

CAP 5510: Introduction to Bioinformatics

Giri Narasimhan

ECS 254; Phone: x3748

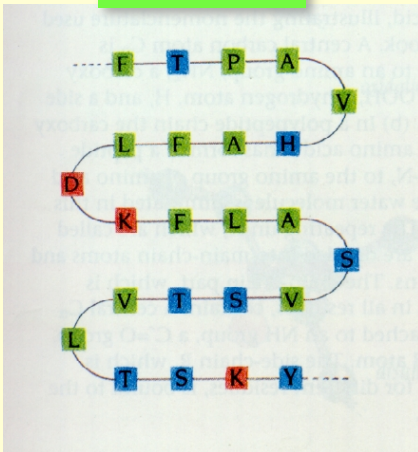
giri@cis.fiu.edu

www.cis.fiu.edu/~giri/teach/BioinfS11.html

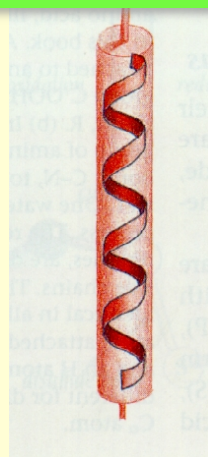
Protein Structures

- Sequences of amino acid residues
- 20 different amino acids

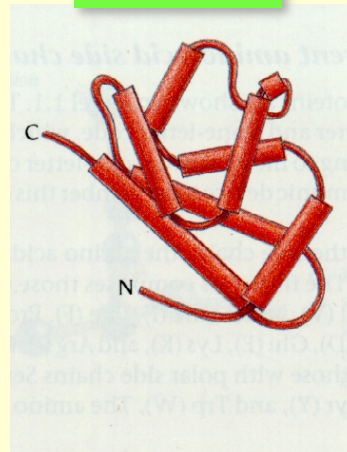
Primary



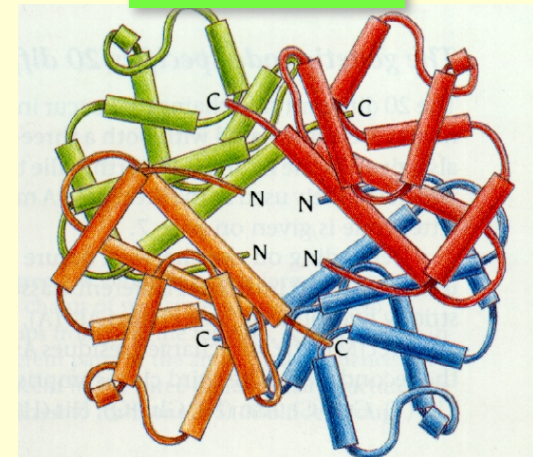
Secondary



Tertiary



Quaternary

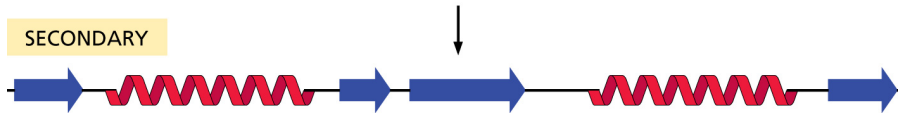


Proteins: Levels of Description

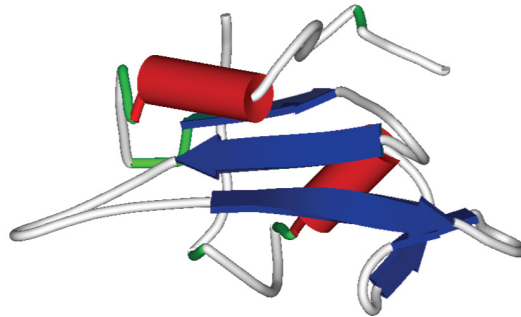
PRIMARY

N terminus-...MYCATISEATINGFISHANDMEATANDWATER...-C terminus

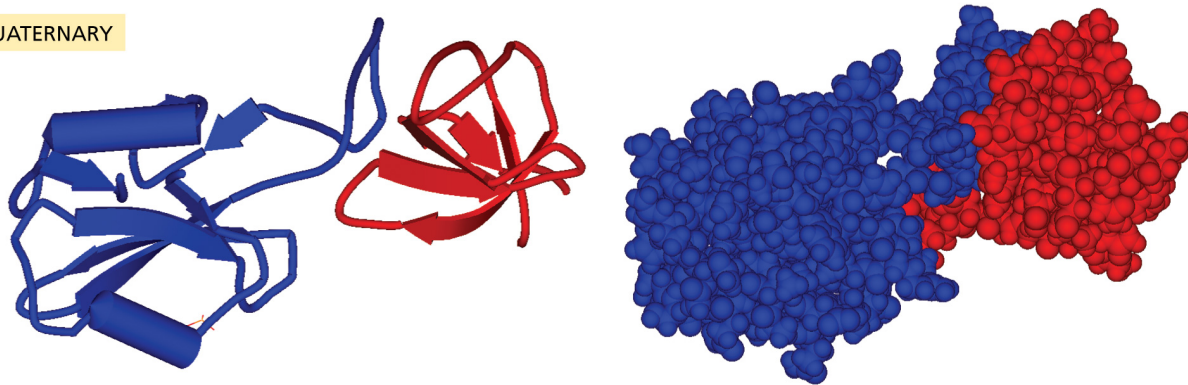
SECONDARY



