Protein Structure

- Why protein folds?
- Secondary structure
  - Alpha-helix (α-helix)
  - Beta-sheet (beta-conformation)
  - Beta-turn (β-turn)
  - Ramanchandran plot
- Tertiary & Quaternary Structure
- Motif & Domain
- Stable conformation
- Visualization of protein
- Denature of protein

Much More About Structure

Structure ⟷ Function

Structure ⟷ Mechanism

Structure ⟷ Origins/Evolution

Structure-Based Drug Design

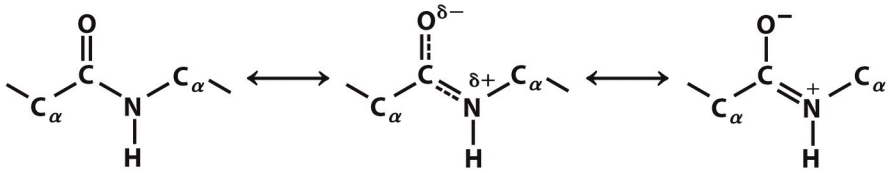
Solving the Protein Folding Problem

Protein Conformation

Native protein

Why Protein Folds?

Peptide bond
Disulfide bond
Hydrogen bonding
Ionic interaction
Hydrophobic-lipophilic interaction
Van der Waals interaction
Rotational prohibition



The carbonyl oxygen has a partial negative charge and the amide nitrogen a partial positive charge, setting up a small electric dipole. Virtually all peptide bonds in proteins occur in this trans configuration; an exception is noted in Figure 4–7b.

Figure 4-2a
Lehninger Principles of Biochemistry, Fifth Edition

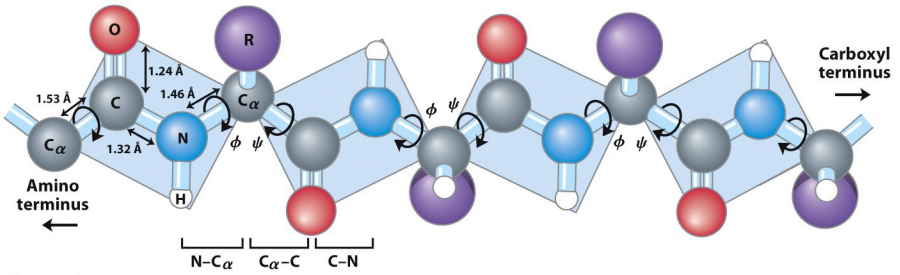


Figure 4-2b
Lehninger Principles of Biochemistry, Fifth Edition

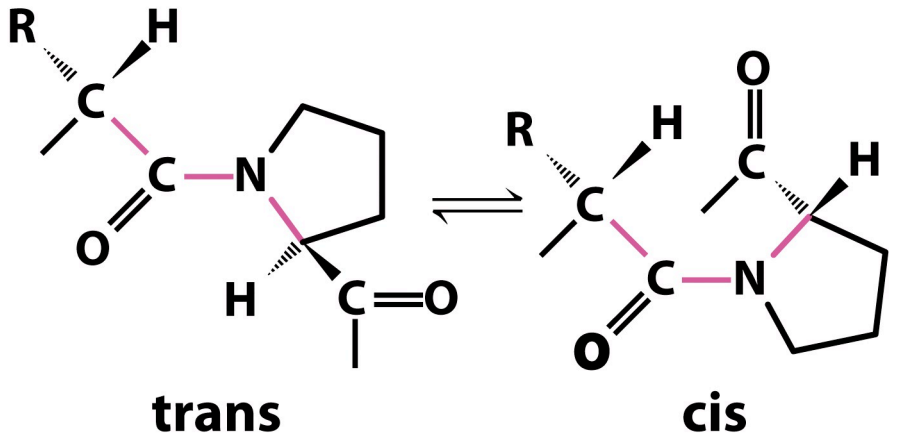
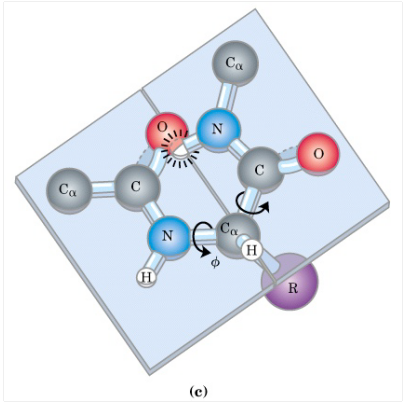
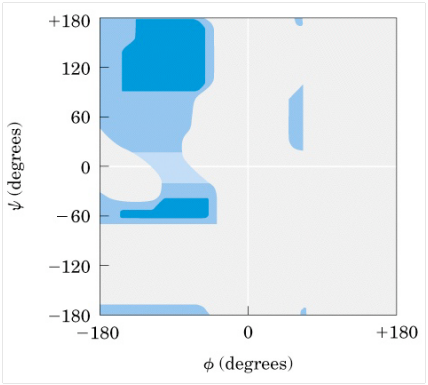


Figure 4-7b
Lehninger Principles of Biochemistry, Fifth Edition



(c)



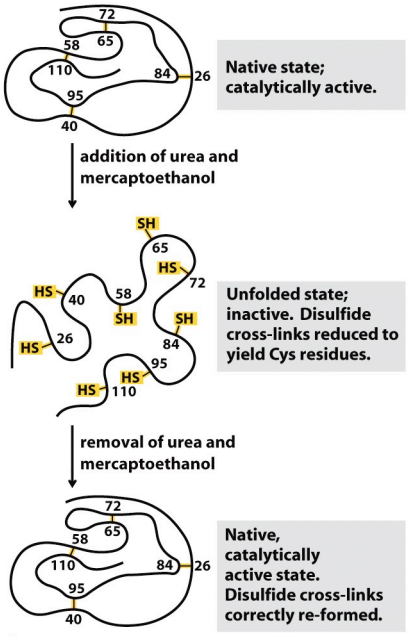
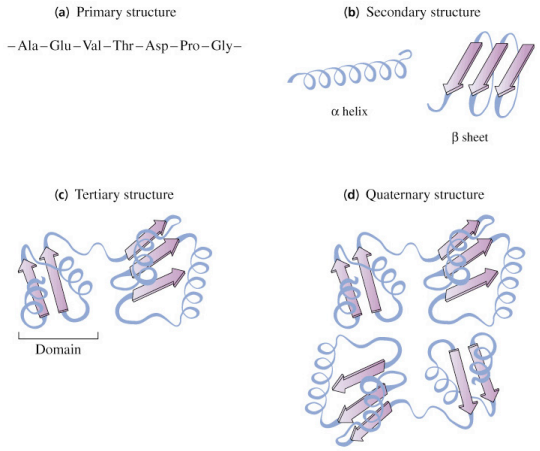


Figure 4-26
Lehninger Principles of Biochemistry, Fifth Edition

Levels of Protein Structure

- Primary
  - Linear AA sequence
  - Covalent bonds
- Secondary
  - Local structure; certain "motifs" are common
  - Mostly H-bonds
- Tertiary
  - Complete 3D shape
  - H-bonds, hydrophobic interactions, ionic bonds, van der Waals interactions, disulfide bonds
- Quaternary
  - >1 peptide chain
  - Mostly H-bonds



Protein Folding

- **Folded shape = conformation**
- **Three-dimensional, functional structure = native**
  - **Energy of native conformation?**
- **Molecular chaperones**
- **There are thousands of possible conformations, but not an infinite amount…**
- **Conformations are restrained by**
  - **planarity of peptide bond**
  - **"allowed" angles**
- **No algorithm predicts the 3D shape with high accuracy**

Primary Structure

- AA sequence of polypeptide chain(s)
- Linked by peptide bonds
- Linear sequence
- Predication of primary structure?
- Experimental determination: protein sequencing

Secondary Structure

- Regular repeating structure
  - Helices
  - Sheets
- Torsion/dihedral angles
  - Angles of rotation around C_α
  - Clockwise (+) and counterclockwise (-)
  - Φ = rotation around C_α-N
  - Ψ = rotation around C_α-C
- How free is rotation?
  - Not very (sterics)
  - Avoid collision of C=O, N-H, R
  - Calculations of allowed values = Ramachandran diagram

Alpha Helix

- (30-35%)
- Φ = -57°, Ψ = -47°
- Discovered by Pauling: 1951
- α-helix formers: A,C,L,M,E,Q,H,K
- Tightly wound, repeating sequence
- "Right-handed"
- Each twist ~ 5.4 Å; 3.6 residues
- Average length = 18 residues
- R-groups are on outside of helix
- Stabilized by H-bonds between C=O (i) and N-H (i + 4)

Alpha helix, cont.

