BSC 4934: Q'BIC Capstone Workshop

Giri Narasimhan

ECS 254A; Phone: x3748

giri@cs.fiu.edu

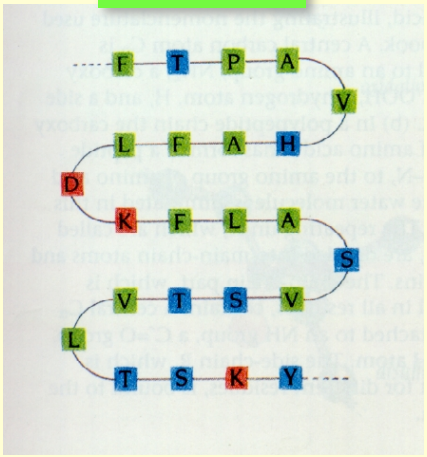
http://www.cs.fiu.edu/~giri/teach/BSC4934_Su10.html

July 2010

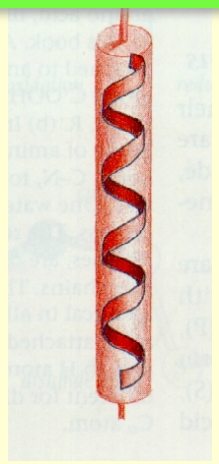
Protein Structures

- Sequences of amino acid residues
- 20 different amino acids

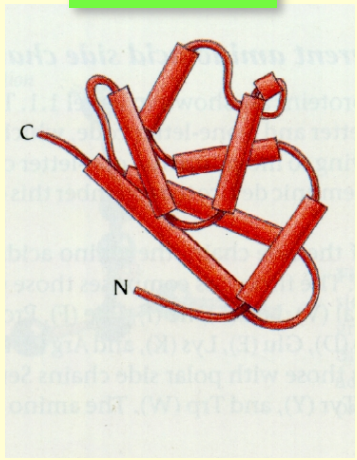
Primary



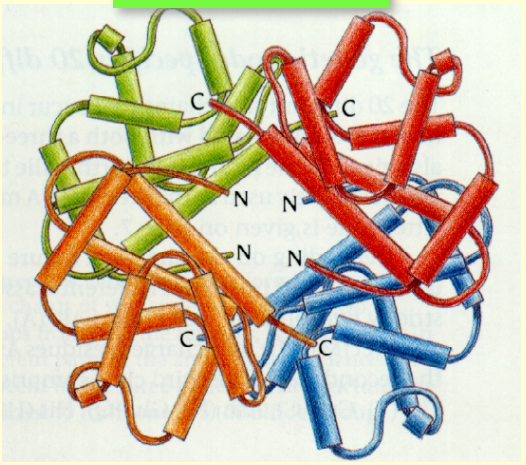
Secondary



Tertiary



Quaternary



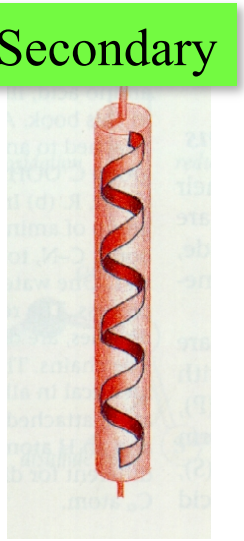
Proteins

- **Primary structure** is the sequence of amino acid residues of the protein, e.g., Flavodoxin:

  AKIGLFYGTQTGVTQTIAESIQQEFGGESIVDLNDIANADA…

- Different regions of the sequence form local regular **secondary structures**, such as
  - Alpha helix, beta strands, etc.

  AKIGLFYGTQTGVTQTIAESIQQEFGGESIVDLNDIANADA…



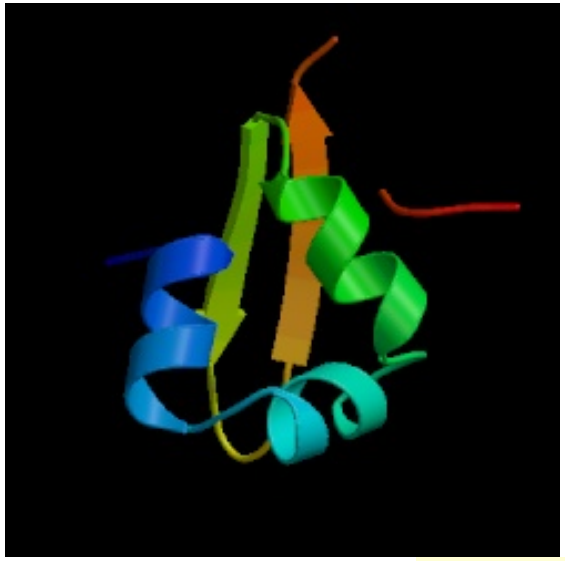
More on Secondary Structures

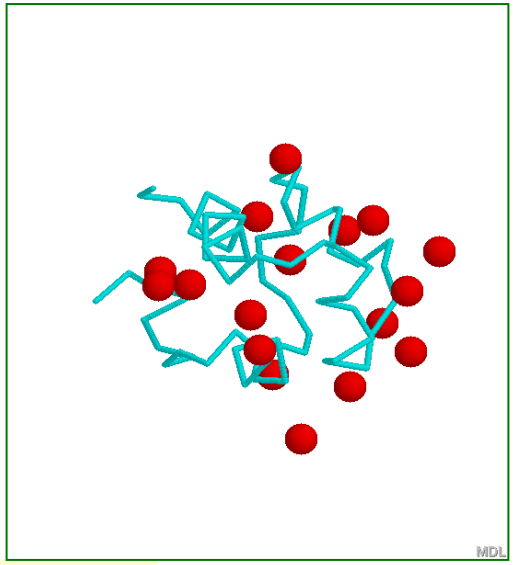
- **α-helix**
  - Main chain with peptide bonds
  - Side chains project outward from helix
  - Stability provided by H-bonds between CO and NH groups of residues 4 locations away.
- **β-strand**
  - Stability provided by H-bonds with one or more β-strands, forming β-sheets. Needs a β-turn.

Proteins

- **Tertiary structures** are formed by packing secondary structural elements into a globular structure.

Myoglobin



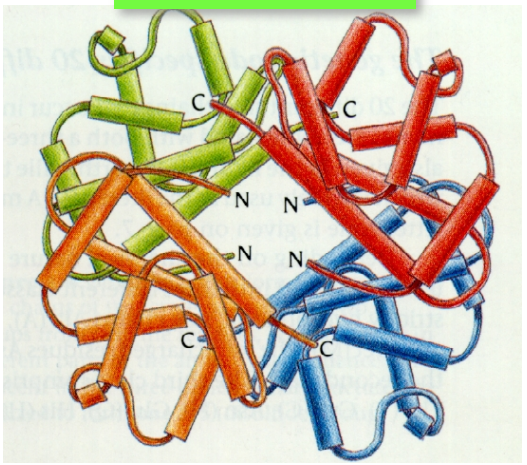


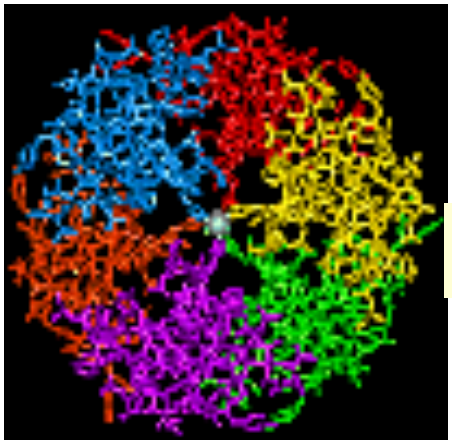
Lambda Cro

Quaternary Structures in Proteins

- The final structure may contain more than one "chain" arranged in a **quaternary structure**.

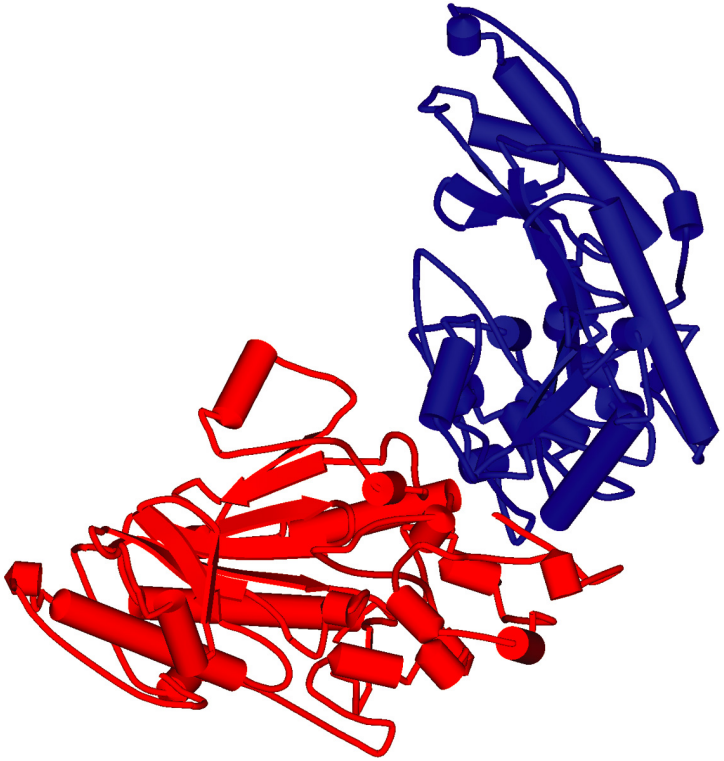
Quaternary





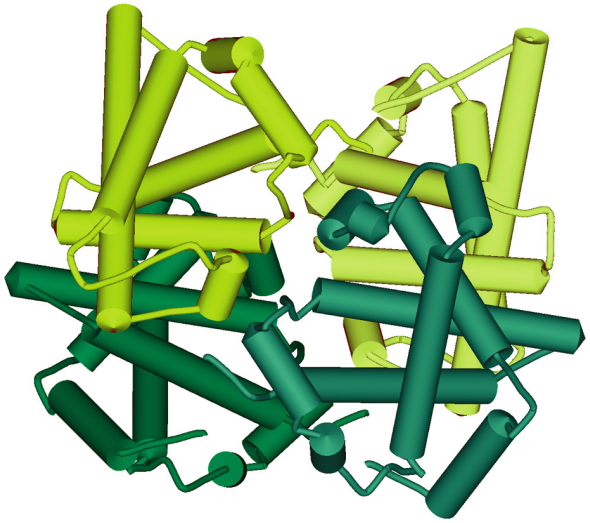
Insulin Hexamer

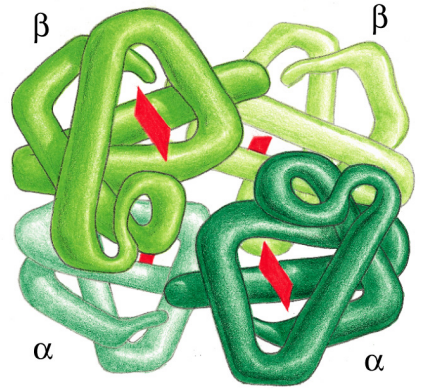
More quaternary structures



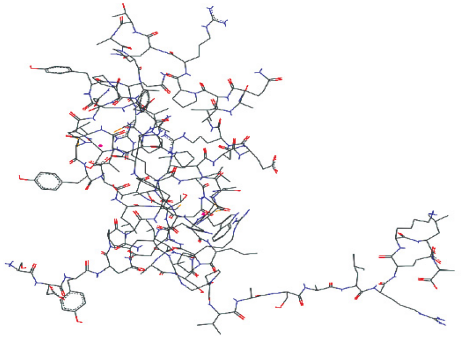
Muscle creatine kinase
(Homodimer)

Bovine deoxyhemoglobin
(Heterotetramer)

