BME 42-620 Engineering Molecular Cell Biology

Lecture 04:

Basics of MATLAB

Structure and Dynamics of Cellular Molecules

Outline

- Basics of MATLAB
- Chemical composition of a cell
- Chemical bonds of cellular molecules
- A brief introduction to protein structures
- A brief introduction to protein folding

Outline

- Basics of MATLAB
- Chemical composition of a cell
- Chemical bonds of cellular molecules
- A brief introduction to protein structures
- A brief introduction to protein folding

MATLAB Overview

- MATLAB stands for "matrix laboratory“, a product of MathWorks Inc. (Natick, Massachusetts).

- It is both a language and a development and application environment.

- History of MATLAB
  - First developed in 1970's by Cleve Moler, coauthor of LINPACK & EISPACK
  - First written in FORTRAN; Later rewritten in C
  - Commercial development initiated by Jack Little
    - MathWorks was founded in 1984.
  - In 2000, rewritten based on LAPACK.

Sources:
http://www.mathworks.com/company/aboutus/founders/clevemoler.html
http://www.mathworks.com/company/aboutus/founders/jacklittle.html
http://www.mathworks.com/company/newsletters/news_notes/clevescorner/dec04.html

Advantages of MATLAB (I)

- MATLAB provides reliable and efficient numerical computation with friendly user interfaces.
  - Maple, Mathematica strong in symbolic computation
- Examples of numerical computation issues
  - Precision; numerical stability
  - Underflow and overflow
  - Code quality (debugging, exception handling)
  - Code efficiency (optimization)
- Visit www.netlib.org for more information about different numerical packages.

Advantages of MATLAB (II)

- Fast prototyping: MATLAB is an interpreted language
- Extensive toolboxes
- Versatile graphics
- Cross-platform: Windows, Unix, Linux, Mac OS
- Support parallel computing
- Support object oriented programming
- Large groups of users

  MATLAB file exchange **(use with caution)**

  http://www.mathworks.com/matlabcentral/

MATLAB Toolboxes (I)

- A large collection of basic math functions are provided in the MATLAB base package.
- Function extensions are packaged as toolboxes.
- Math and optimization
  - Optimization toolbox
  - PDE toolbox
  - Genetic algorithm and direct search algorithm
- Statistics & data analysis
  - Statistics toolbox
  - Curve fitting toolbox
  - Spline toolbox
  - Neural network toolbox

MATLAB Toolboxes (II)

- Signal & image processing
  - Signal processing toolbox
  - Image processing toolbox
  - Wavelet toolbox

- Third party toolboxes
  - Pattern recognition toolbox: www.prtools.org
  - Wavelet toolbox: http://www-stat.stanford.edu/~wavelab/

Limitations of MATLAB

- To a great extent, computation details are hidden.
- Limited efficiency: MATLAB is an interpreted language.
  - Compiler also available
  - Can use MEX (MATLAB executable) to call DLL implemented in C
- Lack of properties to support large scale software development
  - E.g. implicit & dynamic data type
- MATLAB is the required implementation language for this class.

Learning MATLAB

- How to write a MATLAB function
  - Video

- Some references
  - There are many MATLAB tutorials online and books available.
  - Kermit Sigmon, MATLAB Primer, 3rd ed.
  - MATLAB User's Guide.

- Reminder: MATLAB computation results can be saved in .mat files and loaded back.

- Reminder: MAT files are exchangeable on different platforms.

Getting Help with MATLAB

- First, read related references and practice.
- For a specific function, it is often helpful to look in MATLAB online help.
- For a general question, it is often helpful to check related toolbox manuals.
- If none of these works, direct your questions to the instructor.

Outline

- Basics of MATLAB
- Chemical composition of a cell
- Chemical bonds of cellular molecules
- A brief introduction to protein structures
- A brief introduction to protein folding

Chemical Compositions of Cells (I)

- Cells are made mostly of macromolecules.
- Macromolecules dominate the physics and chemistry of a cell.

Table 2–3 Approximate Chemical Compositions of a Typical Bacterium and a Typical Mammalian Cell

COMPONENT	PERCENT OF TOTAL CELL WEIGHT *E. COLI* BACTERIUM	MAMMALIAN CELL
H_2O	70	70
Inorganic ions (Na^+, K^+, Mg^{2+}, Ca^{2+}, Cl^-, etc.)	1	1
Miscellaneous small metabolites	3	3
Proteins	15	18
RNA	6	1.1
DNA	1	0.25
Phospholipids	2	3
Other lipids	–	2
Polysaccharides	2	2
Total cell volume	2×10^{-12} cm^3	4×10^{-9} cm^3
Relative cell volume	1	2000

Alberts *MBoC* 5e

Chemical Compositions of Cells (II)

- Macromolecules in cells are usually polymers.
- Polymer: a natural or synthetic compound of large molecules that are formed by a linked series of repeated structural units.

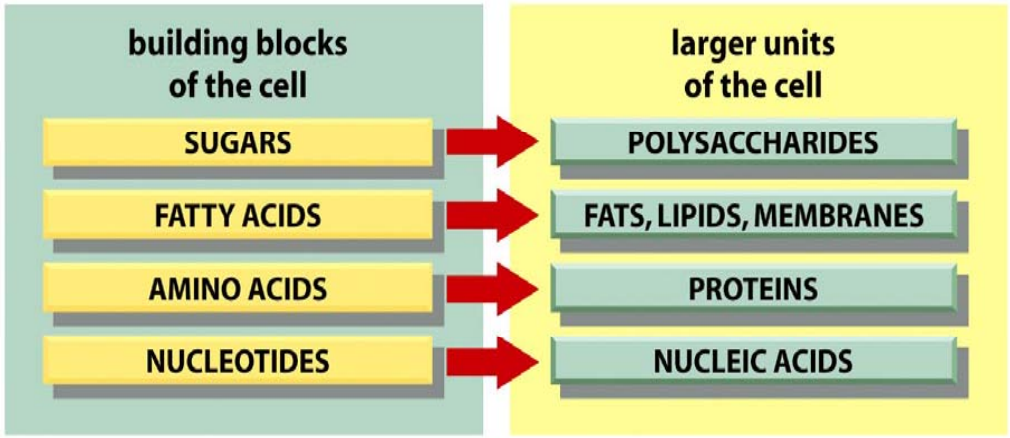


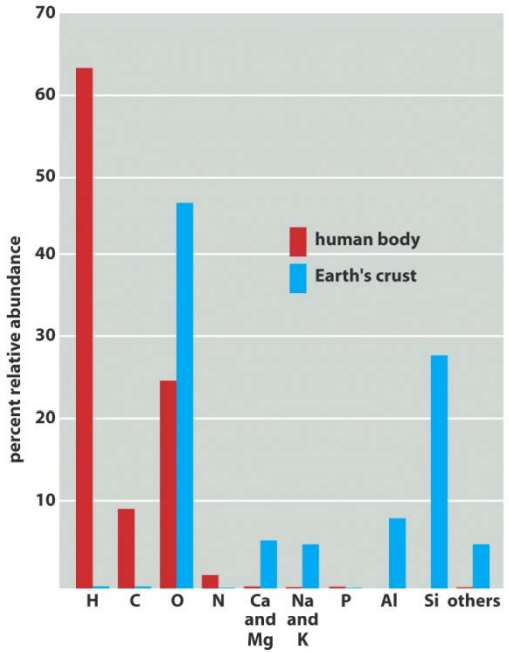
Table 2–2 The Approximate Chemical Composition of a Bacterial Cell

	PERCENT OF TOTAL CELL WEIGHT	NUMBER OF TYPES OF EACH MOLECULE
Water	70	1
Inorganic ions	1	20
Sugars and precursors	1	250
Amino acids and precursors	0.4	100
Nucleotides and precursors	0.4	100
Fatty acids and precursors	1	50
Other small molecules	0.2	~300
Macromolecules (proteins, nucleic acids, and polysaccharides)	26	~3000

Alberts *MBoC* 5e

Atoms Forming a Cell & Chemical Bonds

- All living organisms are fundamentally chemical systems.
- Cells are made primarily of a few chemical elements.
  - Organic chemistry; biochemistry
  - Statistical mechanics, thermodynamics
- Cell chemistry is based overwhelmingly on carbon compounds and reactions in water.
- Atoms that make up a molecule are joined together by different chemical bonds, which define boundaries between molecules.



Alberts *MBoC* 5e

Outline

- Basics of MATLAB
- Chemical composition of a cell
- Chemical bonds of cellular molecules
- A brief introduction to protein structures
- A brief introduction to protein folding

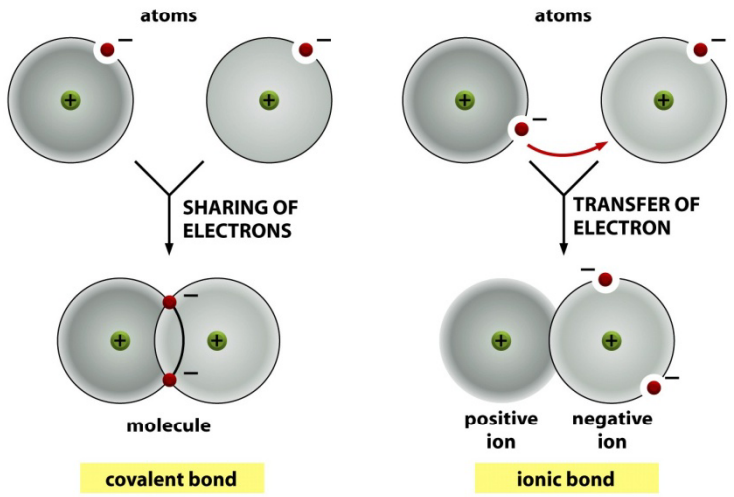
Chemical Bonds of Cellular Molecules (I)

- Chemical bonds of cellular molecules
  - Covalent bonds
  - Noncovalent bonds
- Four types of noncovalent attractions
  - Electrostatic attractions
  - Hydrogen bonds
  - van der Waals attraction
  - Hydrophobic force
- Ionic bond is a type of strong electrostatic attraction between fully charged atoms.