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...`



Secondary



More on Secondary Structures

□ α -helix

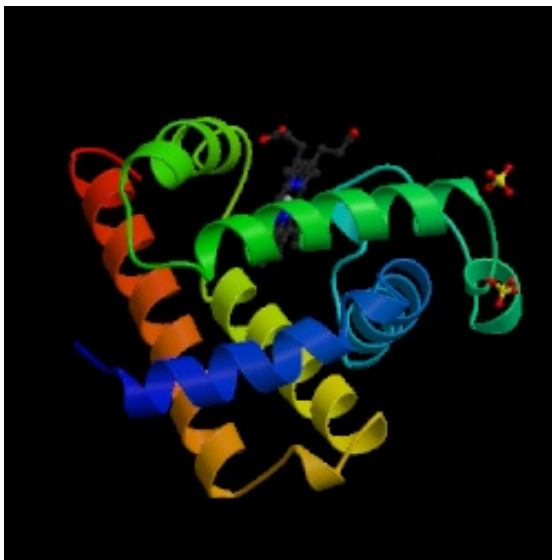
- Main chain with peptide bonds
- Side chains project outward from helix
- Stability provided by H-bonds between CO and NH groups of residues 4 locations away.

□ β -strand

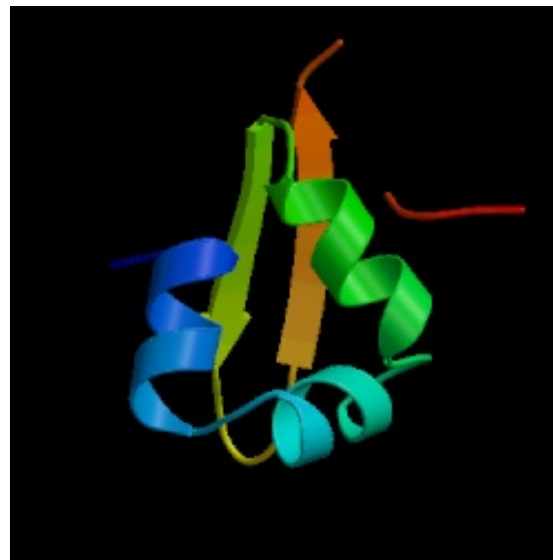
- Stability provided by H-bonds with one or more β -strands, forming β -sheets. Needs a β -turn.

Proteins

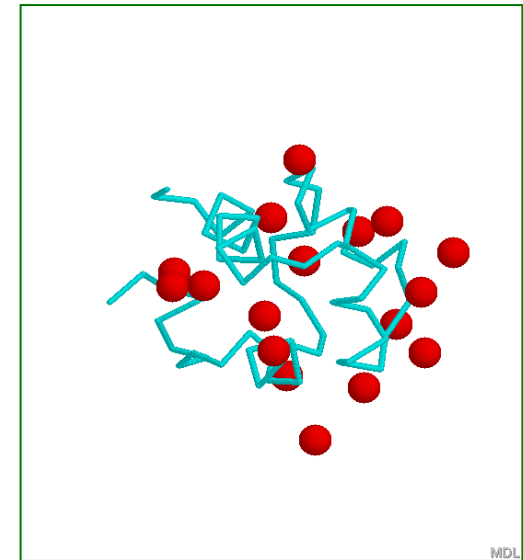
- **Tertiary structures** are formed by packing secondary structural elements into a globular structure.



Myoglobin



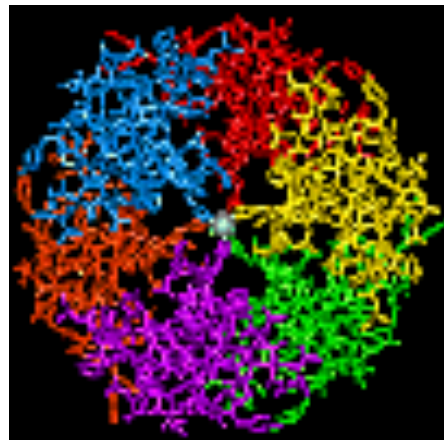
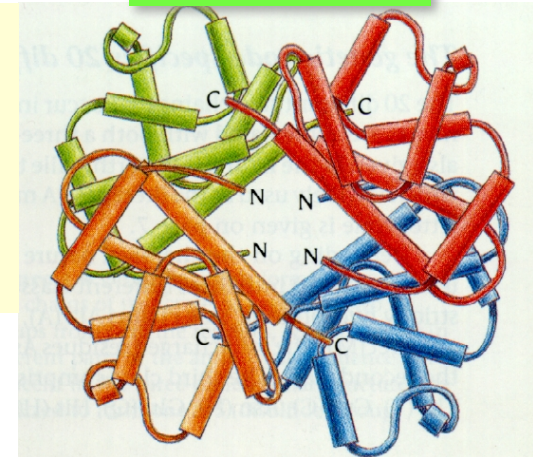
Lambda Cro



