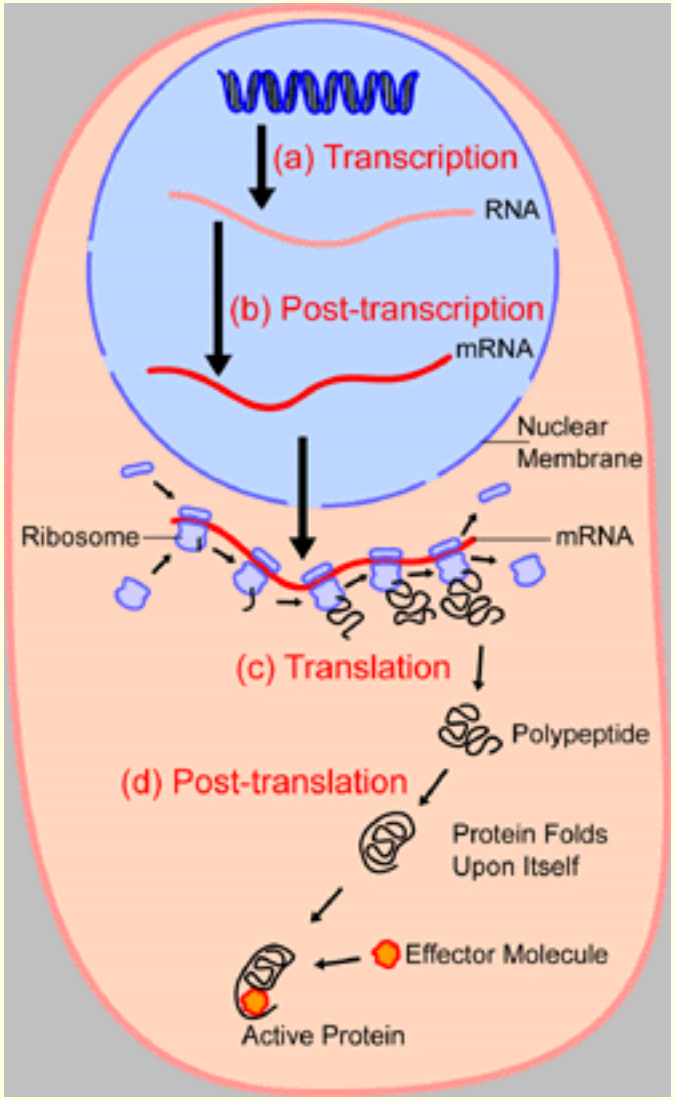
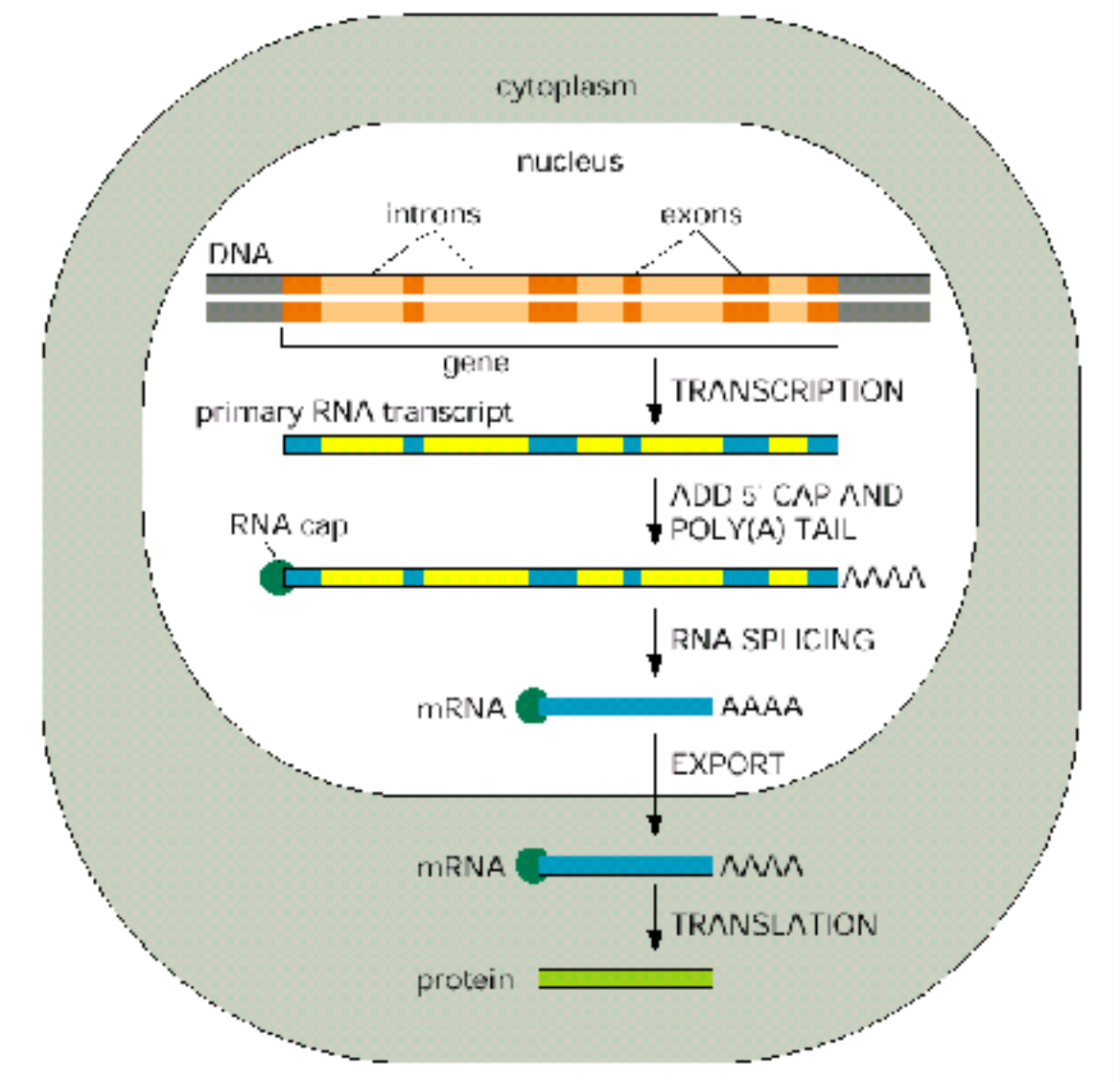
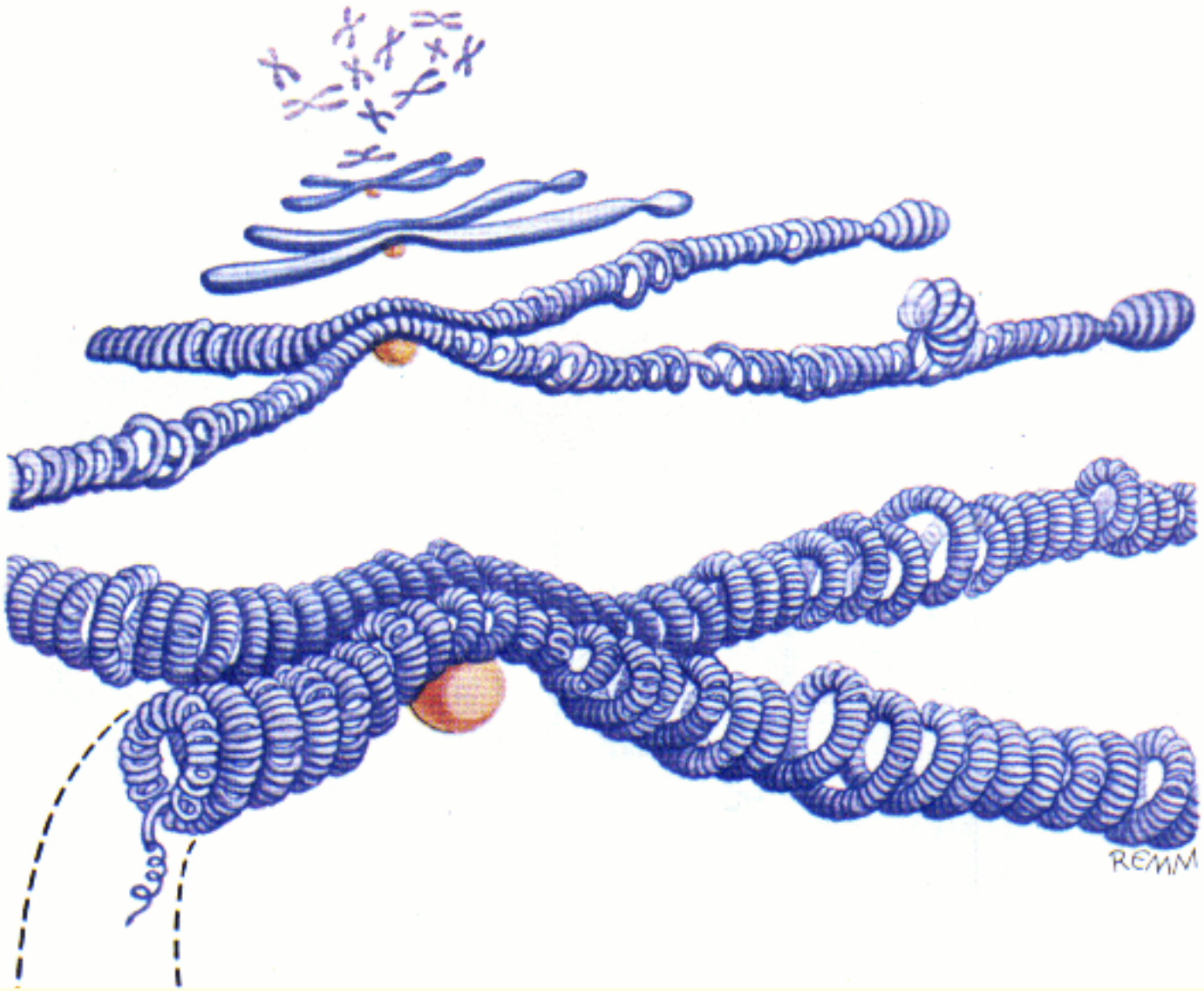
CMPS 6630: Introduction to Computational Biology and Bioinformatics

Protein Structure

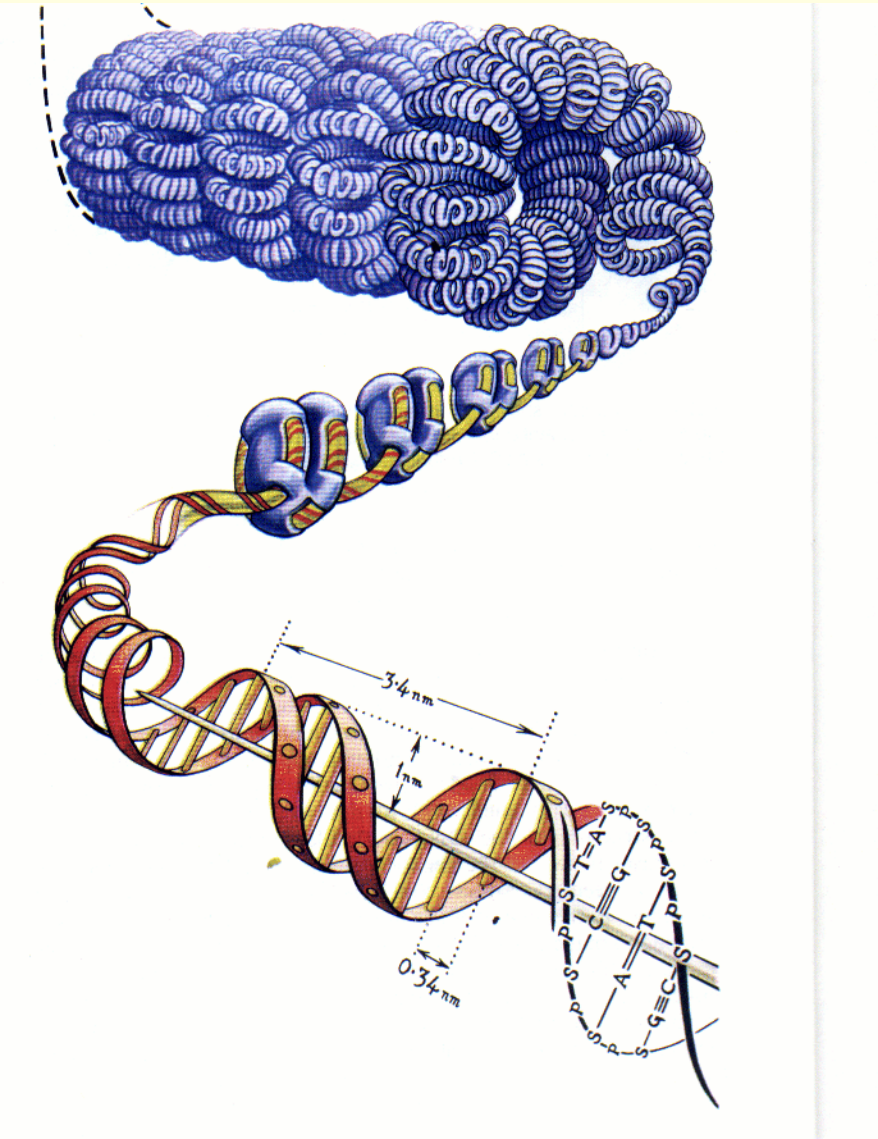
Recap: Central “Dogma”






REMM

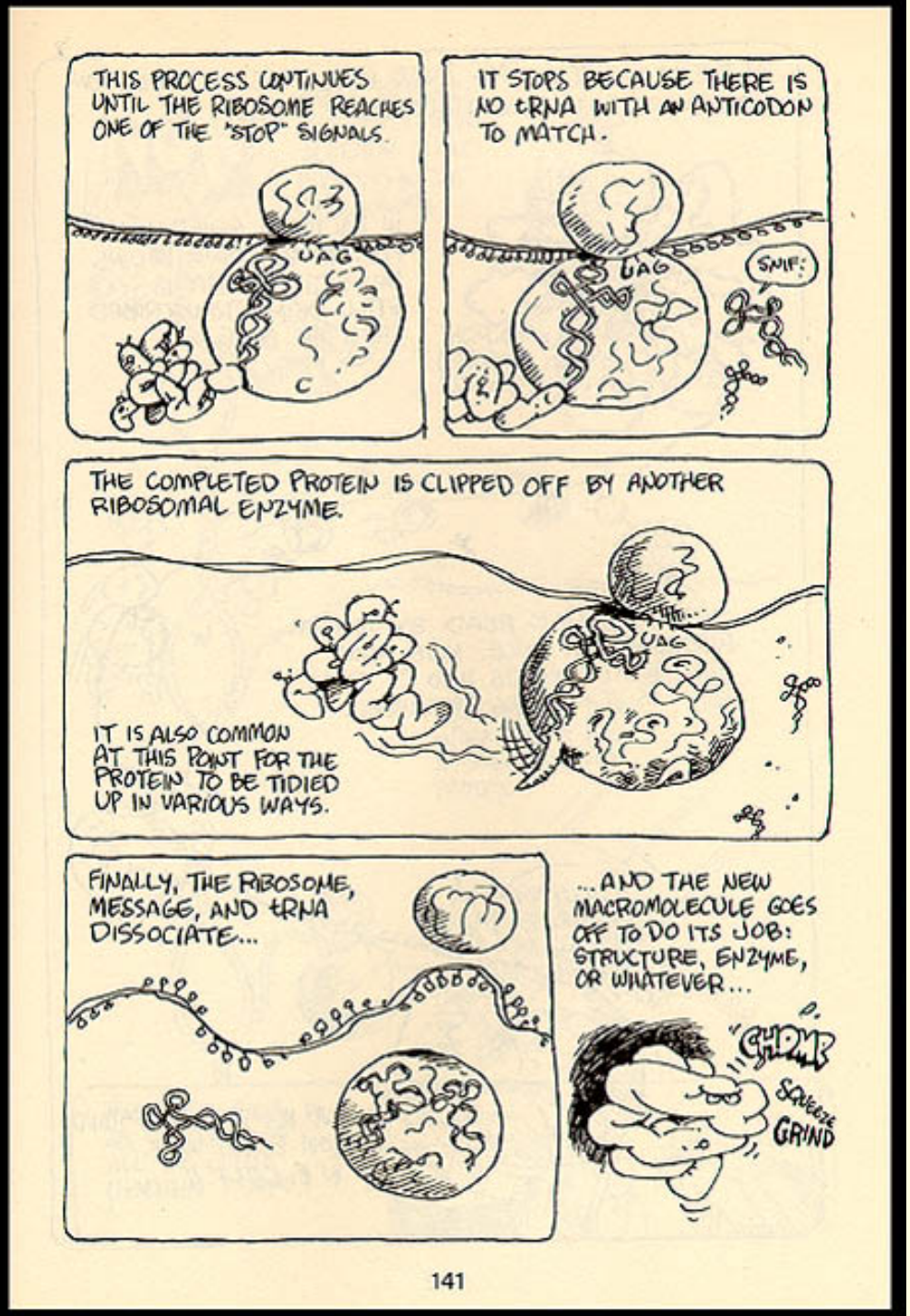
DNA Structure



Franklin/Watson/Crick showed that DNA has a structure that is stable, and facilitates replication.

Enzymes that bind DNA and RNA must have a “compatible” structure (e.g. ribosomes).

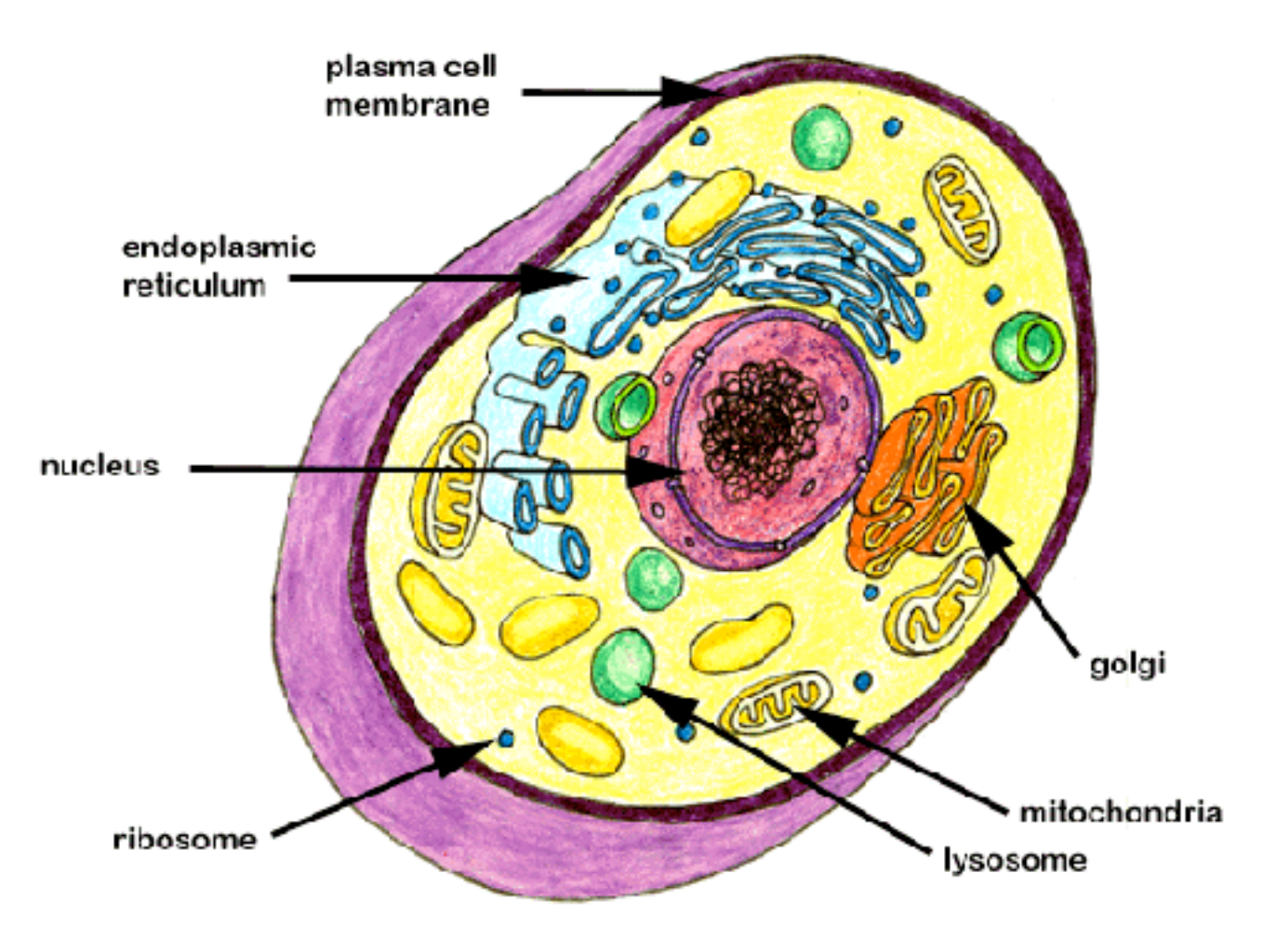
What is the (molecular) nature of this regulation?



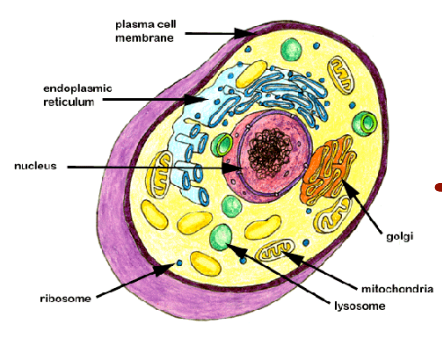
Proteins have 3D shape that is determined by a sequence of amino acids.

Structure is function (e.g. ribosome, hemoglobin, transcription factors).