Table 2–1 Covalent and Noncovalent Chemical Bonds

BOND TYPE	LENGTH (nm)	STRENGTH (kcal/mole) IN VACUUM	IN WATER
Covalent	0.15	90	90
Noncovalent: ionic*	0.25	80	3
hydrogen	0.30	4	1
van der Waals attraction (per atom)	0.35	0.1	0.1

*An ionic bond is an electrostatic attraction between two fully charged atoms.

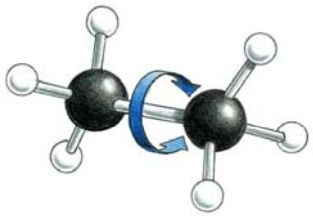


Chemical Bonds of Cellular Molecules (II)

- Covalent bonds are abundant in cellular molecules.

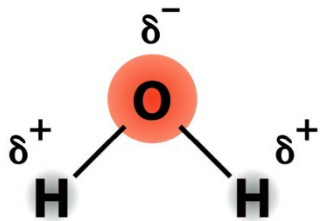
- There are different types of covalent bonds.
  - single bonds, double bonds
  - ploar covalent bonds vs nonpolar covalent bonds

- Polar covalent bonds allow cellular molecules to interact through electrostatic forces.

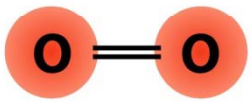


(A) ethane

(B) ethene



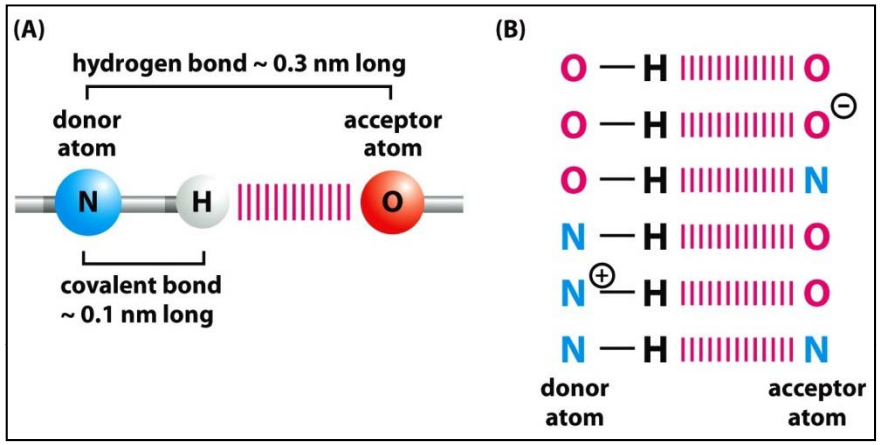
water



oxygen

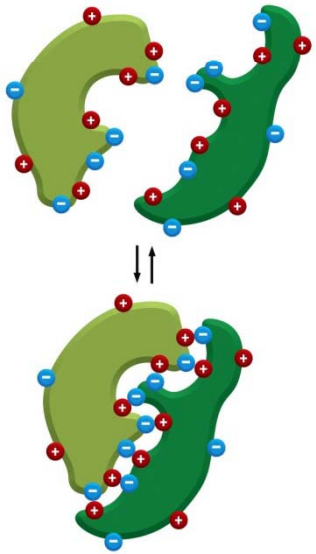
Chemical Bonds of Cellular Molecules (III)

- Hydrogen bonds: an electropositive hydrogen is shared by two electronegative atoms.
  - Highly directional
  - Example: nucleotide base pairing



Alberts *MBoC* 5e

- Nonpolar atoms can become dipoles transiently due to the fluctuation of their electron cloud.

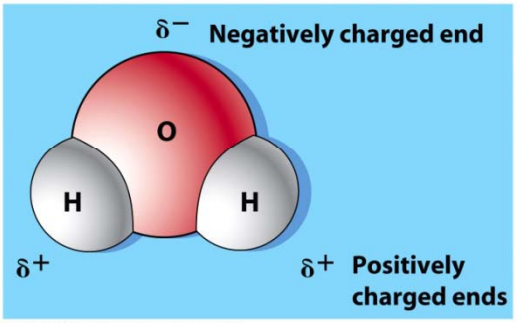


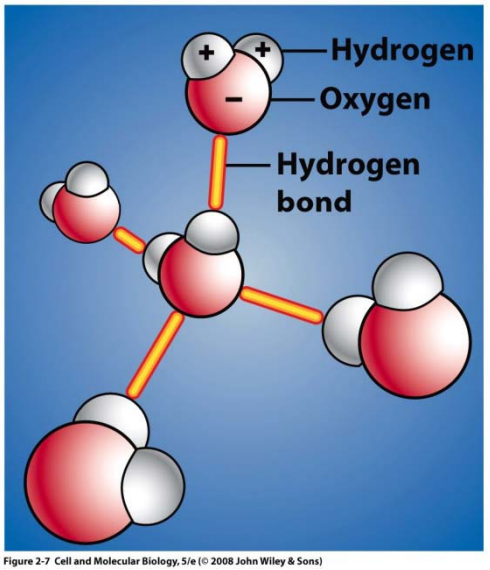
- van der Waals attractions result from attraction of atoms of opposite transient dipoles

- Water weakens hydrogen bonds but not van der Waals attractions.

Chemical Bonds of Cellular Molecules (IV)

- Water molecules are polar.
- Water molecules form hydrogen bonds with each other.
- Functions
  - A solvent for most cellular molecules.
  - Reactant or product in cellular biochemical reactions
- Water molecules forms hydrogen binds with many cellular molecules and generates functionally important complexes.
  - e.g. ion-water complex can affect ion permeability





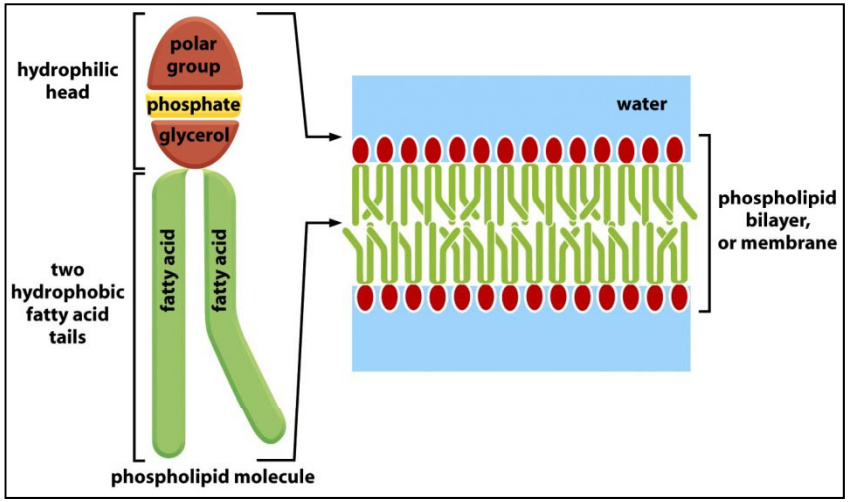
Karp *CMB* 5e

Chemical Bonds of Cellular Molecules (V)

- Molecules can be hydrophilic or hydrophobic.
  - hydrophilic: polar or charged molecules/groups that dissolve easily in water
  - hydrophobic: nonpolar molecules/groups that are insoluble in water

- Nonpolar surfaces tend to be pushed out of the water molecule network
  → Hydrophobic force

- Hydrophobic effect stabilizes biological structure.



Intermolecular Bonding: Noncovalent Bonds

- Hydrophobic interaction results from the pushing of nonpolar surfaces out of the hydrogen-bonded water network.
  - Brings nonpolar surfaces together to reduce contact with water
  - Critical to protein folding

- Noncovalent bonds define interactions between molecules.

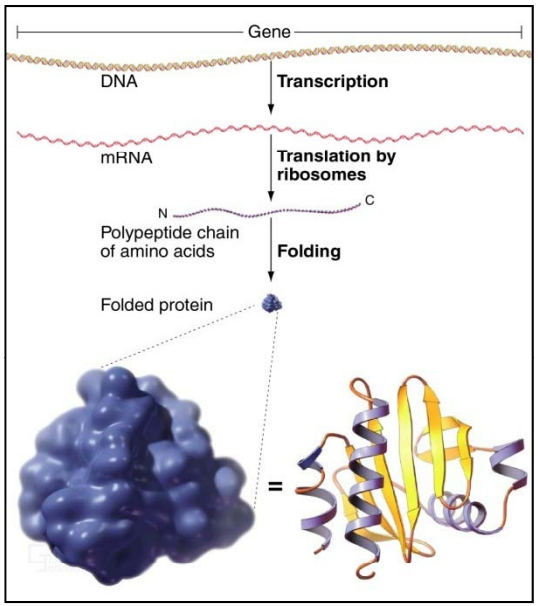
- Interactions between cellular molecules are defined by relative weak noncovalent bonds.
  - sensitivity
  - flexibility
  - transient interaction

Outline

- Basics of MATLAB
- Chemical composition of a cell
- Chemical bonds of cellular molecules
- A brief introduction to protein structures
- A brief introduction to protein folding

Proteins: Overview

- Proteins are the predominant structural and functional components of all cells.
- Proteins vary widely in length, typically in the range of 100~1000 amino acids.
- Determination of protein sequence
  - Genetic approach
  - Mass spectrometry

