).
- London dispersion forces: complex interaction. Essentially, an atom or group may have no net dipole, but may have a transient dipole resulting from temporary asymmetry in the distribution of electrons. This transient dipole can similarly polarize nearby neutral atom. (synchronization of electron flow and distribution in neighboring atoms and groups). Interaction becomes insignificant at distances greater than 50\AA .
- Van der Waals interactions are often represented by an energy potential as a function of distance (d).
- The optimal distance for the interaction of two atoms is usually $0.3\text{-}0.5\text{\AA}$ greater than their combined van der Waals radii.

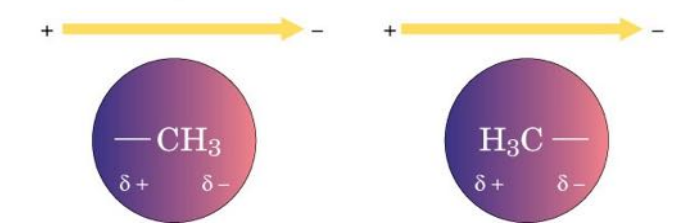
(a) Interactions between permanent dipoles



(b) Dipole-induced dipole interactions

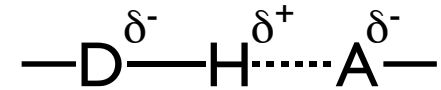


(c) London dispersion forces

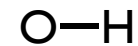


Electrostatic Forces: Hydrogen Bonding

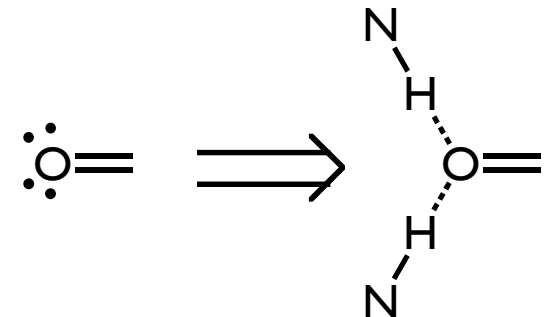
- A hydrogen bond occurs when two electronegative groups compete for the same hydrogen atom.
- In such interactions, the H atom is formally attached to the donor atom via a covalent bond and interacts favorably with the acceptor atom.
- The main component of the hydrogen bond is an electrostatic interaction between the dipole of the covalent D-H bond (H has δ^+) and the δ^- of the acceptor atom.
- The electrostatic and covalent elements of the hydrogen bond make it energetically favorable for the three participating atoms (D,H and A) to be collinear. This is the most common arrangement however some deviation from linearity is observed.
- The lengths and strengths of hydrogen bonds is dependent on the electronegativities of D and A. The greater their electronegativities are, the shorter and stronger the hydrogen bond.



Common
H-donors



Common
H-acceptors



Hydrophobic Interaction

- Electrostatic, hydrogen bonding and van der Waals interactions between two molecules in an aqueous environment are not particularly favorable due to competing interactions with surrounding water molecules.
- In the case of hydrophobic surfaces, interactions with surrounding water are not as favorable.
- The relative absence of favorable interactions with surrounding water molecules increases the favorability of interactions among nonpolar groups themselves (relative to other solvents).
- The preference of nonpolar molecules and groups for nonpolar environments is known as the hydrophobic interaction.

Hydrophobic Interactions

- The magnitude of the hydrophobic interaction is generally measured by the free energy of transfer (ΔG_{tr}) of a nonpolar molecule in the gas, liquid or solid state into water. (positive ΔG_{tr} value indicates that the nonpolar molecule prefers a nonaqueous environment)
- Transferring a solute molecule into a liquid involves:
 - Creating a suitable cavity in the liquid.
 - Introducing the solute molecule into the cavity.
 - Rearranging the solute and liquid molecules to maximize favorable interactions between them.
- The observed thermodynamics of this transition are the net effect of all these factors. Interpretation is not always straightforward.
- At room temperature, the unfavorable transfer of a nonpolar molecule from a nonpolar liquid to water is primarily a result of the unfavorable change in entropy ($\Delta H_{tr} \approx 0$).
- Water molecules cannot form hydrogen bonds with the nonpolar group. Therefore, they are generally believed to satisfy their hydrogen bond potential by forming a hydrogen bonded “iceberg” network among themselves at the nonpolar surface.

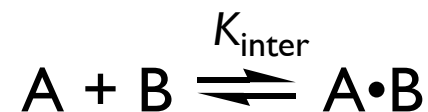
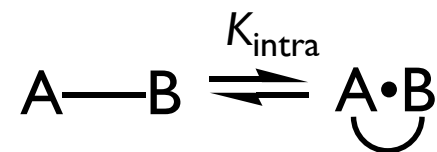
Hydrophobic Interactions:

to summarize

- Hydrophobic interactions do not result from repulsion between water molecules and nonpolar molecules and surfaces.
- While favorable interactions do occur between water molecules and nonpolar molecules and surfaces, the magnitude of these interactions are less than the favorable van der Waals interactions in a nonpolar environment and the hydrogen bonding in liquid water.
- Hydrophobic interaction results in nonpolar atoms, molecules and groups to interact with each other rather than with water.

Intramolecular Interactions

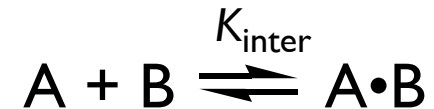
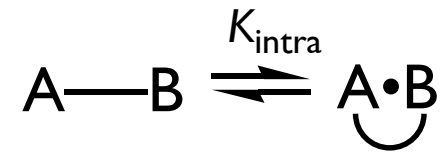
- For molecules to interact with each other they must lose entropy, which is energetically unfavorable.
- The magnitude of the loss in entropy is dependent on the degrees of freedom that become fixed as a result of the interaction.
- In the case of intramolecular interactions, the groups involved are incorporated within the same molecular scaffold, which automatically limits the number of degrees of freedom by fixing the relative distance and orientation of the groups involved.
- Intramolecular and bimolecular interactions can be compared by means of the ratio of their equilibrium constants (the **effective concentration**).



$$\frac{K_{\text{intra}}}{K_{\text{inter}}} = \text{effective concentration of } A-B$$

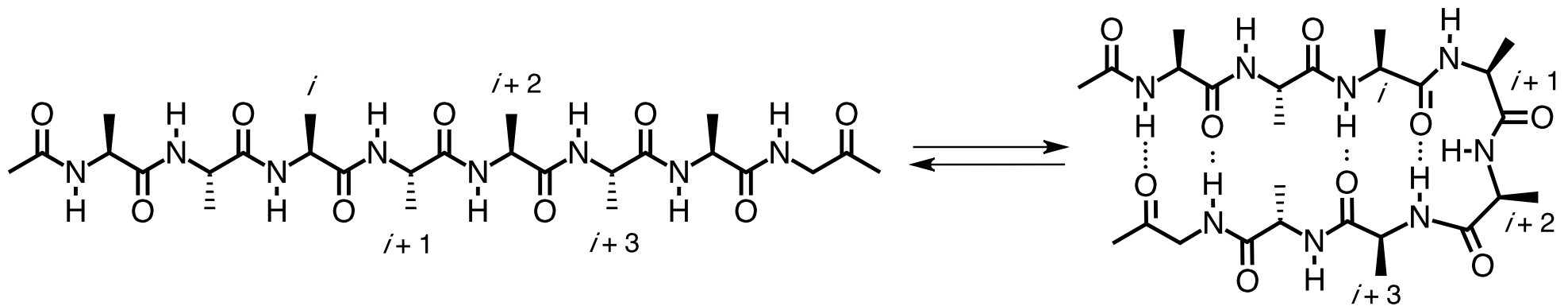
Intramolecular Interactions

- The maximum effective concentration of two groups in an aqueous solution was believed to be 55M, but much greater values for effective concentration are usually observed.
- Covalently linking the interacting moieties through a bond network results in their concentration relative to each other to be much higher than would be possible were the two groups on separate molecules.
- Interaction between the two groups results in the sacrifice of some fraction of the internal flexibility and conformational freedom of the molecule.
- When there is no entropic difference between molecules with and without interaction between the groups, their effective concentration is at its maximum value.



$$\frac{K_{\text{intra}}}{K_{\text{inter}}} = \text{effective concentration of } A-B$$

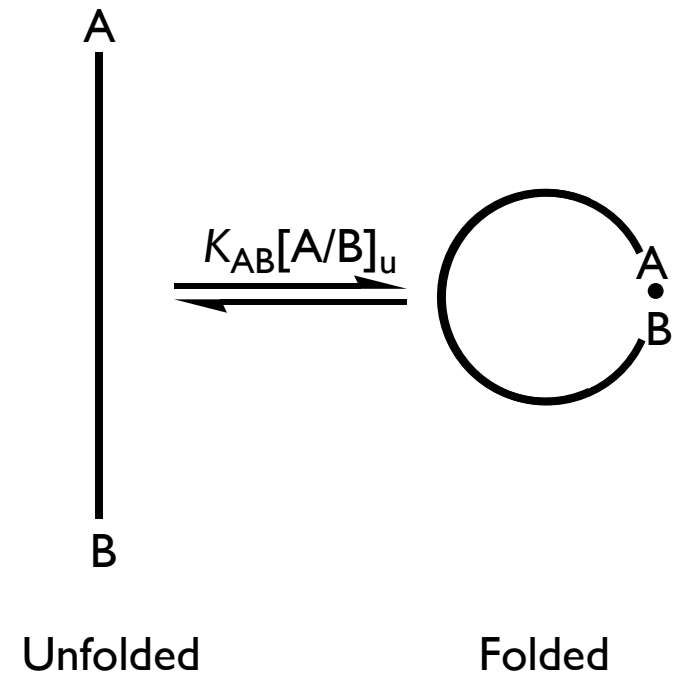
Cooperativity of Multiple Interactions in the Folding of a β Hairpin.



- Formation of a β hairpin is a cooperative process.
- Result of multiple intramolecular hydrogen bonds.

Cooperativity of Multiple Interactions

- Multiple groups within a single molecule can behave differently from the same groups individually in solution.
- The simultaneous presence of multiple interactions within a single molecule results in cooperativity between them. Collectively, these interactions can be much stronger than expected based on their individual strengths.
- **Cooperativity is critical for proteins.**
- Single interactions between groups within a polypeptide chain are not expected to be stable unless these groups lie in close proximity of each other within the covalent structure (resulting in a high effective concentration).
- Due to the size and conformational flexibility of the unfolded protein, groups attached to a moderate sized peptide have effective concentrations in the range of 10^{-2} - 10^{-5} M (depending on proximity).
- Expected values for $K_{obs,u}$ (observed equilibrium constant) for individual hydrogen bonds, salt bridges... etc range from 4×10^{-3} to 10^{-7} M.



$$K_{obs,U} = K_{AB}[A/B]_u$$

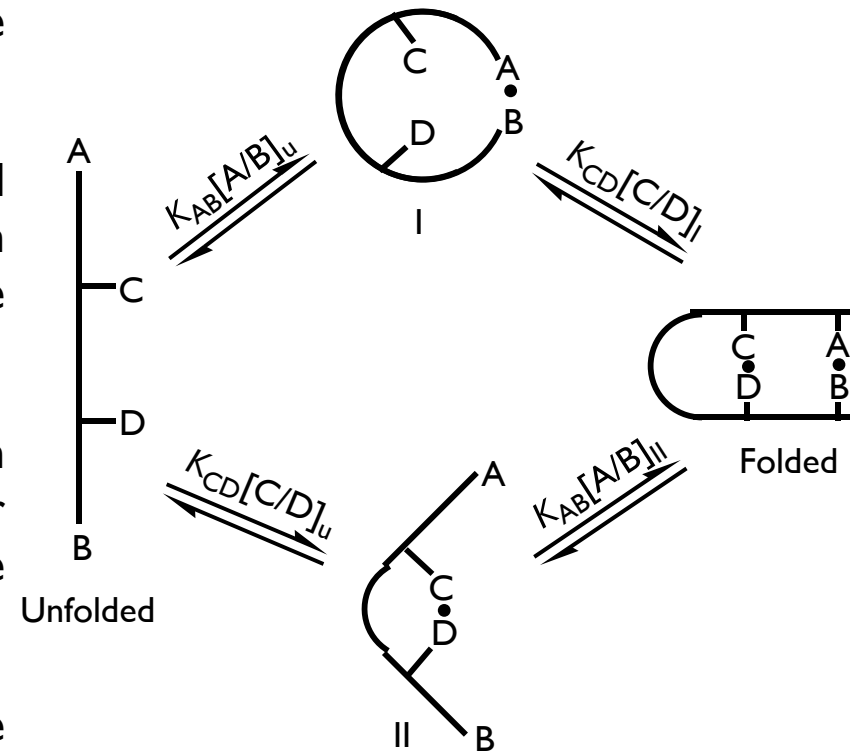
K_{AB} = association constant for free A and B

$[A/B]_u$ = effective concentration of A and B in unfolded peptide

Cooperativity of Multiple Interactions

- Multiple interactions among two or more pairs of groups within the same molecule often do not behave independently, but assist or interfere with each other.
- In the example to the right: if both interactions $A \cdot B$ and $C \cdot D$ are possible simultaneously, the interactions between one pair of groups constrains the peptide, increasing the effective concentration of the other pair.
- This will proceed in a mutual manner, with both interactions having the same effect on each other (factor $Coop$ is the degree of cooperativity between the interactions).
- Each interaction is more stable in the presence of the other than in its absence.
- In polypeptides containing additional groups that interact simultaneously, equilibria are extended:

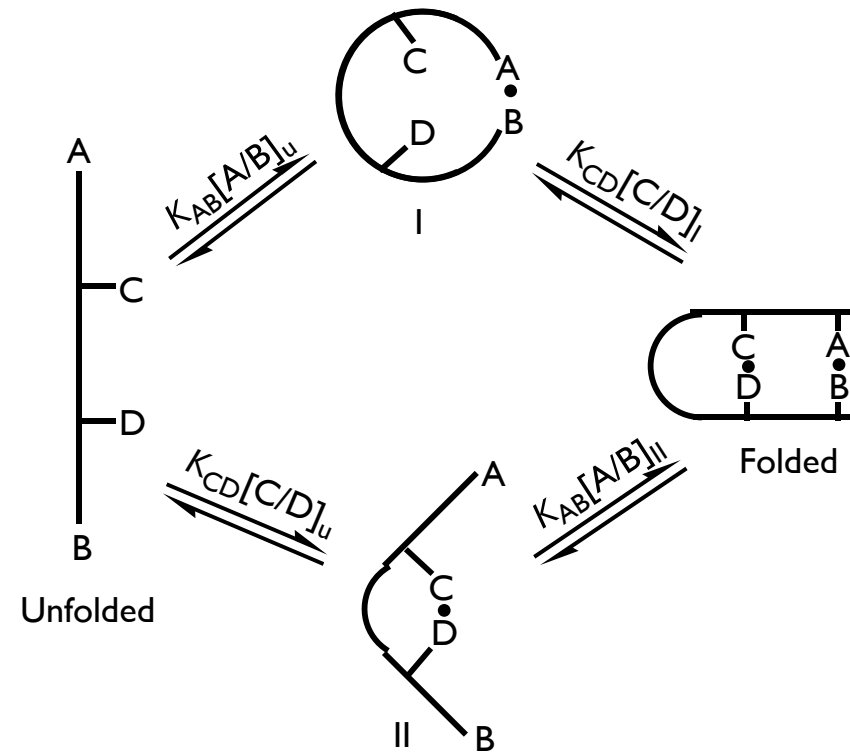
$$K_{net} = (K_{AB}[A/B]_u)(K_{CD}[C/D]_I)(K_{EF}[E/F]_{III})(K_{GH}[G/H]_{IV}) \dots$$
- The value of K_{net} is pathway independent.
- The final folded conformation is stable-only if the value of K_{net} is greater than unity.



$$\frac{[A/B]_{II}}{[A/B]_u} = \frac{[C/D]_I}{[C/D]_u} = Coop$$

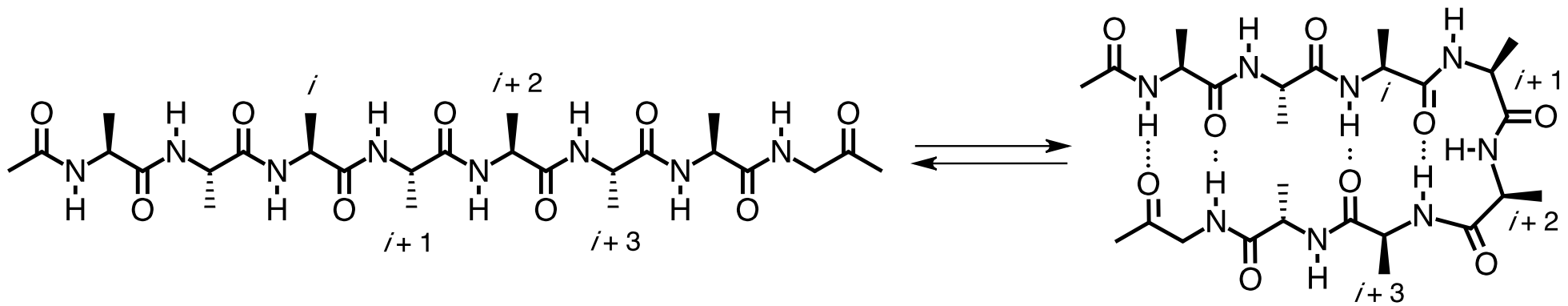
Cooperativity of Multiple Interactions

- In considering a series of weak interactions, the first will be very weak, with an equilibrium constant of 10^{-3} - 10^{-7} .
- The first interaction increases the effective concentration of the next pair of interacting groups, resulting in a slightly larger equilibrium constant for the interaction (by the factor Coop). If the equilibrium constant of the second interaction is less than unity, the product of the two equilibrium constants is lower than that of the first.
- *The net stabilities of conformations with additional weak interactions are even lower than that of the conformation with a single interaction.*
- The process continues until the effective concentrations of additional interacting groups are sufficiently high to make the equilibrium constant for each additional interaction greater than unity, with K_{net} increasing with each additional interaction.
- In this manner, a sufficient number of simultaneous weak interactions can make the value of K_{net} greater than unity and provide a stable folded structure.



$$\frac{[A/B]_{II}}{[A/B]_u} = \frac{[C/D]_I}{[C/D]_u} = \text{Coop}$$

Cooperativity Beyond Effective Concentration



- As weak interactions accumulate, structure becomes less flexible.
- Geometries of interactions involving dipoles and hydrogen bonds on average improve... and are thus strengthened.
- *...not only does the existing network significantly strengthen a new hydrogen bond, but the new hydrogen bond also strengthens the existing hydrogen bonds in the network.*
- Has implications for ligand binding: binding of a ligand is likely to have long-range effects... particularly affects interactions involving neighboring groups.
- Mutations usually cause minor rearrangements in protein structure that propagate throughout the protein.

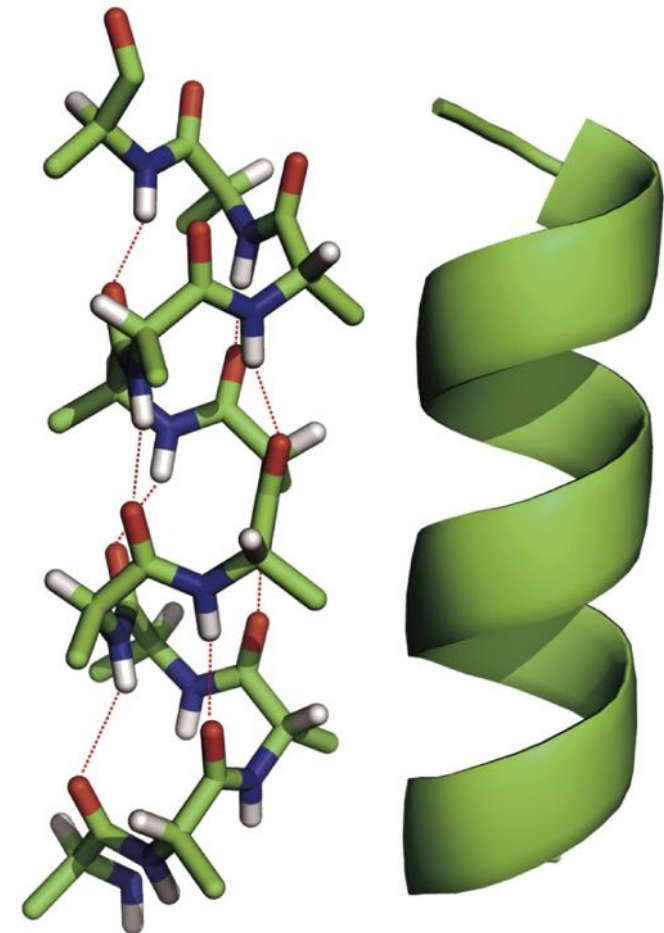


Figure 1.15 How Proteins Work (©2012 Garland Science)

Essential Water

- Water is essential for protein function, but may require only a few key water molecules to achieve activity.
- Enzymes may require only a few water molecules to hydrate essential buried water sites within the protein for onset of function.
- Full function may be achieved with less than a single layer of water molecules surrounding the protein.
- Essential water is required to maintain correct structure and polarity of enzyme active sites...
- Water molecules in hydration layer are less mobile than bulk water.
- Hydration layer may be one to two water molecules thick
- Water molecules outside of this layer of interacting water molecules may be treated as bulk water.

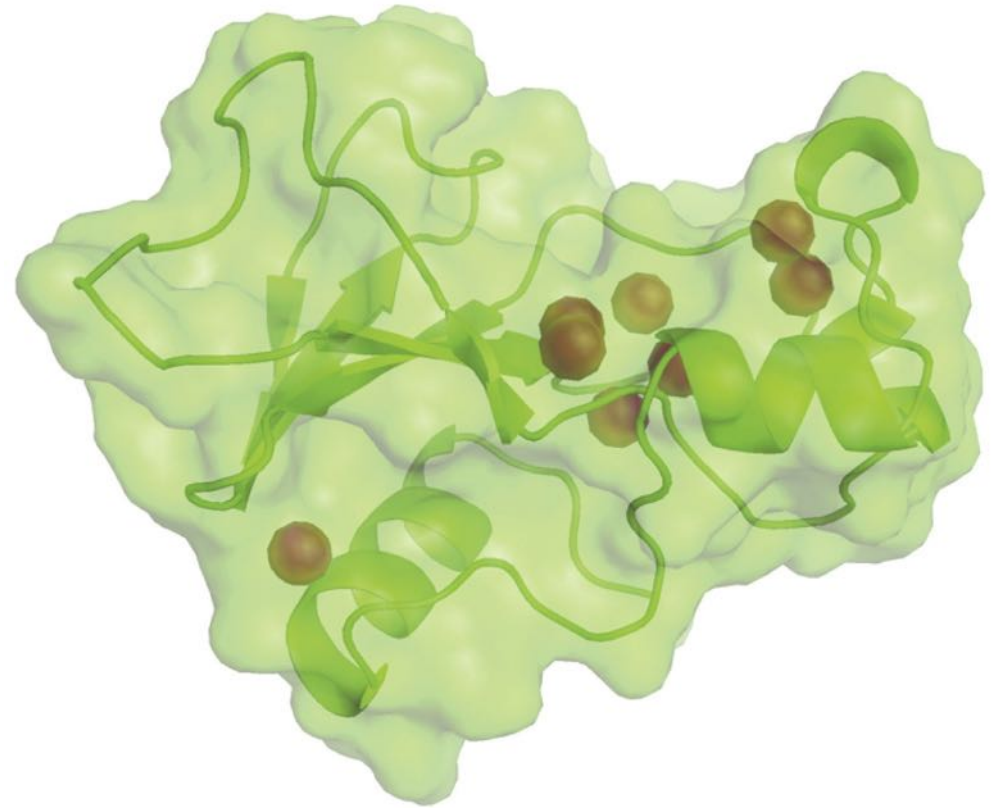


Figure 1.27 How Proteins Work (©2012 Garland Science)

Enthalpy, Entropy and Protein Folding

- Entropy/enthalpy compensation: observation that many protein folding and ligand binding interactions are associated with very large changes in entropy and enthalpy that mostly cancel out resulting in small changes in free energy.
- Historically, associated with release of a large number of water molecules and the enthalpy and entropy associated with the water molecules.
- Alternative view suggests that in the formation of protein-protein interfaces (would apply to folding as well) several strong protein-water interactions are replaced by a network of weaker interactions (such as hydrogen bonds).
- In the latter scenario, the loss of enthalpy that arises from the breaking the strong protein-water interactions is offset by an increase in entropy associated with rearrangement of hydrogen bonding network.
- *At least some and possibly most of the thermodynamic changes are due to the increased cooperativity that is characteristic of larger systems.*

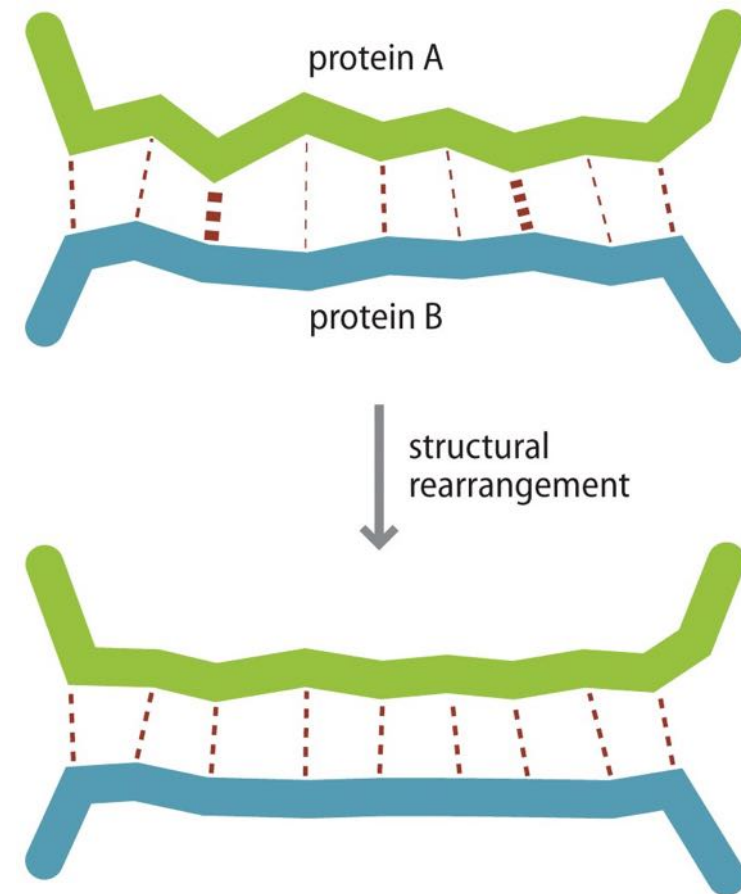


Figure 1.30 How Proteins Work (©2012 Garland Science)

Levels of Organization

- **Primary structure (1° structure):** the amino acid sequence of polypeptide chain.
- **Secondary structure (2° structure):** local spatial organization and arrangement of the peptide backbone. Generally refers to easily localized structural elements (i.e. helices and sheets).
- **Tertiary structure (3° structure):** the comprehensive three-dimensional structure of a protein (single polypeptide chain).
- **Quaternary structure (4° structure):** assembly through noncovalent interactions) of a larger protein structure from 2 or more polypeptide chains (subunits), and the organization of these subunits.