Cell Composition



Cell Composition



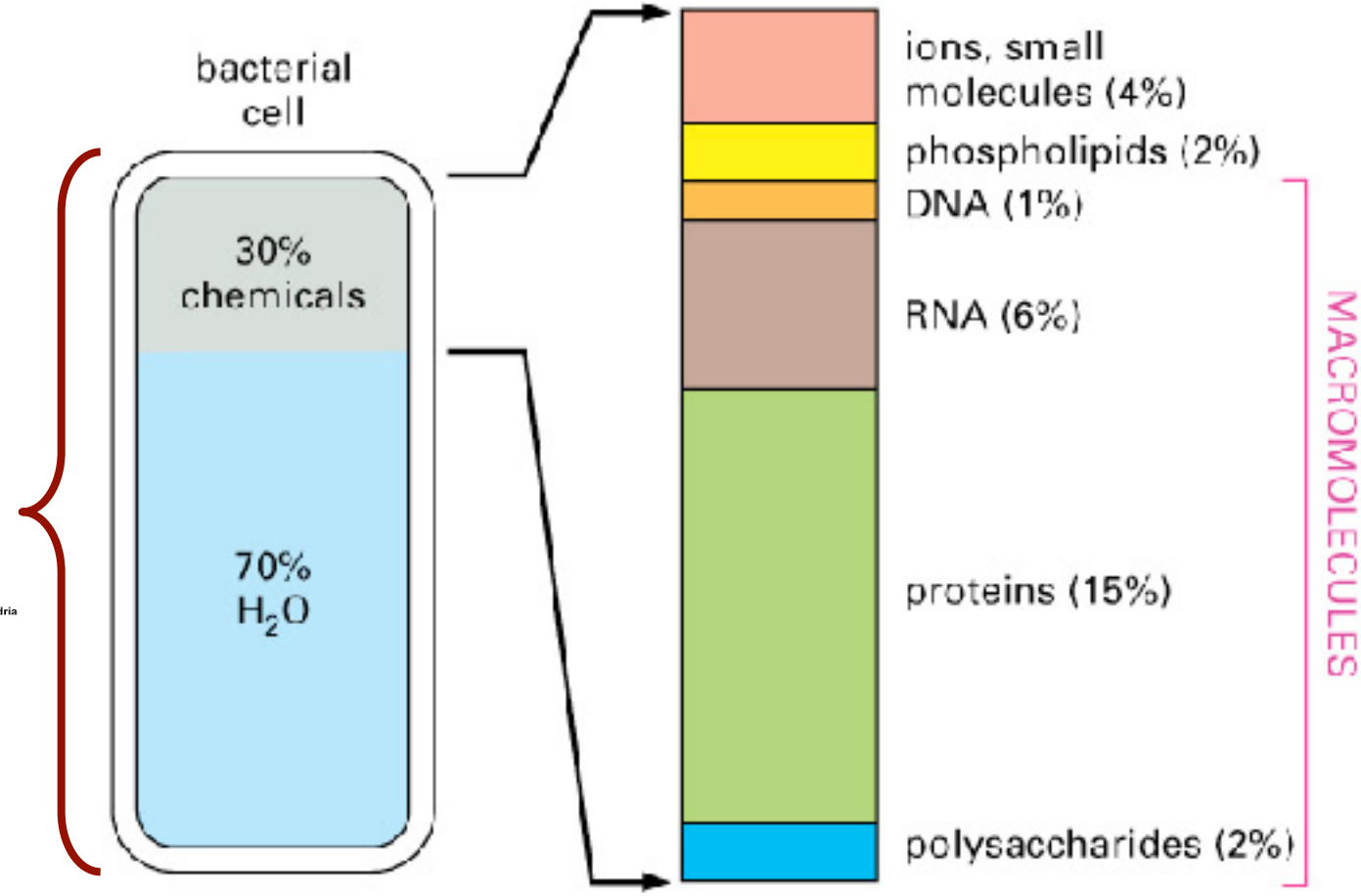
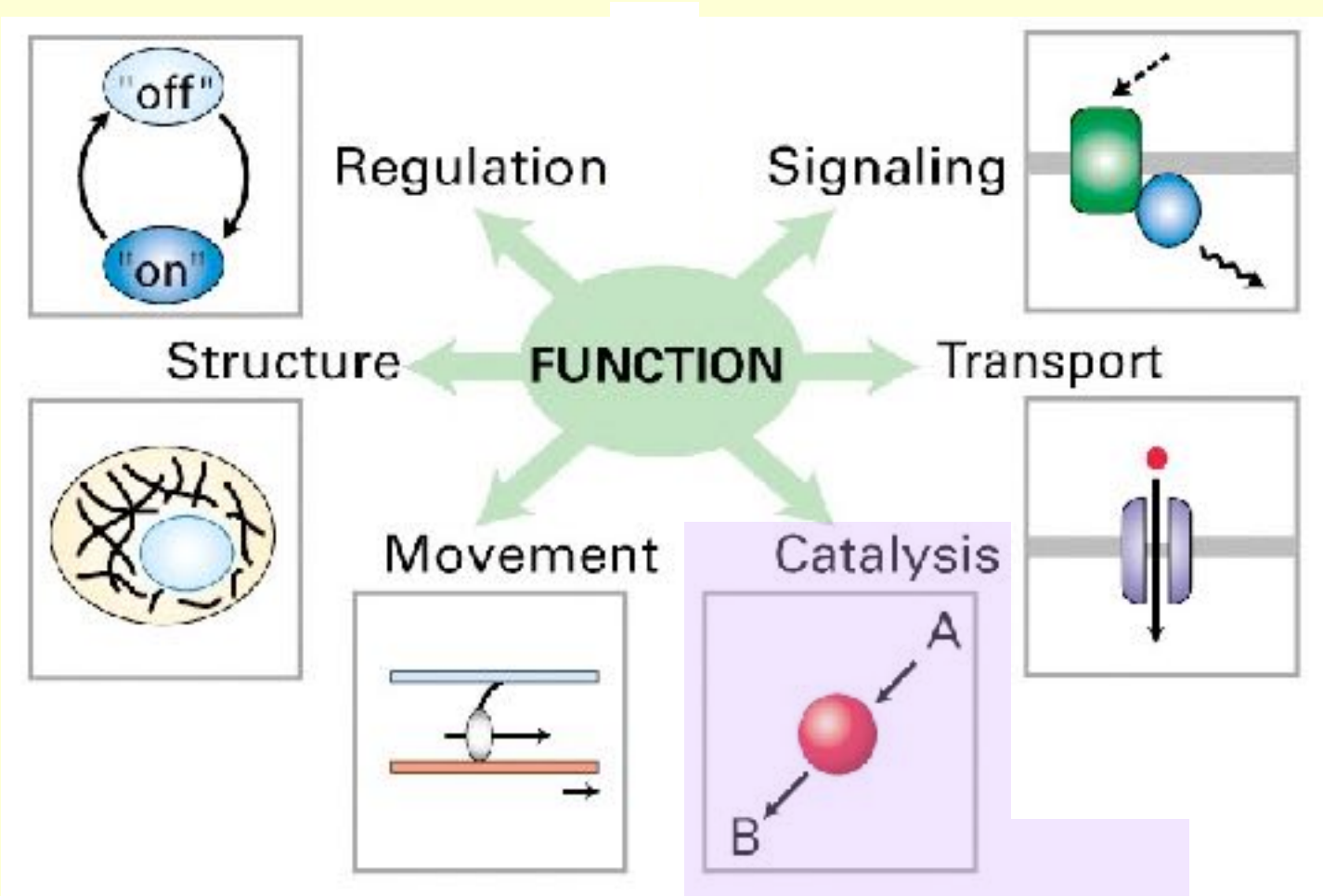


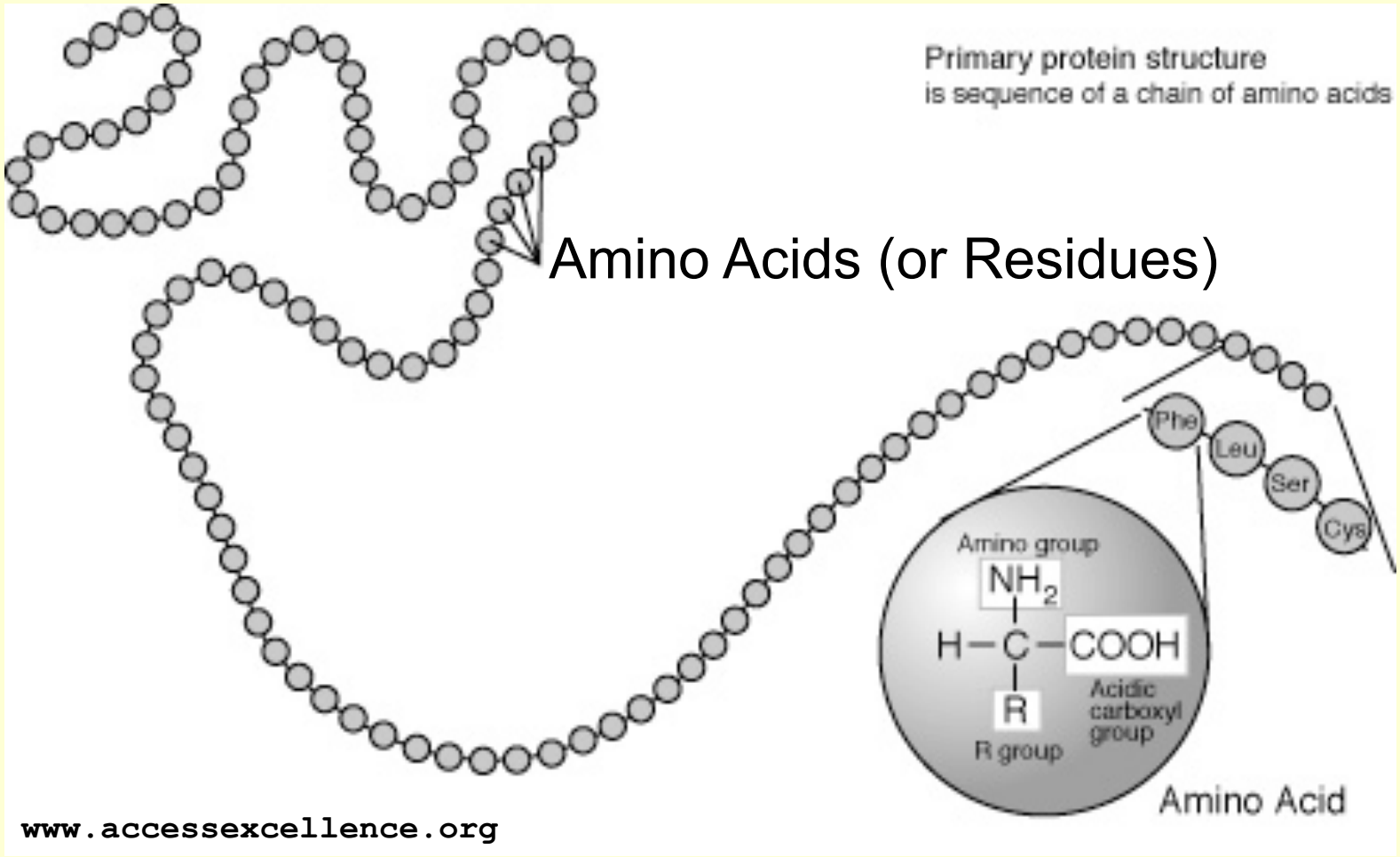
Figure 2-26 Essential Cell Biology, 2/e. (© 2004 Garland Science)

Protein Functions

- Activity
- Specificity
- Regulation

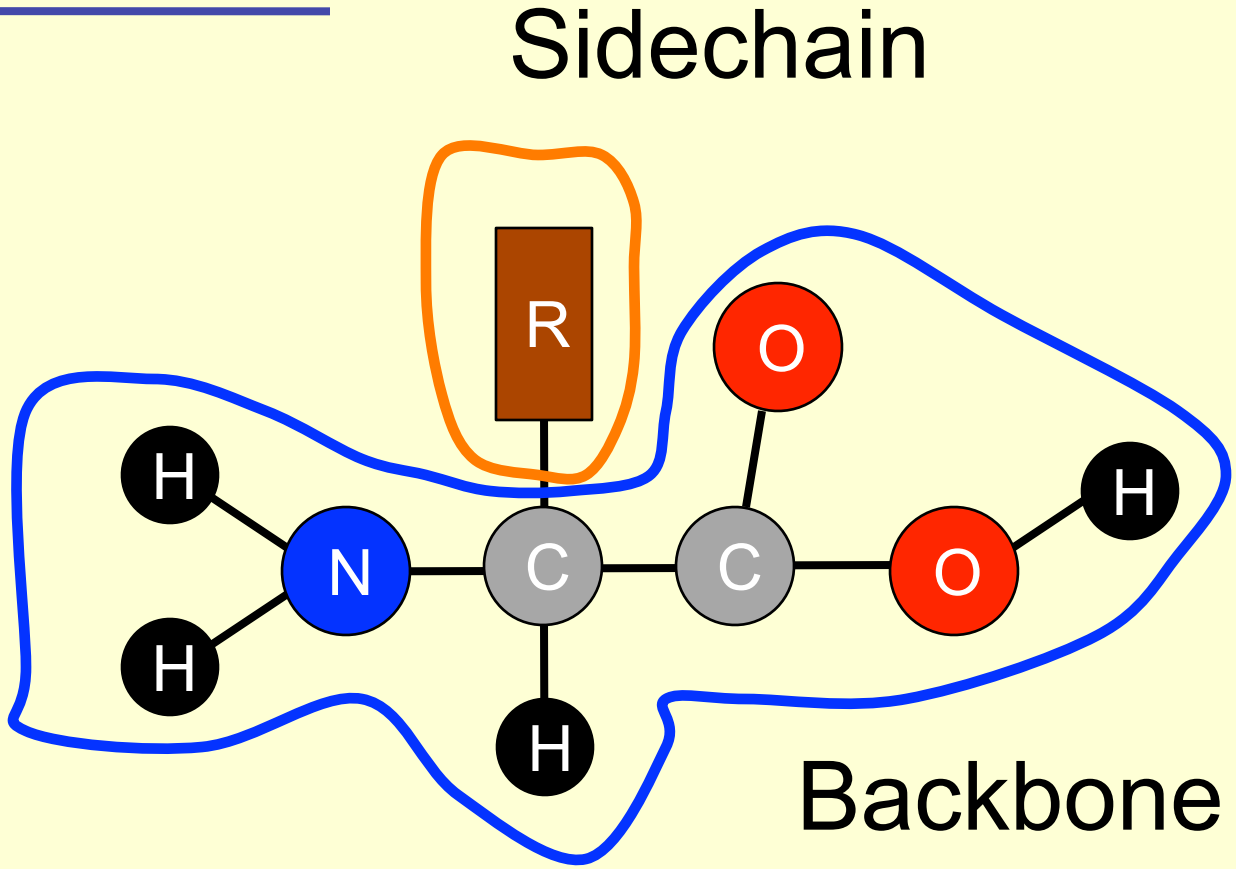


Enzymes

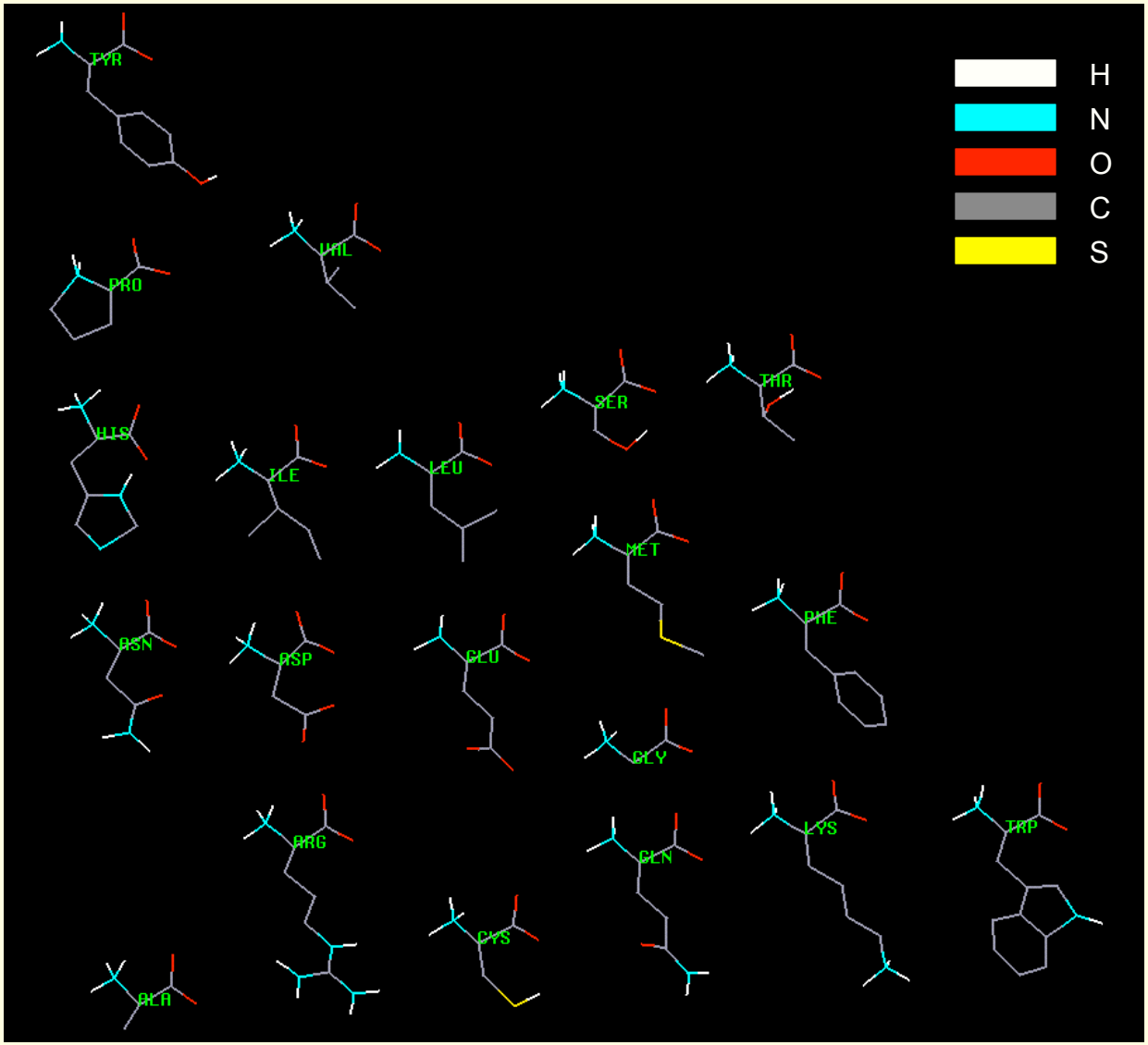


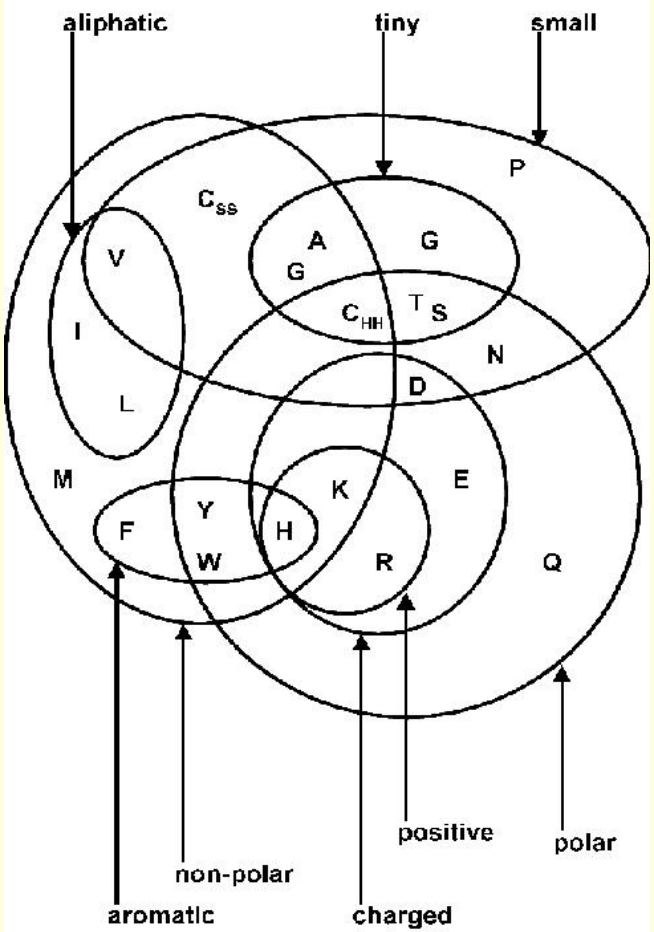
www.accessexcellence.org

Amino Acid



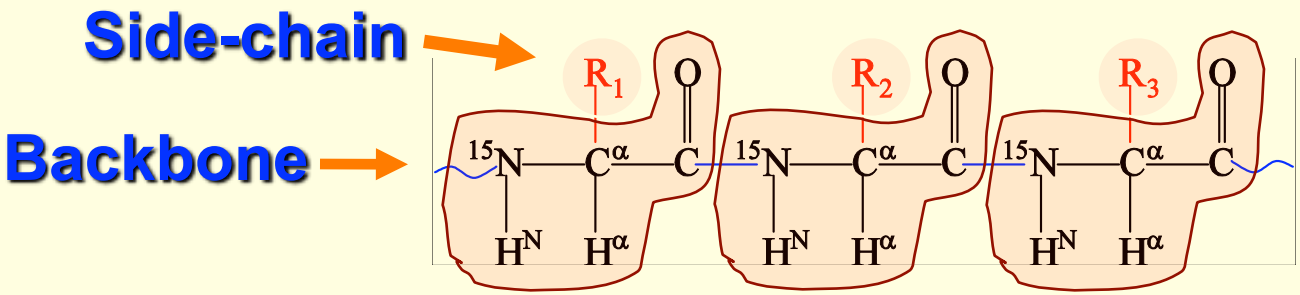
Amino Acids


TYR
PRO
VAL
HIS
ILE
LEU
SER
THR
MET
ASN
ASP
GLU
PHE
GLY
ARG
GLN
LYS
TRP
CYS
ALA
H
N
O
C
S


aliphatic
tiny
small
P
C_{SS}
V
A
G
G
C_{HH}
T
S
I
N
D
L
M
K
E
Y
F
H
W
R
Q
positive
polar
non-polar
aromatic
charged

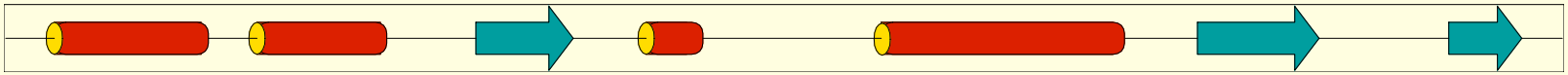
Protein Structure

Primary Sequence: Linear String of Amino Acids

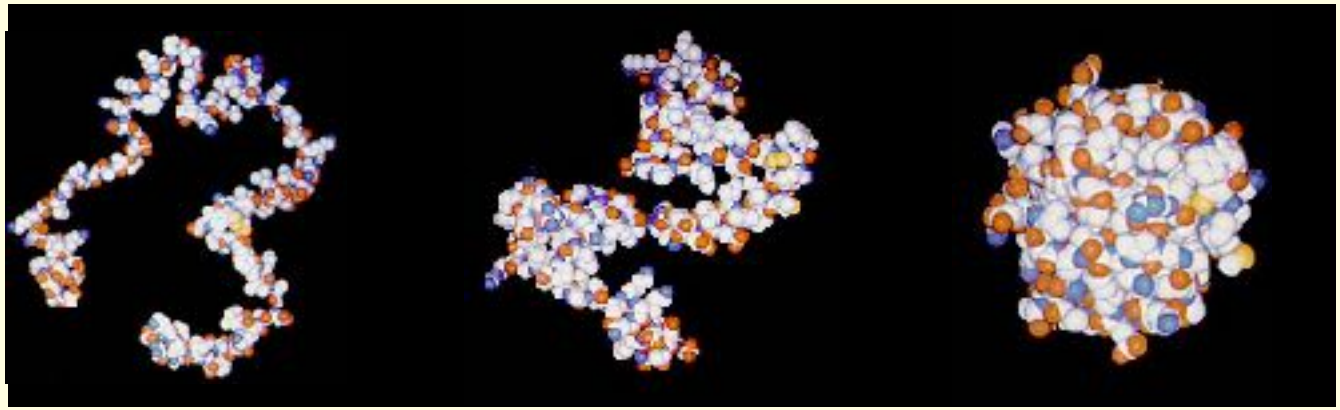


... ALA PHE LEU ILE LEU ARG ...