- Deviates from ideal conformation at ends (less H-bonding)
- Some amino acids are "α-helix breakers"
  - Repeating like-charges
  - Repeating "bulky" groups
  - Pro and Gly
- Effects on helical stability:
  - Electrostatic interactions between adjacent residues
  - Steric interference between adjacent residues
  - Interactions between residues 3-4 amino acids away
  - Polarity of residues at both ends of helix (positive at amino end; negative at carboxyl)

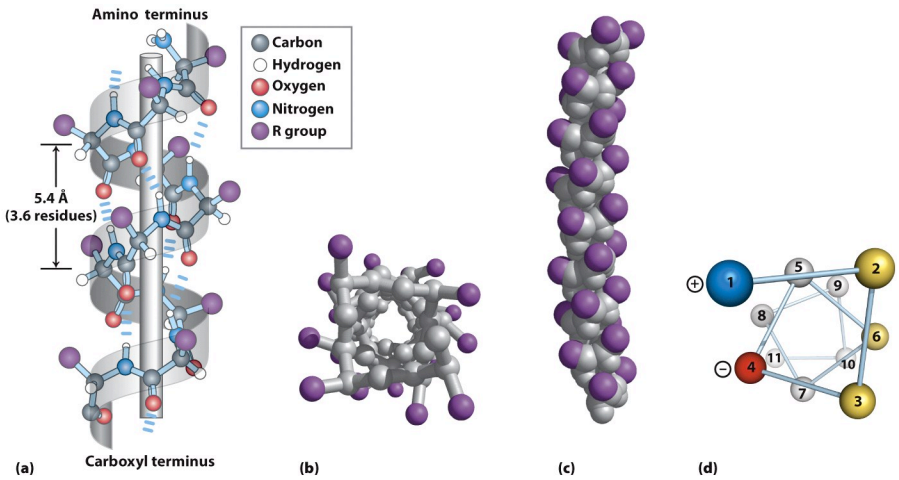
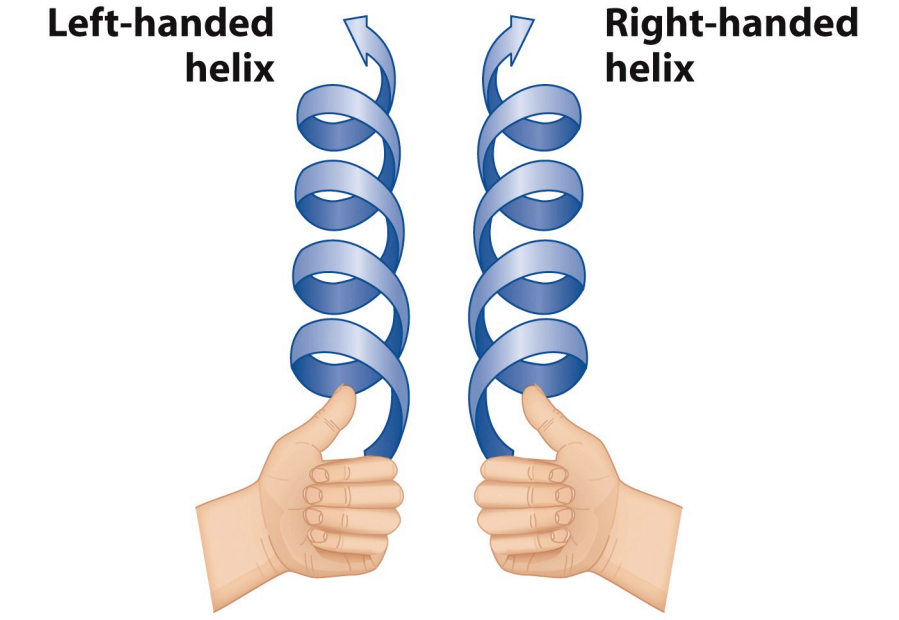


Figure 4-4
Lehninger Principles of Biochemistry, Fifth Edition



Box 4-1
Lehninger Principles of Biochemistry, Fifth Edition

TABLE 4–1 Propensity of Amino Acids to Take Up an α-Helical Conformation

Amino acid	$\Delta\Delta G°$ (kJ/mol)*	Amino acid	$\Delta\Delta G°$ (kJ/mol)*
Ala	0	Leu	0.79
Arg	0.3	Lys	0.63
Asn	3	Met	0.88
Asp	2.5	Phe	2.0
Cys	3	Pro	>4
Gln	1.3	Ser	2.2
Glu	1.4	Thr	2.4
Gly	4.6	Tyr	2.0
His	2.6	Trp	2.0
Ile	1.4	Val	2.1

Sources: Data (except proline) from Bryson, J.W., Betz, S.F., Lu, H.S., Suich, D.J., Zhou, H.X., O'Neil, K.T., & DeGrado, W.F. (1995) Protein design: a hierarchic approach. *Science* **270**, 935. Proline data from Myers, J.K., Pace, C.N., & Scholtz, J.M. (1997) Helix propensities are identical in proteins and peptides. *Biochemistry* **36**, 10,926.

*$\Delta\Delta G°$ is the difference in free-energy change, relative to that for alanine, required for the amino acid residue to take up the α-helical conformation. Larger numbers reflect greater difficulty taking up the α-helical structure. Data are a composite derived from multiple experiments and experimental systems.

Table 4-1
Lehninger Principles of Biochemistry, Fifth Edition

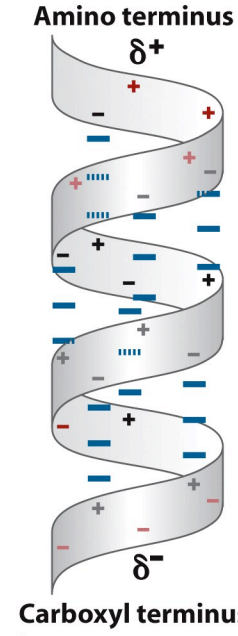


Figure 4-5
Lehninger Principles of Biochemistry, Fifth Edition

β-Sheet / β-Strand

- Extended, zigzag conformation (20-25%)
  - Hydrogen bond between groups across strands
  - Forms parallel and antiparallel pleated sheets
  - Residues alternate above and below β-sheet
  - β-sheet formers: V, I, P, T, W
- Intrastrand H-bonding
- Average 6 residues/strand; up to 15
- 2-12 strands/sheet; average 6
- R-groups alternate on opposite sides of sheet
- Distortions:
  - Beta-bulge = extra residue
  - Kink = Pro

Anti-parallel vs. Parallel

- Anti-parallel β-sheet
  - Opposite orientation
  - Φ = -140°, Ψ = 135°
  - More stable
  - Can be twisted
  - 6.5 Å per two amino acid residues
  - Can withstand distortions and exposure to solvent
- Parallel β-sheet
  - Same amino-carboxyl direction
  - Less twisted
  - Tend to be buried
  - Φ = -120°, Ψ = 115°
  - 7.0 Å per two amino acid
- Can have mix of parallel and anti-parallel