Secondary Structure

- Three basic units of secondary structure:
 - β -sheets
 - α -helices
 - turns
- These are the main secondary structural elements in globular proteins. (other secondary structures can be formed)

β -Sheet

- The most populated region of backbone conformational space is the **β -sheet** region.
- The β -sheet is characterized by peptide chains in extended conformations with a repeating pattern of ϕ and ψ angles (approx. -130° and $+125^\circ$ respectively).
- Extended conformation of an isolated chain is not stable. β -strand is only stable when incorporated within a β -sheet.
- In a β -sheet, hydrogen bonds formed between backbone amide C=O of one strand with amide -NH of an adjacent strand - with near ideal geometry for hydrogen bonds.
- Two flavors of β -sheet:
 - ◆ Antiparallel β -sheet \rightarrow H-bonded β -strands run in opposite directions.
 - ◆ Parallel β -sheet \rightarrow H-bonded β -strands run in the same direction.
 - ◆ β -strands can also combine in mixed β -sheets - strong bias against mixed β -sheets.

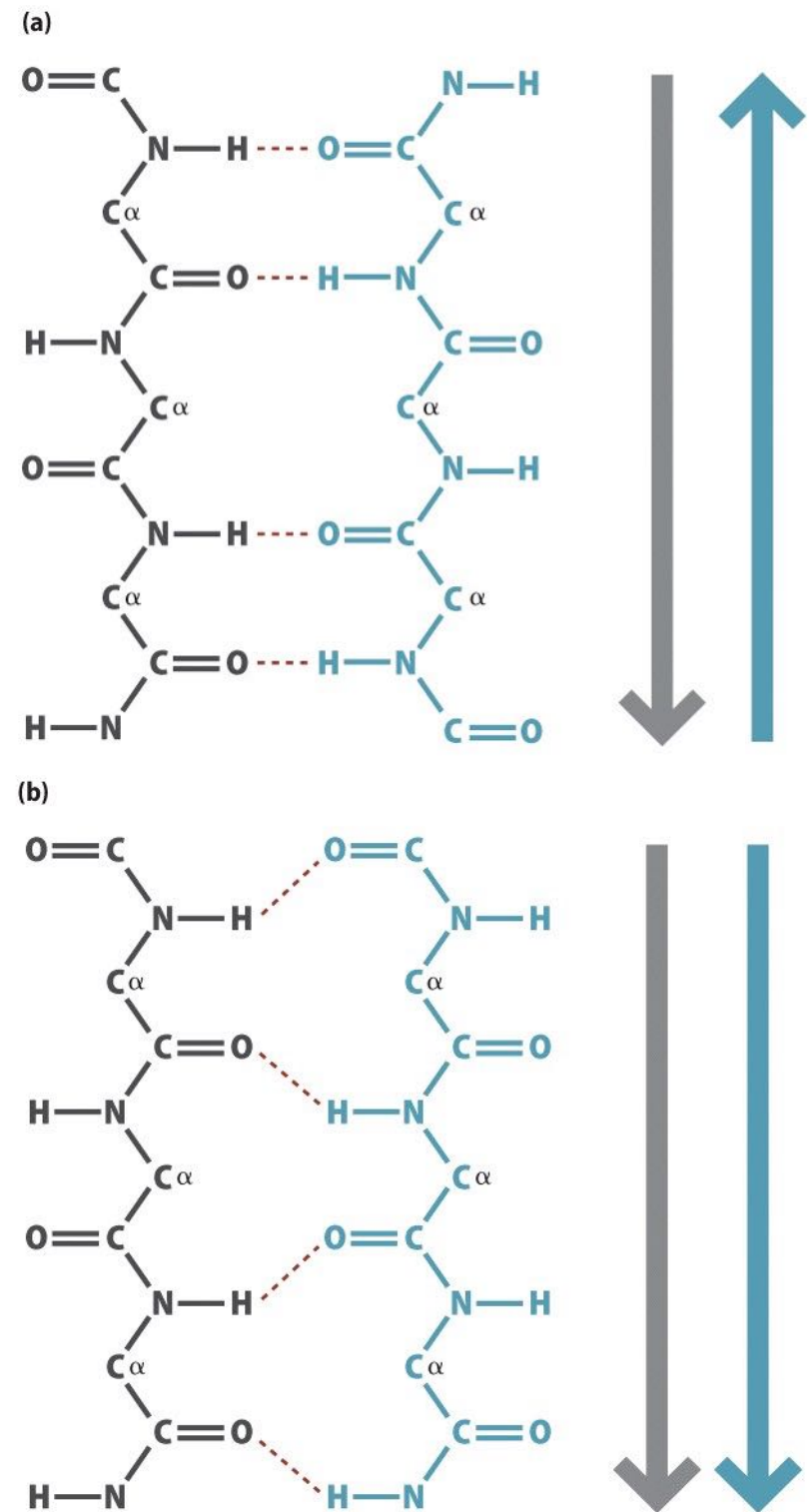
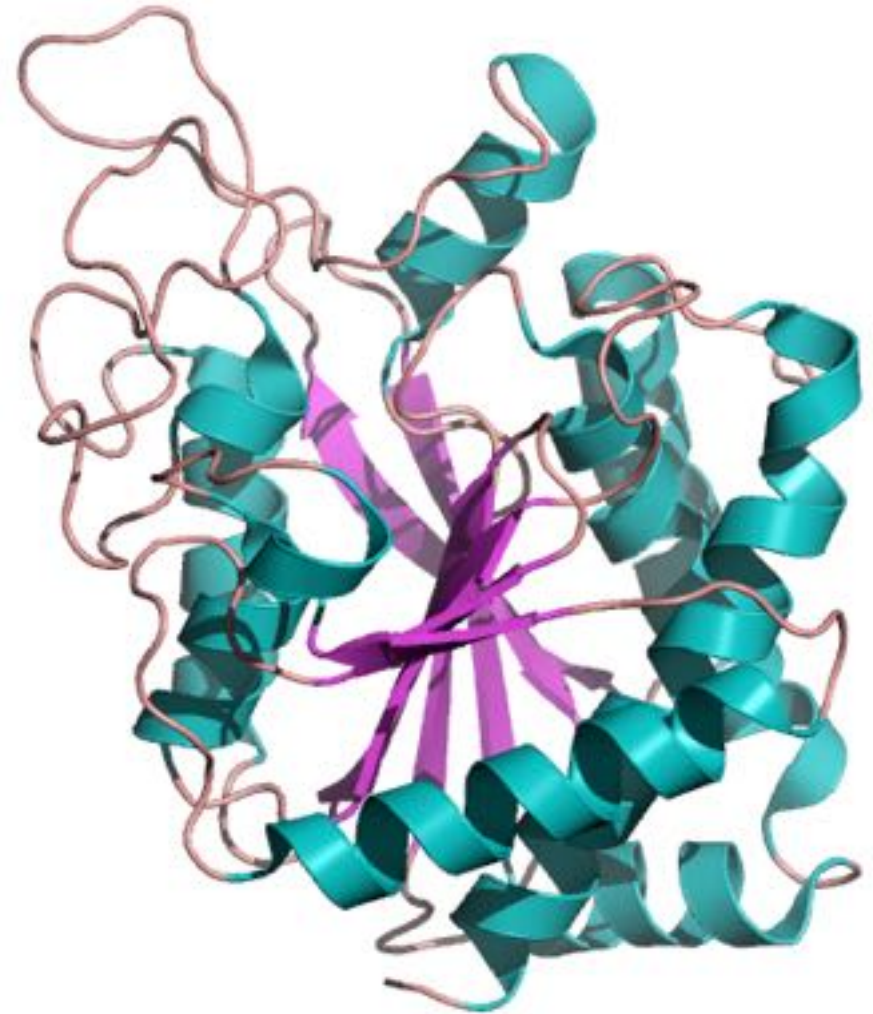


Figure 1.10 How Proteins Work (©2012 Garland Science)

Secondary Structure

- Most β -sheets in globular proteins are twisted rather than planar – with a right-handed twist of 0° - 30° between strands. Likely due inherent chirality of the amino acids and non bonding interactions.
- The conformational parameters of the peptide backbone can also deviate from ideality. More positive ϕ and ψ values are generally observed in twisted sheets.
- Further distortions are also observed in mixed β -sheets because of differences in the backbone conformations of parallel and antiparallel β -sheets.
- Isolated β -sheets have a propensity to aggregate and grow indefinitely from the edges. Therefore, there is no ideal model for isolated β -sheets.



The α -Helix

- The other major region is the α -helical region.
- The right handed α -helix is the best known and most recognizable of the polypeptide regular structures.
- The α -helix combines favorable conformational angles, van der Waals interactions and backbone hydrogen bonding.
 - ◆ $\phi = -57^\circ$ (-62°) and $\psi = -47^\circ$ (-41°).
 - ◆ 3.6 residues/turn with pitch of 5.4\AA .
- H-bonds between N-H group (donor) of n^{th} residue and the C=O group (acceptor) of the $n-4^{\text{th}}$ residue .
- Core of the helix is tightly packed-van der Waals contacts.
- Side chains are directed outward and slightly backwards (towards N-terminus). [restrictions on side chain conformations]
- The detailed geometry of the α -helix is found to vary somewhat in folded proteins.
- Slightly different geometry is adopted by natural proteins with $\phi = -62^\circ$, $\psi = -41^\circ$ and H-bonds directed slightly out-away from helix (believed more favorable than classic conformation).

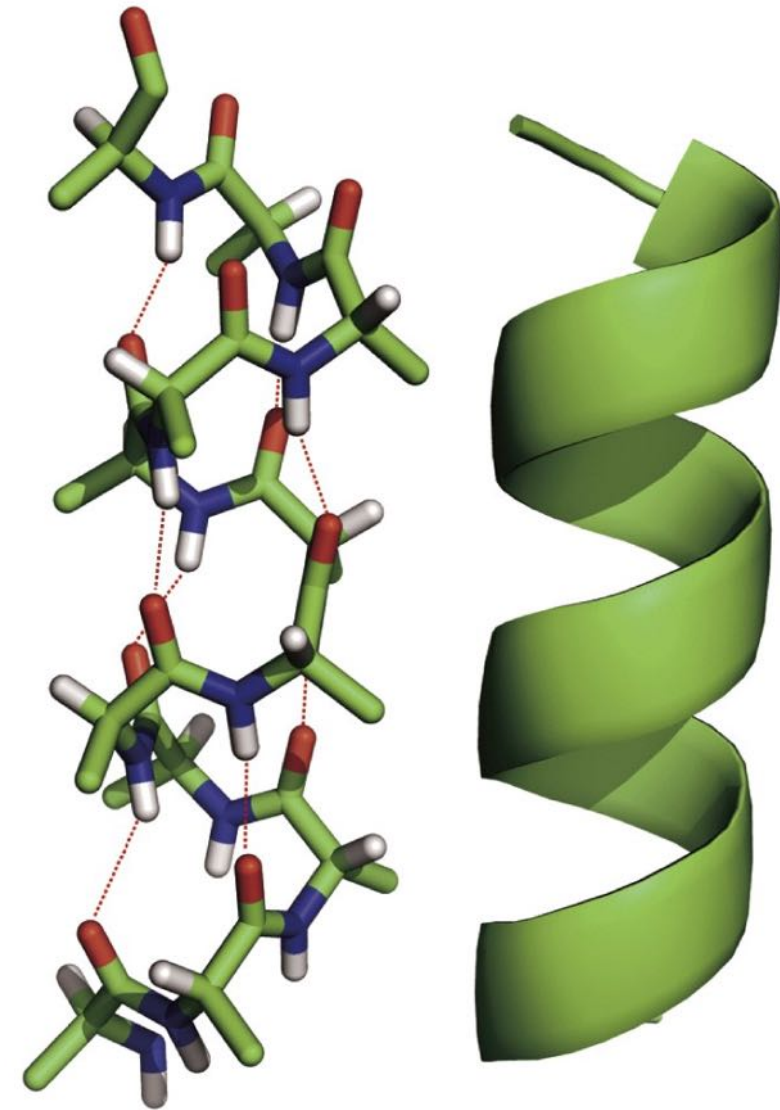


Figure 1.15 How Proteins Work (©2012 Garland Science)

The α -Helix

- All backbone hydrogen bonds and peptide groups point in the same direction in the α -helix.
- Alignment of hydrogen bonds results in helices having a net dipole with the N- and C-termini having partial positive and negative charges (respectively) ~ 0.5 - 0.7 unit charge at each end.
- Frequently negatively-charged groups/species bind at N-terminus of helix, but positively-charged groups only rarely bind at C-terminus of helix.
- Polarization of hydrogen bonding may increase the dipole moment of each peptide bond as much as 50%.



Figure 1.39 How Proteins Work (©2012 Garland Science)

The α -Helix

- Based on theoretical, experimental and statistical studies, amino acids have different tendencies to form α -helices.
 - ◆ **Ala, Glu, Leu and Met are “good” α -helix formers.**
 - ◆ **Pro and Gly are not common in α -helices**
- The distribution of amino acids in the helix reflects whether the helix is completely or partially buried in the protein interior or completely exposed to solvent.
- α -helices are usually at least partially exposed to solvent and are therefore amphipathic, with nonpolar side chains presented predominantly along one side of the helical cylinder and polar residues along the remainder of the surface. (these helices possess significant hydrophobic moments)
- Transmembrane helices tend to be rich in amino acids with hydrophobic side chains - allows for prediction of transmembrane helices based on amino acid sequence.
- A **helical wheel** is used to illustrate amino acid distribution along the helical cylinder (positions a-g).

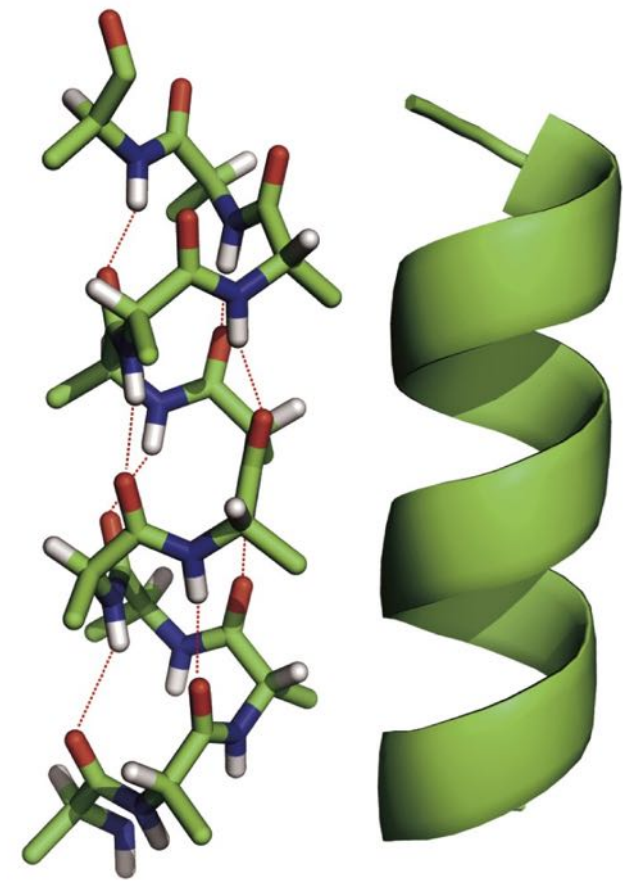
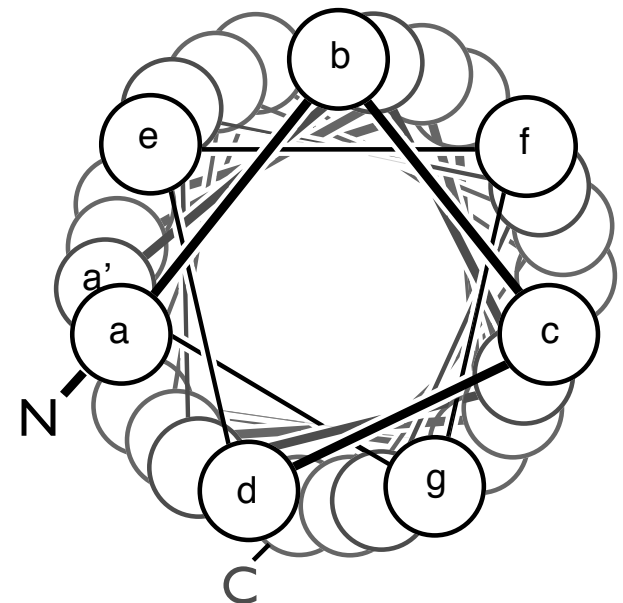


Figure 1.15 How Proteins Work (©2012 Garland Science)



Other Regular Helical Conformations

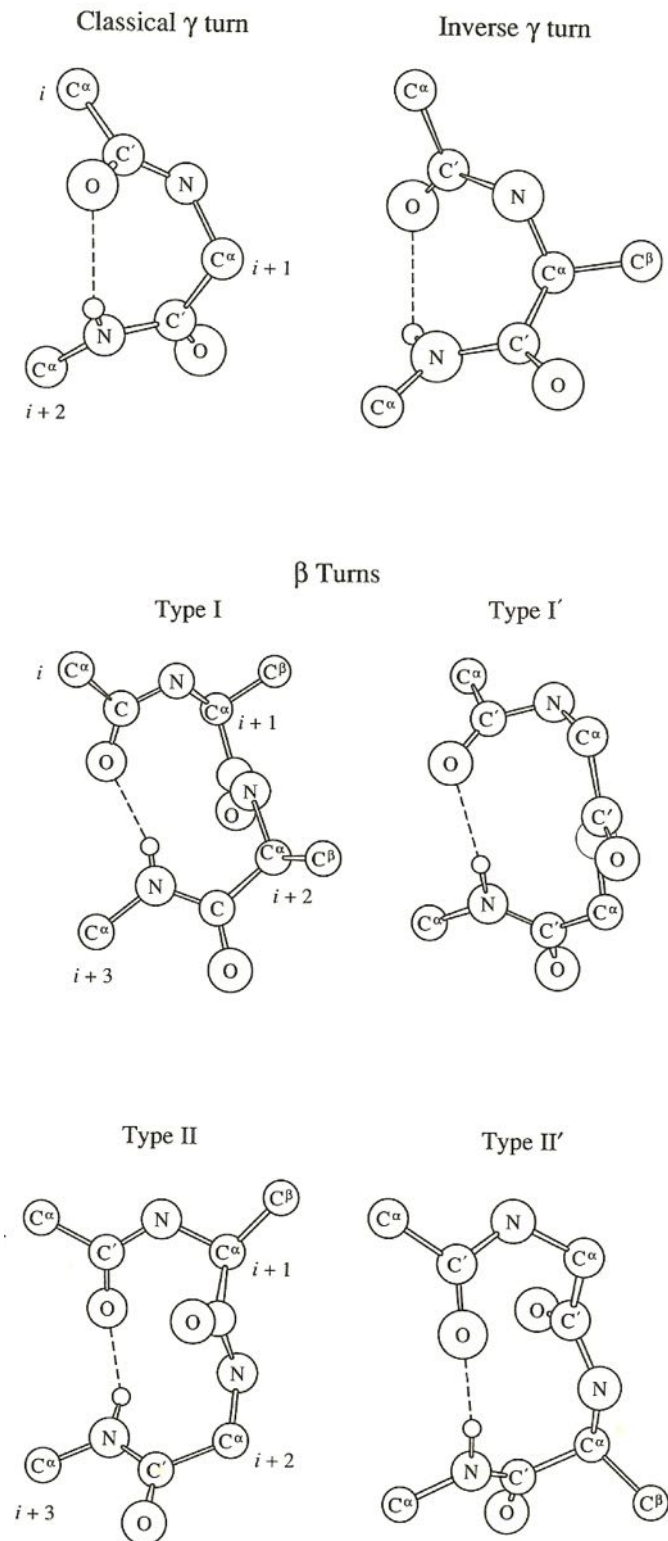
- Less prevalent helical conformations than the α -helix.
- Described using n_m notation:
 - n = number of residues per turn.
 - m = number of atoms (including H) in the ring formed by the backbone hydrogen bond.
 - An α -helix would be a 3.6_{13} helix.
- The **3_{10} helix**: right handed helix with pitch = 6.0\AA (required torsion angles slightly in forbidden range). Packing of backbone atoms is somewhat tight, and hydrogen bonds are nonlinear. Usually observed only for short segments in proteins.
- The **π -helix (4.4_{16} helix)**: wide-flat conformation results in an axial “hole”. A mildly forbidden conformation. In most cases found in short segments (a few residues), within larger helices. Soybean lipoxygenase contains a 43-residue π -helix (~ 3 turns of the helix).
- Polyproline helices (Pro I and Pro II): two helical conformations differ in peptide bond geometry (Pro I = all *cis* and Pro II = all *trans*).
 - ◆ Pro I is a right-handed helix with 3.3 residues per turn.
 - ◆ Pro II is a left-handed helix with 3.0 residues per turn.
 - ◆ Pro I vs Pro II depends on solvent and environment and can interconvert (slow).

