Biophysical Properties

Forces Stabilizing Macromolecular Structure

- Covalent Bonds: Disulfide Bonds (see discussion of Cys residues)
- Noncovalent interactions are key biological forces:

	Distance Dependance	Approx. Bond Energy
★ Hydrophobic Effect	No simple distance	<10 kJ/mol
★ Electrostatic Forces:		
- Charge-Charge (ionic) Interactions	$\propto 1/r^2$	<20 kJ/mol
- Hydrogen bonding	No simple expression	<10 kJ/mol
- Van der Waals	$\propto 1/r^6$	<5 kJ/mol

- With the exception of disulfide bonds, these forces are transient in nature.
- Several factors influence the strength of noncovalent interactions.
  - ★ Electrostatic interactions are greatly impacted by environment, being stronger in nonpolar environments (protein interior) and weaker in polar environment (water)
- Individually noncovalent interactions are weak (C-C bond ~80 Kcal/mol or ~335 kJ/mol), but they add up and collectively can be very strong.

Conformational Landscape

- The tertiary fold of a protein reflects the global energy minimum... the most stable state of the protein.*
- Often represented using a conformational landscape, with conformational freedom in the x-axis and free energy in the y-axis.
- The folded "conformation" of globular proteins likely represent a large population of related/similar folded conformations.

  Globular proteins likely interconvert between these conformations.
- Current opinion is that the observed folded conformation is the most stable kinetically accessible conformation. Other more stable conformations may exist, but are not accessible under environmental conditions.

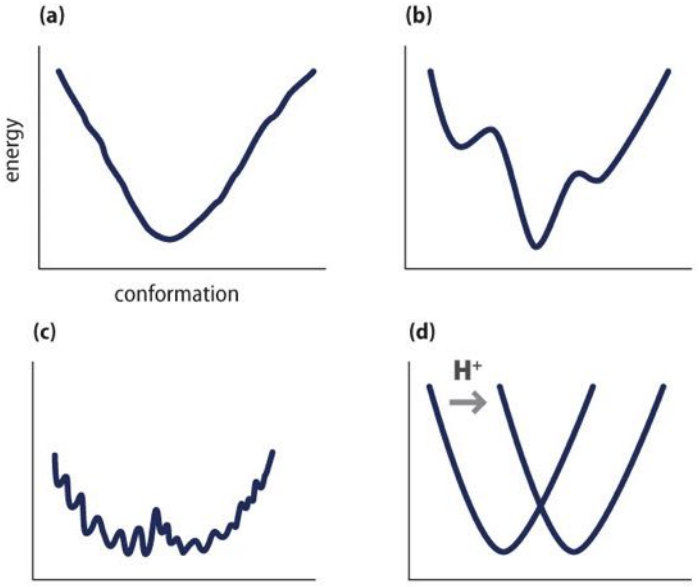


Figure 1.45 How Proteins Work (©2012 Garland Science)

(a) Smooth landscape with a single well-defined energy minimum.

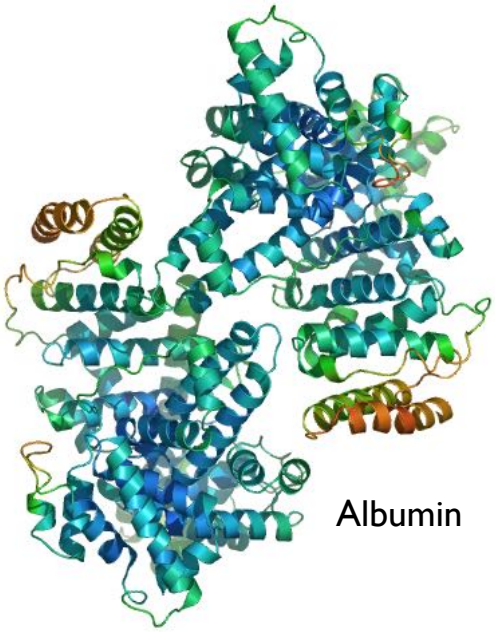
(b) Landscape has higher-energy metastable states w/ low energy barriers to the global minimum.

(c) Locally, the surface of the landscape is very rough, with energy barriers low enough that proteins at room temperature can cross them rapidly.

(d) A change in conditions causes a change in the structure of the global minimum.

Flexibility of Protein Structure

- Protein structures are not static. Both crystal structures and NMR indicate varying degrees of conformational freedom.
- Proteins can be thought of as existing in a range of distinct but closely related microstate conformations that interconvert rapidly at room temperature.
- On a longer time scale, larger backbone conformational movements can occur.
- On the longest time scales, the folded conformation is marginally stable and may transiently sample the unfolded state (10^{-4}-10^{-12}/s).
- Side chains of residues at the protein surface can have significant conformational freedom.
- Close packing of atoms in the protein interior is constraining and requires coordinated motions.



Albumin

Conformational Motility

- **Hydrogen Exchange:** Best evidence for extensive structural mobility is that internal groups in proteins react with appropriate reagents in solution. (buried groups either are occasionally at surface or reagent can permeate the protein)
  - Isotopic exchange with water (H_2O, ^2H_2O and ^3H_2O).
  - Hydrogen atoms covalently attached to various atoms exchange with solvent at different intrinsic rates, depending on tendency of that atom to ionize.
  - Exchange of amide protons most often studied because these hydrogen atoms exchange on a useful time scale.
  - Rates of exchange impacted by temperature, hydrogen bonding, environment and degree of exposure.
  - Rate of exchange of individual H's varies 100-fold.
  - Protons involved in hydrogen bonding in the interior of β-sheets and α-helices tend to exchange least readily.
  - Acid and base catalyzed exchange (via transient protonation of C=O and deprotonation of N–H respectively).
  - Rates of exchange generally increase at elevated temperatures, but in a complex manner.
  - Classical methods (NMR and MS) only provided insights into average number of protons exchanged.
  - Exchange of individual hydrogen atoms can be followed using ^{1}H-NMR.

Interior vs. Exterior

The surface area of a small monomeric protein is 23-45% of the surface area of the unfolded polypeptide.

Representations of Protein Structure

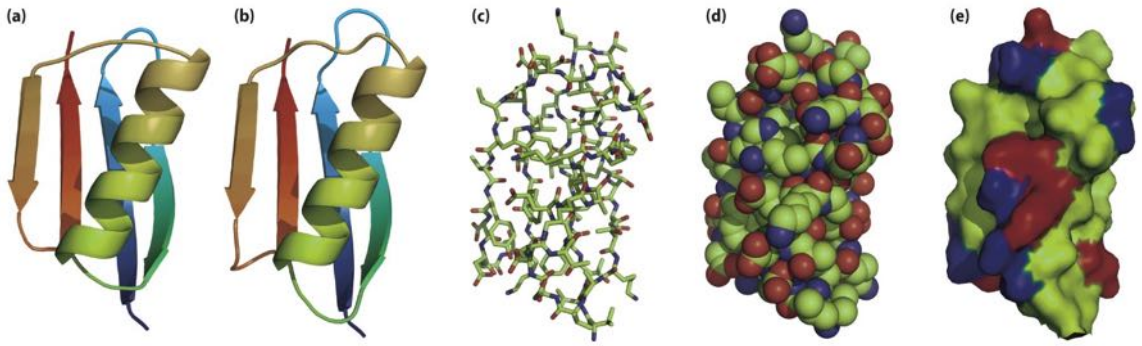
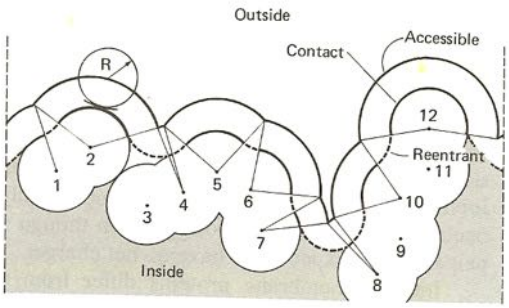


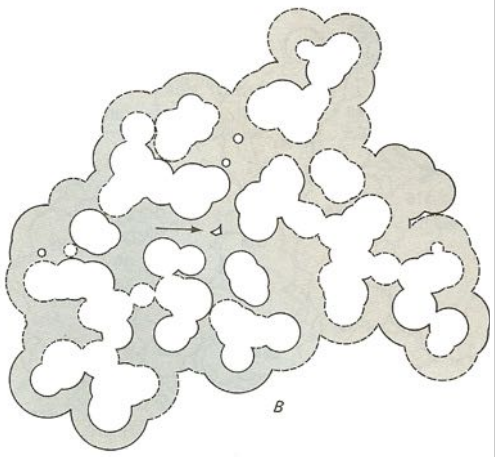
Figure 1.7.3 How Proteins Work (©2012 Garland Science)

- **Streptococcal protein G, B1 domain:** Different ways of representing protein structures.
- While cartoon/ribbon illustrations are commonly used to illustrate protein structures, other representations provide different perspectives and insights into protein structures.

Interiors and Exteriors

- Analysis of the complex surfaces of proteins:
- **van der Waals surface:** Every atom of the protein is depicted as a sphere corresponding to the appropriate van der Waals radius. Spheres are truncated for covalently bonded atoms.
- **Contact surface:** parts of the van der Waals surface that make contact with the surface of a spherical probe.
- **Reentrant surface:** where the probe is in simultaneous contact with more than one atom of the protein.
- **Molecular surface:** contact surface + reentrant surface.
- **Accessible surface:** defined by the center of the probe as it moves over the surface of the protein. Frequently, a sphere with a radius of 1.4Å is used as a probe, simulating a water molecule.
- The total accessible surface area of a protein is approximately proportional to the 2/3 power of their molecular weight.

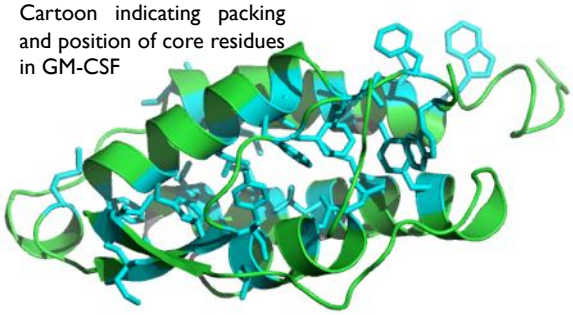




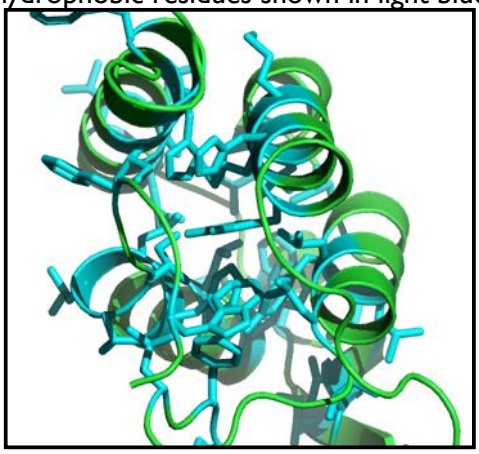
Interiors and Exteriors

- The interiors of proteins are densely packed with adjacent atoms frequently in van der Waals contact. (packing density in the protein interior may vary and is not uniform.)
- ~75% of the interior volume is filled with atoms. (compared to 70-78% for crystals of small organic molecules and 58% for water).
- Nonpolar groups predominate the protein interior (Val, Leu, Ile, Phe, Ala and Gly account for ~63%).
- Polar groups in the protein interior are paired in hydrogen bonds. Most are in the protein backbone.
- Main-chain polar groups (C=O and -NH of peptide bond) in the protein interior must be neutralized by backbone hydrogen bonding and formation of regular secondary structure.
- Water molecules are usually excluded from the protein interior, but when present are usually isolated from bulk solvent and form H-bonds with polar groups.

Cartoon indicating packing and position of core residues in GM-CSF



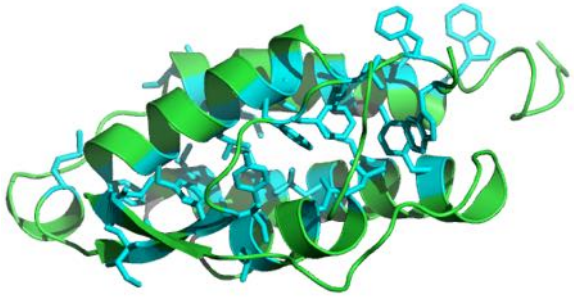
hydrophobic residues shown in light blue



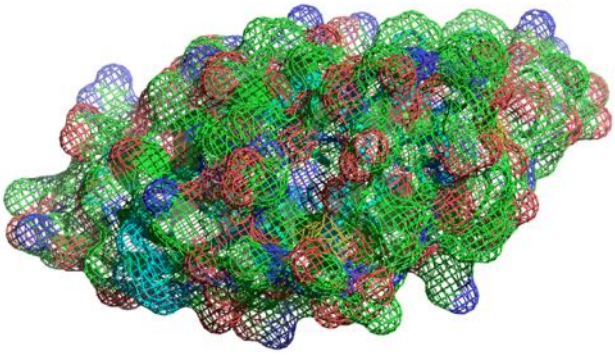
granulocyte macrophage colony stimulating factor (PDB# 2GMF)

Interiors and Exteriors

- Almost all ionized groups in water soluble proteins are on the surface of the protein (Asp, Glu, Lys and Arg residues account for ~27%).
- Charge density on the protein surface varies between 0.5 and 25 charged groups per 100Å^2.
- Oppositely charged groups are usually positioned near each other.
- Integral membrane proteins differ from water soluble proteins. They generally have extremely nonpolar surfaces where the protein comes in contact with the membrane interior.



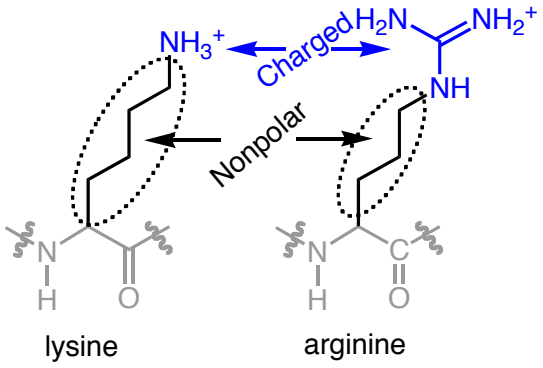
Mesh representation of protein surface. Colors reflect nature of atoms at surface (CHNOS)



granulocyte macrophage colony stimulating factor (PDB# 2GMF)

Interiors and Exteriors

- It is often difficult to classify residues as being buried or exposed.
  - ★ The ionized terminal groups on the long side chains of Lys and Arg are almost invariably exposed to solvent, while the nonpolar methylene groups are often buried.
- Even in large proteins, only ~15% of the residues are totally inaccessible to solvent.
- A residue is usually considered buried if more than 95% of its surface area is inaccessible to solvent.
- The hydrophobic residues are primarily involved in packing together secondary structure elements.
- No simple rules appear to relate conformation to amino acid sequence.



lysine arginine

Domains in Globular Proteins

Levels of Organization

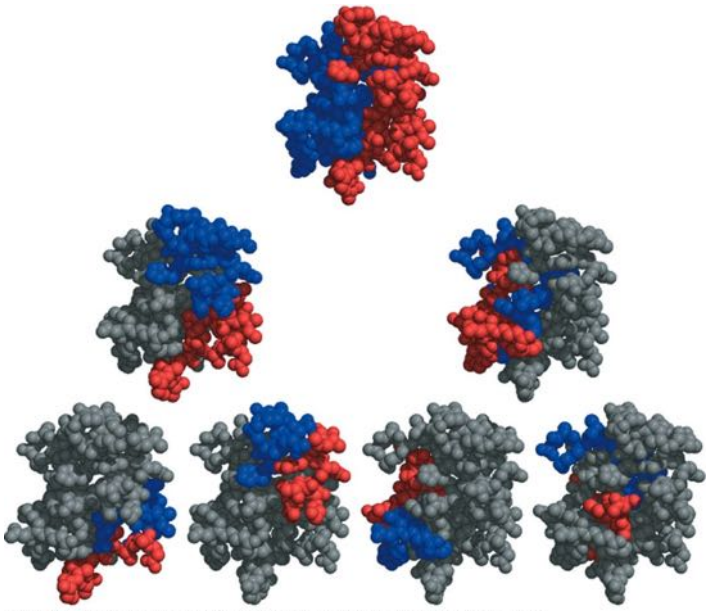
- **Primary structure (1° structure):** the amino acid sequence of polypeptide chain.
- **Secondary structure (2° structure):** local spatial organization and arrangement of the peptide backbone. Generally refers to easily localized structural elements (i.e. helices and sheets).
- **Tertiary structure (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.

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).
  - Supersecondary Structure/Motifs
  - Subdomains
  - Domains
- **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

Hierarchal Organization of Proteins

- **Domains**: large protein subunits consisting of contiguous, compact and physically separable segments.
- **Subdomains** and **sub-subdomains**: smaller discreet structural subunits (combine to form domains and subdomains respectively). Are not generally stable independently.
- **Supersecondary structure/motifs:** combinations of secondary structure. Reflect particularly stable arrangements of secondary structure elements.
- Consistent with the observation that most hydrogen bonding in proteins occurs locally.



Courtesy of George Rose, The Johns Hopkins University of Medicine, and Robert Baldwin, Stanford University School of Medicine

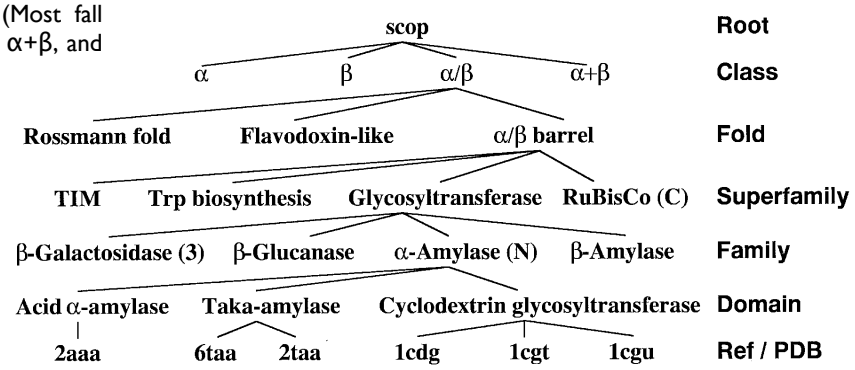
Protein Databases

- Proteomics is still in its early stages and scientific research is already struggling to develop methods to handle the vast amounts of data being generated on top of the information from genomics (enter - bioinformatics).
- Protein databases have been established in an attempt to organize and categorize protein structural and sequence information. (archived at Protein Data Bank)

  http://www.rcsb.org/pdb/home/home.do
- Such databases organize data in hierarchical arrangement -- important in analysis of evolutionary relationships.

SCOP

- SCOP (Structural Classification of Proteins) database contains all known protein structures - sorted according to folding pattern - composition and distribution of 2° structure.
- Hierarchal classification system:
  - Class: Secondary structure composition and packing within a protein (Most fall within five classes - α, β, α/β, α+β, and multi-domain).
  - Common Fold: Share major secondary structure elements, arranged the same and with the same topological connections.
  - Superfamily: Low sequence similarities but structural (and sometimes functional) features suggest probable common evolutionary origin.



  - Family: Grouped into families based on whether they share significant sequence similarity (≥30%) or whether they have very similar structures and/or activities (in the absence of strong sequence similarity)
- Primarily manual classification and somewhat subjective.
- Such databases organize data in hierarchical arrangement -- important in analysis of evolutionary relationships.

http://scop.mrc-lmb.cam.ac.uk/scop/
http://en.wikipedia.org/wiki/Scop

[17]Figure from: *Nucleic Acids Research*, 1997 vol. 25, no. 1, pp 236-239

CATH

http://www.cathdb.info/

- CATH database attempts to classify protein folds based on hierarchal system:
  - Class: secondary structure composition and packing within a protein (Three major classes - α, β, and α-β).
  - Architecture: describes overall shape of the domain - orientations of secondary structure elements (such as β barrel and β-α-β sandwich) - ignores connectivity of secondary structure elements. [done manually]
  - Topology (Folds): assigned based on overall shape and connectivity of secondary structures.
  - Homology: brings together evolutionarily related protein domains. Similarities identified based on sequence then structural comparisons.
- Classification involves both automatic and manual methods.