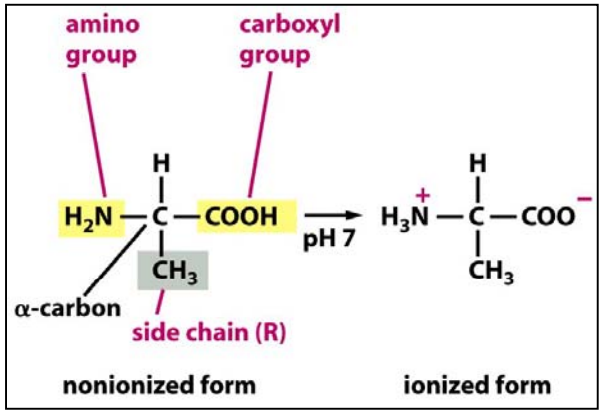


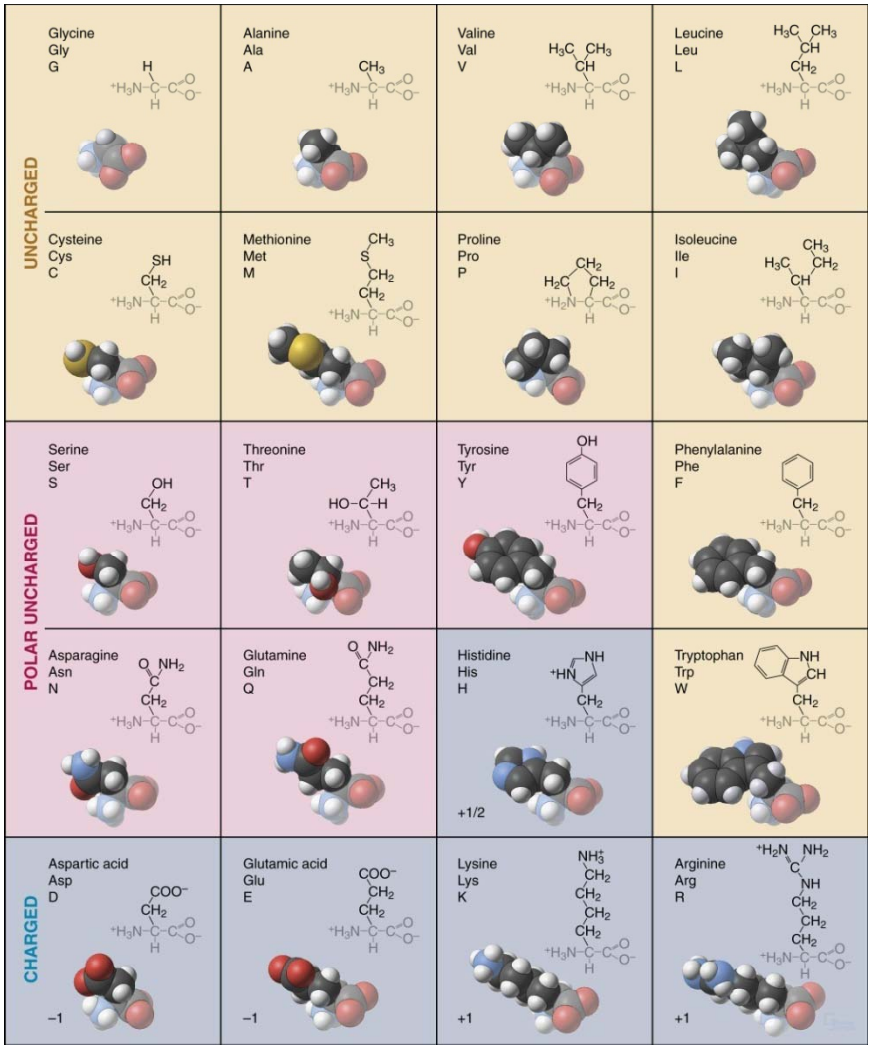
From **Protein Structure and Function** by Gregory A Petsko and Dagmar Ringe

1st position (5' end)	2nd position U	2nd position C	2nd position A	2nd position G	3rd position (3' end)
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Protein Primary Structure (I)

- 20 naturally occurring amino acids
- Amino acids differ in side chains

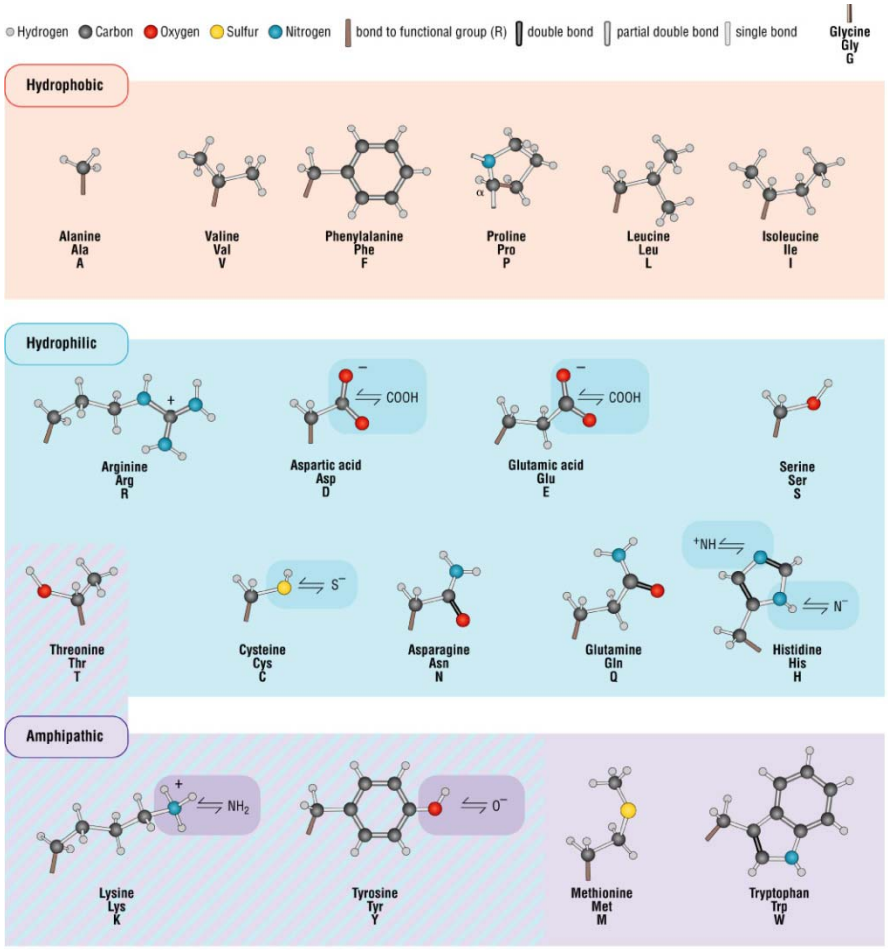




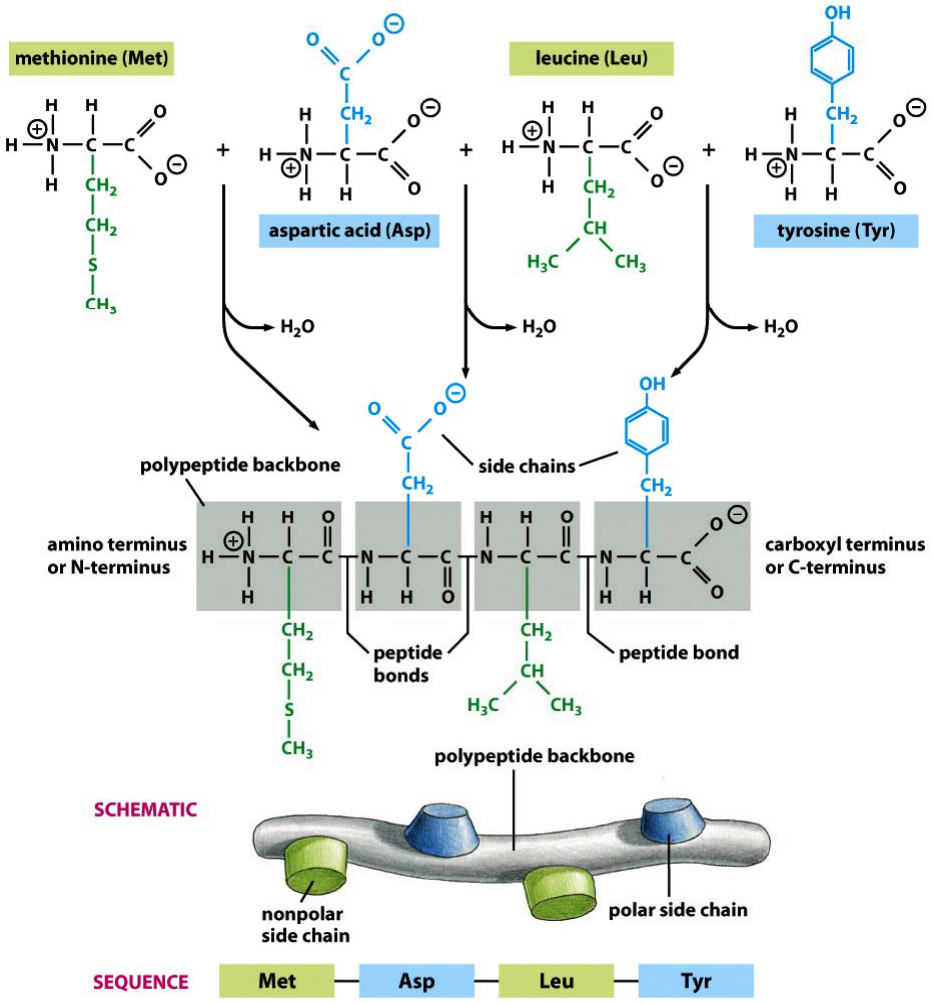
Protein Primary Structure (II)

- Amino acids can be hydrophobic, hydrophilic, or amphipathic

- Amphipathic: residues that have both polar and unpolar properties and are ideal for forming interfaces.