Secondary structure: regular α-helices and β-strands



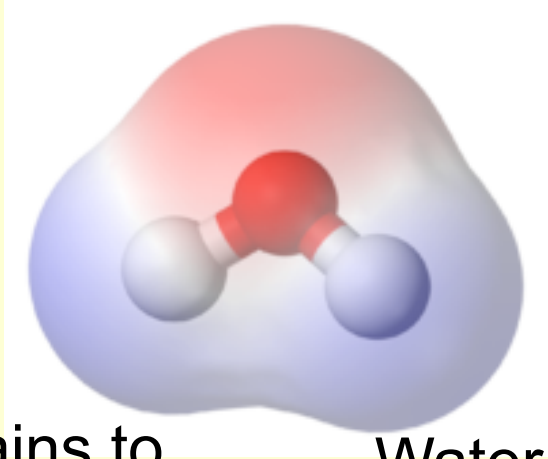
Global Fold



Protein Scale and Size

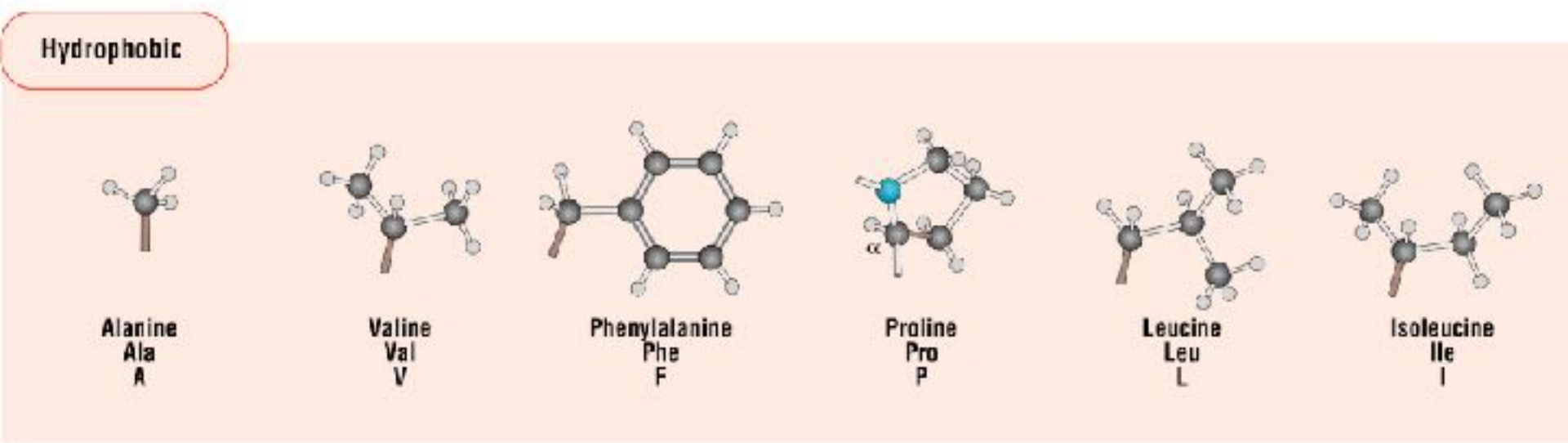
- Size is measured in Daltons (Da). An average residue is ~135 Da.
- Interatomic distances are measured in Angstroms (Å), 1×10^{-10} m.
- Concentration is measured in mol/L (1 mol = 6.022×10^{23}).
- Proteins fold to “native state” in microseconds to seconds.

Hydrophobic Residues



Water

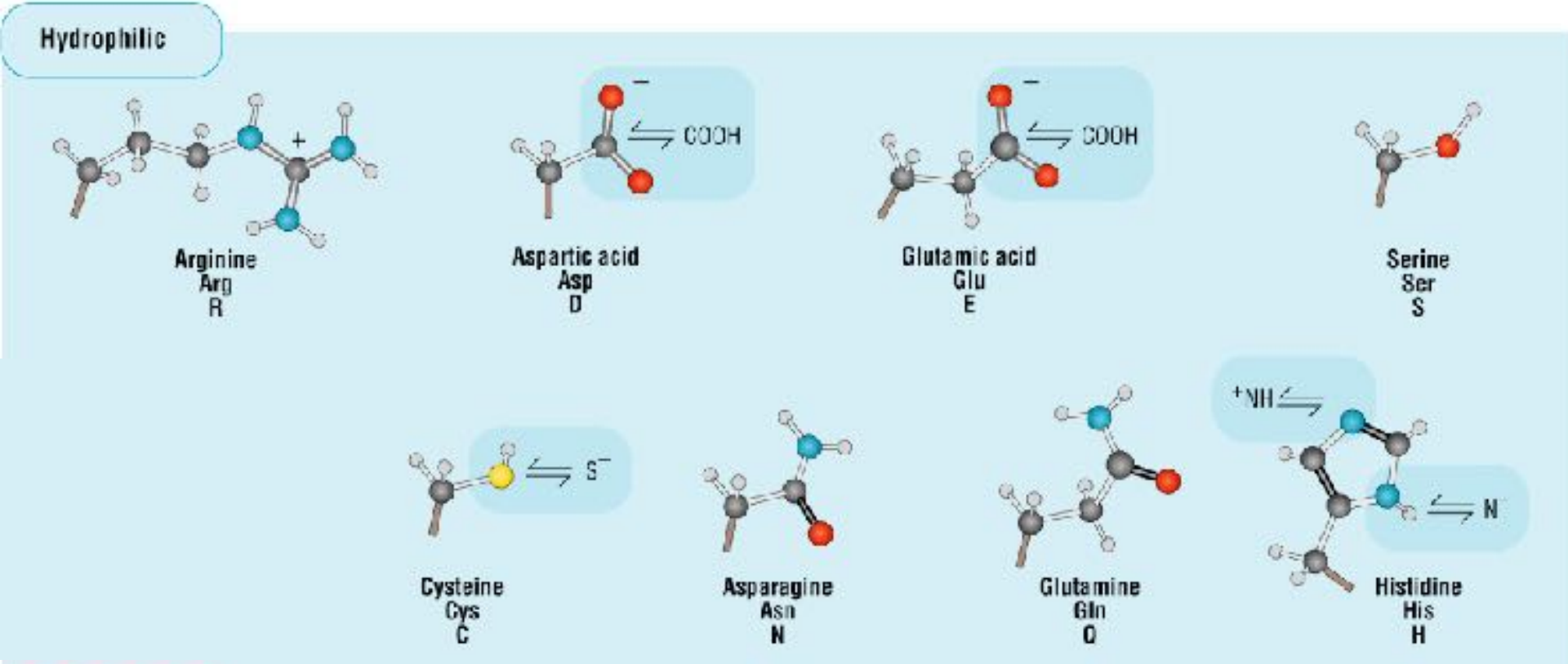
- Nonpolar and uncharged
- Tend to avoid water
- Tend to interact with other nonpolar sidechains to minimize contact with water
- Tend to be buried in the protein core



Petsko, Ringe, Prot Struct and Function, 2004

Hydrophilic Residues

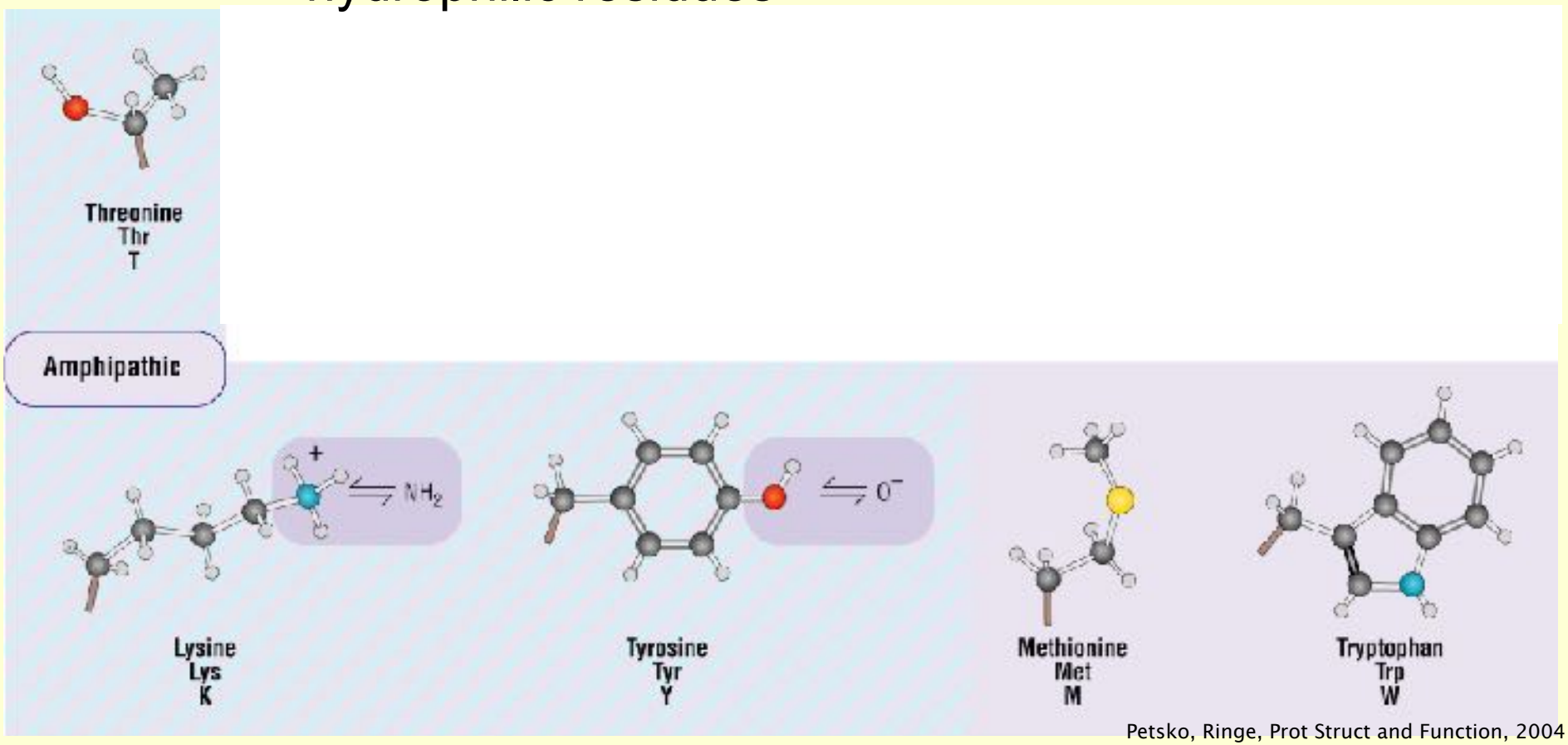
- Polar or charged
- Tend to interact with water or other hydrophilic sidechains



Petsko, Ringe, Prot Struct and Function, 2004

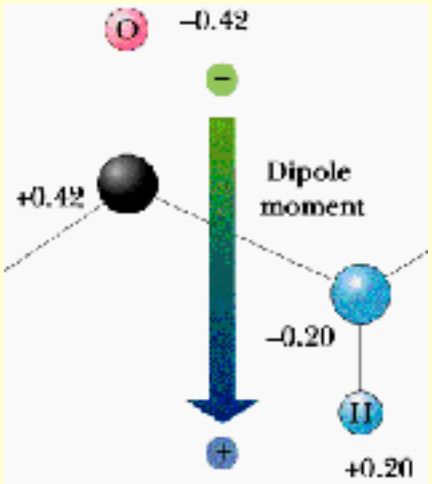
Amphipathic Residues

- Have *both polar and nonpolar* characteristics
- Tend to form interfaces between hydrophobic and hydrophilic residues

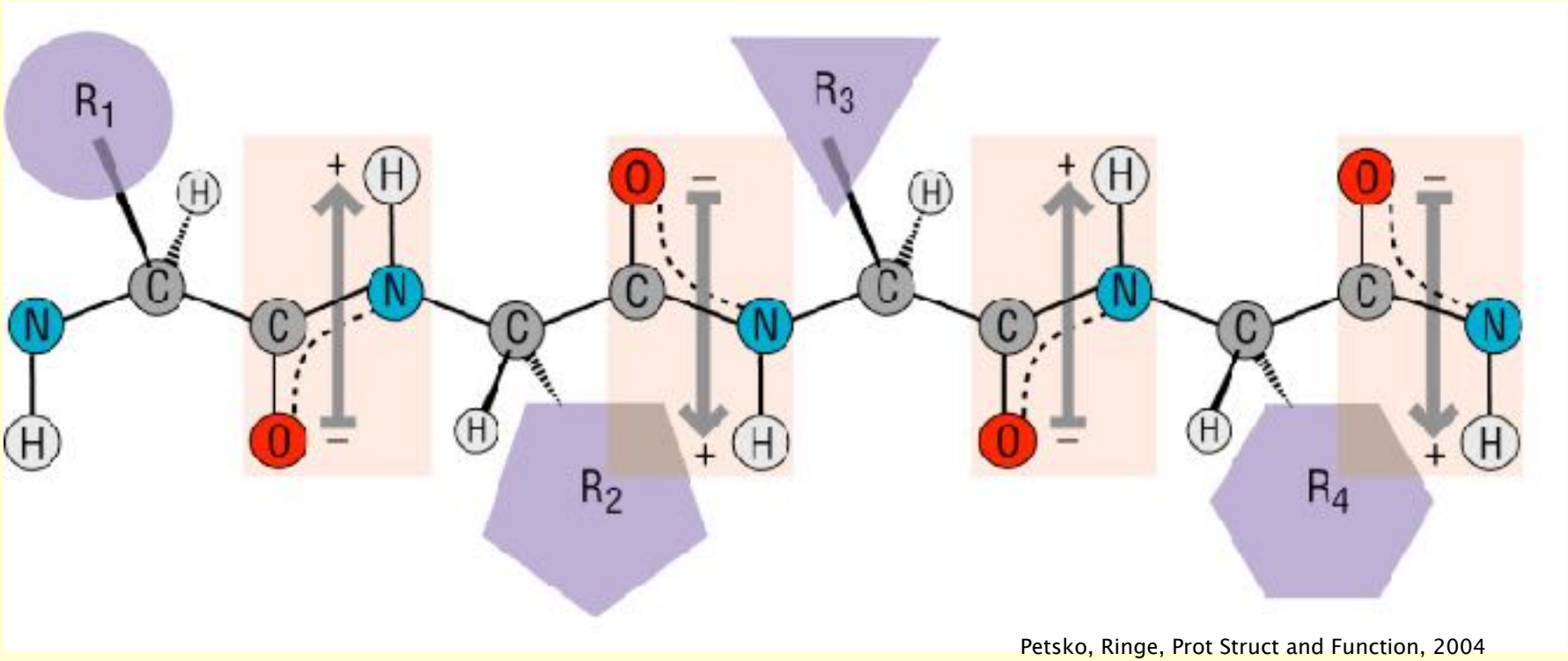


Petsko, Ringe, Prot Struct and Function, 2004

Peptide Backbone



- Peptide Backbone is polar (dipolar)
- Double bonds are not very flexible, but single bonds....



Petsko, Ringe, Prot Struct and Function, 2004

Peptide Backbone

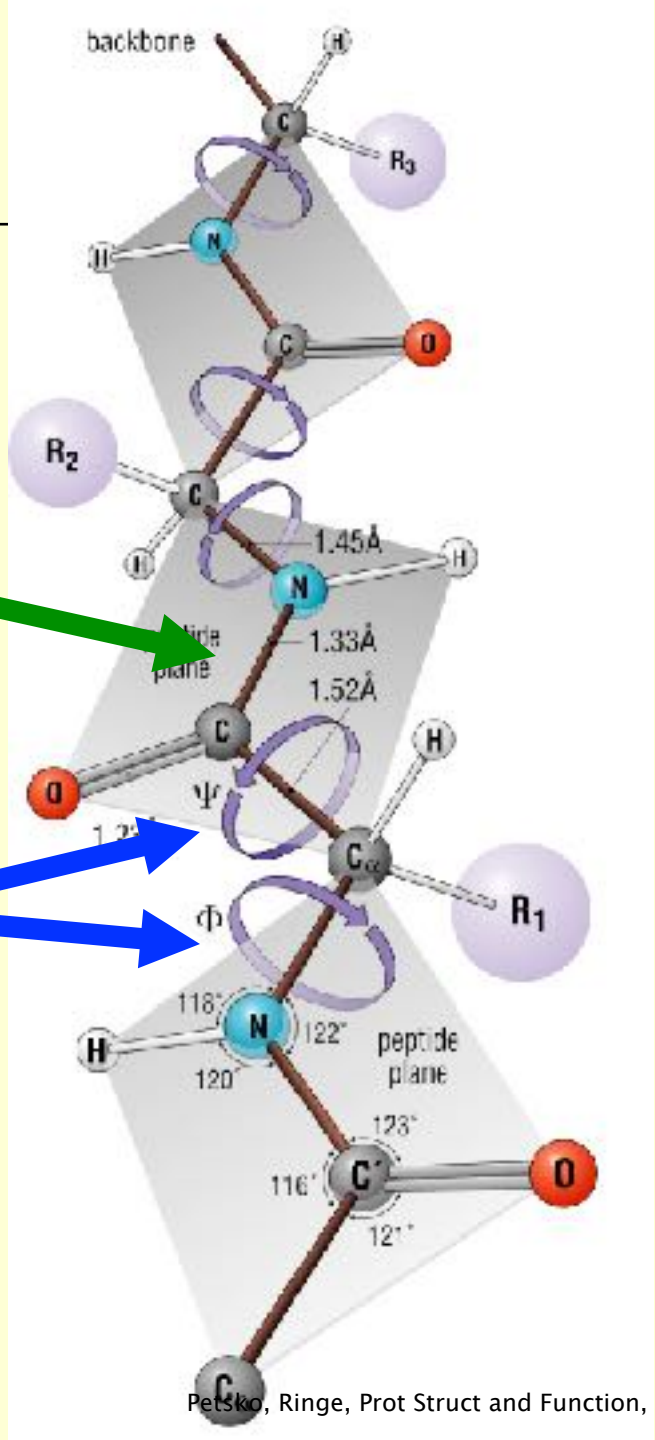
Peptide bond has partial double bond character and is rigid

(ie. it doesn't rotate)

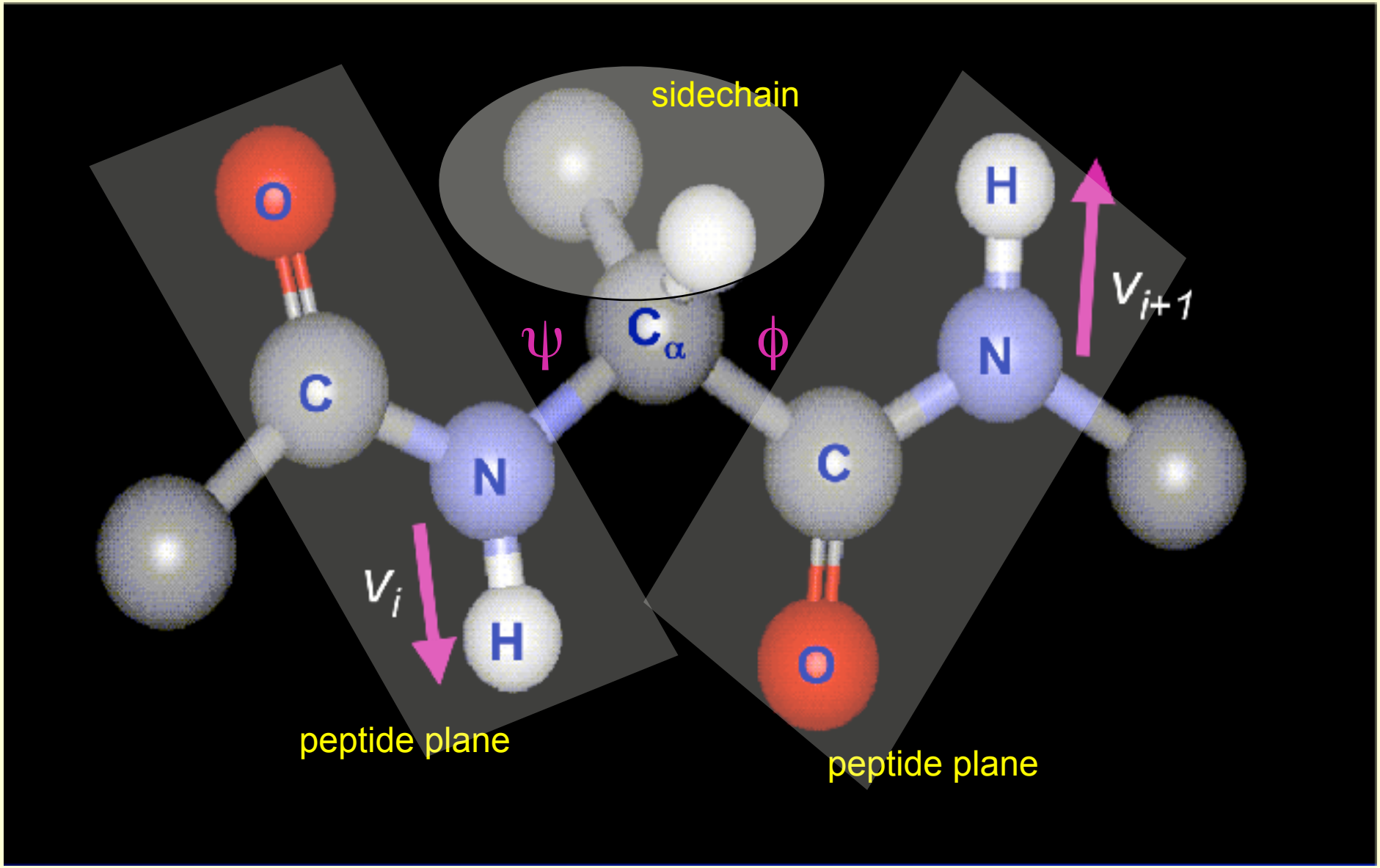
Other backbone bonds are flexible

psi torsion angle

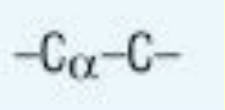
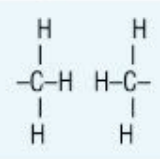
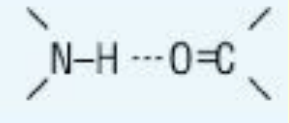
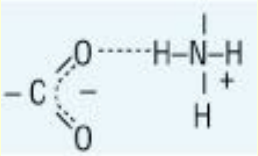
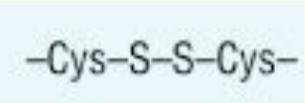
phi torsion angle