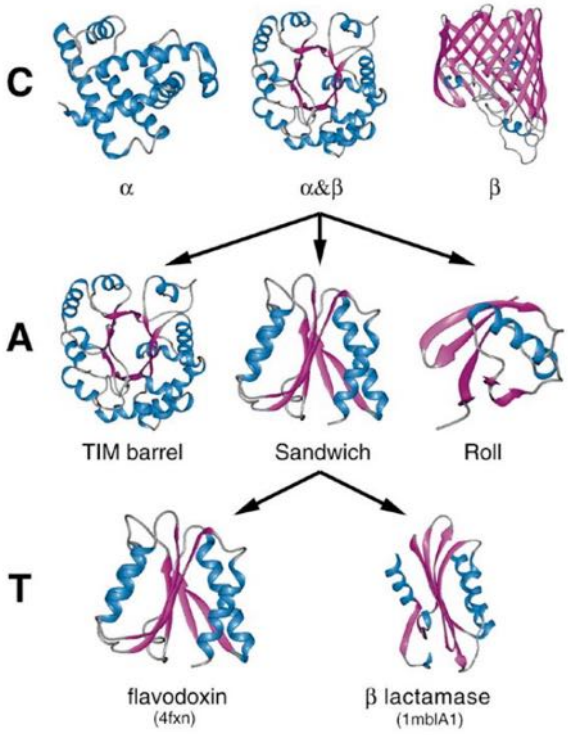
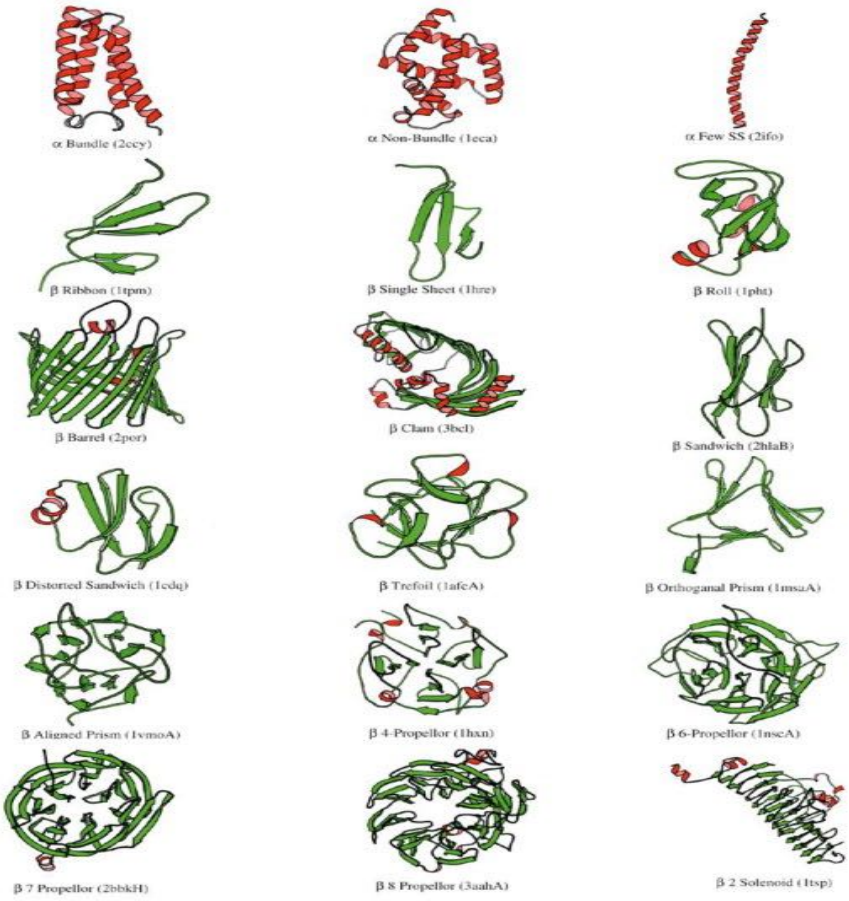


Figure from: *Structure*, 1997 vol. 5 no. 8, pp. 1093-1109.

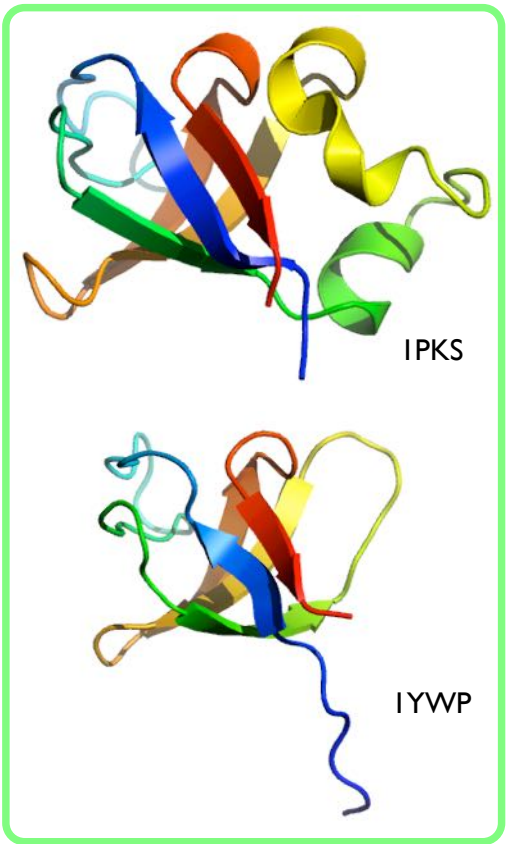


19

Figure from: *Structure*, 1997 vol. 5 no. 8, pp. 1093-1109.

Sequence Homology

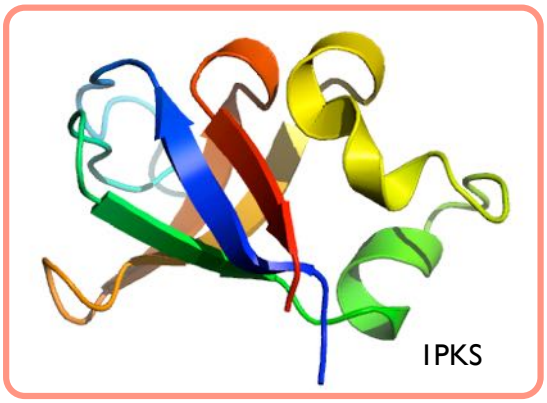
- When sequences are evolutionarily linked, the term homology is used to refer to sequence similarities.
- In order to determine homology it is necessary to establish rules governing potential similarity.
- Alignment of sequences is the first step in determining similarity between two or more sequences.
- Point mutations and larger mutational events occur that give rise to proteins containing different residues, which can obscure the relationships between proteins.
- Comparison to previously characterized proteins aids in the identification newly determined sequences (often from genomic data).
- Domains are key modular elements in proteins, and sequence alignments reveal that gene duplication leads to a proliferation of related domains in different proteins.
- Proteins can be related by the presence of similar domains - SH3 domain is a good example.



20

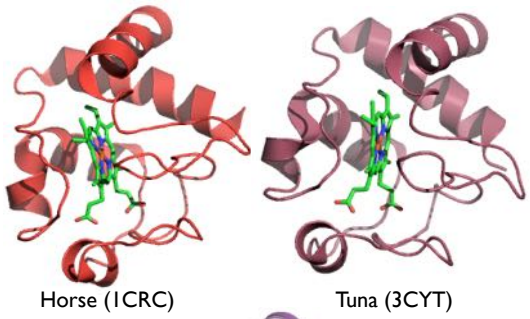
Sequence Homology

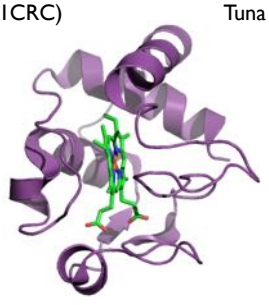
- Proteins can be related by the presence of similar domains - SH3 (Src Homology 3) domain is a good example.
- SH3 domains mediate protein-protein interactions - binding to Pro-rich peptide sequences.
- Present in a diverse range of proteins that have little else in common.
- Found in kinases, lipases, GTPases, structural proteins and regulatory proteins:
  - PI3 Kinase
  - CDC24 and CDC25
  - Ras GTPase activating protein
  - Phospholipase
  - Vav proto-oncogene
  - ZAP70
  - GRB2
  - ... and others

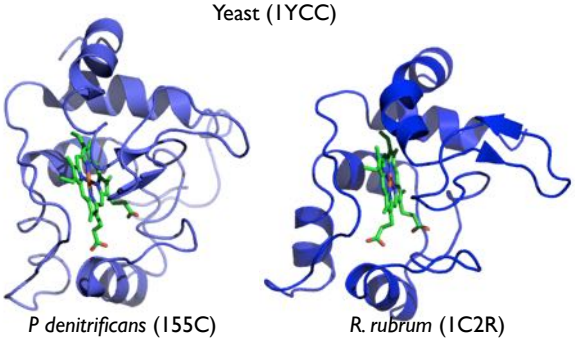


Structural Homology

- The diversity of folded conformations of proteins is less than would be expected based solely on the potential sequence diversity.
- Protein folds have been conserved throughout evolution despite changes in primary sequence.
- In most cases, structural similarities arise from sequence homology, but in some cases structural homology has been observed even though the evolutionary link is not clear.
  - Cytochrome C family of proteins is a good example of sequence and structural homology (see figure).
  - Only 18 residues conserved between cytochrome C from Horse, Tuna and Yeast mithochondria and cytochrome C_2 of *R. rubrum* and cytochrome C-550 of *P. denitrificans*.

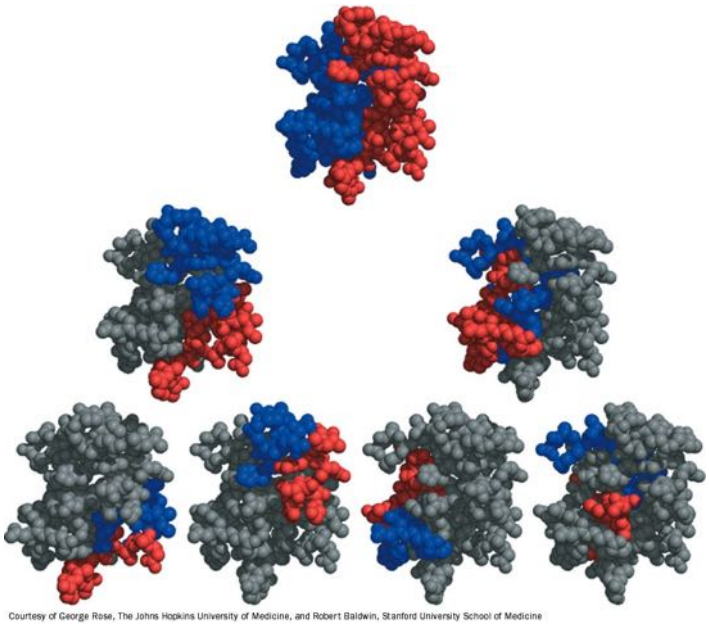






Hierarchal Organization of Proteins

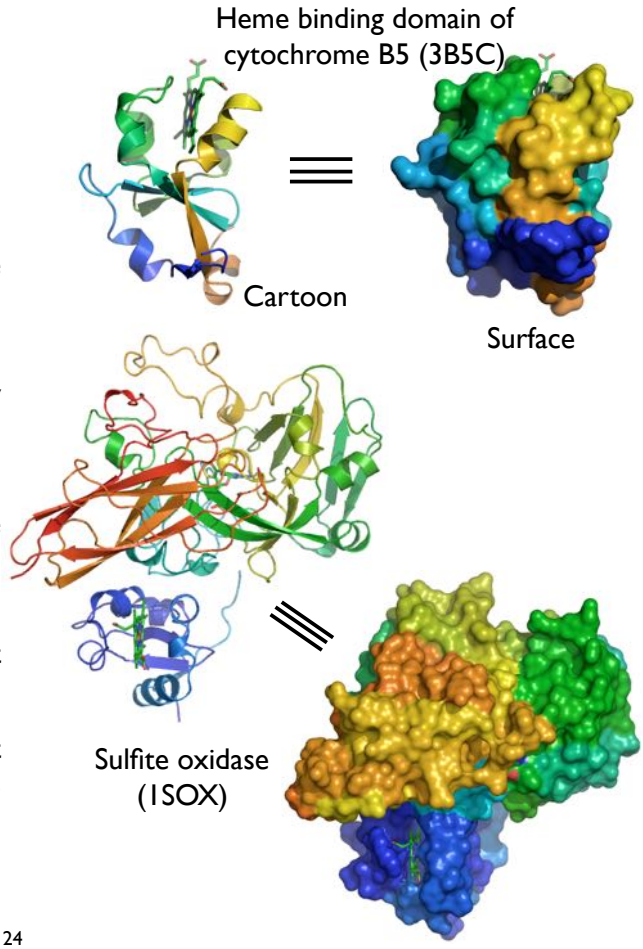
- **Domains**: large protein subunits consisting of contiguous, compact and physically separable segments.
- **Subdomains** and **sub-subdomains**: smaller discreet structural subunits (combine to form domains and subdomains respectively). Are not generally stable independently.
- **Supersecondary structure/motifs:** combinations of secondary structure. Reflect particularly stable arrangements of secondary structure elements.
- Consistent with the observation that most hydrogen bonding in proteins occurs locally.



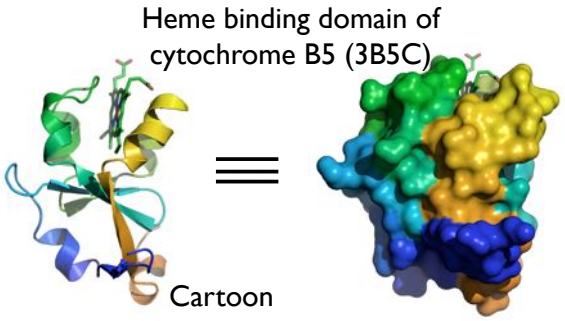
Courtesy of George Rose, The Johns Hopkins University of Medicine, and Robert Baldwin, Stanford University School of Medicine

Protein Domains

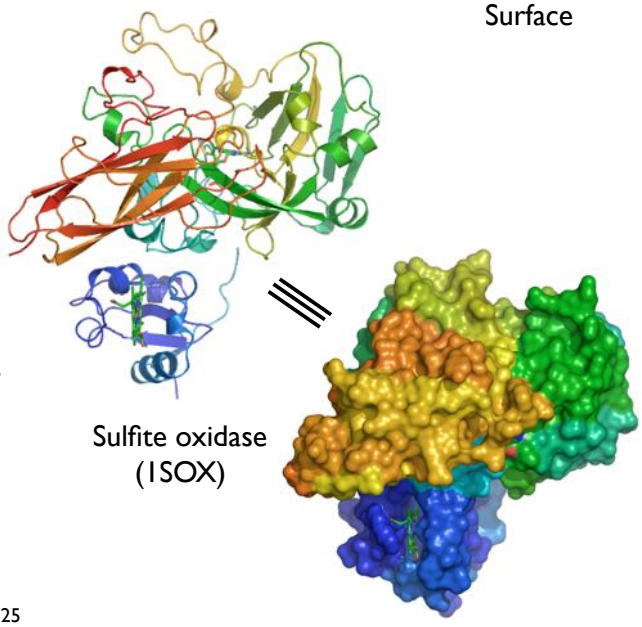
- The folded structures of most soluble (globular) proteins are roughly spherical with irregular surfaces and very compact.
- Proteins that have more than 150 residues may be organized around structural subunits known as **domains**.
- Domains are often associated with identifiable functions.
- The definition of a domain is not rigorous and is somewhat subjective. There are different approaches to identifying domains.
- Domains are assembled from different combinations of secondary structure elements and motifs.



Protein Domains

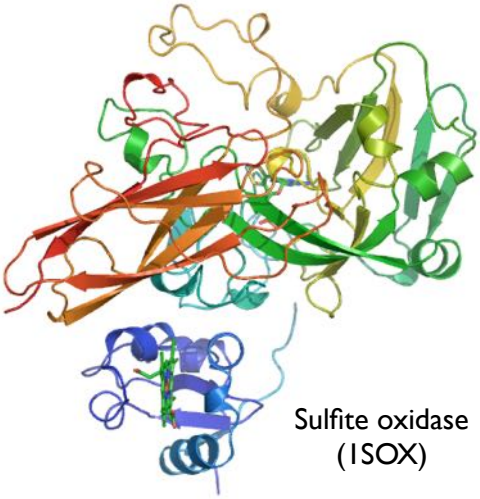


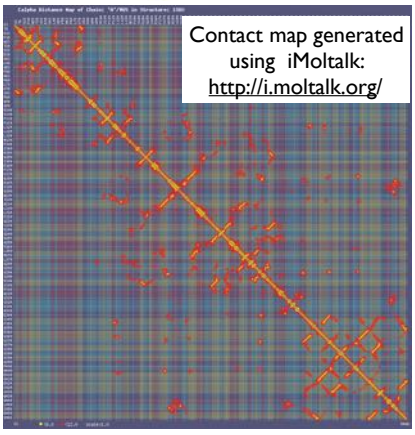
- Domains are generally structurally independent and have characteristics similar to small globular proteins (within a single polypeptide chain).
- The domains of a protein molecule interact with each other but to less of a degree than do the structural elements within each domain.
- Frequently a single segment of the polypeptide chain links domains and each domain is assembled from a single continuous stretch of polypeptide chain.



Protein Domains

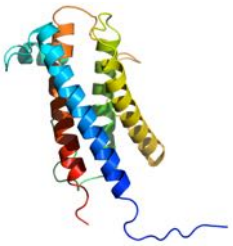
- *The definition of a domain is not rigorous and is somewhat subjective. There are different approaches to identifying domains.*
- Proteins and domains are often further subdivided into subdomains and folding units.
- The course of the polypeptide backbone through a domain is usually irregular, but generally follows a moderately straight course across the entire domain and then makes a U-turn and re-crosses the domain in a more or less direct but different path to the other side.
- These turns are usually at the protein/domain surface.
- Two-dimensional plots of the distances between the C_α of all pairs of residues *i* and *j* are useful descriptions of the overall folding of a polypeptide (contact maps or distance maps).
- The distances between all pairs of C_α in the protein are represented as contours, or only those a certain distance apart may be represented.



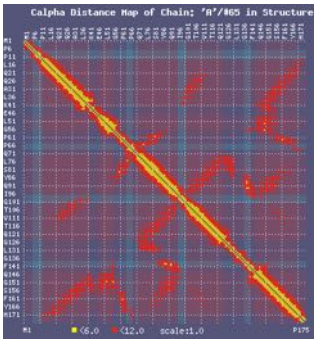


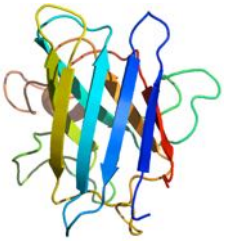
Domains and Secondary Structure

- Secondary structure is most apparent in large proteins, where it comprises most of the interior.
- The various secondary structures provide an efficient means of pairing polar groups of the peptide backbone in hydrogen bonds.
- α-helical and β-sheet regions can be identified based on contact maps.
  - α-helices are characterized by a greater spread of close contacts along the diagonal, because $C_{\alpha i}$ is in close proximity of $C_{\alpha i-4}$, $C_{\alpha i-3}$. $C_{\alpha i+3}$ and $C_{\alpha i+4}$.
  - Parallel β-sheets are characterized by a series of close contacts on a diagonal line parallel to the main diagonal. This is the result of the fact that in parallel β-sheets the residues of adjacent strands are paired $C_{\alpha i}$ with $C_{\alpha j}$, $C_{\alpha i+1}$ with $C_{\alpha j+1}$ ..... etc. (i and j denote first residues of each strand)
  - Antiparallel β-sheets give rise to a series of contacts that are perpendicular to the main diagonal. Residues in adjacent strands pair $C_{\alpha i}$ with $C_{\alpha j}$, $C_{\alpha i+1}$ with $C_{\alpha j-1}$ ..... etc (Here i and j denote the first and last residues respectively of adjacent strands).
  - Structural domains are often apparent as segregated areas of contacts on the distance plots

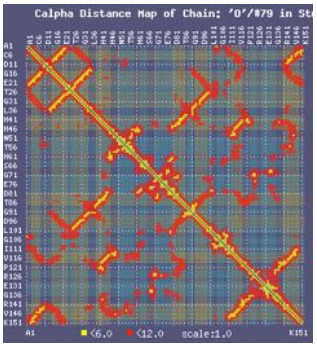


G-CSF: helical bundle





Superoxide dismutase: antiparallel β-sheet



Contact maps generated using iMoltalk: http://i.moltalk.org/

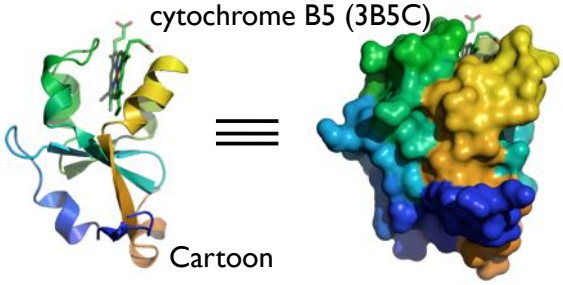
Protein Domains

- Domains are fundamental units of tertiary structure, and are often associated with aspects of protein function.

Structural Aspects of Domains

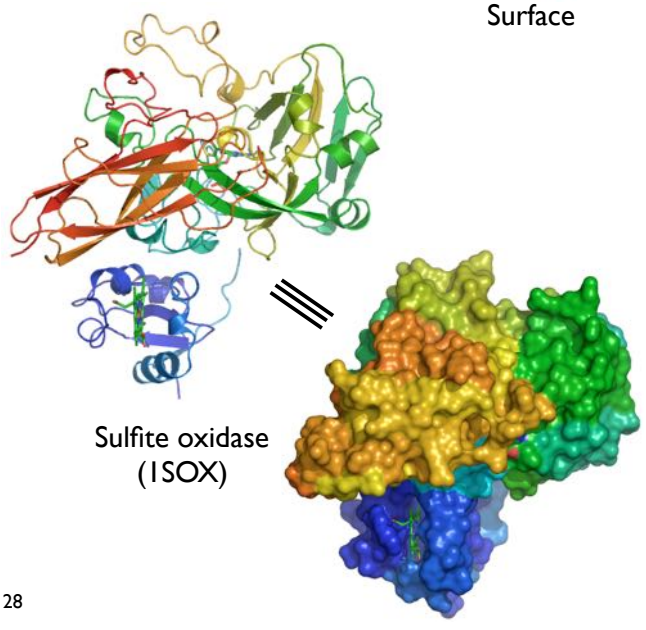
- Rotations about the individual bonds of both the backbone and the side chains are generally close to one of the conformations favored by the isolated structural unit.
- The dihedral angles and bond angles of the backbone lie within limits for the isolated peptide unit.
- Rotations about the side chain bonds are generally close to one of the three conformations, which allow the attached atoms to be staggered.
- *Cis* peptide bonds occur in folded proteins at about 5% of the bonds preceding Pro residues. These occur primarily at tight bends or turns of the polypeptide backbone.