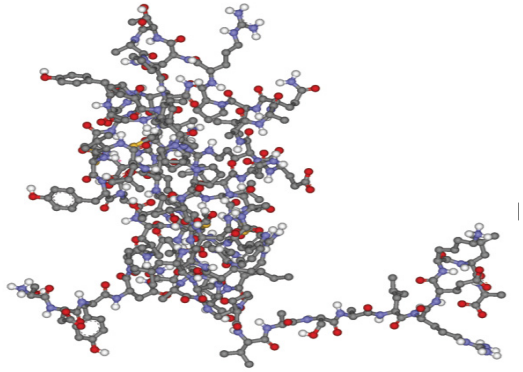




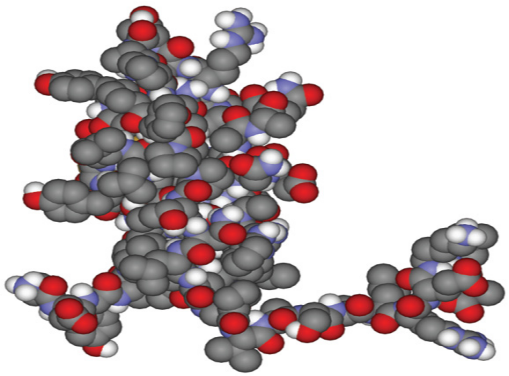
Molecular Representations



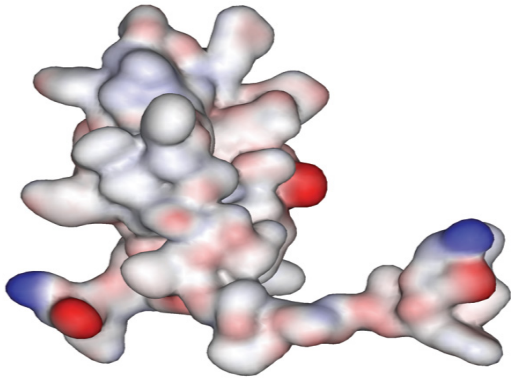
wire-frame



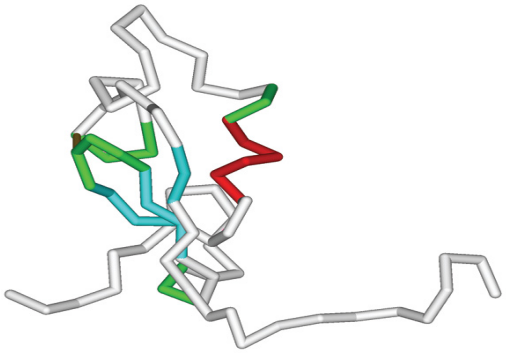
ball and stick



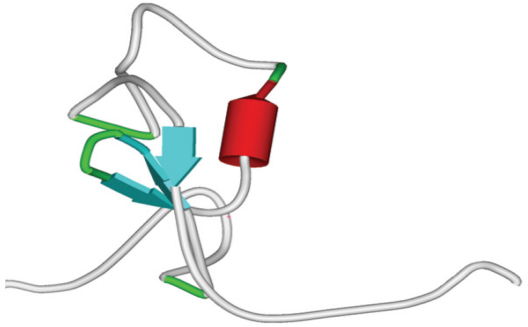
space-filling



surface

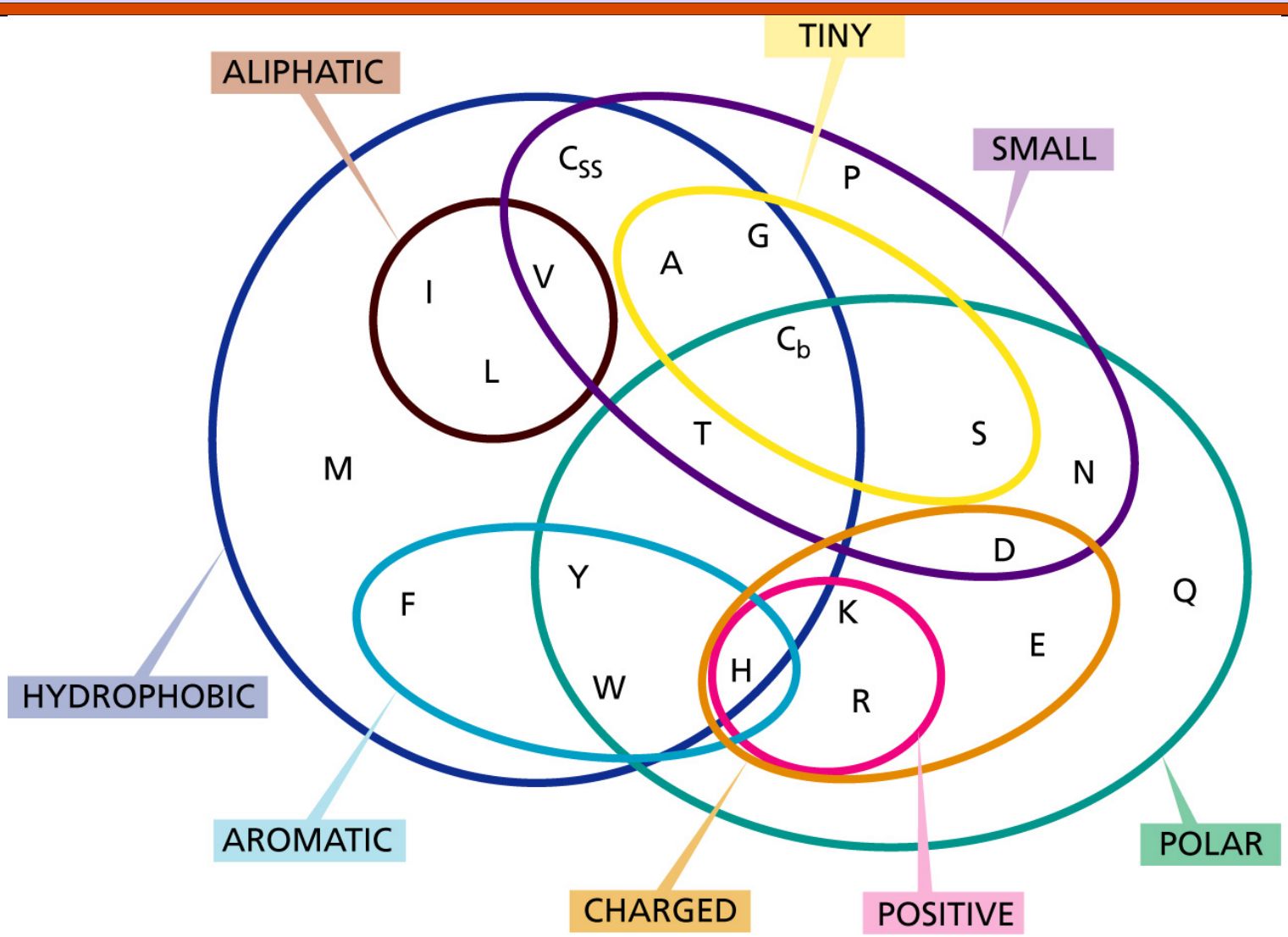


C_α representation



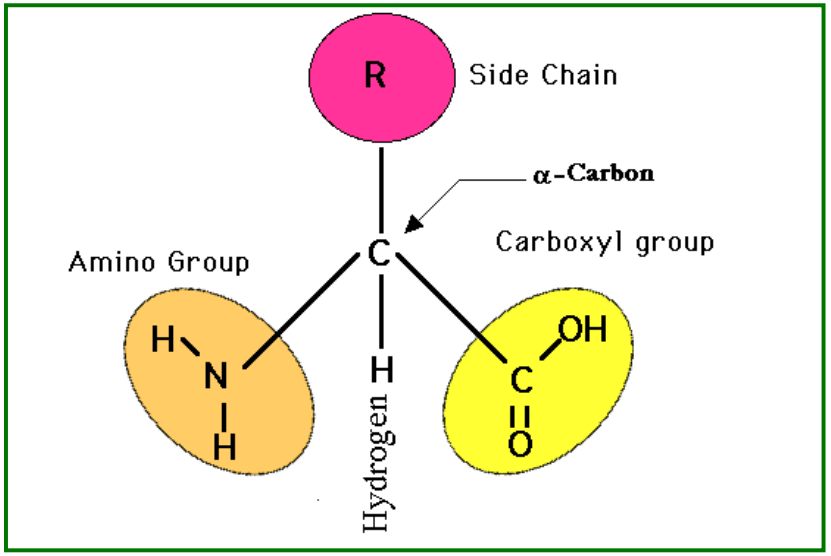
α/β schematic

Amino Acid Types



Structure of a single amino acid

All 3 figures are cartoons of an amino acid residue.



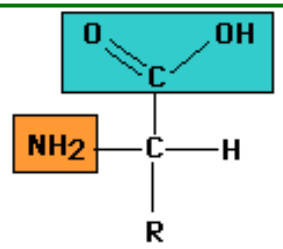
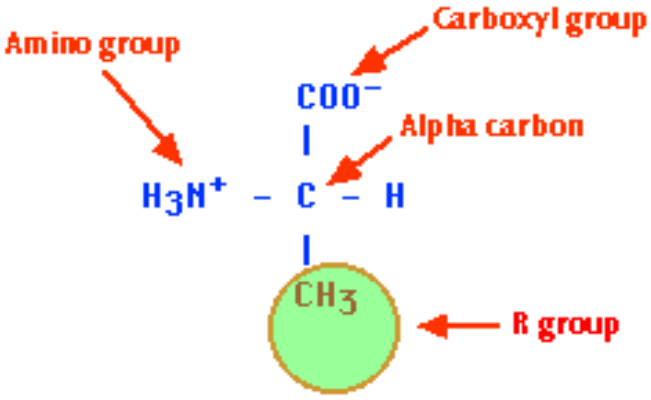
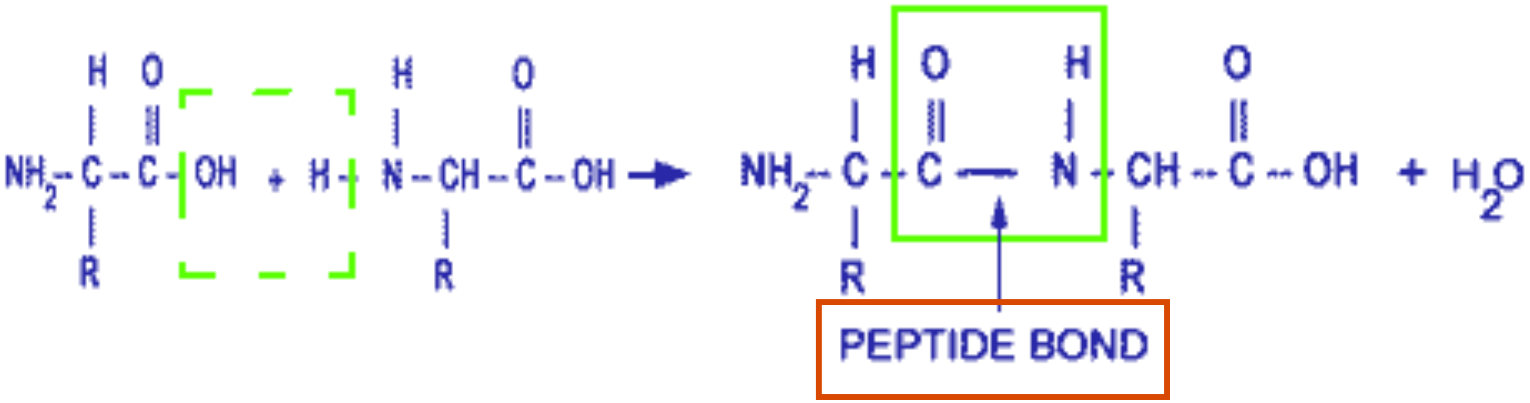


Fig. General formula for an amino acid molecule. "R" represents the variable groups that are attached to this basic molecule to make up the 20 common amino acids