Protein Primary Structure (III)

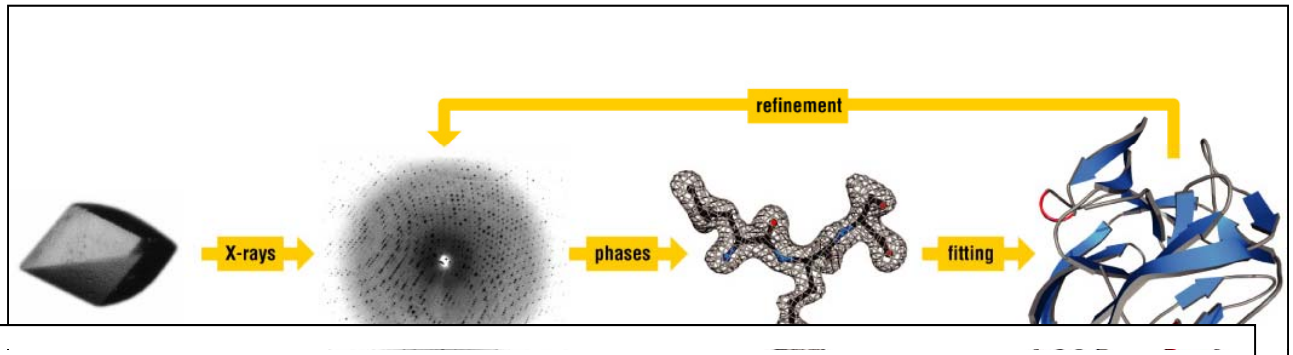


Terminology

- **Genome:** all the genetic information encoded in a cell. → Genomics
- **Proteome:** the complete set of proteins expressed in a cell. → Proteomics
- **Backbone:** the regularly repeating part of a polymer.
- **Residue:** the basic building block of a polymer.
- **Sidechain:** the chemical group that protrudes from the backbone.
- **Polypeptide:** a linear polymer of amino acids.
- **Homologs:** Different forms of a gene/protein that are similar in sequence as a result of derivation from the same ancestral gene.
- **Isoforms:** Different forms of a protein that may be produced from different genes, or from the same gene by alternative splicing in the same cell.

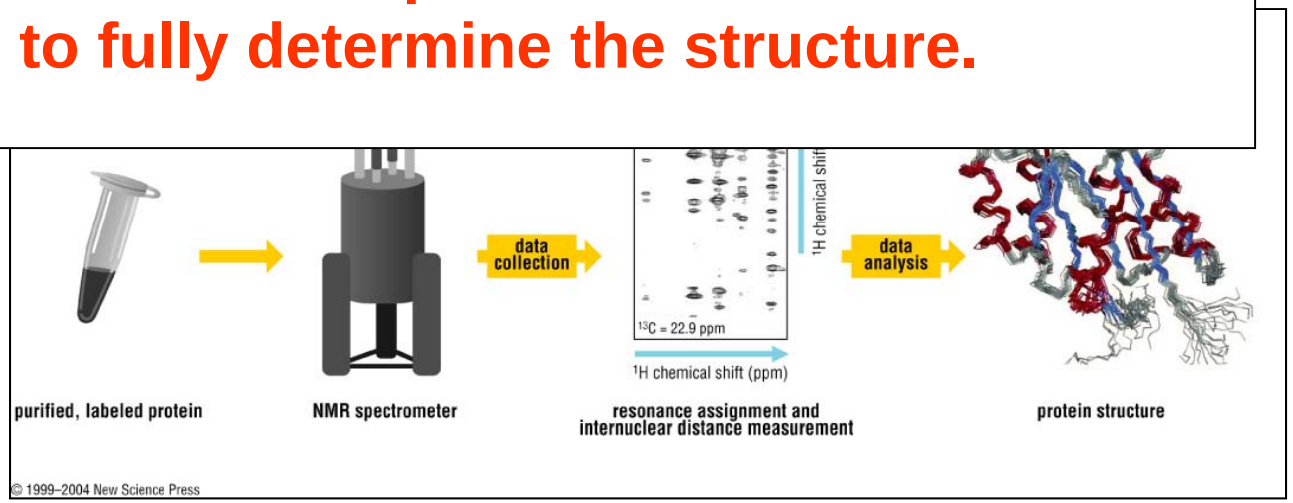
How Protein Structures are Determined

X-ray crystallography



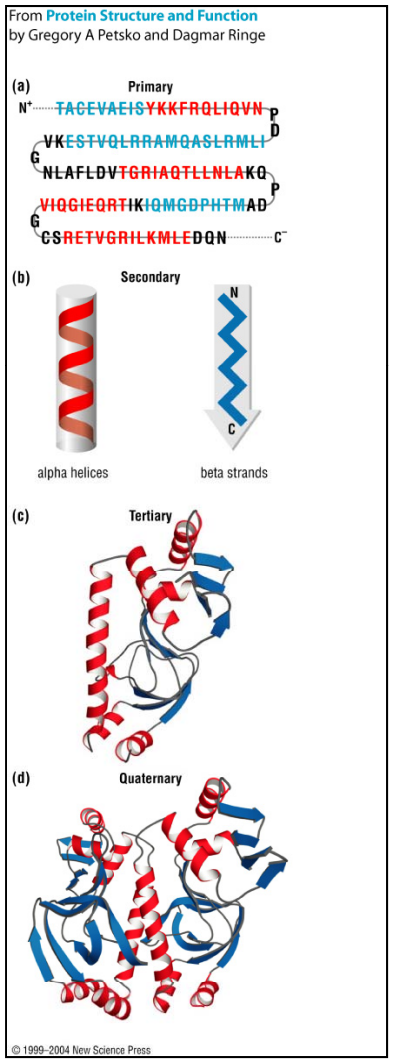
Structure data interpretation is not unambiguous. Information from multiple sources is often required to fully determine the structure.

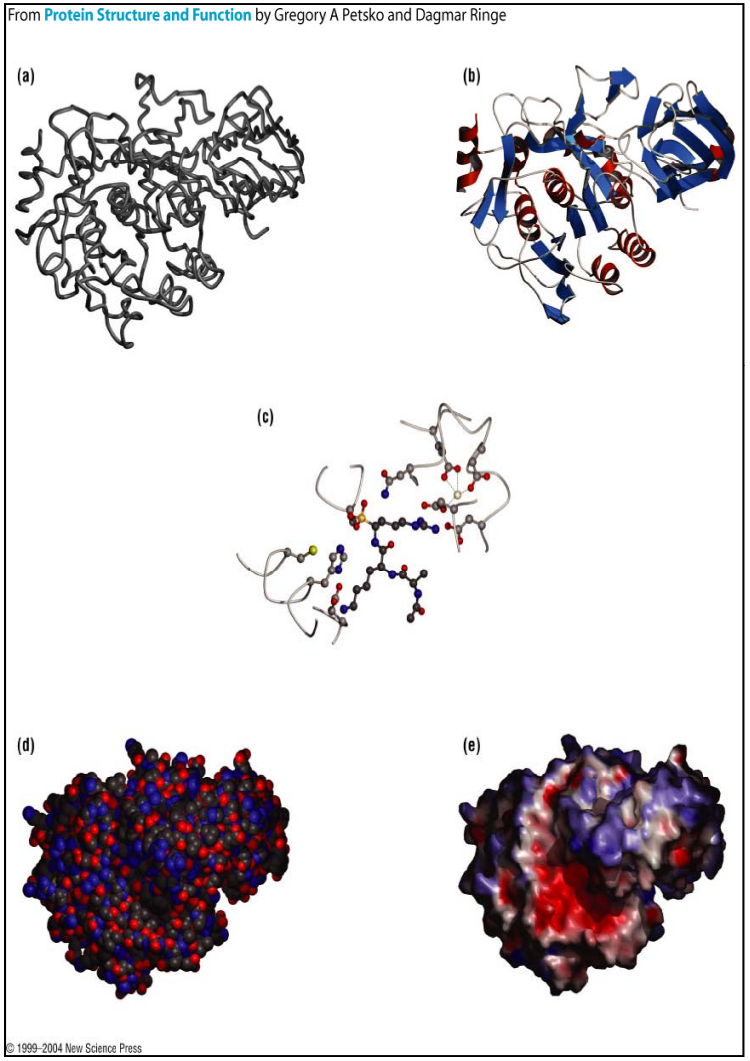
NMR
Nuclear magnetic resonance



Protein Structure Overview

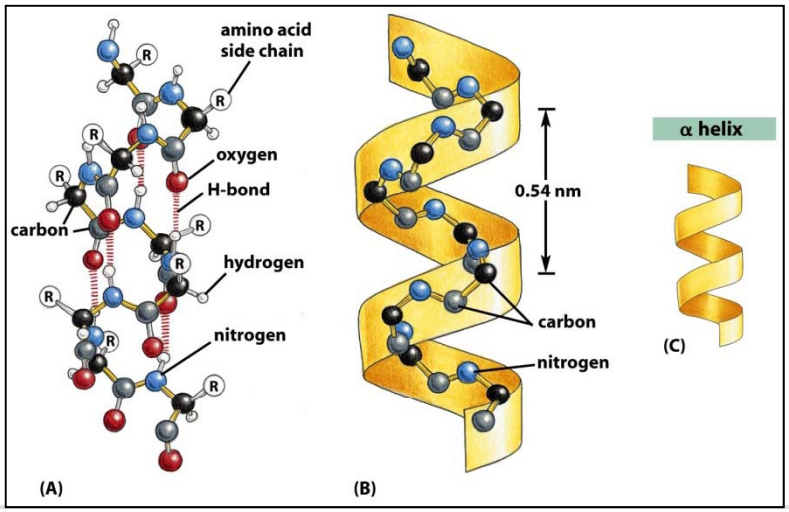
- A structural hierarchy
  - primary structure
  - secondary structure
  - tertiary structure
  - quarternary structure
- Different representations
  - wire diagram
  - ribbon
  - ball-and-stick
  - space filling
  - surface

