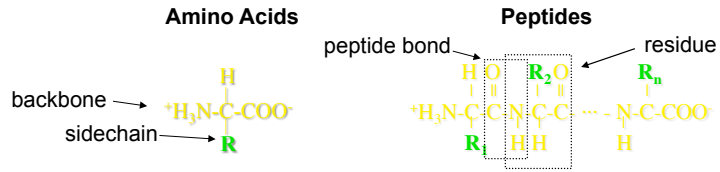
BIOCH 755: Biochemistry I Fall 2015

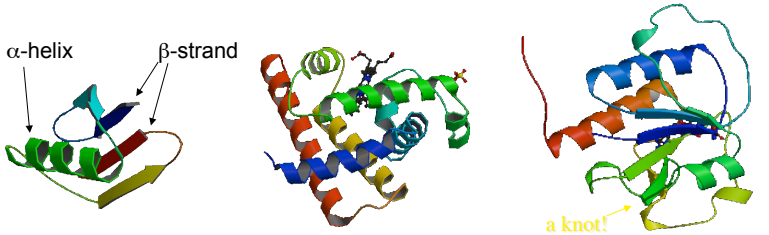
4. Protein Structures (Chapters 5 & 6)

Jianhan Chen
Office Hour: M 1:30-2:30PM, Chalmers 034
Email: jianhanc@ksu.edu
Office: 785-2518

Hierarchical Organization of Proteins



Primary Sequence: MTYKLILNGK TLKGETTTEA VDAATAEKVF KQYANDNGVD GEWTYDDATK TFTVTE

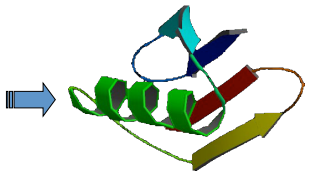


Protein G B1 (3gb1) Myoglobin (1mbc) tRNA transferase (1mxi)

Structure-Function Paradigm

Sequence Folded 3D Structure Functions

MTYKLILNGK
TLKGETTTEA
VDAATAEKVF
KQYANDNGV
DGEWTYDDA
TKTFTVTE …



catalysis
transport
signaling
motors …

- Structural Genomics (NMR, X-ray)
- Protein structure predictions

5.1 Polypeptide diversity

- In theory, the size and composition of a polypeptide chain are unlimited (possible number of sequence: 20^N).

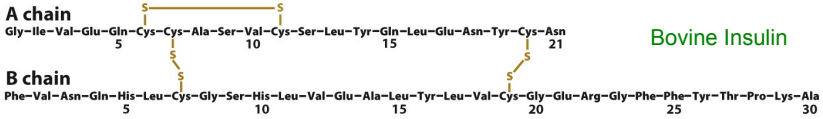


Figure 5-1

- In cells, this potential variety is limited by the efficiency of protein synthesis and by the ability of the polypeptide to fold into a functional structure.
  - Only a very small subset are explored in biology and most of them give rise to specific 3D structures under the proper conditions
  - Some sequences are "intrinsically disordered"! They play particularly important roles in signaling and regulation.

Most Proteins Have 100-1000 Amino Acids

TABLE 5-1 Compositions of Some Proteins

Protein	Amino Acid Residues	Subunits	Protein Molecular Mass (D)
Proteinase inhibitor III (bitter gourd)	30	1	3,427
Cytochrome *c* (human)	104	1	11,617
Myoglobin (horse)	153	1	16,951
Interferon-γ (rabbit)	288	2	33,842
Chorismate mutase (*Bacillus subtilis*)	381	3	43,551
Triose phosphate isomerase (*E. coli*)	510	2	53,944
Hemoglobin (human)	574	4	61,986
RNA polymerase (bacteriophage T7)	883	1	98,885
Nucleoside diphosphate kinase (*Dictyostelium discoideum*)	930	6	100,764
Pyruvate decarboxylase (yeast)	2,252	4	245,456
Glutamine synthetase (*E. coli*)	5,616	12	621,264
Titin (human)	34,350	1	3,816,188

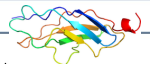


Table 5-1

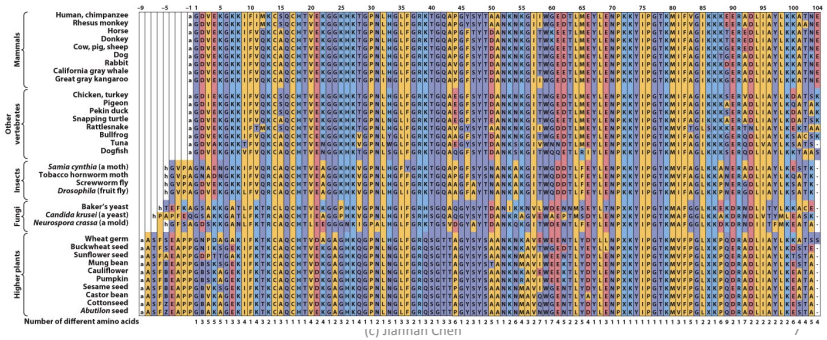
Most Proteins Have 100-1000 Amino Acids

- Peptides shorter than 40 residues can not fold properly;
- Longer than 1000 residues are error prone in transcription and translation and also experience expression efficiency bottleneck.

5.4 Protein Evolution

- **Key Concepts 5.4**
  - Sequence comparisons reveal the evolutionary relationships between proteins.
  - Protein families evolve by the duplication and divergence of genes encoding protein domains.
  - The rate of evolution varies from protein to protein.

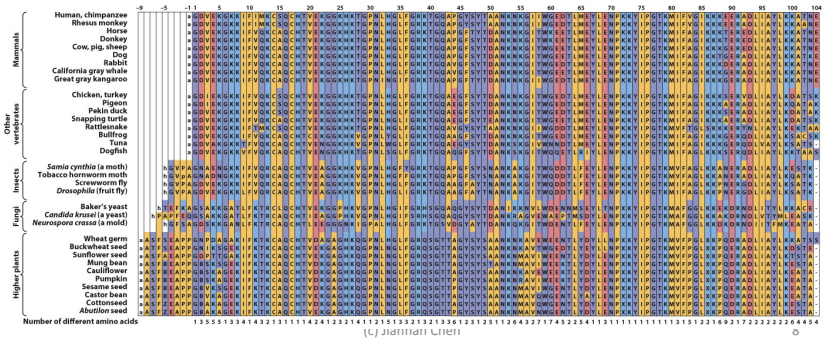
TABLE 5-6 Amino Acid Sequences of Cytochromes *c* from 38 Species[a]



5.4 Protein Evolution

- Most (random) mutations harmful or deleterious
- Mutations propagated only if it increases or does not decrease fitness; this happens only rarely
- Evolution by natural selection

TABLE 5-6 Amino Acid Sequences of Cytochromes *c* from 38 Species[a]