Structural Propensities

Amino Acid	P_{α}	Helix Classification	P_{β}	Sheet Classification
Ala	1.42	H_{α}	0.83	i_{β}
Arg	0.98	i_{α}	0.93	i_{β}
Asn	0.67	b_{α}	0.89	i_{β}
Asp	1.01	I_{α}	0.54	B_{β}
Cys	0.70	i_{α}	1.19	h_{β}
Gln	1.11	h_{α}	1.10	h_{β}
Glu	1.51	H_{α}	0.37	B_{β}
Gly	0.57	B_{α}	0.75	b_{β}
His	1.00	I_{α}	0.87	h_{β}
Ile	1.08	h_{α}	1.60	H_{β}
Leu	1.21	H_{α}	1.30	h_{β}
Lys	1.16	h_{α}	0.74	b_{β}
Met	1.45	H_{α}	1.05	h_{β}
Phe	1.13	h_{α}	1.38	h_{β}
Pro	0.57	B_{α}	0.55	B_{β}
Ser	0.77	i_{α}	0.75	b_{β}
Thr	0.83	i_{α}	1.19	h_{β}
Trp	1.08	h_{α}	1.37	h_{β}
Tyr	0.69	b_{α}	1.47	H_{β}
Val	1.06	h_{α}	1.70	H_{β}

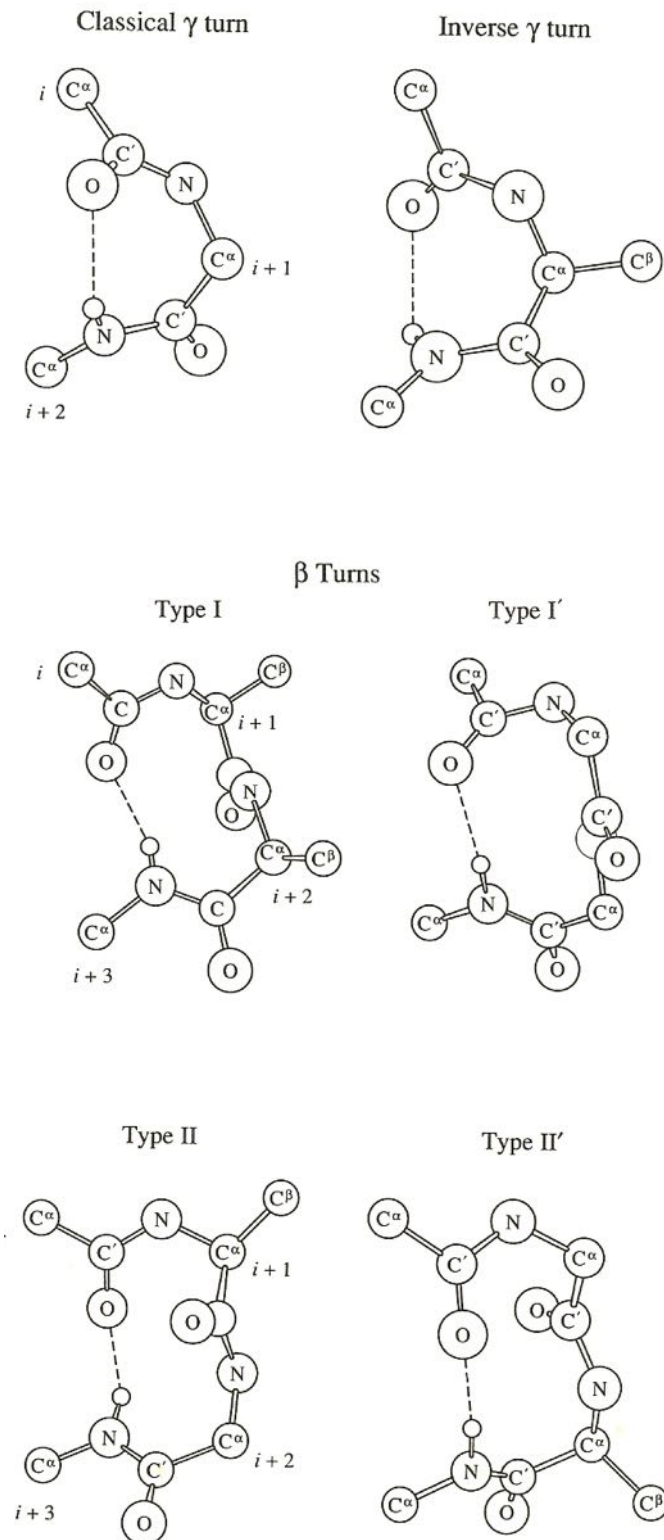
Loop Regions and Reverse Turns

- Most protein structures are assembled from combinations of secondary structure elements connected by loops and turns.
- Loop regions generally found at or near the surface of the protein, with the main-chain C=O and -NH groups exposed to solvent (H-bonding with water molecules).
- Internal main-chain H-bonding is important in tight reverse turns.
- In some globular proteins, ~1/3 of the residues lie in tight turns, reversing the direction of the polypeptide chain.
- Exposed loops and turns generally contain charged and polar amino acids that interact favorably with water.
- Sequences of tight reverse turns usually contain amino acid residues with smaller side chains. (frequently contain: Gly, Ser, Pro, Asn, Asp and Cys)



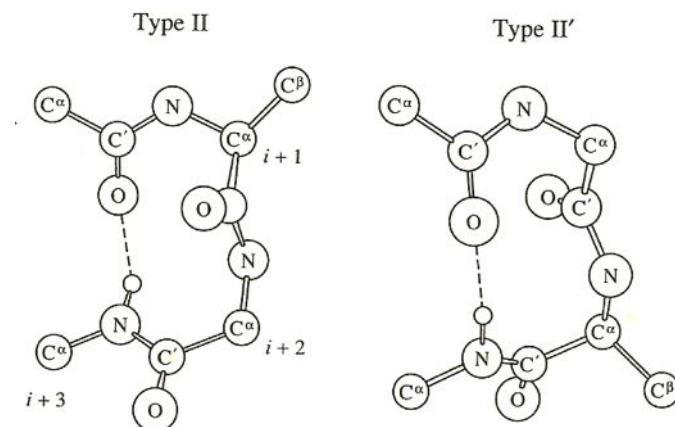
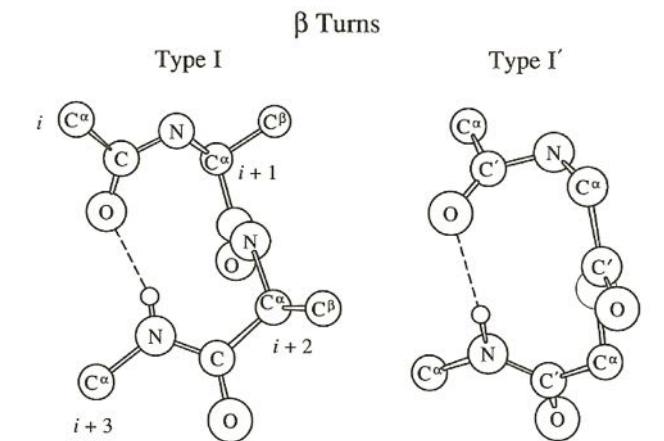
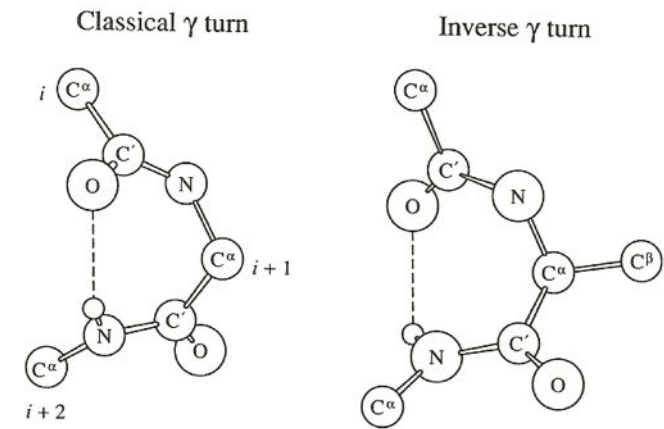
Loop Regions and Reverse Turns

- Tight reverse turns: defined by number of residues and ϕ & ψ angle combinations.
 - γ turn: tight turn connecting adjacent strands in a β -sheet. One turn residue does not participate in H-bonding network of the sheet (two types).
 - β turn: more common than the γ turn. Two turn residues not involved in hydrogen bonding of the sheet. Defined by four residues (i to $i+3$), residues i and $i+3$ participate in hydrogen bonding network of sheet (6 types).
- Conformations of reverse turns in proteins frequently deviate from the predicted ideal turns - *likely due to overestimation of peptide backbone conformational restrictions.* (i.e. $i/i+3$ H-bond is frequently absent.)
- **Long loop regions:** often flexible and can frequently adopt several different conformations. Such long loop regions (and conformational changes in loop) frequently play a role in the function of the protein.
 - Ω loop: longer and less well defined than turns. Usually 6-16 residues in length, and the distance between the ends is typically $<1\text{\AA}$.
- Longer loops are often defined by the rest of the peptide conformation.



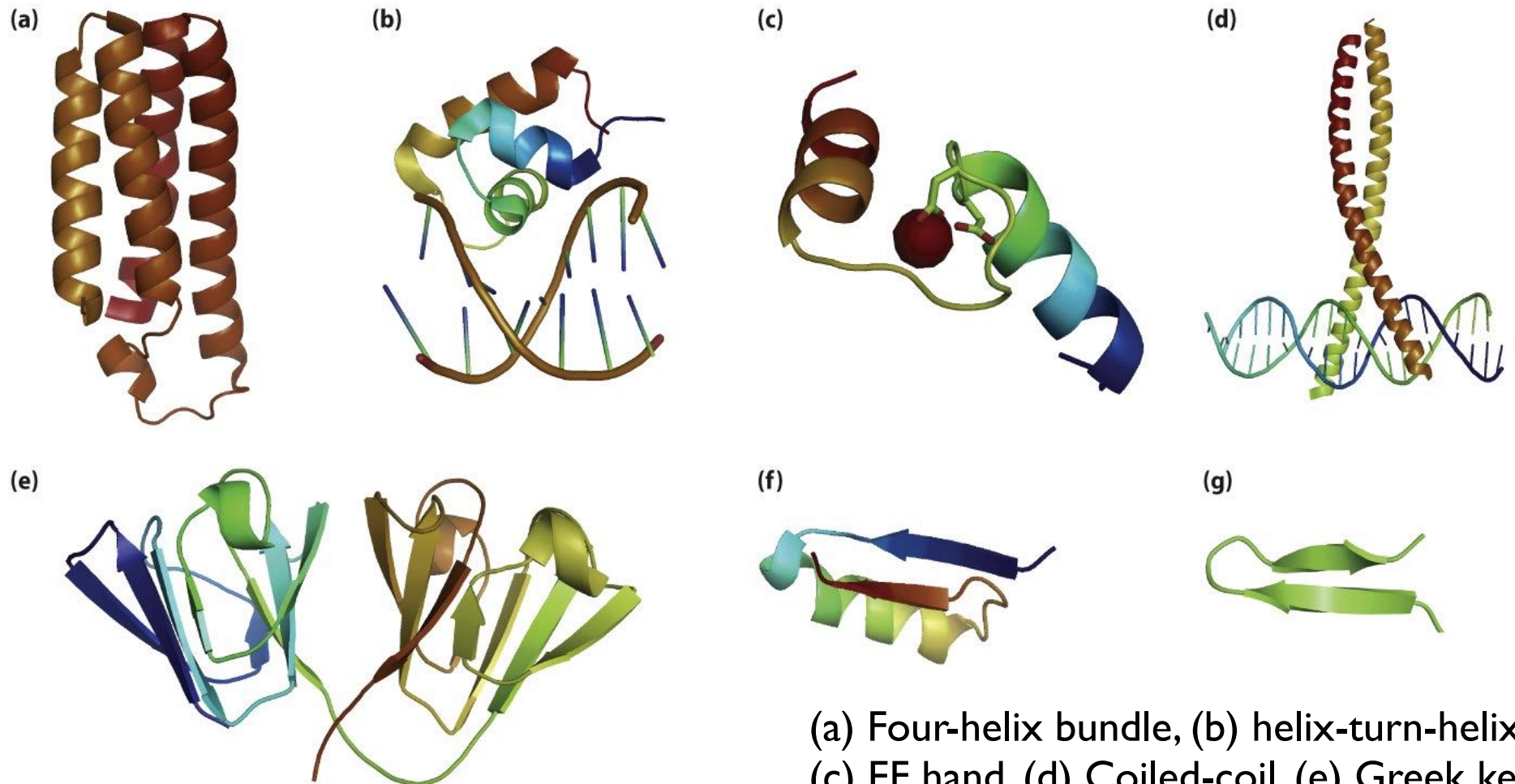
Loop Regions and Reverse Turns

- Three ideal types of ideal β turn are predicted based on favorable backbone geometries (I, II and III with mirror images I', II' and III').
- Type-I β turns occur most frequently, 2-3 times more common than Type II.
- Type-I' is particularly common in β hairpins. Presumably because it accommodates the inherent twist of a β sheet.
- Short loops such as γ and β turns depend on the positions of certain amino acids (such as Asn, Gly or Pro) that allow the chain to assume the turn conformation:
 - Type-I β turn geometry is compatible with any amino acid at positions $i - i+3$ (except Pro cannot occur at $i+2$).
 - In Type-I and II turns Gly predominates at $i+3$, and Pro at $i+1$.
 - Asp, Asn, Ser and Cys often occur at i position. Their side chains can H-bond with backbone N-H of $i+2$.
 - Gly and Asn occur most often at position $i+2$ in type-II turns.
 - Ideally, type-I' turns have Gly at $i+1$ and $i+2$.
 - Type II' turns have Gly at $i+1$.



Structural Motifs

- Secondary structure elements frequently pack/organize into supersecondary structures of motifs.
- Represent particularly stable arrangements of secondary structure.
- Combine and pack together forming larger and more complex assemblies.



(a) Four-helix bundle, (b) helix-turn-helix, (c) EF hand, (d) Coiled-coil, (e) Greek key, (f) b-a-b motif and (g) b hairpin

Simple Motifs

- Simple combinations of secondary structural elements with defined geometric arrangements have been observed as recurring themes in protein structures.
- These structural units are referred to as either supersecondary structures or “motifs”.
- Some motifs are associated with specified functions (such as DNA binding), and other have no specific biological function.
- Simplest motifs with specific functions consist of two α -helices connected by a loop.
 - Helix-turn-helix motif: specific for DNA binding (will be covered later).
 - A helix-loop-helix motif (EF hand): specific for calcium binding (found in parvalbumin, troponin-C, calmodulin and other calcium-binding proteins).

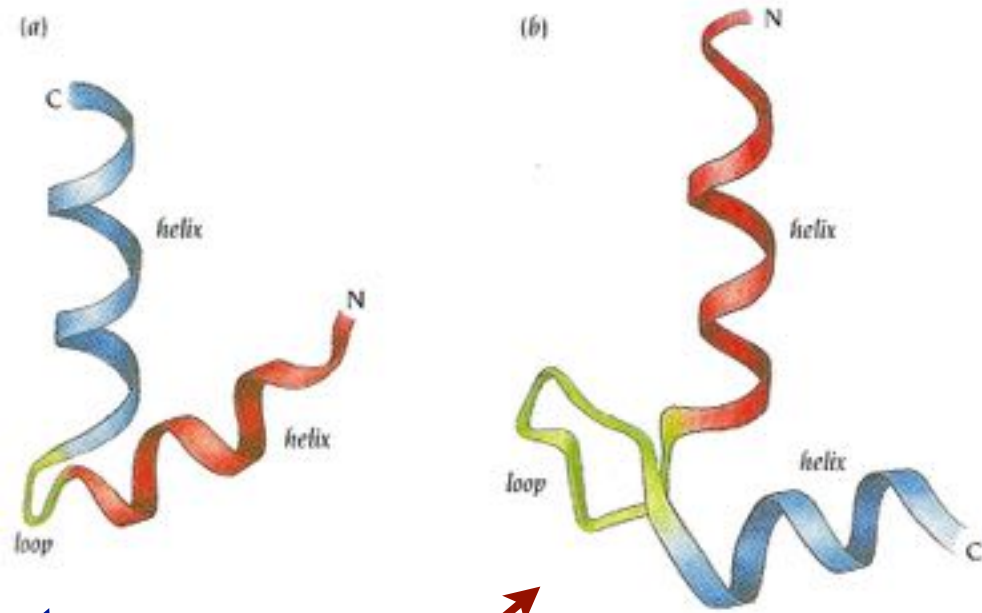
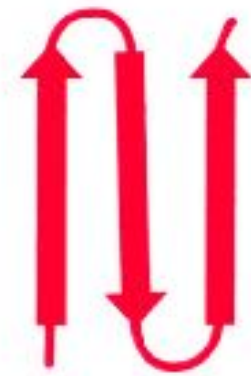


Figure 2.12 Two α helices that are connected by a short loop region in a specific geometric arrangement constitute a helix-turn-helix motif. Two such motifs are shown: the DNA-binding motif (a), which is further discussed in Chapter 8, and the calcium-binding motif (b), which is present in many proteins whose function is regulated by calcium.

β Hairpin Motif



β hairpin motif

- The β hairpin (or a β - β unit) is the simplest motif involving β -strands. Consists of two adjacent antiparallel β -strands connected by a loop/turn.
- β hairpin is common in proteins, and is present in most antiparallel β structures, either as isolated ribbons or as part of larger β -sheet structures.
- Length of turn connecting the strands can vary but is usually 2-5 residues in length. (i.e. a γ - or β -turn)
- Unlike EF hand and helix-turn-helix motifs, the β hairpin motif is not associated with a specific function.

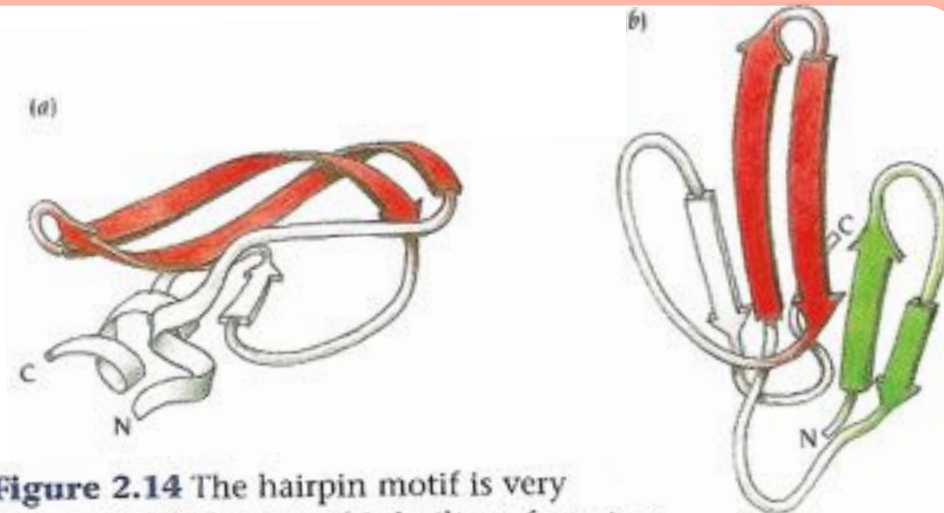
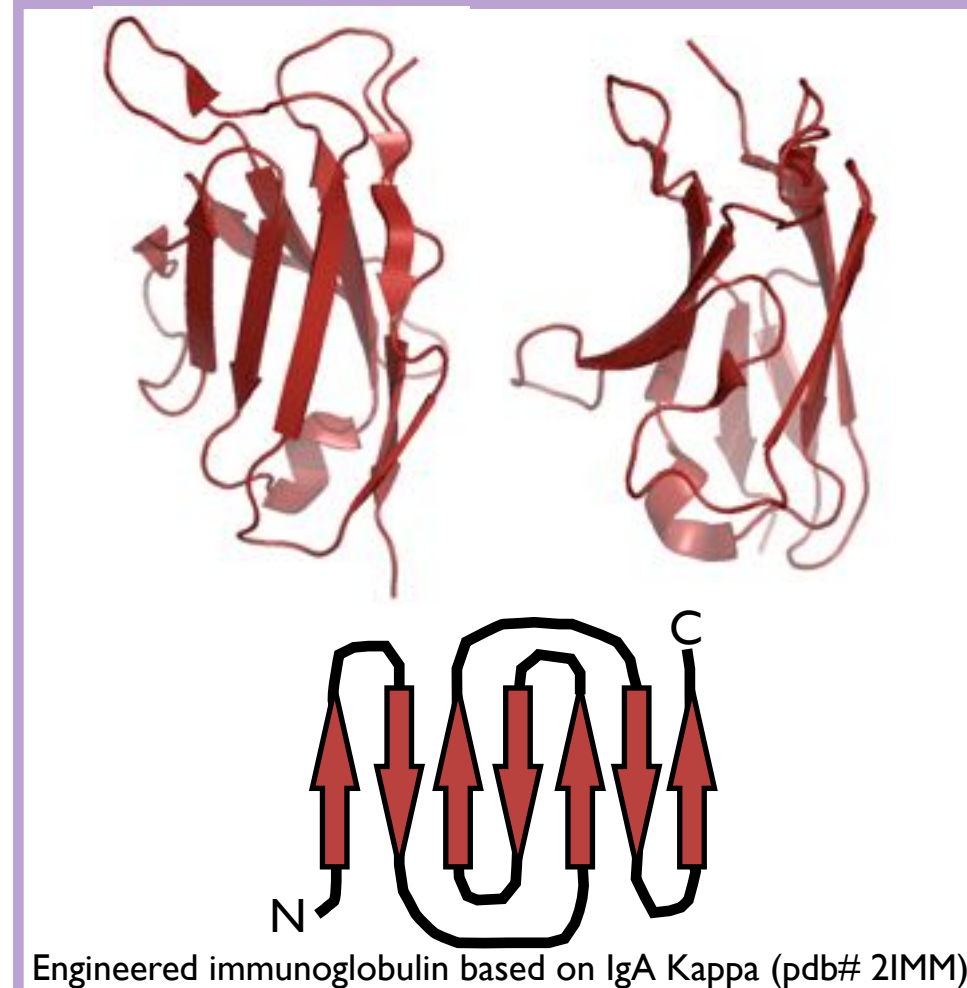
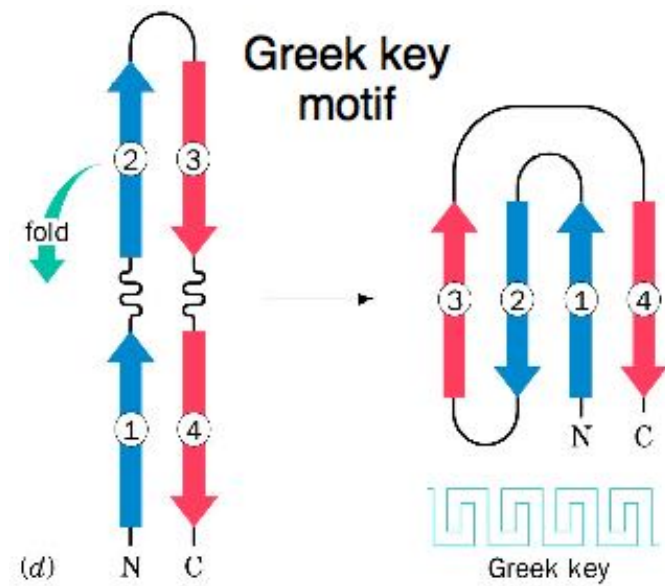


Figure 2.14 The hairpin motif is very frequent in β sheets and is built up from two adjacent β strands that are joined by a loop region. Two examples of such motifs are shown. (a) Schematic diagram of the structure of bovine trypsin inhibitor. The hairpin motif is colored red. (b) Schematic diagram of the structure of the snake venom erabutoxin. The two hairpin motifs within the β sheet are colored red and green. (Adapted from J. Richardson.)

Greek Key Motif

- A Greek motif consists of four adjacent antiparallel β -strands arranged in a pattern similar to the ornamental Greek key from Greek architecture and pottery.
- As with the β hairpin, the Greek key motif occurs frequently in protein structures and is not associated with a specific function.
- The immunoglobulin fold consists of two overlapping Greek keys that fold to form a β -sandwich.
- This is an arrangement shared by all known proteins that contain a β -sandwich motif.
- In a β -sandwich, hydrophobic side-chains are located on the interior facing side of both β -sheets and side-chains that are more hydrophilic are located on the outer face of each β -sheet.
- The two sheets may be aligned with respect to each other or oriented orthogonally to each other.



β - α - β Motif

- The β hairpin motif is a simple and common means for connecting two adjacent antiparallel β -strands.
- Adjacent strands in a parallel β -sheet that are consecutive in sequence require a longer connection (such crossover connections are frequently α -helices).
- β - α - β motif consists of a β -strand followed by a loop, an α -helix, another loop and a second β -strand (running in parallel to the first strand).
- β - α - β motif is found in the structures of almost every protein containing a parallel β -sheet.
- The axis of the crossover helix usually is aligned roughly in parallel to the β -strands, with the helix packing against the face of the strands, shielding hydrophobic residues in the strands from solvent.

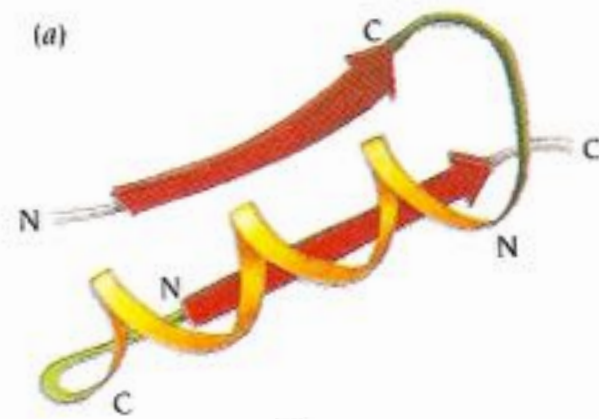


Figure 2.17 Two adjacent parallel β strands are usually connected by an α helix from the C-terminus of strand 1 to the N-terminus of strand 2. Most protein structures that contain parallel β sheets are built up from combinations of such β - α - β motifs. Beta strands are red, and α helices are yellow. Arrows represent β strands, and cylinders represent helices. (a) Schematic diagram of the path of the main chain. (b) Topological diagrams of the β - α - β motif.