Chains of amino acids



Amino acids vs Amino acid residues

Angles ϕ and ψ in the polypeptide chain

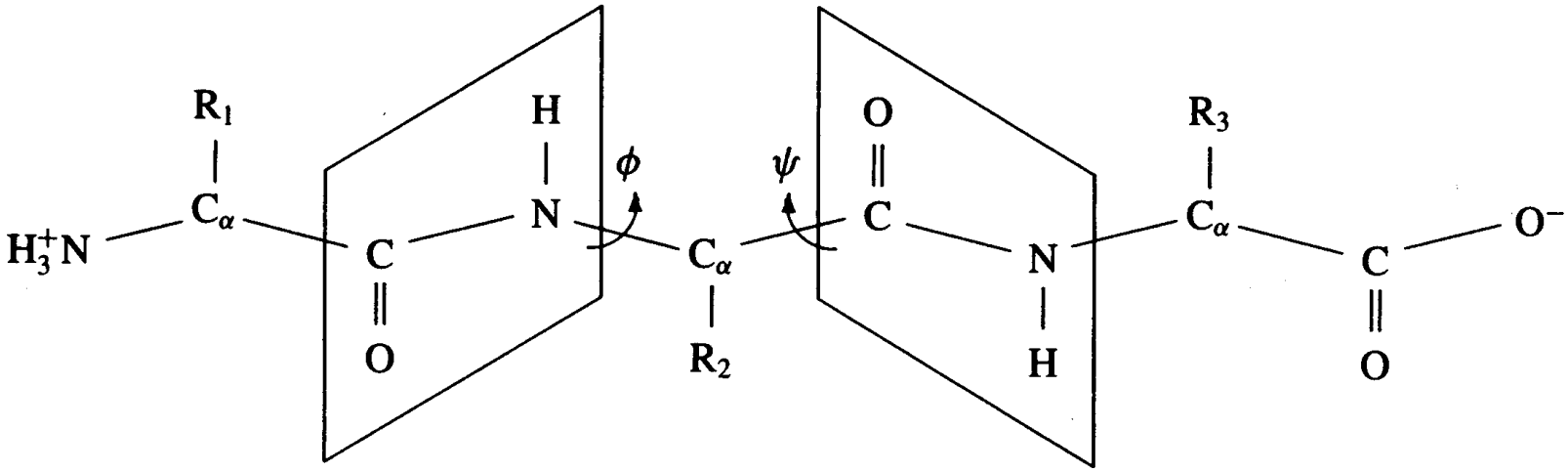
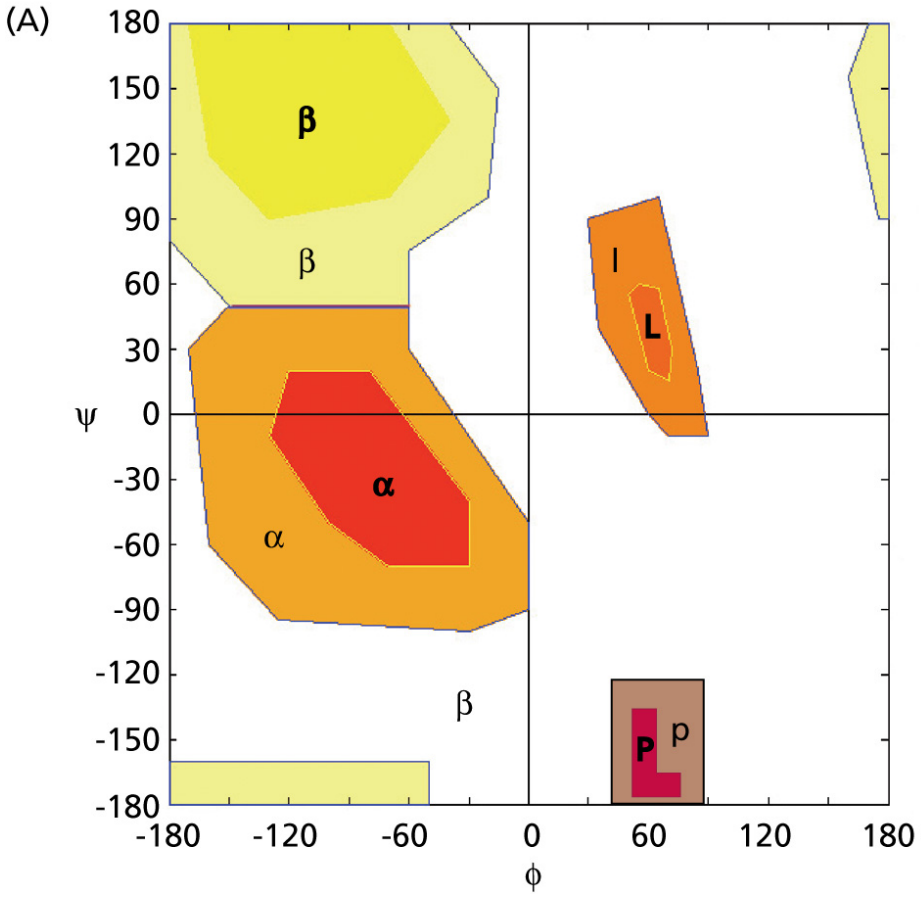
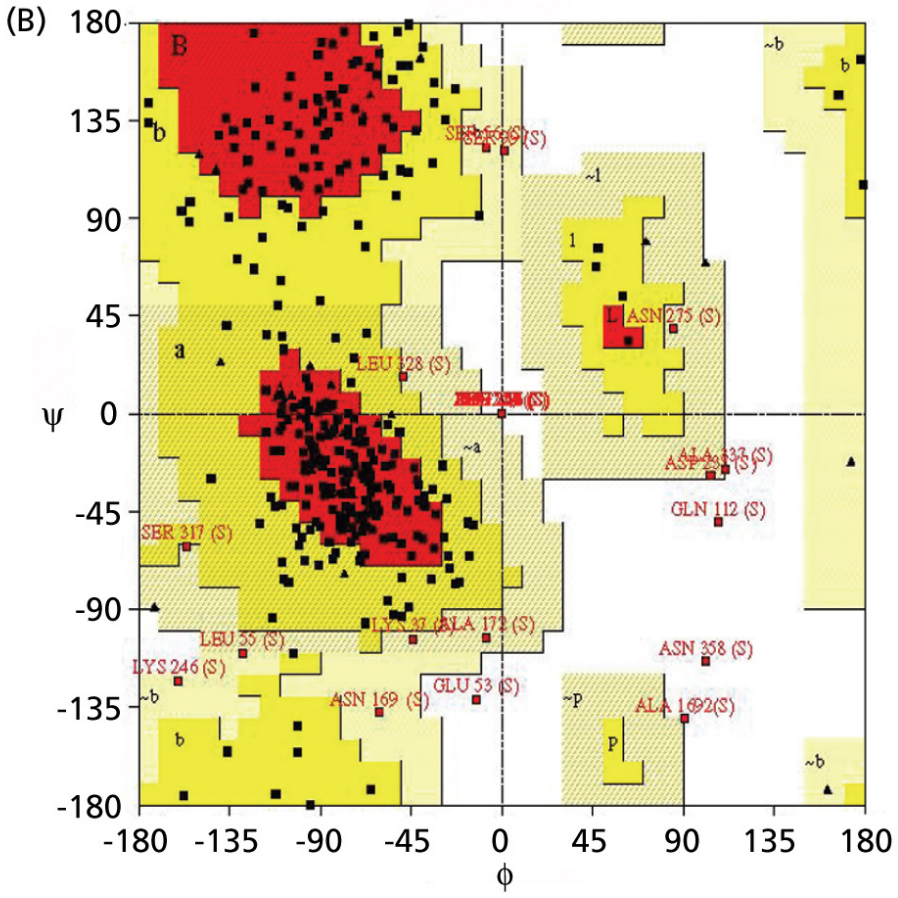


FIGURE 1.2

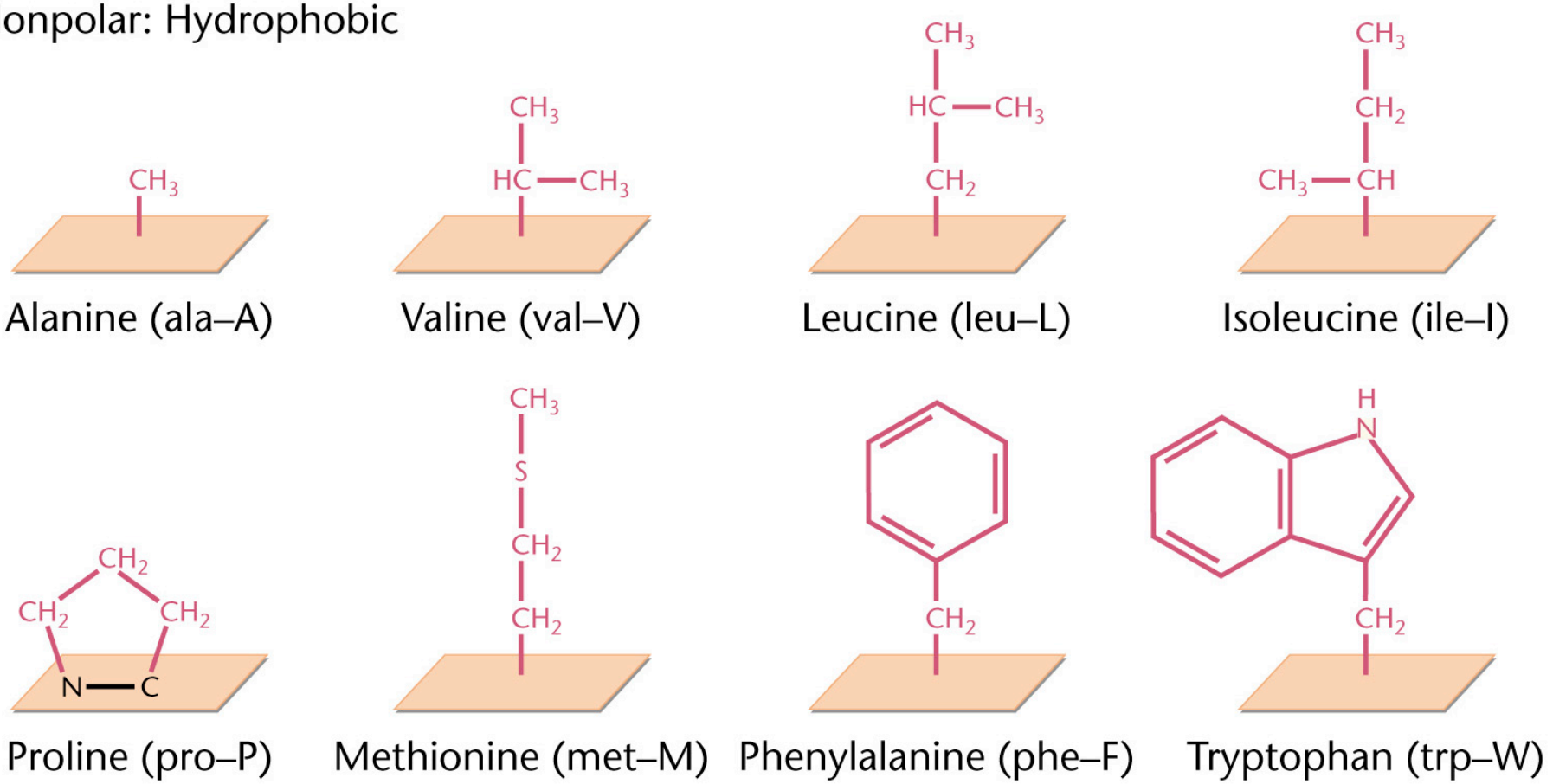
A polypeptide chain. The R_i *side chains identify the component amino acids. Atoms inside each quadrilateral are on the same plane, which can rotate according to angles* ϕ *and* ψ.

Ramachandran Plot





1. Nonpolar: Hydrophobic

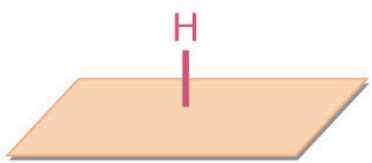


Alanine (ala–A) Valine (val–V) Leucine (leu–L) Isoleucine (ile–I)

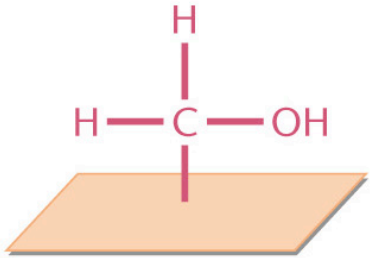
Proline (pro–P) Methionine (met–M) Phenylalanine (phe–F) Tryptophan (trp–W)

Amino Acid Structures from Klug & Cummings

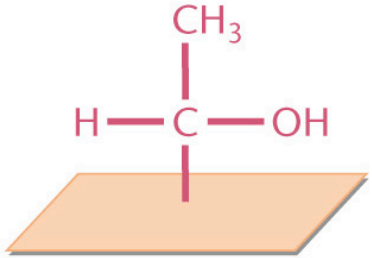
2. Polar: Hydrophilic



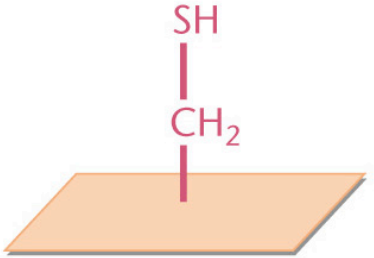
Glycine (gly–G)



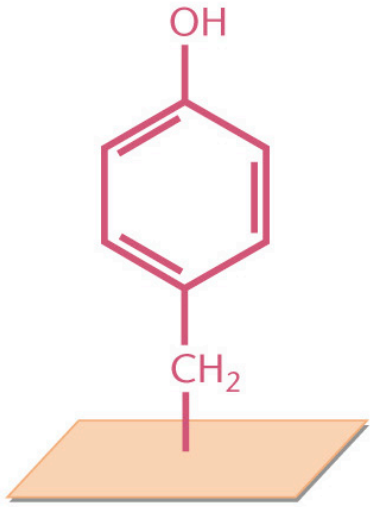
Serine (ser–S)



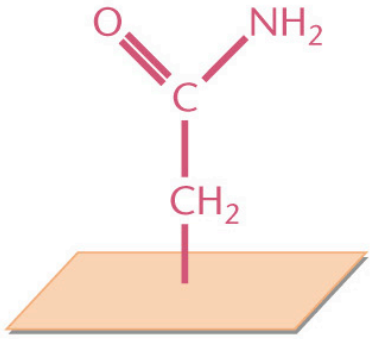
Threonine (thr–T)



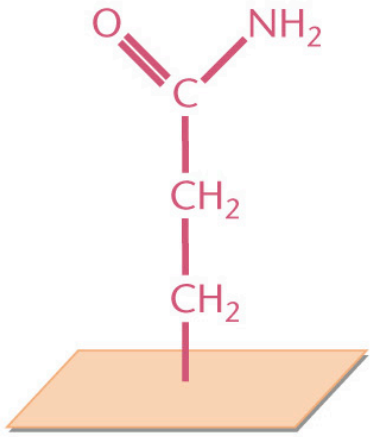
Cysteine (cys–C)



Tyrosine (tyr–Y)



Asparagine (asn–N)

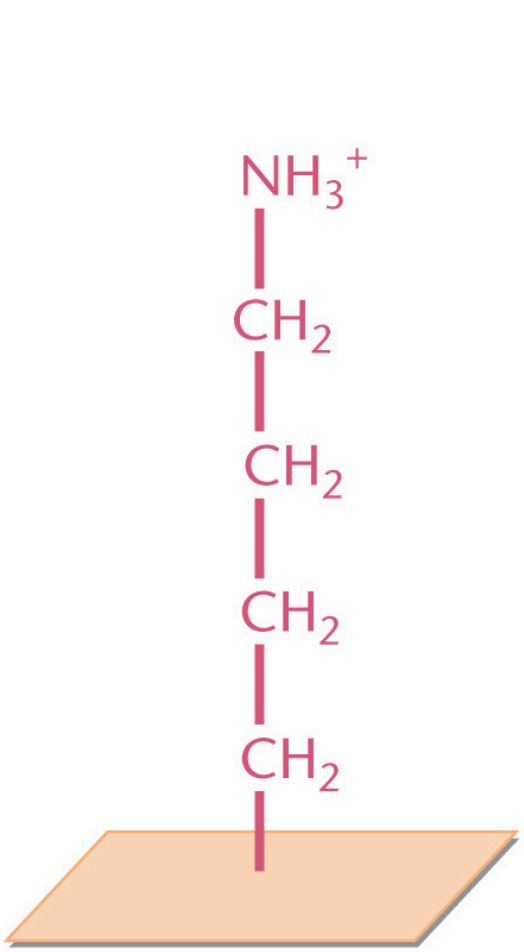


Glutamine (gln–Q)