Quaternary Structures in Proteins

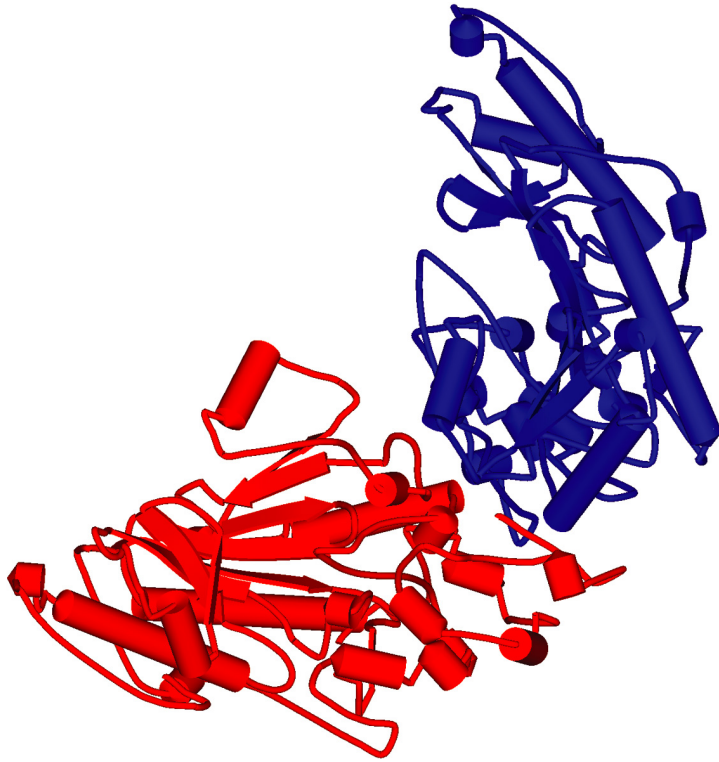
- The final structure may contain more than one “chain” arranged in a **quaternary structure**.

Quaternary



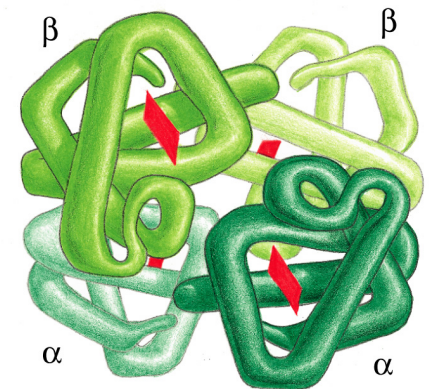
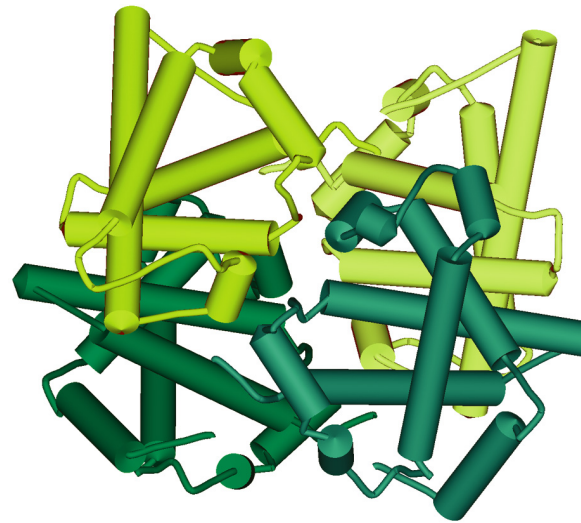
Insulin Hexamer

More quaternary structures



Muscle creatine kinase
(Homodimer)

Bovine deoxyhemoglobin
(Heterotetramer)