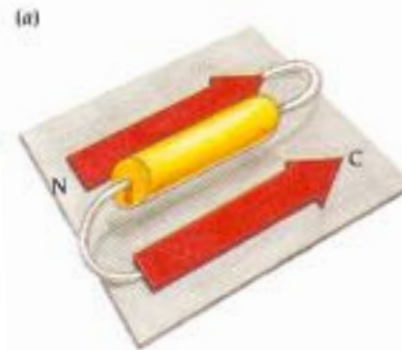
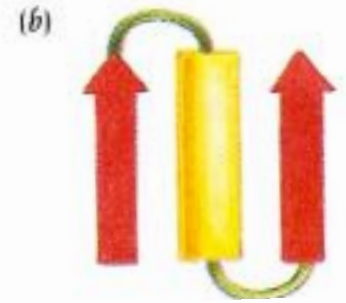


Figure 2.18 The β - α - β motif can in principle have two “hands.” (a) This connection with the helix above the sheet is found in almost all proteins and is called right-handed because it has the same hand as a right-handed α helix. (b) The left-handed connection with the helix below the sheet.

β - α - β Motif

- Loop segments connecting the helix to the strands can vary in length (~2 to >100 residues) ---- and the two connecting loops can have different functions.
- Functional loop segments in homologous proteins usually have conserved residues that are associated with function. Amino acid composition of loops that do not serve a specific function tend to be more variable.
- In some cases the connecting segment bridging the two β -strands is not a helix, but is instead polypeptide segment of poorly-defined structure.
- Essentially every known β - α - β motif has a right-handed twist to it. This is probably favored by the over-all right-handed twist of the sheet and the handedness of the bridging helix.

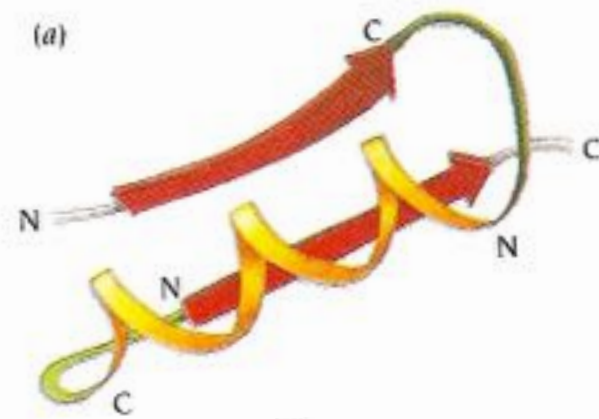


Figure 2.17 Two adjacent parallel β strands are usually connected by an α helix from the C-terminus of strand 1 to the N-terminus of strand 2. Most protein structures that contain parallel β sheets are built up from combinations of such β - α - β motifs. Beta strands are red, and α helices are yellow. Arrows represent β strands, and cylinders represent helices. (a) Schematic diagram of the path of the main chain. (b) Topological diagrams of the β - α - β motif.

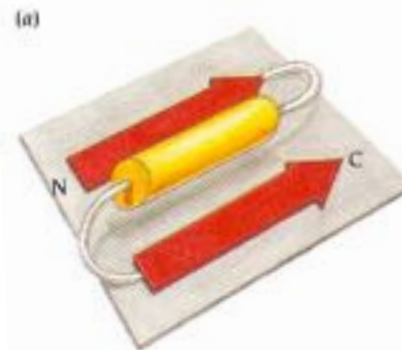
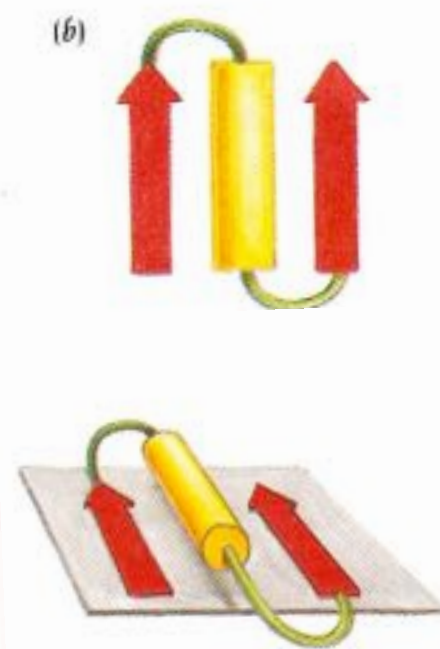
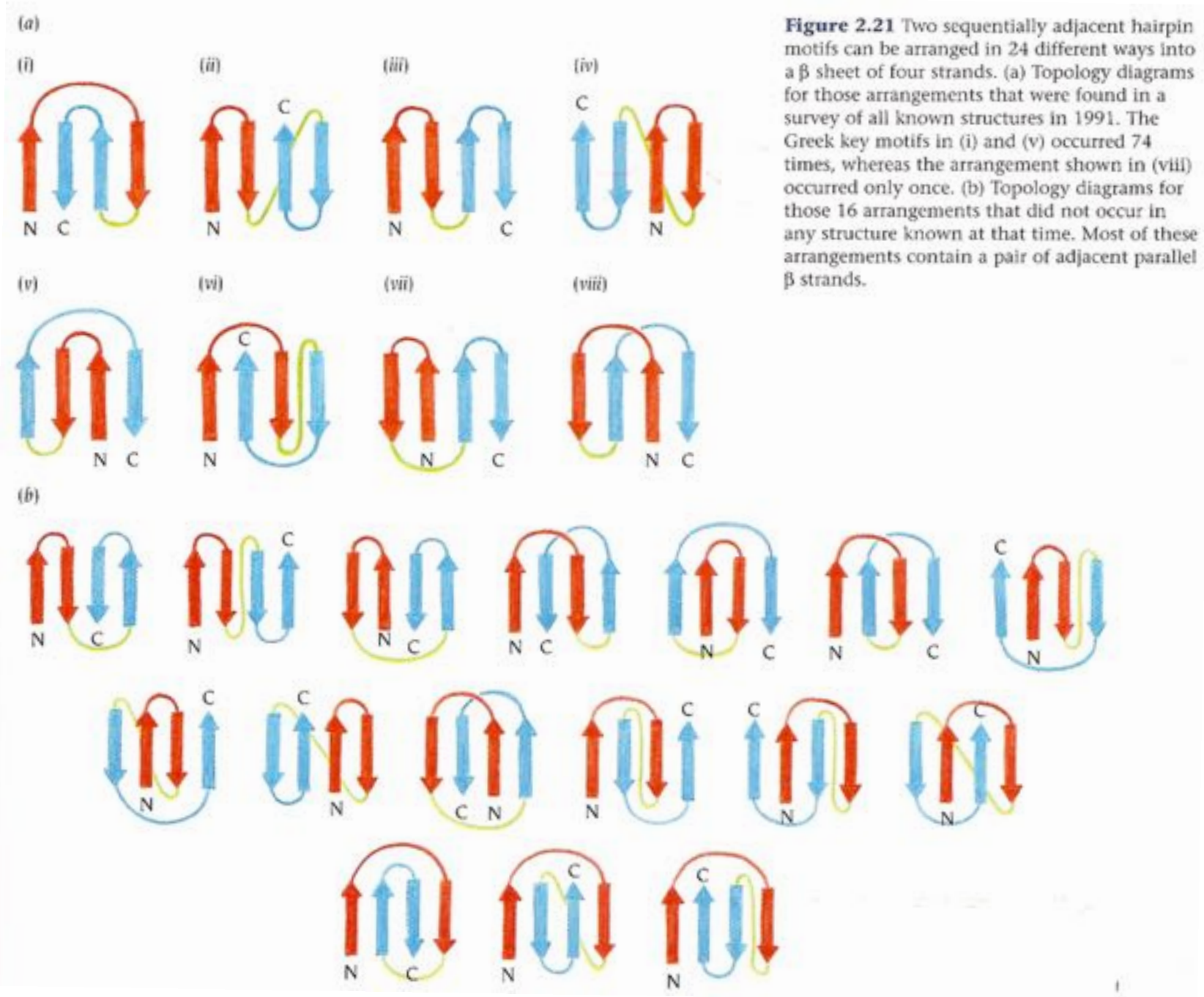


Figure 2.18 The β - α - β motif can in principle have two “hands.” (a) This connection with the helix above the sheet is found in almost all proteins and is called right-handed because it has the same hand as a right-handed α helix. (b) The left-handed connection with the helix below the sheet.

Complex Motifs from Simple Motifs

- Complex motifs can arise from combinations of simple motifs.
- to the right are illustrated 24 different ways two β hairpin motifs could combine to form a four-stranded β -sheet.
- Survey of all known structures in 1991 showed that only (i)-(viii) occur in natural proteins (occur with different frequencies).



Common Folds

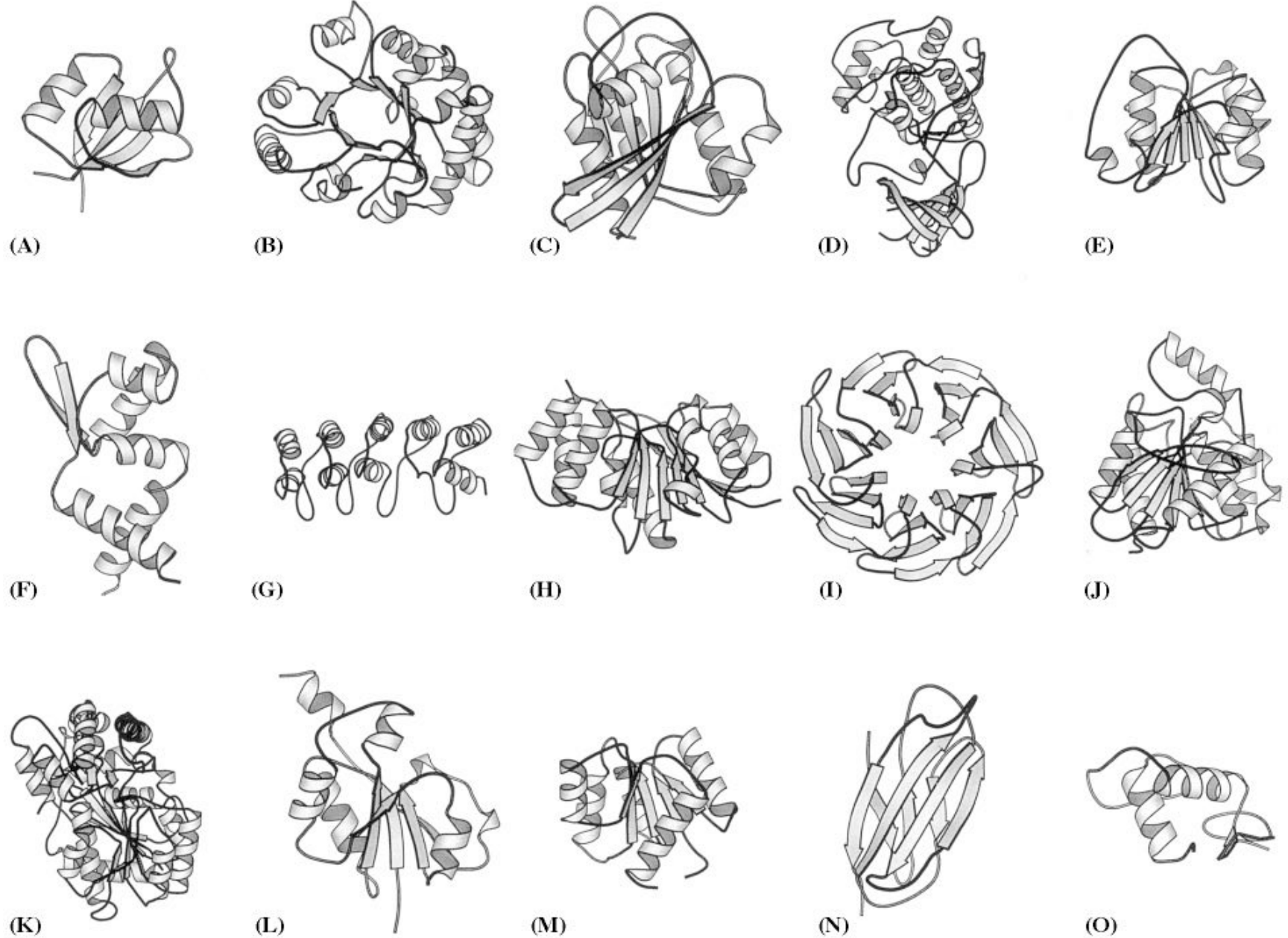


Figure 1. The 15 most populated folds. They were selected on the basis of a structural annotation of proteins from completely sequenced genomes of 20 bacteria, five Archaea, and three eukaryotes [C. Zhang, unpublished data]. From left to right and top to bottom, they are: ferredoxin-like (4.45%) (A), TIM-barrel (3.94%) (B), P-loop containing nucleotide triphosphate hydrolase (3.71%) (C), protein kinases (PK) catalytic domain (3.14%) (D), NAD(P)-binding Rossmann-fold domains (2.80%) (E), DNA/RNA-binding 3-helical bundle (2.60%) (F), α - α superhelix (1.95%) (G), S-adenosyl-L-methionine-dependent methyltransferase (1.92%) (H), 7-bladed beta-propeller (1.85%) (I), α / β -hydrolases (1.84%) (J), PLP-dependent transferase (1.61%) (K), adenine nucleotide α -hydrolase (1.59%) (L), flavodoxin-like (1.49%) (M), immunoglobulin-like β -sandwich (1.38%) (N), and glucocorticoid receptor-like (0.97%) (O), where the values in parentheses are the percentages of annotated proteins adopting the respective folds.

Secondary Structure: Classification of Globular Proteins

- Protein structures have been divided into three main classes based on their secondary structures.

-Michael Levitt and Cyrus Chothia-

- Within each class there are many variations of the basic theme.

- Protein domains dominated by α -helices. (α)

- ~60% of the residues are in helices.
- Helices are usually in contact with each other.

- Protein domains dominated by β -sheet. (β)

- Almost always contain two β -sheets that pack against each other.

- Protein domains containing helices and sheets that interact and often alternate along the peptide chain. (α/β)

- Usually have one major β -sheet of primarily parallel strands; a helix usually occurs in each of the segments of the polypeptide chain connecting β -strands. Helices pack against either or both sides of the sheet.

- Protein domains containing discrete collections of helices and sheets. ($\alpha+\beta$)

- Usually contain a single antiparallel β -sheet with helices clustered at one or both ends of the β -sheet.

Structures containing both α -helices and β -strands.

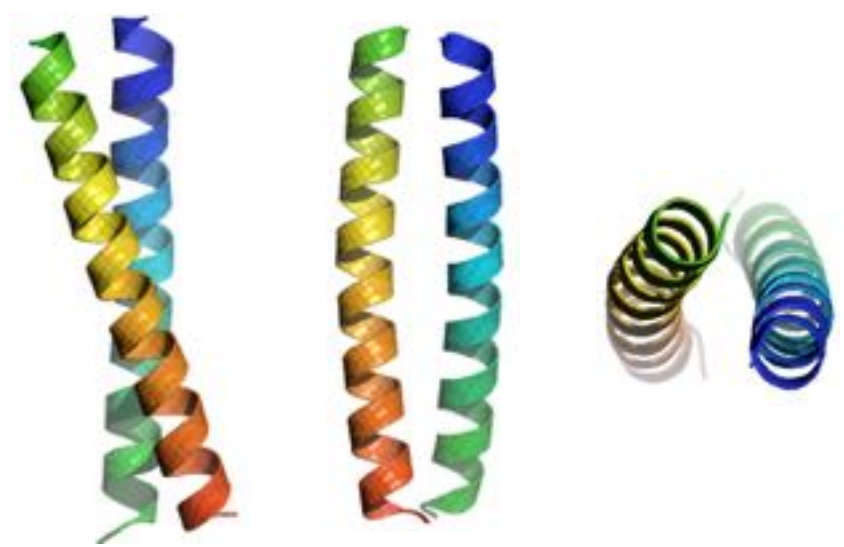
α -helical Structures

Alpha-helical Structures: Helical Bundles

- α helices can form many different classes of structures, but isolated helices are only marginally stable in solution.
- Frequently stabilized by packing together with other helices--stabilized through interactions of hydrophobic side chains.
- *Transmembrane segments are frequently α -helices.*

In α -helical coiled-coils (such as GCN4 Leucine Zipper), side chain interactions are maximized if the two α helices twist around each other (left-handed supercoil):

- Results in tightening of the helix (3.6 \rightarrow 3.5 res./turn). -- heptad repeat pattern (labeled a-g)
- Complementarity of hydrophobic residues in the a and d positions (knobs and holes).
- Salt bridges between residues in the e and g positions.



GCN4 Leucine Zipper (pdb# 2ZTA)
parallel 2-helix coiled-coil

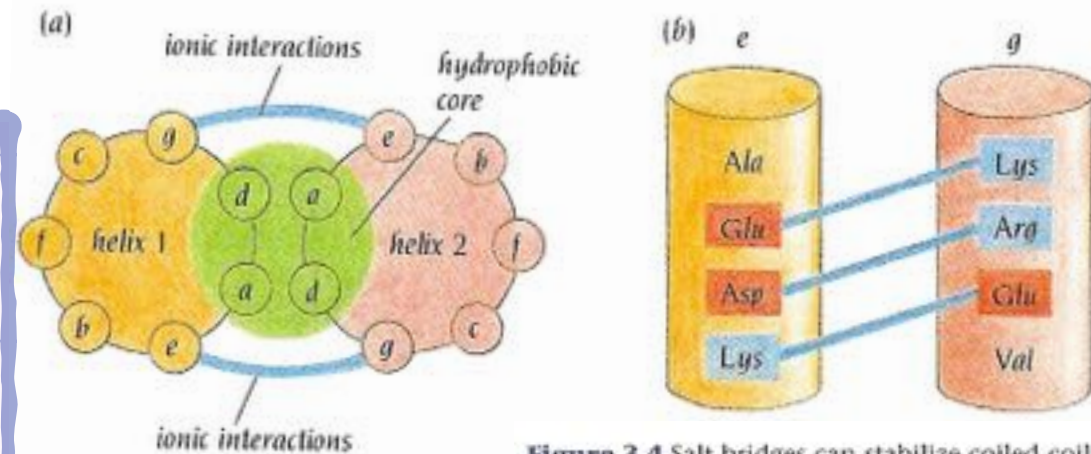


Figure 3.4 Salt bridges can stabilize coiled-coil structures and are sometimes important for the formation of heterodimeric coiled-coil structures. The residues labeled "e" and "g" in the heptad sequence are close to the hydrophobic core and can form salt bridges between the two α helices of a coiled-coil structure, the e-residue in one helix with the g-residue in the second and vice versa. (a) Schematic view from the top of a heptad repeat. (b) Schematic view from the side of a coiled-coil structure.

Side-chain Packing - “Knobs in Holes”

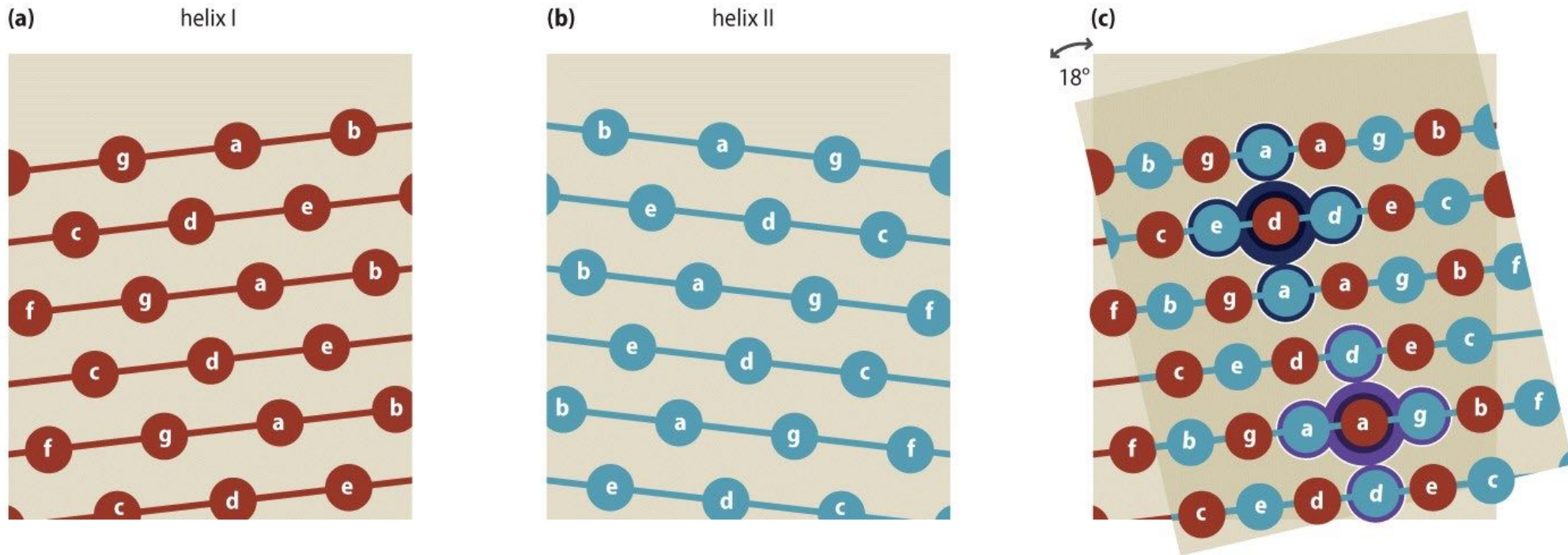


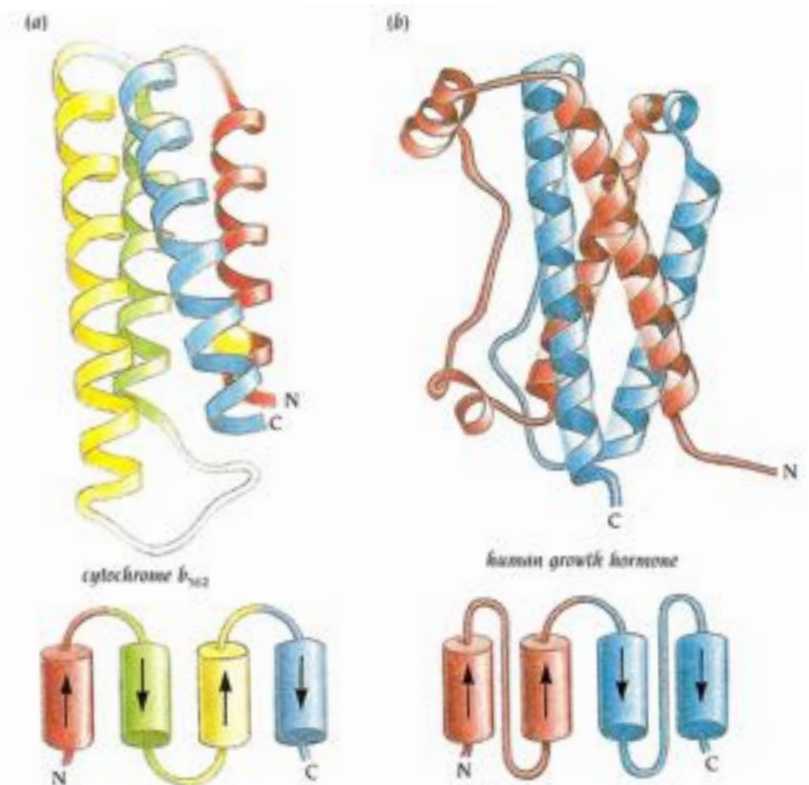
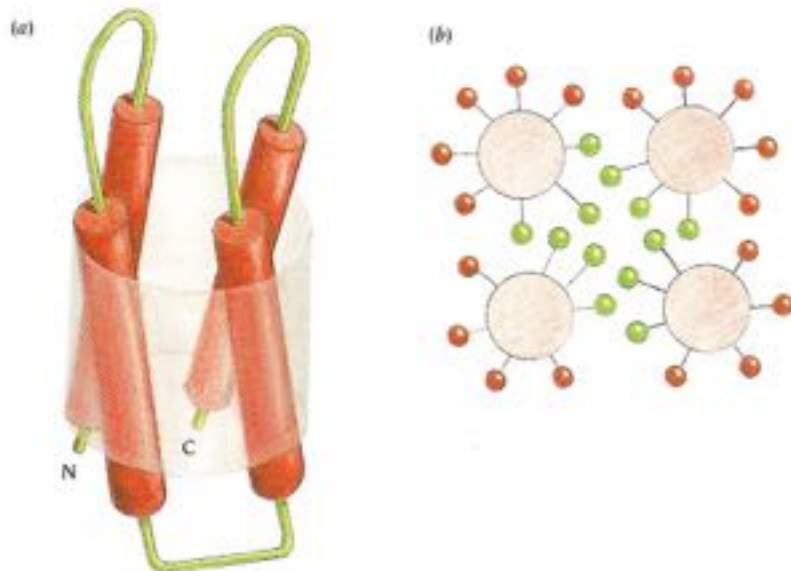
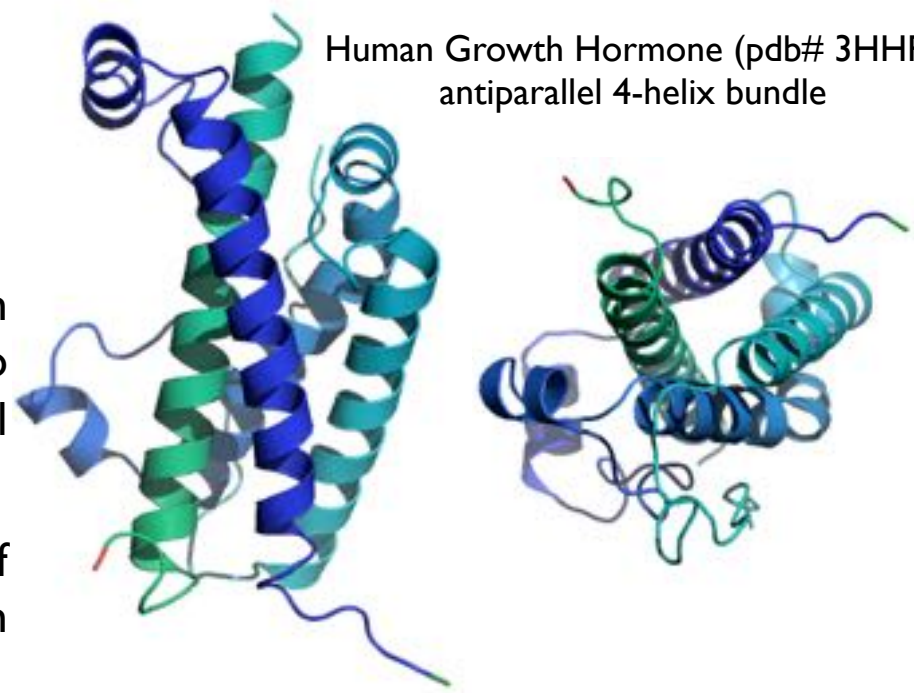
Figure 1.37 How Proteins Work (©2012 Garland Science)

- Complementarity of hydrophobic residues in the *a* and *d* positions (“knobs in holes”).
- In forming coiled-coil, side chains of hydrophobic residues in *d* positions (often Leu or Ile) of each strand pack against each other every second turn of the α -helices.
- Residues in *a* positions are generally hydrophobic and similarly pack against each other.
- In the “knob in hole” packing scheme, hydrophobic side chains in the *a* and *d* positions of one of the helices makes contact with four side chains from the other helix.
- The side chain of a *d*-position residue in one helix is directed into a hole at the surface of the other helix - hole is flanked by d_n, a_{n-3}, a_{n+4} and e_{n+1} of the complementary helix.