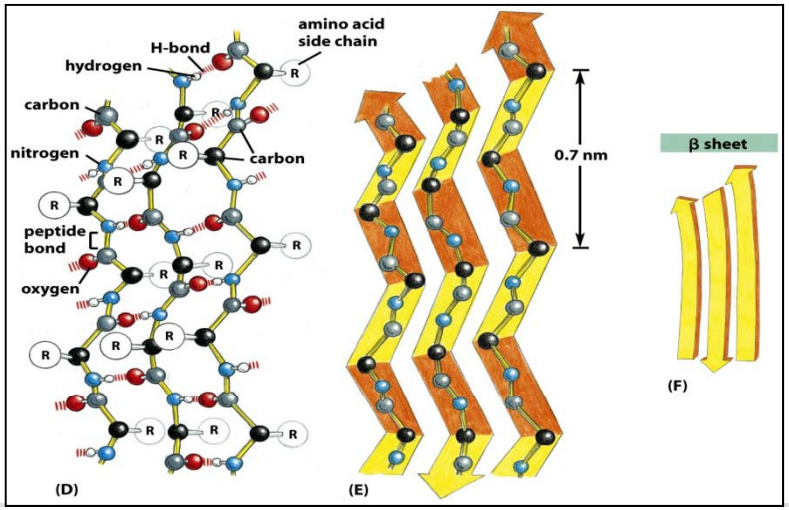




Secondary Structure (I)

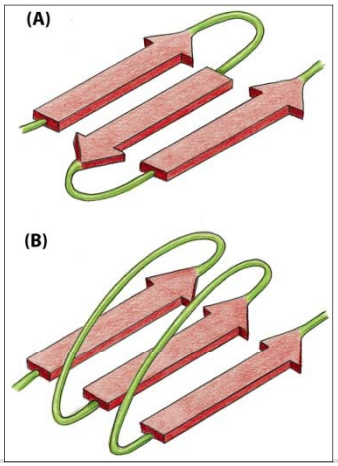
- Secondary structure
  - Local structures of repeated residue conformation
  - Two primary types of secondary structure elements (i.e. folding patterns: alpha helix, beta sheet.
- Alpha helix was first discovered in hair protein keratin.
- Beta sheet was first discovered in fibroin, the silk protein.
- Both patterns result from hydrogen bonding between N-H and C=O group of the backbone.

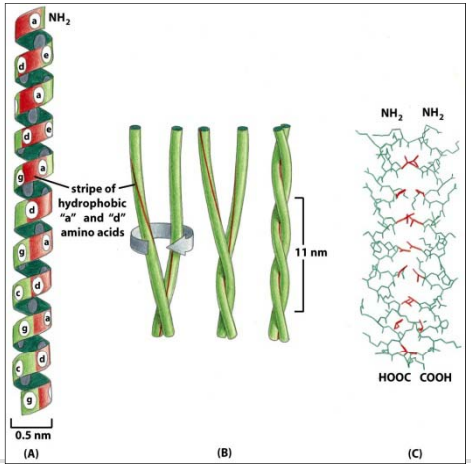


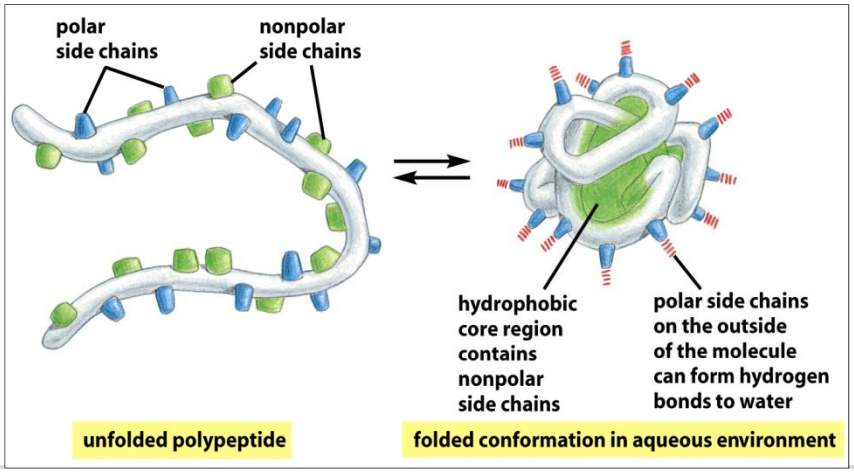


Secondary Structure (II)

- Many cellular proteins contain extensive regions of beta sheets, which provide structural rigidity.
- Alpha helix are abundant in membrane proteins.
- Many proteins contain a hydrophobic core.
- Secondary structure consists of extensive network of hydrogen bonds and contributes significantly to the stabilization of the overall protein structure.





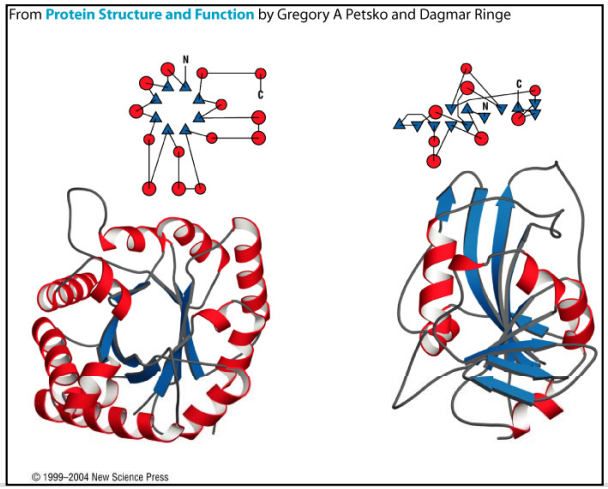


Tertiary Structure (I)

- Tertiary structure

  - The three dimensional conformation of a protein is its native folded state; i.e. the global organization of secondary structures.

- Tertiary structures are not regular. Proteins with similar secondary structure elements can have very different tertiary structures.



Left: triosephosphate isomerase
Right: dihydrofolate

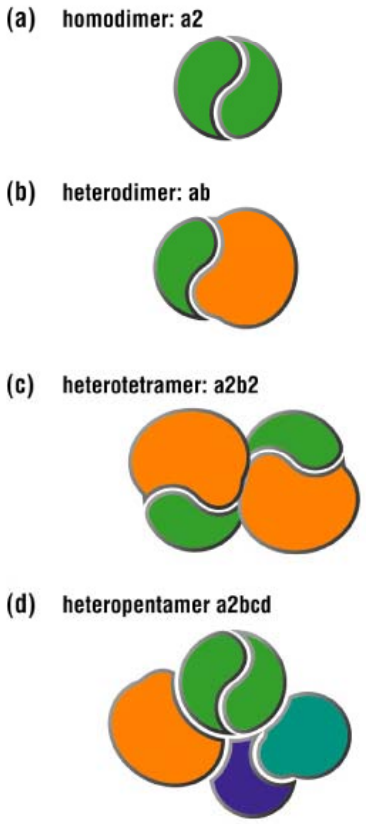
Tertiary Structure (II)

- The folded structure of a protein is directly determined by its primary structure.
- Condensing of multiple secondary structural elements leads to tertiary structure.
- Computational predication of folding is not yet reliable.
- Most folded proteins are marginally stable to allow flexibility.
- Conformation changes tend to be local.

Quarternery Structure (I)

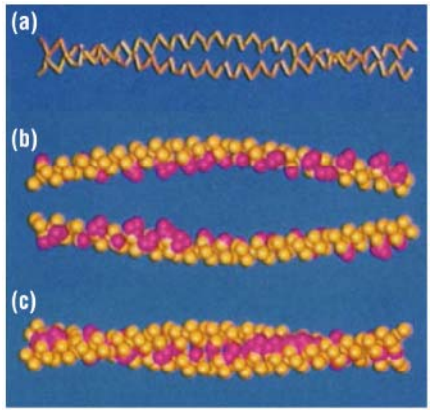
- Many proteins have more than one polypeptide chain. These proteins are called oligomers.
- Individual polypeptide chains are referred to as monomers.
- Quaternary structure is the arrangement of different polypeptide chains.