Amino Acid Types

Hydrophobic **I, L, M, V, A, F, P**

Charged

Basic **K, H, R**

Acidic **E, D**

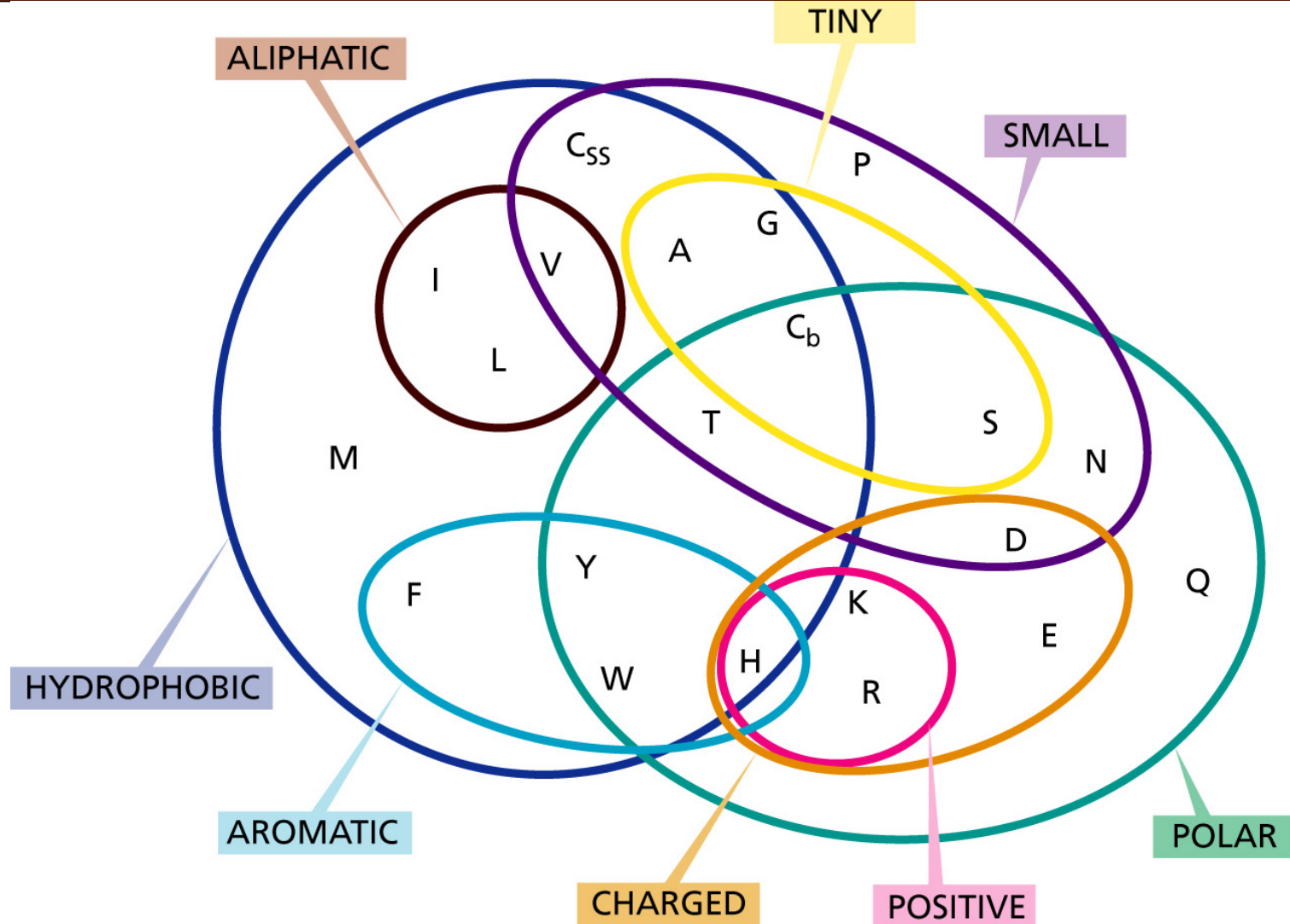
Polar **S, T, Y, H, C, N, Q, W**

Small **A, S, T**

Very Small **A, G**

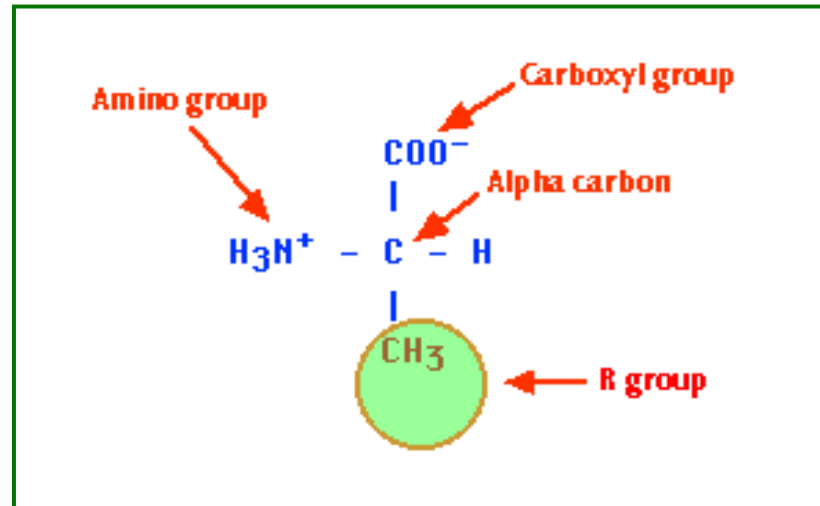