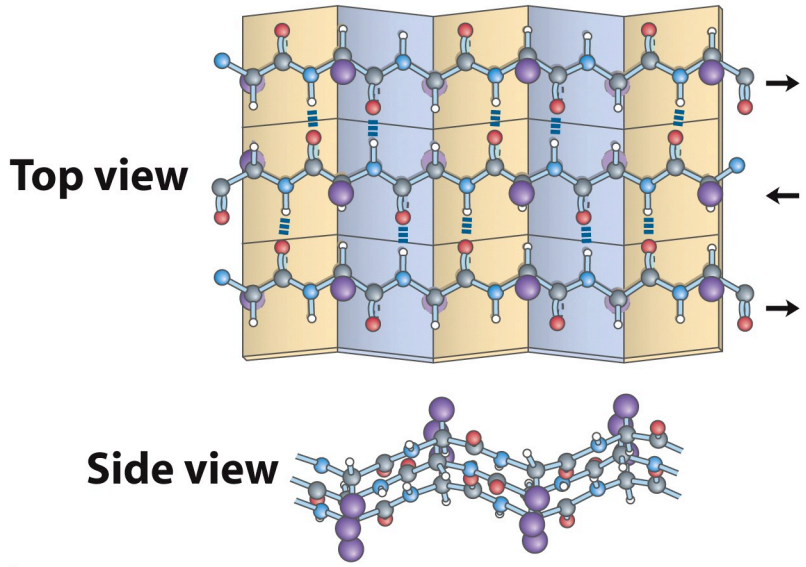


Figure 4-6a
Lehninger Principles of Biochemistry, Fifth Edition

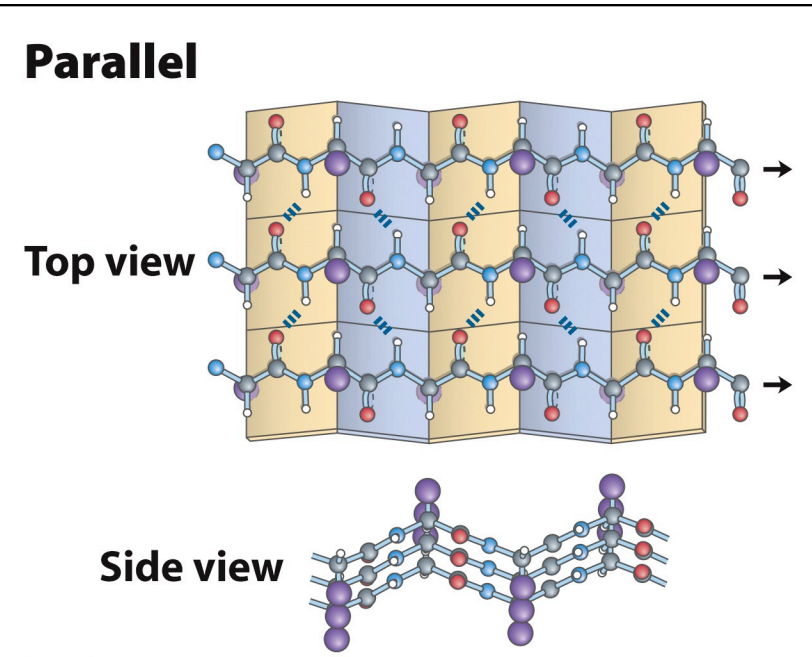


Figure 4-6b
Lehninger Principles of Biochemistry, Fifth Edition

β-turns

- Interacting strands can be many amino acids apart
- Turns are 180°; "connect" strands in folded (globular) proteins
- Interaction is between carbonyl oxygen of AA 1 and amino hydrogen of AA 4
  - Short turn (4 residues)
  - Hydrogen bond between C=O & NH groups within strand (3 positions apart)
  - Usually polar, found near surface
  - β-turn formers: S, D, N, P, R
- Pro and Gly are often present
  - Gly: small and flexible (Type II turns)
  - Pro: cis conformation makes inclusion in tight turn favorable

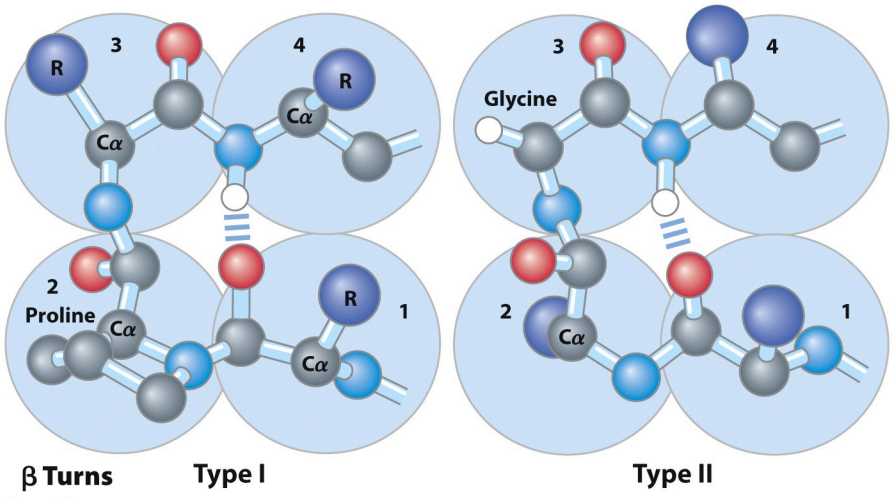


Figure 4-7a
Lehninger Principles of Biochemistry, Fifth Edition

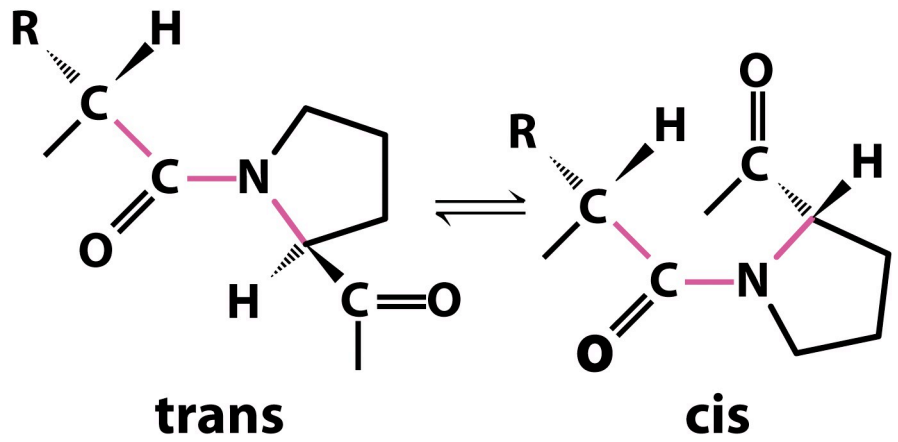


Figure 4-7b
Lehninger Principles of Biochemistry, Fifth Edition

Others

- **Loop**
  - **Regions between α-helices and β-sheets**
  - **On the surface, vary in length and 3D configurations**
  - **Do not have regular periodic structures**
  - **Loop formers: small polar residues**
- **Coil (40-50%)**
  - **Generally speaking, anything besides α-helix, β-sheet, β-turn**