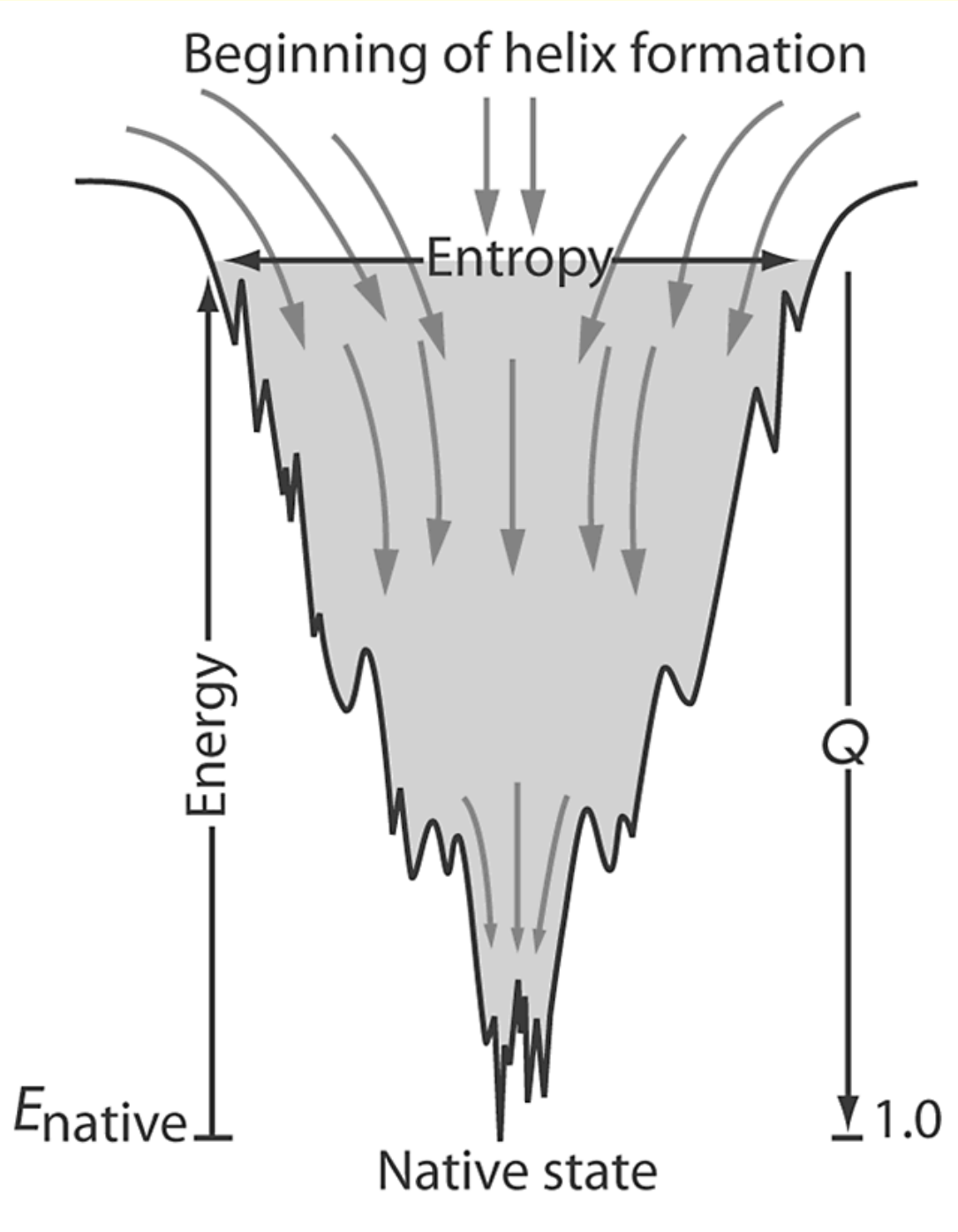


Petsko, Ringe, Prot Struct and Function, 2004


sidechain
O
C
N
H
ψ
Cα
ϕ
C
O
N
H
V_i
V_{i+1}
peptide plane
peptide plane

Stabilizing Forces

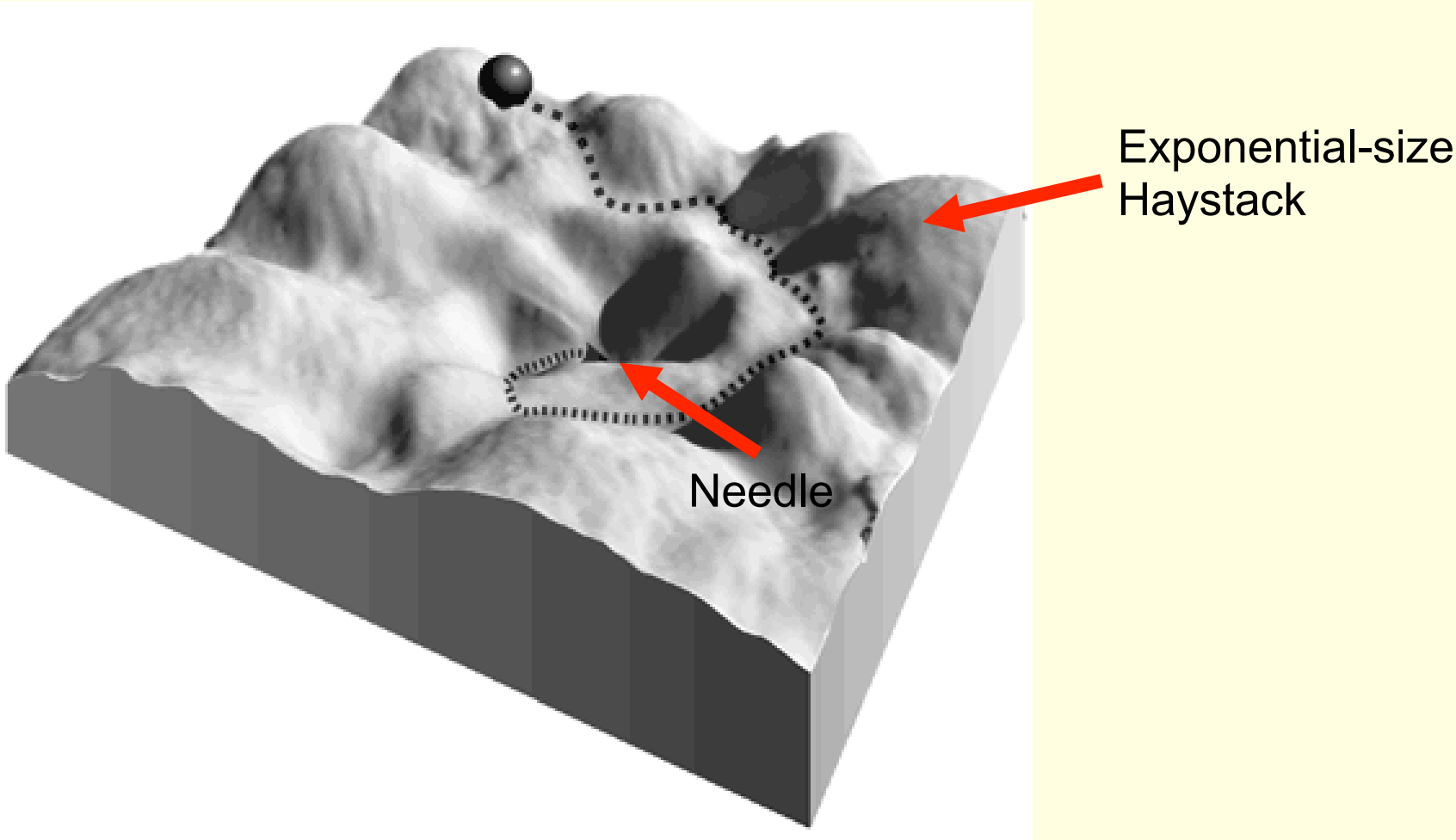
Interaction	Typical Distance	Free Energy
*Covalent Bond 	1.5 A	356 kJ/mol
*Van der Waals Interaction 	3.5 A	4-15 kJ/mol
*Hydrogen Bond 	3.0 A	2-6 kJ/mol Up to 20 kJ/mol if one atom is charged
Salt Bridge 	2.8 A	13-17 kJ/mol
Disulfide Bond 	2.2 A	167 kJ/mol
*Long Range Electrostatic	variable	variable



Protein folding is a physical/chemical process.

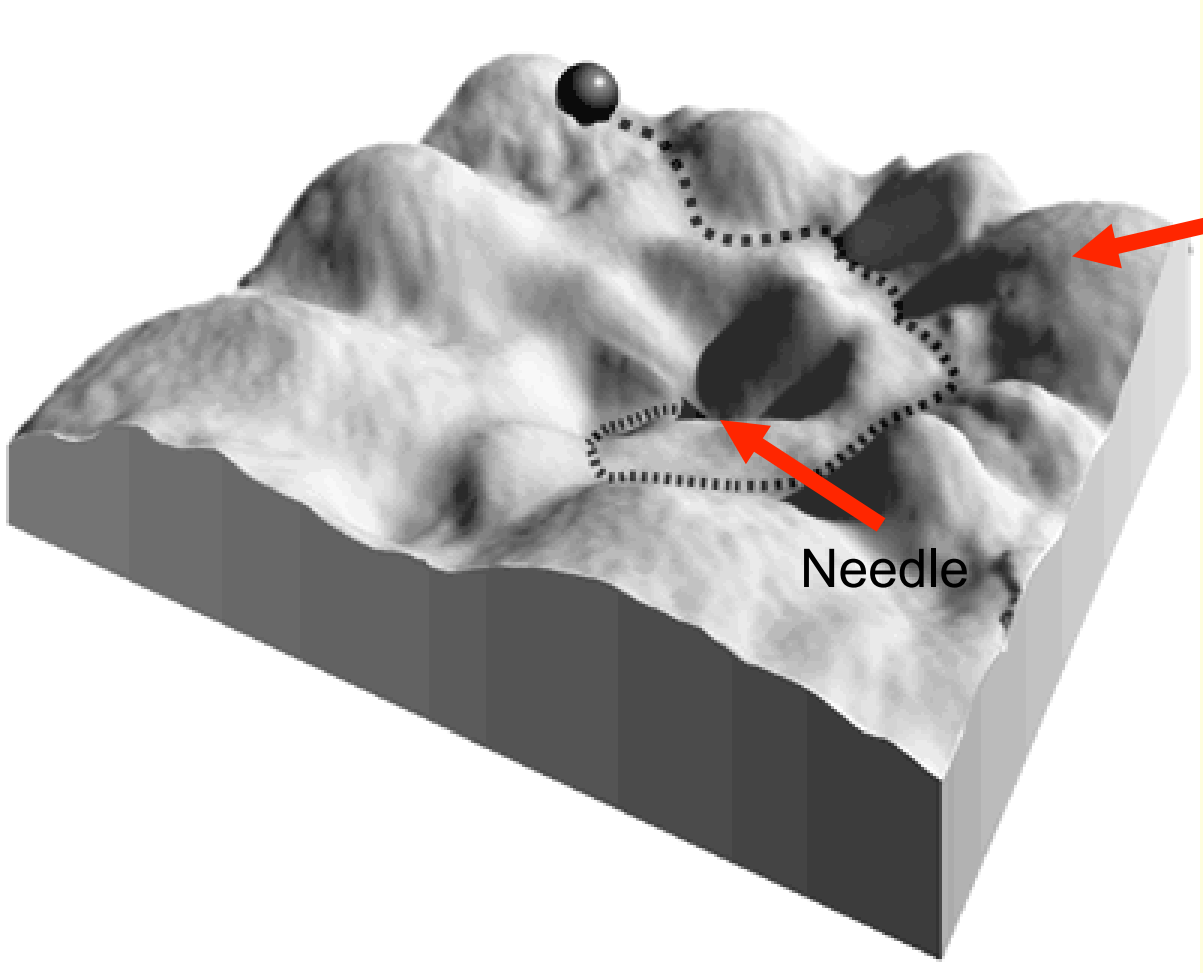
Thesis: Proteins adopt low energy conformations in their native state.

Conformational energy is determined by solvent (water), van der Waals forces, charge, electrostatic forces, etc.



Levinthal's Paradox (1969): The conformation space is exponential in the number of amino acids, but the folding pathway must be relatively short.

De novo Structure Prediction



Does this landscape have a more compact representation?
Can we quickly find low-energy conformations?

Backbone Flexibility

- Protein with n residues has $2n$ rotatable backbone dihedrals
- $2n$ degrees-of-freedom (DOF)
- Number of protein conformations is exponential in its length (consider kinematic chain)

Large Search Spaces!

www.active-robots.com

How Many Possible Proteins Are There?

Sequences

Avg Protein 300 Amino Acids (AAs) long

$$20^{300} \approx 10^{390}$$

There are only 10^{80} particles estimated in the universe

Human Proteins

Current Estimate ~$20{,}000$ Genes *conservative

Each gene has alternate splicing (say 2 / protein)

Each protein can be post-synthetically modified (say 2 / protein)

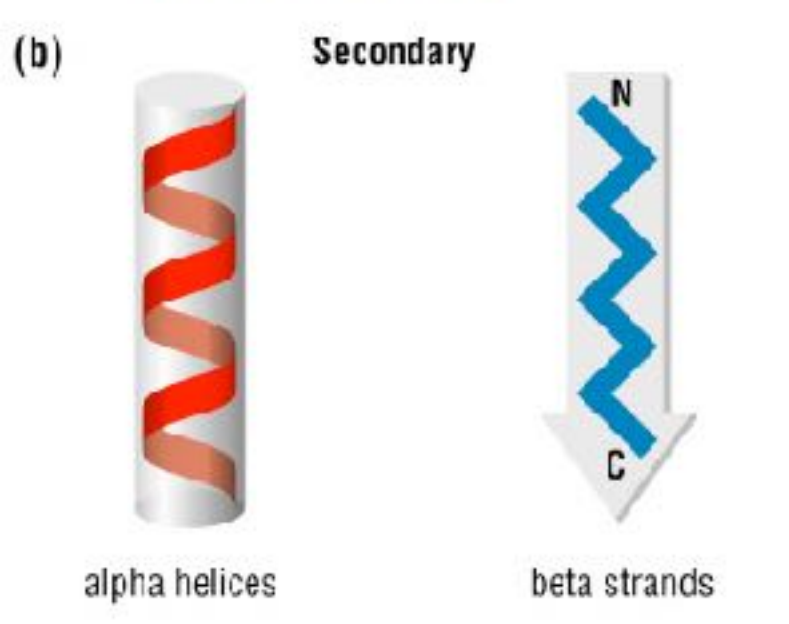
cleavage (insulin), phosphorlyated (kinase), glycosylated (sugars), ...

SO

$20{,}000 \times 2 \times 2 = 80{,}000$

Levels of Protein Structure

GLY	SER	MET	SER
GLY	ILE	ALA	LEU
SER	ARG	LEU	ALA
GLU	GLY	GLY	LEU
PHE	LYS	LEU	ARG
MET	LEU	LEU	ASN
VAL	GLU	TYR	GLU
LYS	ARG	VAL	ARG
ALA	GLN	ALA	LYS



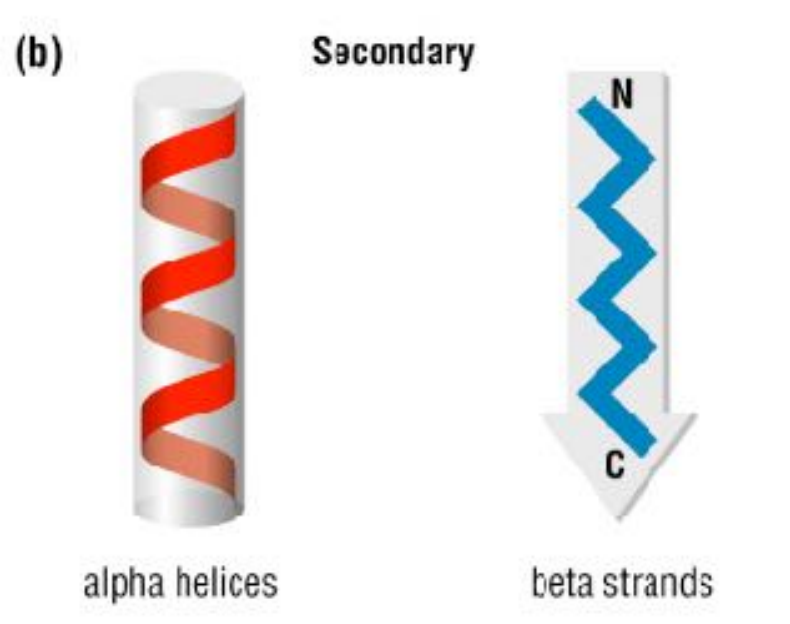
Petsko, Ringe, Prot Struct and Function, 2004