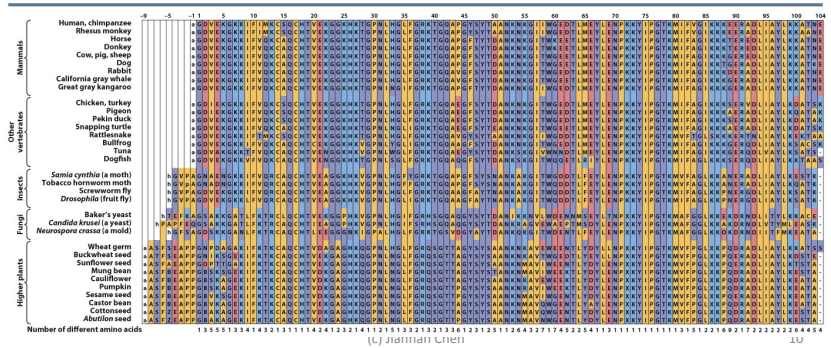
5.4 Protein Evolution

- Most (random) mutations harmful or deleterious
- Mutations propagated only if it increases or does not decrease fitness; this happens only rarely
- Evolution by natural selection
- Evolutionary relationships can be derived from sequences: related species have evolved from a common ancestor, so it follows that the genes specifying each of their proteins must likewise have evolved from the corresponding gene in that ancestor.
- Sequence conservation provide insights on protein function (and structure): invariant (conserved) residues likely essential for function; variant (or hypervariable) residues not specific/required for function
- Neutral drift: random mutations that do not affect function will naturally accumulate over time
- Predict function of novel proteins: sequence homologs
- Structure prediction: homology modeling

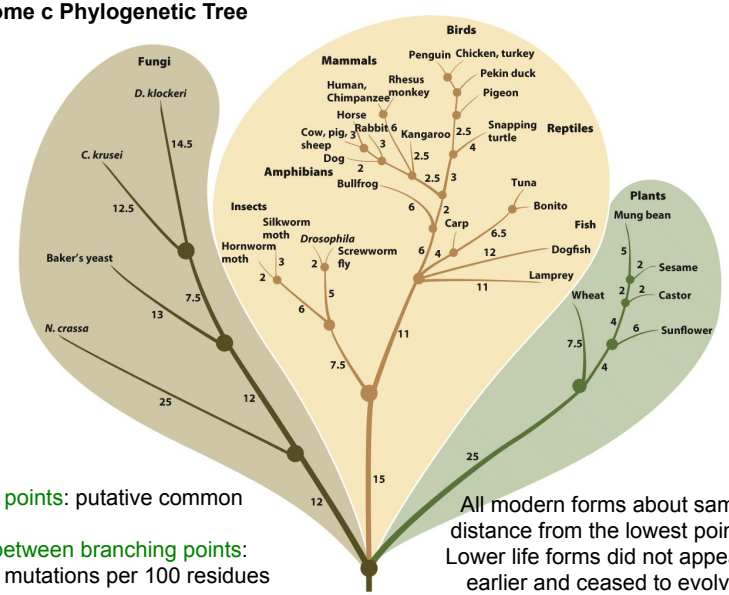
Phylogenetic Trees

- Depict evolutionary history
- Levels of differences parallel classical taxonomy
- Many online tools:
  - http://www.ebi.ac.uk/Tools/phylogeny/clustalw2_phylogeny/
  - http://www.phylogeny.fr (also contains a list of other phylogeny programs)

TABLE 5-6 Amino Acid Sequences of Cytochromes *c* from 38 Species[a]



Cytochrome c Phylogenetic Tree



Branching points: putative common ancestors

Distance between branching points: number of mutations per 100 residues

All modern forms about same distance from the lowest point: Lower life forms did not appear earlier and ceased to evolve.

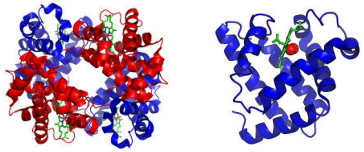
Figure 5-22
© 2013 John Wiley & Sons, Inc. All rights reserved. After Dayhoff, M.O., Park, C.M., and McLaughlin, P.J., in Dayhoff, M.O. (Ed.), *Atlas of Protein Sequence and Structure*, p. 8, National Biomedical Research Foundation (1972).

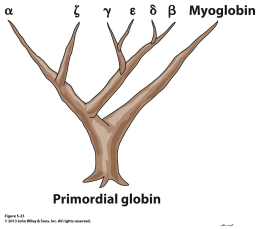
Protein Domains

- Domains: evolutionary conserved sequences are often segments of about 40-200 residues
  - Related to structural domains to be discussed in Chapter 6
- Protein domains can be grouped into 1000-1400 families, with half of them in ~200 most populated families
- Level of sequence conservation within family is correlated with functional conservation
  - >40% sequence conservation for most functionally conserved proteins
  - <20% sequence conservation: likely different function
- Protein families: likely arise from gene duplication

Evolution by Gene Duplication

- Orthologs: homologous proteins with the same function
- Gene duplication: aberrant genetic recombination event in which one member of a chromosome pair receives both copies of the primordial gene
  - The sequences may diverge through evolution: one copy of the gene may evolve a new function (paralogs)
  - In human, ~98% of protein domains have been duplicated!
  - Pseudogenes: dead ends of protein evolution
- Example: globin family
  - Hemoglobin: transports O2, tetramer (α2β2)
  - Myoglobin: O2 diffusion in muscle, single-chain





Rate of Sequence Divergence Varies

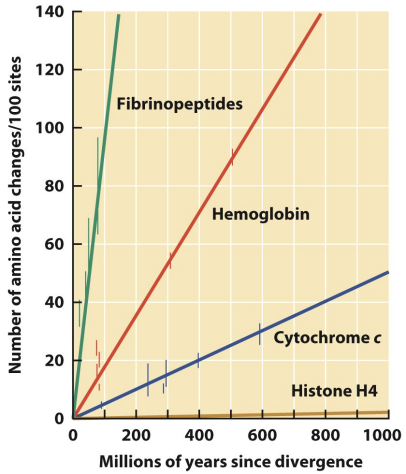


Figure 5-24
Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.

- Evolution rates largely constant over time
- Rates (of mutations accepted) vary from protein to protein
- Functional (and structural) constrains
- Histone H4: most conserved, extremely intolerant of mutations
- Fibrinopeptides: cleaved from fibrinogen to induce blood clotting, little selective pressure for conservation
- Intrinsically disordered proteins: highly mutation rates

Sub-summary: Primary Sequence

- Protein sequence: highly selective (structural, functional restrains; efficiency of synthesis)
- Mutation and evolution:
  - Natural selection
  - Gene duplication
  - many useful insights that can be derived from examine sequence conservation and variations

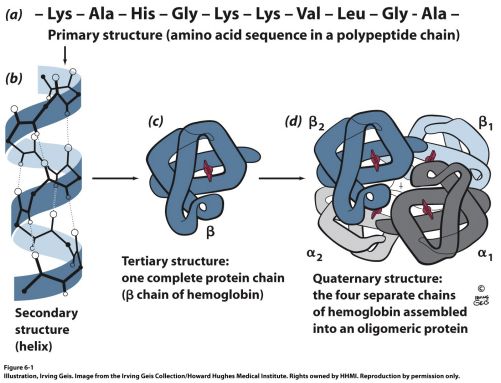
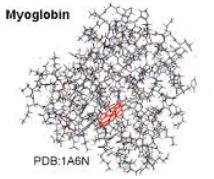


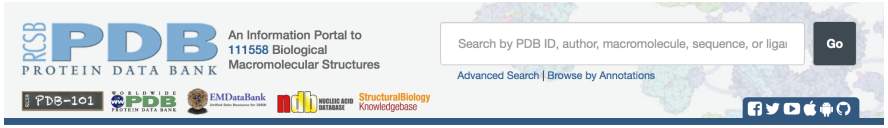
Figure 6-1
Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.