Heme binding domain of cytochrome B5 (3B5C)



Cartoon

Surface



Sulfite oxidase (1SOX)

Domains: Eukaryotes vs Prokaryotes

- Majority of proteins consist of 2 or more domains.
- Nearly 2/3 of all prokaryote proteins contain 2 or more domains.
- ~80% of eukaryotic proteins contain 2 or more domains.
- Most evident in enzymes.

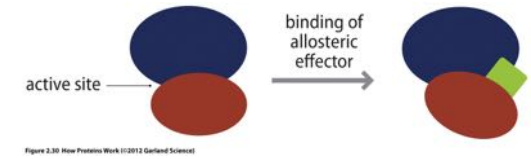
Genome Group	Number of Domains in Proteins*					
	1	2	≥2	3	≥3	≥4
Archae	36	9	43	2	9	2
Bacteria	35	10	42	2	10	2
Yeast	22	5	57	1	5	3
Metazoa	23	4	52	1	4	7

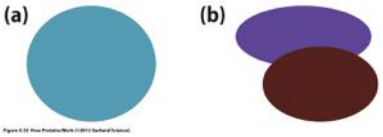
* in terms of % of total proteins.

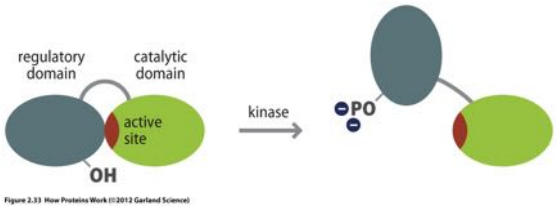
Domain Construction and Control

Multidomain Construction, Control & Regulation

- Every biochemical process needs regulating.
  - Simplest means of regulating is to control the amount of enzyme that is present.
  - Allosteric regulation and covalent modification (such as phosphorylation) provide greater flexibility.
- Most allosteric enzymes have multiple domains, with many being oligomers.
- Allostery requires that a ligand binding at some distant site affects the activity of the protein.
- Active site is almost always at the interface between two domains, with allosteric effector binding site located elsewhere.
- Binding of effector causes a change in the structure of the interface/active site.





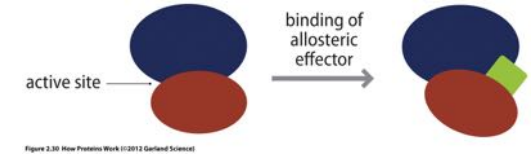


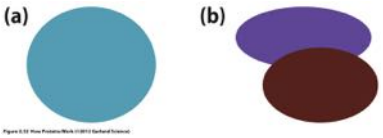
Multidomain Construction, Control & Regulation

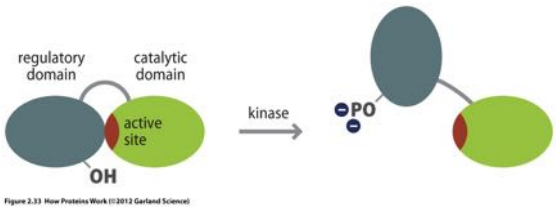
- First approximation, single domain protein is spherical with a completely convex surface.
- A two domain or dimeric protein has a much more interesting shape, with clefts or pockets.
- Expected that it is much easier to develop binding site for a small molecule or protein partner form a two-domain (or multidomain) protein.

Covalent regulation of protein.

- Most common covalent protein modification that alters enzyme activity is phosphorylation.
- Phosphorylation can affect regulation in many ways.
- Common approach is to add a regulatory domain that controls activity of the catalytic domain.
- Phosphorylation of a side-chain group on the regulatory domain leads to its dissociation or reorganization of the domain interface - altering enzyme activity.







Multidomain Construction, Control & Regulation

- Two-component bacterial signaling system.
- External stimulus results in phosphorylation of a His residue on interior position of a membrane protein.
- Phosphorous is transferred to an intracellular response regulator, leading to DNA binding.

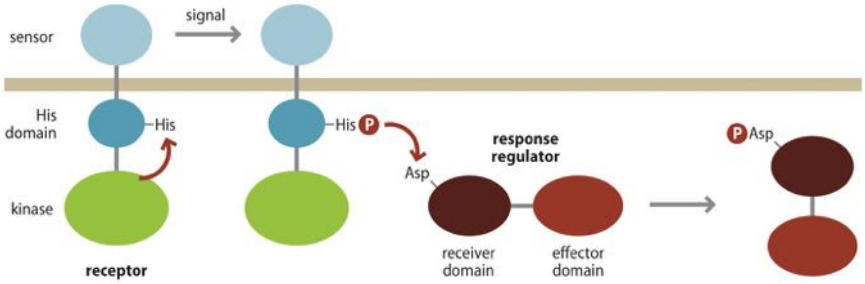


Figure 2.34 How Proteins Work (©2012 Garland Science)

- Phosphorylation of response regulator leads to conformational changes, liberating the effector domain from the receiver domain.
- Unphosphorylated response regulator is inactive: DNA recognition helix is blocked and cannot dimerize.
- Regulator can only dimerize when effector domain is has been freed from receiver domain.
- Domain binds more tightly to DNA as a dimer.

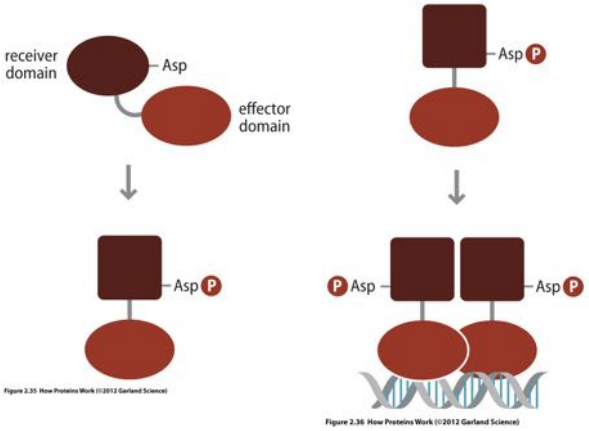


Figure 2.35 How Proteins Work (©2012 Garland Science)

Figure 2.36 How Proteins Work (©2012 Garland Science)

Multidomain Construction, Control & Regulation

- Another mechanism for phosphorylation leading to a change in activity is one in which the phosphorylated residue is recognized by a second protein.
- Here phosphorylation serves primarily to bring the two proteins together.

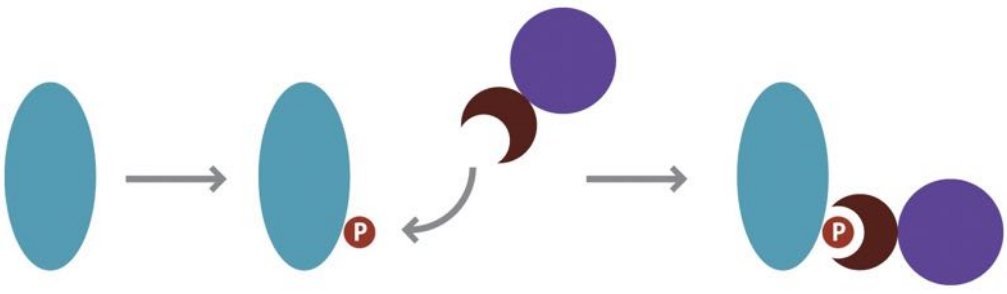


Figure 2.37 How Proteins Work (©2012 Garland Science)

Increased Binding Specificity Through Additional Domain Interactions

- Very few new enzyme functions or new structures in eukaryotes compared with prokaryotes.
- Eukaryotic enzymes tend to be more specialized.
- Eukaryotic proteins tend to be more complicated, with more domains.
- Greater number of domains in eukaryotic proteins has been suggested to be reason they are on average 50% larger than their prokaryotic counterparts.
- Domain architectures lend themselves to evolution of new biochemical function. Domains analogous to independent subsystems.
- Easiest way to evolve a new pathway is to duplicate and existing one and modify it so as to link it to a new stimulus... How important is specificity?
- More specific binding can evolve, but requires increased number of residues involved at the interface. (larger interface associated with slower on and off rates.

Increased Binding Specificity Through Additional Domain Interactions

- More specific binding can evolve, but requires increased number of residues involved at the interface. (larger interface associated with slower on and off rates.
- Alternatively, cost of increased specificity can be reduced by adding domains...
- Recognition between an SH2 domain and a phosphotyrosine.
- Typically recognizes phosphotyrosine plus three residues C-terminal to the phosphotyrosine.
  - Recognition site could evolve to recognize more of the sequence surrounding the phosphotyrosine.
  - This would generally be costly and require additional modification of existing systems.
  - Alternatively, adding extra domains (modular approach) can increase specificity by recognizing other parts of the receptor, or causing dimerization, or orients binding site, etc.

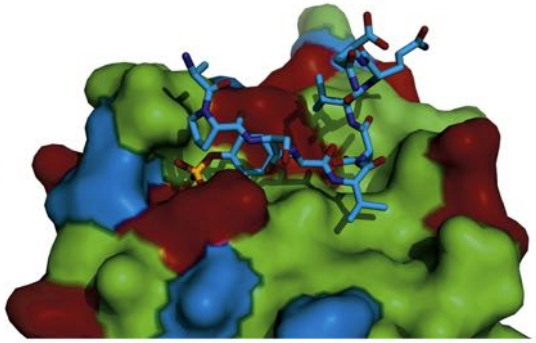


Figure 2.18 How Proteins Work (©2012 Garland Science)

Intramol. Domain:Peptide Binding & Autoinhibition

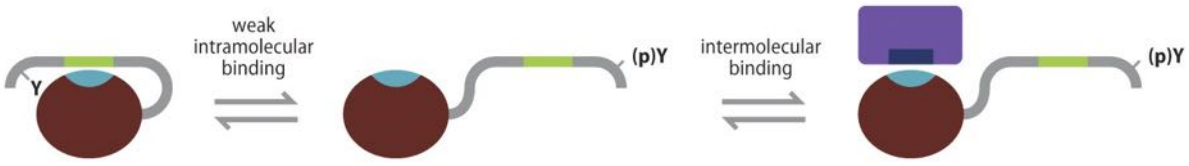


Figure 2.22 How Proteins Work (©2012 Garland Science)

- Intramolecular binding of two domains can have a wide range of affinities, but the dissociation constant is often in the nanomolar range - strongly favorable due to the large number of interactions at the interface.
- Intermolecular binding of a peptide to a domain is much weaker - smaller interface with fewer interactions and free peptides tend to be unstructured.
- Low affinity of free peptide for domain binding site - typically in μM range.
- Intramolecular domain:peptide binding exhibits higher affinity, benefitting from higher effective concentration depending on length of linking segment (typically ~1mM for 10-20 residue separation).

Intramol. Domain:Peptide Binding & Autoinhibition

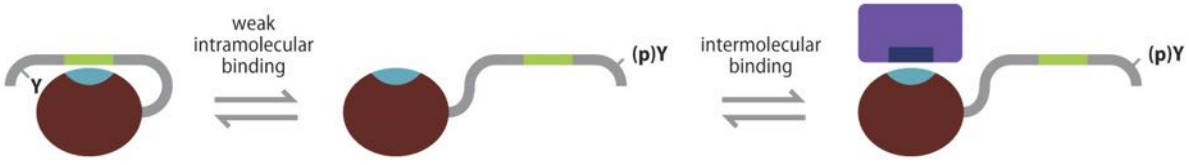


Figure 2.22 How Proteins Work (©2012 Garland Science)

- Intramolecular domain/peptide binding is often used to regulate protein function/activity, with the peptide binding to its target domain effectively shielding it from further interactions.
- Activation of the protein by some upstream event or (de-)phosphorylation liberates the domain from its intermolecular interaction.
- A protein with a domain that recognizes a specific peptide sequence can be prevented from binding the target peptide by binding intramolecularly to a similar peptide sequence.

Autoinhibition

- Because the binding of the regulating peptide segment is intramolecular, the sequence of the segment can differ significantly from that of the actual target peptide and can be very short.
- Such segments can be present in the protein "by chance" and can evolve "easily", where single point mutations can create or destroy an intermolecular binding interaction.

HMGB1g & Autoinhibition

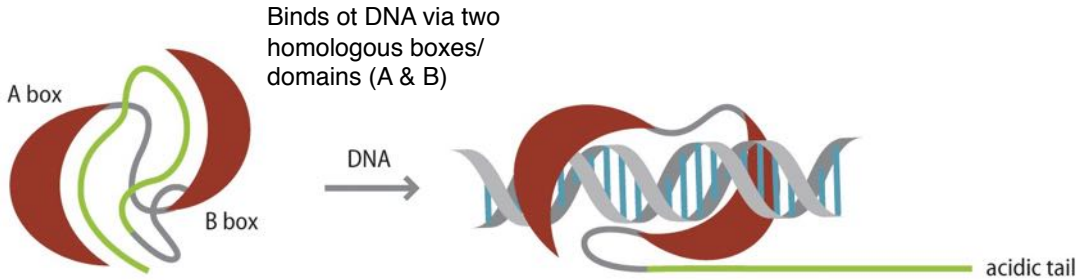


Figure 2.23 How Proteins Work (©2012 Garland Science)

- High-mobility group protein B1 (HMGB1) is an abundant nonhistone chromatin-binding protein that binds nonspecifically to DNA.
- HMGB1 binding results in bending of DNA and it is believed to aid in the bending of DNA during formation of nucleoprotein complexes.
- Contains a 30-residue C-terminal tail segment composed entirely of acidic Glu and Asp residues. This segment down-regulates HMGB1 activity.
- The C-terminal Glu/Asp segment is believed to function in an autoinhibitor capacity by binding to DNA-binding surfaces and making them inaccessible.
- The binding of the peptide segment is inherently weak due to its dependence solely on electrostatic interactions.

Scaffold Proteins

- Linking a binding domain in a protein to its cognate peptide sequence can dramatically increase their effective concentration.
- More generally, linking two recognition elements together (either in same peptide chain or within a complex) increases their binding on-rate with little effect on the off-rate effectively increasing affinity as $K_d = k_{off}/k_{on}$.
- Use of **scaffold** proteins: nh
  - Incorporate multiple recognition domains that hold binding partners. Bring the bound proteins together in a complex.
  - They increase the effective concentration of the bound proteins (including peptides and other biomolecules)
  - Holding partners together in the complex ensures that they interact with each other over other proteins that may be present.
  - Specificity of transmission is increased by a factor of $1/K_{intra}$.
  - Mainly discussed in reference specificity of signal transduction, but relevant to many other systems.
- The poor specificity of most kinases and phosphatases makes them poor drug targets... but scaffold proteins **are** potentially good targets.

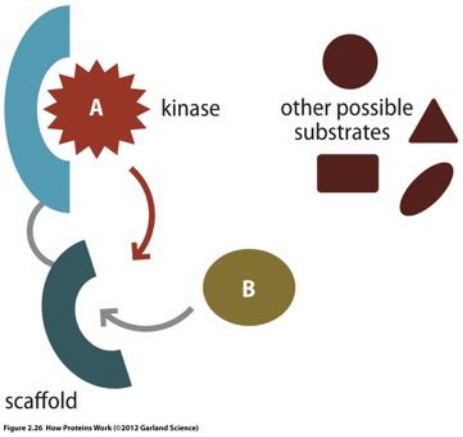


Figure 2.26 How Proteins Work (©2012 Garland Science)

Scaffold Proteins and Blood Clotting

- Blood clotting cascade ultimately results in the production of thrombin, the protease that converts fibrinogen into fribrin to cause clotting.
- Once damage has been repaired, thrombin is deactivated by thrombomodulin, a scaffold protein.
  - Binds thrombin by means of the fifth EGF domain (EGF5) near the fibrinogen binding site.
  - Blocks the fibrinogen binding site with EGF6 domain.
  - Slows fibrinogen activation by reducing availability/accessibility of the fibrinogen binding site.
  - Binds Protein C using EGF4 domain.
  - Protein C is poor thrombin substrate, but binding by thrombomodulin increases rate of Protein C cleavage.

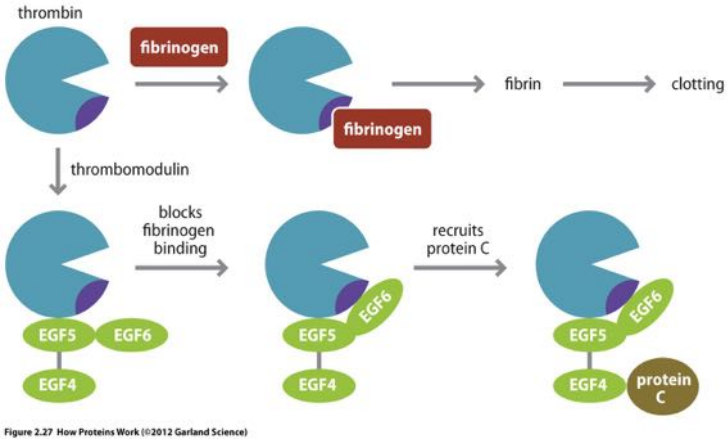


Figure 2.27 How Proteins Work (©2012 Garland Science)

- The scaffold protein, thrombomodulin, does not affect the thrombin proteolytic function… alters availability to other substrates.

Enzymes & Multidomain Construction

- Multidomain construction contributes to making an effective enzyme.
- Enzymes catalyze chemical reaction by lowering the transition-state energy ($\Delta G^{\ddagger}$).
- Before enzyme is ready to catalyze the reaction, usually necessary for some conformational change to occur in order to orient functional groups in the active site ($E \longrightarrow E^{*}$).
- Enzyme lowers the activation energy by being complementary to the transition state (referring to the E^{*} state).

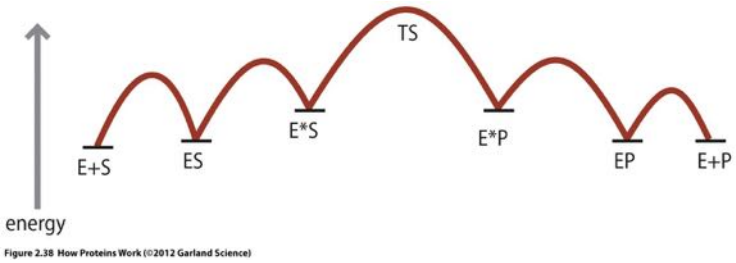


Figure 2.38 How Proteins Work (©2012 Garland Science)

- Structural rearrangement within domains is possible, but is slow and not easy feature to evolve because protein function normally requires a stable fold.
- More practical to have active site at interface between two domains, where $E \longrightarrow E^{*}$ rearrangement is simpler... Rearrangement of a few residues in a hinge rather than restructuring of the domain.

Multidomain Construction Simplifies Folding

- Proteins are synthesized as unfolded chains and subsequently fold to adopt their native three dimensional structure.
- Folding needs to be a rapid process, otherwise unfolded chains could interact with other cellular components and not fold properly.
- Small proteins can fold quickly without assistance, but larger proteins often require help.
- Evolutionary pressure for proteins to fold efficiently to their native structure.
- Folding is easier if the protein is organized in smaller domains.
- Generally, smaller proteins fold faster than larger ones. In multidomain proteins each module usually folds independently, which can dramatically accelerate folding.
- Thus pressure for smaller domains, median size ~120 residues. Median protein size is ~200 residues.

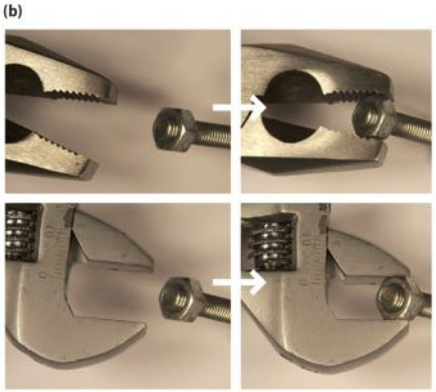
Multidomain Construction Simplifies Folding

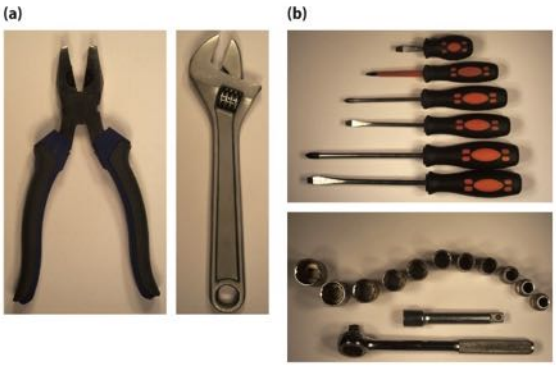
- In general, the stability of a protein is determined by the buried hydrophobic surface, which depends on the volume of the protein.
- A protein's stability generally increases in proportion to its radius.
- If there is an optimum stability for proteins (ΔG = ~20-60 kJ mol^{-1}), must be an optimum size. (ratio of volume to surface area is r/3).
- Minimum Size: Proteins less that 40 residues without disulfide bonds generally lack well defined three dimensional folded structures.
- Largest known domains are ~35kDa, but folded proteins can be much larger (megadaltons range). Such large proteins are assembled from multiple domains.
- Because most protein degradation mechanisms require access to protein surface, multiple domains help to stabilize proteins by reducing accessible surface area.