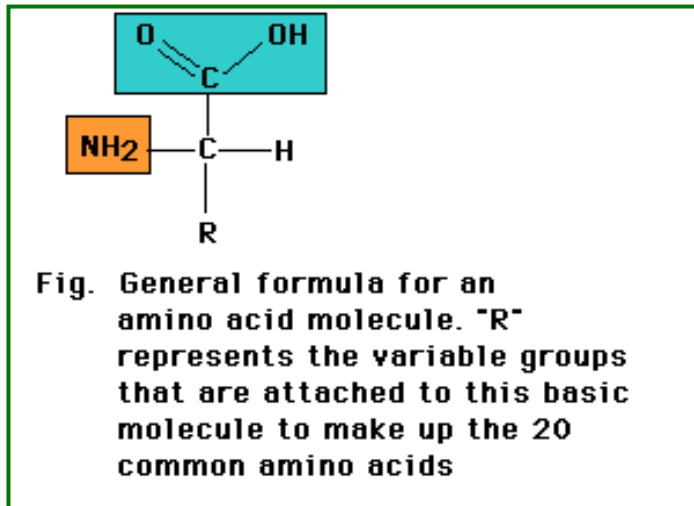
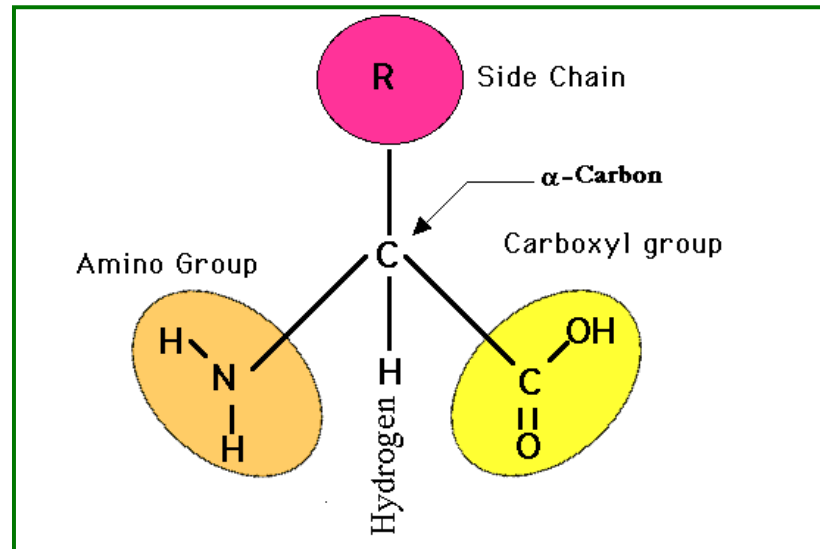
Aromatic **F, Y, W**

Amino Acid Types

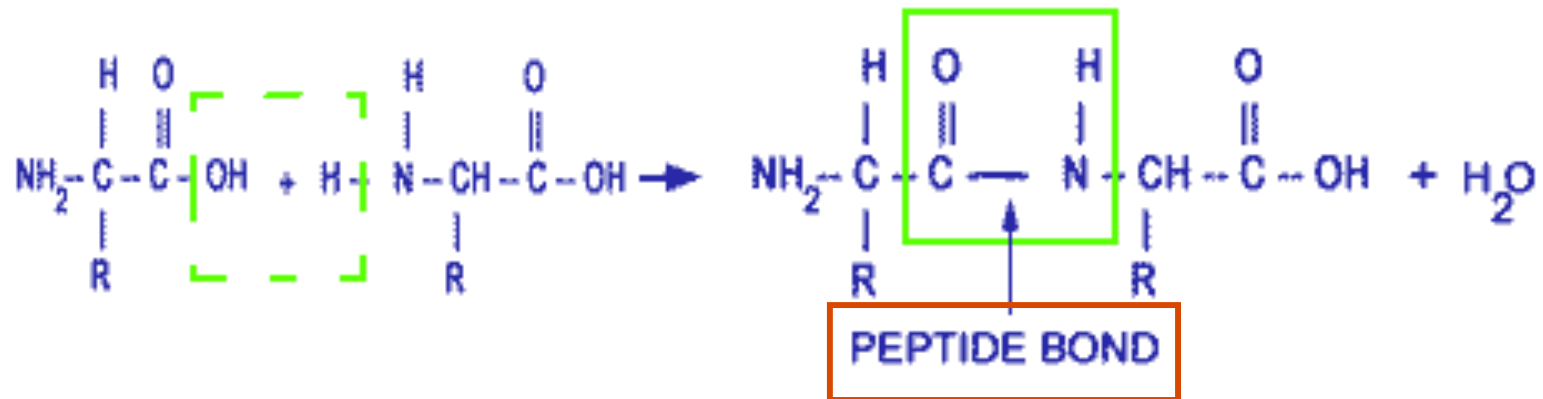


Structure of a single amino acid

All 3 figures are cartoons of an amino acid residue.