PROTEIN STRUCTURE (CHAPTER 6)

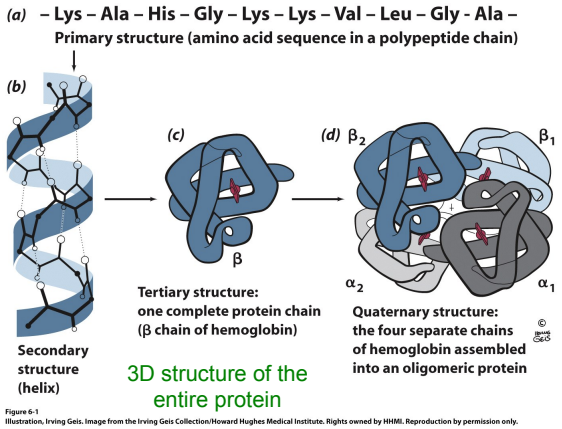
A little history

- The fact that proteins are usually structured was not recognized until 1934, when Bernal and Hodgkin observed discrete diffraction pattern of a crystal of pepsin.
- First structure of protein: myoglobin in 1958 by Kendrew and coworkers; it was depressingly complex and nothing like the beautiful DNA double helix!
  - The structural complexity of proteins is actually consistent with the extremely diversity in function
- Structural proteome efforts
  - As of Aug 2015, 111,558 structures publically available
  - Regularity of protein structure (on both 2nd and tertiary levels!)





Levels of Protein Structure

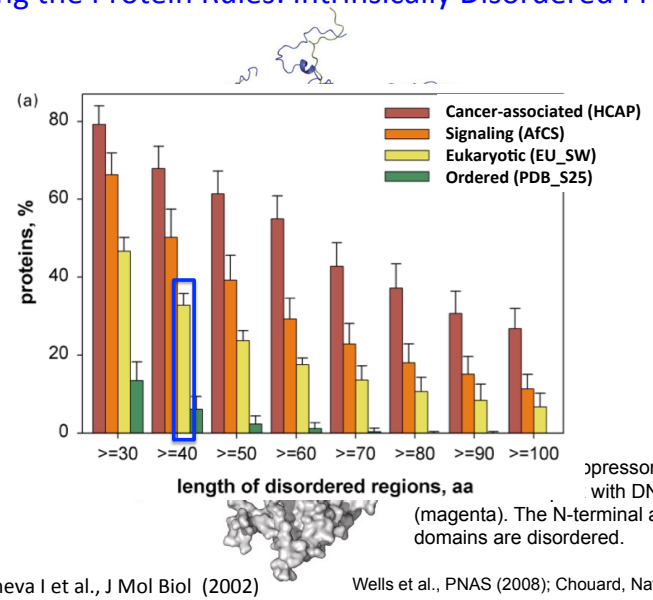


Local conformational arrangements

3D structure of the entire protein

Spatial arrangement of multiple subunits

Breaking the Protein Rules: Intrinsically Disordered Proteins



opressor p53
with DNA
(magenta). The N-terminal activation
domains are disordered.

Iakoucheva I et al., J Mol Biol (2002)

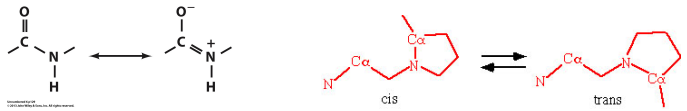
Wells et al., PNAS (2008); Chouard, Nature (2011).

6.1 Secondary Structure

- **Key Concepts 6.1**
  - The planar character of the peptide group limits the conformational flexibility of the polypeptide chain.
  - The α helix and the β sheet allow the polypeptide chain to adopt favorable ϕ and ψ angles and to form hydrogen bonds.
  - Fibrous proteins contain long stretches of regular secondary structure, such as the coiled coils in α keratin and the triple helix in collagen.
  - Not all polypeptide segments form regular secondary structure such as α helices or β sheets.

A. Peptide Backbone Has Limited Flexibility

- The peptide plane is rigid due to resonance



  - Almost always in trans- (that is, CA on opposite sides of peptide bond)
  - Only proline has finite (~10%) probability of adopting cis-conformation
- Two major rotatable torsion angles (before and after each peptide plane)

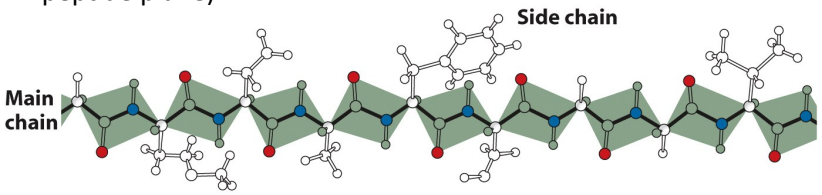
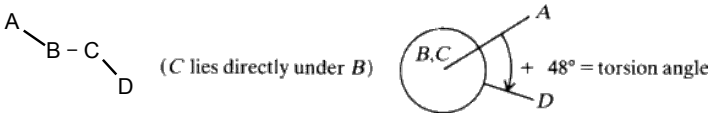


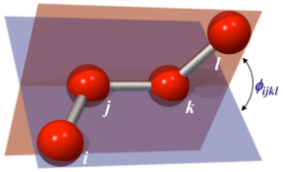
Figure 6-3
Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.

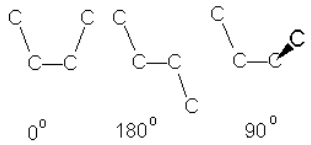
Torsion (Dihedral) Angle

- Peptide conformations are mainly defined by key backbone torsion angles
  - Also referred to as dihedral angle: the angle between two groups on either side of a freely rotating chemical bond
  - Defined by four consecutively linked atoms: ϕ(A-B-C-D)



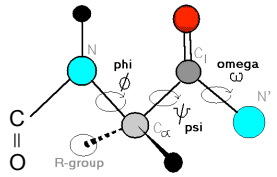
  - Right-hand rule: positive sense is clockwise





Peptide Backbone Torsions

- ϕ: C(O)-N-Cα-C(O)
- ψ: N-Cα-C(O)-N



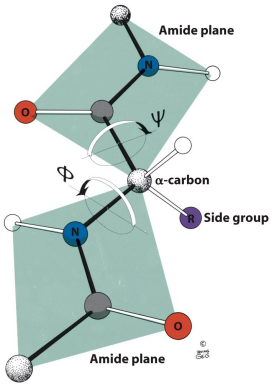


Figure 6-4
Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.

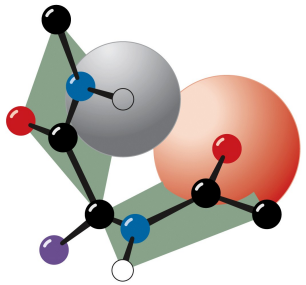


Figure 6-5
Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.