Primary Structure

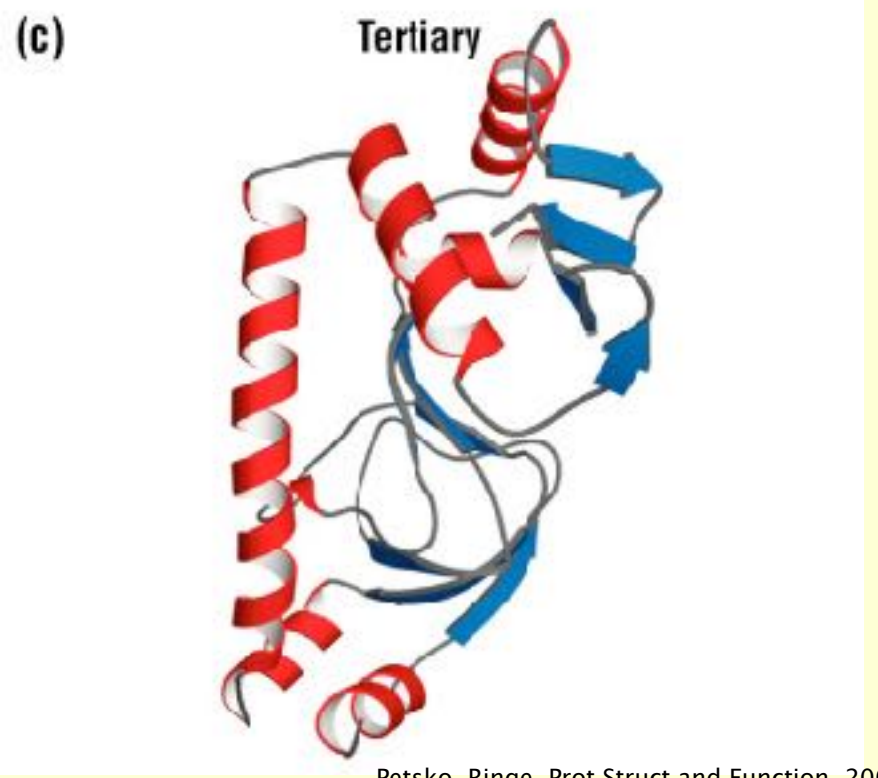
Aka Primary Sequence

Secondary Structure

Levels of Protein Structure



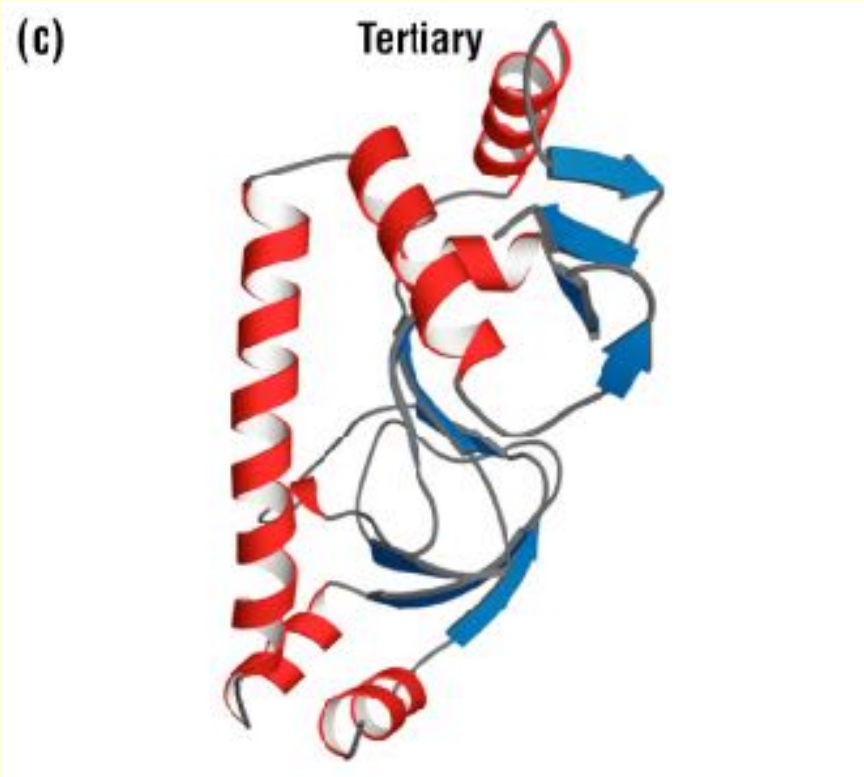
Secondary Structure

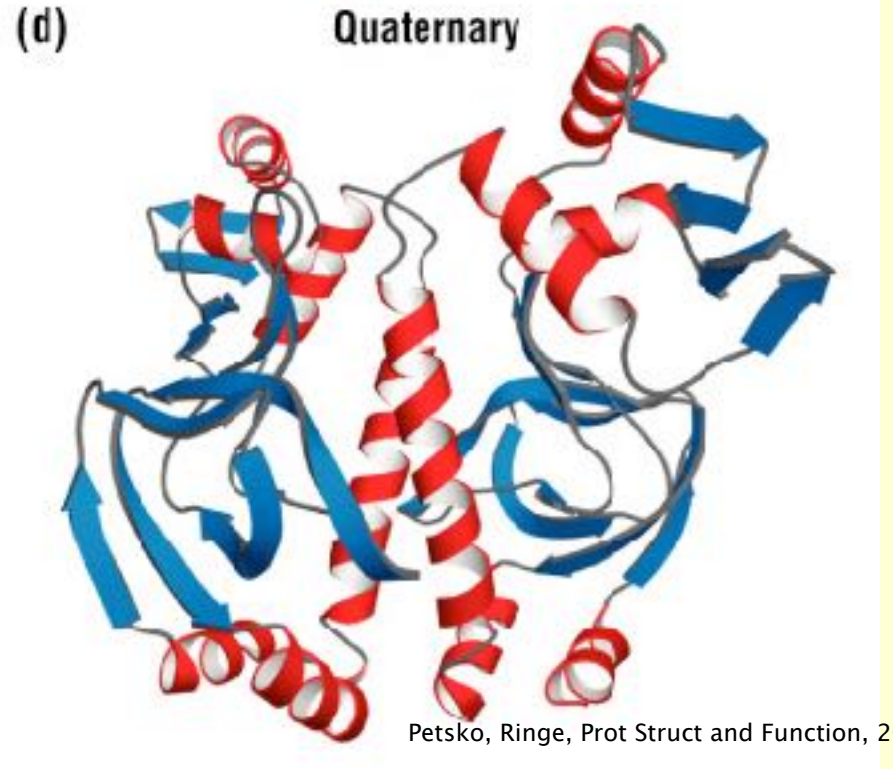


Petsko, Ringe, Prot Struct and Function, 2004

Tertiary Structure

Levels of Protein Structure



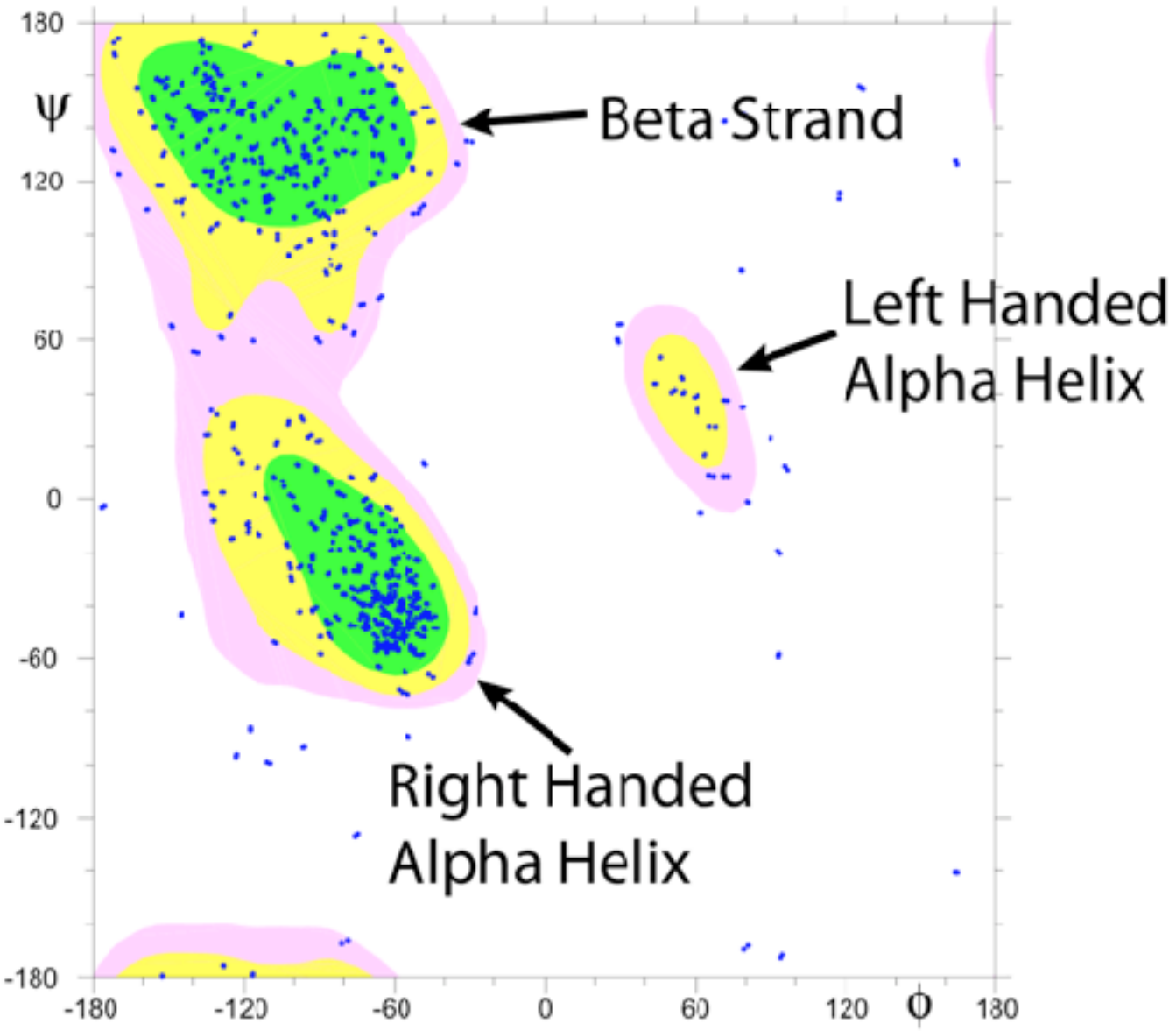


Petsko, Ringe, Prot Struct and Function, 2004

Tertiary Structure

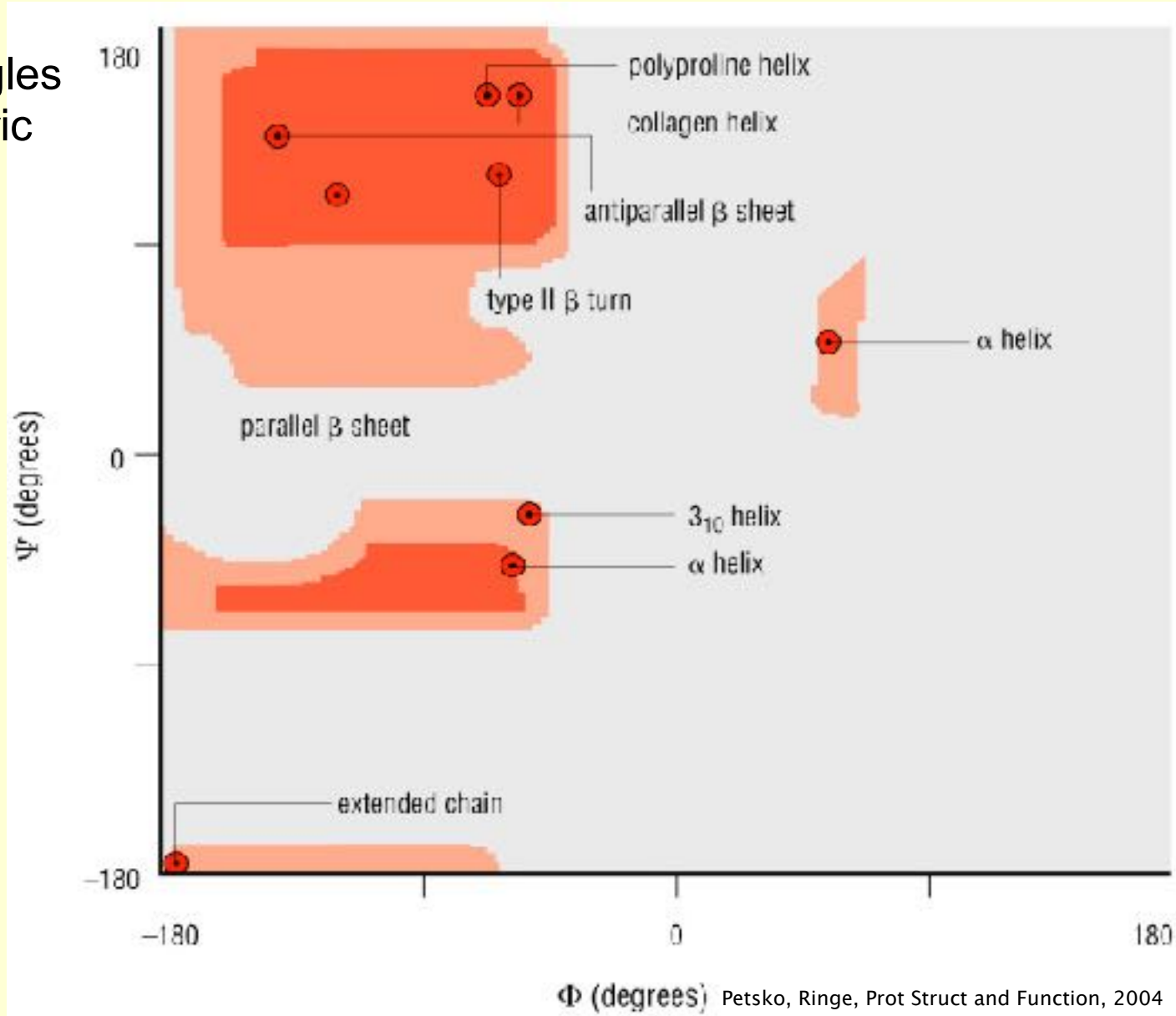
Quaternary Structure

Phi/Psi Histogram



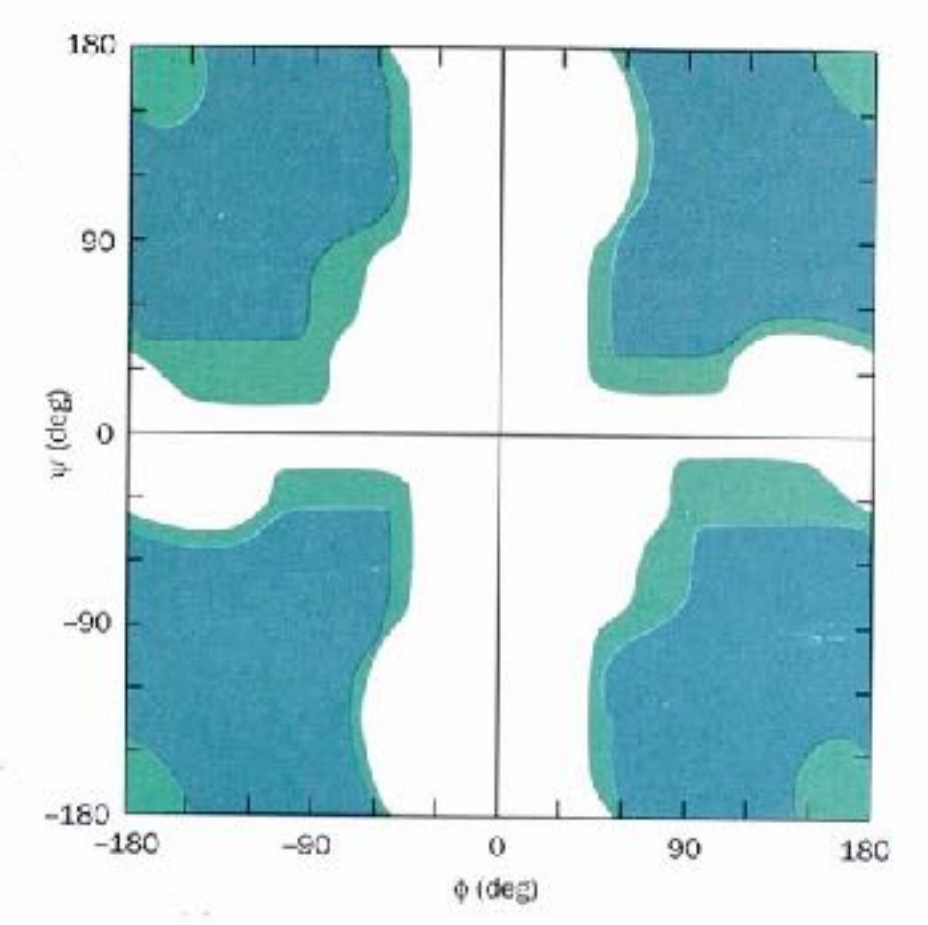
Ramachandran Plot

Allowed phi/psi angles do not result in steric interference

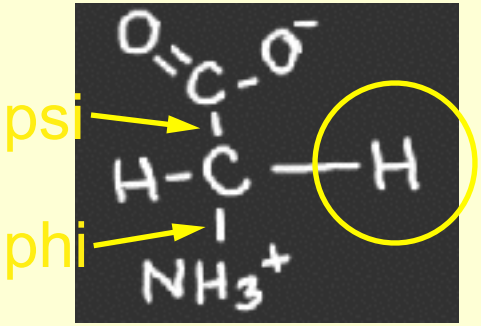


Petsko, Ringe, Prot Struct and Function, 2004

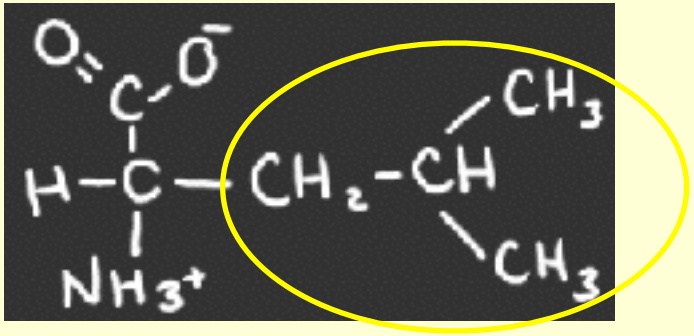
Phi / Psi Histogram for Glycine



Glycine

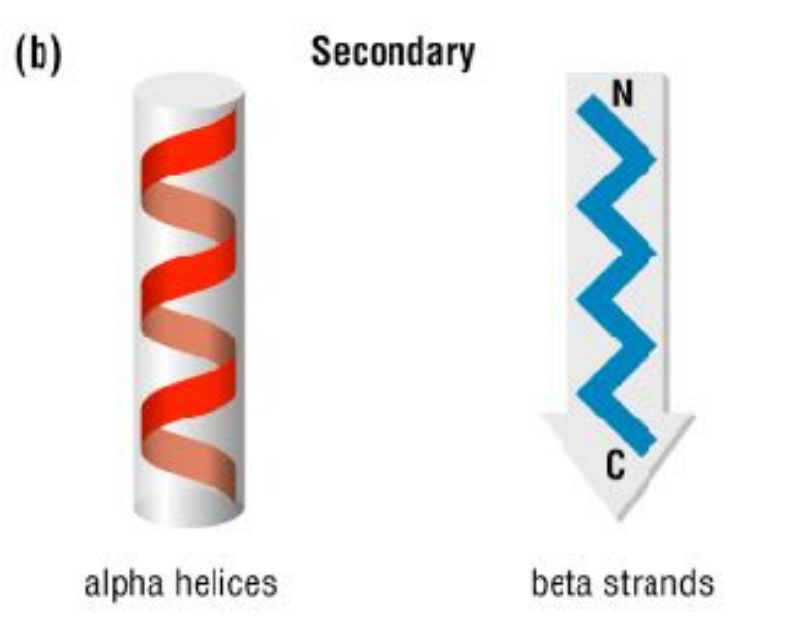


Leucine



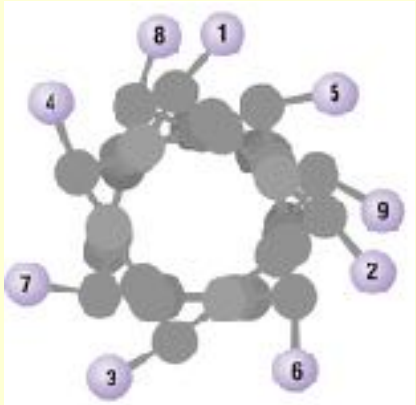
Secondary Structure

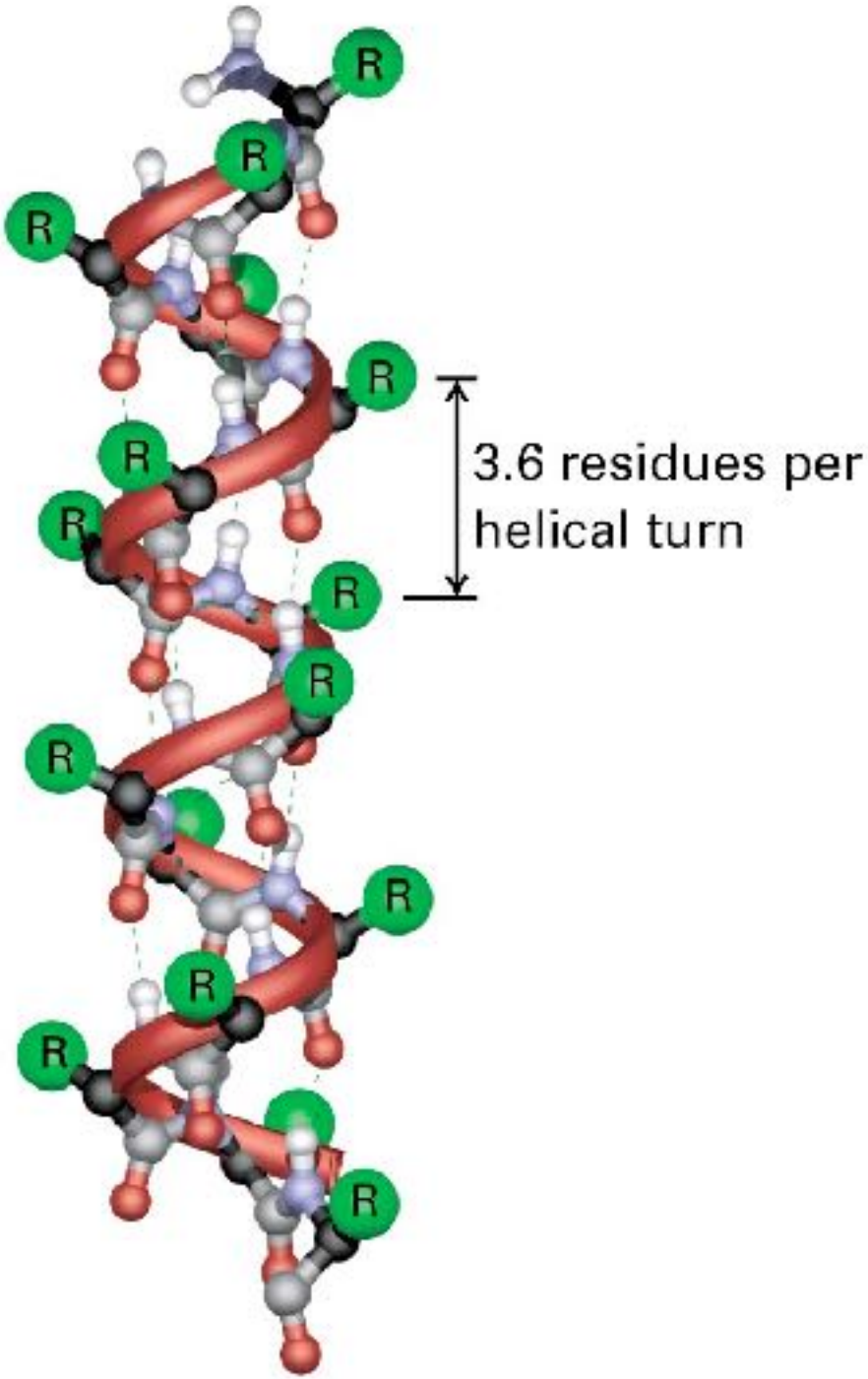
No true gold standard definition
May be dynamic - change with changing protein state



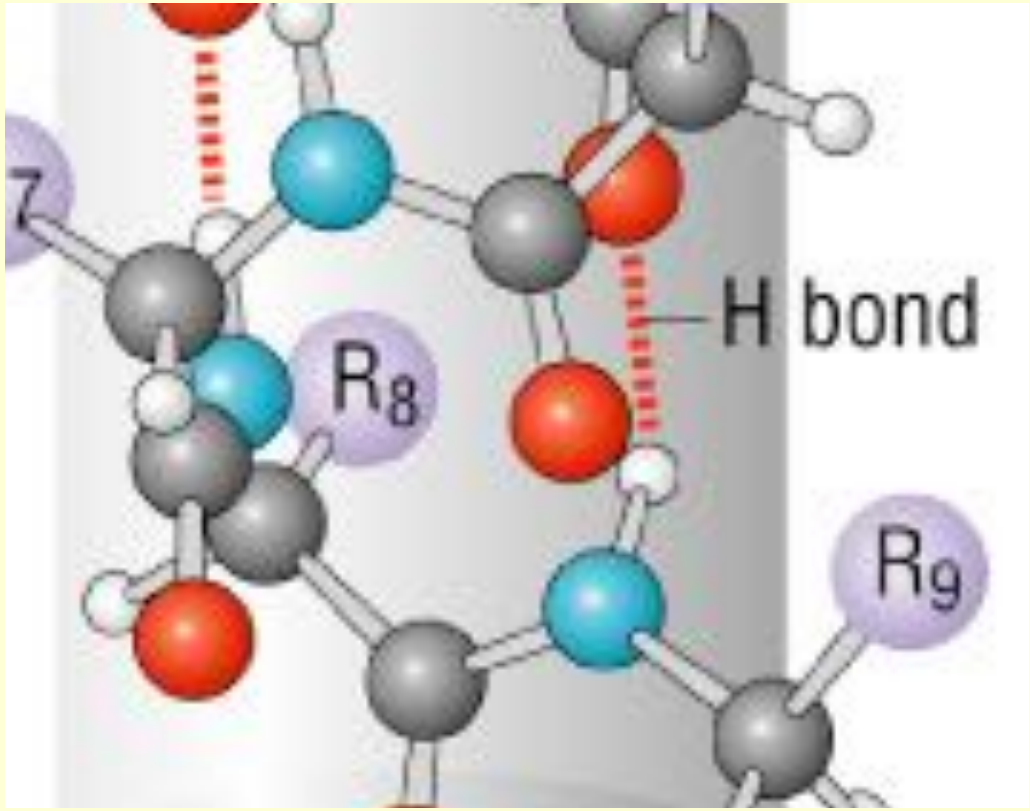
The Alpha Helix

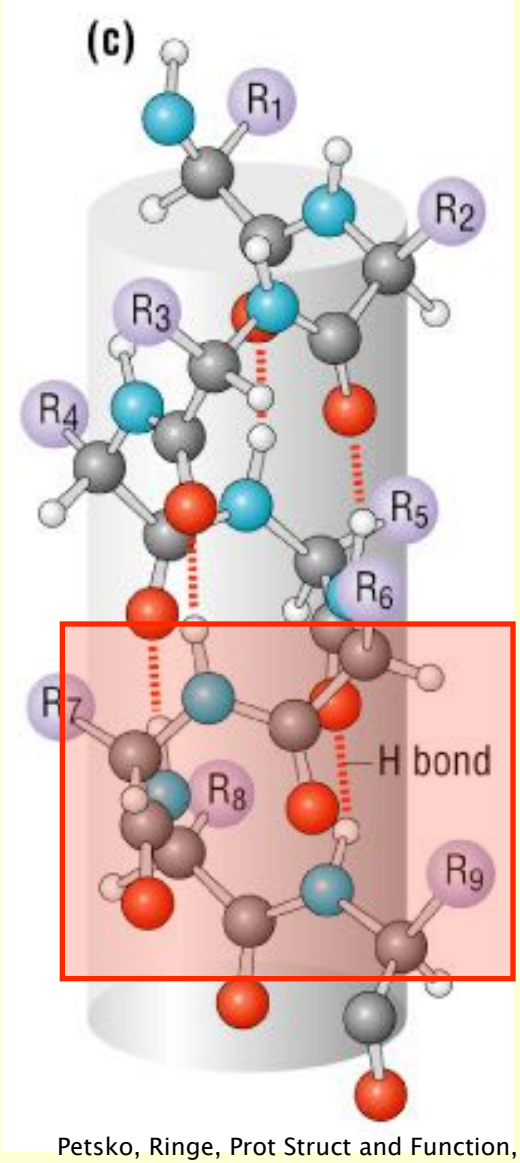
- Most common secondary structure type
- Hydrogen bonding between carbonyl oxygen atom of residue *n* and amide nitrogen of residue *n*+4.
- 3.6 Residues per turn
- Cylindrical structure with hydrogen-bonded wall and outside studded with side chains