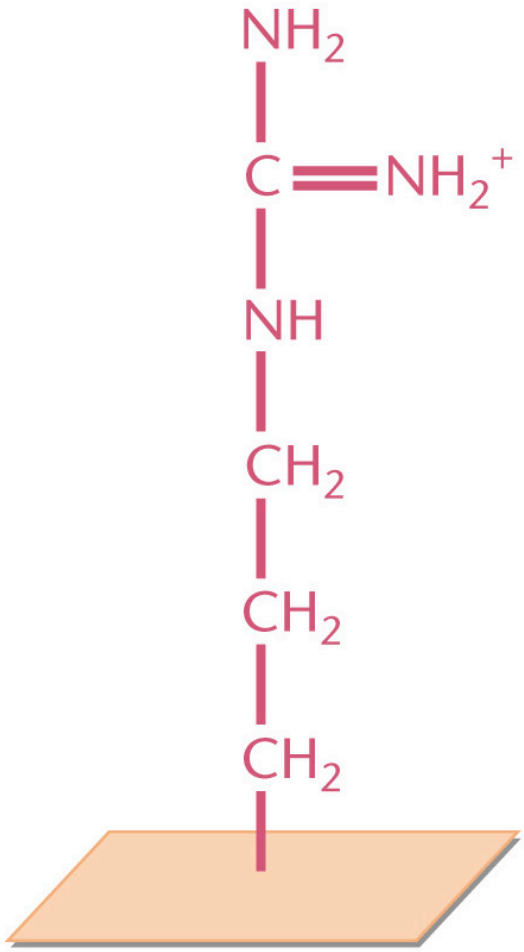
Amino Acid Structures from Klug & Cummings

3. Polar: positively charged (basic)

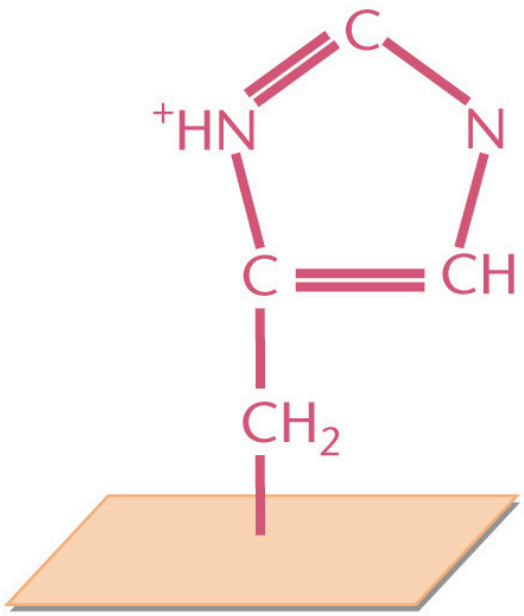
Amino Acid Structures
from Klug & Cummings



Lysine (lys–K)

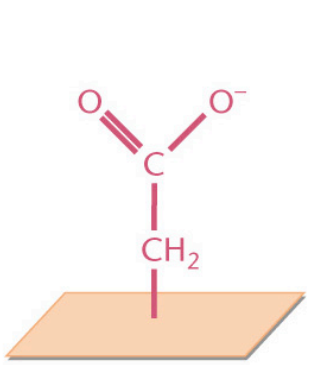


Arginine (arg–R)

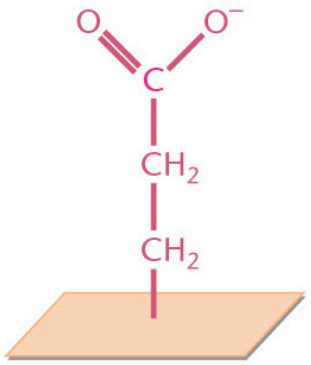


Histidine (his–H)

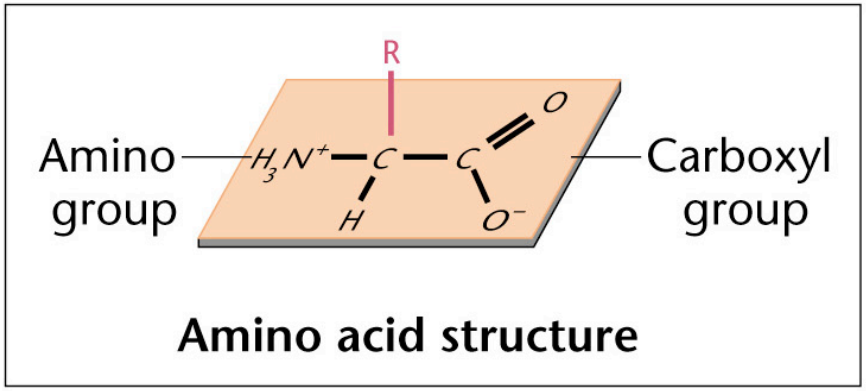
4. Polar: negatively charged (acidic)



Aspartic acid (asp–D)

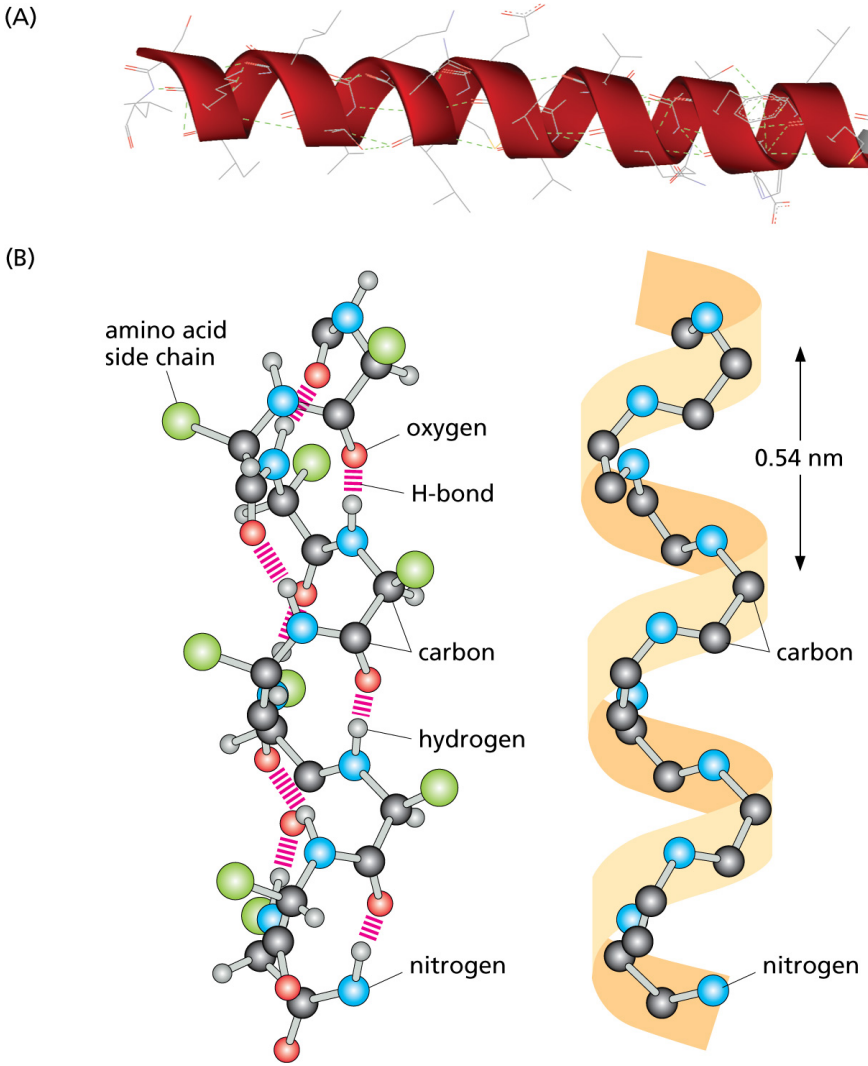


Glutamic acid (glu–E)



Amino Acid Structures from Klug & Cummings

Alpha Helix


(A)
(B)
amino acid
side chain
oxygen
H-bond
carbon
hydrogen
nitrogen
0.54 nm
carbon
nitrogen

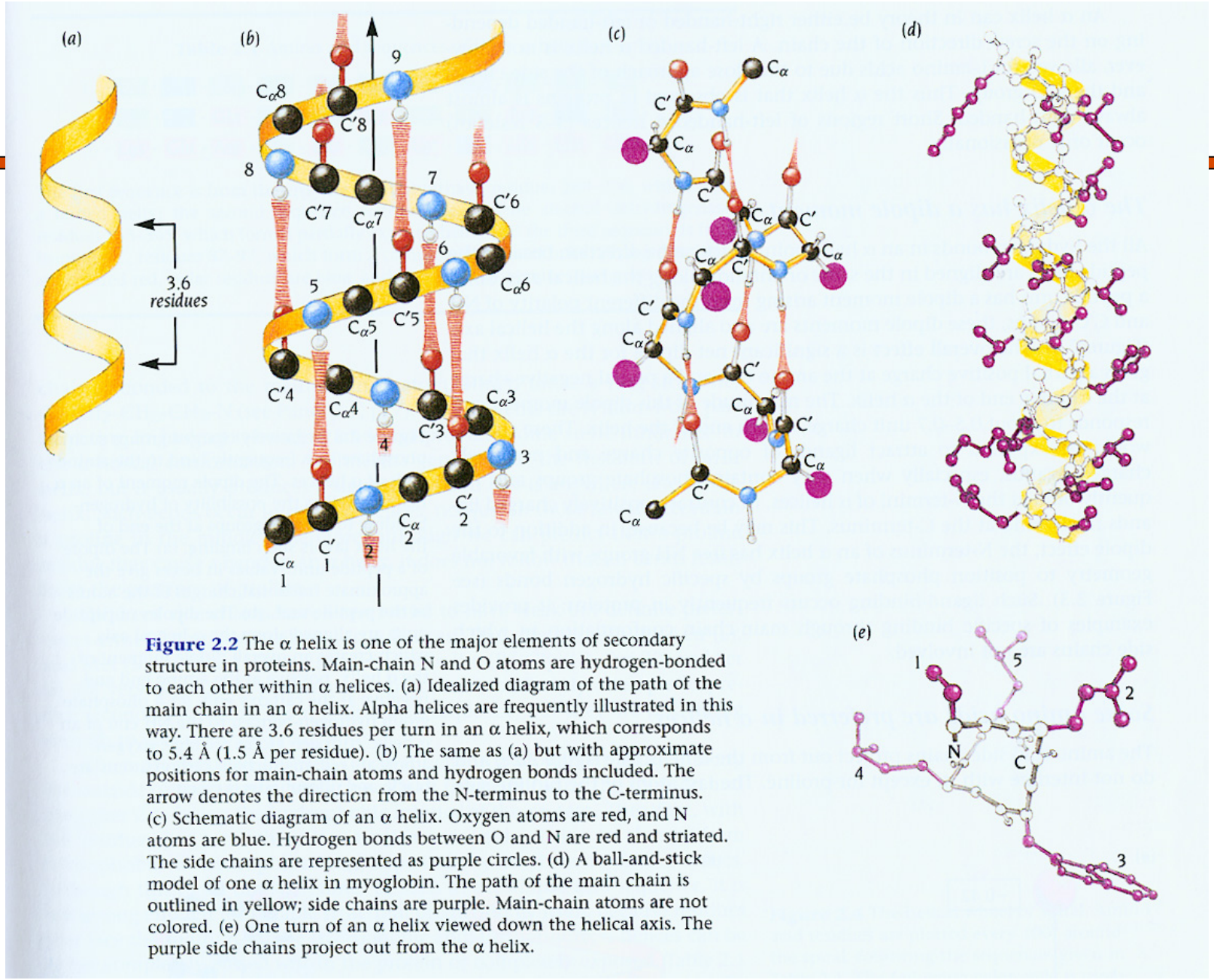
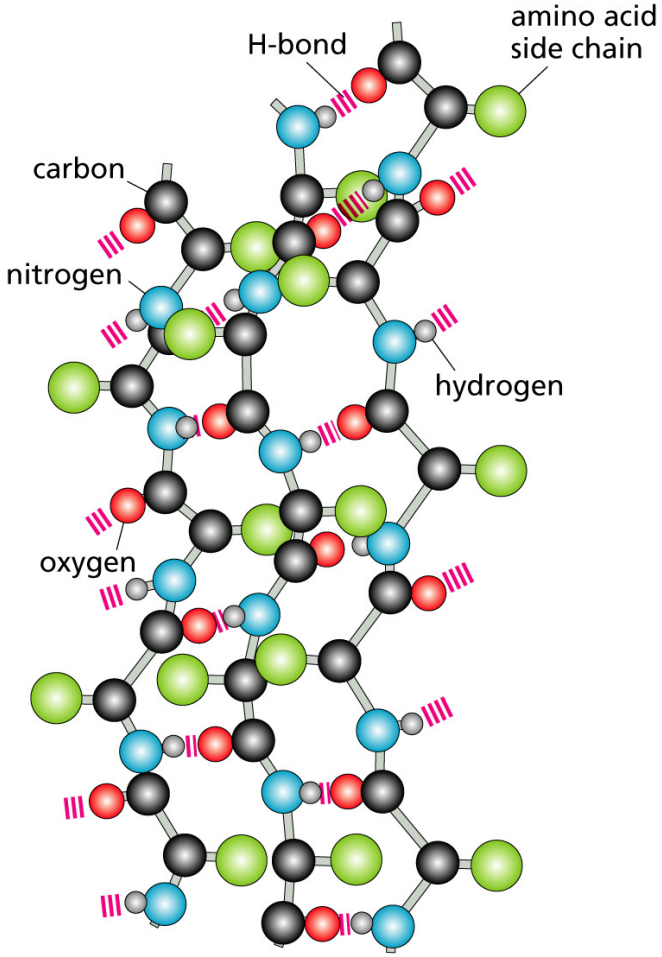