Chains of amino acids



Amino acids vs **Amino acid residues**

Angles ϕ and ψ in the polypeptide chain

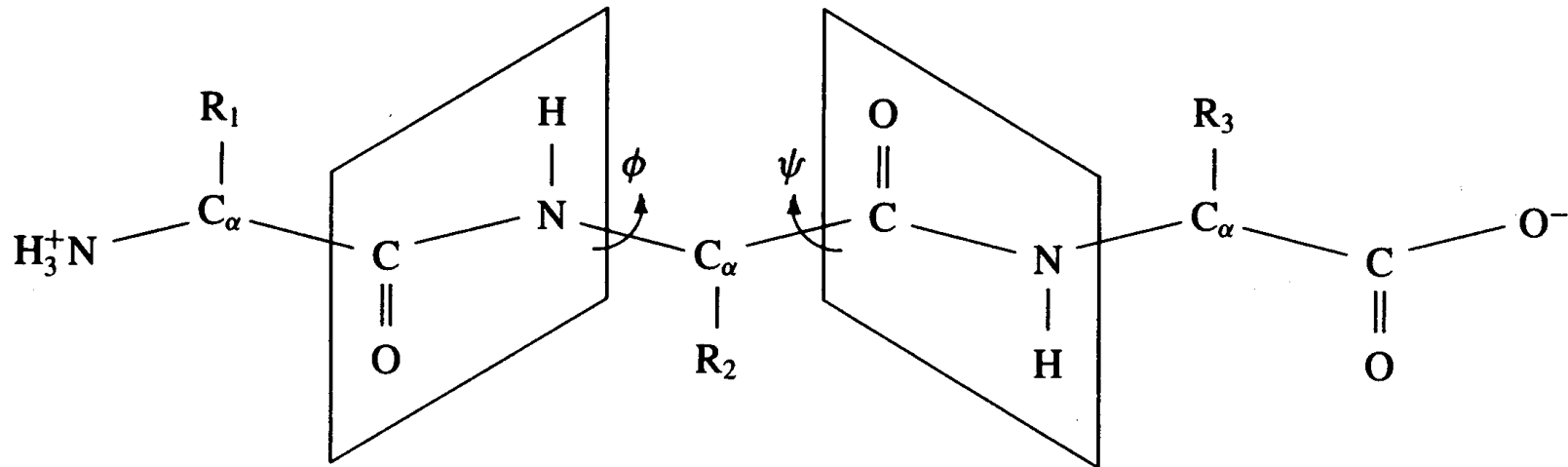


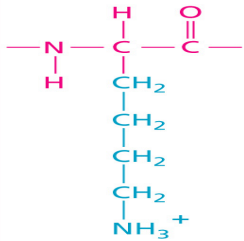
FIGURE 1.2

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 .

BASIC SIDE CHAINS

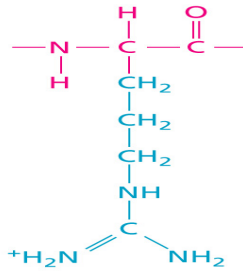
lysine

(Lys, or K)



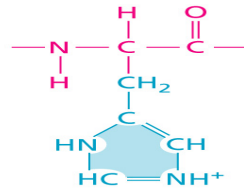
arginine

(Arg, or R)



histidine

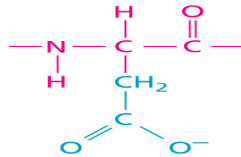
(His, or H)



ACIDIC SIDE CHAINS

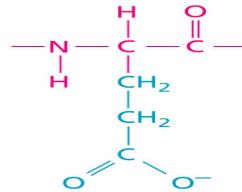
aspartic acid

(Asp, or D)



glutamic acid

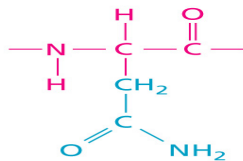
(Glu, or E)



UNCHARGED POLAR SIDE CHAINS

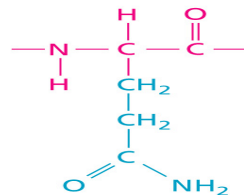
asparagine

(Asn, or N)



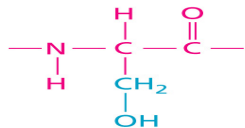
glutamine

(Gln, or Q)



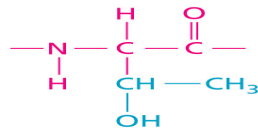
serine

(Ser, or S)