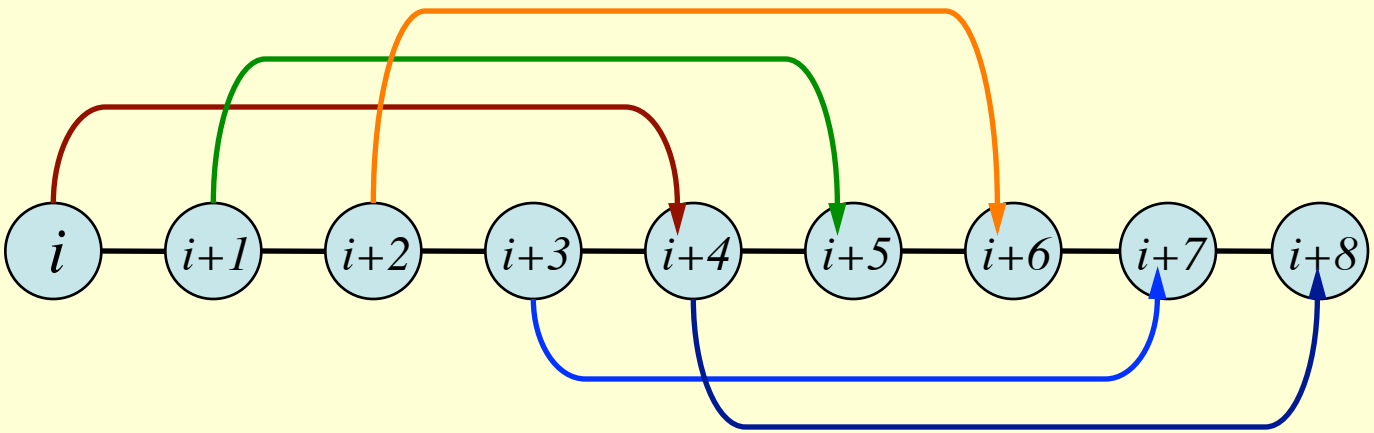
The Alpha Helix



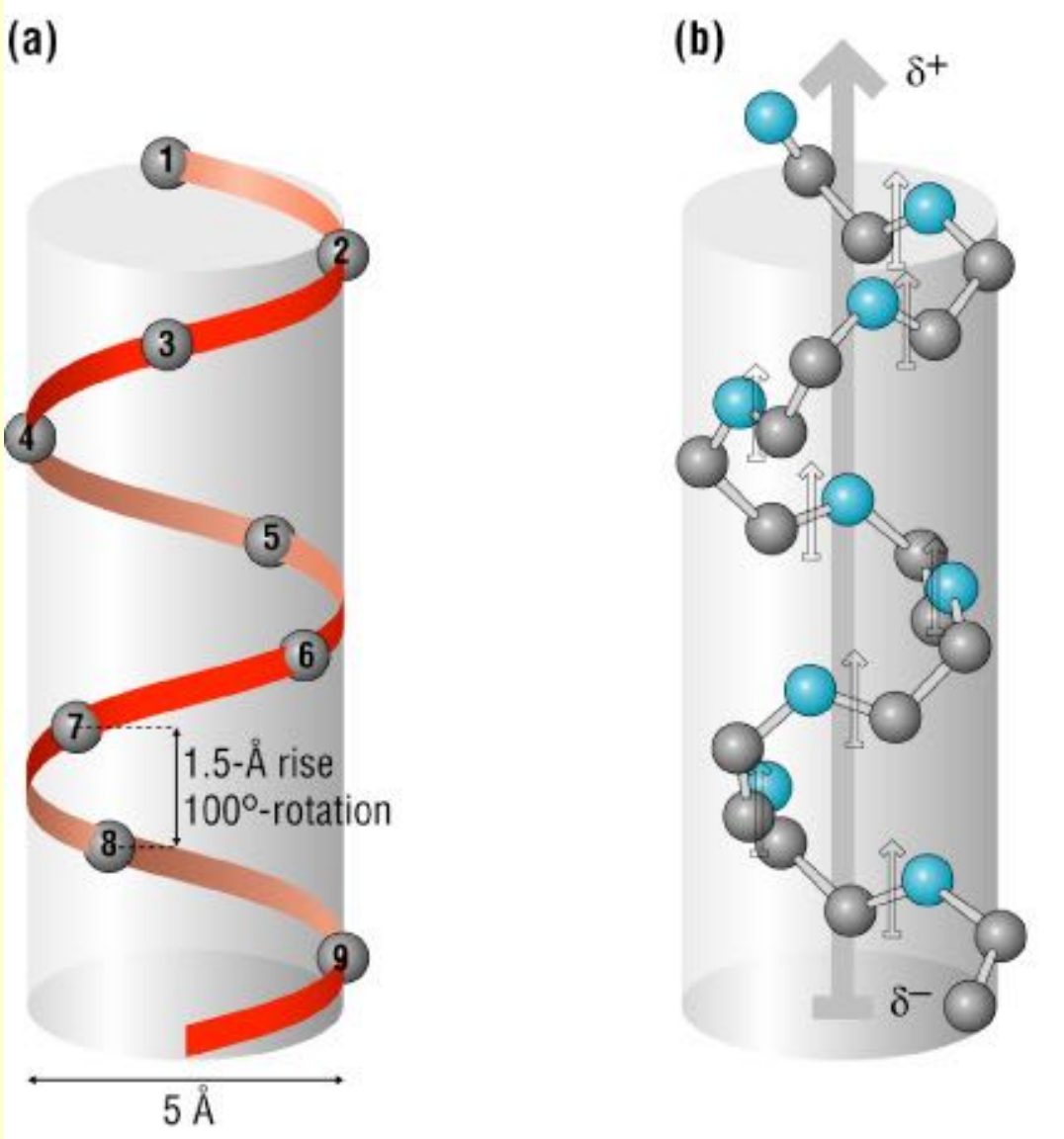


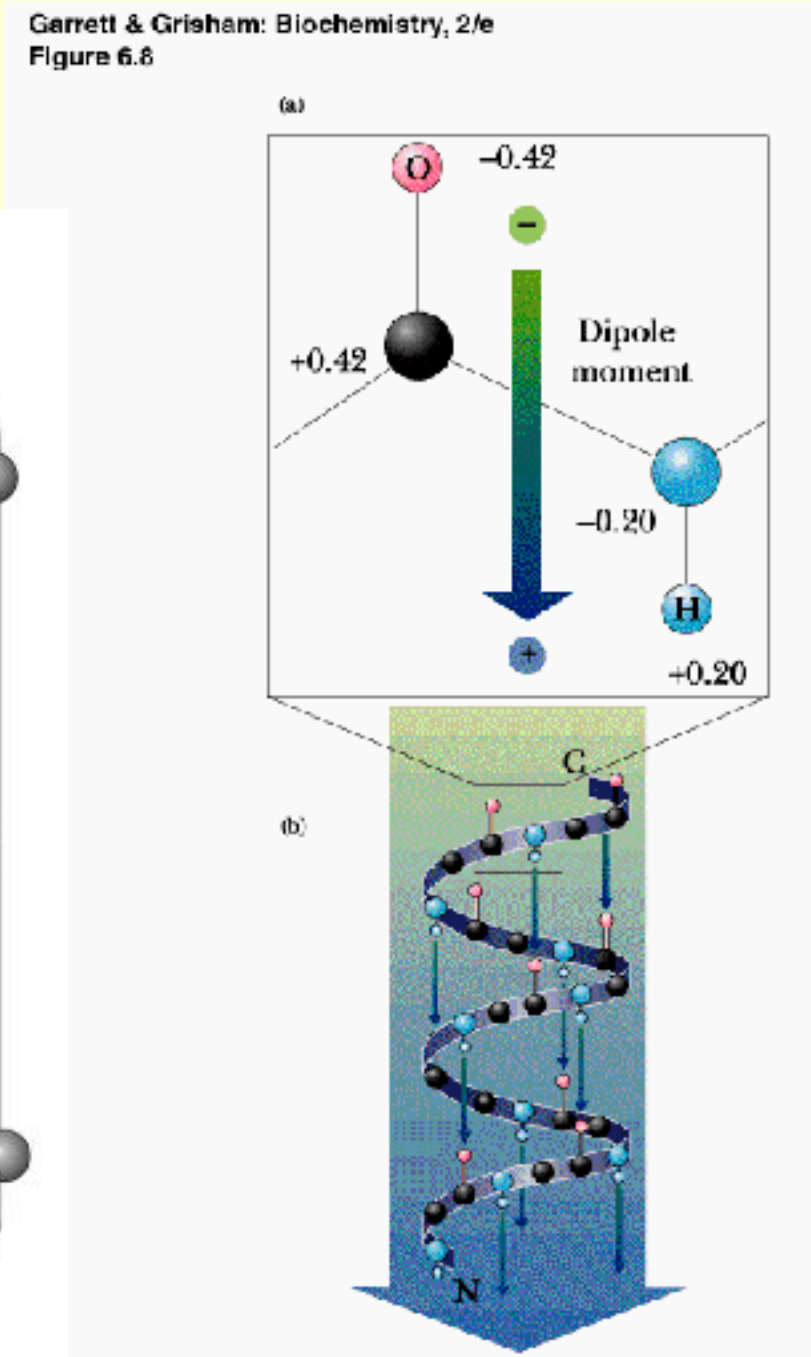
Petsko, Ringe, Prot Struct and Function, 2004

Hydrogen Bonding Pattern



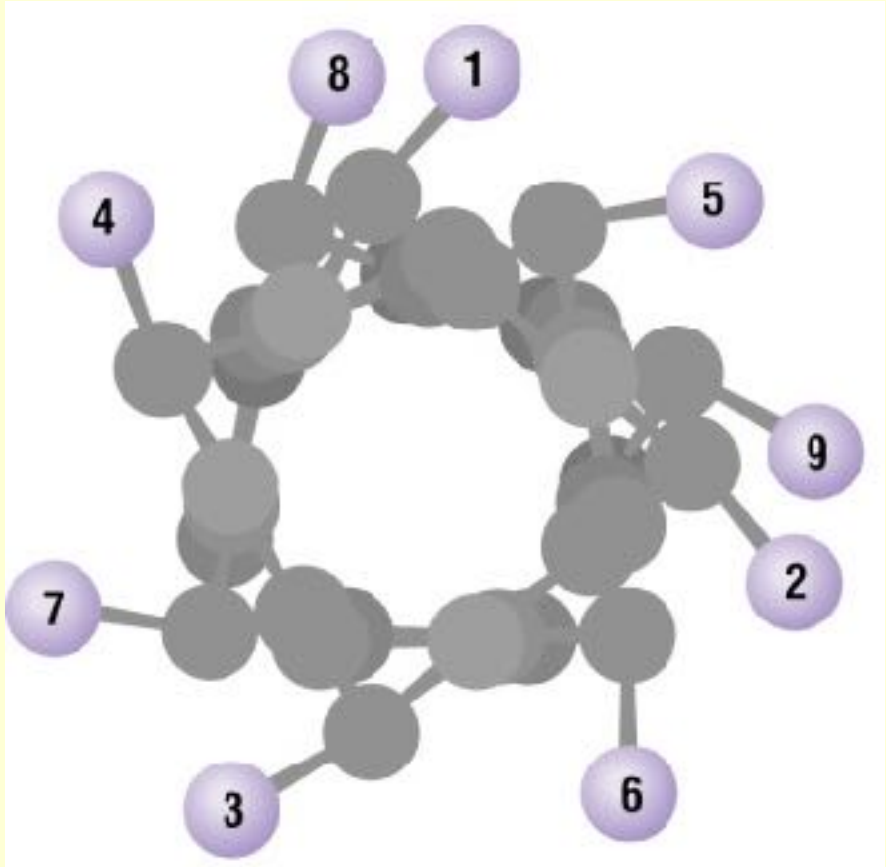
The Alpha Helix

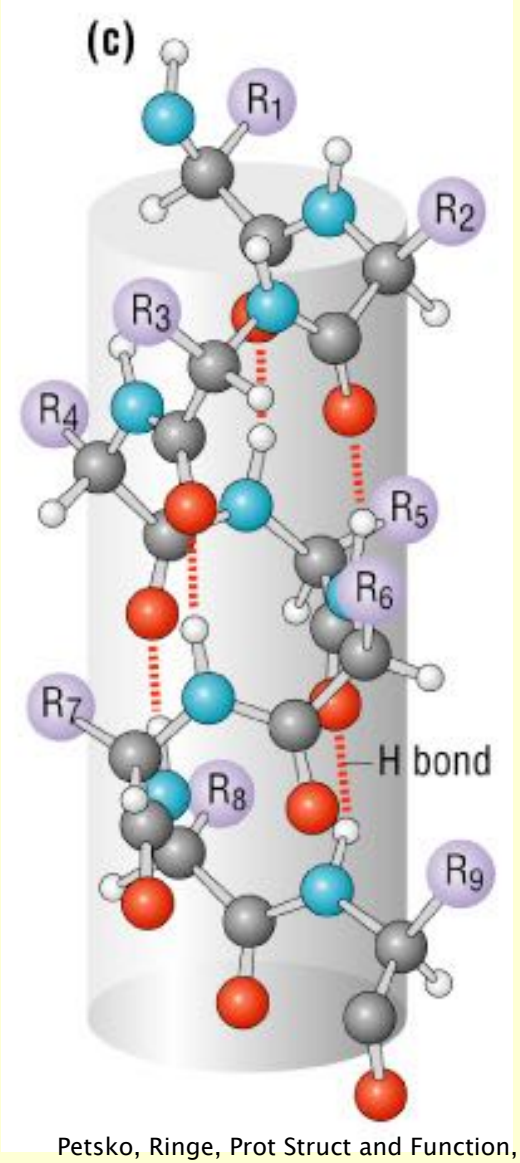




Petsko, Ringe, Prot Struct and Function, 2004

The Alpha Helix

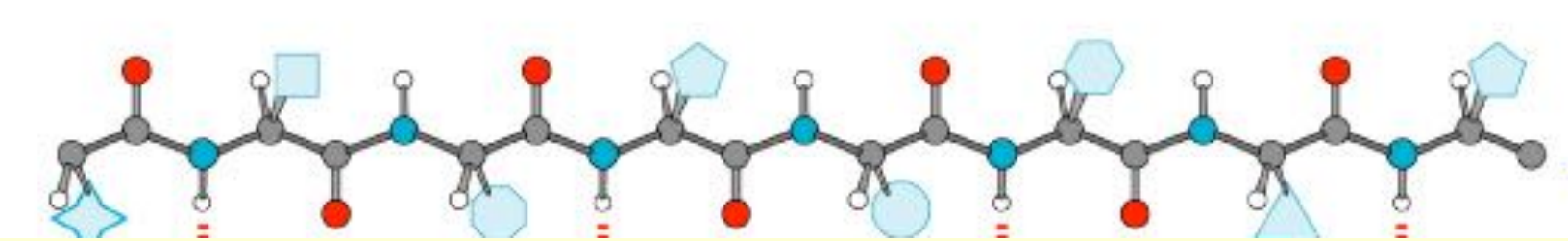




Petsko, Ringe, Prot Struct and Function, 2004

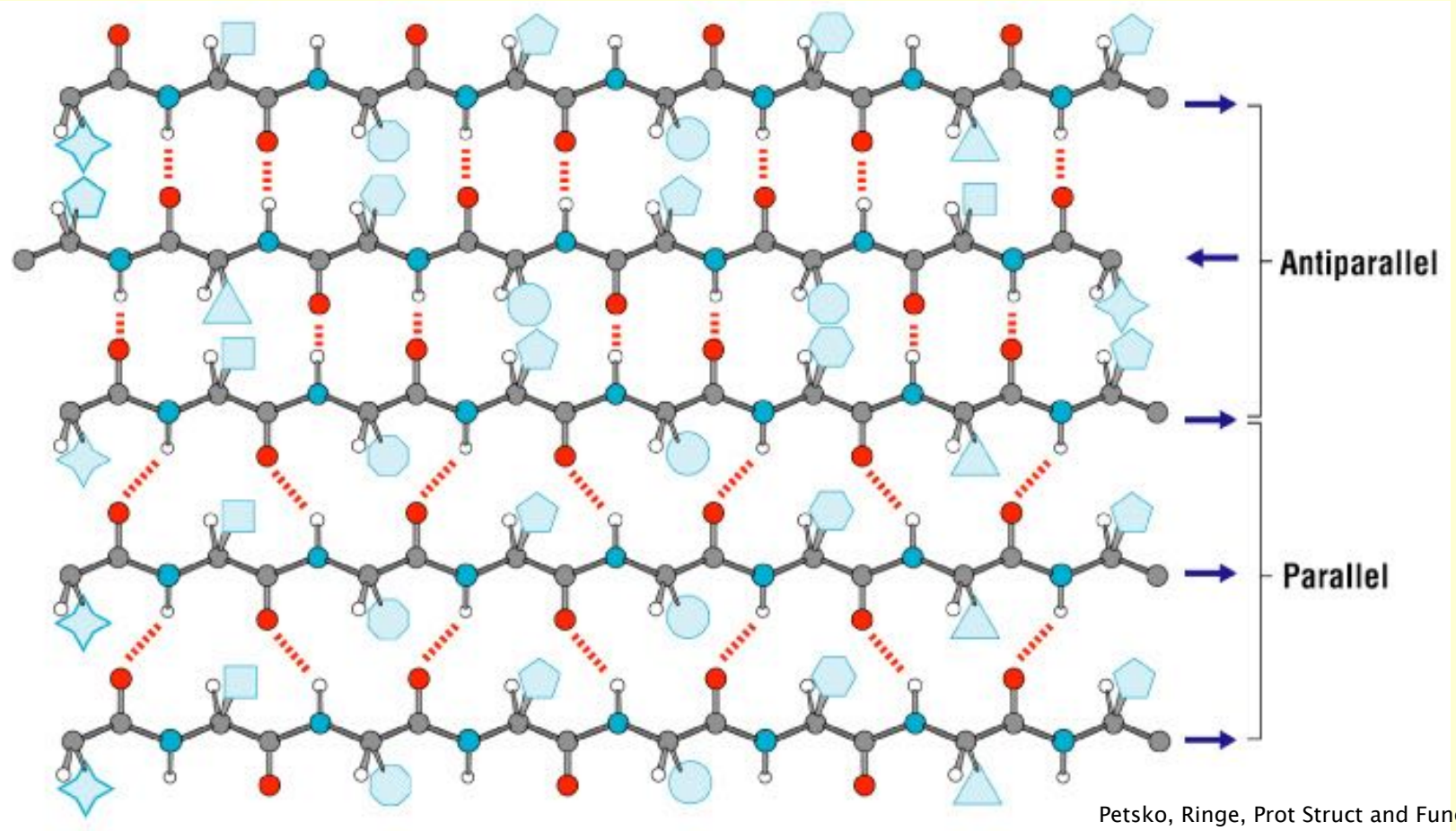
Beta Strands

- Side-chains protrude in opposite directions
- Relatively linear backbone



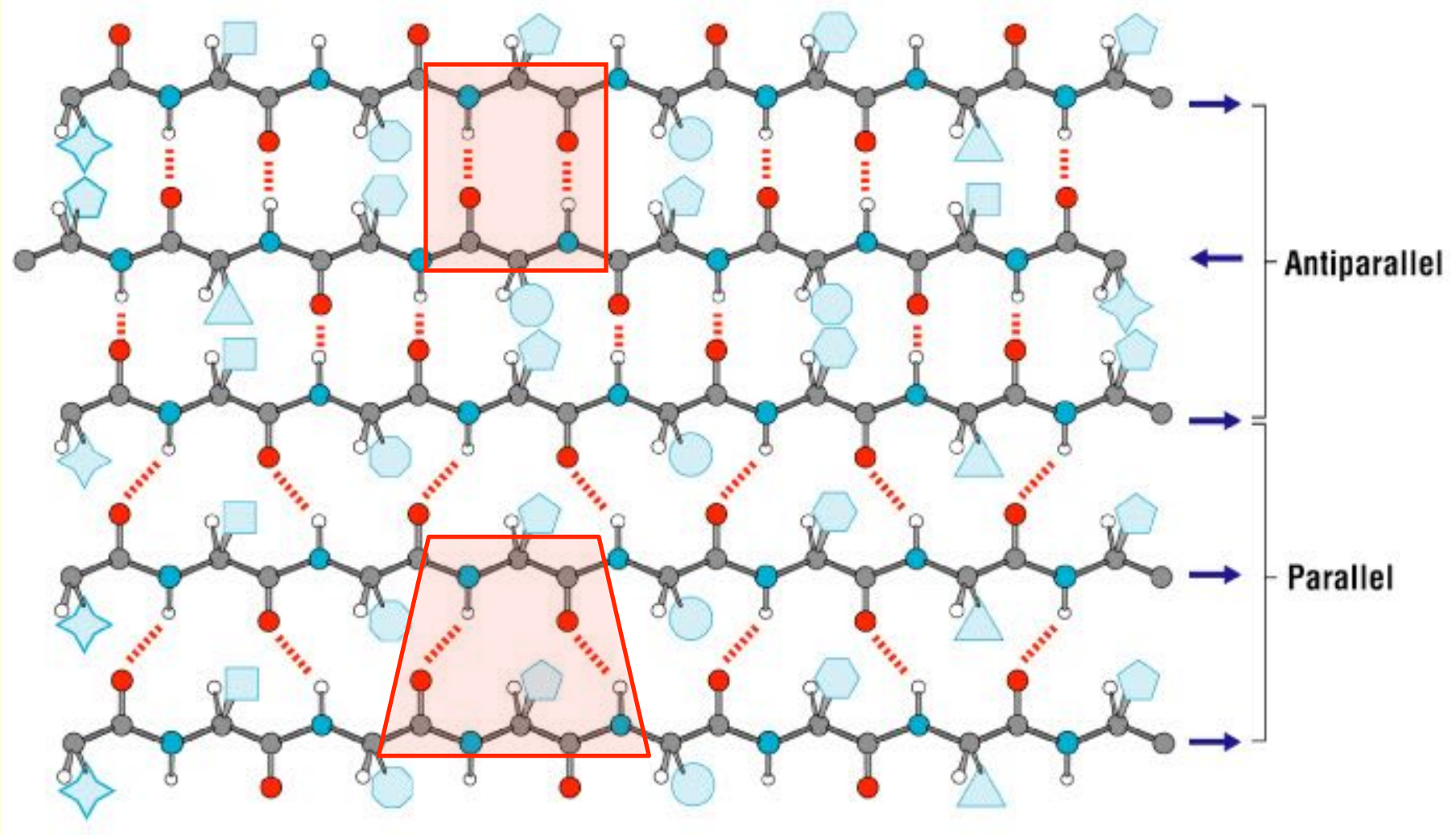
Beta Sheets

- Hydrogen bonding between backbone of strands
- Strands of a sheet may be separated by arbitrary number of amino acids