Proteins as Tools

Proteins as Tools

- Properties of tools:
  - They recognize their substrate.
  - They change in response to substrate and environment.
  - They act on substrate to change it.
- Tools work on design principles
  - Two parts are made to move independently.
  - There are a range of sizes - share common design.
  - Share common parts
  - Where possible, they are symmetric (or nearly symmetric)
  - Specialist tools





Tools with Specialized Use

- Any reasonably common fold has at some point been used for multiple applications.
- Generation of new enzymes is a relatively rare event.
- Only 25% of fold superfamilies have members of different enzyme types based on EC classification.
- If two sequences share ≥ 40% identity, it is highly unlikely that the enzymes have different functions.
- ≥30% identity - the first three digits of the EC classification number can be predicted with ≥ 90% accuracy.
- Enzymes have been divided into 6 major classes based on the chemistry of their reactions.

  Each enzyme is given a 4-digit classification number.
  - 1st digit denotes the class (type of reaction)
  - 2nd digit denotes the subclass (different versions of the reaction)
  - 3rd and 4th digits describe more reaction specifics (subcategories)
- Thousands of enzymes have been identified.

Classification	Type of Reaction Catalyzed
1. Oxidoreductases	Oxidation–reduction reactions
2. Transferases	Transfer of functional groups
3. Hydrolases	Hydrolysis reactions
4. Lyases	Group elimination to form double bonds
5. Isomerases	Isomerization
6. Ligases	Bond formation coupled with ATP hydrolysis

Independently Moving Parts

- In proteins, many parts can move independently.
- Movement of fixed units.
- Widespread use of two-domain structure for binding substrates between domains (such as in pliers or nutcracker).
- Common use of lid/flap that closes over substrate. (common in proteases and nucleases)

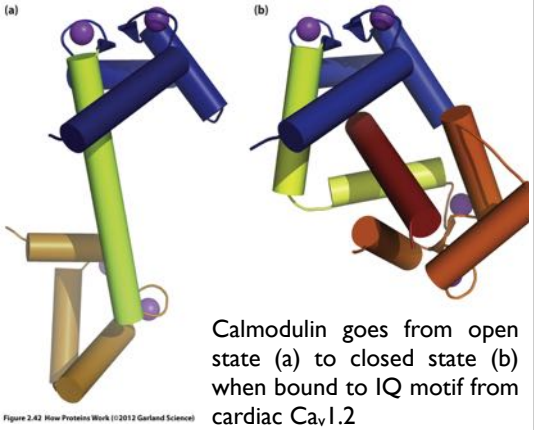


Figure 2.42 How Proteins Work (©2012 Garland Science)

Calmodulin goes from open state (a) to closed state (b) when bound to IQ motif from cardiac $Ca_V1.2$

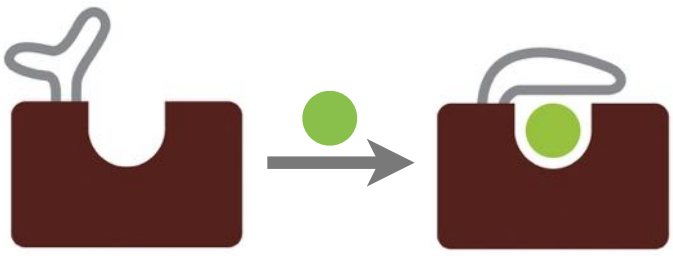
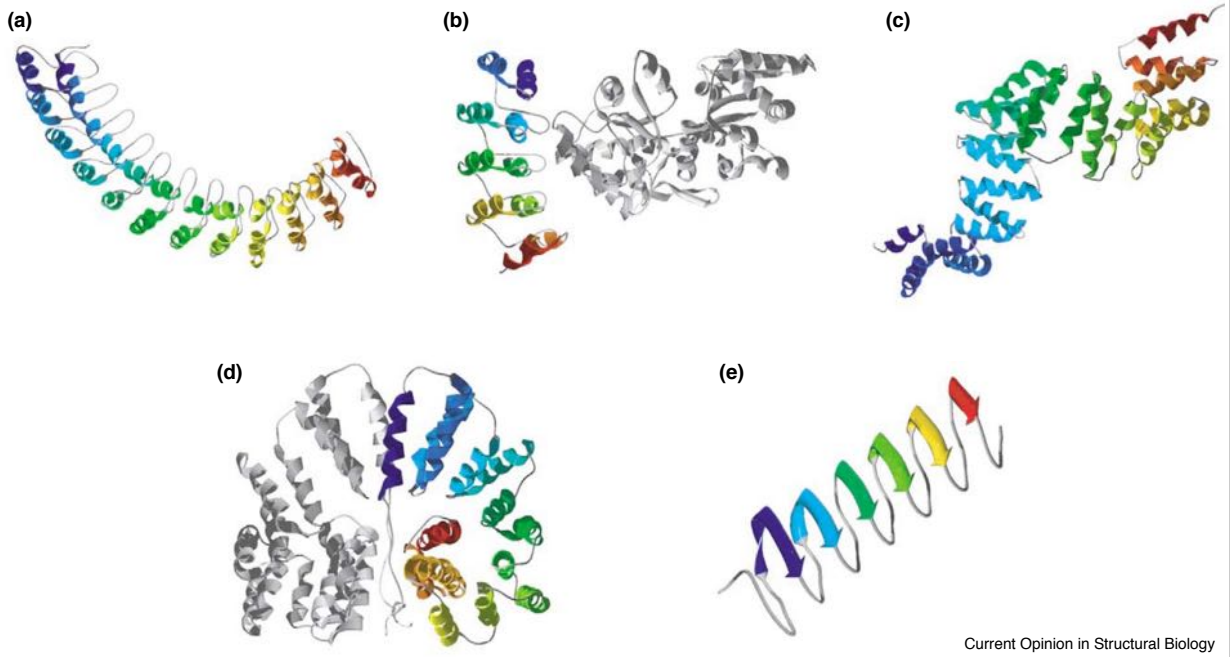


Figure 2.43 How Proteins Work (©2012 Garland Science)

Common Design but Different Sizes

- Proteins can extend in one dimension...
- Examples:
  - Zinc finger:
  - Tetratricopeptide repeat:
  - Ankyrin repeat
  - Leucine-rich repeat
  - ...β helix and armadillo repeat.

Repeat Proteins



Current Opinion in Structural Biology

Ribbon illustrations of various repeat proteins. **(a)** The 12 ANK stack from human ankyrinR, demonstrating the curved interaction surfaces [39]. **(b)** A complex showing the interaction between the designed ANK repeat protein off7 and MBD (grey) [19••]. **(c)** The superhelical TPR repeat domain of *O*-linked GlcNAc transferase, which comprises 11.5 repeats [40]. **(d)** The globular TPR protein NlpI from *Escherichia coli* K-12. Only one of the monomeric units of NlpI is coloured for clarity [11•]. **(e)** The *Tenebrio molitor* beetle antifreeze protein (TmAFP) [41]. Repeat proteins are coloured by secondary structure, progressing from the N terminus (blue) to C terminus (red). Figure prepared using Swiss-PDB viewer v3.7 SP5 (http://www.expasy.ch/spdbv) and POVRay v3.5 (http://www.povray.org), with PDB coordinates 1N11 (ankyrinR), 1SVX (off7), 1W3B (*O*-linked GlcNAc transferase), 1XNF (NlpI) and 1EZG (TmAFP). 50

Curr. Opin. Struct. Biol. 2005, vol 15, p.p. 464-471.

Table 1

Examples of commonly occurring repeat protein architectures and their interactions.

Repeat type[a]	Architecture	Example interaction	PDB codes
Antifreeze protein (AFP)	A 12-residue motif forming a regular α helix, which resembles a rectangle in cross-section	Prevents ice crystals forming in plant/animal tissues [42]	1EZG [41]
ANK repeat	A 33-residue motif forming a helix-loop-helix-α turn motif, which is L-shaped in cross-section	Amongst the best-characterised interactions are those of the INK4 proteins with cyclin-dependent kinases [43,44]. Further examples include the interaction of IκB with NF-κB [45]	1A5E [46], 1BI7, 1BI8 [43], 1G3N [43], 1NFI [45]
Armadillo repeat (ARM)	An ~40-residue motif forming a three-helix bundle	The armadillo domain of β-catenin interacts with the cytosolic domain of E-cadherin [47]	1BK5 [48], 1I7X [47]
HEAT repeat	A 37- to 47-residue motif; each module comprises a pair of antiparallel helices	Importin-β interacts with sterol regulatory element binding protein (SREBP-2) [49]	1F59 [50], 1UKL [49]
Hexapeptide repeat	A hexapeptide motif comprising a β strand and loop, which forms a continuous β helix resembling an equilateral prism in cross-section	Galactoside acetyltransferase in complex with coenzyme A and β-galactoside [51]	1KRV [51]
LRR	A 20- to 29-residue motif forming a β strand-loop-helix structure	Human placental RNase inhibitor, an LRR protein, binds to human angiogenin with high affinity [52]	1FO1 [53], 1A4Y [52]
TPR	A 34-residue motif; each module comprises a pair of antiparallel helices	The adaptor protein Hop contains two TPR regions, which bind Hsp70 and Hsp90 [54]	1NAO [12], 1W3B [40], 1ELW [54]
WD40 repeat	A 40- to 50-residue motif forming a four- stranded β sheet	The β subunit of the G protein heterotrimer Gi $\alpha_1\beta_1\gamma_2$ contains a sevenfold WD40 β propeller, which interacts with both α and γ subunits [55]	1QHU [56], 1GP2 [55]

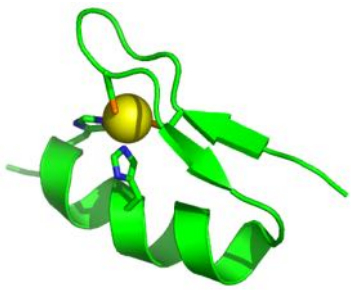
[a] Many of the repeat proteins listed can be further split into subfamilies [3].

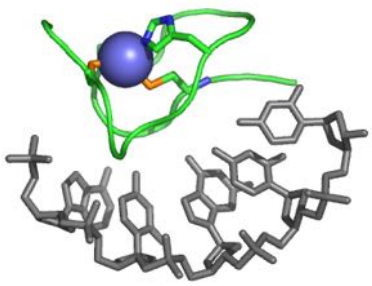
51

Curr. Opin. Struct. Boil. 2005, vol 15, p.p. 464-471.

Common Design but Different Sizes

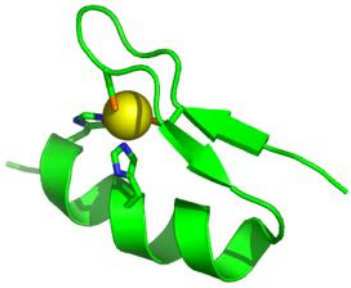
- Proteins can extend in one dimension...
- Examples:
  - Zinc finger:
    - Cys_2His_2 family is largest class of eukaryotic transcription factors.
    - Commonly used to recognize DNA
    - Each finger binds in major groove.
    - Proteins contain 1-28 (or more) zinc fingers.
    - Increasing the number of zinc finger units allows protein to recognize longer DNA sequences.
  - Tetratrico peptide repeat:
  - Ankyrin repeat
  - Leucine-rich repeat
  - ...β helix and armadillo repeat.

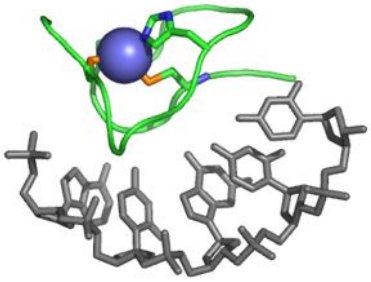




DNA-Binding Proteins

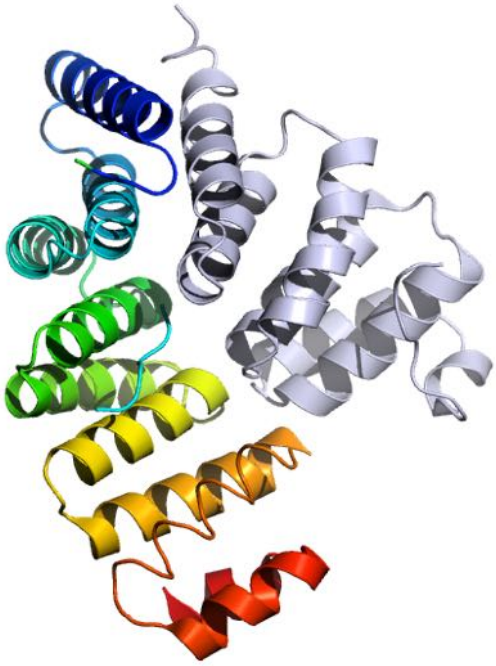
- The zinc finger structure is stabilized by a coordinated Zn^{2+} ion chelated by the side chains of two Cys and two His residues, α-helix to interact with DNA.
- Consist of 28-31 amino acids with the characteristic pattern:
  $-X_3-C-X_{2-4}-C-X_{12}-H-X_{3-4}-H-X_4-$
- Each zinc finger appears to be a stable, autonomous structural unit (so long as bound Zn^{2+} is present).
- Hundreds of zinc finger motifs have been identified in proteins involved with aspects of gene regulation in eukaryotes.
- Multiple zinc fingers are usually present in tandem along the polypeptide chain linked by only a few residues (often -T-G-E-K-)
- Simultaneous interactions involving multiple zinc fingers greatly stabilizes the interaction between the protein and DNA.
- Each Zinc finger makes contact with three adjacent nucleotide base pairs in the major groove.
- Proteins containing multiple zinc fingers are of modular construction and alteration of the combination and placement of the zinc fingers would alter DNA sequence specificity of the protein.





Common Design but Different Sizes

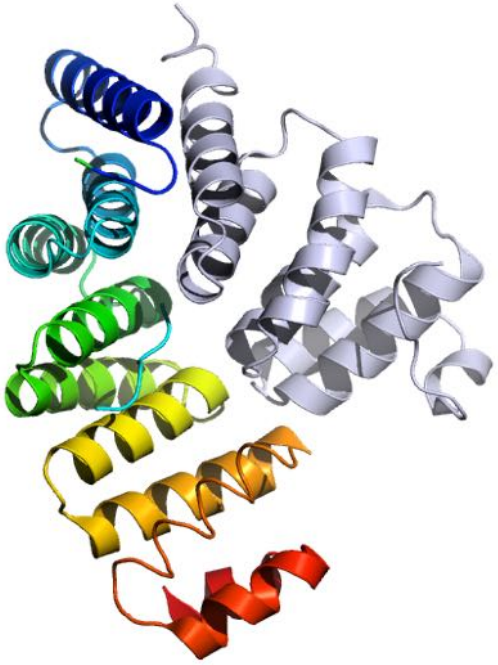
- Proteins can extend in one dimension...
- Examples:
  - Zinc finger:
  - Tetratrico peptide repeat:
    - 34-residue motif, consisting of a pair of anti-parallel helices (helix-turn-helix motif).
    - Overall topology consists of a spiral of of repeating helices, with right-handed twist.
  - Ankyrin repeat
  - Leucine-rich repeat
  - ...β helix and armadillo repeat.



Structure of Peroxisomal Targeting Signal 1 (PTS1) binding domain of Trypanosoma brucei Peroxin 5 (TbPEX5)complexed to PTS1 peptide (10-SKL)

Tetratrico Peptide Repeat

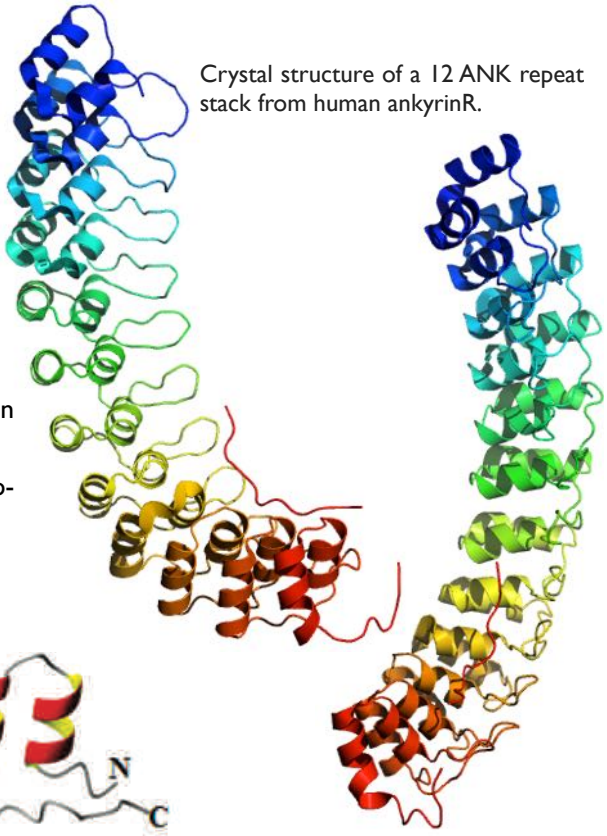
- First identified in 1990, and named based not he 34-residue repeat pattern.
- The organization of the helices and overall construction of TPR domains generates two protein faces:
- An inner, concave surface
- An outer, convex surface.
- Found in a range of proteins and are involved with formation of multi-protein complexes.
- TPR domains mediate protein-protein interactions.
- Specific details of the interactions between some target proteins and TPR domains have been established, but not the case for most.



Structure of Peroxisomal Targeting Signal 1 (PTS1) binding domain of Trypanosoma brucei Peroxin 5 (TbPEX5)complexed to PTS1 peptide (10-SKL)

Common Design but Different Sizes

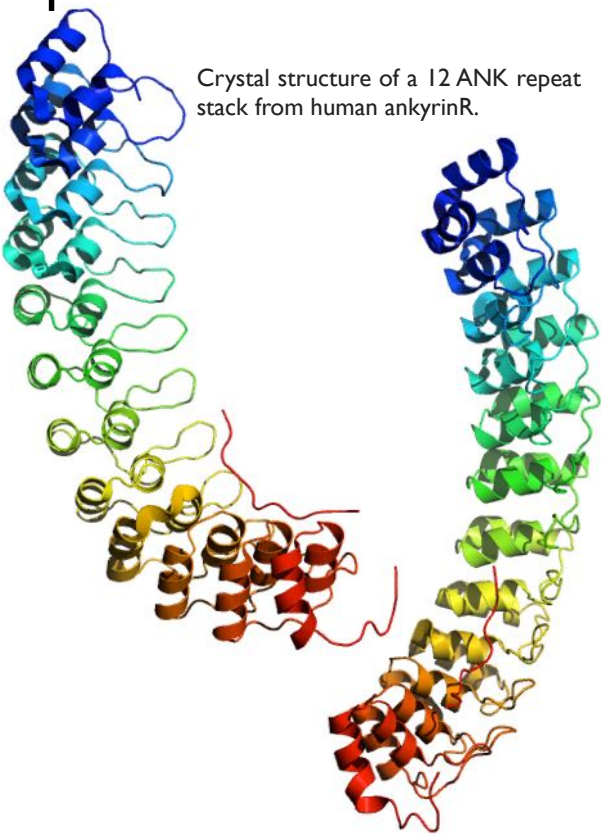
- Proteins can extend in one dimension...
- Examples:
  - Zinc finger:
  - Tetratrico peptide repeat:
  - Ankyrin repeat:
    - ✤ One of the most widely used protein structural motifs.
    - ✤ 33-residue motif consisting of a helix-loop-helix.
    - ✤ L-shaped cross-section
  - Leucine-rich repeat:
  - ...β helix and armadillo repeat.



Crystal structure of a 12 ANK repeat stack from human ankyrinR.

Ankyrin Repeat

- Sequence analysis of the human erythrocyte cytoskeletal protein ankyrin revealed an N-terminal domain consisting primarily of 22 tandem repeats of 33-residue helix-loop-helix units.
- Ankyrin repeats consist of strings of tandem repeats of helix-loop-helix units packed together to form an elongated structure.
- Structure stabilized by inter- and intra-unit hydrophobic and hydrogen bonding interactions.
- Analysis of a protein database in 2004 revealed 19,276 ankaryn sequences in 3608 proteins.
- Found in a variety of proteins where they mediate specific protein-protein interactions.
- They do not recognize specific sequences, but instead recognize specific contacts between residues distributed on the surface of the repeat and those on the surface of the folded binding partner.
- Ankyrin repeats are not associated with enzymatic activity nor are proteins containing them.