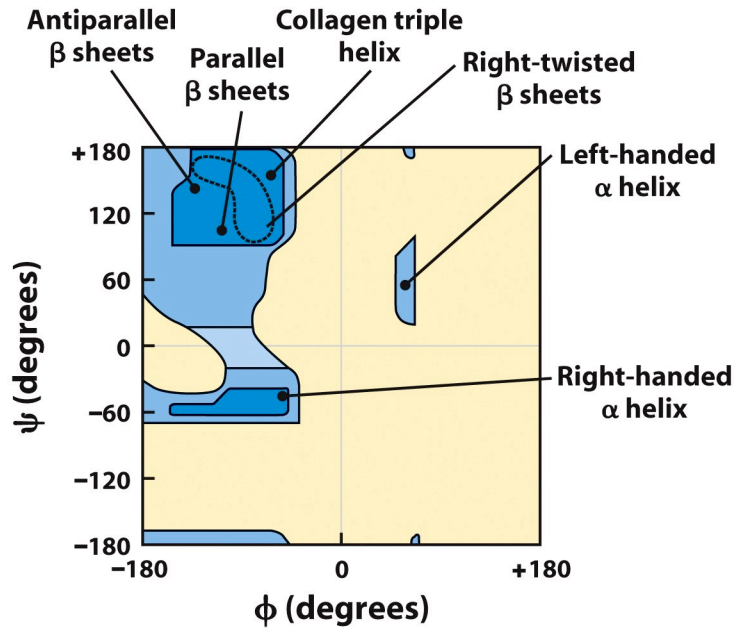


Figure 4-8a
Lehninger Principles of Biochemistry, Fifth Edition

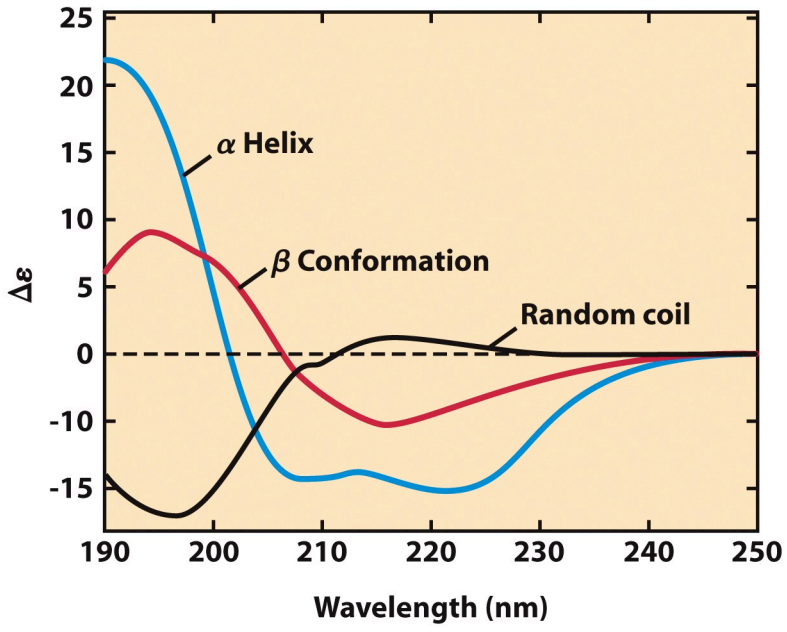
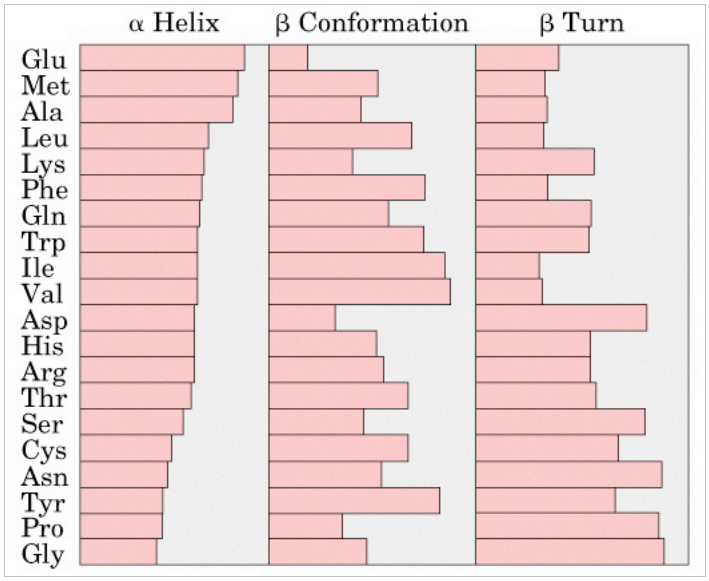
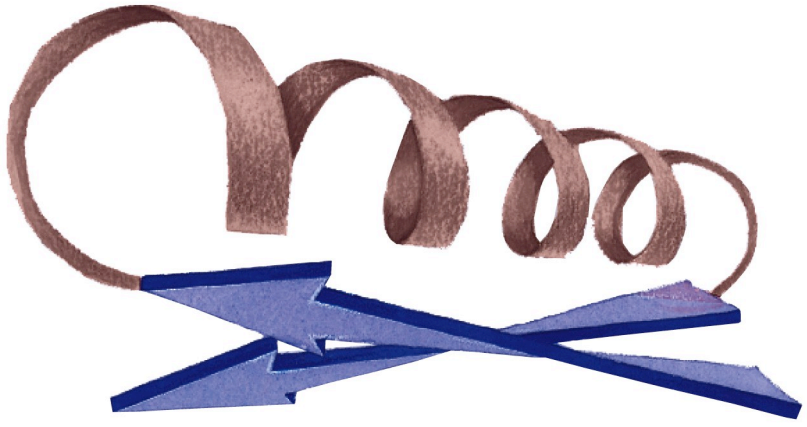