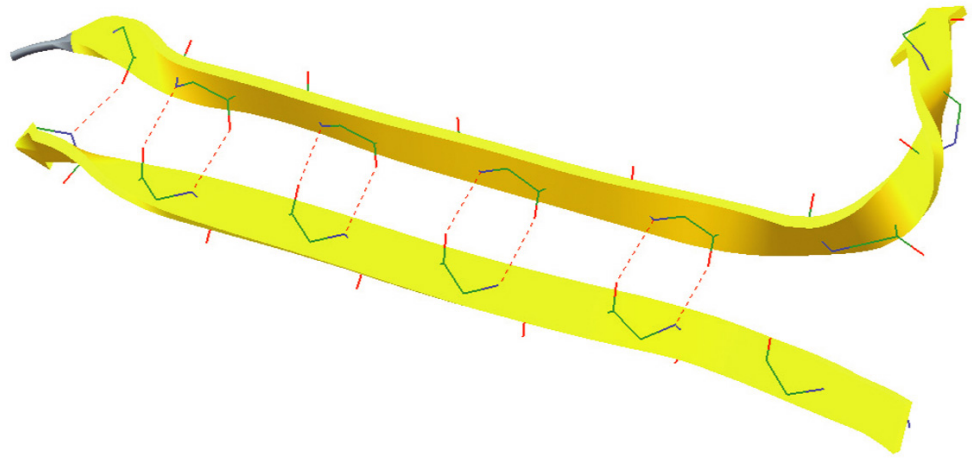
Figure 2.2 The α helix is one of the major elements of secondary structure in proteins. Main-chain N and O atoms are hydrogen-bonded to each other within α helices. (a) Idealized diagram of the path of the main chain in an α helix. Alpha helices are frequently illustrated in this way. There are 3.6 residues per turn in an α helix, which corresponds to 5.4 Å (1.5 Å per residue). (b) The same as (a) but with approximate positions for main-chain atoms and hydrogen bonds included. The arrow denotes the direction from the N-terminus to the C-terminus. (c) Schematic diagram of an α helix. Oxygen atoms are red, and N atoms are blue. Hydrogen bonds between O and N are red and striated. The side chains are represented as purple circles. (d) A ball-and-stick model of one α helix in myoglobin. The path of the main chain is outlined in yellow; side chains are purple. Main-chain atoms are not colored. (e) One turn of an α helix viewed down the helical axis. The purple side chains project out from the α helix.

Beta Strands and Sheets

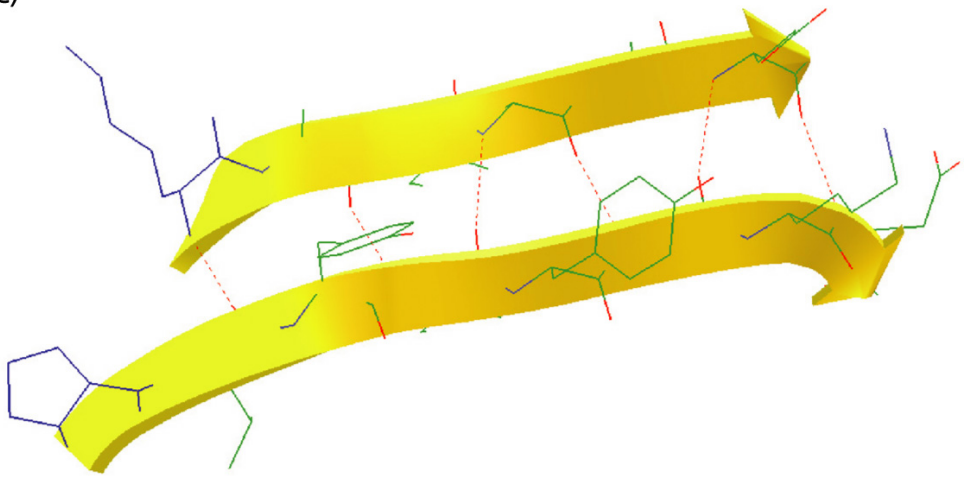
(A)



(B)



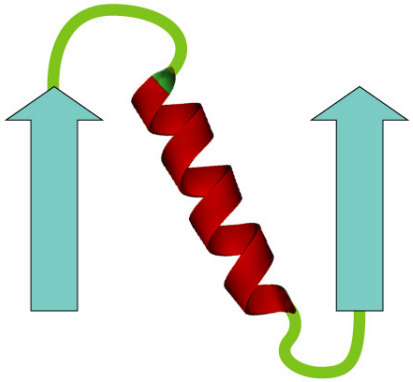
(C)



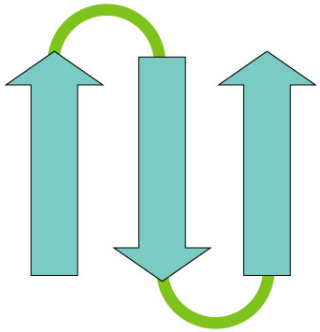
Supersecondary structures

(A)

βαβ repeat



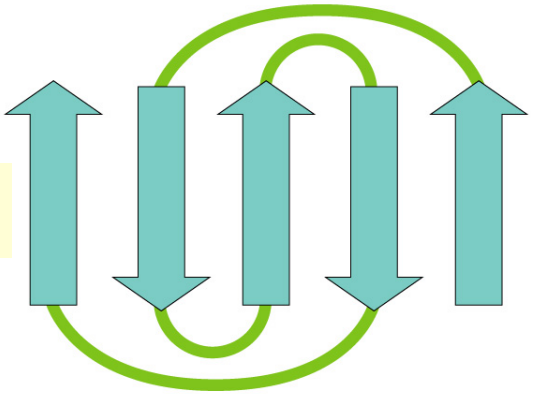
(B)



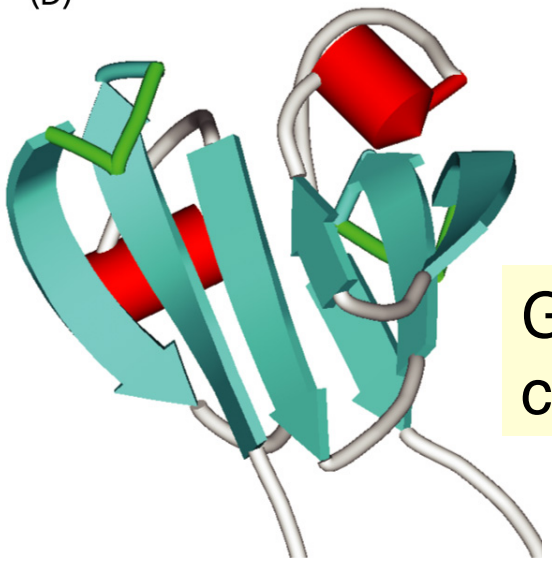
βαβ-meander

(C)

Greek Key



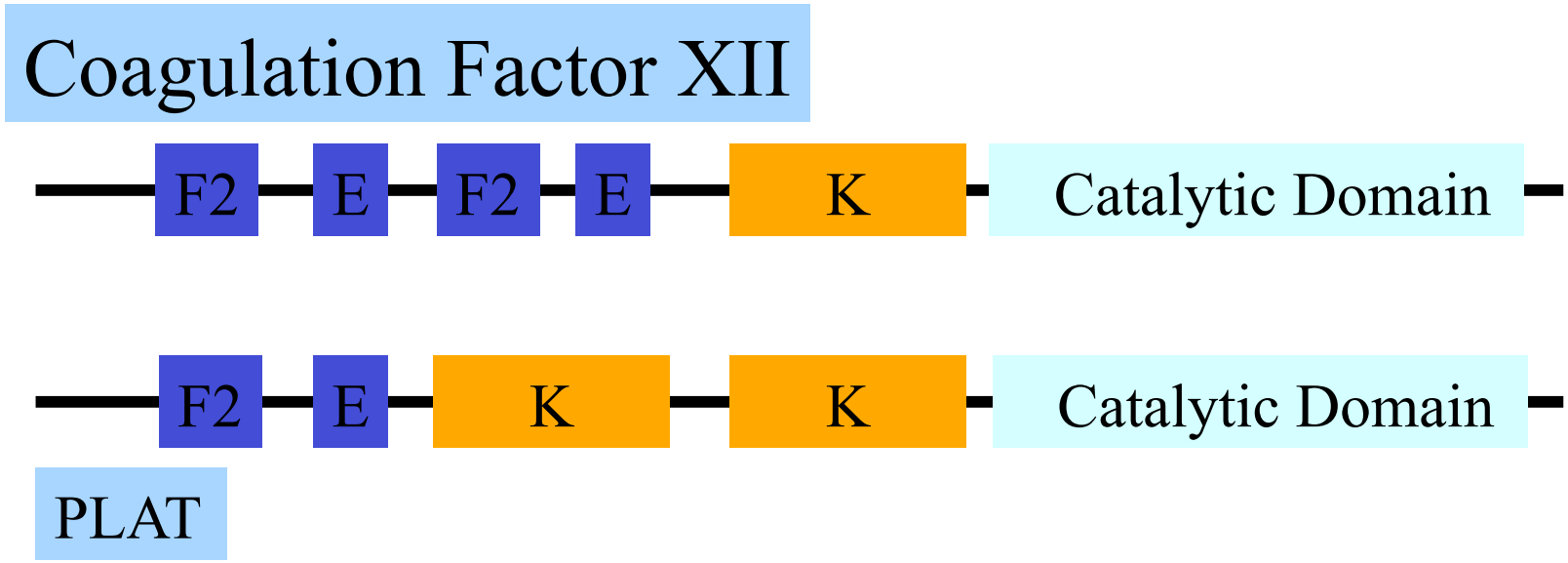
(D)



Gamma β crystallin

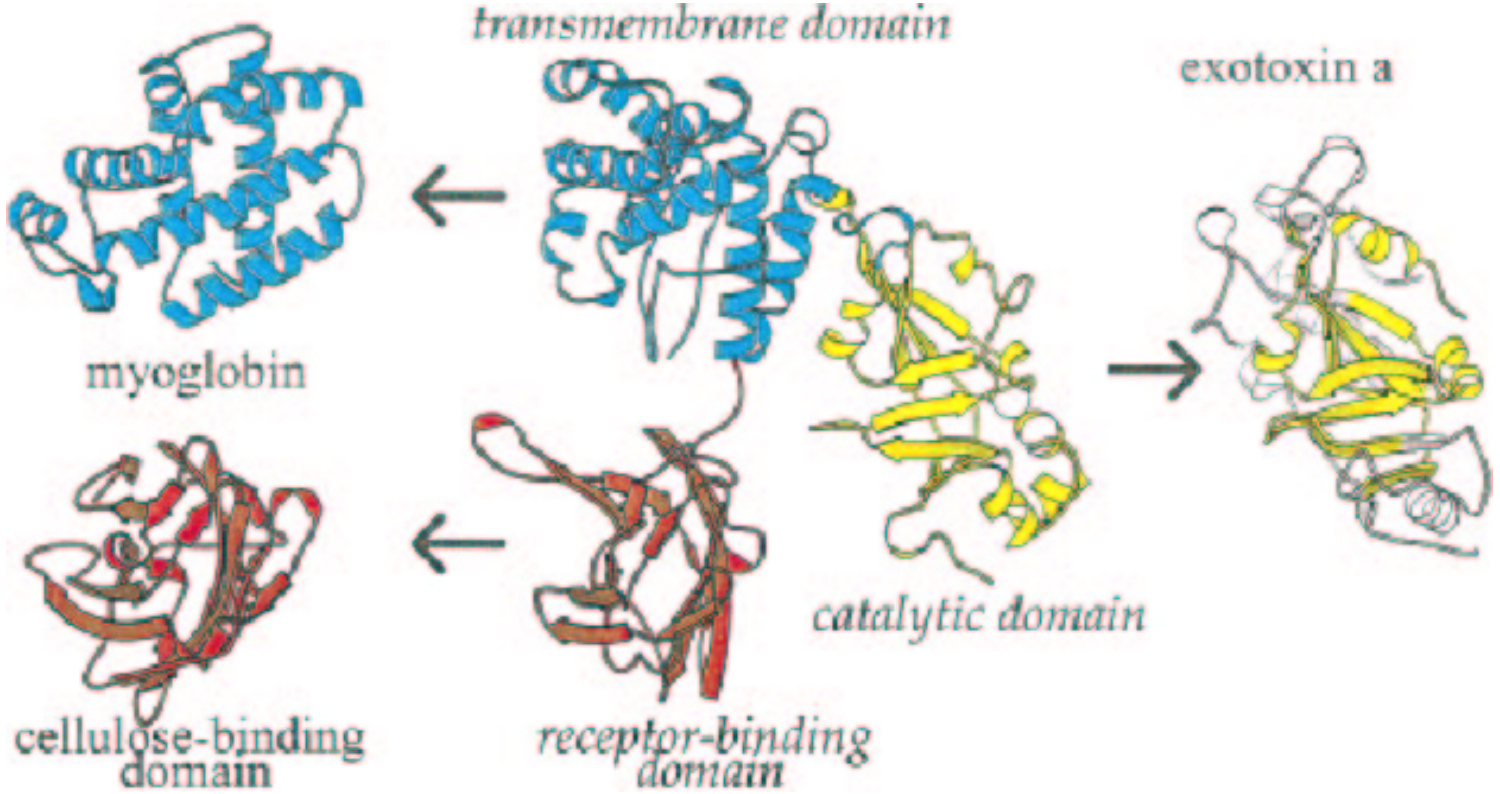
Modular Nature of Proteins

- Proteins are collections of "modular" domains. For example,



Modular Nature of Protein Structures

Example: Diphtheria Toxin



Domain Architecture Tools

- CDART
  - Protein AAH24495; Domain Architecture;
  - It's domain relatives;
  - Multiple alignment for 2nd domain
- SMART

Active Sites

Active sites in proteins are usually hydrophobic pockets/ crevices/troughs that involve sidechain atoms.