Alpha-Domain Structures: Helical Bundles

- 4-helix bundle consisting of four α helices arranged in a bundle with the helical axes aligned nearly parallel to each other is the simplest and most common α -helical domain. (such as Human Growth Hormone)
- Hydrophobic side-chains are buried in the interior of the bundle, and charged and polar side-chains are on the outer surface.
- 4-helix bundles can be formed using different topological arrangements of the α helices.
- Adjacent helices in the bundle are generally aligned antiparallel to each other.

Human Growth Hormone (pdb# 3HHR)
antiparallel 4-helix bundle



Helix Packing: Ridges and Grooves

In most helical-bundle structures, the helices are packed against each other based on a “ridges and grooves” model.

- Side-chains are arranged in a helical row along the surface of an α helix forming ridges separated by shallow grooves.
- Ridges and grooves formed by amino acids 3 or 4 residues apart.
-influences optimal helix orientation-
- Pack with the ridges of one helix occupying the grooves of another.

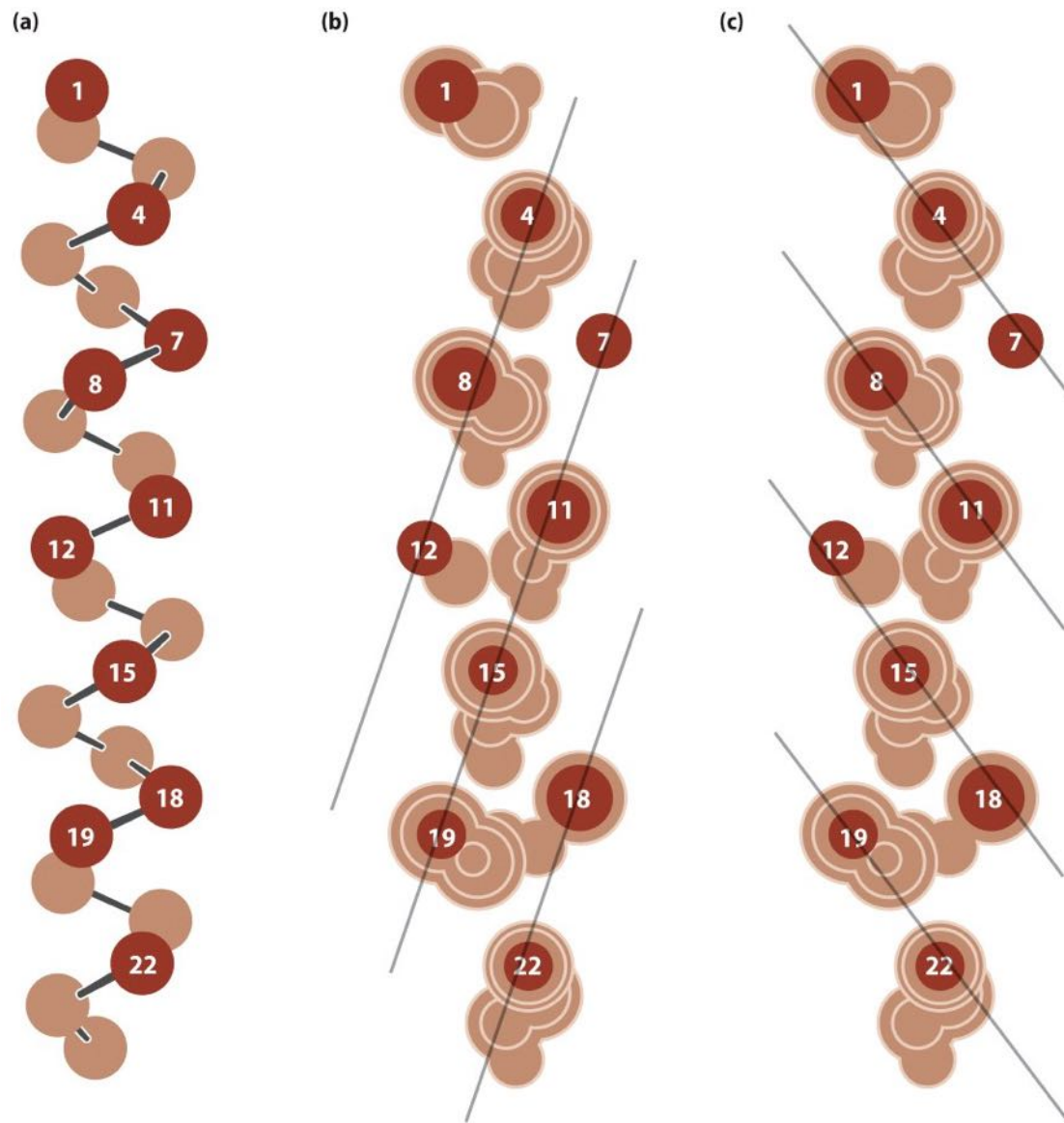
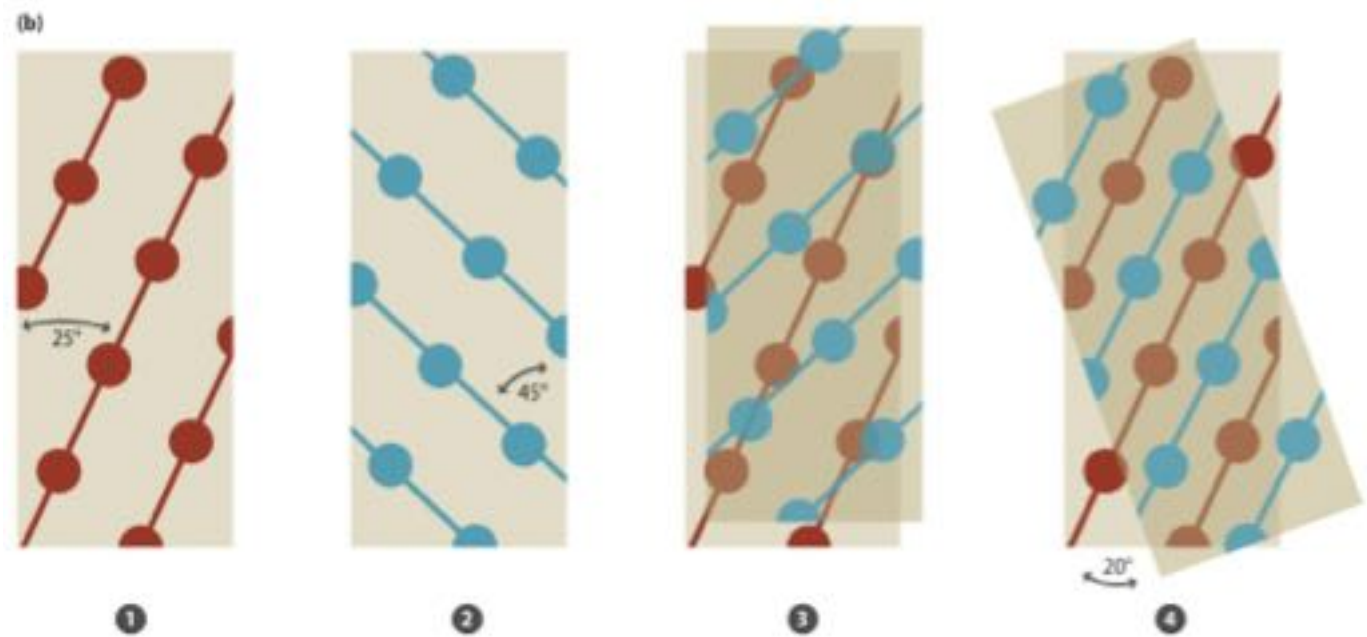


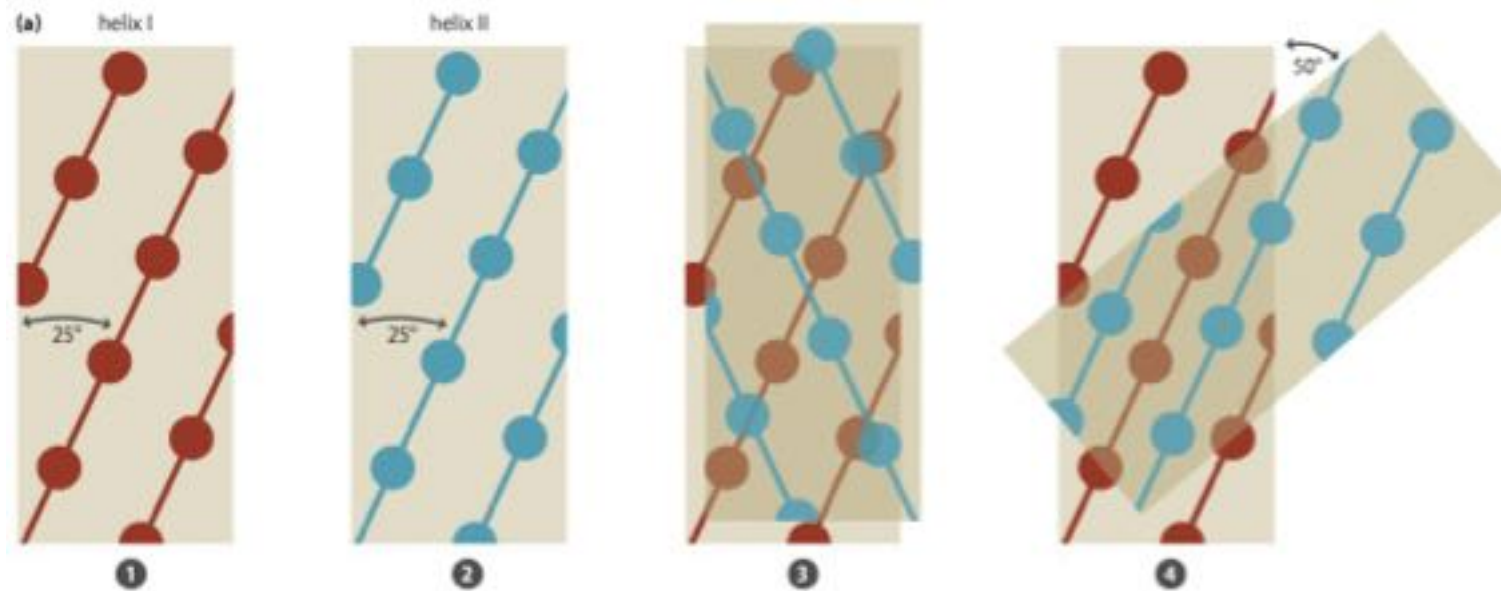
Figure 1.35 How Proteins Work (©2012 Garland Science)

Helix Packing: Ridges and Grooves

- Optimal core packing geometry in most helical bundles is achieved by an antiparallel alignment of helices oriented at a 20° angle relative to each other. (b)

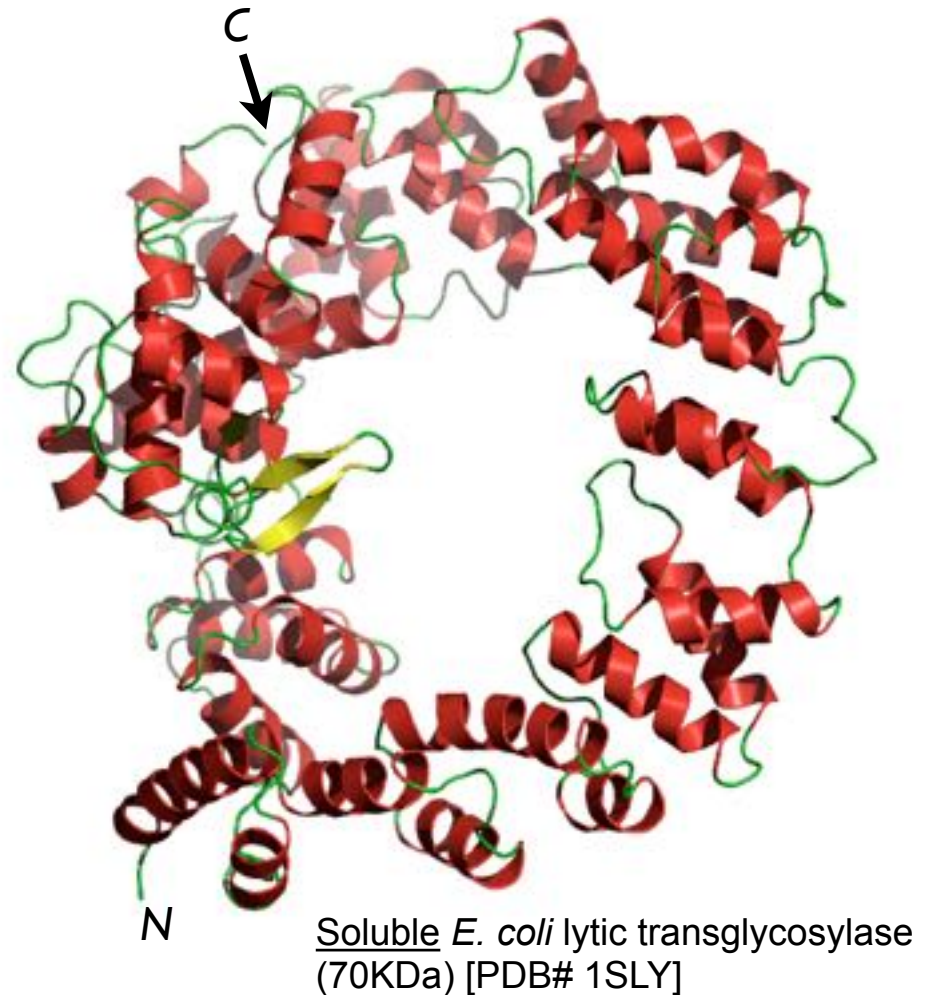


- An alternative packing scheme arises when one of the helices utilizes the ridge/groove scheme illustrated in (a). In this scheme helices are again oriented antiparallel, but with their helical axes at a 50° angle relative to each other.



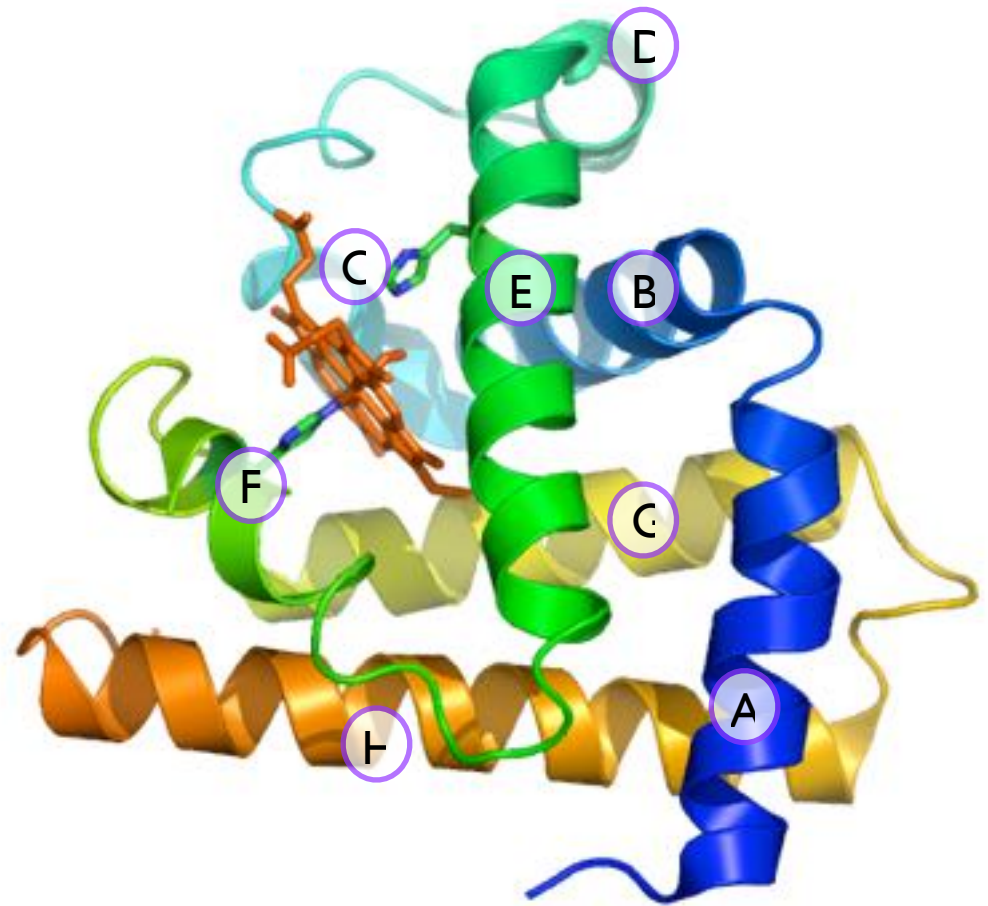
Large & Complex α -helical Domains

- Several enzymes have been found to contain domains as large as 300-400 residues consisting of more than 20 α -helices packed together to form a complex structure resembling a doughnut or horseshoe.
- Observed in bacterial muramidase. Where *N*-terminal 450 residues form such a structure containing 27 α -helices.
- In muramidase, the catalytic domain lies at the top of the ring, the function of the ring structure is not known, but it may play a role in substrate specificity.



The Globin Fold

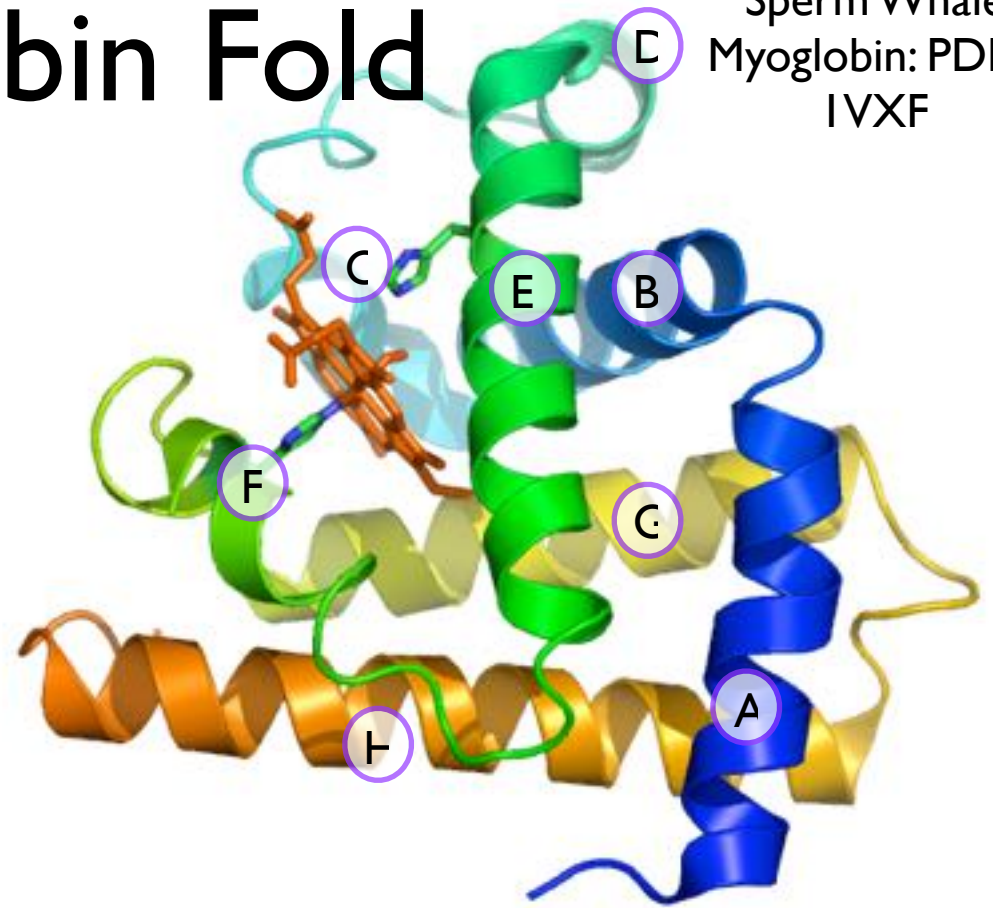
- The Globin fold is one of most important α structures.
- Found in large group of proteins: myoglobin, hemoglobin, phycocyanins and algal light-harvesting centers.
- Relative orientation of interacting helices in the globin fold is distinct from that found in coiled-coils and helical bundles.
- Globin fold is a bundle of 8 helices (labeled A-H) linked by relatively short connecting loops.
- Orientation and positioning of helices creates a pocket for the active site