Ramachandran Plot

- 2D plot of (ϕ,ψ) distribution
- Not all possible (ϕ,ψ) accessible to peptides
- Steric clash of C=O with side chain prohibits positive ϕ (right half)
- Allow assignment of secondary structures
- Statistics of known protein structures reveal common regions of (ϕ,ψ) distributions
- An important way of structural evaluation

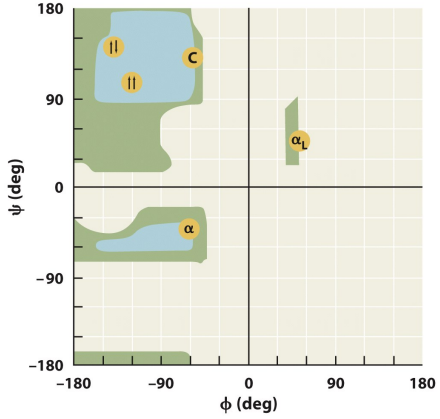
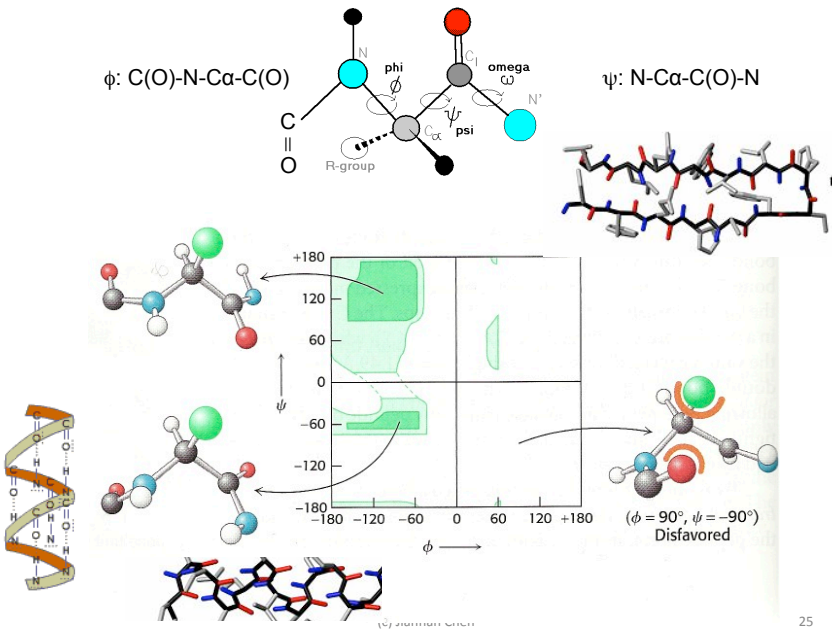
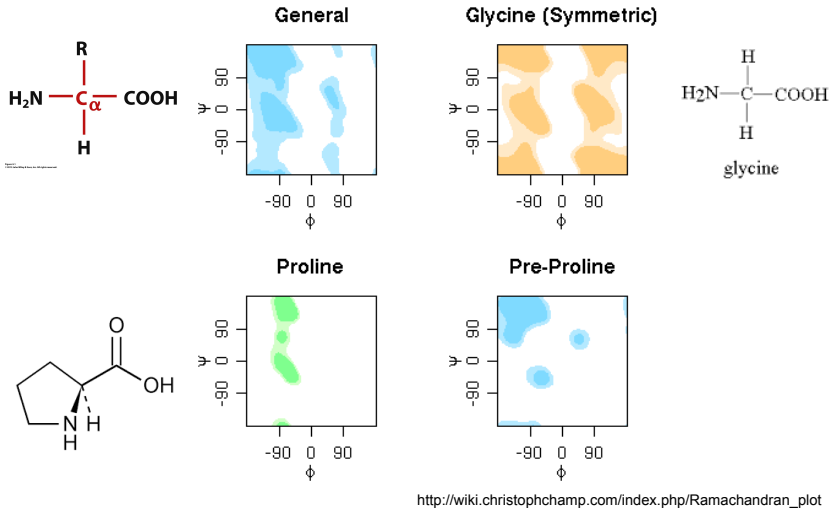


Figure 6-6
© 2013 John Wiley & Sons, Inc. All rights reserved.



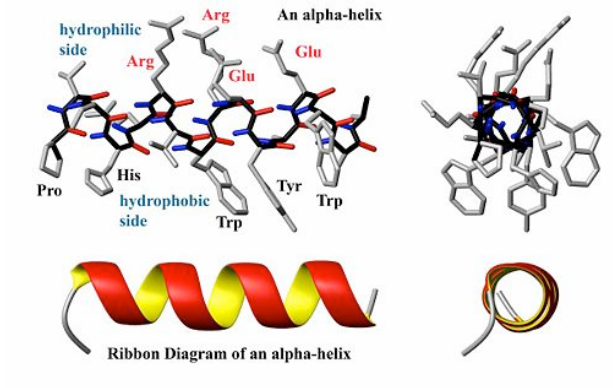
Glycine and Proline



http://wiki.christophchamp.com/index.php/Ramachandran_plot

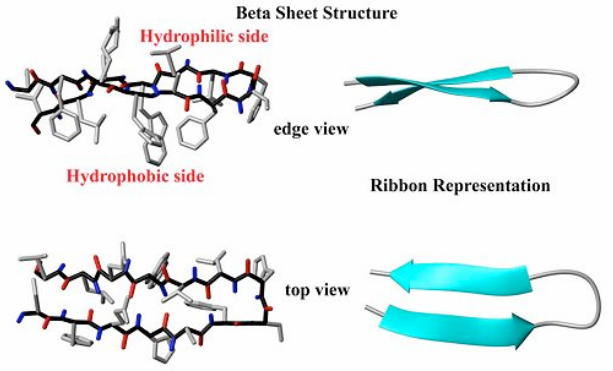
B. Secondary Structures of Proteins

- α-helix:
  - (i,i+4) hydrogen bonds, 3.6 residues/5.4 Å per turn, (ϕ,ψ) ~ (-55°, -50°)
  - Average helix length in proteins: ~12 residues (or ~18 Å)



Secondary Structures of Proteins

- β-strands:
  - (near) trans- backbone torsions: (ϕ, ψ) ~ (–135°, 135°), Right-hand twist
  - CA-CA distance along the strand ~ 7 Å
  - Hydrogen bonds between neighboring strands



Beta Sheets

Note the distinct hydrogen bonding patterns!

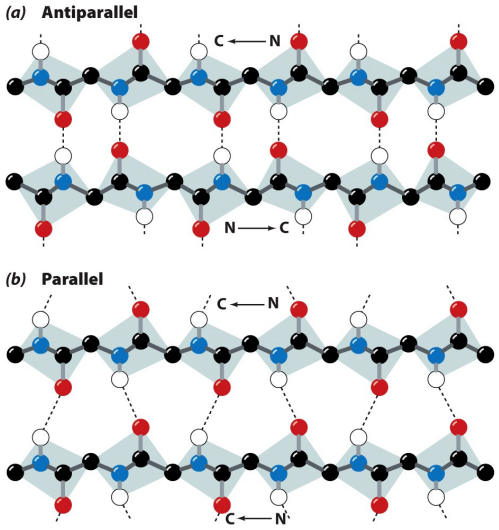
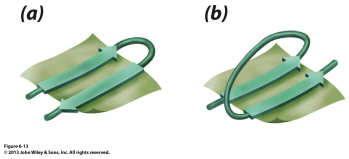


Figure 6-9
Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.

Turns Connecting Secondary Structures

- Links between secondary structures vary greatly!
- Very tight turns possible between strands
- Types I and II beta-bends: 4-residue tight turn
  - Differ in the central peptide plane
  - 2nd residue is often Pro
  - 3rd residue in type II is often Gly



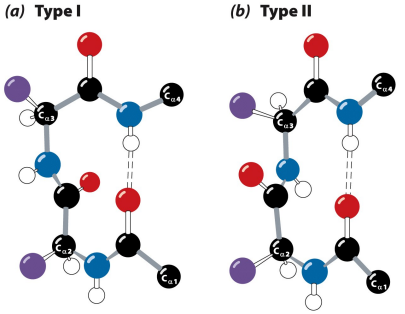


Figure 6-14
Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.