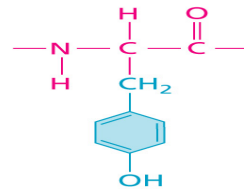
threonine

(Thr, or T)



tyrosine

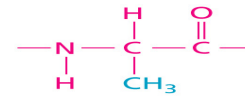
(Tyr, or Y)



NONPOLAR SIDE CHAINS

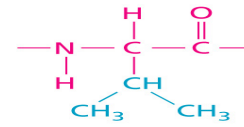
alanine

(Ala, or A)



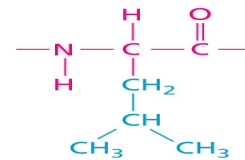
valine

(Val, or V)



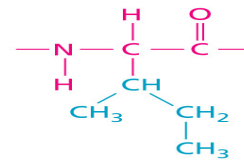
leucine

(Leu, or L)



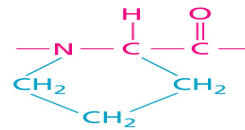
isoleucine

(Ile, or I)



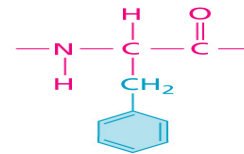
proline

(Pro, or P)



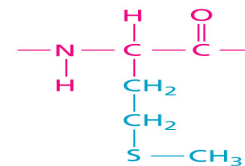
phenylalanine

(Phe, or F)



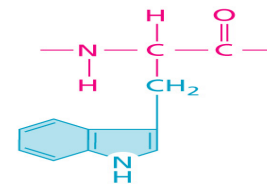
methionine

(Met, or M)



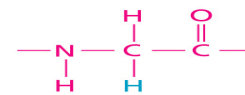
tryptophan

(Trp, or W)



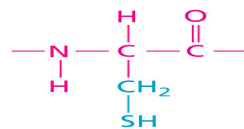
glycine

(Gly, or G)

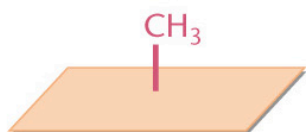


cysteine

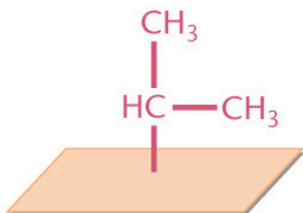
(Cys, or C)



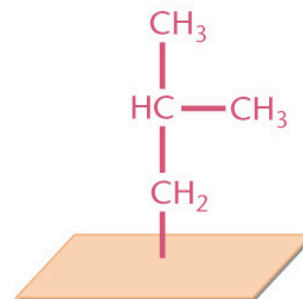
1. Nonpolar: Hydrophobic



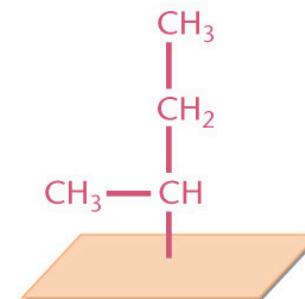
Alanine (ala-A)



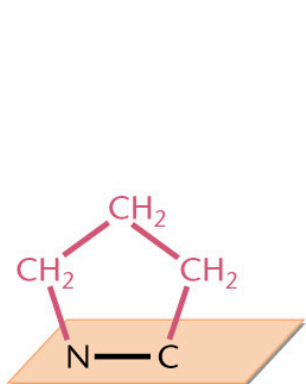
Valine (val-V)



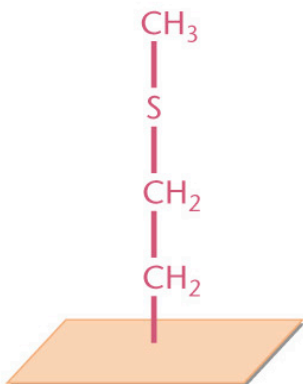
Leucine (leu-L)



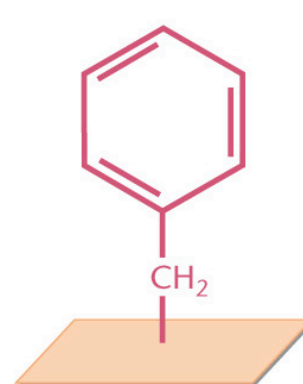
Isoleucine (ile-I)



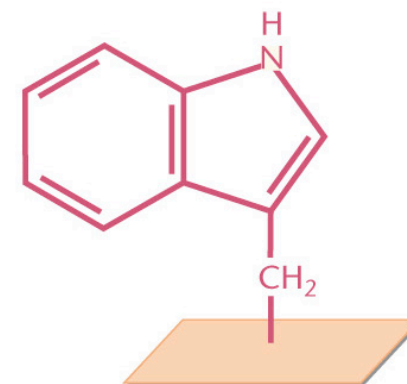
Proline (pro-P)