Figure 4-9
Lehninger Principles of Biochemistry, Fifth Edition





β-α-β Loop

Figure 4-17a
Lehninger Principles of Biochemistry, Fifth Edition

β Barrel

Figure 4-17b
Lehninger Principles of Biochemistry, Fifth Edition

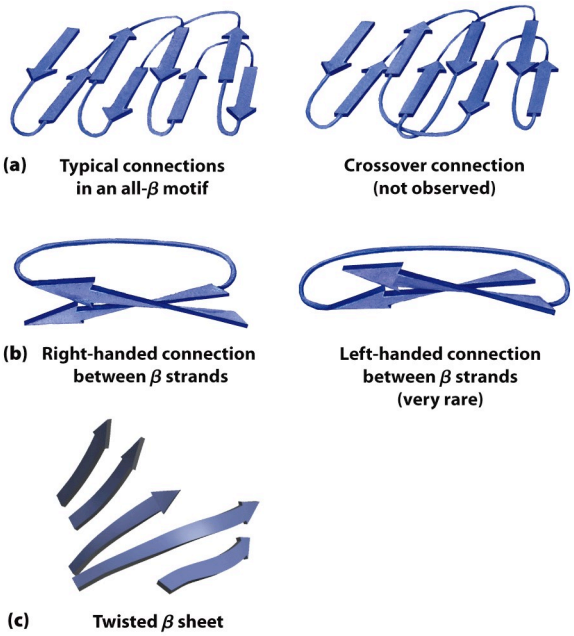


Figure 4-19
Lehninger Principles of Biochemistry, Fifth Edition

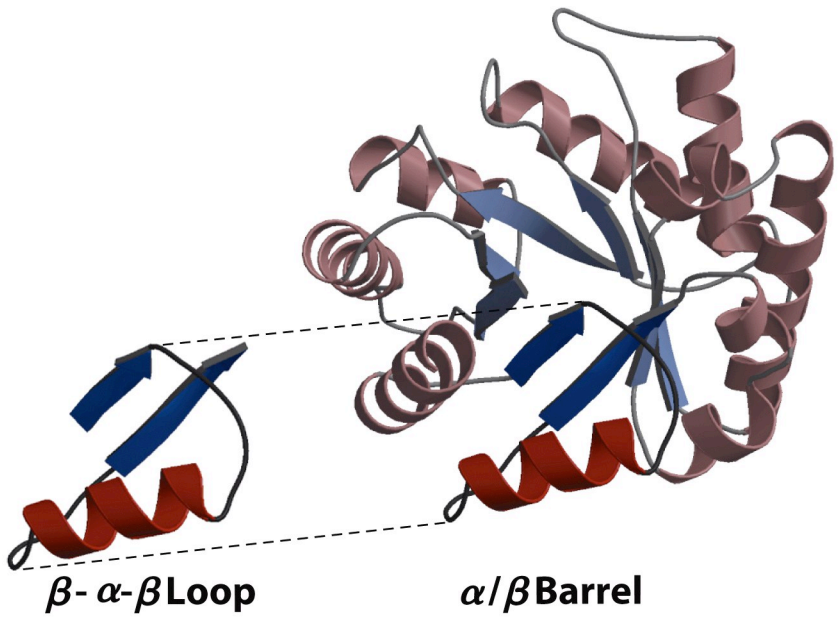


Figure 4-20
Lehninger Principles of Biochemistry, Fifth Edition

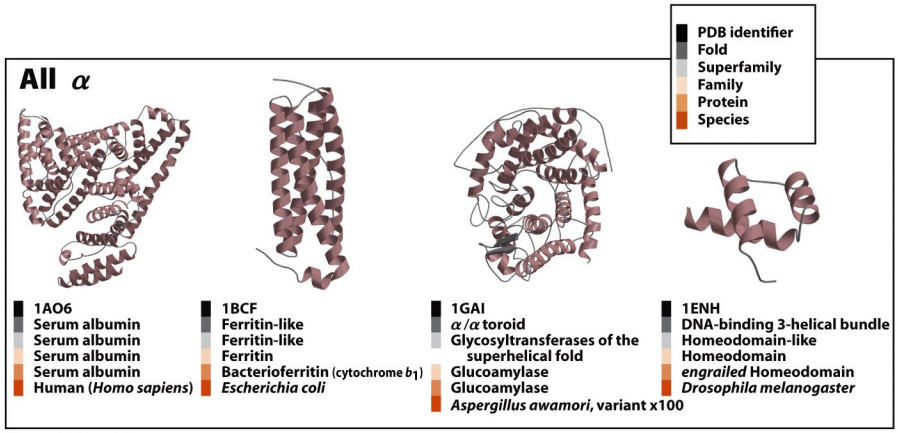


Figure 4-21 part 1
Lehninger Principles of Biochemistry, Fifth Edition

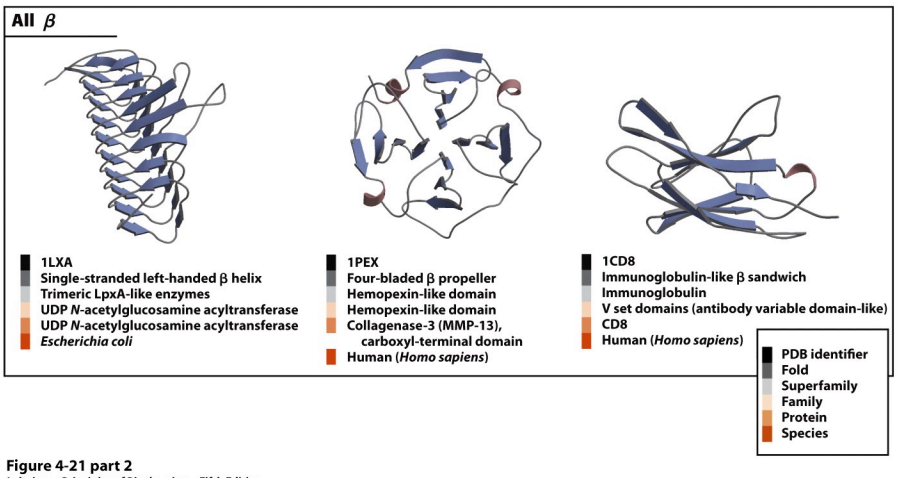


Figure 4-21 part 2
Lehninger Principles of Biochemistry, Fifth Edition

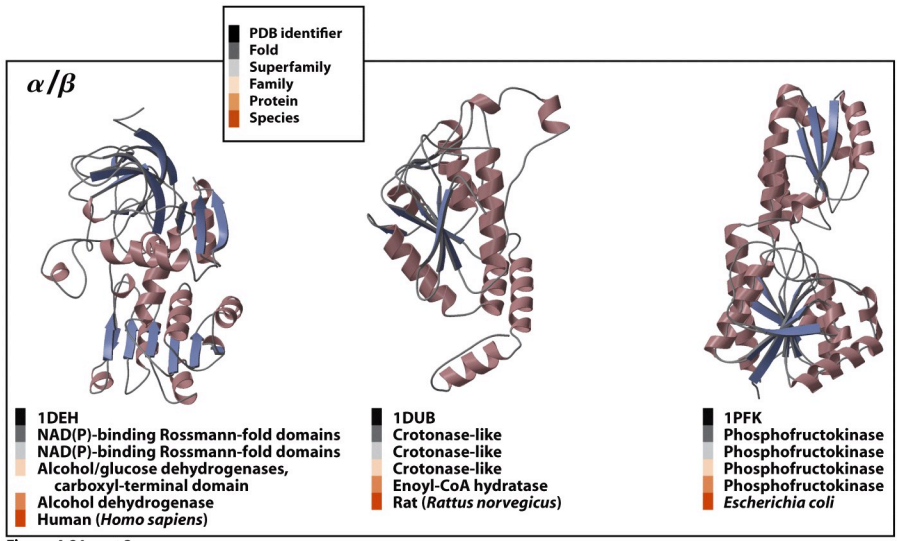


Figure 4-21 part 3
Lehninger Principles of Biochemistry, Fifth Edition

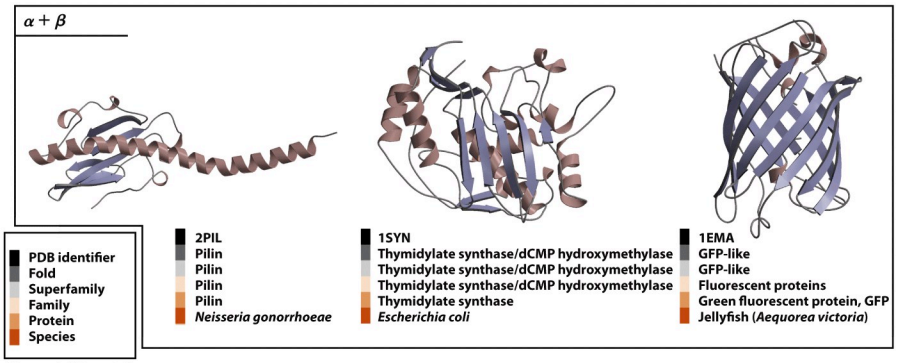


Figure 4-21 part 4
Lehninger Principles of Biochemistry, Fifth Edition

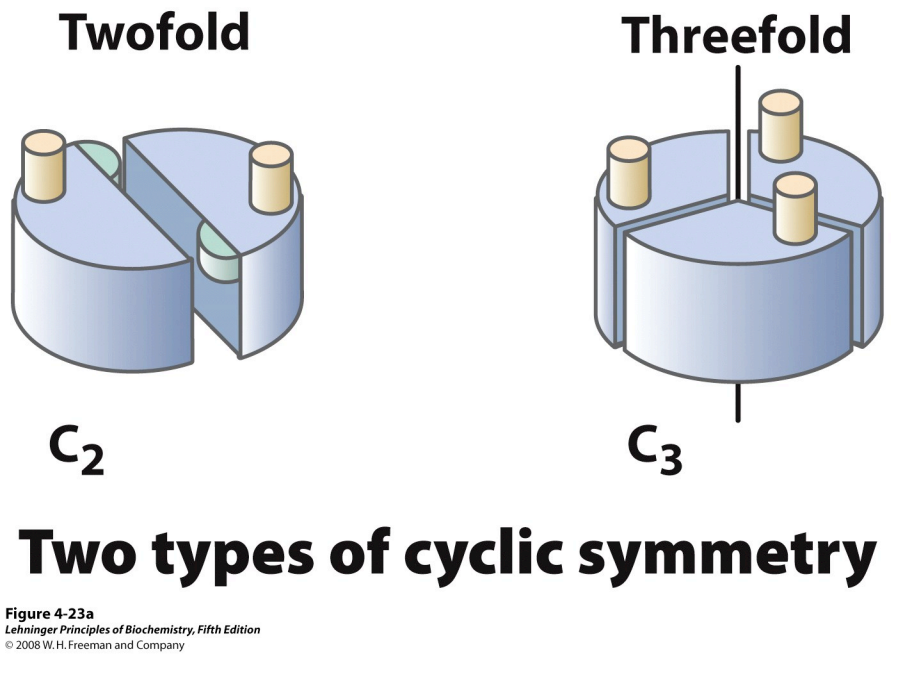
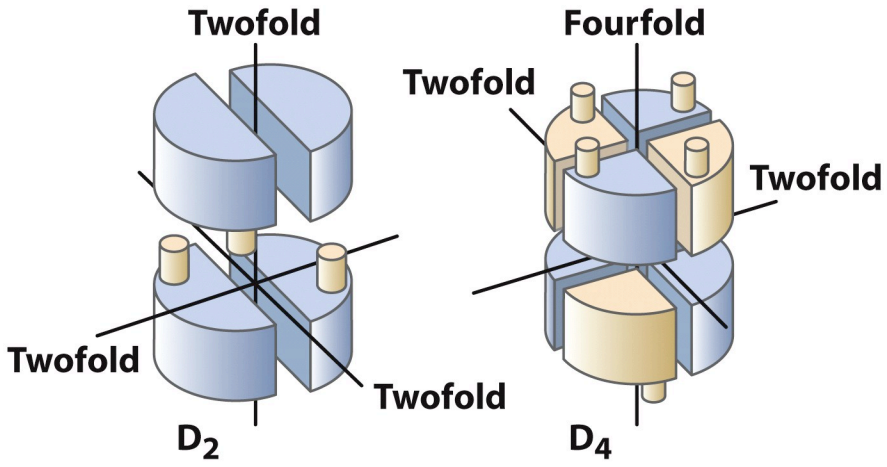
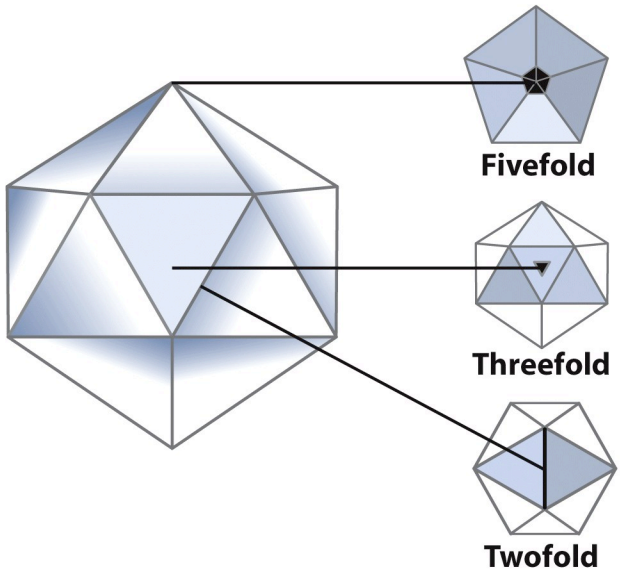