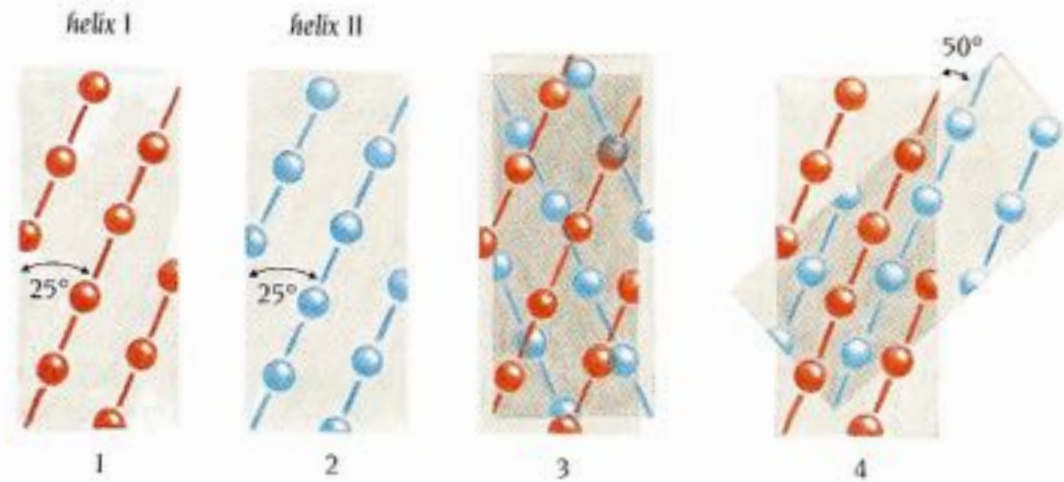
Sperm Whale Myoglobin: PDB# 1VXF

The Globin Fold

Sperm Whale
Myoglobin: PDB#
IVXF

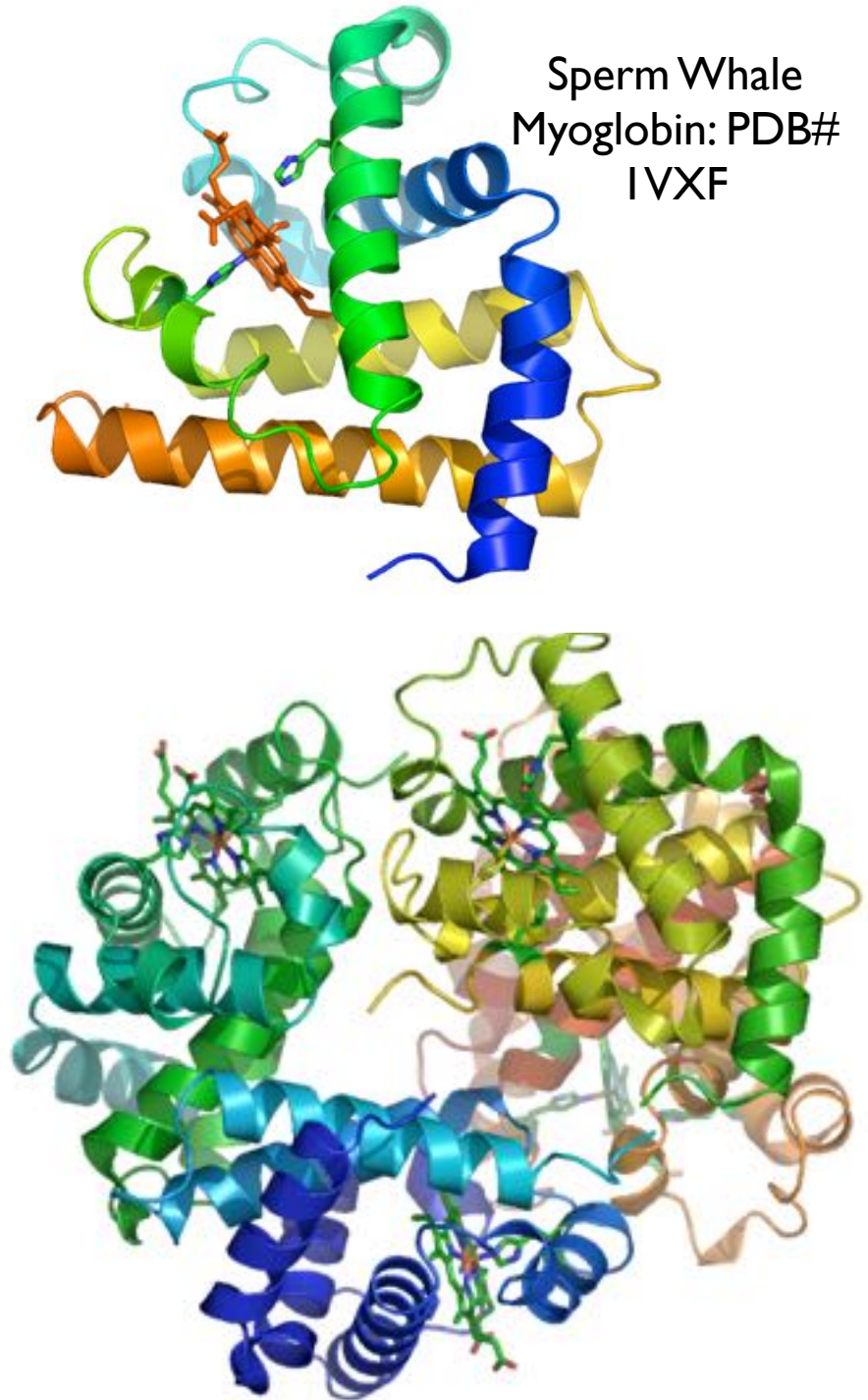


- In myoglobin, helix length ranges from 7 residues in helix (C) to 28 residues in helix (H).
- Majority of packing interactions occur between pairs of helices that are not sequentially adjacent, with helices (G) and (H) being an exception.
- Geometry of helix orientation in globin fold reflects packing scheme: ridge and groove - helices pack at $\sim 50^\circ$ relative to each other.



Globin Fold has been Preserved

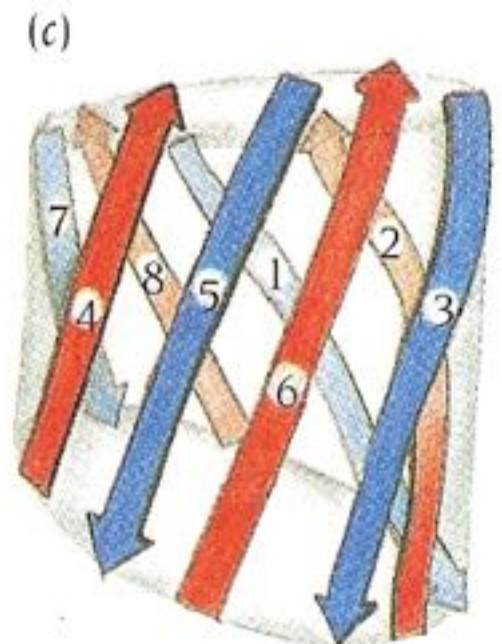
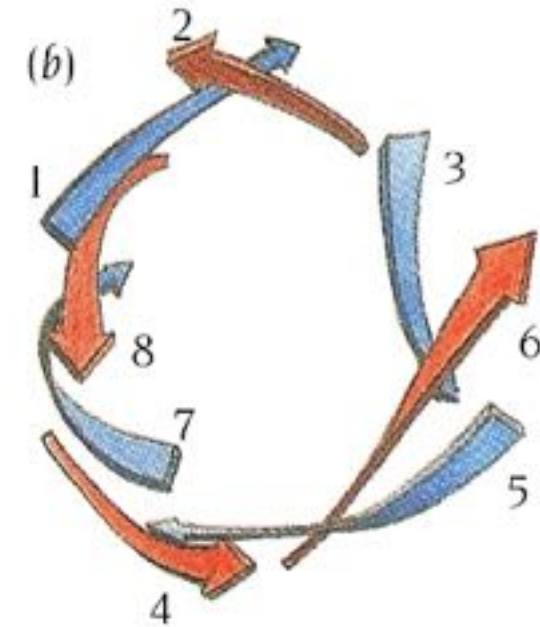
- Hemoglobin and myoglobin figure prominently in our understanding of proteins.
- Globin domains have been observed in proteins from diverse organisms: mammals, insects, and plants.
- Demonstrate varied amino acid sequence homology (16% - 99%). Provides an example of conservation of fold.
- Hydrophobic interior is preserved in each case.
- In comparing sequences of proteins with globin fold:
 - Sequence conservation and size-compensatory mutation in hydrophobic core are not important.
 - Strong preference for hydrophobic residues in buried positions (59 such positions in globin). -no such conservation in surface/exposed positions.*
 - Helix movement can accommodate mutations at buried positions.
 - Loop flexibility allows one or more helices to shift without causing other helices to shift as well.
 - Important to maintain integrity of active site.



β Structures

Introduction to β Structures

- Antiparallel β structures comprise the second large group of protein domain structures.
- Functionally diverse group:
 - enzymes, transport proteins, antibodies, cell surface proteins and viral coat proteins. (not a complete list)
- Cores of these proteins consist of β strands arranged antiparallel, usually forming two joined β sheets that pack against each other.
- In β structures, β sheet is twisted & when two sheets pack against each other it usually results in a barrel-like structure.
- Antiparallel β structures generally have a core of hydrophobic side chains from sheets, and the surface formed by residues from the loops and the strands.
- The number of possible topologies for forming a antiparallel β sheet rapidly increases with the number of strands. However, the number of topologies that have been observed is very small... with most β structures falling within a few groups of common/similar topologies.



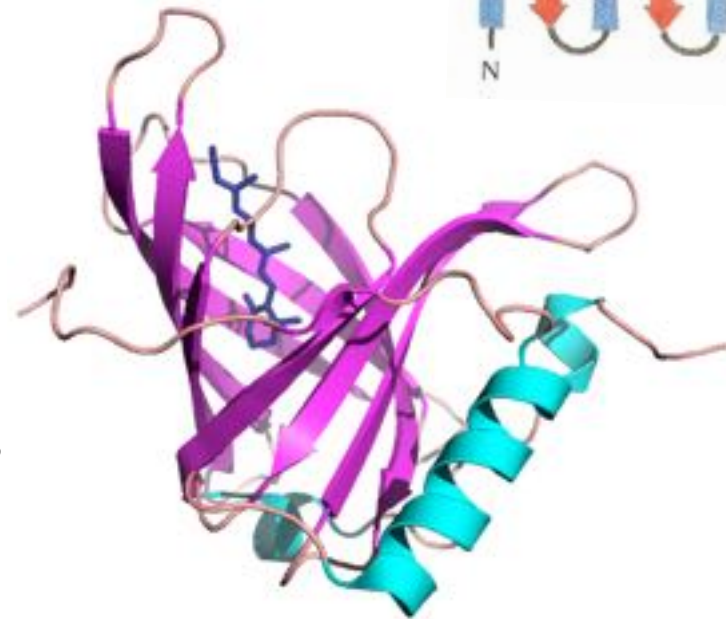
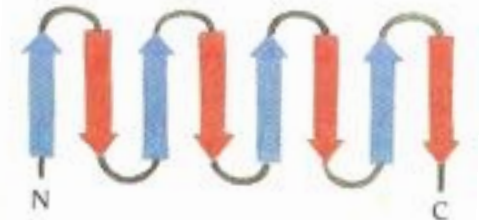
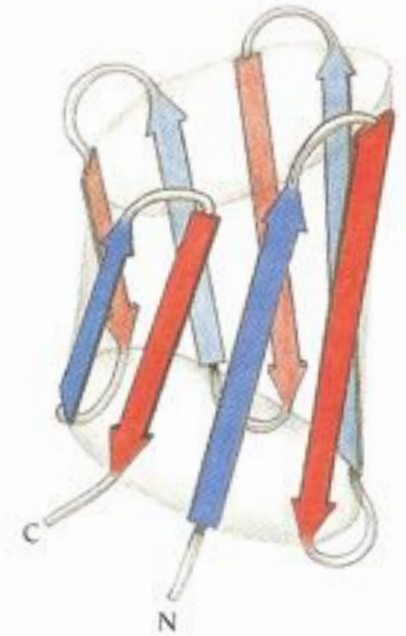
Common β Structure Topologies.

Most β structures fall within a few groups of common/similar topologies.

1. Up-and-down barrels
2. Greek key motif
3. jelly roll barrels

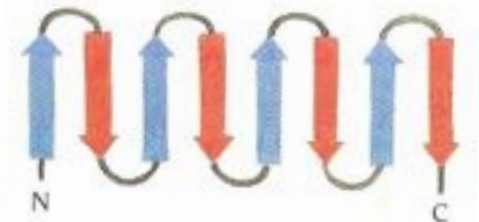
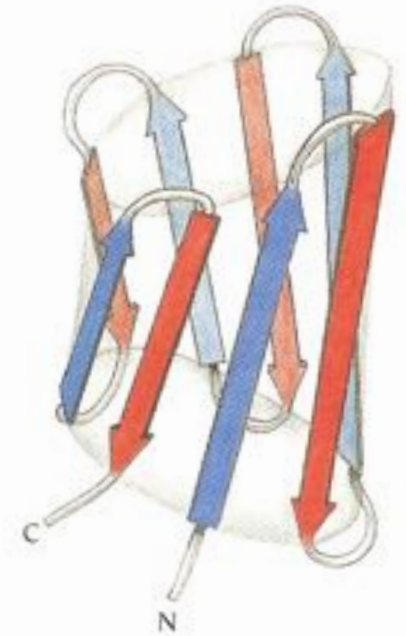
Up-and-Down Barrels

- The simplest β structure topology is the up-and-down β barrel, assembled from successive β hairpins.
- This arrangement allows considerable structural and functional versatility.
- Retinol-binding protein (RBP) is a good example of a up-and-down β barrel containing protein.
- RBP consists of a single 182 residue polypeptide chain. It is responsible for the transport of retinol (vitamin A) from its storage site in the liver to vitamin-A-dependent tissues. (1:1 stoichiometry)
- RBP contains a β -barrel core consisting of 8 β strands aligned antiparallel to each other, and a short C-terminal helix that packs against the outside of the barrel.



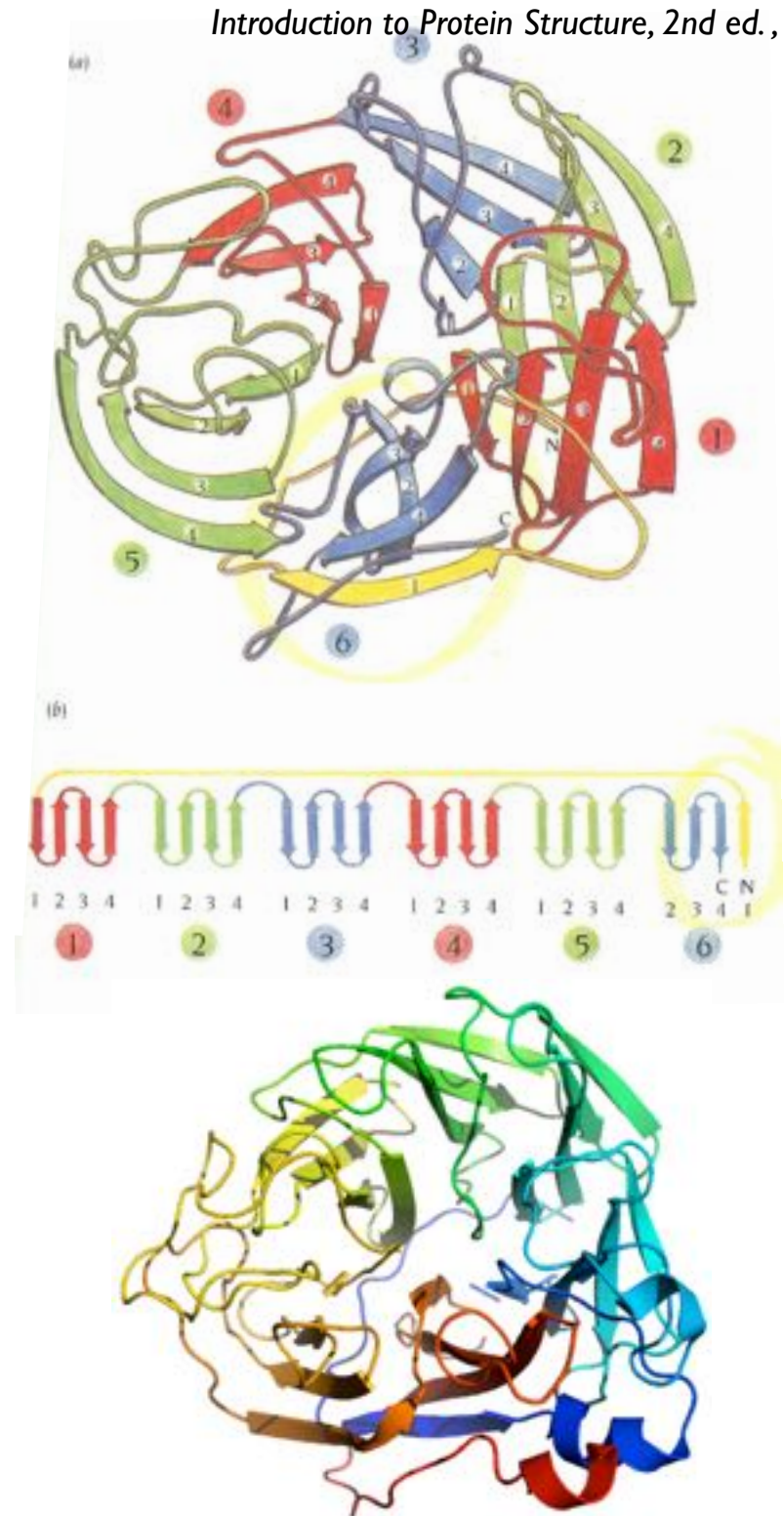
Up-and-Down Barrels

- The β strands of the barrel are curved and twisted, and one end of the barrel is open to solvent (the other end of the barrel is closed by tight packing of hydrophobic side-chain packing).
- Hydrophobic retinol binds at the open end of the barrel, its hydrophobic tail inserted into the hydrophobic interior of the barrel.
- The three-dimensional structure of the apo protein does not differ significantly from that of the protein with bound retinol.
- A large part of the RBP surface consists of side chains from residues in the β strands. As a result the strands show an amphipathic pattern of amino acid side chains: hydrophobic residues directed into the barrel core alternating with polar and charged residues displayed on the outer surface of the barrel.
- The structure of the protein also indicates that the barrel is build up from two β sheets.



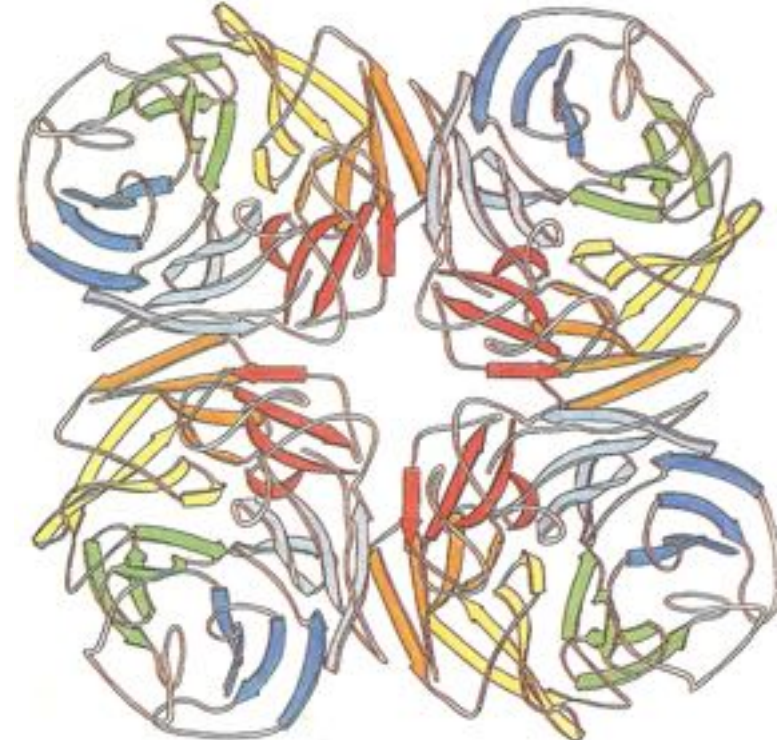
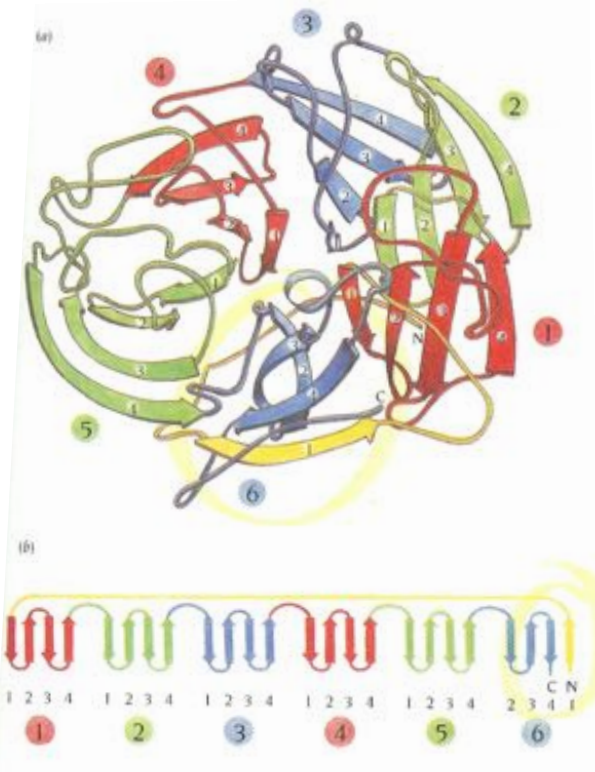
Neuraminidase

- Neuraminidase from influenza virus is another example of an up-and-down β sheet arrangement. The enzyme removes sialic acid residues from carbohydrate chains on viral hemagglutinin and glycosylated cell membrane proteins, facilitating the release of progeny virions from infected cells.
- The arrangement and packing of the β sheets in neuraminidase is dramatically different from that observed in RBP.
- The β sheets in neuraminidase do not form a simple barrel, instead forming 6 small (4 stranded) sheets arranged in a pattern resembling the blades of a propeller.
- The loop regions between the β strands in the middle of the propeller structure form the active site.
- The soluble head component can be proteolytically separated from the stalk (membrane anchor).



Neuraminidase

- Neuraminidase is a tetrameric proteins, consisting of 4 identical subunits.... each chain ~470 amino acids in length.
- Each of the four subunits of the tetrameric neuraminidase head is folded into single a domain assembled from six closely-packed similarly folded four-stranded antiparallel β sheet.
- The strands in the β sheets have a large twist such that the direction of the 1st and 4th strands in each sheet are oriented at 90° angles relative to each other.
- The tetrameric protein consists of ~1600 residues, and is composed of four identical polypeptide chains, each of which is folded into the “propeller-like” super barrel structure to the right.



Neuraminidase

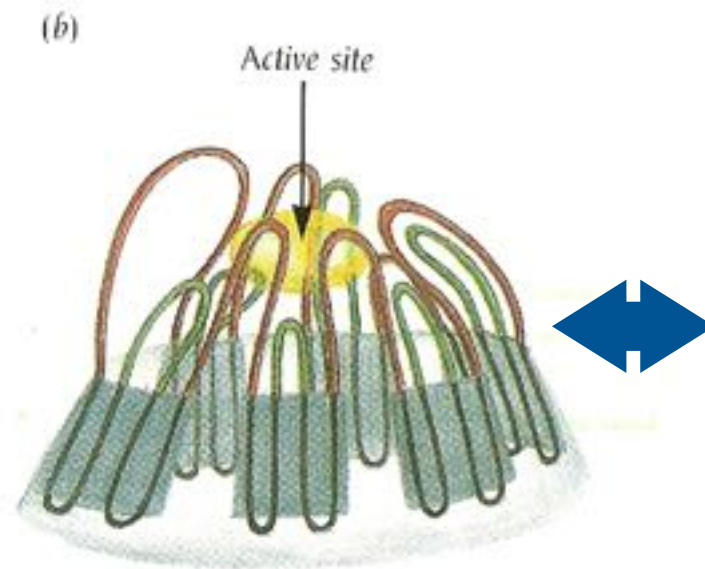
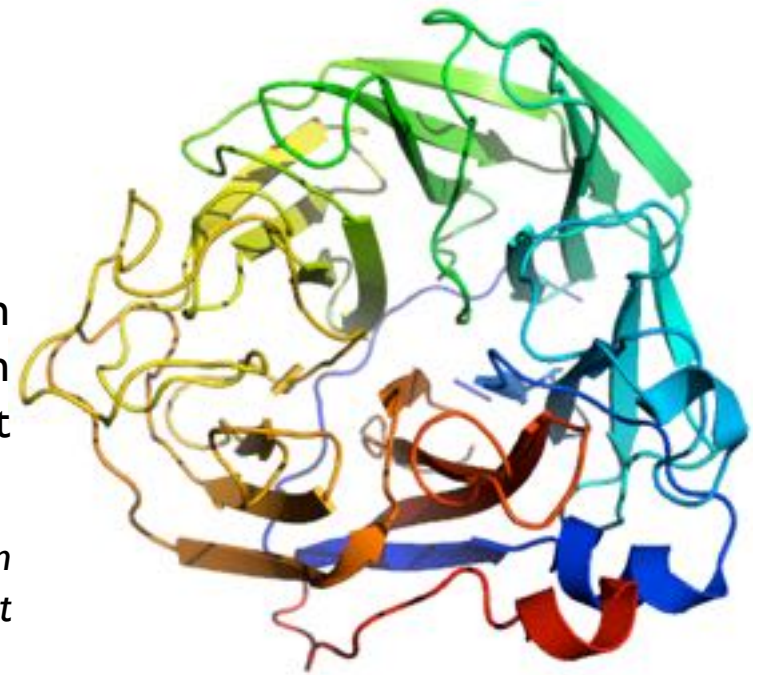
- The 6 β sheets within each subunit are connected to each other by a loop that runs from the end of the 4th strand in the preceding sheet, across the top of the subunit, to the first strand in the next sheet.