From **Protein Structure and Function**
by Gregory A Petsko and Dagmar Ringe

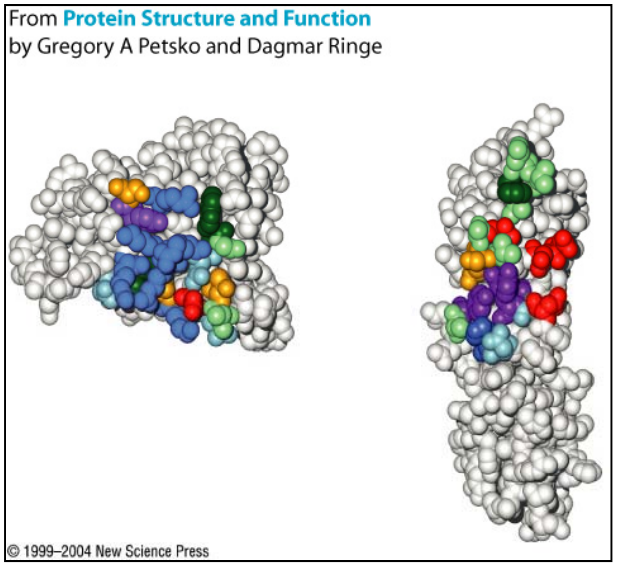


Quarternery Structure (II)

- Irregular protein surfaces enables specific binding.
- Specific intermolecular interactions depend on complementarity.
- Protein binding can trigger large conformational changes

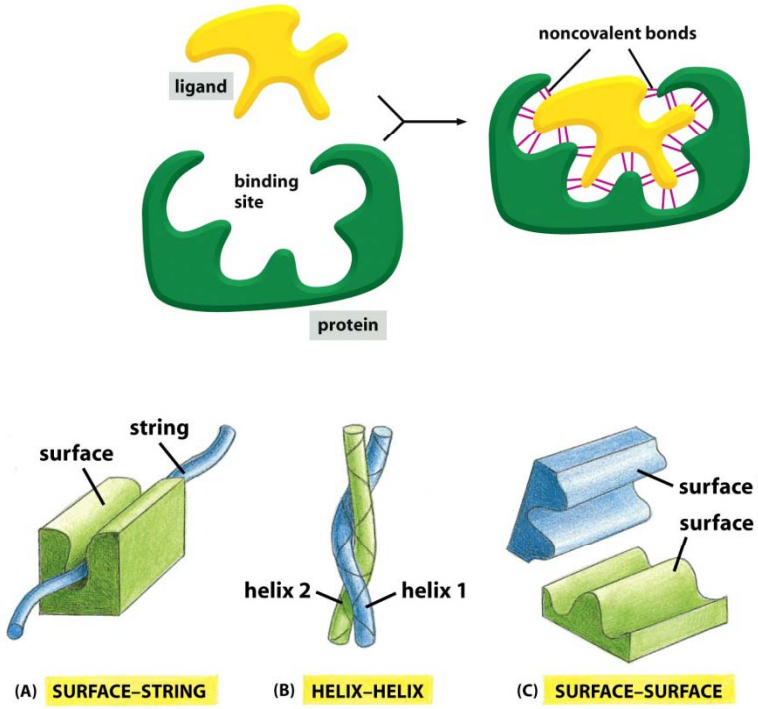


Alpha helical coiled coil



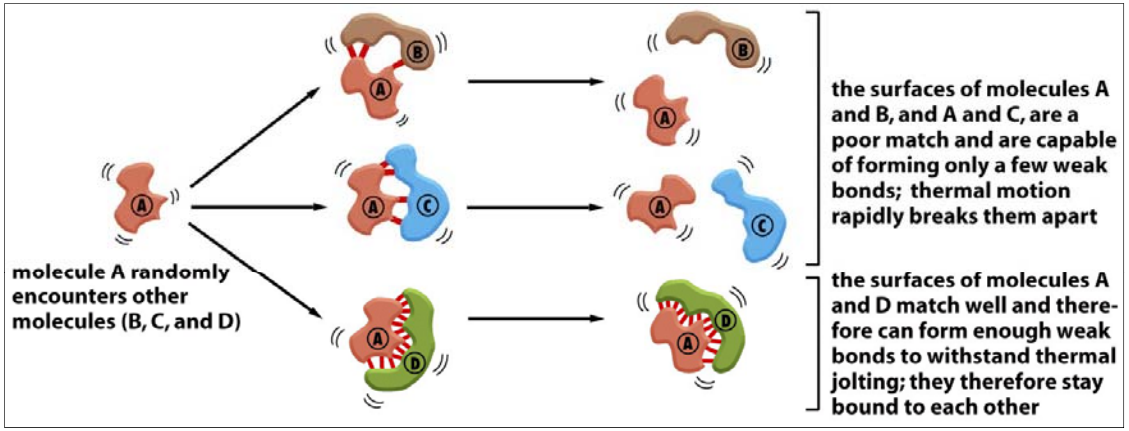
Protein Interactions (I)

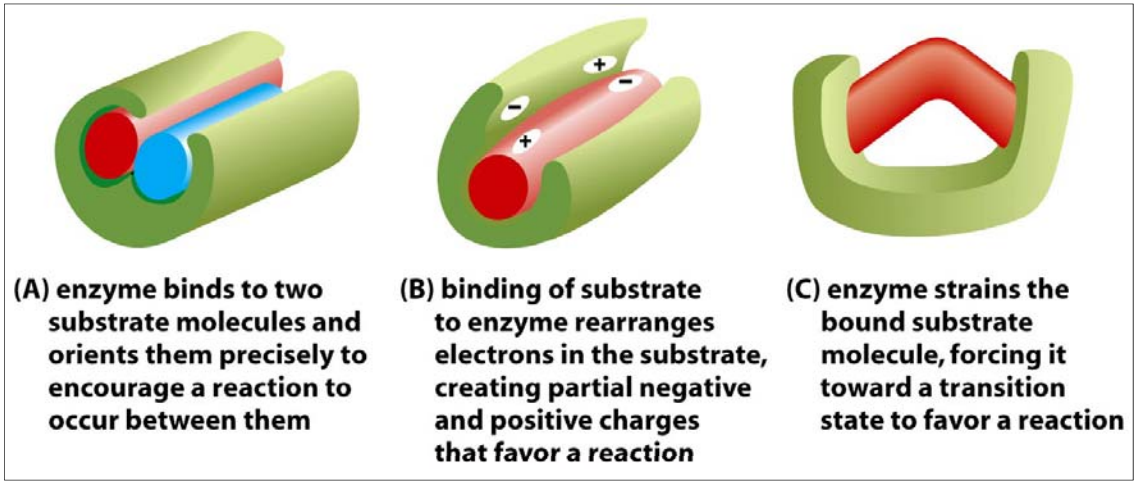
- Selectivity and affinity of protein binding depend on weak noncovalent bonds.
- Surface conformation of a protein defines its chemistry.
- The most common way of protein interaction is through precise matching of surfaces.
- Protein interactions often require catalysis by enzymes.



Protein Interactions (II)

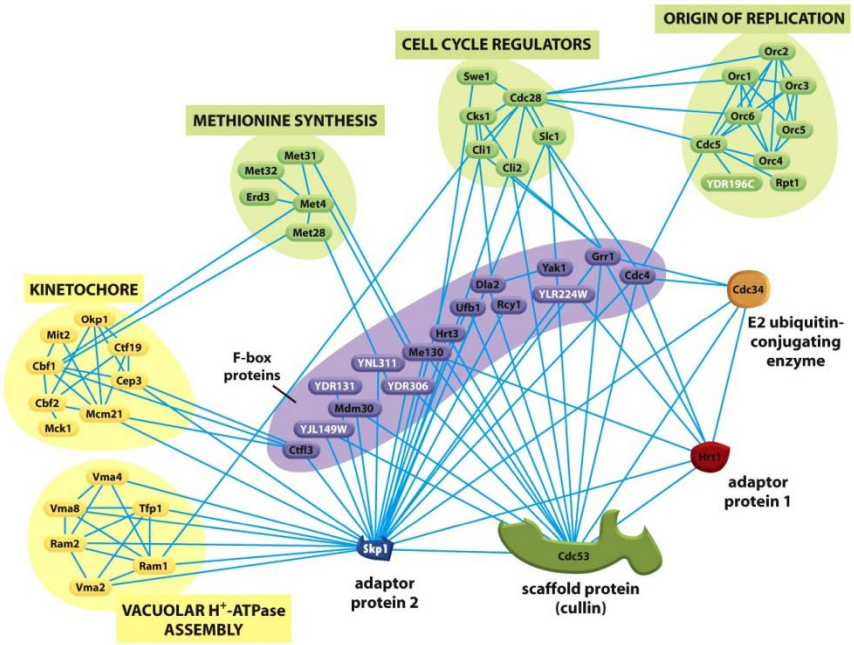
- Protein interactions often require catalysis by enzymes.

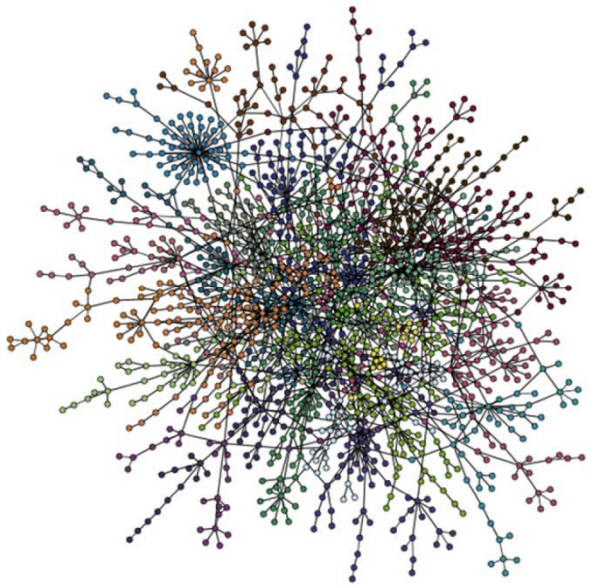




Protein Interactions (II)

- A complex network of protein interactions underlies cell function.