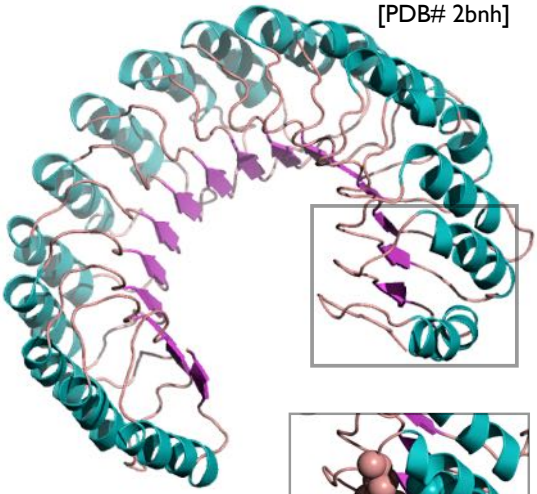


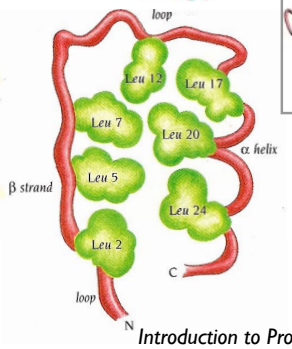
Crystal structure of a 12 ANK repeat stack from human ankyrinR.

Common Design but Different Sizes

- Proteins can extend in one dimension...
- Examples:
  - Zinc finger:
  - Tetratrico peptide repeat:
  - Ankyrin repeat:
  - Leucine-rich repeat:
    - ✤ Leucine-rich motifs consist of tandem repeats of homologous 20-30 residue right-handed β-loop-α structures.
    - ✤ 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".
  - ...β helix and armadillo repeat.

Ribonuclease inhibitor
[PDB# 2bnh]



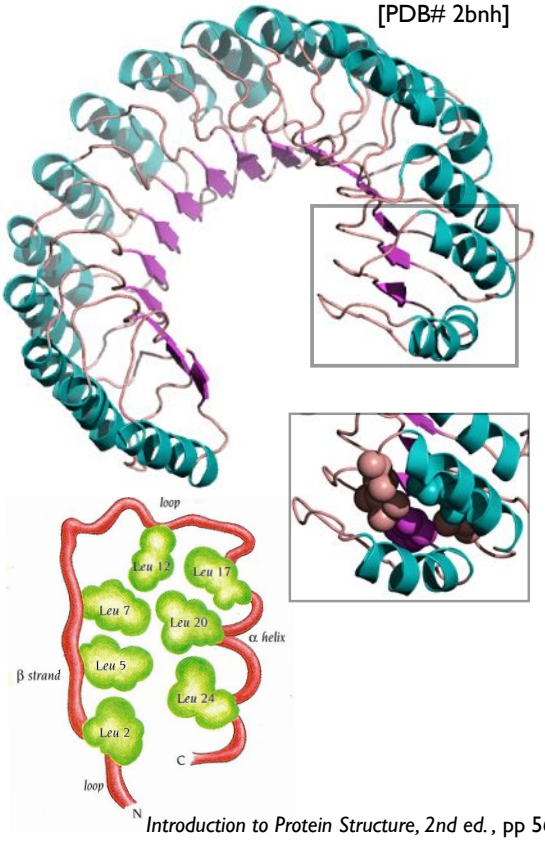


Introduction to Protein Structure, 2nd ed., pp 56

Leucine-rich Motifs

Ribonuclease inhibitor
[PDB# 2bnh]

- 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.
- With the β strands lining the inner surface of the 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.
- Ribonuclease inhibitor has been used as the basis for constructing plausible models for other proteins with leucine-rich motifs.

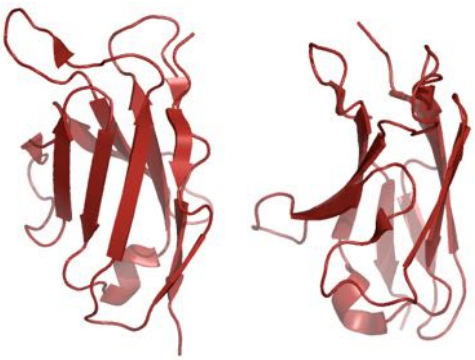


Introduction to Protein Structure, 2nd ed., pp 56

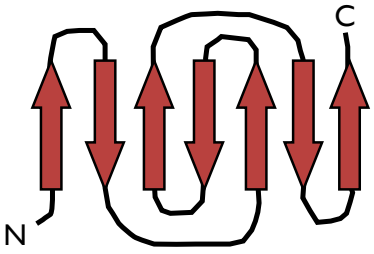
Tools have Common Parts with Variable Ends

- In human tools, use of common parts with variable end functionalities is used to save on materials, cost and weight.
- Biological systems more concerned with compactness and adaptability.
- Recurring theme of using common motifs and domains... modular nature of some proteins.
- The immunoglobulin or immunoglobulin-like fold represents a modular element utilized by proteins that perform a variety of functions.

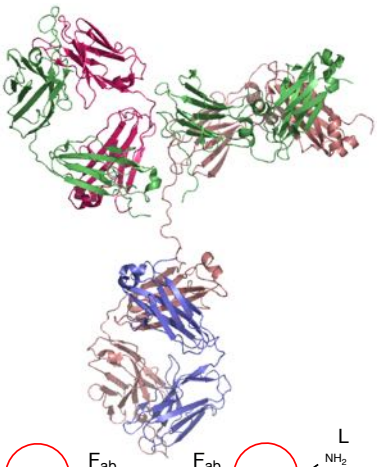
  Found in variety of proteins including immunoglobulins, cell surface receptors and the kinase titin.
- The immunoglobulin fold consists of two overlapping Greek keys that fold to form a β-sandwich.
- This is as arrangement shared by all known proteins that contain a β-sandwich motif.
- In a β-sandwich, 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.



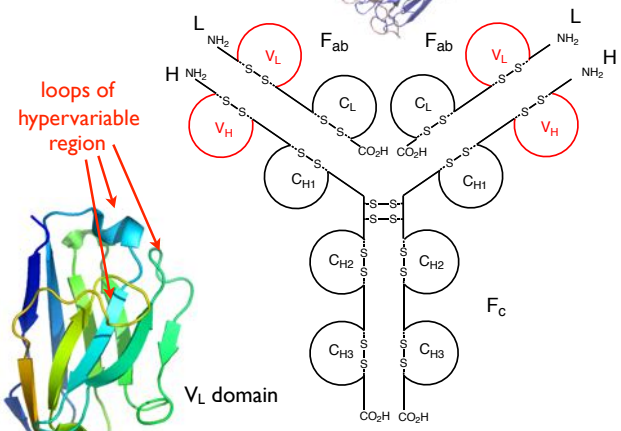
Engineered immunoglobulin based on IgA Kappa (pdb# 2IMM)



Immunoglobulins

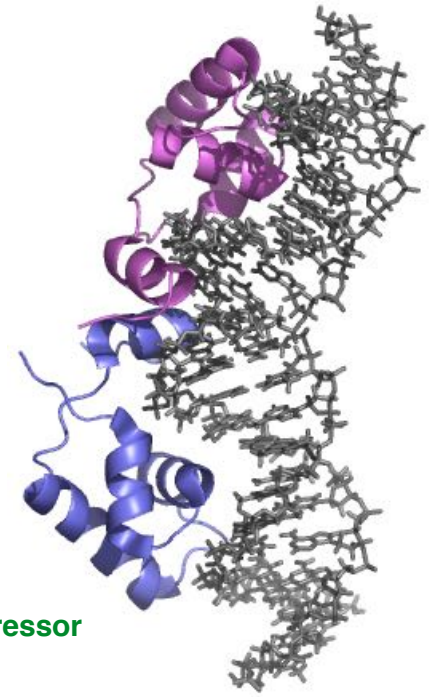


- Antibodies are capable of recognizing and binding a very diverse spectrum of antigens (virtually any molecule).
- Antibodies share common structural and functional properties.
- All immunoglobulins of a given class demonstrate these common qualities linked to different antigen specificities.
- The characteristic structure for an intact immunoglobulin consists of a Y-shaped molecule assembled from four polypeptide chains (two H and two L chains)
  - ✤ Each L chain contains two domains (V_L and C_L).
  - ✤ Each H chain contains 4 domains (V_H, C_{H1}, C_{H2} and C_{H3}).
  - ✤ These domains are approximately 100 residues and share significant sequence homology. Thus, it is not surprising that they share a common folded conformation (immunoglobulin fold).
- The antigen binding sites are located at the tips of the F_{ab} arms, situated between the V_L and V_H domains.
- Binding site formed by residues in the loops between β-strands of both of the V_L and V_H domains.
- Different binding sites and specificities are generated at the complementarity determining regions (CDR's) or hypervariable regions.



Some Tools are Symmetric

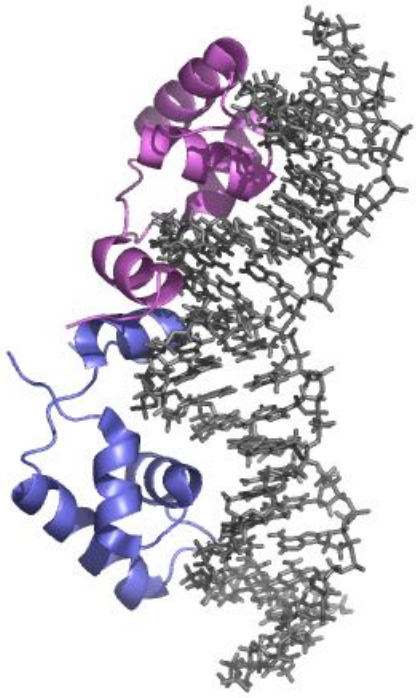
- Proteins assembled from only L-amino acids, which precludes mirror symmetry.
- Oligomeric proteins can exhibit rotational symmetry.
- Proteins, such as lac repressor, that recognize and bind palindromic DNA sequences exhibit rotational symmetry.



***lac* repressor**

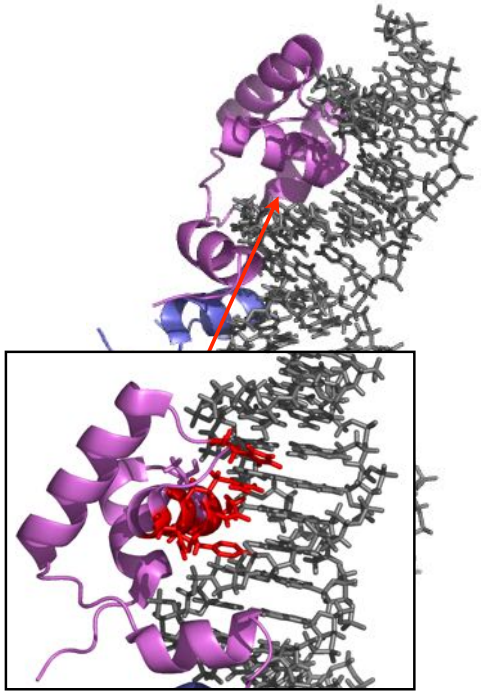
DNA-Binding Proteins

- Proteins that specifically bind DNA are of great biological importance: replication, gene regulation, gene expression.... etc.
- Only the edges of the nucleotide bases are accessible to solvent and the protein (primarily in major groove).
- In order for a protein to discriminate among base-pairs by interacting with their edges in the major groove, it must display interacting groups that project substantially from the protein surface.
- The helix-turn-helix is a common motif used by proteins for interaction with DNA.



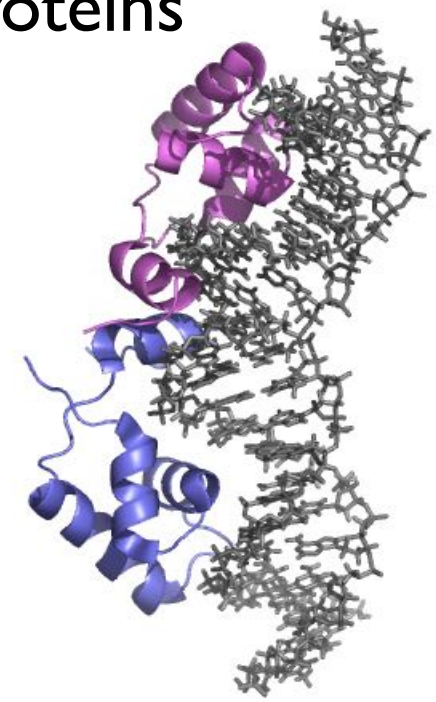
DNA-Binding Proteins

- The helix-turn-helix is a common motif used by proteins for interaction with DNA.
- The specificities of the various helix-turn-helix motifs arises from the different amino acid side chains displayed near the amino terminus of the helix projected into the groove (known as recognition helix).
- These side chain groups participate in hydrogen bonds, electrostatic and van der Waals interactions with the edges of the nucleotides exposed in the major groove.
  - The residues involved with interactions with DNA generally have polar side chains (Arg, Asn, Gln, Asp and Glu).
  - Water molecules are frequently recruited into hydrogen bond networks.
- The other helix crosses the major groove and participates primarily in nonspecific interactions.
- No simple code relating amino acid sequence to the recognized nucleotide sequence.



DNA-Binding Proteins

- Many helix-turn-helix DNA-binding proteins bind as dimers, with both equivalent binding helices making the same interactions with DNA with the same sequence.
- Such proteins tend to recognize palindromic DNA sites. In palindromic sites, the nucleotide sequence of one DNA strand is repeated in a complementary fashion and in reverse order.
- Palindromic sites define areas of localized twofold symmetry. The dimeric protein binds with its twofold axis of symmetry coinciding with that of the DNA binding site.
- Such a strategy should improve binding specificity.



5' TGTGTGGAATTGTX$_9$ACAATTTCACACA
3' ACACACCTTAACA Y$_9$TGTTAAAGTGTGT

⇐ Recognized by *lac* repressor

Multimeric Proteins

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.

Quaternary Structure

- Many proteins consist of more than one polypeptide chains; they consist of multiple subunits (monomers or protomers).
- A protein's quaternary structure consists of the spatial arrangement of these subunits.
- The subunits may be identical or different.
- Each subunit is usually folded into an apparently independent globular conformation. There are exceptions where the polypeptide chains are intertwined.
- The centers of the interfaces between the monomers are usually similar to the closely-packed interiors of globular proteins (predominantly hydrophobic and closely packed).
- Interface is usually surrounded by polar and charged groups and many hydrogen bonds and salt bridges.
- Some interfaces involve interactions between secondary structure elements.
- Interfaces between subunits are highly complementary.

Figure 3.1 How Proteins Work (©2012 Garland Science)

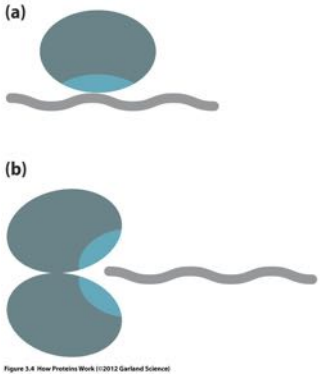
Why Oligomerize

- Simple and effective strategy for generating proteins with multiple domains.
- Evolution of a stable oligomeric protein from a monomer facilitated by the fact that mutations affect both monomers, doubling their impact. (homodimers)
- Benefits of oligomerization similar to those associated with multiple domains.

Figure 3.1 How Proteins Work (©2012 Garland Science)

Oligomers = Better Enzymes

- Binding at oligomeric interface allows more restrictive access to binding/ active site.
- Flexibility within the subunit interface allows active site to open to bind substrate or release product.
- Flexibility also allows subunits to close over active site during reaction - shielding from solvent.
- Decreases the energy input need to change the active site and its environment.
- Structure and behavior of the active site can be altered through the dimer (oligomeric) interface - minor changes at interface will probably impact function.
- Active sites in enzymes are usually hydrophobic - prone to aggregation and nonspecific binding of hydrophobic molecules. *Placing at oligomeric interface helps to shield the activesite.*
- Oligomeric proteins allow for allostery.



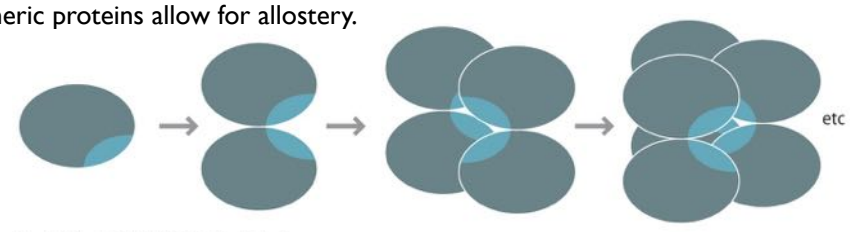


Figure 3.6 How Proteins Work (©2012 Garland Science)

Coding Errors

- Many of the advantages associated with dimeric proteins are equally true for single chain proteins containing two domains.
- There are advantages associated with assembling a protein from a single chain.
- One advantage of assembling proteins from two or more polypeptide chains relates to transcription/translation errors.
  - Error rate in translation, while low, is potentially not insignificant.
  - Statistically, error rate (1:4000 AAs) in large two domain protein is more significant than for a short single domain protein.
  - Less severe repercussions in multimeric assemblies of smaller single-domain proteins than for larger multidomain proteins.... and less cost in producing peptide.
- Linkers in multidomain single-chain proteins a potential liability.
  - In oligomeric proteins, interface between subunits has greater flexibility than those between domain in multidomain proteins.
  - In single-chain multidomain proteins long and conformationally flexible linkers are susceptible to proteoliytic cleavage.

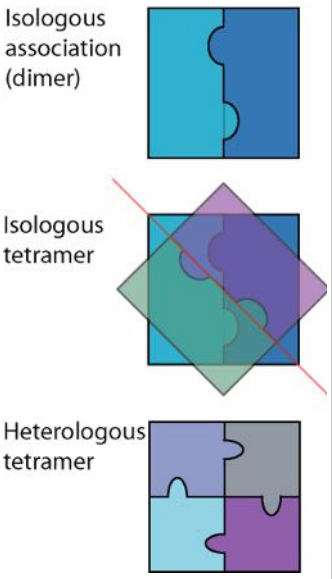
Organization of Oligomers

- The dimer formed by association of two monomers is almost always symmetric. (axis of symmetry)
- Subunits can assemble head-to-head, where interactions between subunits are symmetric.
- Subunits can assemble in a cyclic head-to-tail manner... Essentially forming a ring.
- Cyclic head-to-tail oligomers may contain an even or odd number of subunits.
- Oligomers containing an even number of subunits are more common than those containing odd numbers.

Alternatively....

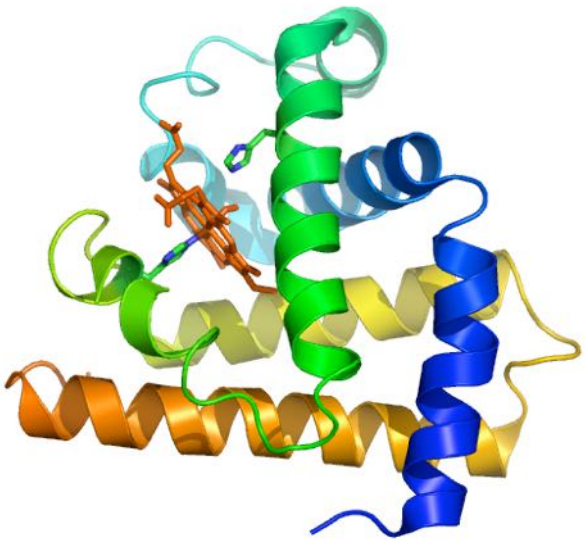
Quaternary Structure

- Association of two molecules only requires spatial and physical complementarity of interacting surfaces.
- Two fundamental types of interaction can occur between identical monomers:
- **Isologous**: association between monomers involves the same faces on both monomers.
  - Monomers associate to form a dimer with a twofold rotation axis of symmetry.
  - Association to produce tetramers or higher ordered assemblies requires that each monomer have another binding surface. Formed by two sets of isologous interactions (tetramer).
- **Heterologous**: association between monomers involves interaction between two different faces that are complementary and largely nonoverlapping.
  - Indefinite polymerization can occur unless geometry of association results in a closed ring.
  - Oligomeric proteins with a fixed number of identical subunits that is not a power of two likely involves heterologous association.
- More complex associations often occur in higher order structures such, such as icosahedral (60 monomers) viral coat assemblies.



Myoglobin and Hemoglobin

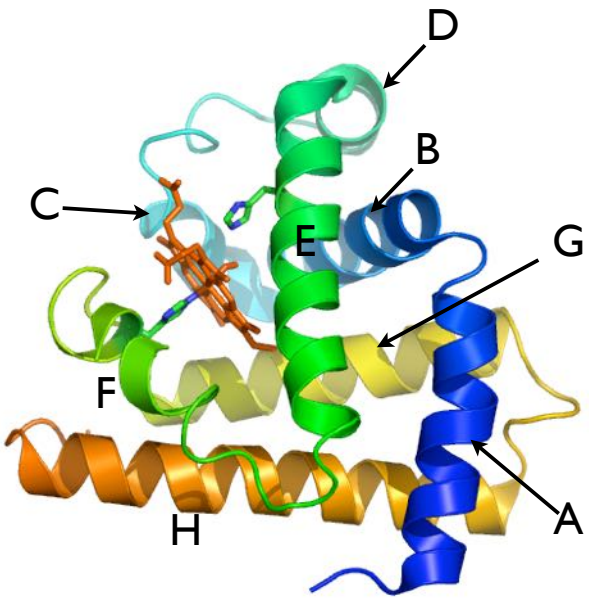
- Hemoglobin and myoglobin figure prominently in our understanding of proteins.
  - Hemoglobin was the first protein to be associated with a specific function.
  - Hemoglobin was the first protein to have its molecular mass accurately determined.
  - Hemoglobin and myoglobin were the first protein x-ray structures elucidated.