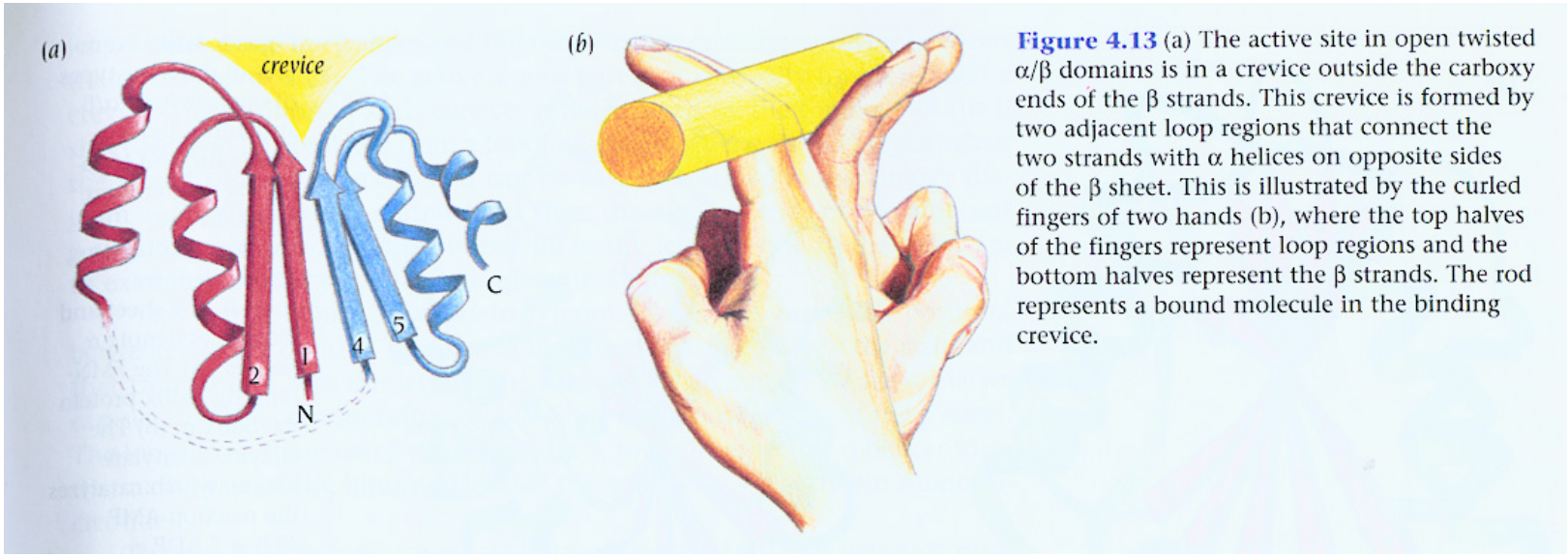
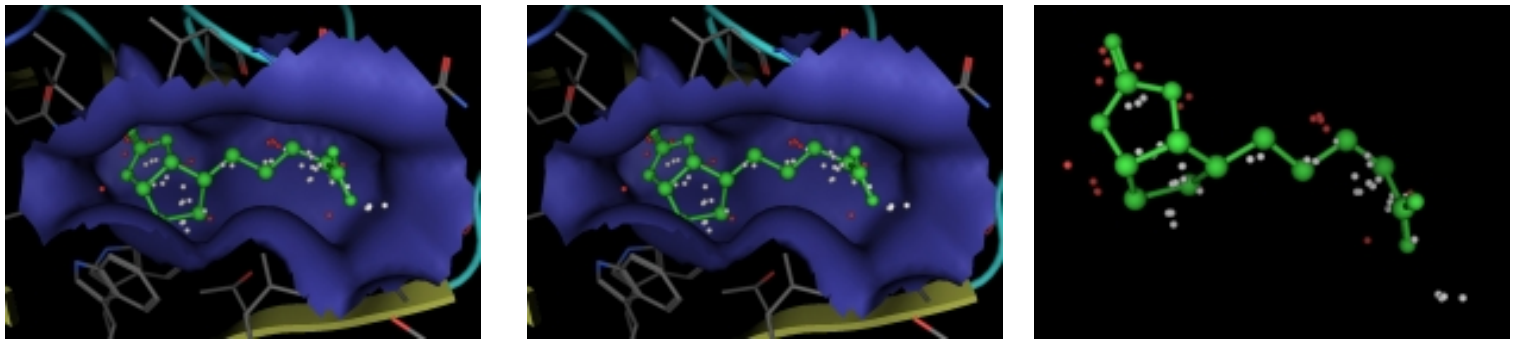


Figure 4.13 (a) The active site in open twisted α/β domains is in a crevice outside the carboxy ends of the β strands. This crevice is formed by two adjacent loop regions that connect the two strands with α helices on opposite sides of the β sheet. This is illustrated by the curled fingers of two hands (b), where the top halves of the fingers represent loop regions and the bottom halves represent the β strands. The rod represents a bound molecule in the binding crevice.

Active Sites



Left PDB 3RTD (streptavidin) and the first site located by the MOE Site Finder. **Middle** 3RTD with complexed ligand (biotin). **Right** Biotin ligand overlaid with calculated alpha spheres of the first site.

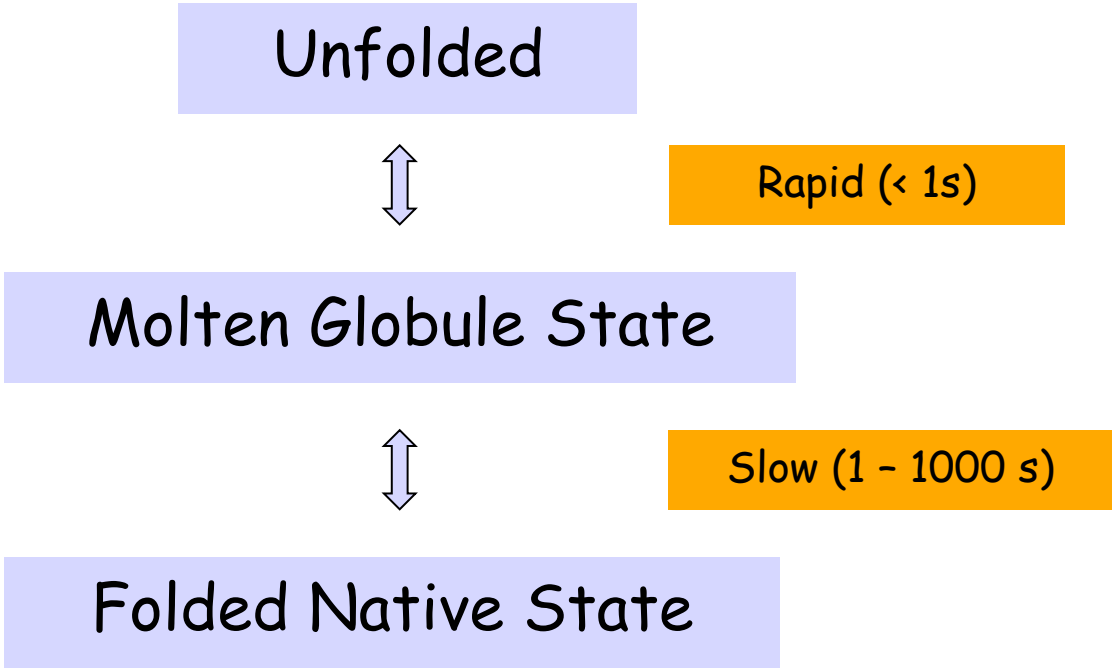
Secondary Structure Prediction Software

Figure 11.3 Comparison of secondary structure predictions by various methods. The sequence of flavodoxin, an α/β protein, was used as the query and is shown on the first line of the alignment. For each prediction, H denotes an α helix, E a β strand, T a β turn; all other positions are assumed to be random coil. Correctly assigned residues are shown in inverse type. The methods used are listed along the left side of the alignment and are described in the text. At the bottom of the figure is the secondary structure assignment given in the PDB file for flavodoxin (1OFV, Smith et al., 1983).

PDB: Protein Data Bank

- Database of protein tertiary and quaternary structures and protein complexes. http://www.rcsb.org/pdb/
- Over 29,000 structures as of Feb 1, 2005.
- Structures determined by
  - NMR Spectroscopy
  - X-ray crystallography
  - Computational prediction methods
- Sample PDB file: Click here [ ]

Protein Folding



- How to find minimum energy configuration?

Protein Structures

- Most proteins have a hydrophobic core.
- Within the core, specific interactions take place between amino acid side chains.
- Can an amino acid be replaced by some other amino acid?
  - Limited by space and available contacts with nearby amino acids
- Outside the core, proteins are composed of loops and structural elements in contact with water, solvent, other proteins and other structures.

Viewing Protein Structures

- SPDBV
- RASMOL
- CHIME

Secondary Structure Prediction Software

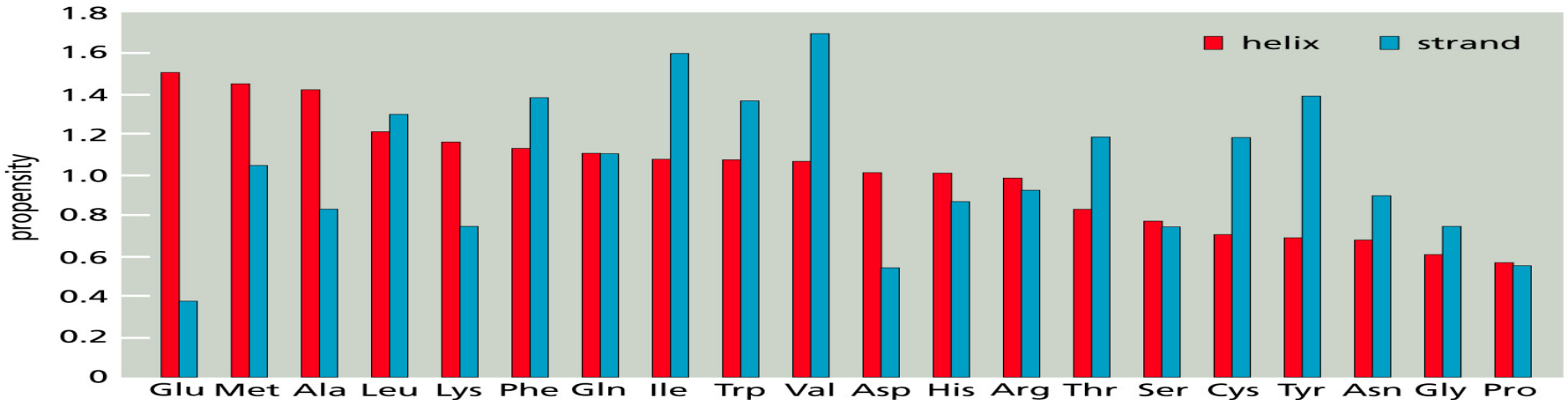
Figure 11.3 Comparison of secondary structure predictions by various methods. The sequence of flavodoxin, an α/β protein, was used as the query and is shown on the first line of the alignment. For each prediction, H denotes an α helix, E a β strand, T a β turn; all other positions are assumed to be random coil. Correctly assigned residues are shown in inverse type. The methods used are listed along the left side of the alignment and are described in the text. At the bottom of the figure is the secondary structure assignment given in the PDB file for flavodoxin (1OFV, Smith et al., 1983).