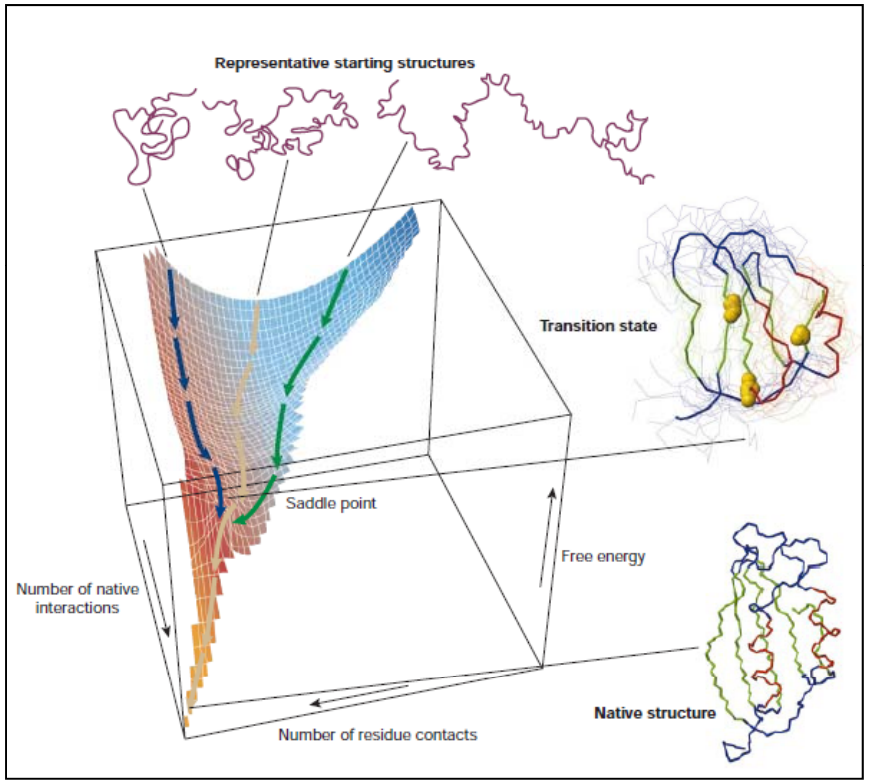


Outline

- Basics of MATLAB
- Chemical composition of a cell
- Chemical bonds of cellular molecules
- A brief introduction to protein structures
- A brief introduction to protein folding

Protein Folding: the Energy Landscape Theory

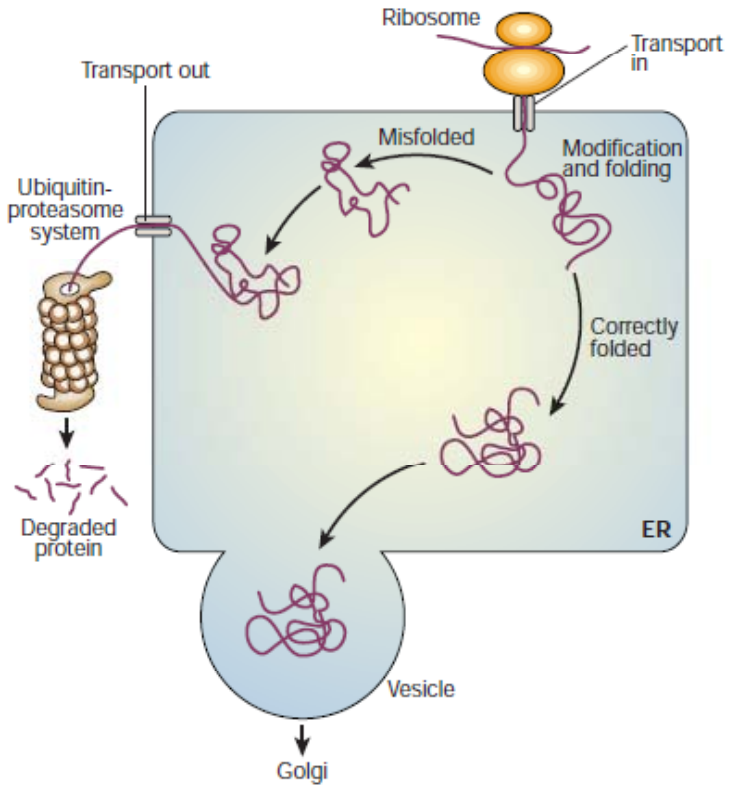
- First proposed by Joseph Bryngelson and Peter Wolynes.
- Principle of minimal frustration: the native folded state is favored by evolution.
- The energy landscape is encoded by the amino acid sequence and represents all the possible energy states.



Dobson, *Nature*. 426:884, 2003.

Protein Folding in Cells

- Protein folding in cells can happen before the completion of synthesis (co-translational).
- Complex protein structures often fold after exit from ribosomes.
- Incorrectly fold proteins are detected by a quality-control mechanism and sent for degradation.

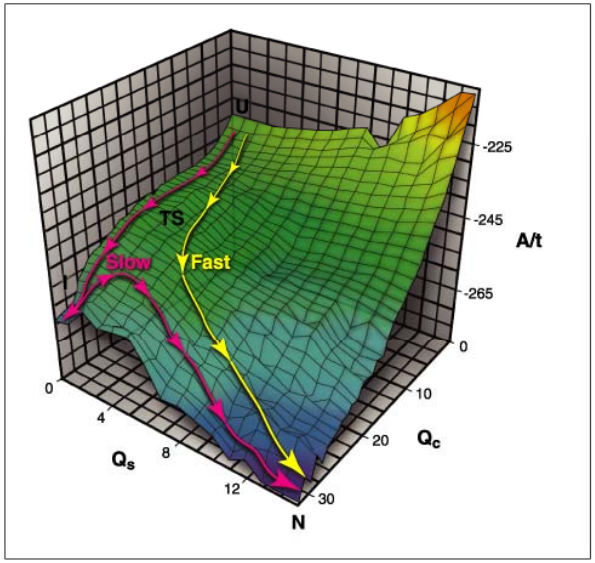


Dobson, *Nature*. 426:884, 2003.

Chaperone-Assisted Protein Folding

- Chaperons increase the efficiency of protein folding by avoiding unfavorable folding paths.

- Typical functions
  - To prevent aggregations.
  - To prevent interference



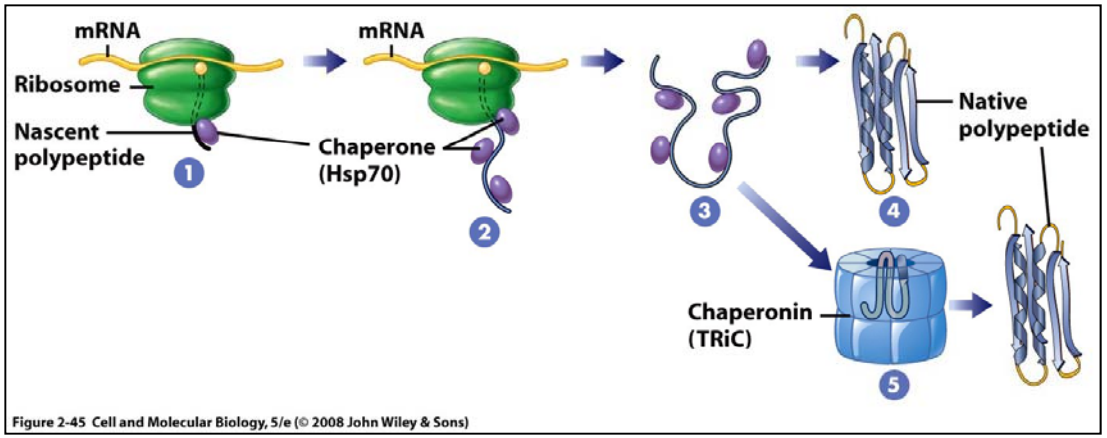
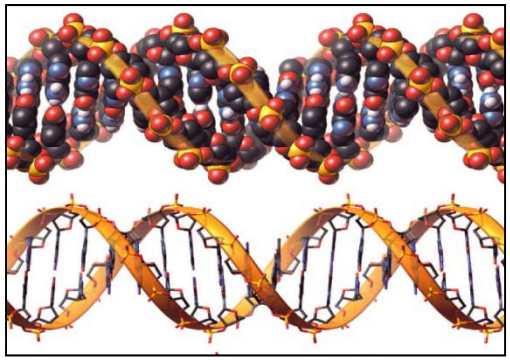


Figure 2-45 Cell and Molecular Biology, 5/e (© 2008 John Wiley & Sons)

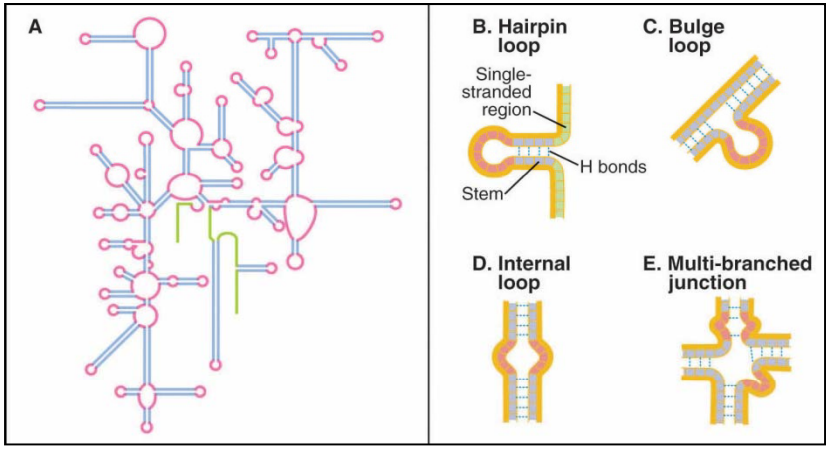
Structure of DNA and RNA

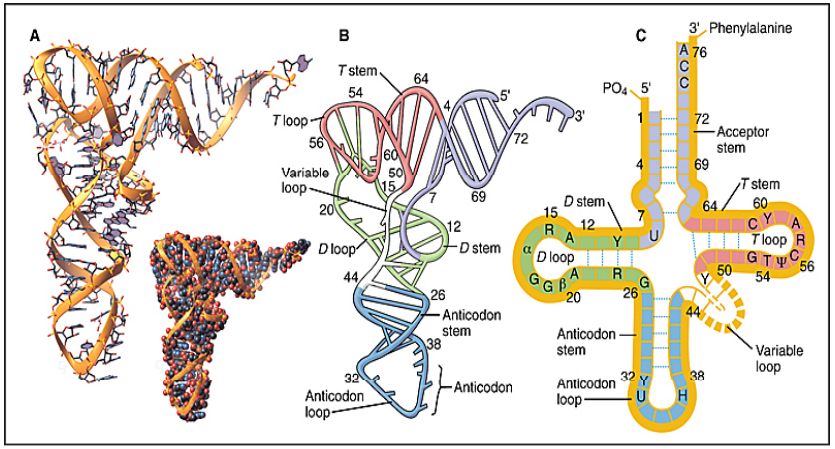
- DNA secondary structure

  - Purine: adenine (A) guanosine (G)
  - Pyrimidines: thymine (T) cytosine (C)