Fibrous Proteins: Repeating 2nd Structures

- Fibrous vs globular
- Keratin & collagen: highly elongated
- Keratin: coiled coil
  - Mechanically durable
  - Horny outer epidermal layer (hair, horn, nail and feather)
  - Alpha-helix with 5.1 Å spacing
  - 7-residue pseudo-repeats (3.5 residues per turn)
  - Higher order structures not fully understood
  - Rich with cys: chemical inter coil cross-linking

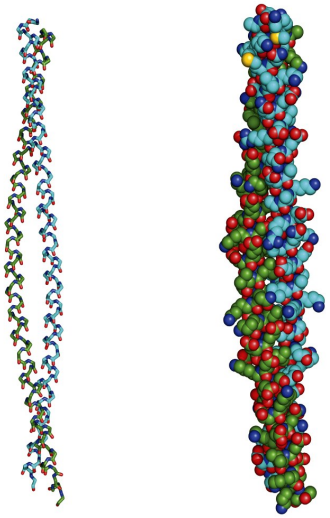


Figure 6-15b
© 2013 John Wiley & Sons, Inc. All rights reserved.

Higher Order Structure of α Keratin

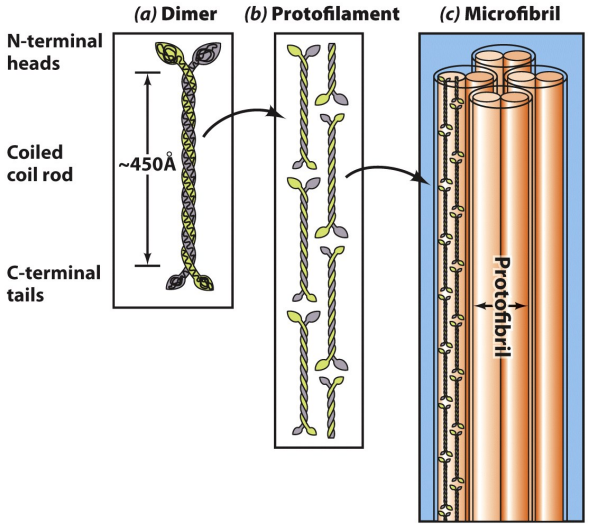
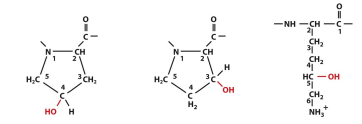


Figure 6-16
© 2013 John Wiley & Sons, Inc. All rights reserved.

Collagen: Triple Helix

- The most abundant vertebrate protein
- The major stress-bearing components of connective tissues
- Each collagen consists of three peptide chains
- Many collagen chains/varieties (fort different tissues)
- Type I: the most common one
  - Two α1 + one α2; ~285 KD;
  - ~14 Å in width and ~3000 Å in length!
- Very distinctive amino acid compositions
  - 1/3 are Gly; 15-30% are Pro and its derivatives



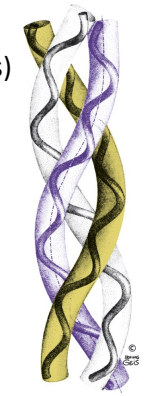
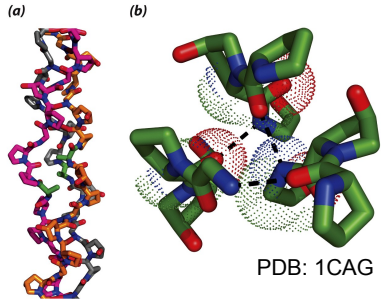


Figure 6-17
Illustration, Irving Geis. Image from the Irvi
Medical Institute. Rights owned by HHMI. R

Collagen: Triple Helix

- Typical sequence: -(Gly-X-Y)- repeats (X ~ Pro & Y ~ Hyp)
- Left-handed helix with right-handed twist
  - Three residues per turn
- Covalent linkages: Lys and His
  - Increase with age (tissues becoming stiffer)



- Gly required for packing
- H-bonds between Gly NH and Pro CO
- Pro/Hyp rings rigid and well packed

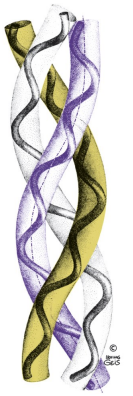
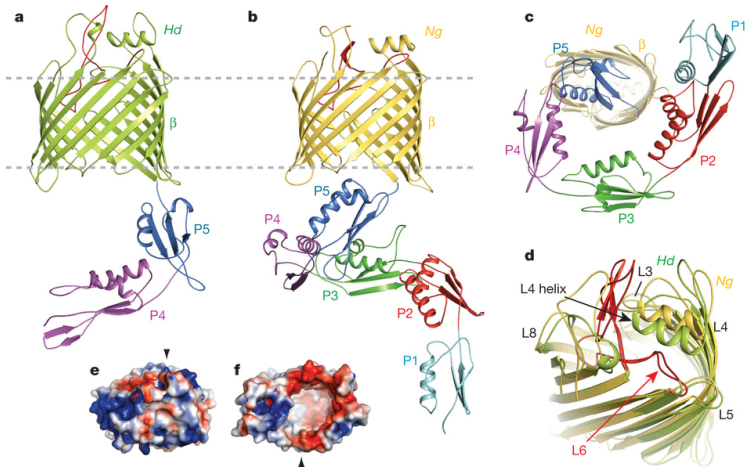


Figure 6-17
Illustration, Irving Geis. Image from the Irvi
Medical Institute. Rights owned by HHMI. R

Other, Nonrepetitive Local Structures



The structure of BamA from the BAM complex (Nature 2013) .

2nd Structure Propensities

- Different amino acids have different inherent 2nd structure propensities
- Tertiary structure context also affect 2nd structure
- Ala: the most helical residue
- Gly and Pro: strong helix breakers
- Certain residues tend to cap helices and strands
  - Asn and Gln: side chains could fold back and h-bond with backbone ("helix

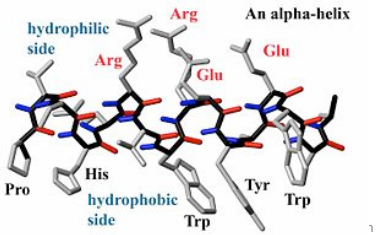


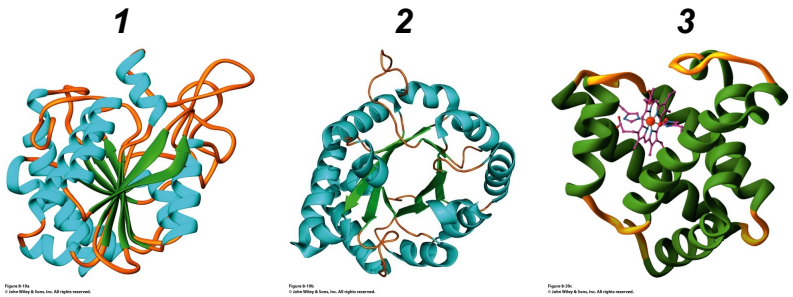
TABLE 6-1 Propensities of Amino Acid Residues for α Helical and β Sheet Conformations

Residue	P_α	P_β
Ala	1.42	0.83
Arg	0.98	0.93
Asn	0.67	0.89
Asp	1.01	0.54
Cys	0.70	1.19
Gln	1.11	1.10
Glu	1.51	0.37
Gly	0.57	0.75
His	1.00	0.87
Ile	1.08	1.60
Leu	1.21	1.30
Lys	1.16	0.74
Met	1.45	1.05
Phe	1.13	1.38
Pro	0.57	0.55
Ser	0.77	0.75
Thr	0.83	1.19
Trp	1.08	1.37
Tyr	0.69	1.47
Val	1.06	1.70

Source: Chou, P.Y. and Fasman, G.D., *Annu. Rev. Biochem.* **47**, 258 (1978).