Sperm Whale Myoglobin: PDB# 1VXF

Myoglobin

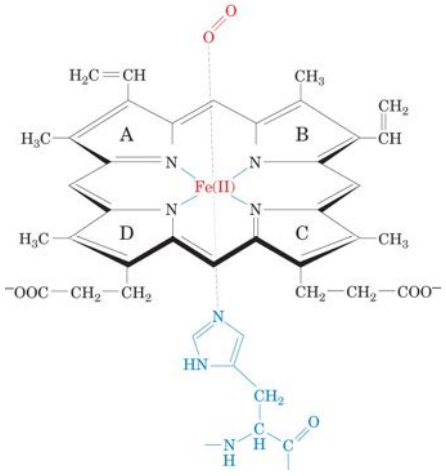
- Found primarily in muscle cells, where it is believed to facilitate O_2 transport in rapidly respiring muscle tissue/cells.
- Also may be involved with the detoxification of nitric oxide (NO).
- Typical member of the globin family.
- A single polypeptide of 153 amino acids.
- All α-helical protein. Consists of 8 α-helices (identified as helices A-H).
- Little structural difference between oxy and deoxy forms.



Sperm Whale Myoglobin: PDB# 1VXF

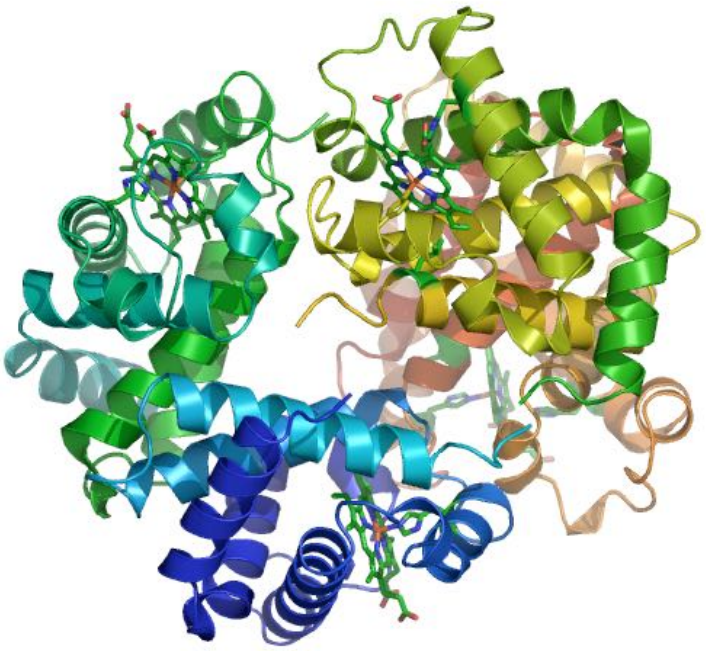
Heme and Oxygen Binding

- Amino acid side chains are not suitable for the reversible binding of O_2.
- Organisms use transition metal complexes of **Cu** and **Fe** for oxygen transport.
- The **heme ligand**, protoporphyrin IX, is a macrocyclic ring system consisting of 4 linked pyrole groups.
- The ring nitrogens of the heme ligand **occupy 4 of the coordination sites of a bound Fe(II)**. A fifth site is occupied by a nitrogen in the side chain imidazole of a histidine provided by the protein.
- A sixth Fe(II) coordination site remains available for O_2 binding.
- The Fe-heme complex is buried within the protein. Protects the Fe(II) complex.
- Heme also binds small molecules such as CO, NO and H_2S (they have higher affinity than does O_2).

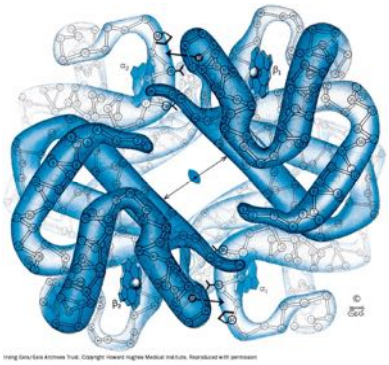


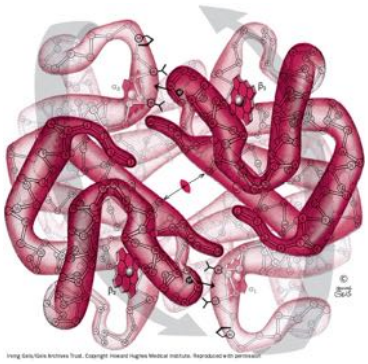
Hemoglobin

- Hemoglobin functions as an oxygen transport protein.
- At the lungs Hb is 95% saturated with O_2, upon return to the lungs it is 55% saturated.
- Hb is a tetramer containing 2 α and 2 β subunits. (It can also be considered an ab dimer.)
- The structures of the α and β subunits are very similar.
- Mb and the two subunits of Hb are very similar in structure despite only 18% amino acid identity (the α subunits of hemoglobin lack a D helix).
- There are extensive interactions between unlike subunits in Hb. (i.e. α_2-β_2, α_2-b_1, α_1-β_2 and α_1-β_1)



O_2 Binding and Reorganization



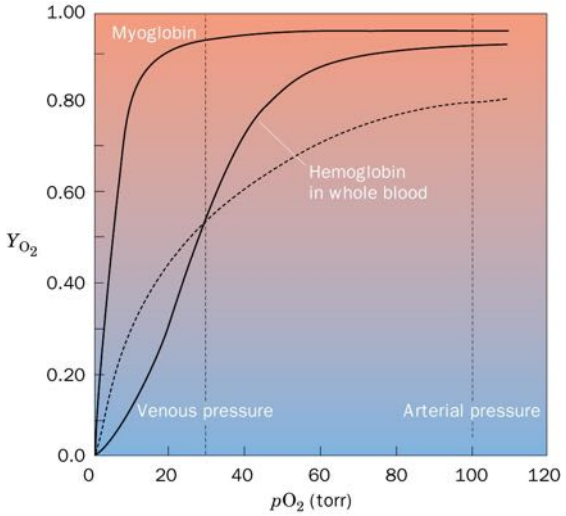


- Hemoglobin (Hb) has two major conformations, the T state (T=tense) and the R state (R=relaxed).
- The T state is more stable in the absence of O_2, while the R state is more stable with bound O_2.
- Transition from the T to the R state involves sliding of the αβ subunits past each other (shifting along the α_1-β_2 and α_2-β_1 interface).
- In the T state the porphyrin ring of the heme group is slightly puckered.

(from Voet and Voet, Biochemistry 3rd ed.)

Hemoglobin and Cooperativity in O_2 Binding

- Cooperative binding indicates that the binding of O_2 to one subunit influences the binding of O_2 to other subunits.
- Binding of a ligand at one site influencing binding of a second ligand at a distal site referred to as allostery.
- Cooperative binding gives rise to a sigmoidal saturation curve for Hb.



Oxygen dissociation curves of hemoglobin and myoglobin in whole blood.

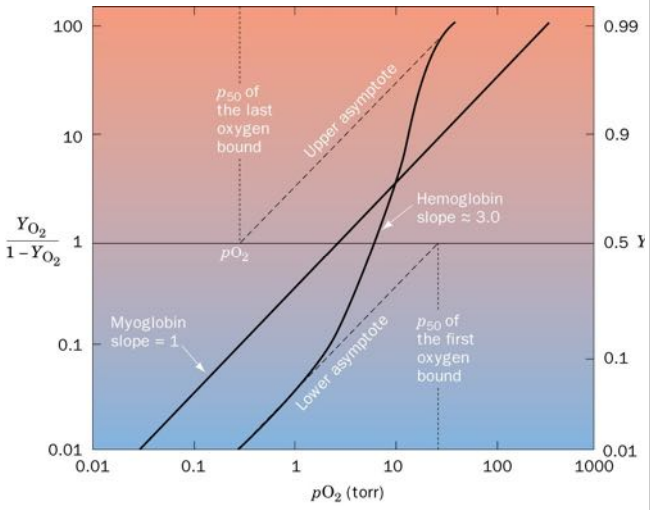
(from Voet and Voet, Biochemistry 3rd ed.)

Hill Equation

- The cooperative binding of O_2 by Hb can be described quantitatively.
- The equilibrium expression becomes:

$$E + nS \leftrightarrow ES_n$$

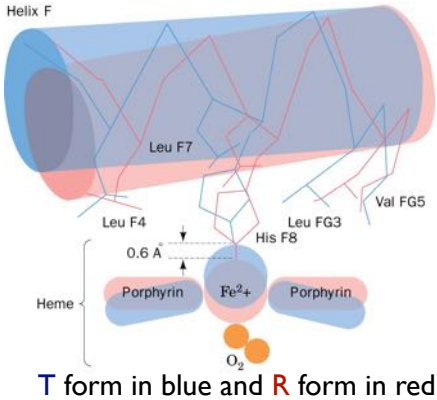
- A Hill coefficient of 1 means no cooperativity.
- A Hill coefficient greater than 1 means cooperativity is occuring.
- For normal Hb the Hill constant is between 2.8 and 3.0



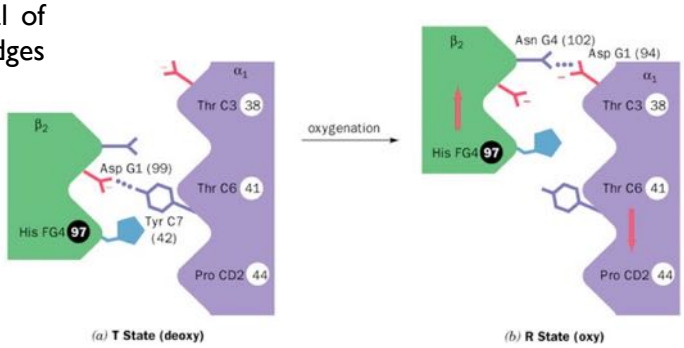
(from Voet and Voet, Biochemistry 3rd ed.)

Structural Basis for Cooperativity

- Movement of Fe into the ring pulls the proximal histidine with it. Cooperativity is reduced if this histidine is mutated to glycine.
- This causes the F helix to translate 1 Å across the heme plane.
- This translation changes the contacts between α_1-β_2 and α_2-β_1.
- Salt bridges that stabilize the T-state are broken upon this translation. Removal of residues involved in these salt bridges abolishes cooperativity.



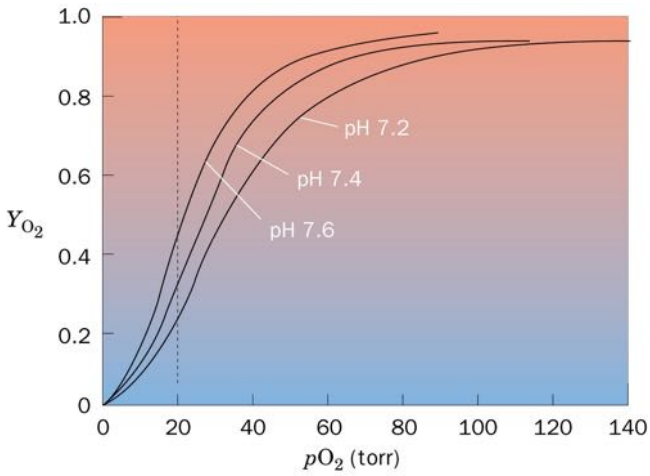
T form in blue and R form in red



(from Voet and Voet, Biochemistry 3rd ed.)

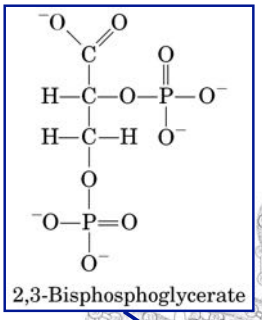
Bohr Effect

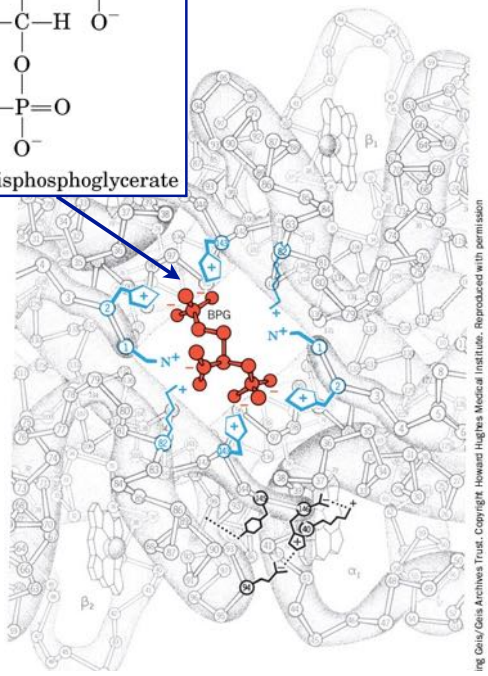
- Lowering the pH results in decreased oxygen affinity (the Bohr effect).
- Hb binds H+ in the peripheral tissues and transports it to the lungs.
- The transition from T to R affects the pK value of different amino acids.
- His-146 is protonated at lower pH, and once protonated forms a salt bridge with Asp-94. This salt bridge favors the R to T transition.
- The amino terminus also has an unusually low pK value in the T state and binds H^+ to transport to the lungs. This low pK is influenced by Cl^-.



2,3-Bisphosphoglycerate (BPG)

- BPG binds to Hb and promotes formation of the T state.
- In the lungs O_2 concentration causes a transition to the R state, releasing BPG.
- The level of BPG in the peripheral tissue is important in O_2 release in cells. (higher concentration in peripheral tissues)
- BPG concentrations are varied at higher altitudes.
- Fetal Hb has a lower affinity for BPG, and thus can obtain O_2 from the maternal Hb.
- In adult Hb, the quintuply charged BPG molecule interacts with two histidine residues, a lysine residue, and the amino terminus of both b subunit.
- In fetal Hb, one of these histidine residues are replaced by serine, reducing the number of contacts between Hb and BPG and lowering the affinity.





Allosteric Models

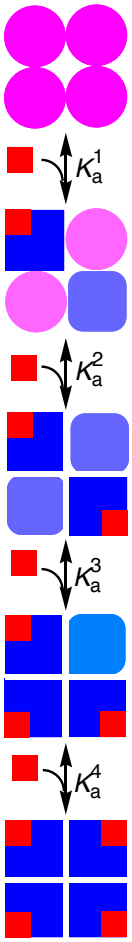
- Both the **sequential** and **concerted** models imply that ligand binding has effects on the protein conformation.
- Sequential model:
  - Binding of ligands at one site directly affects the affinity of other sites.
  - Predicts conformational change upon ligand binding correlates with degree of ligand binding.
- Concerted model:
  - Conformational changes associated with ligand binding must only extend to the interface between domains/subunits, which alters the conformational equilibrium between T and R.
  - Model is restrictive. Only parameters that can be varied are *L* and *c*.
  - Concerted model does not predict negative cooperativity.

Allosteric Models

- The interaction between sites on a protein is linked to protein flexibility.
- Two models have been proposed to address allosteric interactions.
- **The sequential model**: The protein is sufficiently flexible that the binding of one ligand at one site can directly alter the conformation at another site.
  - Binding of ligands at one site directly affects the affinity of other sites.
  - Predicts conformational change upon ligand binding correlates with degree of ligand binding.
- **The concerted model**: Ligand binding at one site has no direct affect on other sites, but alters conformational equilibrium between two alternative quaternary structures.
  - Conformational changes associated with ligand binding must only extend to the interface between domains/subunits, which alters the conformational equilibrium between T and R.
  - Model is restrictive. Only parameters that can be varied are *L* and *c*.
  - Concerted model does not predict negative cooperativity.

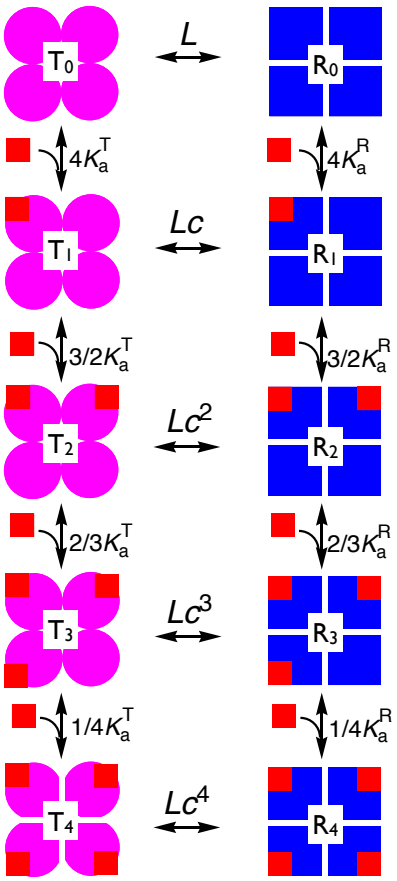
Allosteric Models

- **The sequential model**: *The protein is sufficiently flexible that the binding of one ligand at one site can directly alter the conformation at another site.*
- Observed changes in affinity are usually only a few orders of magnitude, which corresponds to changes of only a few kcal/mol.
- Such changes in affinity could be achieved through only minor conformational changes in the binding site.
- The binding of any ligand at any site on a protein could affect independently every other site (increase or decrease affinity).
- Scheme proposed by Koshland, Nementhy and Filmer (KNF or sequential model).
  - Accounts for known allosteric propreties of proteins.
  - Explains affinity changes in monomeric proteins.
- Conformational changes associated with ligand binding are usually small, but small changes can have significant impact on affinity at other sites.



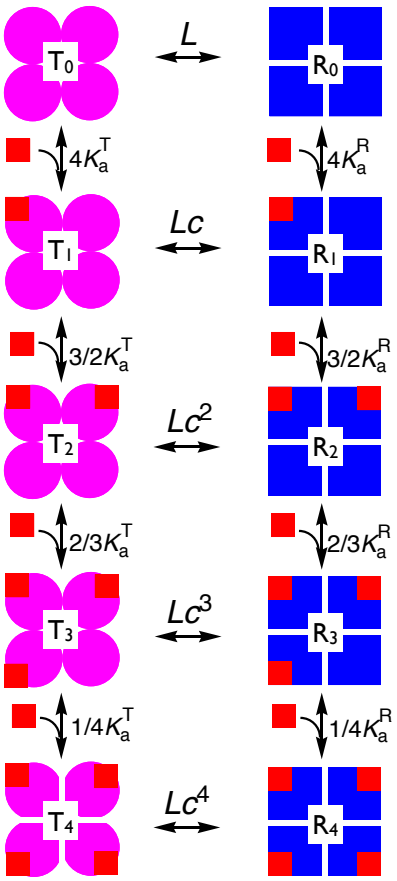
Allosteric Models

- **The concerted model**: *Ligand binding at one site has no direct affect on other sites, but alters conformational equilibrium between two alternative quaternary structures.*
- *Also known as MWC, symmetric or two-state allosteric model.*
- Proteins demonstrating interactions between nonoverlapping binding sites are invariably oligomeric.
- With ligand binding, proteins undergo very large conformational changes involving rearrangement of relatively unaltered subunits.
- Monod, Wyman and Changeux proposed a model for allosteric interactions.
  - Ligand binding at one site does not directly affect other sites.
  - Ligand binding alters the equilibrium between two alternative quaternary conformations.
  - The binding sites in one conformation having low intrinsic affinity for the ligand (designated T, *tense*).
  - The binding sites in the other conformation having high affinity (designated *R*, relaxed).
  - The two forms with *i* bound ligand molecules referred to as T_i and R_i.

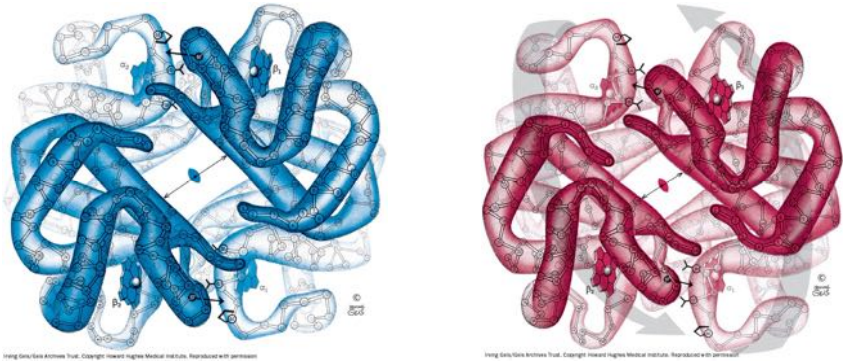


Concerted Model contd.

- According to the concerted model, the protein exists in equilibrium between the T and R states, and the two forms coexist even in the absence of ligand (T_0 and R_0 - with equilibrium constant *L*).
  - T_0 form would normally be the preferred conformation.
  - The R conformation having the greater ligand affinity by a factor of *c*.
  - Ligands preferentially bound by protein in R conformation which will drive equilibrium towards the R form.
- In this model, heterotropic interactions involving other ligands arise because the these ligands also bind preferentially to protein in either the T or the R state.
- Each ligand affects the apparent affinity of the protein for other ligands by shifting equilibrium between T and R states.