Figure 4-23a
Lehninger Principles of Biochemistry, Fifth Edition



Two types of dihedral symmetry

Figure 4-23b
Lehninger Principles of Biochemistry, Fifth Edition



Icosahedral symmetry

Figure 4-23c
Lehninger Principles of Biochemistry, Fifth Edition

Proteins are Complex

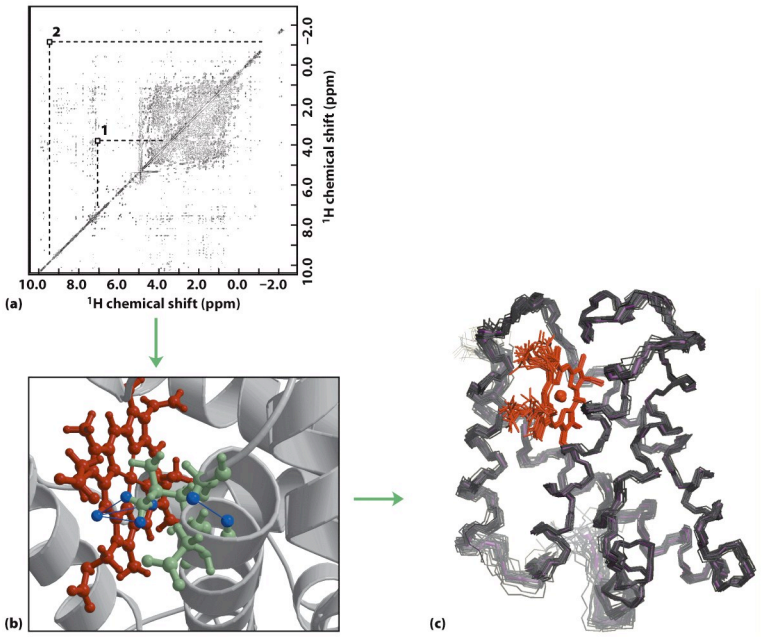
- Average residue contains 8 “heavy” atoms
- Average protein contains 300 amino acids
- Average structure contains 2400 atoms
- First structure (sperm whale myoglobin) took ~5 years with a team of ~15 key punch operators working around the clock to solve
- Most structures still take 1 year to solve

Solving Protein Structures

- Only 2 kinds of techniques allow one to get atomic resolution pictures of macromolecules
- X-ray Crystallography (first applied in 1961 - Kendrew & Perutz)
- NMR Spectroscopy (first applied in 1983 - Ernst & Wuthrich)

X-ray Crystallography

- Crystallization
- Diffraction Apparatus
- Diffraction Principles
- Conversion of Diffraction Data to Electron Density
- Resolution
- Chain Tracing



Box 4-5 figure 3
Lehninger Principles of Biochemistry, Fifth Edition

Classification of Proteins by Secondary Structure

- **Fibrous**
  - **High composition of single secondary structure**
  - **Strong and flexible**
  - **Collagen**
    - **Triple helix (left-handed helices, right-handed super helix)**
  - **Silk fibroin**
    - **Anti-parallel β-sheet**
  - **α-Keratin**
    - **Left-handed coiled coil**
- **Globular**
  - **Majority of all proteins**
  - **Contain several types of secondary structure (regular and non-regular)**
  - **Percentage of protein (on average):**
- **31% α-helix**
- **28% β-sheet**
- **13% turns/bends**
- **28% loops and random coil**

TABLE 4–2 Secondary Structures and Properties of Some Fibrous Proteins

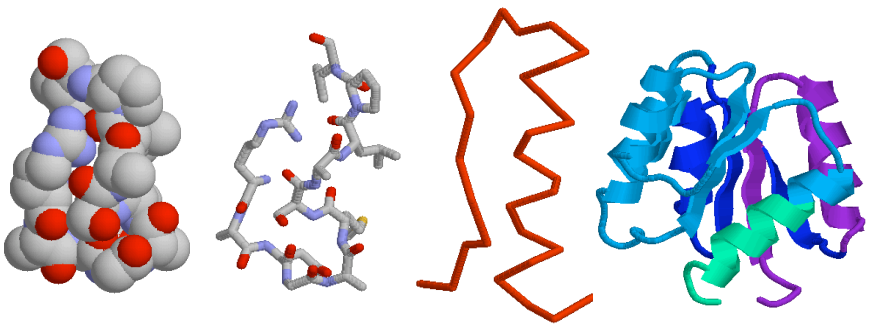
Structure	Characteristics	Examples of occurrence
α Helix, cross-linked by disulfide bonds	Tough, insoluble protective structures of varying hardness and flexibility	α-Keratin of hair, feathers, and nails
β Conformation	Soft, flexible filaments	Silk fibroin
Collagen triple helix	High tensile strength, without stretch	Collagen of tendons, bone matrix

Table 4-2
Lehninger Principles of Biochemistry, Fifth Edition

Aspects Which Determine Tertiary Structure

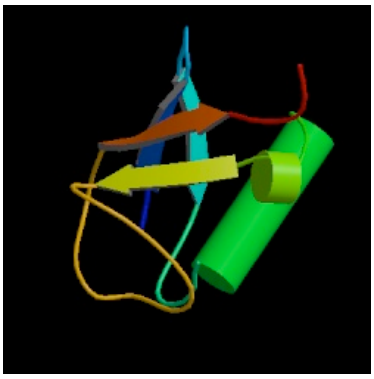
- Covalent disulfide bonds from between closely aligned cysteine residues form the unique Amino Acid cystine.
- Nearly all of the polar, hydrophilic R groups are located in the surface, where they may interact with water
- The nonpolar, hydropobic R groups are usually located inside the molecule

Protein Visualization Model

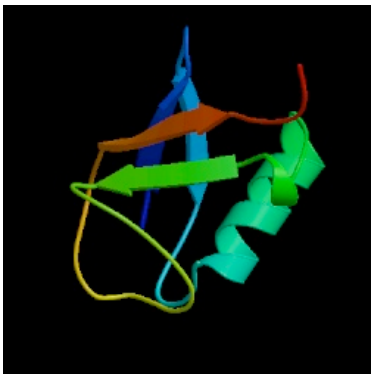


CPK Lin e Trace Cartoon

Protein Visualization Model



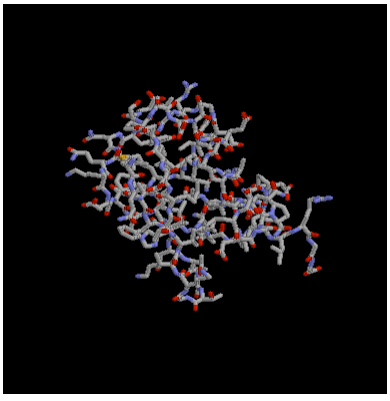
Cylinder



Ribbon (N-C gradient)

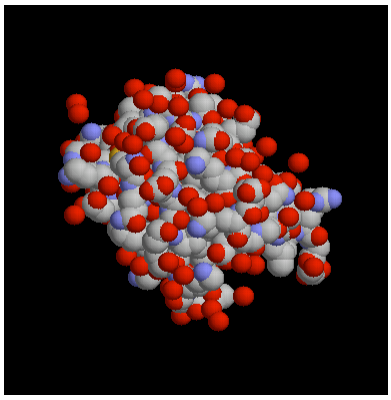
Protein Visualization Model

Ribbon (2º structure)

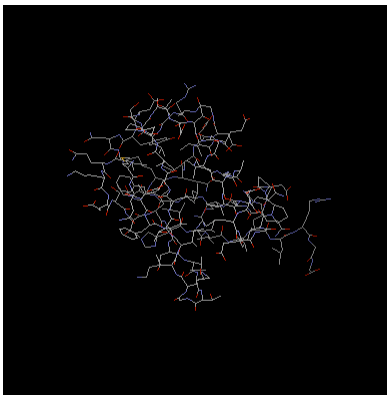


Stick

Protein Visualization Model

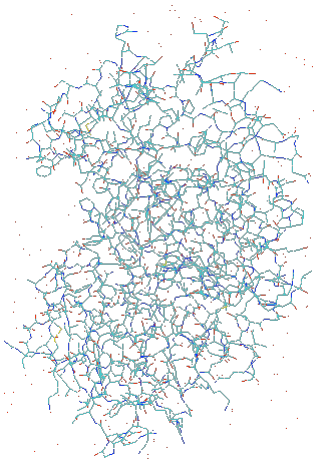


Space Filling

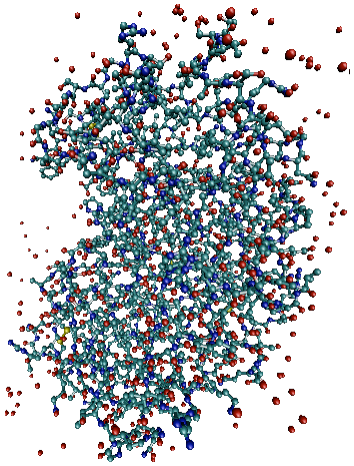


Wire Frame (Vector)