Table 6-1

Which of the following statements is FALSE with respect to the ribbon diagrams shown below ?



A. Protein 2 contains parallel β-sheet strands.
B. Protein 3 contains a prosthetic group.
C. The secondary structure of protein 3 is all α-helix.
D. These proteins have both secondary and tertiary structure.

6.2 Tertiary Structure

- **Key Concepts 6.2**
  - X-Ray crystallography and NMR spectroscopy are used to determine the positions of atoms in proteins.
  - Nonpolar residues tend to occur in the protein interior and polar residues on the exterior.
  - A protein's tertiary structure consists of secondary structural elements that combine to form motifs and domains.
  - Over time, a protein's structure is more highly conserved than its sequence.
  - Bioinformatics databases store macromolecular structure coordinates. Software makes it possible to visualize proteins and compare their structural features.

High Resolution Structure Determination

- Two main techniques: X-ray crystallography and NMR
- X-ray: crystallization is critical; accounts for ~85%
- NMR: very tedious, but very powerful in functional studies
- Cryo-EM: recent advances allow Angstrom resolution; extremely powerful for large, live complexes!

NMR	**X-ray crystallography**
short time scale, protein folding	long time scale, static structure
solution, purity	single crystal, purity
< 20kD, domain	any size, domain, complex
functional active site	active or inactive
atomic nuclei, chemical bonds	electron density
resolution limit 2-3.5Å	resolution limit 1.5-3.5Å

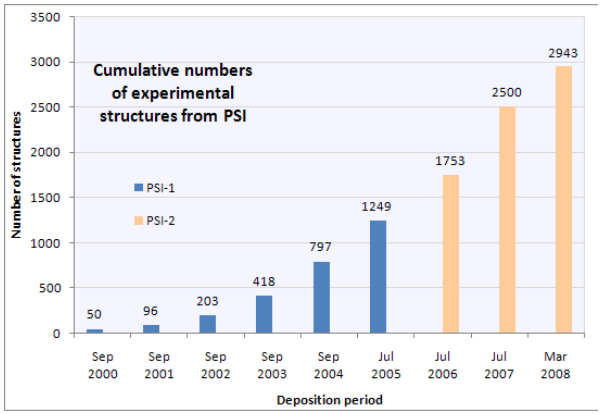
Dynamics
Transient interactions

Protein Structure Initiatives

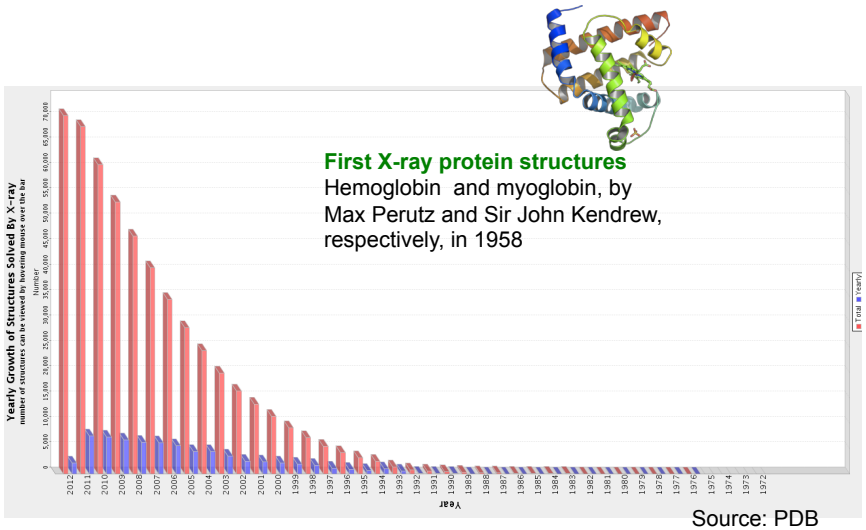
PSI Protein Structure Initiative

- Aimed at covering as much fold space as possible
- Three phases, $764M invested (from NIH)

(more structures were actually solved than those finally deposited)



Over 85% of Known Protein Structures from X-ray

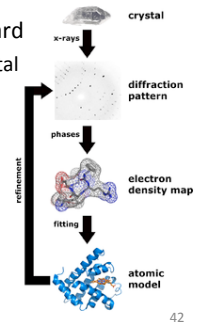


First X-ray protein structures
Hemoglobin and myoglobin, by Max Perutz and Sir John Kendrew, respectively, in 1958

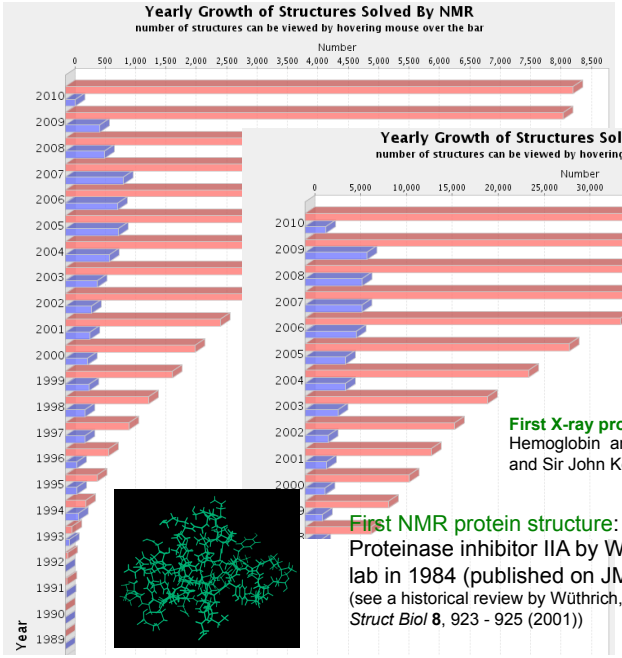
Source: PDB

Protein Structure Determination by X-ray Crystallography

- Protein purification: first step of all structural determination efforts
- Protein Crystallization: the most challenging part!
  - Produce well-ordered protein mono-crystals without any inclusion and large enough to diffract X-Ray beam
  - More art than science: largely impossible to predict crystallization conditions
- Crystal mounting: transfer of crystals from solution to detector
  - Proper alignment can improve completeness
- Diffraction data collection: central but somewhat standard
  - Record monochromatic reflections while rotating the crystal
  - Multi-wavelength data collection on synchrotrons
- Data processing: indexing, integration and scaling
- Structure construction: phase problem
  - Molecular replacement; Heavy atom methods
- Structure refinement and analysis:
  - Resolution, R-factor, what about biology!?



http://proteincrystallography.org/ & Wikipedia



Source: PDB

First X-ray protein structures
Hemoglobin and myoglobin, by Max Perutz and Sir John Kendrew, respectively, in 1958

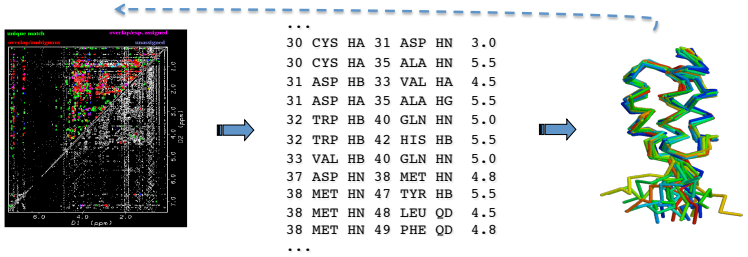
First NMR protein structure:
Proteinase inhibitor IIA by Wüthrich lab in 1984 (published on JMB 1985).
(see a historical review by Wüthrich, *Nature Struct Biol* **8**, 923 - 925 (2001))