Recent Ones:

GOR V
PREDATOR
Zpred
PROF
NNSSP
PHD
PSIPRED
Jnet

Chou & Fasman Propensities

Amino Acid	helix			
	Designation	*P*	Designation	*P*
Ala	F	1.42	b	0.83
Cys	I	0.70	f	1.19
Asp	I	1.01	B	0.54
Glu	F	1.51	B	0.37
Phe	f	1.13	f	1.38
Gly	B	0.61	b	0.75
His	f	1.00	f	0.87
Ile	f	1.08	F	1.60
Lys	f	1.16	b	0.74
Leu	F	1.21	f	1.30
Met	F	1.45	f	1.05
Asn	b	0.67	b	0.89
Pro	**B**	**0.57**	**B**	**0.55**
Gln	f	1.11	h	1.10
Arg	I	0.98	I	0.93
Ser	I	0.77	b	0.75
Thr	I	0.83	f	1.19
Val	f	1.06	F	1.70
Trp	f	1.08	f	1.37
Tyr	b	0.69	F	1.4



GOR IV prediction for 1bbc

```
AFAGVLNDADIAAALEACKAADSFNHKAFFAKVGLTSKSADDVKKAFAII
CCCCCCCHHHHHHHHHHHHHHCCCCCHHHHEEECCCCCCHHHHHHHHHHH
AQDKSGFIEEDELKLFLQNFKADARALTDGETKTFLKAGDSDGDGKIGVD
HHCCCCCHHHHHHHHHHHHHHHHHHHCCCCCEEEEEECCCCCCCCEEECC
DVTALVKA
CEEEEEEC
```

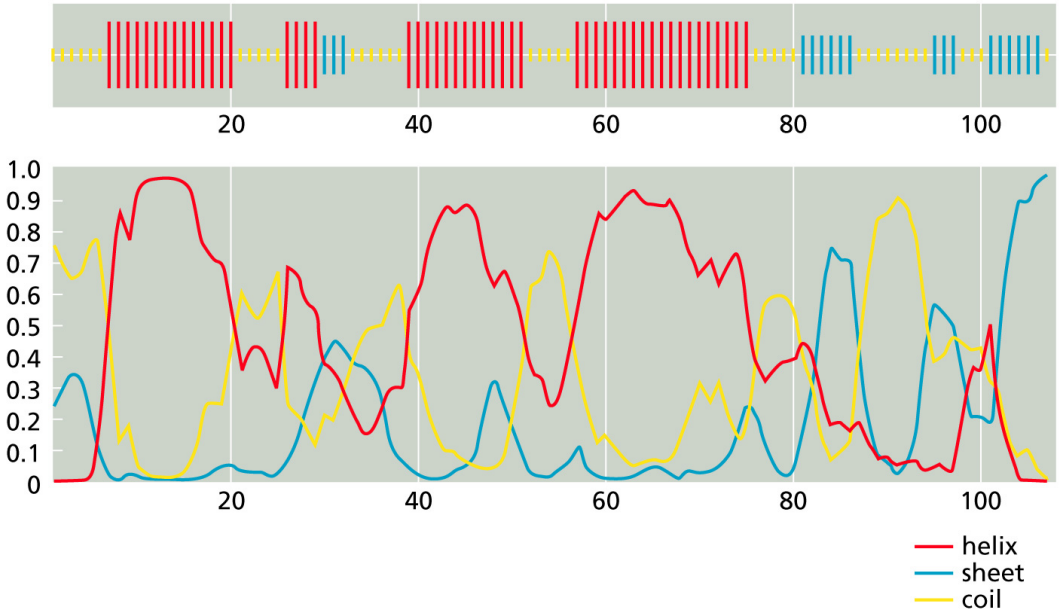
sequence length: 108

GOR IV:

alpha helix (Hh) : 50 is 46.30%

beta sheet (Ee) : 18 is 16.67%

random coil (Cc) : 40 is 37.04%

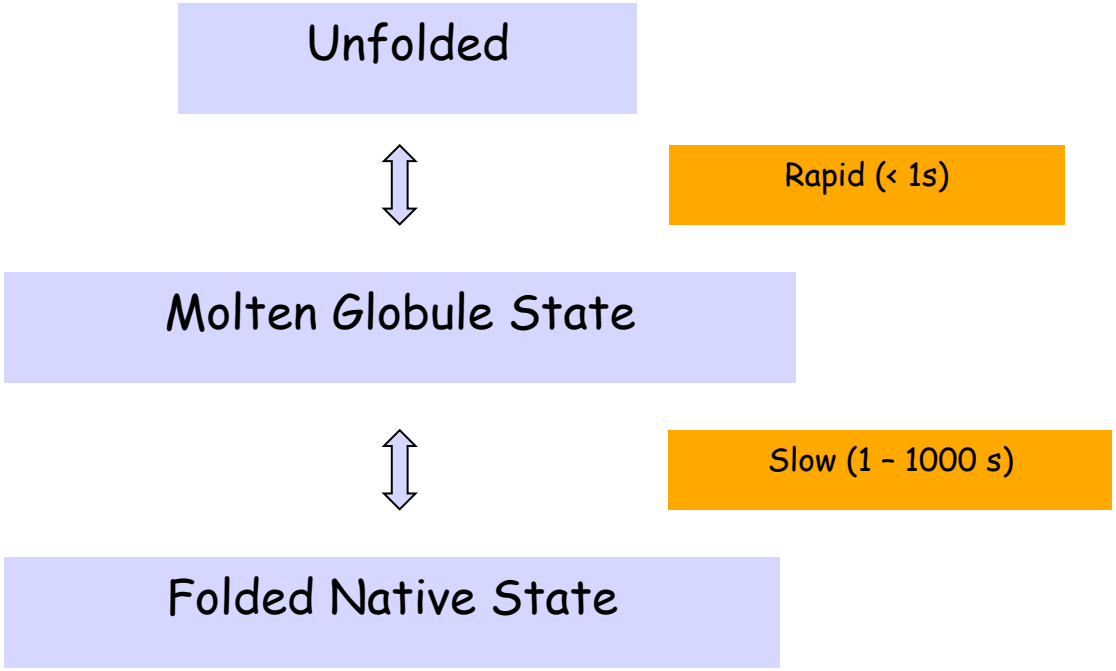


PDB: Protein Data Bank

- Database of protein tertiary and quaternary structures and protein complexes. http://www.rcsb.org/pdb/
- Over 29,000 structures as of Feb 1, 2005.
- Structures determined by
  - NMR Spectroscopy
  - X-ray crystallography
  - Computational prediction methods
- Sample PDB file: Click here [ ]

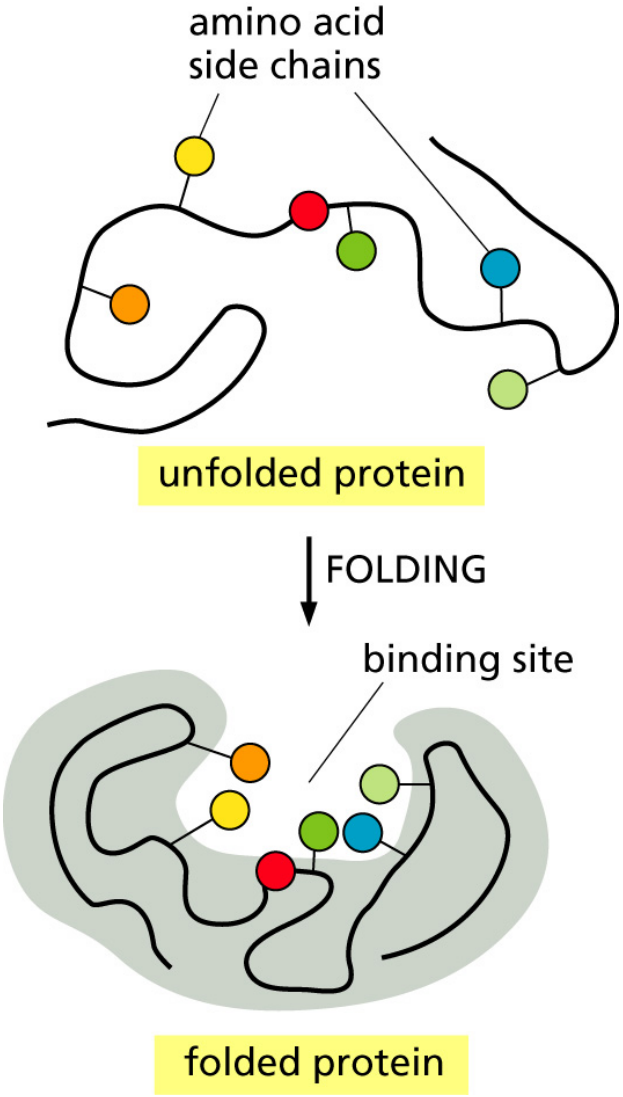
PDB Search Results

Protein Folding

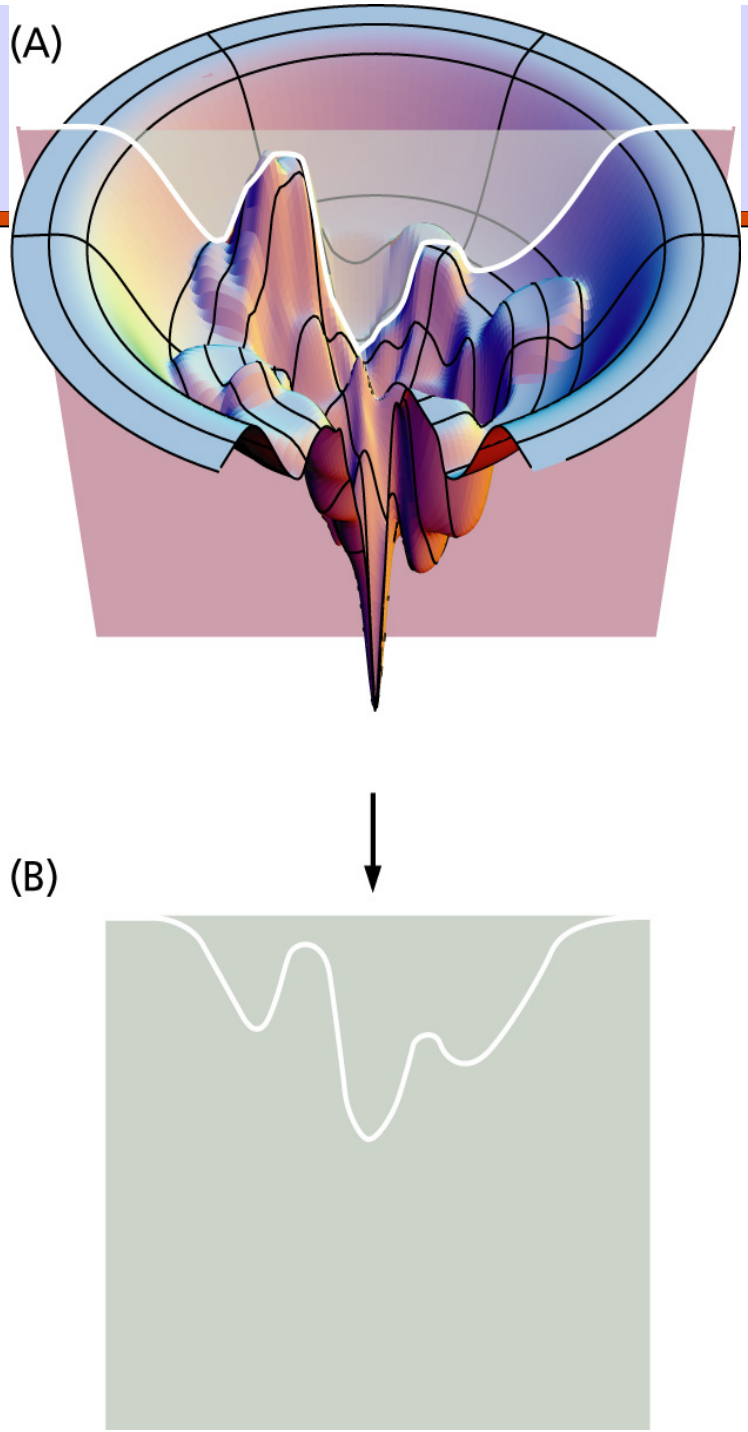


- How to find minimum energy configuration?

Protein Folding



Energy Landscape



Protein Structures

- Most proteins have a hydrophobic core.
- Within the core, specific interactions take place between amino acid side chains.
- Can an amino acid be replaced by some other amino acid?
  - Limited by space and available contacts with nearby amino acids
- Outside the core, proteins are composed of loops and structural elements in contact with water, solvent, other proteins and other structures.

Viewing Protein Structures

- SPDBV
- RASMOL
- CHIME

Structural Alignment

- What is structural alignment of proteins?
  - 3-d superimposition of the atoms as "best as possible", i.e., to minimize RMSD (root mean square deviation).
  - Can be done using VAST and SARF
- Structural similarity is common, even among proteins that do not share sequence similarity or evolutionary relationship.