Protein Visualization Model



Wireframe



Ball and stick

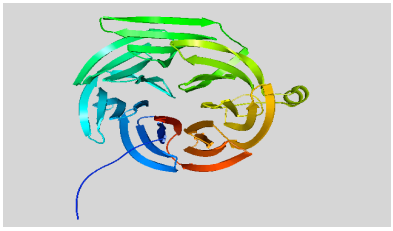
Visualization Software

- Biodesigner
- CACTVS
- Chemdraw net Plugin
- Chemical2vmd
- **Chime**
- Chimera
- Cn3D
- CONSCRIPT
- Dino
- Flex
- FlexV
- Garlic
- Gdis
- gOpenMol
- GRASP
- Hyperactive Molecules Using Chemical MIME
- ICMLite
- ImageMagick
- INTERCHEM modelling software
- Jmol
- Kinemage
- MacMolecule 2 and PCMolecule 2
- Maestro
- Marvin Applets and JavaBeans
- Mercury
- MindTool
- Molden
- MOLEKEL
- MOLMOL
- MolPov
- MolScript
- MolView and MolView Lite
- MSP
- NIH Image
- O
- ORTEX
- POV-Ray
- Pov4Grasp
- PovChem
- Protein Explorer
- PSI88
- PyMOL
- Qmol
- QTree
- **RasMol**
- Raster3D
- Ribbons
- RNA Movies
- RNA Movies
- RnaViz
- Spock
- Swiss-PdbViewer
- Tachyon
- VEGA
- Viewmol
- VMD
- WebLab ViewerLite
- WebMol
- WinMGM
- XMol

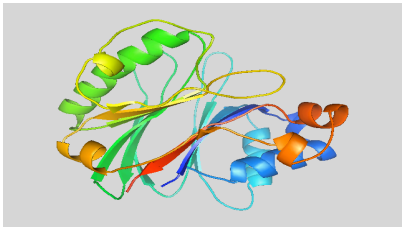
MolScript

- **Kraulis, P. J. (1991). "MolScript: A Program to Produce Both Detailed and Schematic Plots of Protein Structure." J. Appl. Cryst. 24: 946-950**
- **Web Site: http://www.avatar.se/molscript/**
- **MolScript is a program for displaying molecular 3D structures, such as proteins, in both schematic and detailed representations.**
- **Helps from scripting language, we can render the representation by ourselves.**

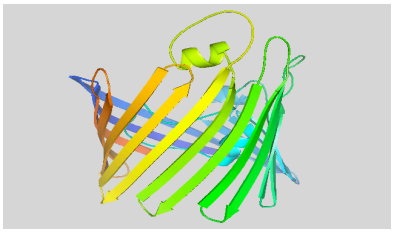
Molscript Representations of Typical Protein Architectures.



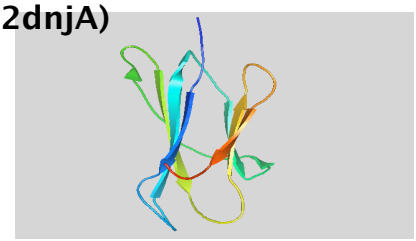
β 7 Propellor (2bbkH)



α 4-Layer Sandwich (2dnjA)

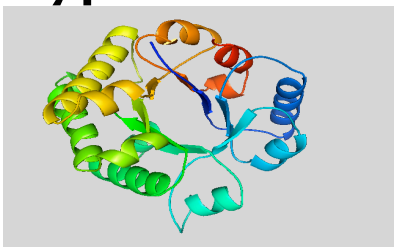


β Barrel (2por)

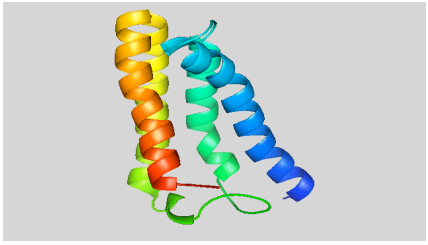


Sandwich (2hlaB)

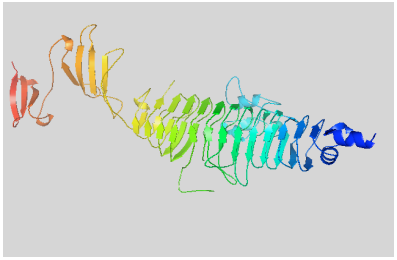
Molscript Representations of Typical Protein Architectures.



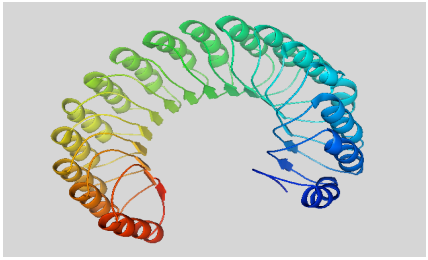
α β Barrel (4timA)



α Helix Bundle (2ccy)



2 Solenoid (1tsp)

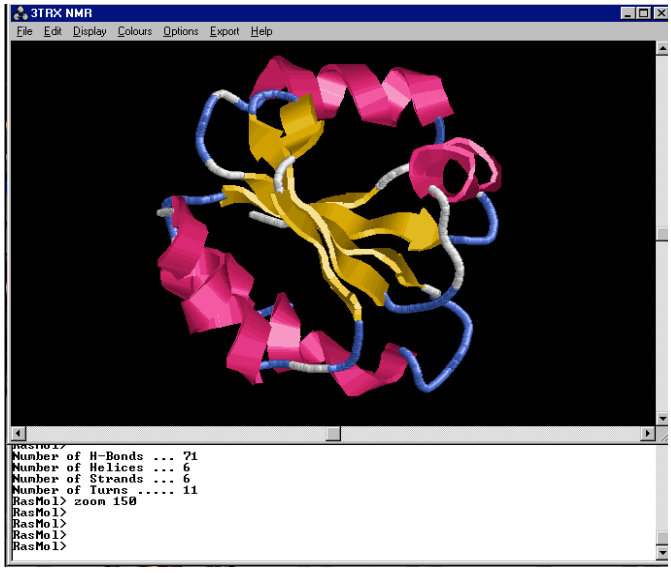


α Horseshoe (1bnh)

RasMol/OpenRasMol

- RasMol is a program for molecular graphics visualization originally developed by Roger Sayle.
- Web Site: http://www.openrasmol.org/
- "Grand-daddy" of all visual freeware
- The most popular one viewer.

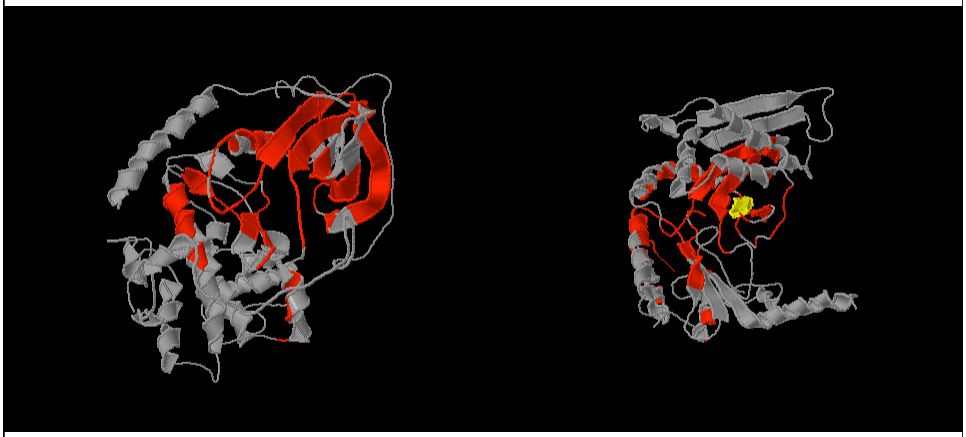
Rasmol



Jmol

- Jmol is a free, open source molecule viewer for students, educators, and researchers in chemistry and biochemistry.
- It's cross-platform, running on Windows, Mac OSX, and Linux/UNIX systems.
- Web Site: http://jmol.sourceforge.net/
- It has more detailed materials on how to write a script language and an interactive web UI to demonstrate.
- It's a Java based application.