High-resolution NMR structure determination

- NMR is one of the only two methods for high-resolution structure determination besides X-ray crystallography
- Advantages of NMR
  - In solution: more physiologically relevant conditions
  - Provide information on protein dynamics
  - Room temperature or (almost) any temperatures of interest
  - Direct monitoring of biophysical and biochemical processes
- Limitations of NMR
  - Time consuming and labor intensive: difficult for high-throughput, prone to human errors, less precise
  - Need to be highly soluble: stable with near mM concentration
    - Sometimes "strange" NMR buffers have to be used
  - Limited to protein of moderate sizes: < 200 residues in general
- X-ray and NMR can be complementary

Basic Steps of NMR Structure Determination

- Sample preparation and data collection
- Chemical shift assignments: backbone and sidechain
  - Chemical Shift Indexing and J-coupling constants: 2nd structures
- Distance and other structural restraints: NOESY
- Structural calculations: restrained molecular dynamics
- Refinement and validation

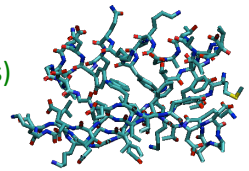


Also see: http://en.wikipedia.org/wiki/Protein_nuclear_magnetic_resonance_spectroscopy

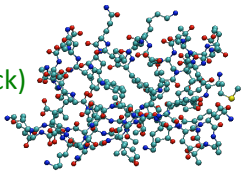
Protein Structure Visualization

- Many free software: Pymol, vmd etc
  - http://www.bioinformatics-made-simple.com/2012/09/10-free-software-for-protein-3d.html
  - They all come with excellent tutorials!

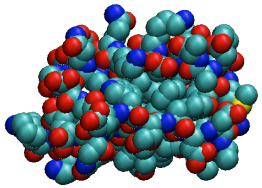
Sticks (bonds)



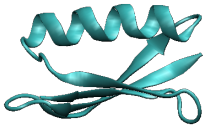
CPK (ball-n-stick)



VDW

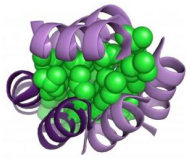


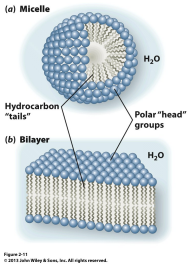
Cartoon



Sidechain Distributions in Protein Structure

- Protein folding is a self-assembly process that is largely driven by hydrophobic effects
- Nonpolar residues form "hydrophobic core"
- Charged residues decorate the protein surface: solubility and aggregation resistance
  - Buried charges are typically complemented
- Polar residues can be on surface or buried
  - Buried polar sidechains typically in h-bonds
- "Buried" backbone typically involved in h-bonds (helices, sheets etc)
- Proteins are very well packed!
- "Structural" waters





α-helix and β-sheet secondary structure is located mainly in the interior of folded proteins, whereas irregular loops occur on the outside. Why?

A. The loops are needed for the polypeptide to "turn corners" and this occurs on the outside of the protein.

B. α-helix and β-sheet structures are more compact than loops, and so they fit better in the interior of the folded protein.

C. The side chains in α-helix and β-sheet structures are mostly non-polar.

D. The H bonding requirements of the polypeptide backbone in irregular loops are not fully satisfied, and so these structures interact with water at the surface.

Tertiary Structure

- Mostly formed by combination of 2nd structures
- Arrange of secondary structures provide a convenient way of classifying "folds"
- Super-secondary structure: common (local) grouping

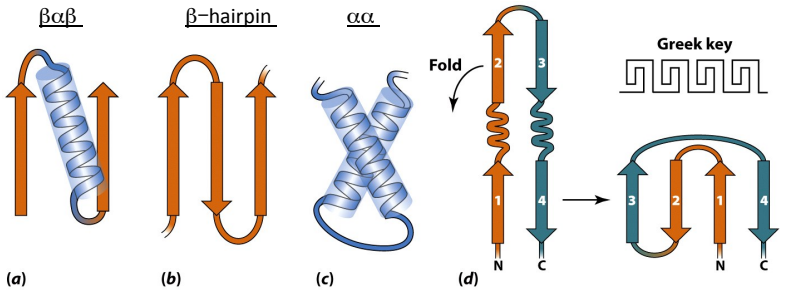
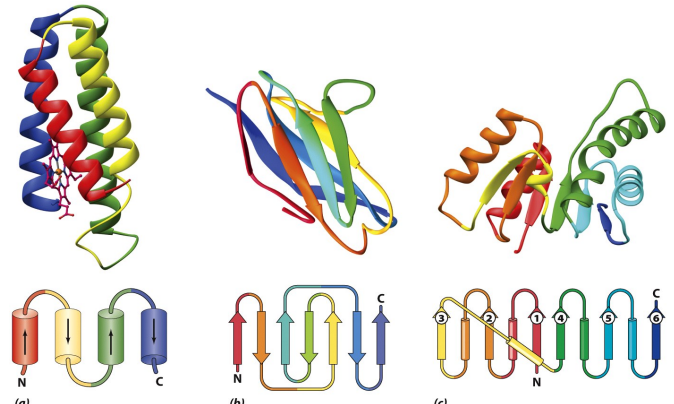


Figure 6-28

Most Proteins are α, β, or α/β

- These classes can be further divided into more specific folds



Cytochrome *b*562
PDBid 256B

Human immunoglobulin fragment
PDBid 7FAB

Dogfish lactate dehydrogenase
PDBid 6LDH

8-Stranded β Barrels

Human retinol binding protein
PDBid 1RBP

Peptide-N⁴-(N-acetyl-β-D-glucosaminyl) asparagine amidase
PDBid 1PNG

Triose phosphate isomerase
PDBid 1TIM

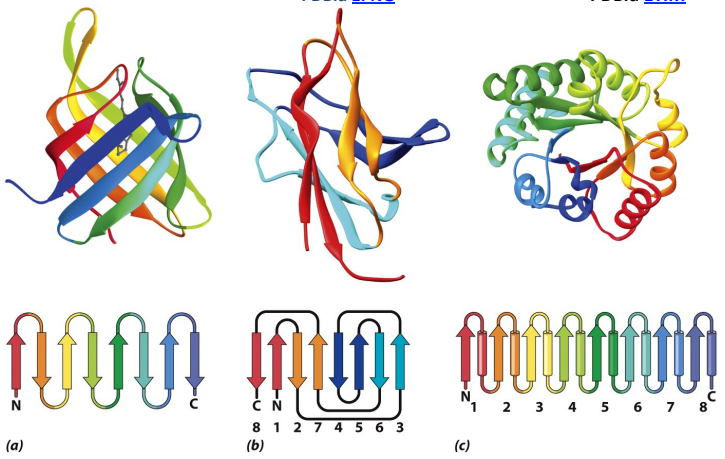
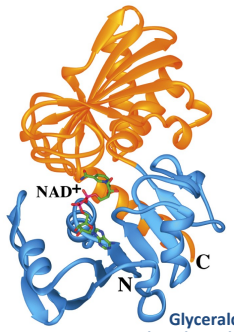


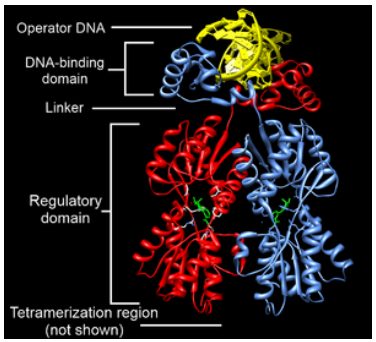
Figure 6-30

Multi-Domain Proteins

- Larger proteins (e.g., >200 residues) often fold into multiple domains
- Domains often linked by one or two segments: flexibility, regulation
- Domains are often structurally independent
- Domains are also often responsible for different aspects of the function, e.g., regulatory and catalytic domains