Other databases & tools

- MMDB contains groups of structurally related proteins
- SARF structurally similar proteins using secondary structure elements
- VAST Structure Neighbors
- SSAP uses double dynamic programming to structurally align proteins

Protein Structure Prediction

- Holy Grail of bioinformatics
- Protein Structure Initiative to determine a set of protein structures that span protein structure space sufficiently well. WHY?
  - Number of folds in natural proteins is limited. Thus a newly discovered proteins should be within modeling distance of some protein in set.
- CASP: Critical Assessment of techniques for structure prediction
  - To stimulate work in this difficult field

PSP Methods

- homology-based modeling
- methods based on fold recognition
  - Threading methods
- *ab initio* methods
  - From first principles
  - With the help of databases

ROSETTA

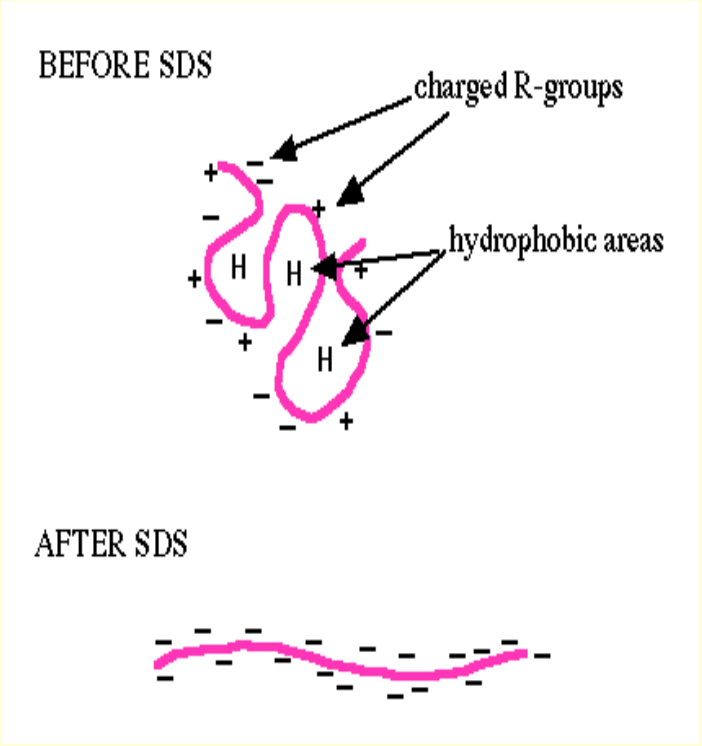
- Best method for PSP
- As proteins fold, a large number of partially folded, low-energy conformations are formed, and that local structures combine to form more global structures with minimum energy.
- Build a database of known structures (I-sites) of short sequences (3-15 residues).
- Monte Carlo simulation assembling possible substructures and computing energy

Modeling Servers

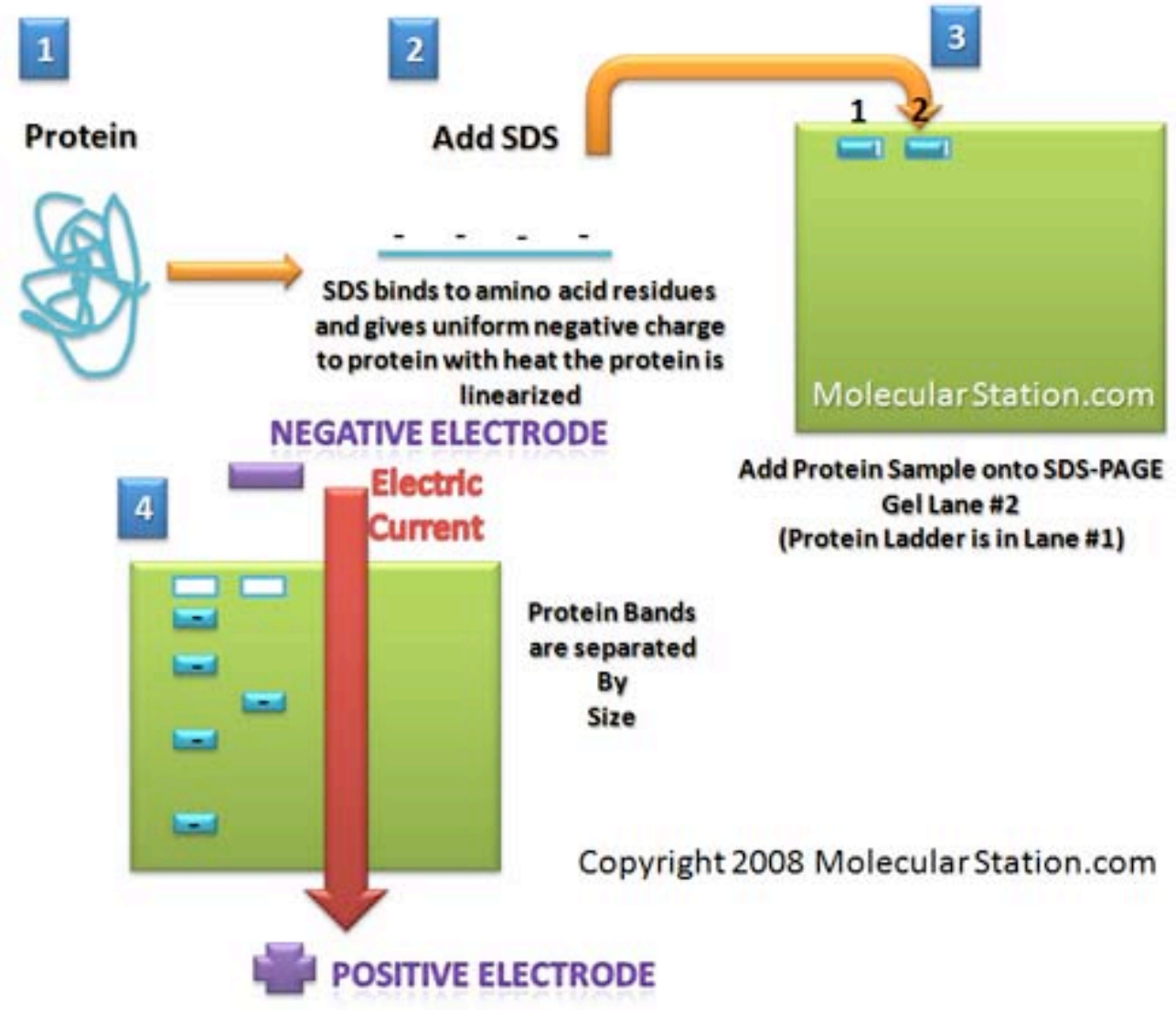
- SwissMODEL
- 3DJigsaw
- CPHModel
- ESyPred3D
- Geno3D
- SDSC1
- Rosetta
- MolIDE
- SCWRL
- PSIPred
- MODELLER
- LOOPY

Gel Electrophoresis for Protein

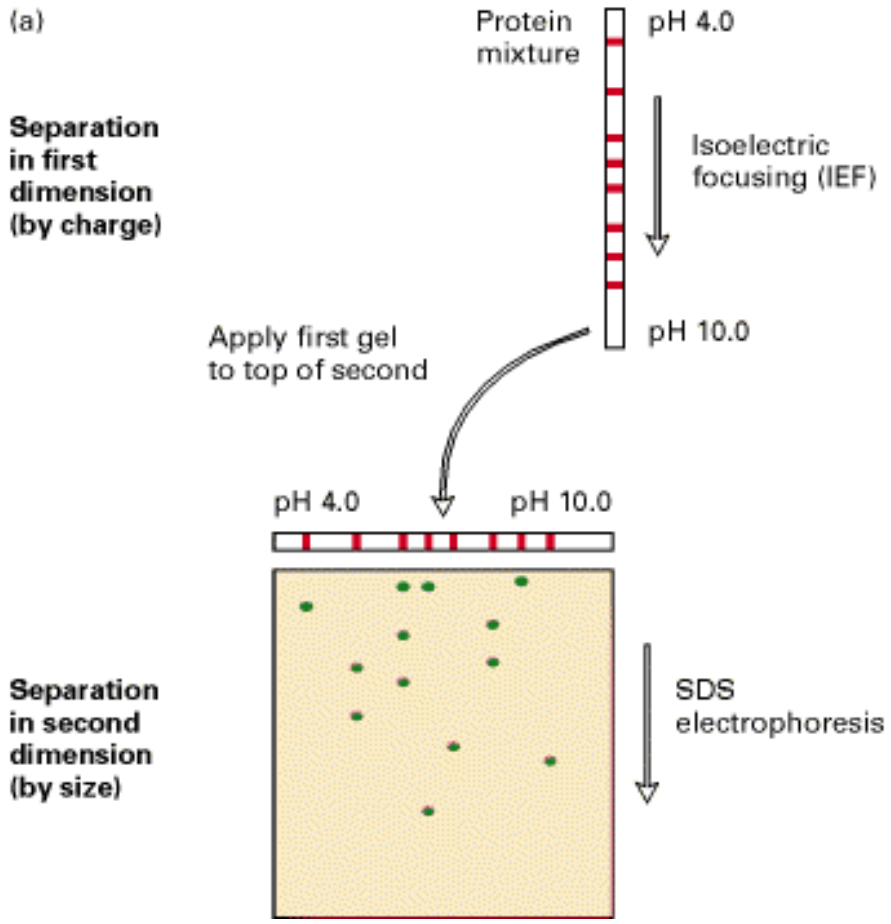
- Protein is also charged
- Has to be denatured - WHY
- Gel: SDS-Polyacrylamide gels
- Add sample to well
- Apply voltage
- Size determines speed
- Add dye to assess the speed
- Stain to see the protein bands



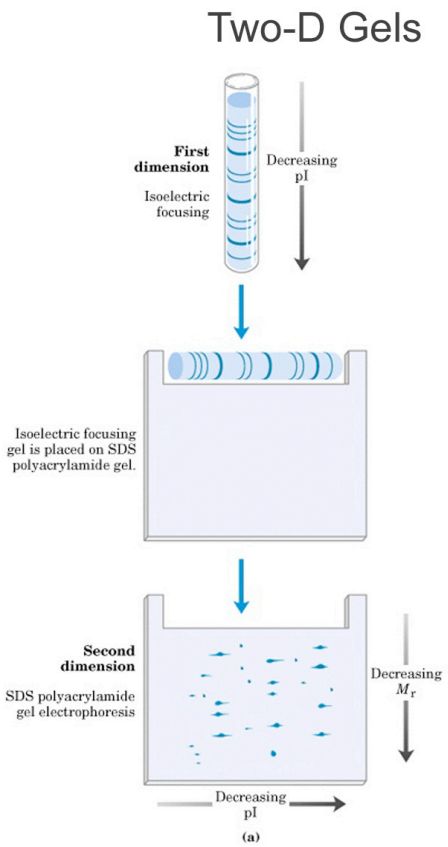
Protein Gel


1
Protein
2
Add SDS
SDS binds to amino acid residues and gives uniform negative charge to protein with heat the protein is linearized
3
1
2
MolecularStation.com
Add Protein Sample onto SDS-PAGE Gel Lane #2
(Protein Ladder is in Lane #1)
NEGATIVE ELECTRODE
4
Electric Current
Protein Bands are separated By Size
Copyright 2008 MolecularStation.com
POSITIVE ELECTRODE

2D-Gels


(a)
Protein mixture
pH 4.0
Separation in first dimension (by charge)
Isoelectric focusing (IEF)
pH 10.0
Apply first gel to top of second
pH 4.0
pH 10.0
Separation in second dimension (by size)
SDS electrophoresis

2D Gel Electrophoresis



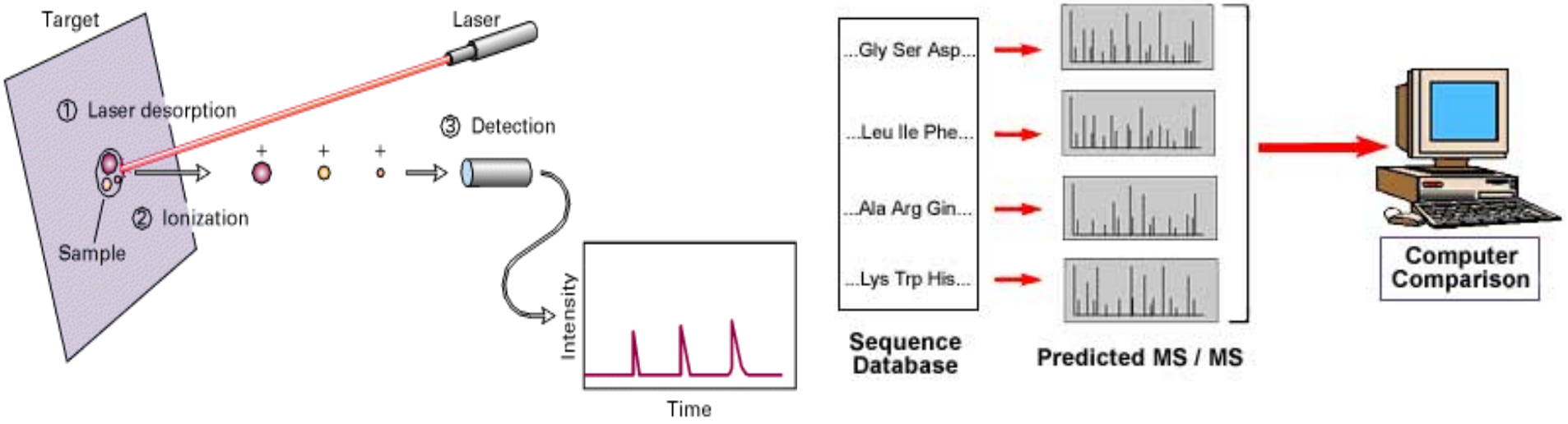
(b)

Mass Spectrometry

- **Mass measurements By Time-of-Flight**
  Pulses of light from laser ionizes protein that is absorbed on metal target. Electric field accelerates molecules in sample towards detector. The time to the detector is inversely proportional to the mass of the molecule. Simple conversion to mass gives the molecular weights of proteins and peptides.

- **Using Peptide Masses to Identify Proteins:**
  One powerful use of mass spectrometers is to identify a protein from its peptide mass fingerprint. A peptide mass fingerprint is a compilation of the molecular weights of peptides generated by a specific protease. The molecular weights of the parent protein prior to protease treatment and the subsequent proteolytic fragments are used to search genome databases for any similarly sized protein with identical or similar peptide mass maps. The increasing availability of genome sequences combined with this approach has almost eliminated the need to chemically sequence a protein to determine its amino acid sequence.

Mass Spectrometry



Protein Sequence

- 20 amino acids
- How is it ordered?
- Basis: Edman Degradation (Pehr Edman)
  - Limited ~30 residues
  - React with Phenylisothiocyanate
  - Cleave and chromatography
- First separate the proteins – Use 2D gels
- Then digest to get pieces
- Then sequence the smaller pieces
- Tedious
- Mass spectrometry