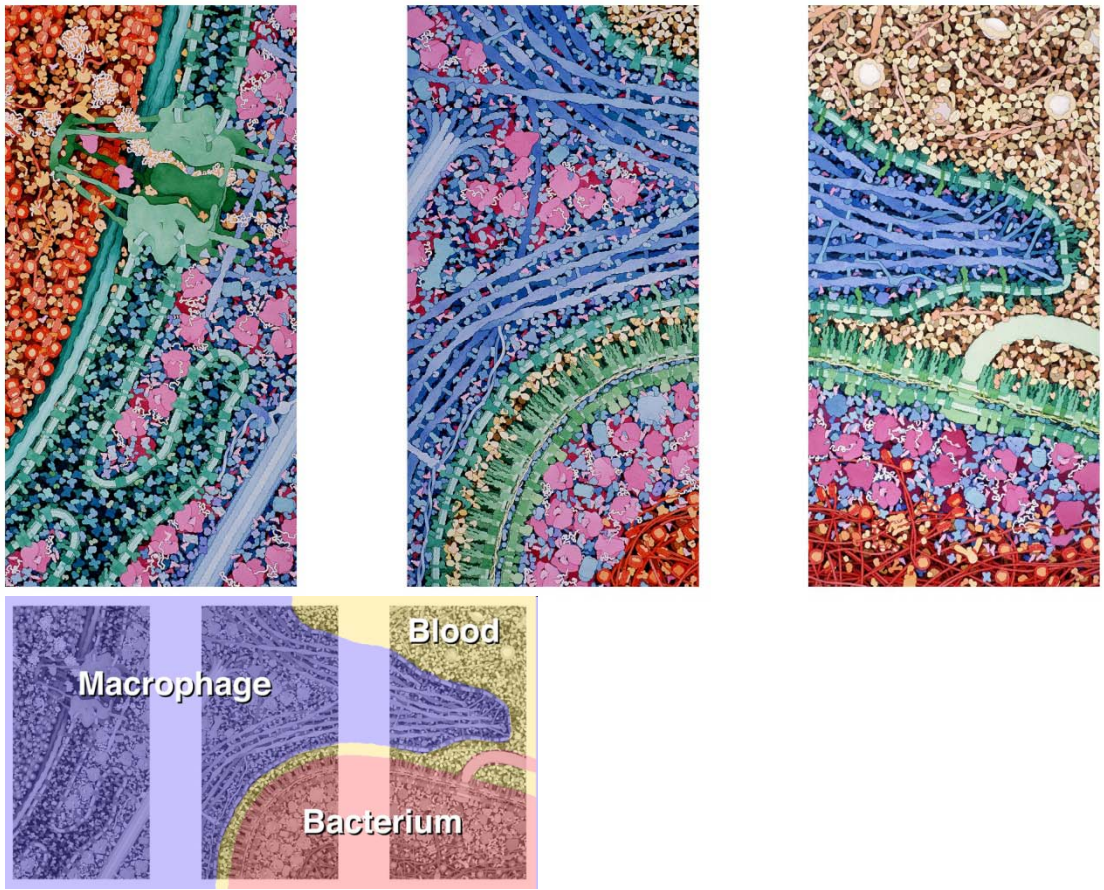
- RNA secendary structure
  - Purine: adenine (A) guanosine (G)
  - Pyrimidines: uracil (U) cytosine (C)





The Cytoplasm

- The global view: the cytoplasm is densely populated.
- Only correctly folded proteins have long-term stability.



Some Comments

- Biochemistry

  - The level of molecular details really depends on the question to be addressed.

- Structural biology

  - Provides critical insights into cellular processes.

  - Needs to be integrated with other approaches.

  - Structural genomics aims to determine the primary and tertiary structures of all proteins of a given organism.

Questions?

Small Molecules

- Definition of small molecules is not firm.
- Can be natural or synthesized. Not a polymer.
- Low molecular weight permits fast permeation across membranes.
- Used to induce immediate functional perturbations.
- Important targets of pharmacology and chemical biology research.

Table 2–2 The Approximate Chemical Composition of a Bacterial Cell

	PERCENT OF TOTAL CELL WEIGHT	NUMBER OF TYPES OF EACH MOLECULE
Water	70	1
Inorganic ions	1	20
Sugars and precursors	1	250
Amino acids and precursors	0.4	100
Nucleotides and precursors	0.4	100
Fatty acids and precursors	1	50
Other small molecules	0.2	~300
Macromolecules (proteins, nucleic acids, and polysaccharides)	26	~3000

Small Molecules References

- B. R. Stockwell, Exploring biology with small organic molecules, Nature, 432:846,2004.
- S. L. Schreiber, Small molecules: the missing link in the central dogma, Nature Chemical Biology, 1:64, 2005.
- S. Ding, P. G. Schultz, Small molecules and future regenerative medicine, Curr. Top. Med. Chem., 5:383, 2005.

An Overview of Cell Biology Literature

- Journals
  - General purpose journals
    - Science
    - Nature
    - PNAS
  - Specialized journals
    - Cell
    - Journal of Cell Biology
    - Nature associated journals
- Commercial vs noncommercial journals
- Review journals and review articles

How to Read Cell Biology Papers (I)

- To be able to critically read and evaluate contemporary *biology* papers - Why so critical?

- General guidelines
  - Fundamentally, it is about original data and ideas
  - Not that different from a mathematical proof:
    Logical coherence and rigor

- Highly stereotyped structures of biology papers

- Organization (I): biology papers are result-driven
  - Introduction: However, …
  - Results: To…, we did …
  - Discussion: We speculate …

How to Read Cell Biology Papers

- Organization (II):
  - Every figure must tell
  - Logical flow: connection between result sections

- Our aims

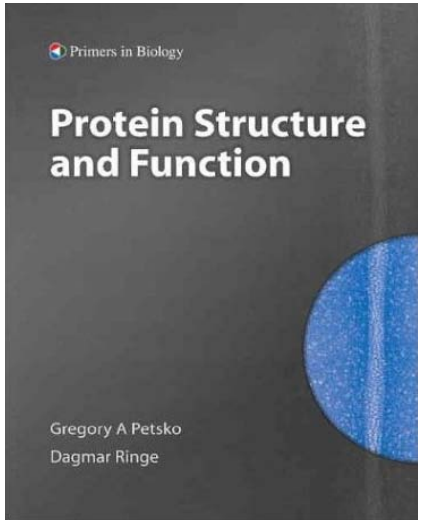
  - To be able to effectively read papers in cell biology
  - To be able to effectively communicate cell biology research results

Process of Publication

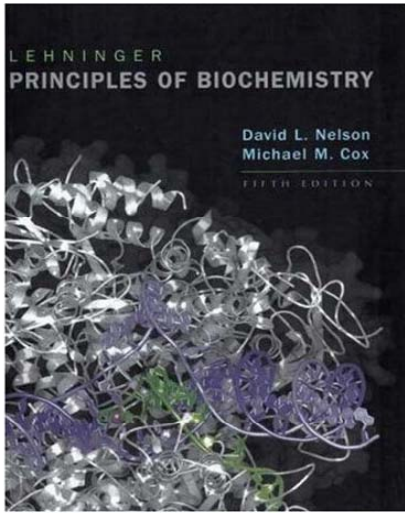
- Journal selection
  - What are the messages: *short vs long format*
  - Usually several comparable journals to choose from
    Similar paper formats
    Similar review standards
  - Keep a rational perspective: *vanity journals*
  - Keep doing good science, your record will show

- Submission and review process
  - Pre-submission inquiry: usually for vanity journals
  - Editorial review
  - External review
  - Outcome I: preliminary acceptance
    Point-to-point response to reviews
  - Outcome II: rejection
  - Peer-review system not perfect but generally works

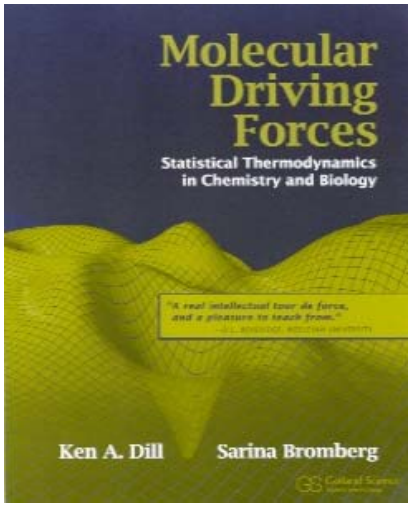
References



Petsko & Ringe
New Science, 2004



Nelson & Cox
W.H. Freeman, 2008



Dill & Bromberg,
Garland Sciences, 2002

Project Assignment 1