1ATP vs. 1BO1



PDBID: 1ATP PDBID: 1BO1

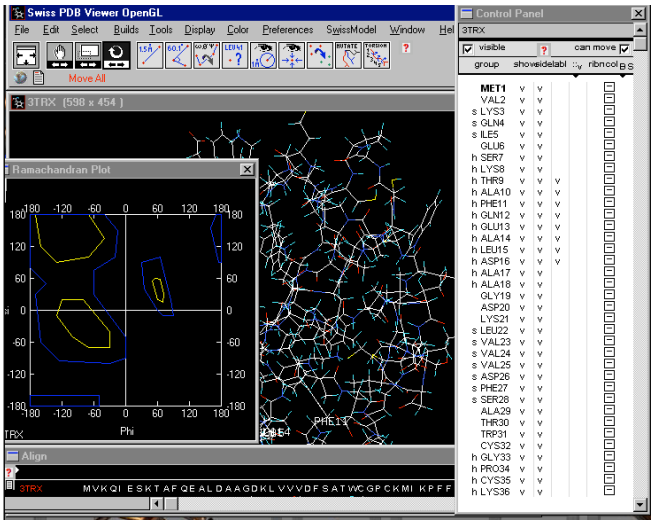
PyMOL

- PyMOL is a user-sponsored molecular visualization system on a open-source foundation.
- Web Site: http://pymol.sourceforge.net/
- It's a Python based application.

DeepView (Swiss-PDB Viewer)

- Swiss-PDB viewer is an application that provides a user friendly interface allowing to analyze several proteins at the same time.
- Web Site: http://www.expasy.org/spdbv/
- Swiss-PDB viewer is not just a viewer and more powerful than the other viewers.
- It also supports SWISS-MODEL, a fully automated protein structure homology-modeling server.
- SWISS-MODEL: An Automated Comparative protein Modeling Server.
- It also provides a function of structure alignment.

Swiss PDB Viewer



Protein Denaturation

The loss of secondary, tertiary, or quaternary protein structure due to disruption of noncovalent interactions and/or disulfide bonds that leaves peptide bonds and primary structure intact.

Factors That Cause Protein to Denature

Heat

pH change

Hydrogen bonding reagents

Detergents

Non-polar solvents

Extremely high or low salt concentration

Oxidants/reductants

Mechanical Agitation

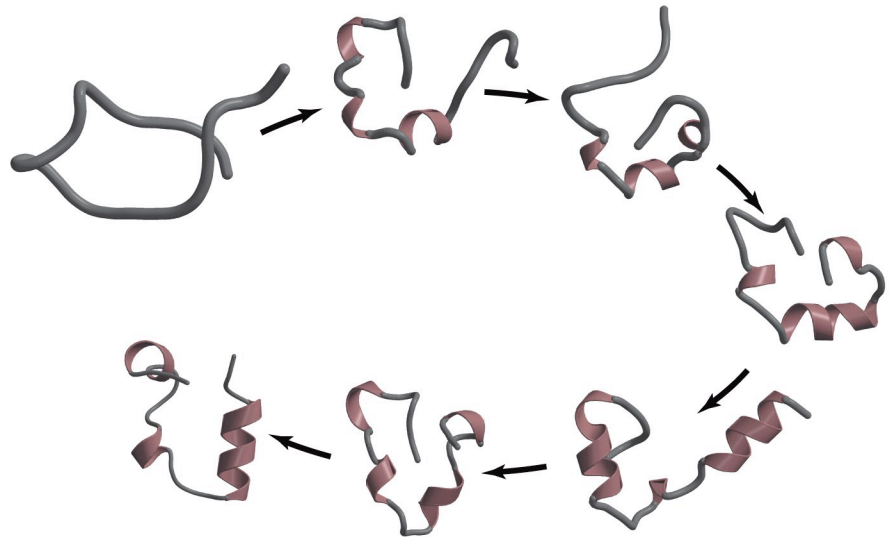
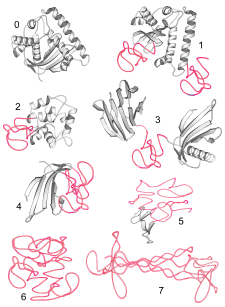


Figure 4-27
Lehninger Principles of Biochemistry, Fifth Edition

What is protein disorder?

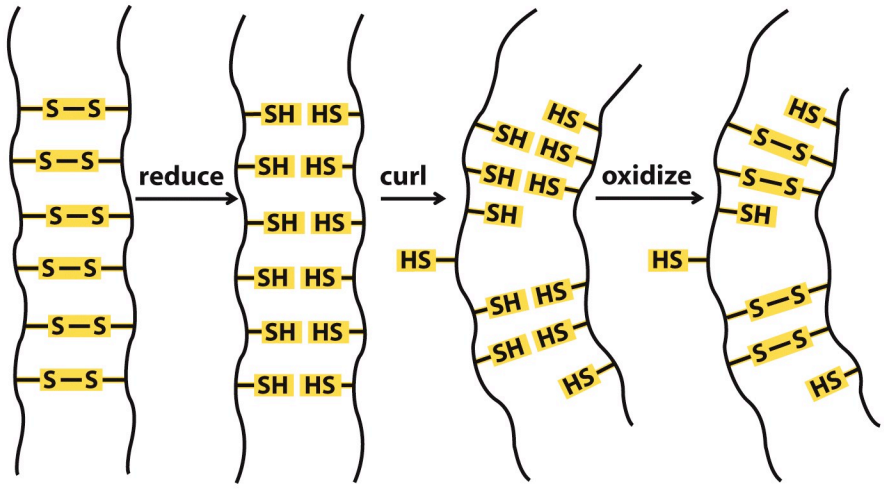
Disordered regions (DRs) are entire proteins or regions of proteins which lack a fixed tertiary structure, essentially being partially or fully unfolded. Such disordered regions have been shown to be involved in a variety of functions, including DNA recognition, modulation of specificity/affinity of protein binding, molecular threading, activation by cleavage, and control of protein lifetimes. Although these DRs lack a defined 3-D structure in their native states, they frequently undergo disorder-to-order transitions upon binding to their partners.



Depiction of degrees of protein disorder:
0 – totally ordered
1-5 – partial disorder
6,7 – total disorder

- **Heat:** The weak side-chain attractions in globular proteins are easily disrupted by heating, in many cases only to temperatures above 50°C.
- **Mechanical agitation:** The most familiar example of denaturation by agitation is the foam produced by beating egg whites. Denaturation of proteins at the surface of the air bubbles stiffens the protein and causes the bubbles to be held in place.
- **Detergents:** Even very low concentrations of detergents can cause denaturation by disrupting the association of hydrophobic side chains.

- **Organic compounds:** Organic solvents can interfere with hydrogen bonding or hydrophobic interactions. The disinfectant action of ethanol results from its ability to denature bacterial protein.
- **pH change:** Excess H^+ or OH^- ions react with the basic or acidic side chains in amino acid residues and disrupt salt bridges. An example of denaturation by pH change is the protein coagulation that occurs when milk turns sour because it has become acidic.
- **Inorganic salts:** Sufficiently high concentrations of ions can disturb salt bridges.
- Most denaturation is irreversible. Hard-boiled eggs do not soften when their temperature is lowered.



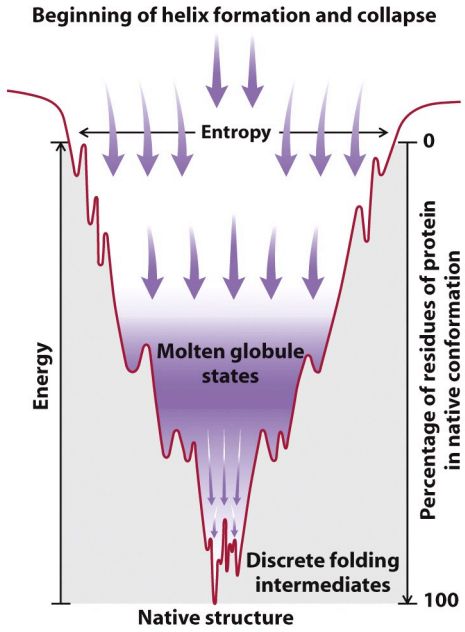


Figure 4-28
Lehninger Principles of Biochemistry, Fifth Edition

Conformation A $\rightleftharpoons^{\Delta G}$ Conformation B

kcal/mol	kJ/mol	$\ln K_{eq}$	K_{eq}
1	4.184	1.69	5.41
2	8.368	3.38	29.23
5	20.92	8.44	4619.71
10	41.84	16.88	21341729.78
15	62.76	25.31	98592623925.87
20	83.68	33.75	455469429828309.00
25	104.6	42.19	2104137137724550000.00
30	125.52	50.63	9720505492587280000000.00
35	146.44	59.07	44905926204791000000000000.00
40	167.36	67.50	207452401508125000000000000000.00
45	188.28	75.94	958370142399971000000000000000000.00
50	209.2	84.38	442739309435167000000000000000000000.00
55	230.12	92.82	2045327660440840000000000000000000000000.00
60	251.04	101.26	944882270314195000000000000000000000000000.00