O_2 Binding and Reorganization



- Hemoglobin (Hb) has two major conformations, the T state (T=tense) and the R state (R=relaxed).
- The T state is more stable in the absence of O_2, while the R state is more stable with bound O_2.
- Transition from the T to the R state involves sliding of the $\alpha\beta$ subunits past each other (shifting along the α_1-β_2 and α_2-β_1 interface).
- In the T state the porphyrin ring of the heme group is slightly puckered.

(from Voet and Voet, Biochemistry 3rd ed.)

Negative Cooperativity

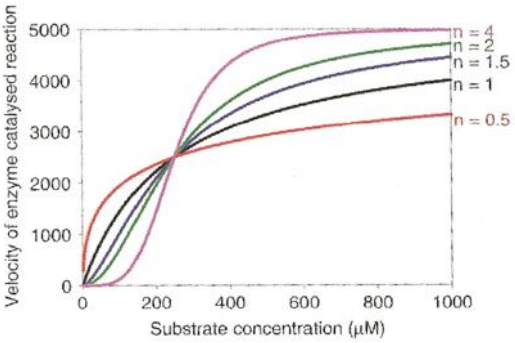
- Homotropic allosteric interactions in which binding of a ligand molecule decreases ligand affinity at other sites for the same ligand are unlikely with the concerted model.
- Negative cooperativity is observed in some proteins.
- In some cases, half-of-the sites reactivity is observed (only half of the expected binding sites are occupied).
- One explanation for such behavior would be that initially identical and equivalent binding sites in an oligomeric protein, are made nonidentical as a result of ligand binding to one of the sites.
- In another explanation, the equivalent binding sites on a symmetric protein either overlap or are sufficiently close together to interact sterically or electrostatically (such as the case of phosphate binding by hemoglobin).

Allosteric Control

- The activity of enzymes can be regulated by means of interactions with effector molecules that are unrelated to substrate.
- These molecules regulate and modulate enzyme activity through allosteric interactions.
- They bind to sites on the protein that are distinct from the enzyme active site.
- Allosteric enzymes are multimeric proteins incorporating at least two subunits/polypeptide chains (quaternary structure) with multiple catalytic and binding sites.
  - ★ They can be composed of identical or combinations of non-identical subunits.
  - ★ In enzymes composed of identical subunits, each polypeptide chain contains at least one catalytic site and regulatory site.
  - ★ In enzymes assembled from non-identical subunits the catalytic and regulatory sites may be located on different polypeptide chains.
- Binding of effector molecules at sites that are distinct from the active site favor the active or inactive conformations of the enzyme.
- Effector molecules are frequently components of the metabolic pathway, and their presence serves to stimulate or inhibit activity, regulating flow in the pathway.

Allosteric Regulation

- Allosteric modulators, effectors, bind non-covalently to the enzyme and affect either K_M or V_{max} of the enzyme.
- Allosteric enzymes can be identified by plotting initial velocity (v_0) versus substrate concentration [S]. The correlation between v_0 and the concentration of at least one substrate will demonstrate a sigmoidal profile.
  - Traces back to the Hill equation, cooperativity and the Hill coefficient. (v_0 vs [S] defined by Hill equation)
  - Hill coefficient (n) of 1 results in a hyperbolic plot (v_0 vs [S]).
  - Hill coefficients of >1 or <1 will result in curves indicative of cooperativity.
- Many enzymes demonstrate allosteric regulation, but it is particularly prevalent in enzymes that are components of long metabolic pathways (i.e. biosynthetic processes, glycolysis, glycogenesis and β-oxidation).



Proteins: Structure and Function, 2005, by David Whitford, John Wiley & Sons publishers, ltd., p231.

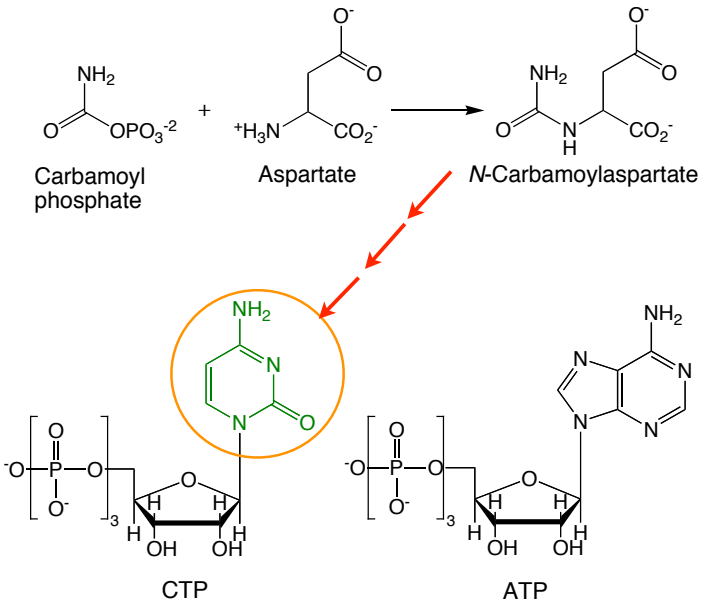
Hill equation

$$v = \frac{V_{max}[S]^n}{K_{0.5}{}^n + [S]^n}$$

$K_{0.5}{}^n$ = substrate concentration at half V_{max}.

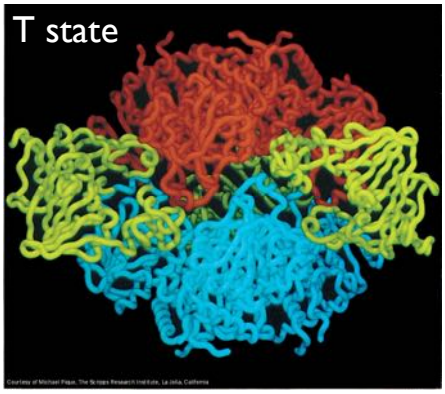
Allosteric Regulation and Aspartate Transcarbamylase

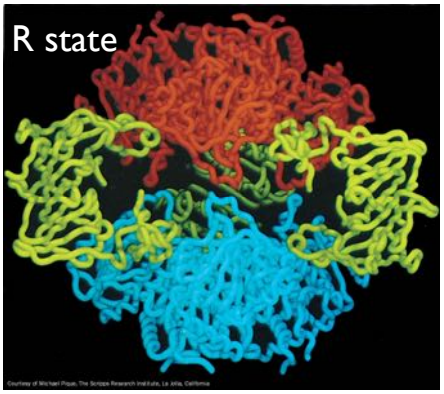
- Aspartate transcarbamylase (ATCase) catalyzes the carbamylation of the α-amino group of aspartic acid. (the first unique step in pyrimidine biosynthesis)
- The activity of ATCase is influenced by the presence of substrate and downstream products.
  - ✳ The binding of substrates (Carbamoyl phosphate and aspartate) increases enzyme activity in a cooperative manner.
  - ✳ Binding of CTP decreases enzyme activity, while ATP has opposite affect.
  - ✳ CTP and ATP bind to same site in regulatory dimer.
- ATCase consists of 12 protein subunits: 6 catalytic and 6 regulatory (c_6r_6).



Allosteric Regulation and ATCase

- ATCase is involved with the first unique step in pyrimidine biosynthesis.
- CTP is a downstream product of this synthetic pathway.
  - When the concentration of CTP becomes high, CTP binds to ATCase and inhibits enzyme activity.
  - At lower CTP concentrations, CTP dissociates and the enzyme is in its activated form.
  - Similarly, ATP functions as a regulator and helps to coordinate purine and pyrimidine biosynthesis (cellular ATP conc. is generally higher than the conc. of CTP.)
- The X-ray structure of ATCase indicates that it consists of a pair of c_3 trimers and three r_2 dimers.
- When the catalytic dimers are separated from the regulatory dimers, they retain catalytic activity but are unaffected by the presence of either ATP or CTP.
- The regulatory dimers bind both ATP and CTP, but have no catalytic activity.

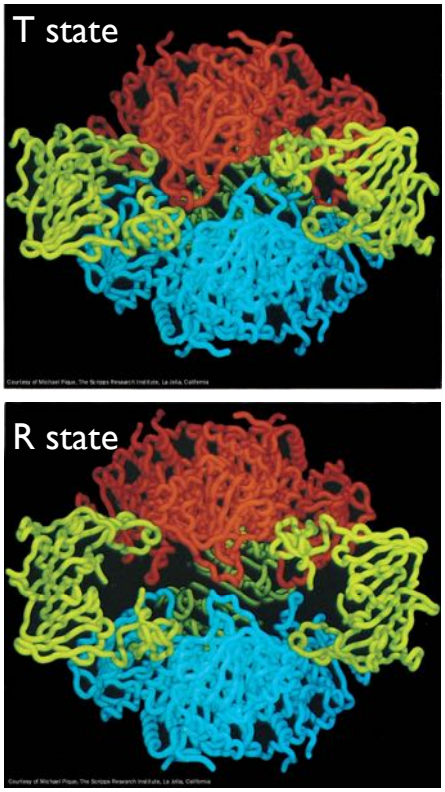




from *Biochemistry, 3rd ed.* Voet and Voet

Allosteric Regulation and ATCase

- ATCase has an active (R state) conformation and an inactive (T state) conformation.
  - ATP is preferentially bound by ATCase in the R state.
  - CTP is preferentially bound by ATCase in the T state.
- Conversion from T to R involves separation of the two catalytic trimers by ~11Å and a slight reorientation about their axis. Associated with clockwise rotation of the regulatory dimers.
- As in other allosteric systems, tertiary and quaternary structures are tightly coupled. Minor shifts in tertiary structure induce larger changes at the quaternary level.
- Carbamoyl phosphate and aspartate bind to separate domains of the catalytic subunit, which induces active site closure. Favored in R state.
- Appears to be all or nothing. Do not see one subunit in R state while others in T (symmetry model of allostery).



from *Biochemistry, 3rd ed.* Voet and Voet

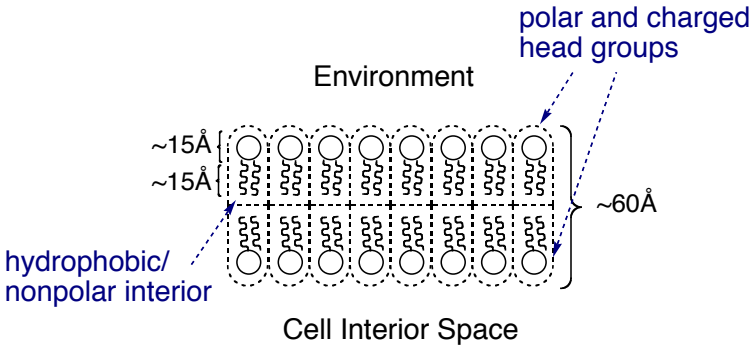
Membrane Proteins

Proteins in Membranes

- Membranes provide a physical and insulating barrier between the cell interior and its environment.
  - Membranes are essentially impermeable to charged molecules.
  - Polar but uncharged molecules can traverse membranes (low rates)
- Membranes are essentially two-dimensional fluids. Above a certain temperature, lipid molecules diffuse freely within the plane of the bilayer.
- Membranes are not homogenous. Proteins typically comprise ~50% of the mass of most natural membranes.

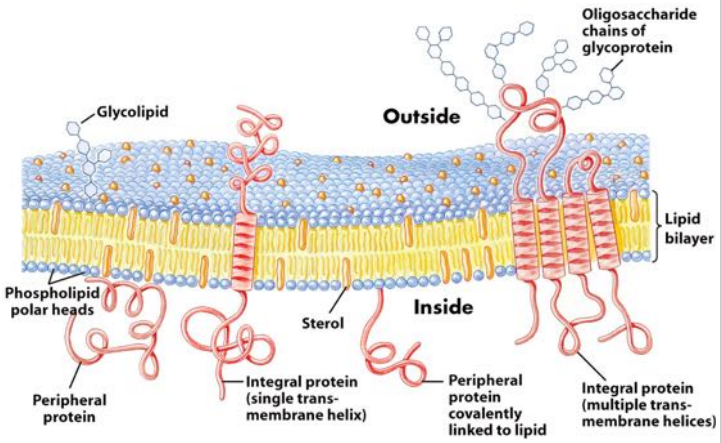
Proteins mediate various functions of the membrane:

- Transport of molecules in and out of the cell.
- Catalysis of chemical reactions.
- Receiving and transducing chemical signals.
- Maintaining membrane structure.



Fluid Mosaic Model

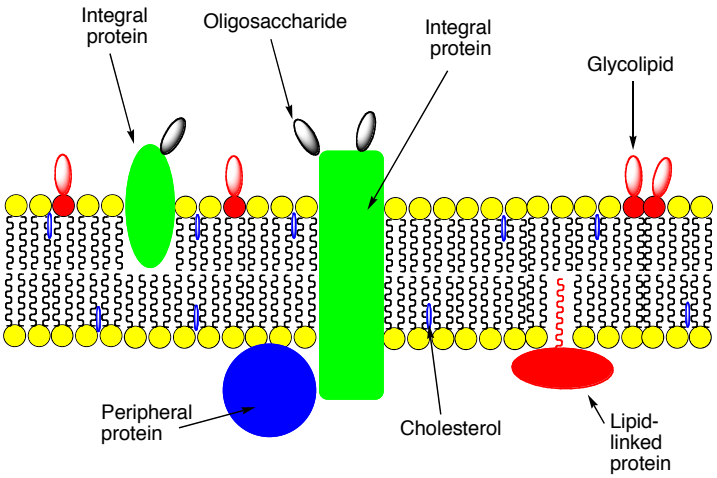
- Biological membranes much like synthetic lipid bilayers are fluid.
- In the fluid mosaic model, integral proteins are like "icebergs" in a two dimensional "sea" of lipids.
- 30-90% of membrane proteins diffuse freely.
- Typically taking 10-60 minutes to traverse the length of a eukaryotic cell (~20 μm)
- Other proteins may diffuse more slowly or or essentially be immobile.



Membrane Structure from *Lehninger; Principles of Biochemistry, 4th ed.* **2005**, p 372.

Association with Membranes

- Membrane proteins are classified as being integral or peripheral (intrinsic-extrinsic respectively).
- Some integral proteins are exposed only at one face of the membrane. Others span the membrane (called transmembrane proteins).
- Integral proteins are amphipathic: containing both hydrophobic and hydrophilic regions.
- Membrane proteins tend to be oriented asymmetrically.
- Integral proteins may move within plane of membrane but have infinitesimal flip-flop rates.

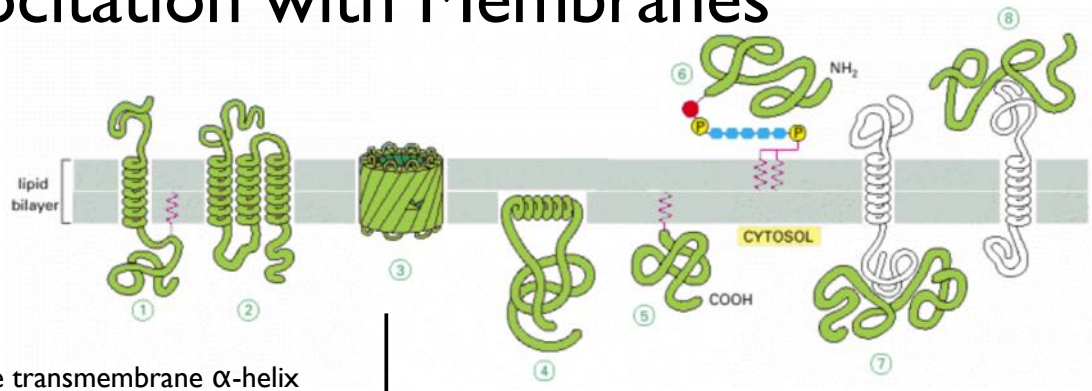


Many integral proteins utilize an α-helical transmembrane segment.

Can have numerous membrane-spanning regions. Some of the best studied examples contain 7 transmembrane segments.

It takes a helix of 20-25 amino acids to span the membrane, and can generally be identified based on hyrophobicity.

Association with Membranes



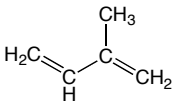
1. a single transmembrane α-helix
2. multiple transmembrane α-helices
3. a transmembrane β-barrel

"single-pass" and "multi-pass" proteins may have covalently attached fatty acid groups inserted into the cytosolic face of the membrane.

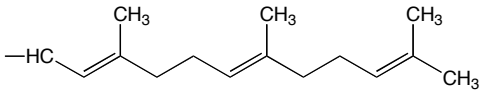
4. anchored to cytosolic face by an amphipathic α-helix that partitions into cytosolic monolayer of membrane via hydrophobic face of helix.
5. Bound to cytosolic face through a covalently attached lipid chain (fatty acid or prenyl group).
6. Linked through polysaccharide connected to phoshpatidylinositol at the outer face of bilayer.
7. Through noncovalent interactions with integral proteins (cytosolic face).
8. Through noncovalent interactions with integral proteins (outer face).

Prenylated and Fatty Acylated Proteins

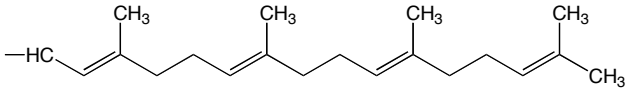
- Prenylated proteins have covalently attached isoprenoid groups.
- Most common site is Cys in a C-terminal sequence CaaX
- Fatty acylted proteins:
  - ✳ Myristic acid
  - ✳ Palmitic acid
- Palmitoylated proteins found on the cytoplasmic face of the cell membrane.
- Myristoylated proteins often targeted to intracellular compartments.



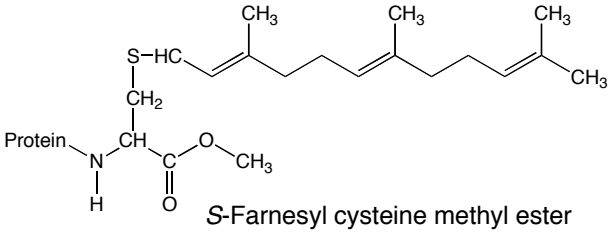
Isoprene



Farnesyl residue



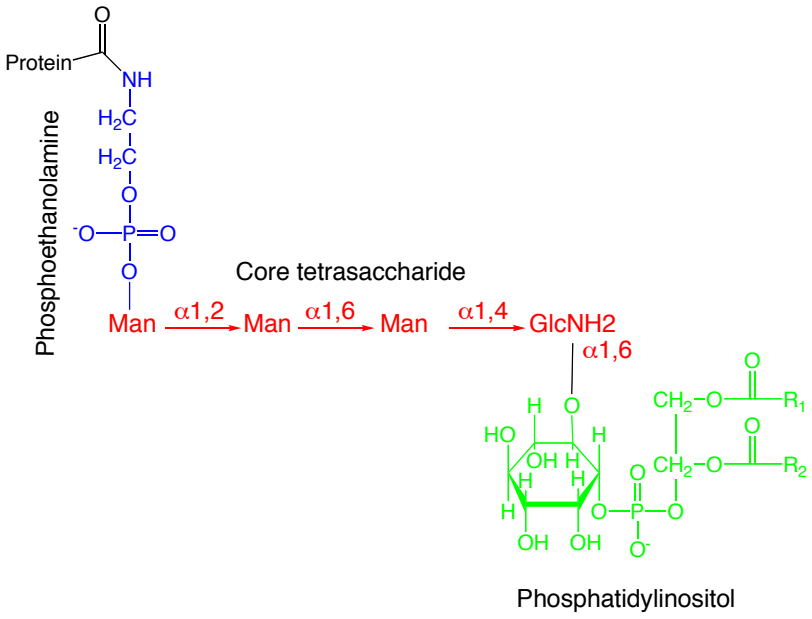
Geranylgeranyl residue



S-Farnesyl cysteine methyl ester

GPI-Linked Proteins

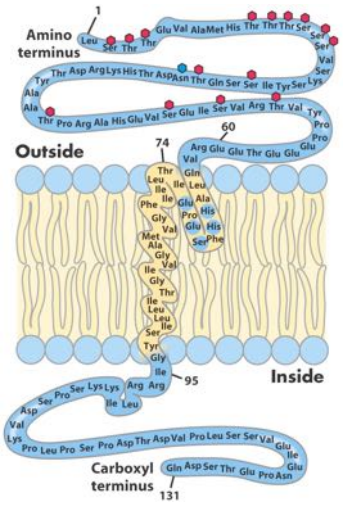
- Glycophosphatidylinositol (GPI) anchor a wide range of proteins to eukaryotic plasma membrane.
- GPI groups appear to replace transmembrane domains in anchoring proteins to membranes.
- Proteins that are to be GPI anchored are synthesized with membrane-spanning domains that are removed in the GPI addition process.



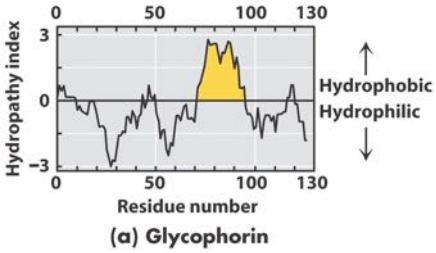
Phosphatidylinositol

Glycophorin A

- Integral membrane proteins can be distinguished from soluble proteins by the abundance of hydrophobic amino acids (Ile, Leu, Val) they contain.
- In integral proteins, hydrophobic residues often occur in blocks or segments ~20-39 res. in length.
- It is often possible to identify membrane proteins from their primary sequence based on the distribution of res. with hydrophobic side chains.
- Hydropathy plots: amino acid residues are assigned numerical values based on their hydrohpobicity. Plot hydropathy of amino acids and their position in sequence.
- The erythrocyte protein glycophorin was the first membrane protein to be sequenced, and a hydropathy plot for glycophorin reveals the presence of a transmembrane segment flanked by N- and C-terminal segments that are rich in polar residues.