In the last (6th) β sheet in the subunit, the outermost strand (the 4th in the other sheets) is the overall first strand in the chain and is not connected to strand 3 as is the case in the other 5 sheets.

- The loops connecting the 2nd and 3rd strands in each sheet also lie at the top of the structure.

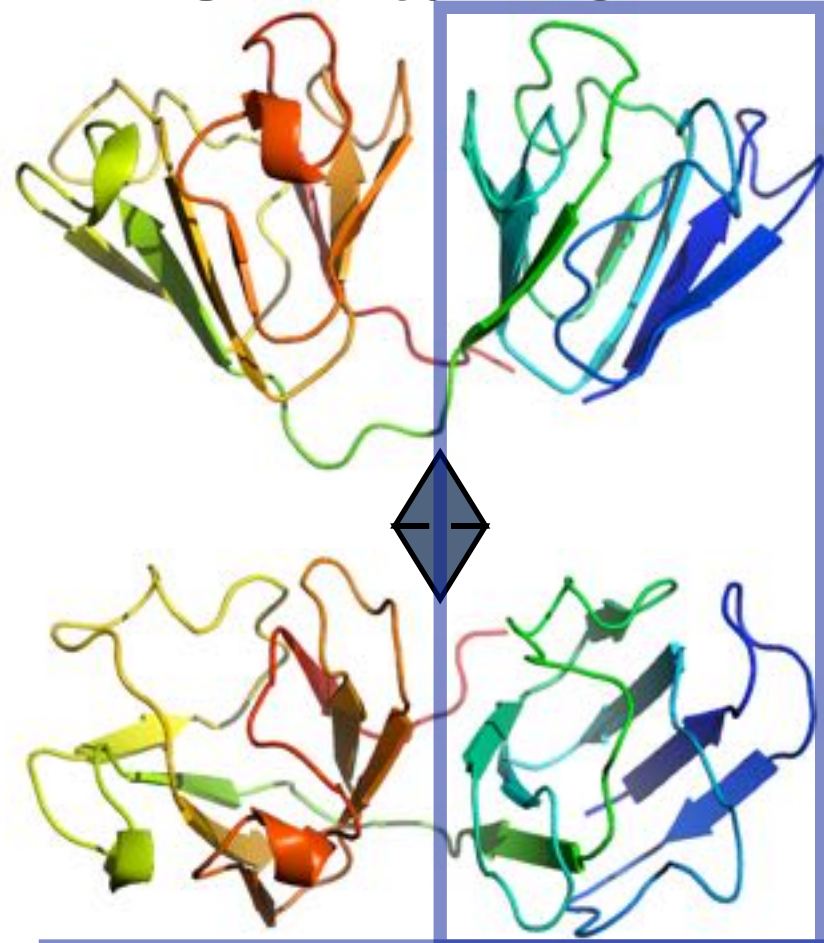
- The β sheets are arranged cyclically around the axis of the “propeller”, and these loops at the top of the barrel.

- Collectively, these loops form a wide funnel-shaped pocket containing the active site.

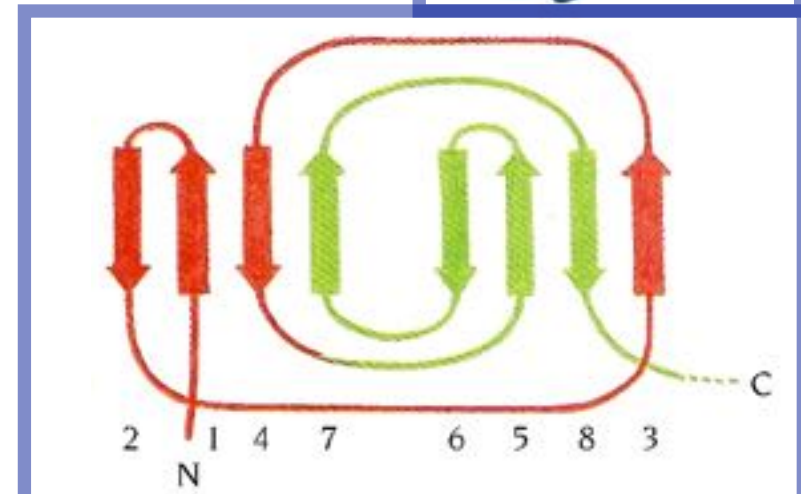


γ-crystallin and Domains

- Crystallins are located in the lens of the eye, and they contribute to the optical properties of the lens.
- Three classes of crystallins have been discovered (α , β and γ) -- and there may be more.
- The α and β crystallins are heterogeneous assemblies of different subunits, while γ -crystallins are monomeric \sim 170-residue proteins.
- The crystal structure of a γ -crystallin reported by Blundell, et al. reveals that the protein is folded into two distinct domains that are similar in size and are connected by a flexible tether.

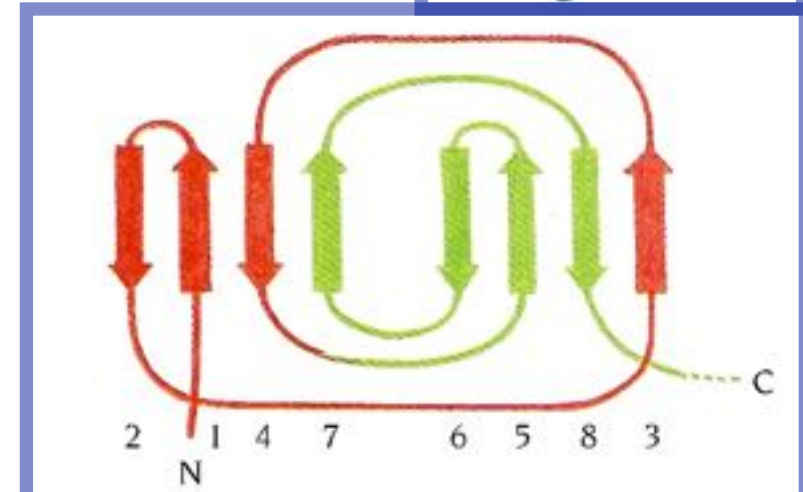
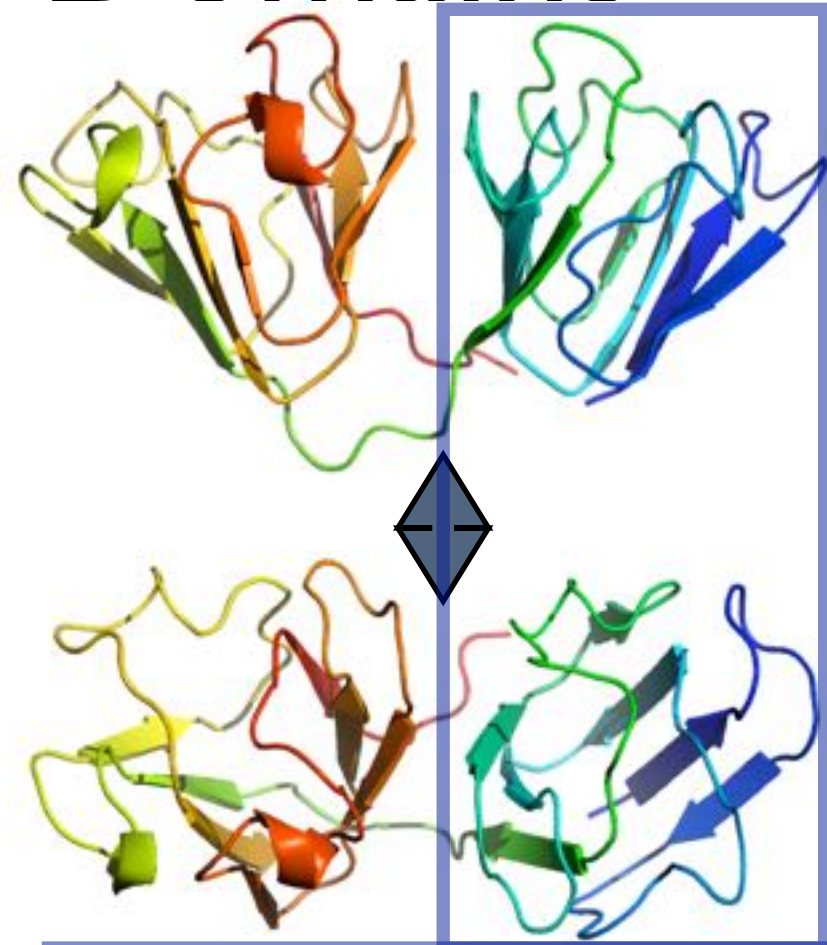


- The secondary structure of the N-terminal domain is dominated by β strands, which are arranged into two antiparallel sheets (Sheet: 2, 1, 4 and 7 & sheet: 6, 5, 8 and 3).
- While strands 6 and 7 are adjacent in sequence they share no hydrogen bonds.



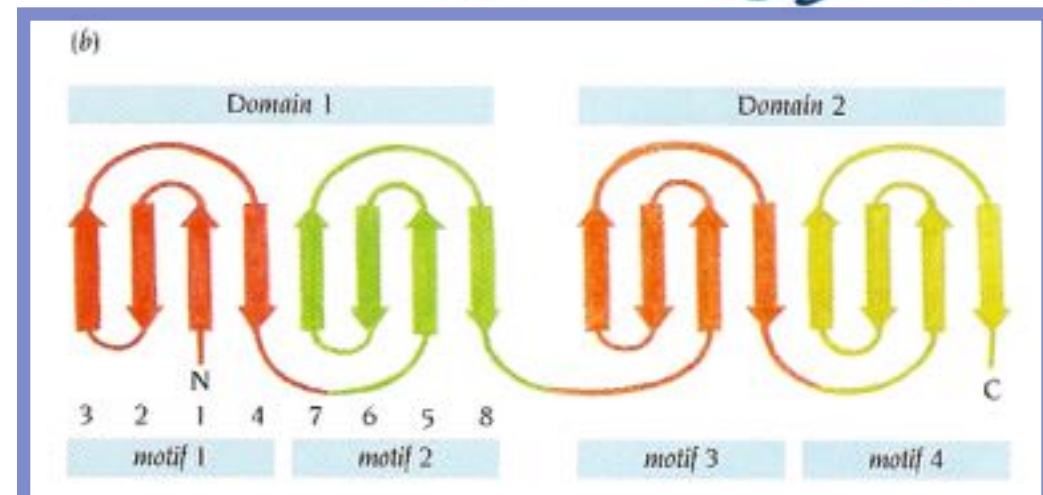
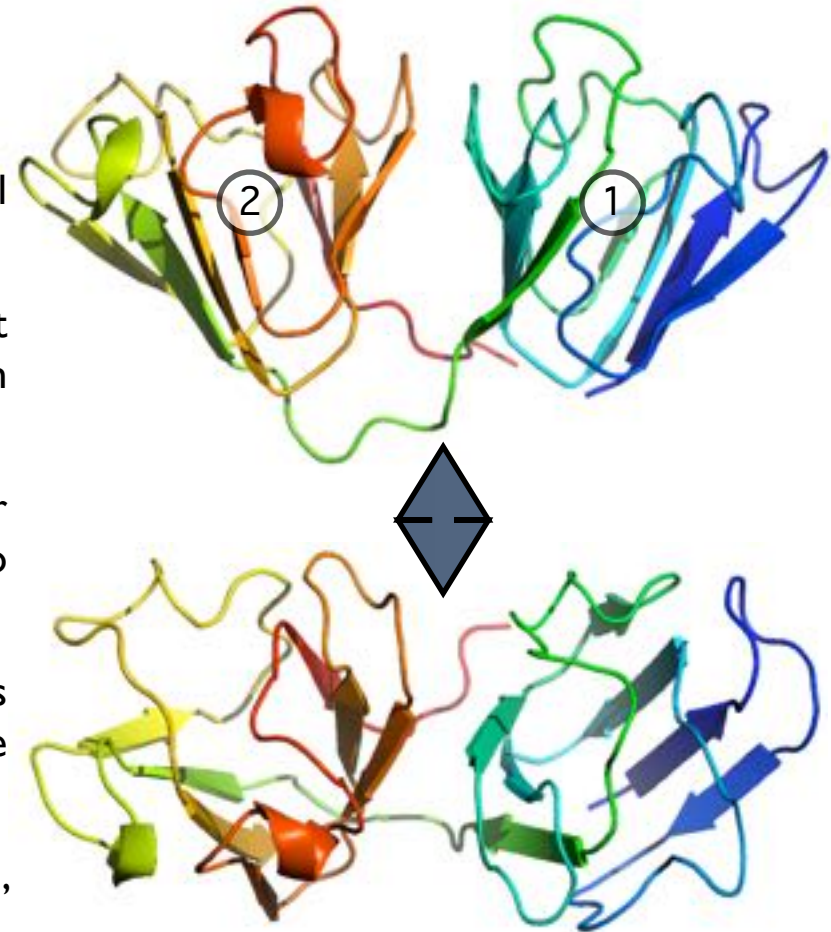
γ -crystallin and Domains

- The highlighted N-terminal domain consists of two four-stranded β sheets. The arrangement of the β strands of this domain is illustrated in the provided topological diagram.
- The two β sheets in this domain are packed against each other, forming a distorted barrel structure.
- Strands 1, 2, 3 and 4 form a Greek key motif, as do strands 5, 6, 7 and 8.
- The N-terminal and the C-terminal domains have identical topologies, and very similar structures.



γ -crystallin and Domains

- The N-terminal and the C-terminal domains have identical topologies, and very similar structures.
- While the structural similarity of the two domains is not immediately obvious from the crystal structure, the similarity in their topologies is clearly evident in the topology diagram.
- In the folded protein, the polypeptide chain is divided into four consecutive Greek key motifs, grouped pairwise in the two domains.
- Overlaying the C α atoms of residues in the two domains results in a mean deviation of $<2\text{\AA}$, indicating that the two domains are structurally equivalent.
- The motif structures within each domain superpose equally well, but they have lower sequence homology.
- Similarity between the four Greek-key motifs suggests that they are evolutionarily related.
- Walter Gilbert in 1978 suggested that genes for larger proteins may have evolved by the accidental juxtaposition of exons coding for specific functions.



Jelly Roll Barrels

- In antiparallel barrel structures built from the Greek key motif, one of the connections in the motif crosses one end of the barrel.
- The jelly roll motif is a common motif in which four of the inter-strand connecting segments cross the ends of the barrel.
- The jelly roll motif is found in a variety of proteins including coat proteins of most spherical viruses, concanavalin A (a plant lectin) and influenza virus hemagglutinin.
- In the jelly roll motif, the polypeptide chain has 8 β strands connected by loop regions.
- Essentially the 8 β strands are arranged in a long antiparallel hairpin such that strand 1 is H-bonded with strand 8, and strand 2 is paired with strand 7... and so on.
- The extended hairpin is then folded into a barrel configuration with the β strands forming the sides of the barrel and the connecting segments crossing both the top and the bottom of the barrel.

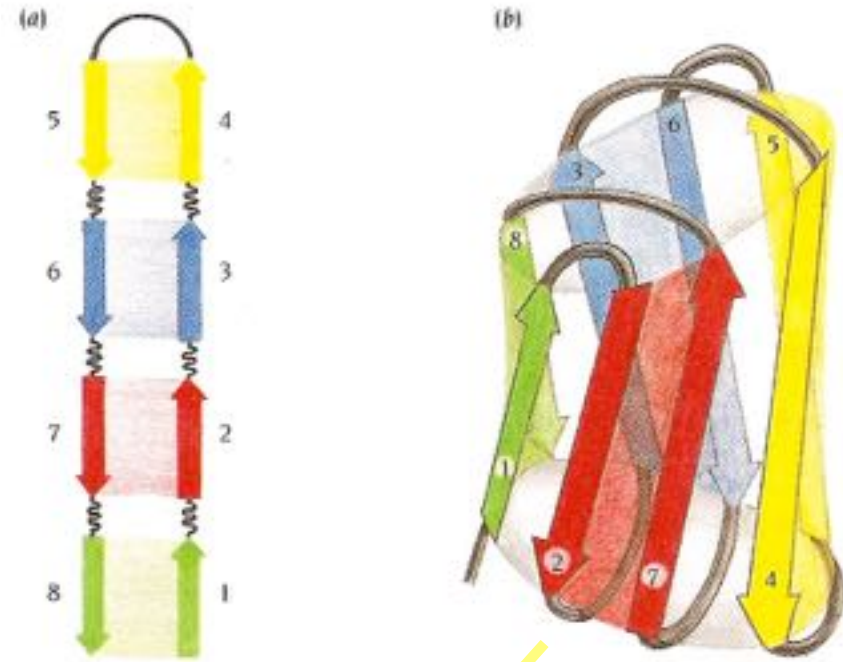
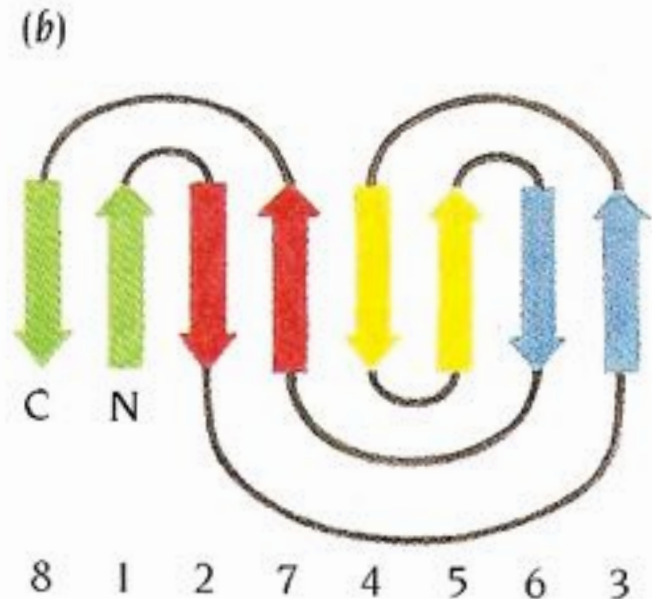
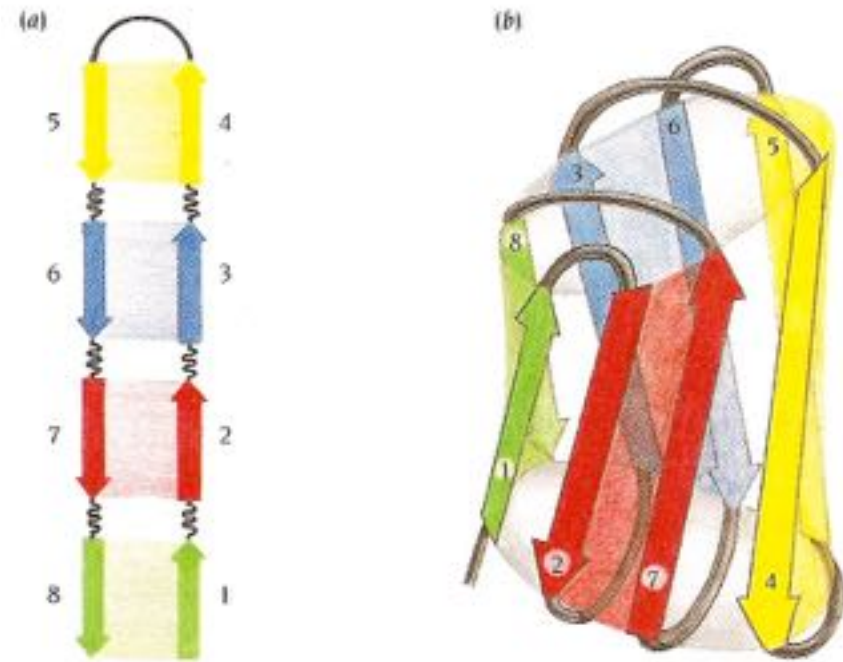


Figure 5.17 A simple illustration of the way eight β strands are arranged in a jelly roll motif. (a) The eight β strands are drawn as arrows along two edges of a strip of paper. The strands are arranged such that strand 1 is opposite strand 8, etc. The β strands are separated by loop regions. (b) The strip of paper in (a) is wrapped around a barrel in the same way as the string in Figure 5.16, such that the β strands follow the surface of the barrel and the loop regions (gray) provide the connections at both ends of the barrel. The β strands are now arranged in a jelly roll motif.

Jelly Roll Barrels

- In the barrel structure, antiparallel H-bonded β strand pairs 1:8, 2:7, 3:6 and 4:5 are arranged such that β strand 1 is adjacent to strand 2, strand 7 is adjacent to 4, 5 to 6 and 3 to 8.
- In the barrel structure, all adjacent strands are aligned antiparallel.
- This barrel arrangement results in two connecting loops (3-4 and 7-8) crossing the top of the barrel and two (2-3 and 6-7) across the bottom of the barrel.
- The described arrangement is for a 8-stranded jelly roll barrel, but a jelly roll barrel can contain any even number of β strands greater than 4. (8-stranded barrels are most common).

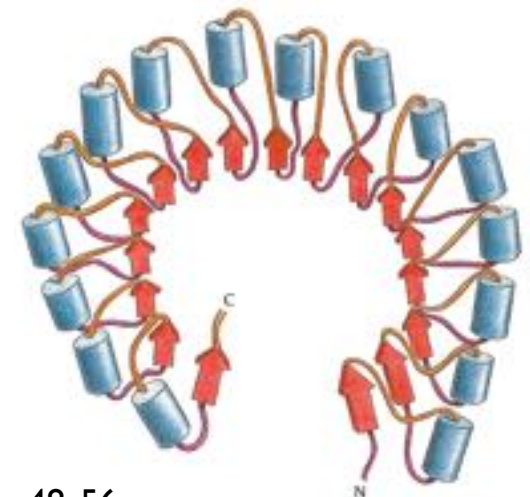
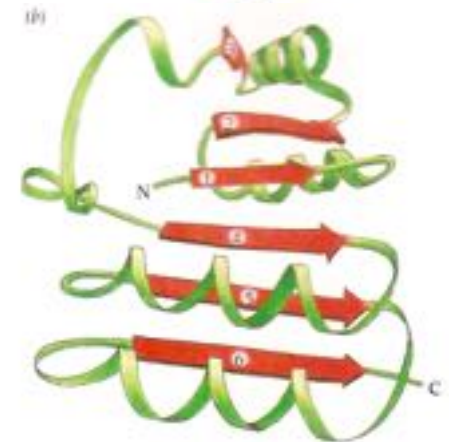
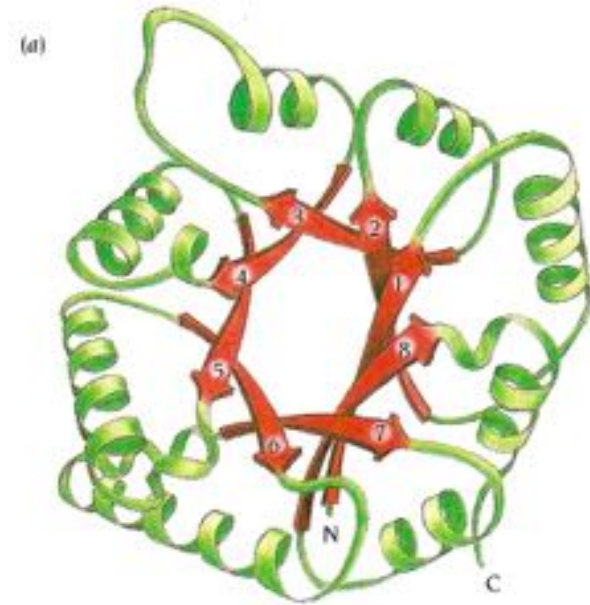
- Analysis of the H-bonding patterns in jelly roll barrels reveals that they usually can be broken down to two sheets with few if any H-bonds between strands belonging to the different sheets.
- Barrel is distorted from the ideal, with gaps separating two pairs of adjacent β strands in the barrel.
- The β strands in jelly roll barrels are often arranged in two sheets that are packed against each other.