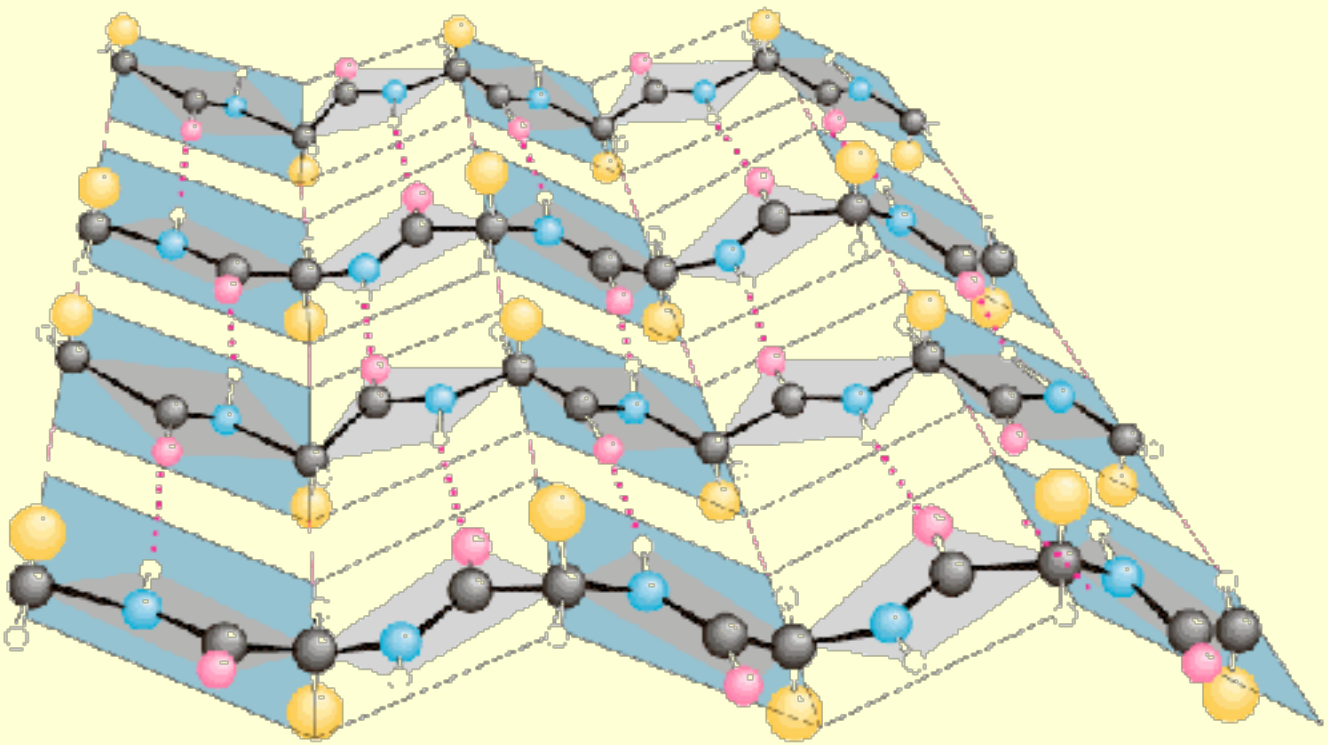
Petsko, Ringe, Prot Struct and Function, 2004

Hydrogen Bonding



Beta Sheets

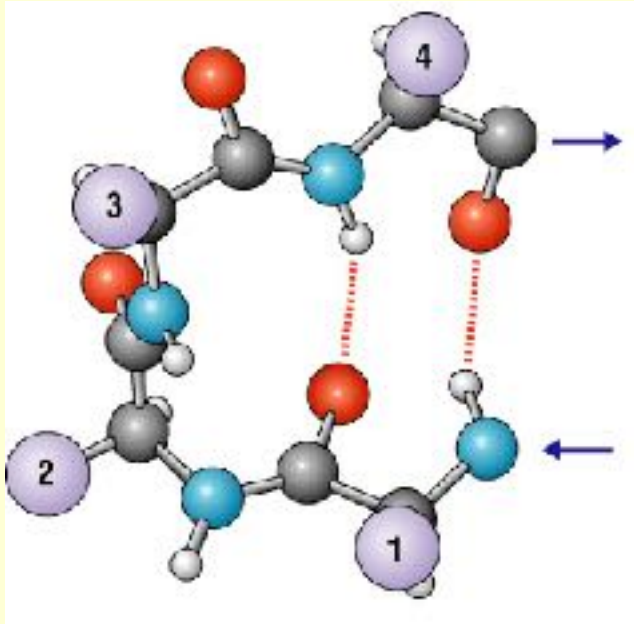
- Hydrogen bonding between backbone of strands
- Strands of a sheet may be separated by arbitrary number of amino acids



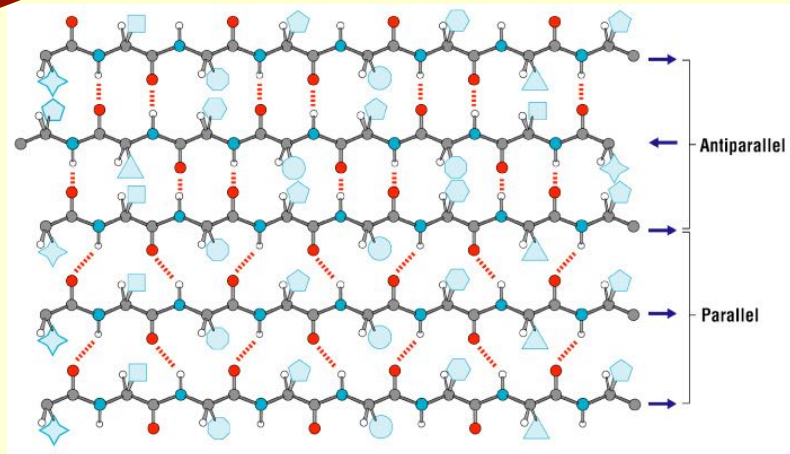
Garrett & Grisham: Biochemistry, 2/e

AntiParallel Beta Sheets

- Antiparallel sheets most commonly have beta-turns (aka. hairpin turn) connecting strands



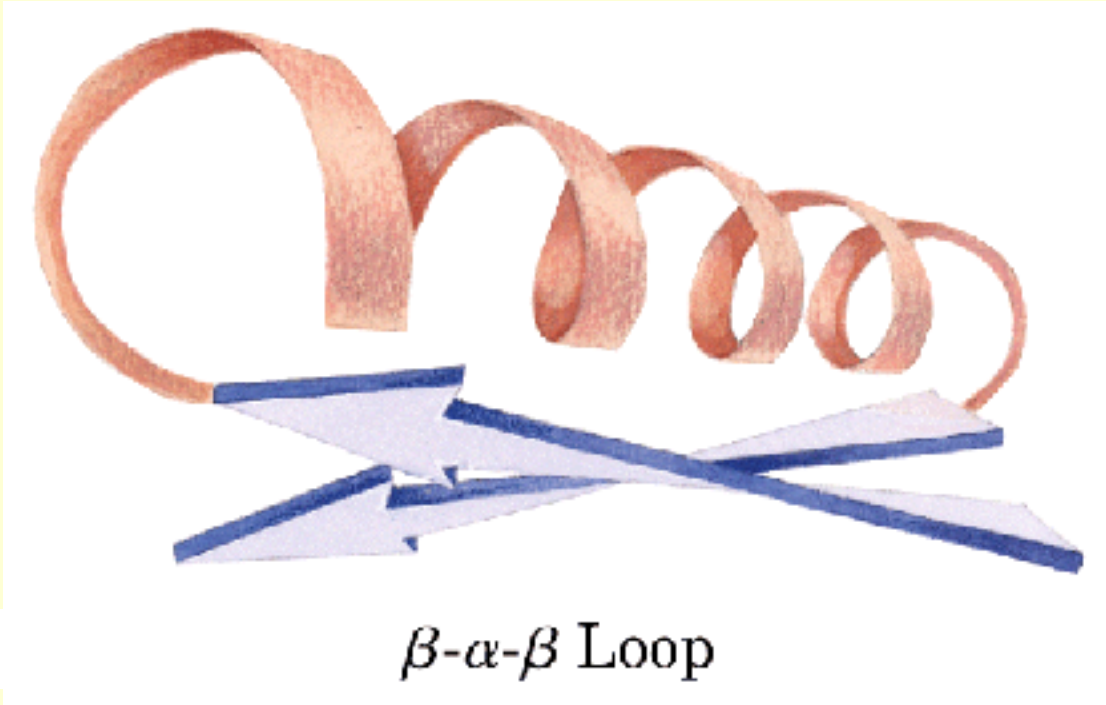
Hydrogen bonds between residues 1 and 4

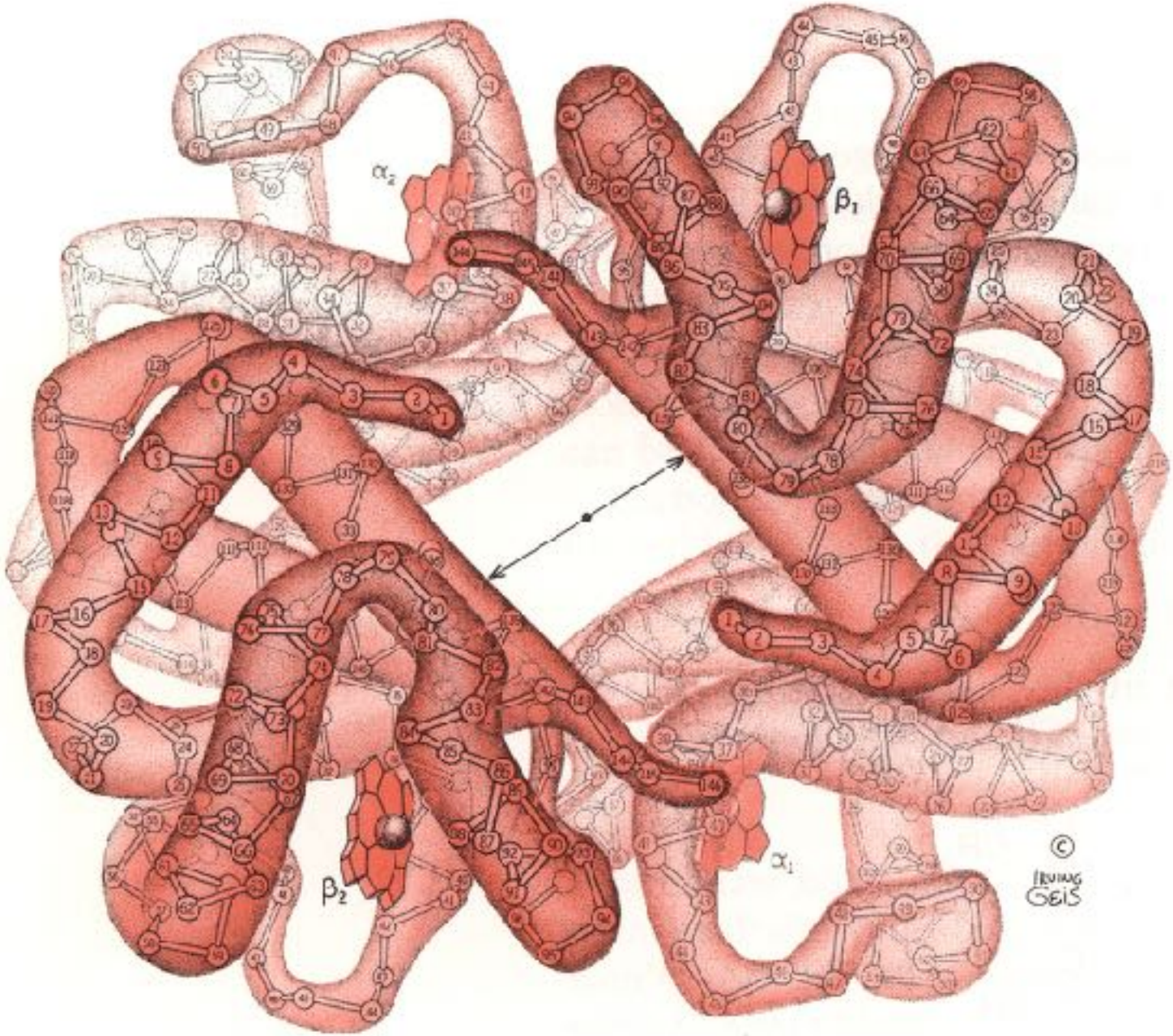


Petsko, Ringe, Prot Struct and Function, 2004

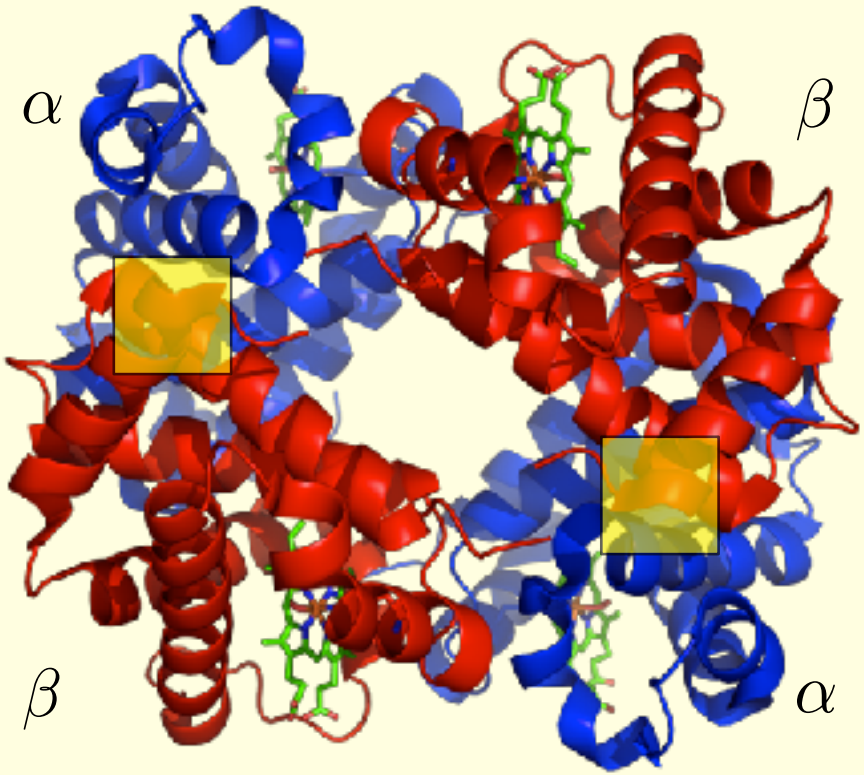
Parallel Beta Sheets

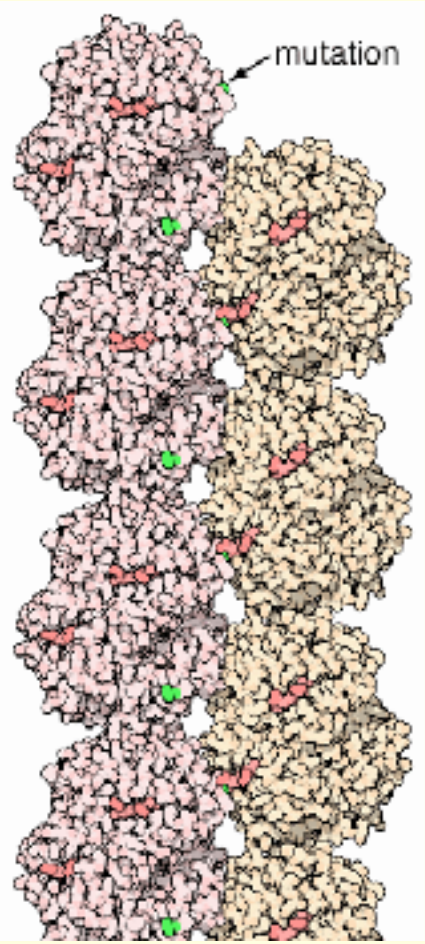
- Discontiguous by necessity
  Often connected by alpha-helix
- Less twisted than antiparallel sheets

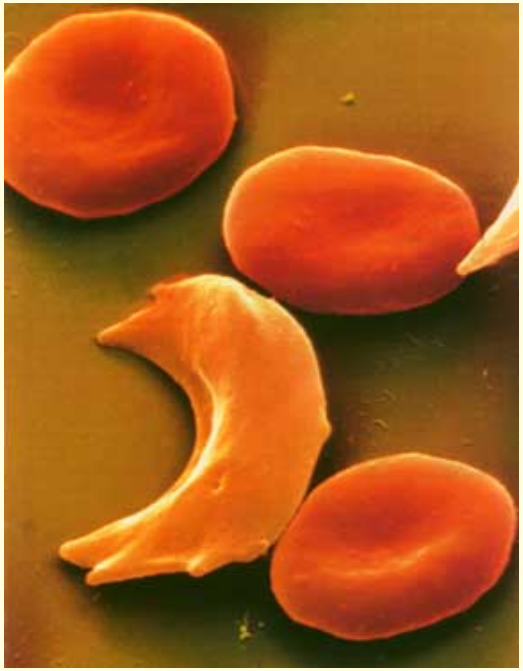




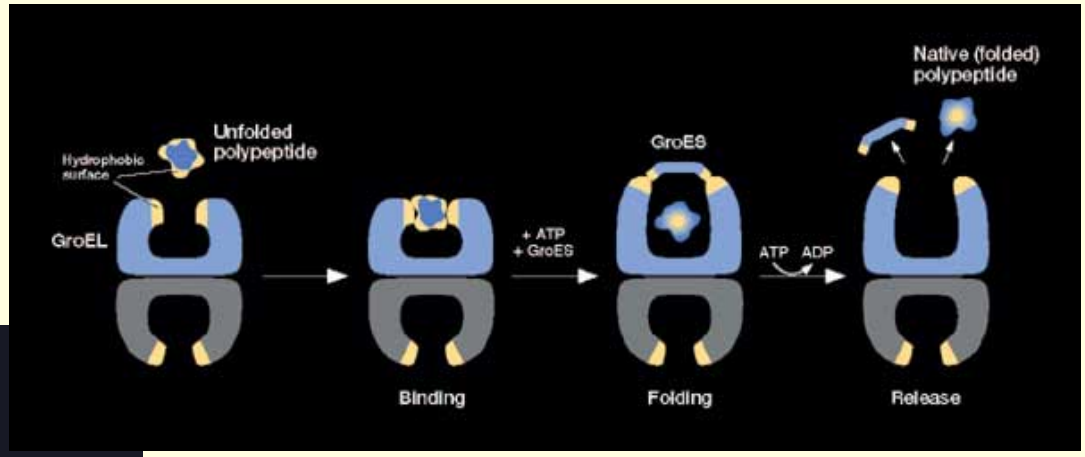
Deoxyhemoglobin

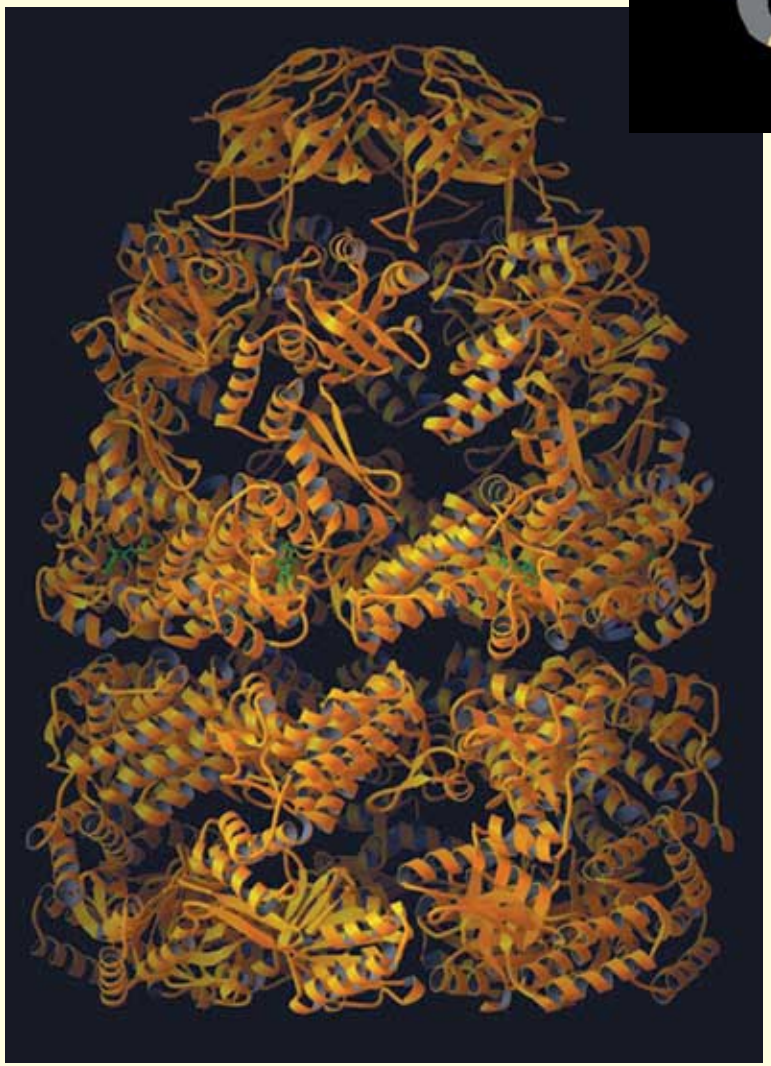






A GLU to VAL mutation at 6th amino acid in the β-subchains causes hemoglobin to aggregate, resulting in sickle-cell anemia.

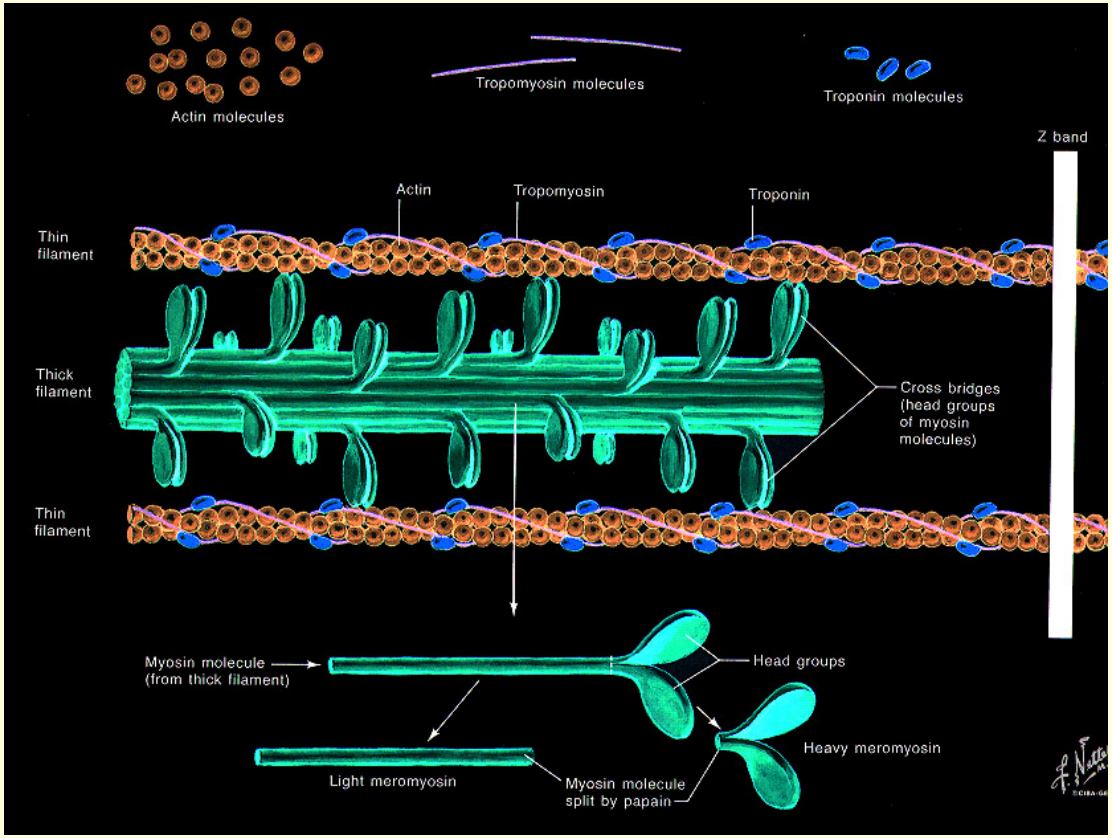


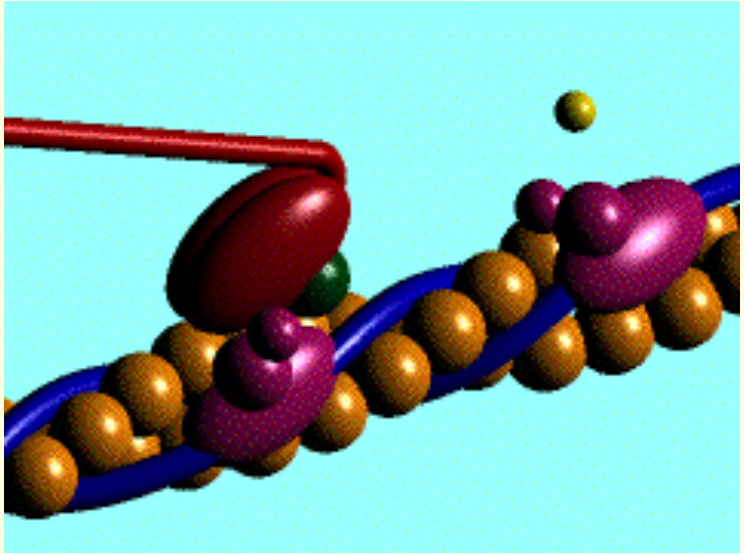


Some proteins need “help” during the folding process from “chaperonins”.