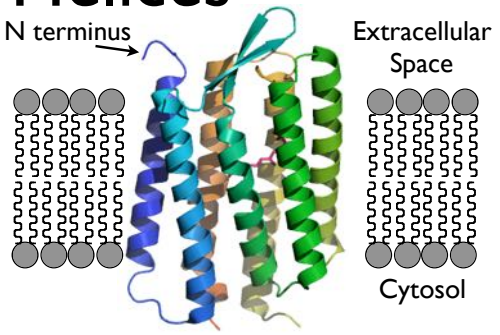
Glycophorin from *Lehninger; Principles of Biochemistry, 4th ed.* **2005**, p 374.

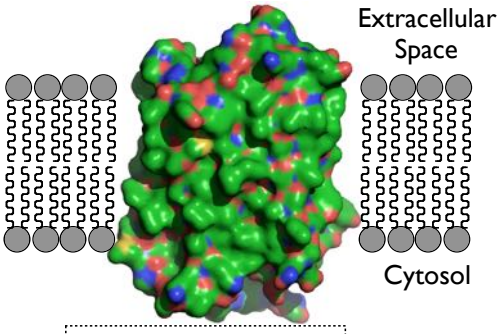


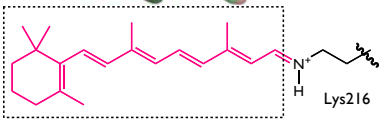
Hydropathy Plot from *Lehninger; Principles of Biochemistry, 4th ed.* **2005**, p 377.

Bacteriorhodopsin and Seven Transmembrane Helices

- Bacteriorhodopsin is a 247-residue membrane protein from the extremophile *H. halobium*.
- It accumulates in the membranes, forming patches ~0.5 μm across that consist primarily of bacteriorhodopsin.
- Bacteriorhodopsin harnesses light energy for the production of ATP using a light driven proton pump.
- Determination of the structure of bacteriorhodopsin, at low resolution in 1975, provided the first opportunity to study the organization of secondary and tertiary structure in a membrane protein.
- Bacteriorhodopsin contains a retinal chromaphore that is covalently linked to Lys216 by means of a Schiff base. [Proton pump is driven by the photo-induced *trans*/*cis* (13-*cis*) isomerization of retinal].



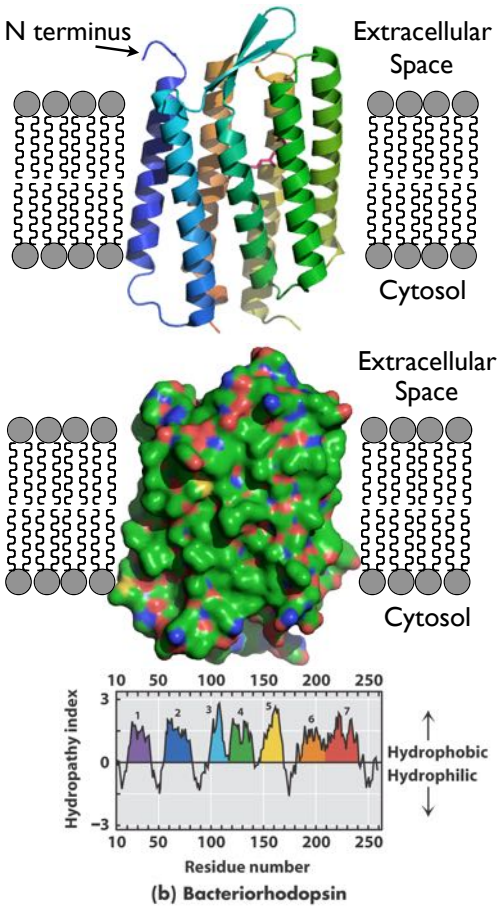




Retinal cofactor

Bacteriorhodopsin and Seven Transmembrane Helices

- Henderson and Unwin used electron crystallography to provide a low resolution structure of bacteriorhodopsin.
- The structure indicated that bacteriorhodopsin contained seven transmembrane helices.
- The amino acid sequence of bacteriorhodopsin was solved soon afterwards and the structural and sequence data were compared.
- Hydropathy plots of the amino acid sequence indicated the presence of seven hydrophobic segments... a block of seven transmembrane helices.
- 1.55 Å resolution structure of bacteriorhodopsin was later reported by Luecke et al.
- Structure confirmed the presence and relative position of the seven transmembrane helices, as well as short connecting loops.



Hydropathy Plot from *Lehninger; Principles of Biochemistry, 4th ed.* **2005**, p 377.

Fibrous Proteins

Fibrous Proteins

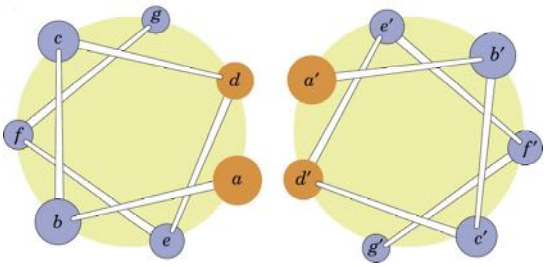
- Most fibrous proteins play structural roles and have regular, extended structures that represent a level of organization intermediate between pure secondary structure and tertiary structures of globular proteins.
- Fibrous proteins are highly elongated molecules/assemblies.
- In fibrous proteins, the secondary structure becomes the dominant structural motif.
- Fibrous protein are structurally simple relative to globular proteins.
- Information regarding the structures and interactions involved in fibrous proteins are limited due to their limited solubility in water and the fact that they do not readily form crystals for X-ray diffraction.

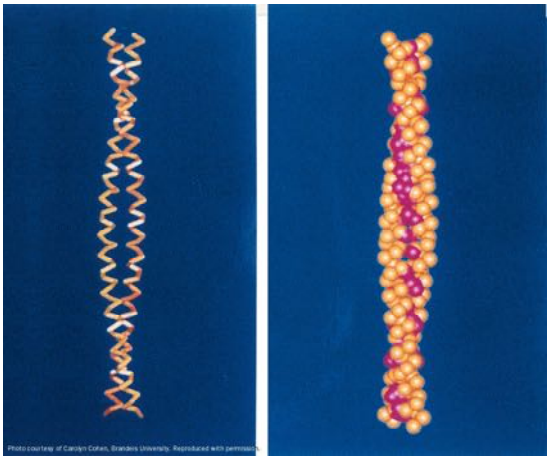
Keratins

- Found in hair, wool, nails, claws, quills, hooves, the outer layer of skin. Keratin is mechanically durable and chemically unreactive.
- Classified as α keratins (mammals) or β keratins (birds and reptiles).
- May comprise as much as 85% of cellular protein. Keratins are the most abundant proteins in epithelial cells.

α Keratin: a Helix of Helices

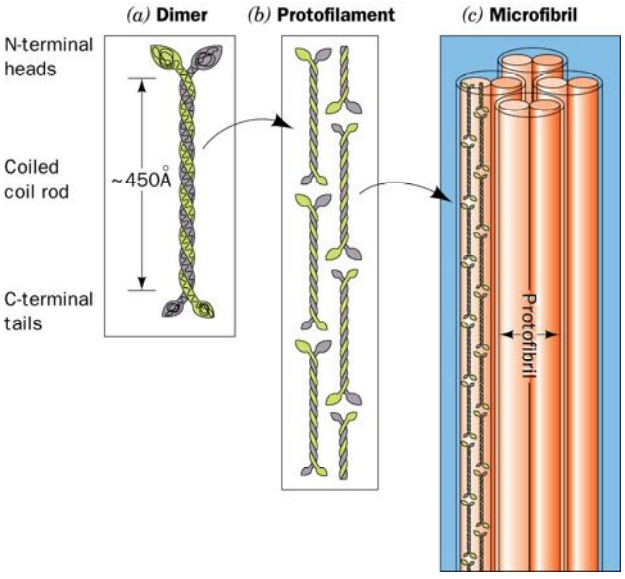
- A wide variety of the structural proteins involved in maintaining cell shape, organizing cytoplasm and movement are coiled coils of two or three α-helices wound around each other forming a left-handed super-helix.
- In forming a coiled-coil structure, the α-helices are distorted slightly from their normal geometry.
- In forming a coiled-coil structure, α-helices come together and interact with each other through hydrophobic residues that form apolar stripes along one face of each helix.
- Distortion of the α-helices results in a slightly tighter winding of the helix (3.5 residues per turn). Therefore, α-helices, which form coiled coils are characterized by a regularity of the amino acid sequence, which repeats every seven residues (heptad repeat: *a-b-c-d-e-f-g*).





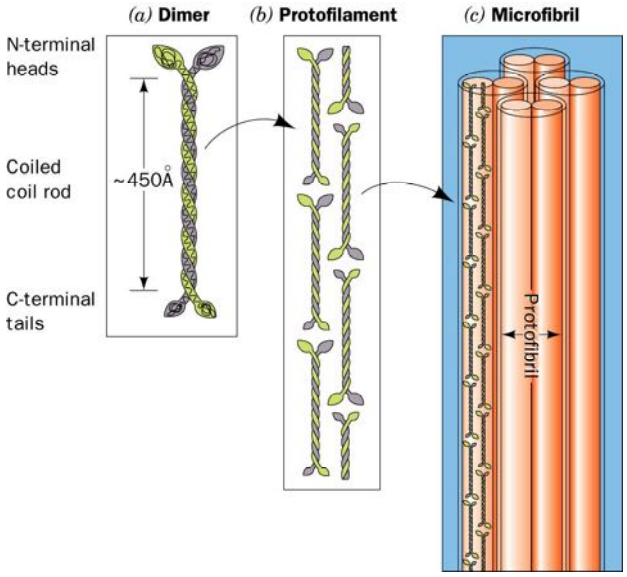
α Keratin: a Helix of Helices

- Higher order construction of α keratin is poorly understood.
- Structure of α keratin is dominated by a coiled-coil extending 300-330 residues on average. The coiled coil is flanked by *N*- and *C*-terminal domains.
- *N*- and *C*-terminal domains vary greatly in size (10 - 500 residues), and they show greater sequence variability than the coiled-coil regions.
- Terminal domains may contribute to functional specificity.
- Keratins are often classified as intermediate filaments (IF) -- components of the large group of cytoskeletal filamentous proteins.
- At least six different IF have been identified. Types I and II identified as "acidic" and "basic" keratins based on the nature of the *N*- and *C*-terminal domains of the keratin coiled-coils.



α Keratin: a Helix of Helices

- Acidic and basic monomers come together to form a heterodimeric coiled-coil.
- Two protofilaments come together to a ~50Å protofibril -- coiled-coils align head to tail forming two staggered rows.
- Four protofibrils come together to form a ~80Å microfibril.
- Microfibrils come together to form ~2000Å macrofibrils.
- α keratin is rich in cysteine residues forming both inter- and intra-strand disulfide bonds.
- Disulfide bonds contribute to mechanical robustness and rigidity.
- Over 30 types of keratin have been identified. Not uniformly distributed -- different keratin have different preferred locations.

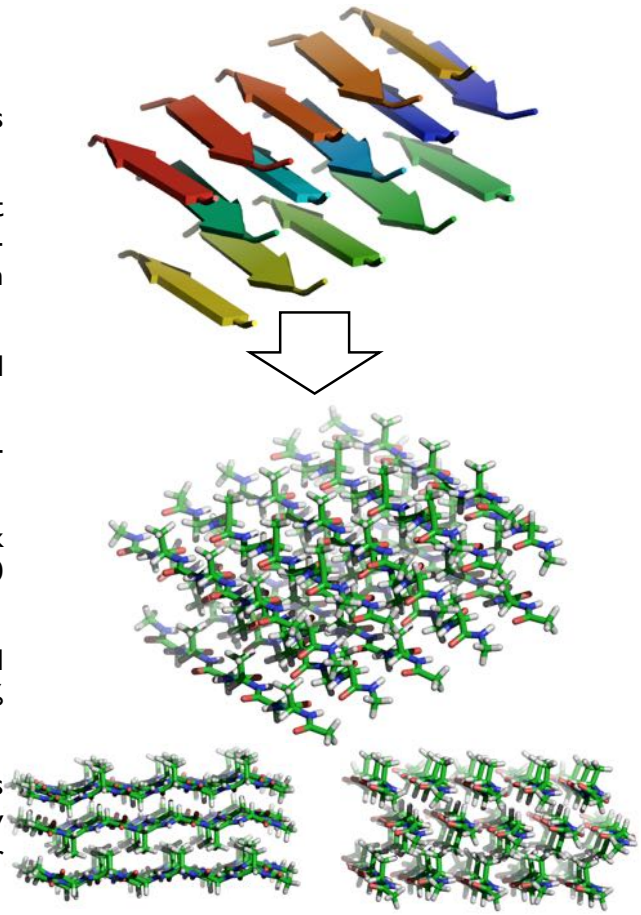


Silk Fibroin

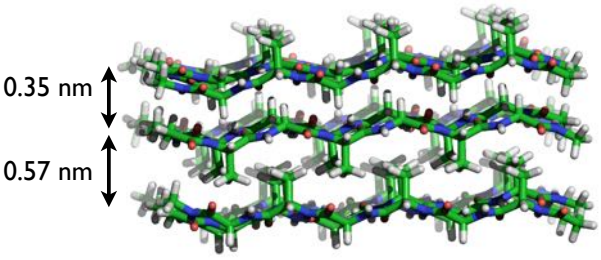
- The structural protein synthesized by spiders for webs and by silkworms for cocoons.
- The fibroin proteins, such as silk, are thought to consist of extended arrays of antiparallel β-sheets, with irregular regions of unknown structure linking the strands.
- The β-sheet regions consist of a repeated sequence:

  -(Gly-Ala)$_2$-Gly-Ser-Gly-Ala-Ala-Gly-(Ser-Gly-Ala-Gly-Ala-Gly)$_8$-Tyr-

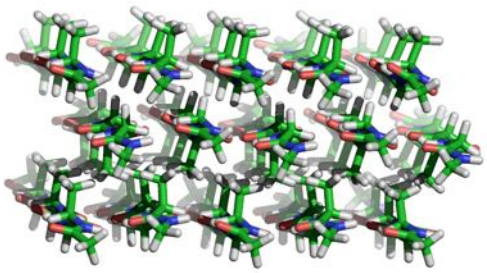
- May be repeated 50 times resulting in silk polypeptides with massed ranging from 300 and 400 kD.
- Gly/Ala/Ser residues may make up 85% of total amino acid compostion (Gly = 45%, Ala = 30% and Ser = 15%).
- The order of Gly and Ala/Ser residues suggests that the β-sheets have predominantly Gly residues on one face of the sheet and Ala/Ser on the other.

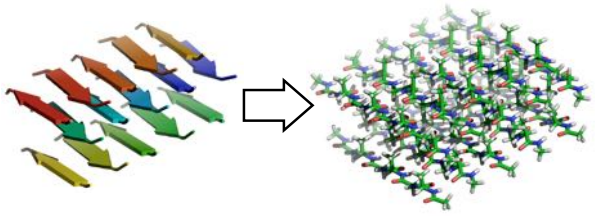


Silk Fibroin

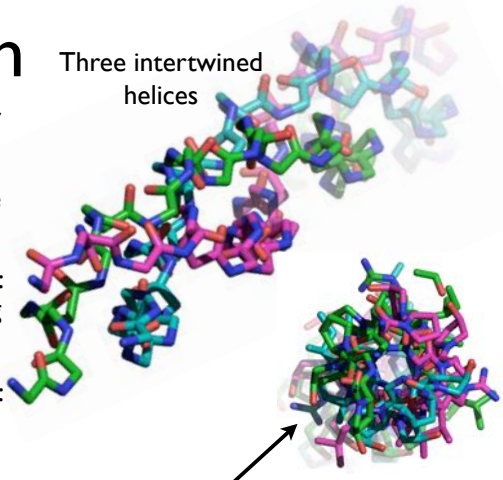


- The sheets are thus stacked on top of each other, with the Gly-rich surfaces packed against each other with the Ala/Ser faces packed similarly.
- Larger side chains can't be accommodated by the tight backing between the sheets, and are usually located in regions linking the β-strands. (These regions have not been clearly defined)
- The silk polypeptide is stored initially as a concentrated concentrated aqueous solution, with a structure resembling random coil.
- Silk is extremely strong because the polypeptide strands in the antiparralel β-sheets are in a fully extended conformation.

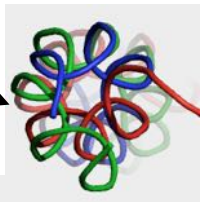


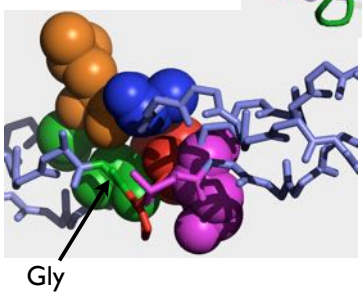


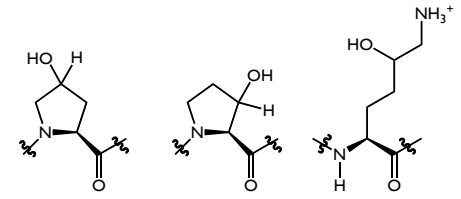
Collagen



- Collagen, an extracellular protein found in all multicellular organisms.
- Found in connective tissue such as tendons, cartilage, the organic matrix of bone, and the cornea of the eye.
- Collagen is the main constituent of higher animal frameworks: bones, tendons, skin, ligaments, blood vessels and supporting membranous tissues.
- Collagen peptides are distinct in their repetitive sequences: every third res. is Gly [i.e. (-Gly-Xaa-Yaa-)$_n$-].
  - ★ Pro residues are frequently in the Xaa and Yaa positions.
  - ★ Many of the Pro (and Lys) residues in the Yaa position are hydroxylated.
  - ★ Nearly 1/3 of the amino acid residues in collagen are Gly, while 15-30% are Pro or 4-hydroxyprolyl (Hyp) residues
- The reason for the repeating sequence becomes apparent from the structure of collagen.



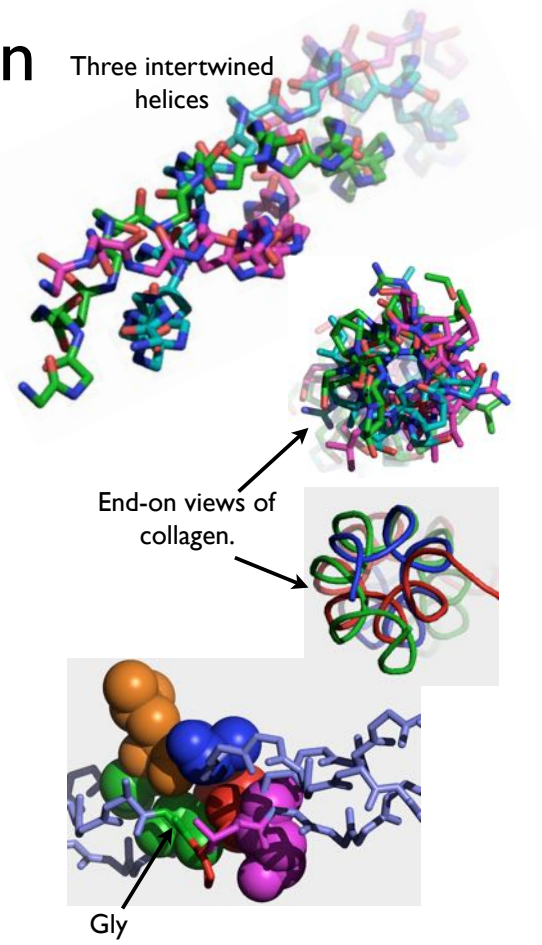




4-hydroxyprolyl residue | 3-hydroxyprolyl residue | 5-hydroxylysyl residue

Collagen

- The reason for the repeating sequence becomes apparent from the structure of collagen.
- Three collagen polypeptide chains (left-handed helical conformation) wind around each other (aligned in parallel) in a right-handed fashion to form a triple helical coiled coil.
- Every third residue must pass through the center of the triple helix, which explains why every third amino acid residue must be Gly.
- The polypeptide chains are staggered so that Gly residues are not all in a single line.
- Hydrogen bonding occurs between the N-H of each Gly and the carbonyl oxygen of a residue in the Xaa position.
- Pro and Hyp confer rigidity on the chain, and the well-packed helical structure gives collagen great tensile strength.
- Collagen assembly is aided by the presence of 100 to 300 residue pro-segments that are cleaved following assembly.



Collagen

- Some types of collagen assemble to form distinctive banded fibrils.
- Driving force for fibril formation is related to the hydrophobic interactions associated with the resulting fiber.
- The packing of the collagen molecules results in the banding pattern observed in collagen fibrils (gaps between aligned collagen molecules).
- Cross-linking: collagen is inter and intramolecularly cross-linked near the N- and C-termini (formed between Lys and His residues from four chains).
- Collagen fibrils are organized in tissues based on the function and the nature of the stress experienced by the tissue.
- O-glycosylation (mostly glucose, galactose and disaccharides) of Hyl residues (primarily in the gap regions).

END