Glyceraldehyde-3-phosphate dehydrogenase
PDBid 1GD1



Lactose Repressor

Structure Conserved More Than Sequence

- Only ~1400 protein domain families
- Domains/folds persist if
  - Form stable structures (against degradation/aggregation)
  - Tolerate (random) mutations
  - Support essential biological function
- Structural conservation: Cytochrome c as an example

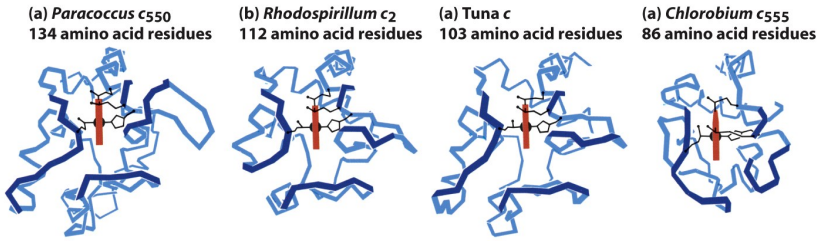


Figure 6-32
Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.

(Structural) Bioinformatics

- Powerful tools for mining sequence/structural data to infer the structure and function of new proteins
- The Protein Data Bank (PDB): structure depository
- Visualization: VMD, PyMol etc
- Sequence search and alignment
- "Big Data" era!

www.taverna.org.uk

Internet Bioinformatics Tools

TABLE 6-2 Structural Bioinformatics Internet Addresses

Structural Databases
Protein Data Bank (PDB): http://www.rcsb.org/
Nucleic Acid Database: http://ndbserver.rutgers.edu/
Molecular Modeling Database (MMDB): http://www.ncbi.nlm.nih.gov/sites/entrez?db=structure
Most Representative NMR Structure in an Ensemble: http://www.ebi.ac.uk/msd-srv/olderado
Molecular Graphics Programs
Cn3D: http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml
FirstGlance: http://molvis.sdsc.edu/fgij/index.htm
Jmol: http://jmol.sourceforge.net/
KiNG: http://kinemage.biochem.duke.edu/software/king.php
Proteopedia: http://www.proteopedia.org/
Swiss-Pdb Viewer (DeepView): http://spdbv.vital-it.ch/
Structural Classification Algorithms
CATH (Class, *A*rchitecture, *T*opology, and *H*omologous superfamily): http://www.cathdb.info/index.html
CE (Combinatorial *E*xtension of optimal pathway): http://cl.sdsc.edu/
Pfam (protein families): http://pfam.sanger.ac.uk/ or http://pfam.janelia.org/
SCOP (Structural *C* lassification *O*f *P*roteins): http://scop.mrc-lmb.cam.ac.uk/scop/
VAST (*V*ector *A*lignment *S*earch *T*ool): http://www.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml

Table 6-2

Quaternary Structure and Symmetry

- Many proteins consist of multiple subunits
- Subunits are usually arranged symmetrically

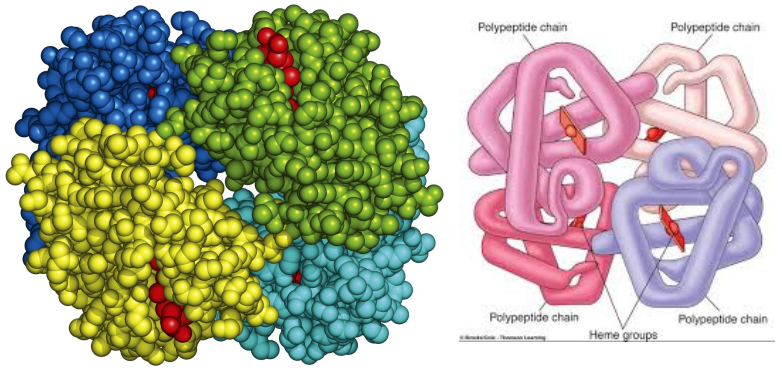


Figure 6-33

Quaternary Structure and Symmetry

- Many proteins consist of multiple subunits
- Subunits are usually arranged symmetrically
- Why multi-unit assembly?
  - More efficient for synthesis
  - More robust to maintain (against damages)
  - Requires smaller genes
  - The assembly-disassembly equilibrium can be regulated
- Same principles determining folding governs assembly
- Typically rotational symmetry (Cn)
- Dihedral symmetry for more complex assemblies (Dn)

Symmetries of Oligomeric Proteins

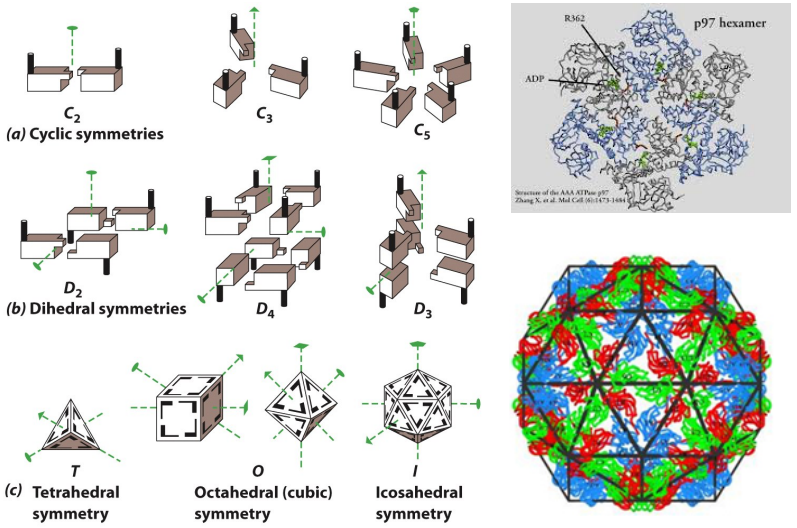
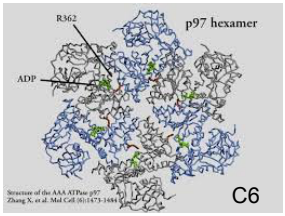
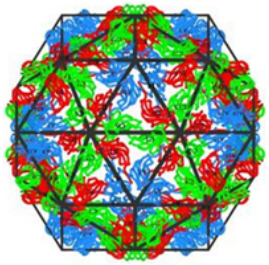


Figure 6-34
Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.

Beauty of Symmetry

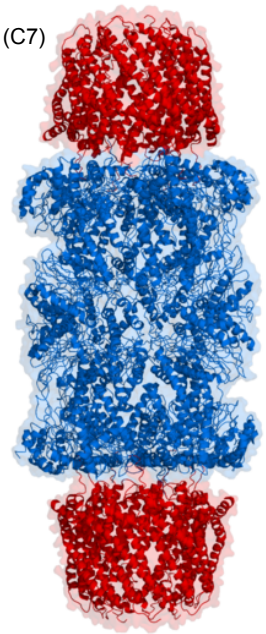




Viral capsid (icosahedral symmetry)

Actin filament

Proteasome (C7)



Sub-summary: 3D Sequence

- Protein backbone: restrict flexibility, key torsions
- Common 2nd structures: basic features
- Two key techniques for solving protein structure
- Tertiary and quaternary structures
  - Basic features
  - Classification
  - Symmetry