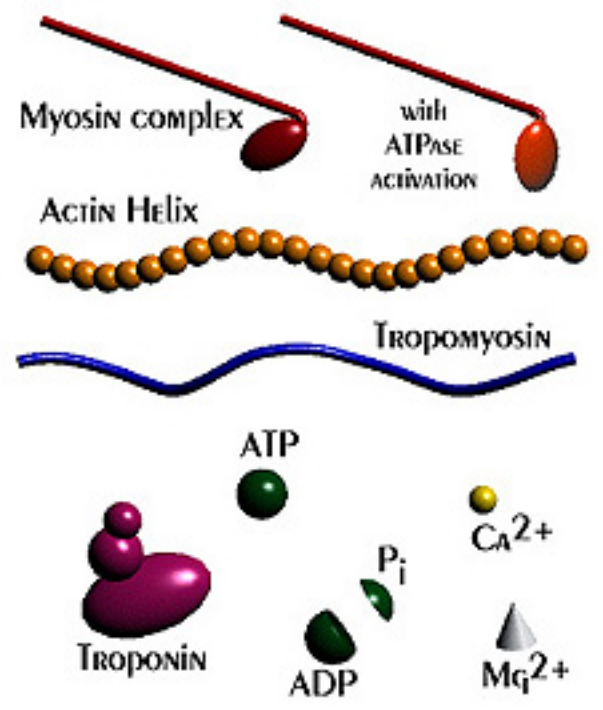
GroEL-GroES complex (Horwitch *et al*.)

Actin and Myosin





www.sci.sdsu.edu/movies/actin_**myosin**.html



Titin is the largest known protein, with 38,138 residues (4200 kDA).

“Computing” Protein Structure

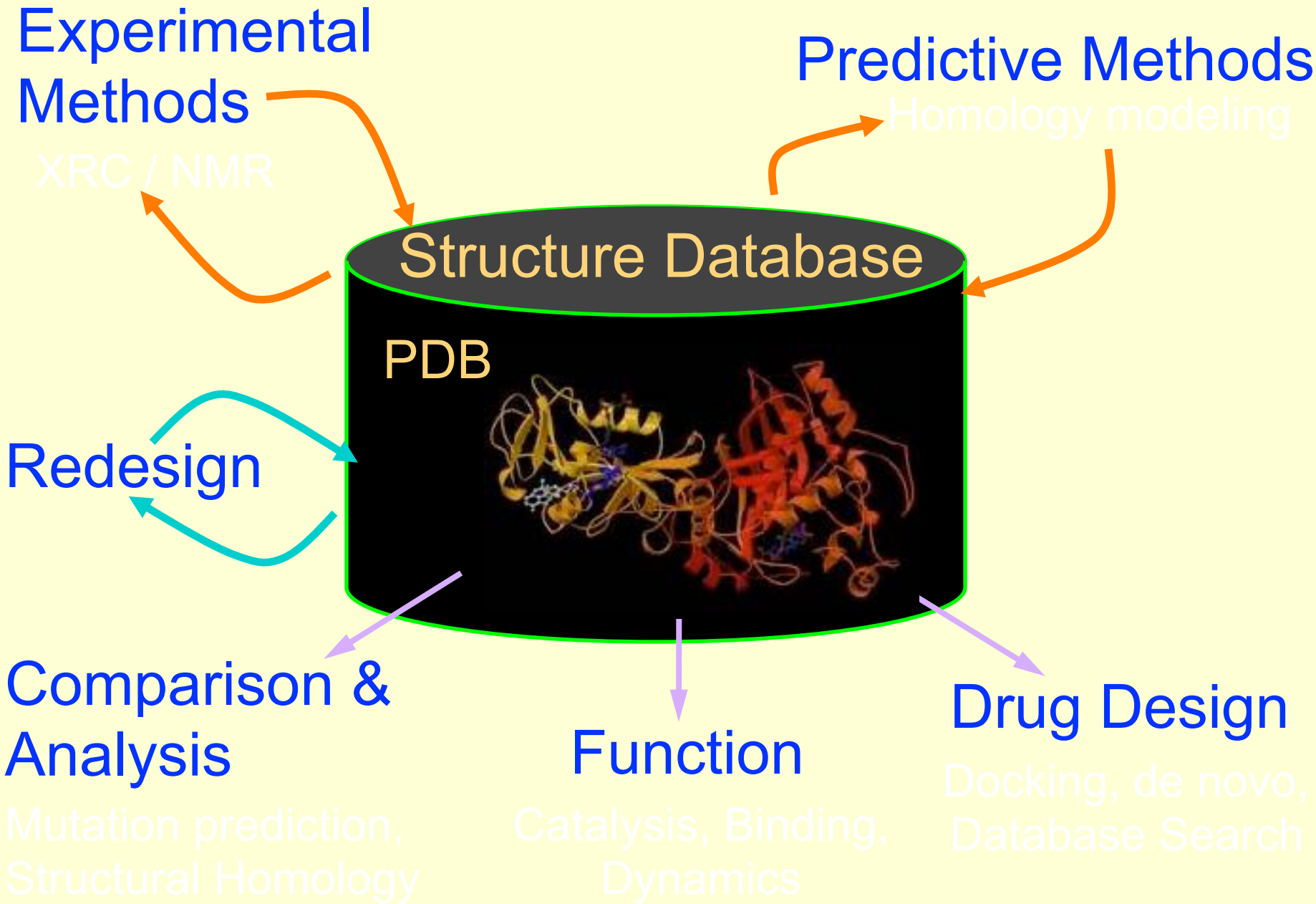
- Molecular Dynamics Simulations
- *De novo* Prediction
- Homology Modeling
- X-ray diffraction
- Nuclear Magnetic Resonance Spectroscopy

Molecular Dynamics

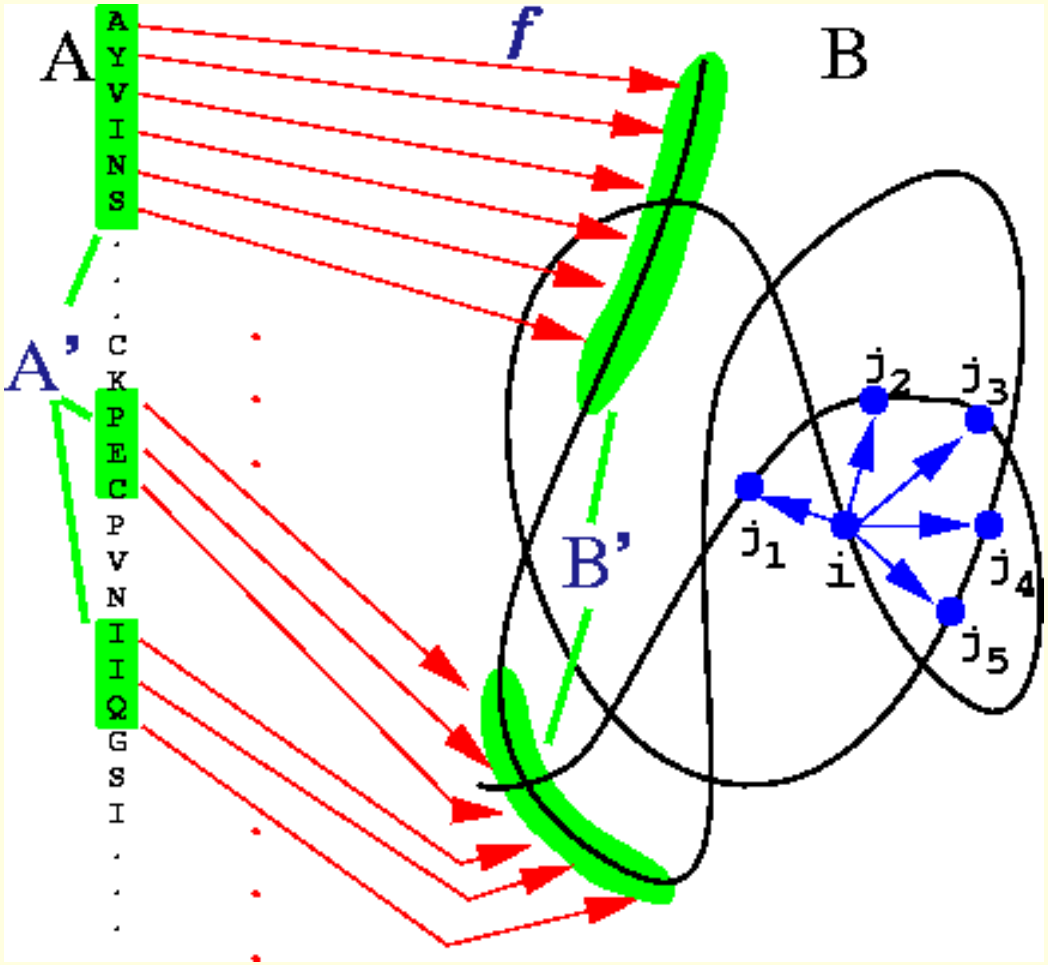
- We can also use folding techniques to study proteins of known structure.
- By simulating physics/chemistry, we can study normal modes, binding, interaction, etc.

http://www.stanford.edu/group/pandegroup/folding/villin/

Formation of a 30-residue α-helix


Experimental Methods
XRC / NMR
Predictive Methods
Homology modeling
Structure Database
PDB
Redesign
Comparison & Analysis
Mutation prediction, Structural Homology
Function
Catalysis, Binding, Dynamics
Drug Design
Docking, de novo, Database Search

Leveraging Existing Structures



primary sequence

known structure

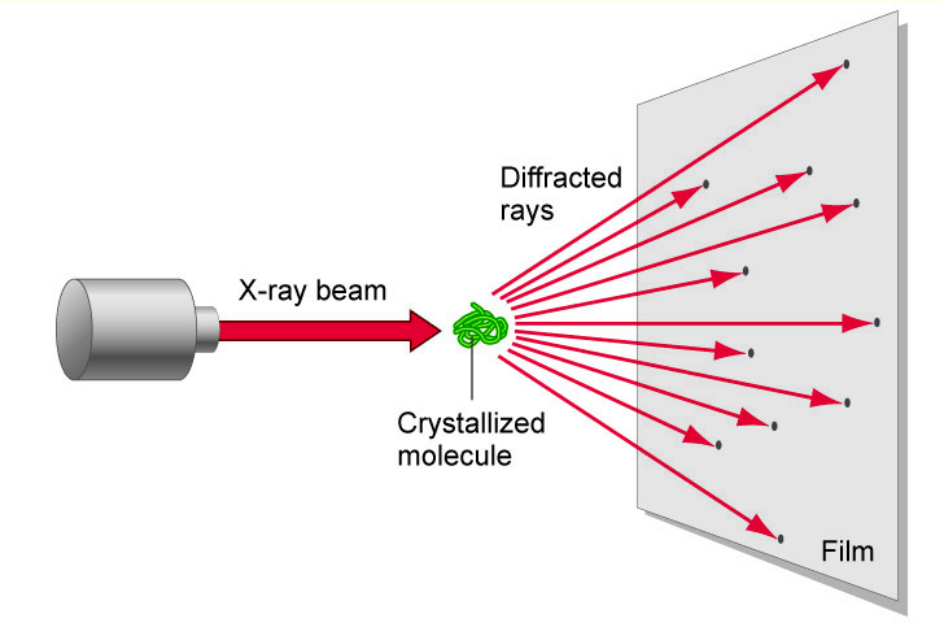
Input:
1) Primary sequence P for unknown structure.
2) Known homologous structure.

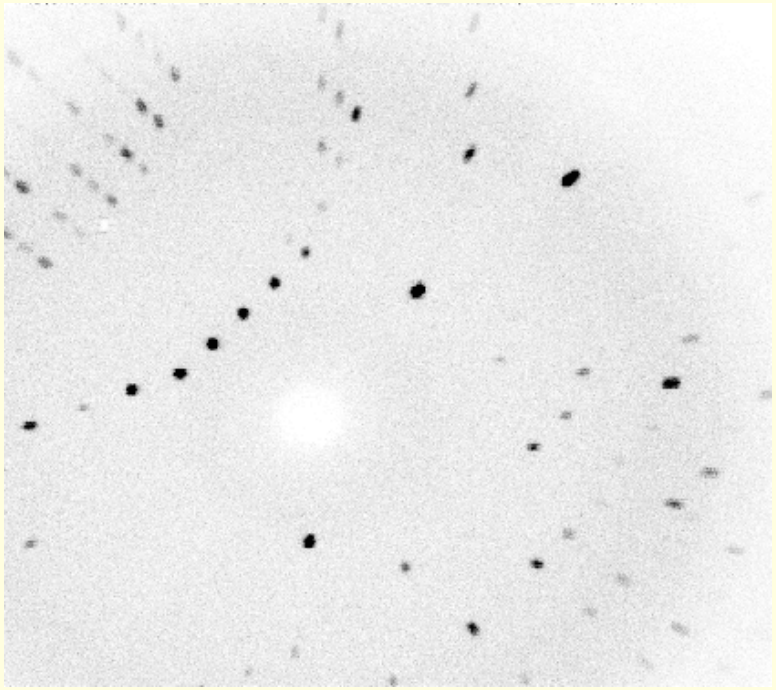
Output:
3D structure of A.

Goal:
Minimize energy of A.

Threading is NP-hard!

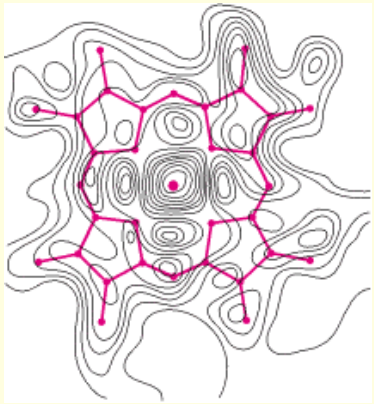
X-ray Crystallography





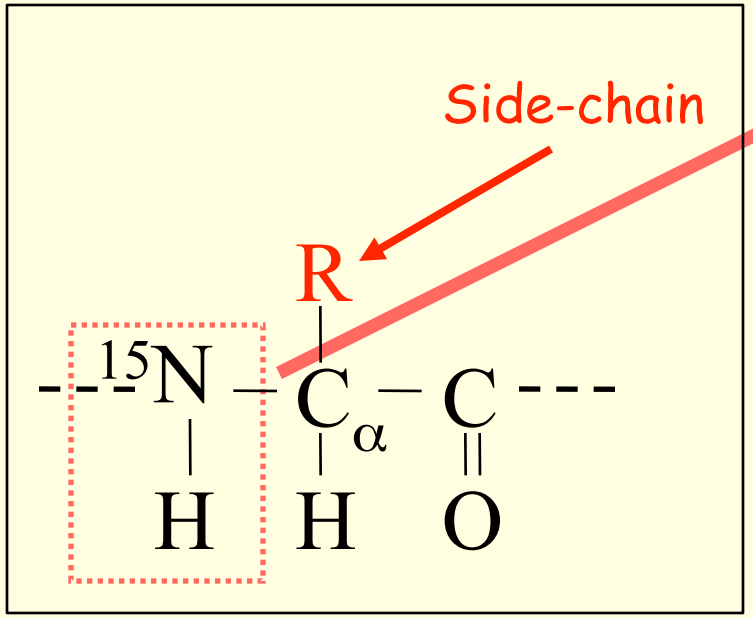
The diffraction pattern is the Fourier transform of electron density. To compute electron density, we can “invert” the diffraction pattern.

But we cannot measure the phase of diffracted wave! We must simulate it…

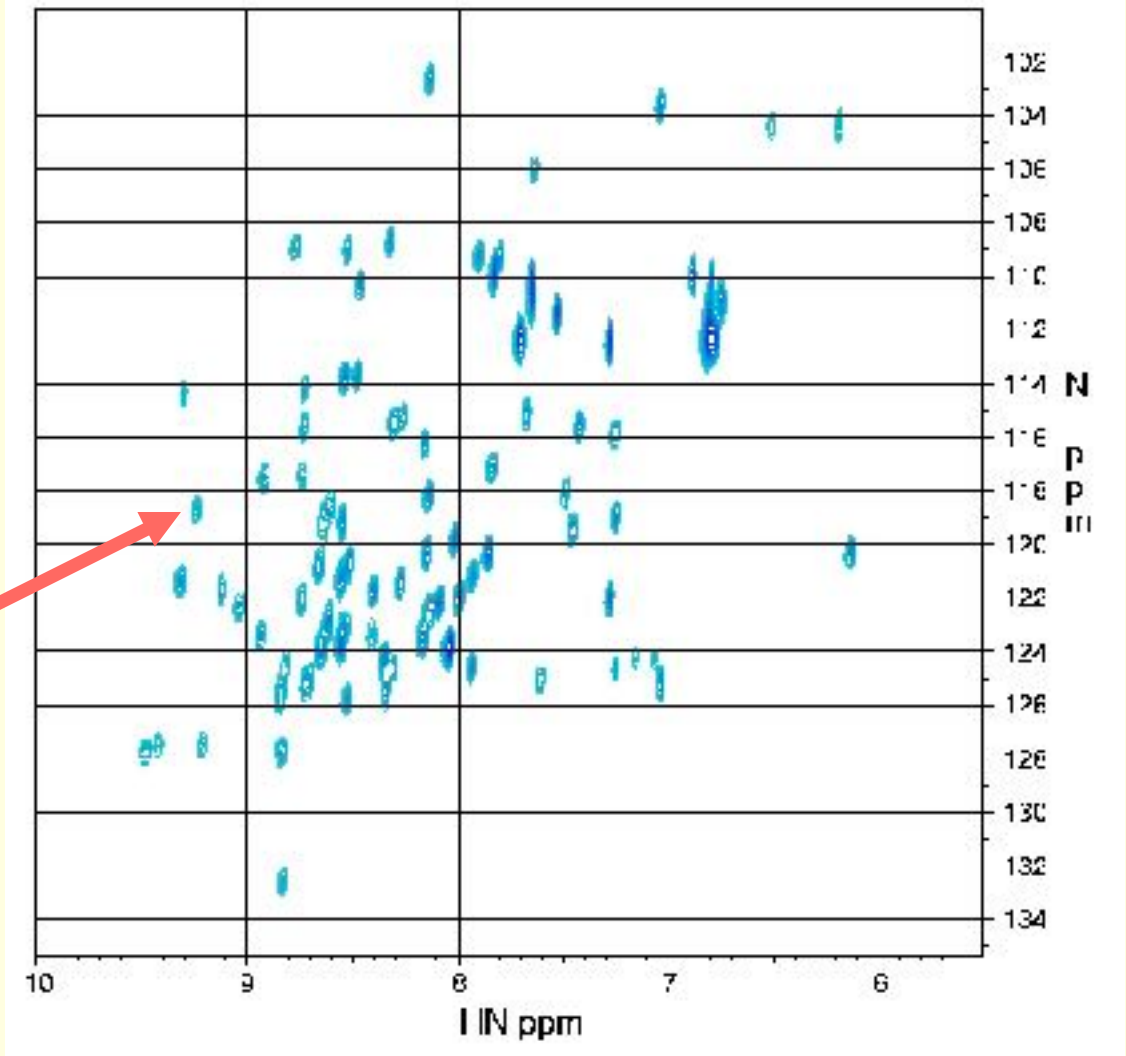


Myoglobin structure was solved by X-ray crystallography (Perutz and Kendrew 1958)

NMR Spectroscopy



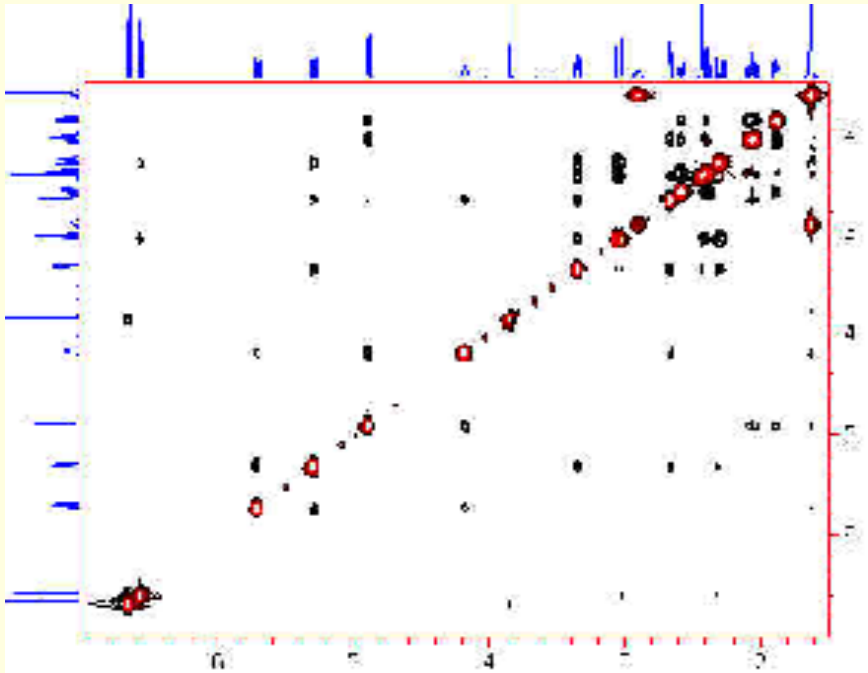
Amino Acid



Ubiquitin: H^N - ^{15}N HSQC

NMR Structure Determination

Distance-based methods use *assigned* NOE restraints.



2D NOE spectrum

	a_1	a_2	a_3	. . .	a_n
a_1	0	?	<6	. . .	
a_2	?	0	<6	. . .	
a_3	?	?	0	. . .	
. . .	. . .	. . .	. . .	. . .	
a_n					

NMR Structure Determination

Structure determination from NOE restraints is NP-hard! [Saxe '79; Hendrickson '92, '95]

	a_1	a_2	a_3	. . .	a_n
a_1	0	?	<6	. . .	
a_2	?	0	<6	. . .	
a_3	?	?	0	. . .	
⋮	⋮	⋮	⋮	⋱	
a_n					

Distance Geometry Method [Crippen&Havel'88]

Exponential Time