α/β Structures

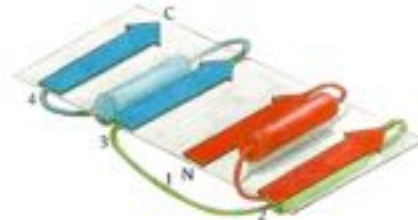
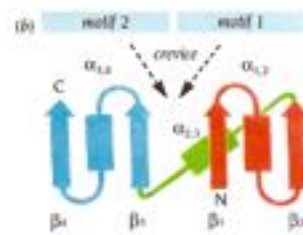
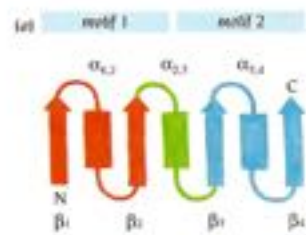
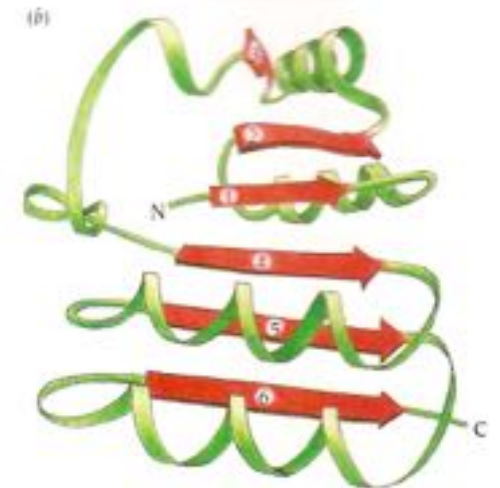
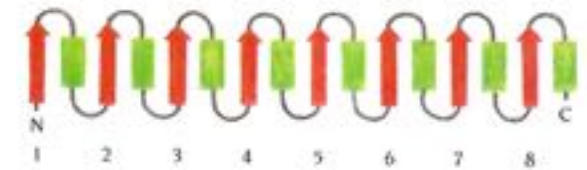
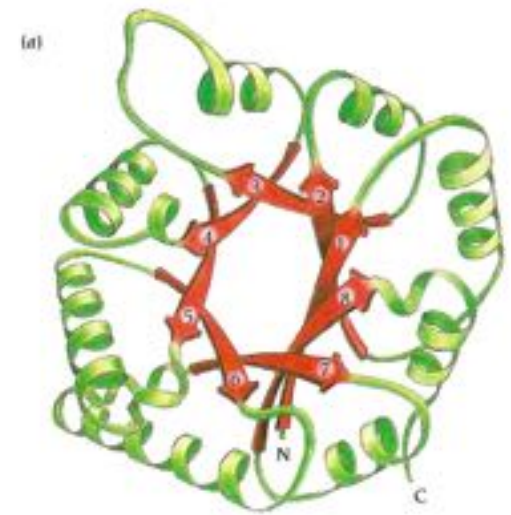
α/β Structures

- In known proteins, α/β domains, consisting of a central parallel or mixed β sheet surrounded by α helices, are the most common domain structures.
- In most α/β domains, binding/catalytic sites are formed by loop regions.
- Three main classes of α/β domains:
 1. α/β barrel - central β sheet forms a barrel structure with all β strands twisting around like barrel staves. Connecting α helices are on outside of the barrel.
 2. Twisted β sheet flanked on both faces by connecting α helices.
 3. Leucine-rich motif - contain repetitive regions with a conserved pattern of Leu residues. Consists of a central twisted β sheet with α helices on outer surface (horseshoe fold).
- All are assembled from β - α - β motifs.



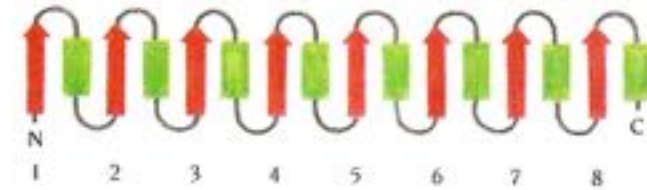
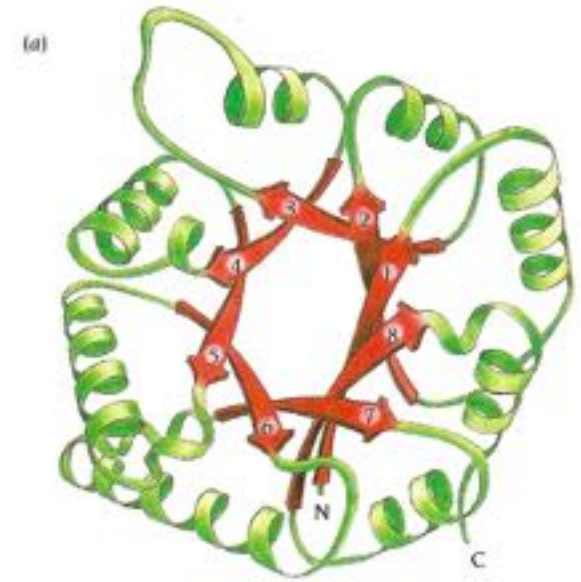
α/β Structures

- All α/β structures are assembled from β - α - β motifs.
- Two fundamentally different ways that two β - α - β motifs can be arranged to form a 4-stranded parallel β sheet.
 1. In the barrel and horseshoe structures, consecutive β - α - β motifs connected in such a way that the motifs are similarly oriented.
 2. In the open-twisted sheet, consecutive β - α - β motifs are connected in such a way that the second β - α - β motif is flipped and turned around resulting in α helices being placed on both faces of the twisted β sheet.
- Nearly all β - α - β motifs are right-handed.



α/β Barrel

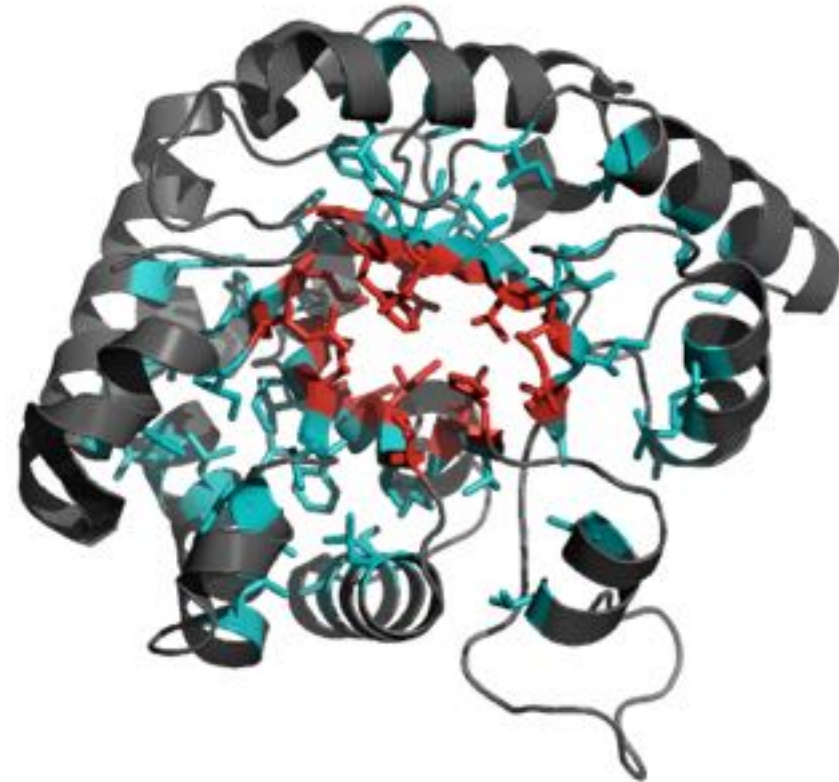
- In α/β structures where the strand order is 1 2 3 4, all connecting helices lie on the same face of the sheet. Such an arrangement would leave the other face of an isolated open twisted sheet exposed to solvent.
- Usually forms a closed barrel (assembled from twisted β strands) with connecting α helices on the outside of the barrel.
- More than four β strand are required to form a closed barrel (usually 8 strands and sometimes 10 strands).
- In almost all α/β barrels, the cross-connections between strands consist of α helices, with there often being an additional α helix after the last strand.
- At a minimum of 200 residues, the eight-stranded α/β barrel is one of the largest domain structures.
 1. Very common structure, found in many different proteins. Amino acid sequence varies significantly, but the core α/β barrel structure is conserved.
 2. Depending on protein, connecting loop regions can have varied lengths and amino acid composition. (in some loops may fold into distinct domains.



Triosephosphate Isomerase from
Giardia lamblia [PDB#2DP3]

α/β Barrel Core

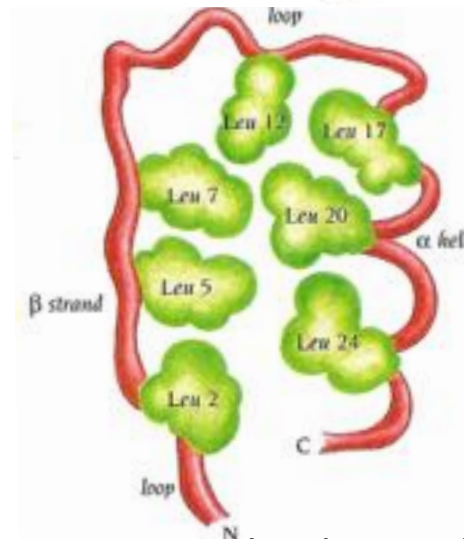
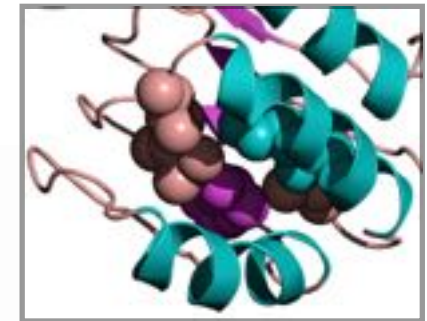
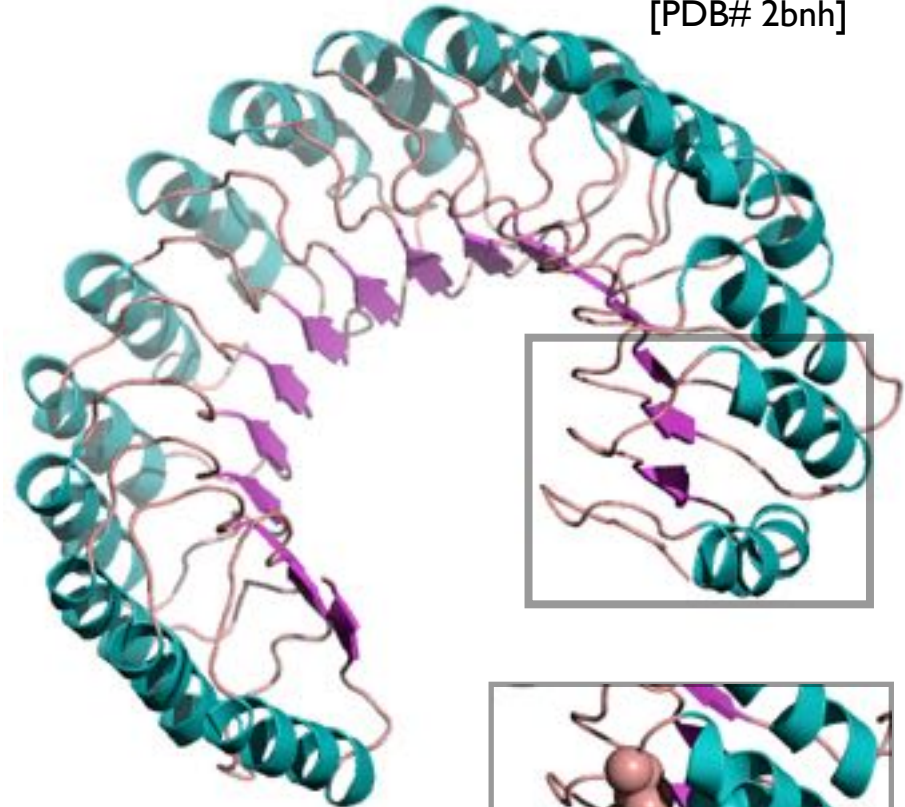
- In α/β barrels, hydrophobic side chains from the α helices pack against hydrophobic side chains on the outer surface of the β sheet.
- As is characteristic of all β sheets, the side chains of consecutive residues are oriented on opposite faces of the β sheet.
- Every second residue contributes to the hydrophobic surface that packs against the helices.
- Other side chains of the β strands are directed inwards towards the hydrophobic core of the barrel.
- Packing interactions between α helices and β strands are dominated by residues with branched hydrophobic side chains (i.e. V, I and L account for $\sim 40\%$ of residues in the β strands in the parallel β sheets).
- Bulky hydrophobic residues from positions 1, 3 and 5 of the β strands fill the interior of the barrel. Form a tightly-packed hydrophobic core.
- Polar and charged residues (R, K and Q) terminate some of the strands. The hydrophobic/aliphatic portion of these side chains of residues in these positions contribute to hydrophobic core.



Leucine-rich Motifs

Ribonuclease inhibitor
[PDB# 2bnh]

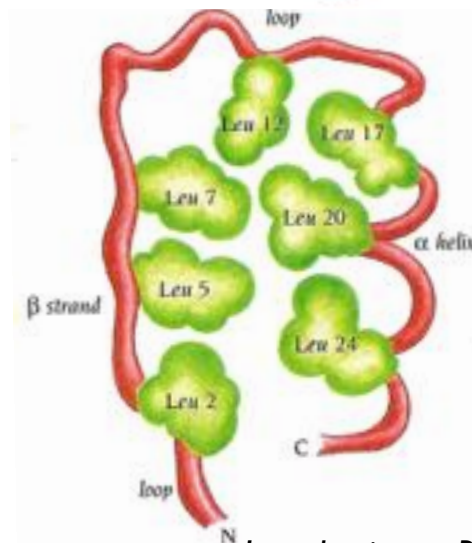
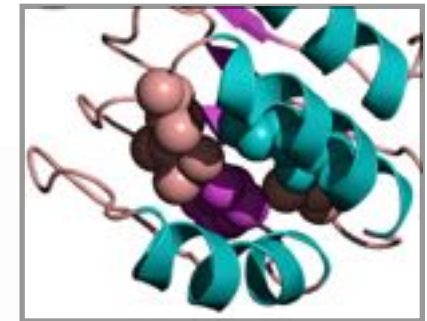
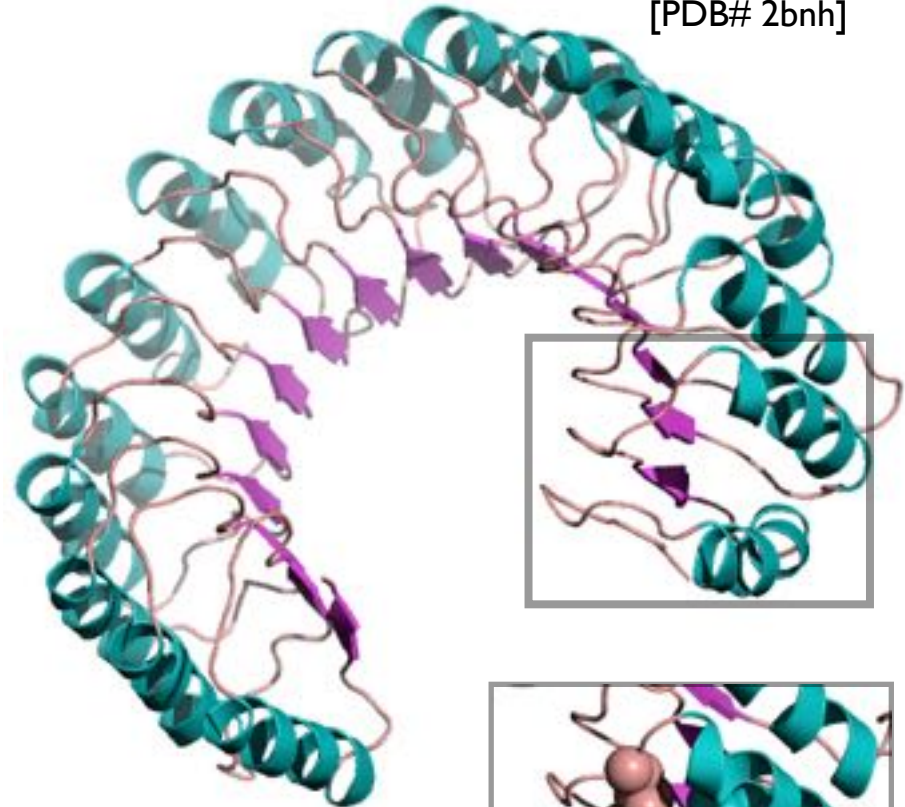
- Leucine-rich motifs consist of tandem repeats of homologous 20-30 residue right-handed β -loop- α structures.
- Such leucine-rich motifs have been identified in over 60 different proteins ranging from receptors, cell adhesion molecules, bacterial virulence factors and proteins involved with RNA splicing and DNA repair.
- In ribonuclease inhibitor, a polypeptide chain consisting of 456 amino acid residues has been arranged to form a tandem repeat of 15 leucine-rich motifs (2 flavors: type A with 29 residues and type B with 28). --also contains two short regions with nonhomologous sequences at the termini--
- Both type A and type B demonstrate similar pattern of Leu.
- Sequential β -loop- α repeats are joined together in similar way to those in the α/β barrel structures.
- The strands form a parallel β sheet - interior surface of a curved open structure, a “horseshoe”.



Leucine-rich Motifs

Ribonuclease inhibitor
[PDB# 2bnh]

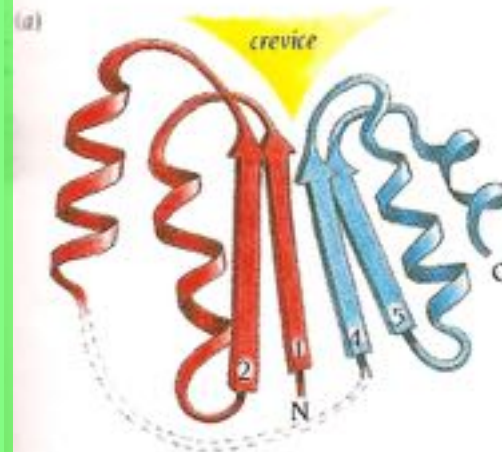
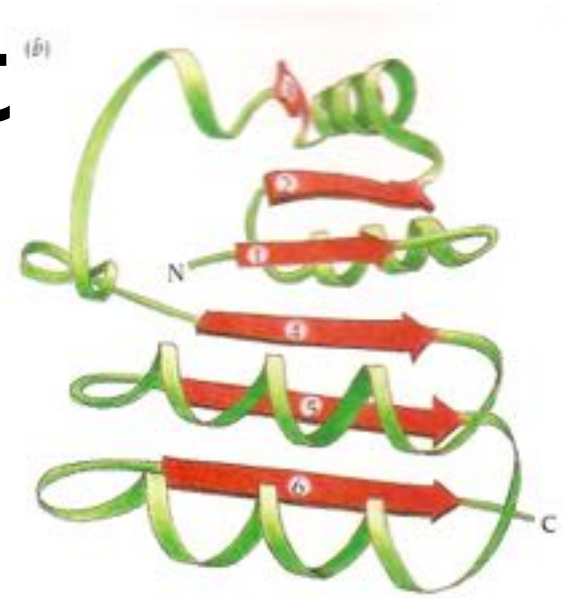
- With the β strands lining the inner surface and α helices adorn the outer surface of the open curved structure.
- α helices are aligned antiparallel to the β strands.
- Hydrophobic packing between outer face of β strands and the inner surface of the α helices. Unlike α/β barrels, the inner surface is exposed to solvent.
- The Leu residues of the leucine-rich motifs form the hydrophobic core between the strands and the helices.
- Leu residues at positions 2, 5, 7, 12, 20 and 24 are invariant between type A and Type B repeats. Examination of more than 500 tandem repeats from 68 different proteins indicates that residues at positions 20 and 24 may be other hydrophobic residues (i.e. Ile and Val).
- Ribonuclease inhibitor has been used as the basis for constructing plausible models for other proteins with leucine-rich motifs.



α/β Twisted Open Sheet^(b)

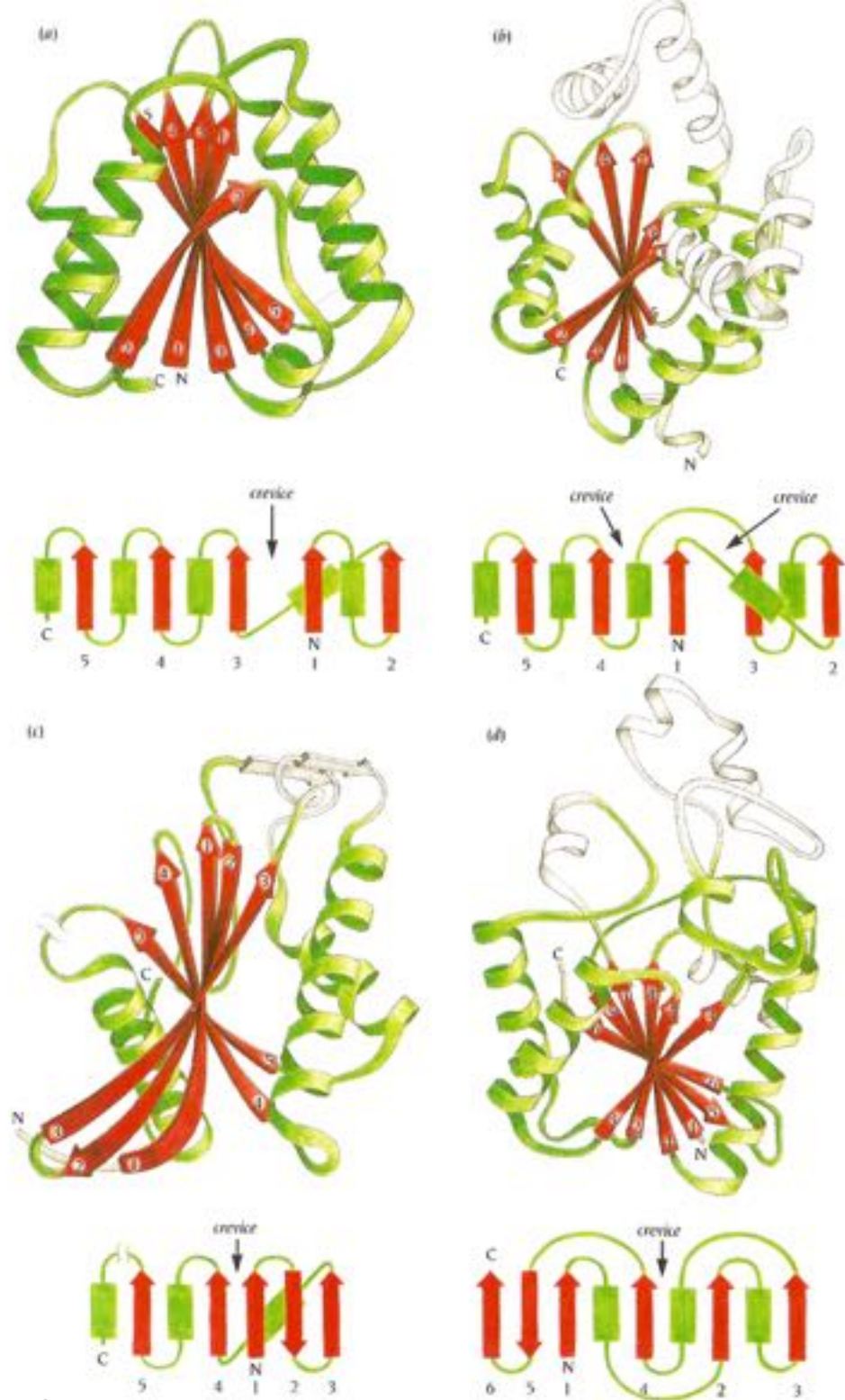
- In α/β twisted open-sheet structures, the connecting α helices are positioned on both faces of the β sheet.
- This arrangement:
 1. Precludes formation of a barrel structure.
 2. Results in two adjacent β strands in the interior of the sheet (i.e. β_1 and β_4) whose connections to the flanking β strands are on opposite faces of the sheet.
 3. α helices are packed against both faces of the sheet with each β strand contributing to hydrophobic side chains to pack against α helices in two similar core regions --- one on each side of the β sheet.

- One of the connecting loops from one of these two strands goes above the sheet, while the other loop goes below the sheet.
- Pattern of connecting segments results in a crevice outside the edge of the β sheet between these loops (β_1 and β_4).
- Almost all binding sites in α/β twisted sheets occur in such crevices found at the carboxy edge of the sheet.



α/β Twisted Open Sheet Topologies

- While α/β -barrel domain structures show the same basic arrangement of 8 α helices and 8 β strands, open β -sheet structures show a variety of topologies
- Because the β strands form an open twisted β sheet, there are no geometric restrictions of the number of strands. [generally ranges from 4-10 strands]
- Two consecutive strands joined by a crossover connection need not be adjacent to each other in the folded β sheet. [β - α - β motif with strands being adjacent is preferred]
- Allows for mixed parallel/antiparallel β sheets that incorporate β hairpin connections.



End Lecture 2 Notes

Hemoglobin

- Hemoglobin is a tetrameric protein, consisting of two different kinds of polypeptide chains (2 α subunits and 2 β subunits).
- Single erythrocyte contains hemoglobin at a concentration of 340 mg/ml.
- Sickle-cell anemia is an inherited disease involving a Glu→Val mutation at a surface position (6th residue) in hemoglobin β subunits.
- The position is solvent exposed tetrameric protein.
- Mutation promotes polymerization of deoxygenated hemoglobin within erythrocytes (does not promote aggregation of oxygenated form).
- Hemoglobin aggregate fibers give erythrocytes their characteristic cycle shape.
- Lethal for homozygotes for disease.
- Gives increased resistance to malaria.