Methionine (met-M)



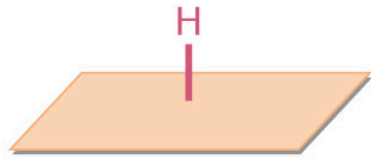
Phenylalanine (phe-F)



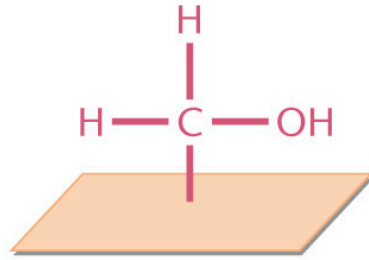
Tryptophan (trp-W)

Amino Acid Structures from Klug & Cummings

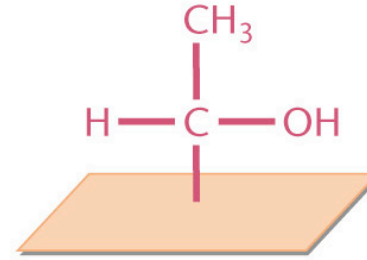
2. Polar: Hydrophilic



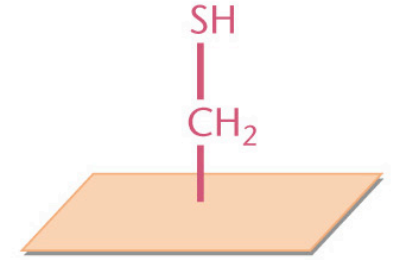
Glycine (gly-G)



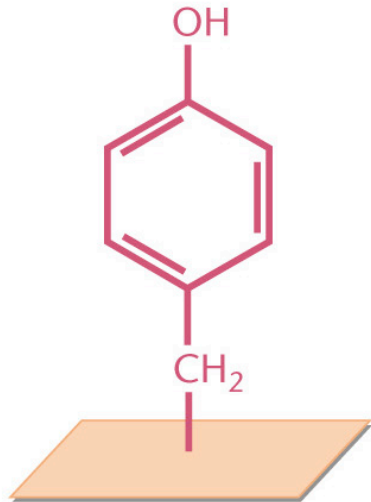
Serine (ser-S)



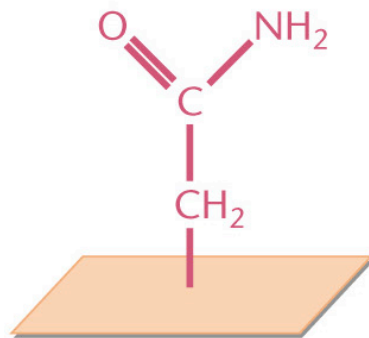
Threonine (thr-T)



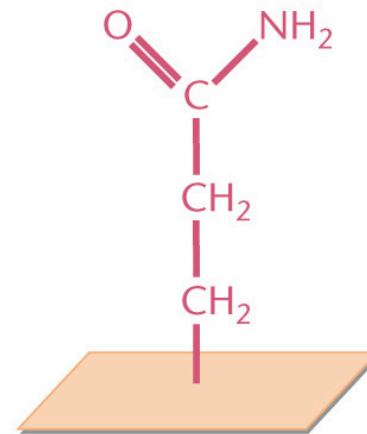
Cysteine (cys-C)



Tyrosine (tyr-Y)



Asparagine (asn-N)

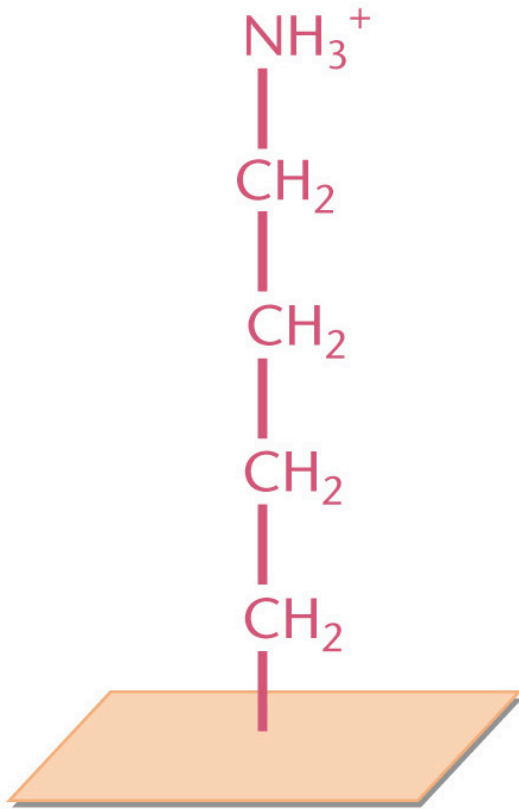


Glutamine (gln-Q)

Amino Acid Structures from Klug & Cummings

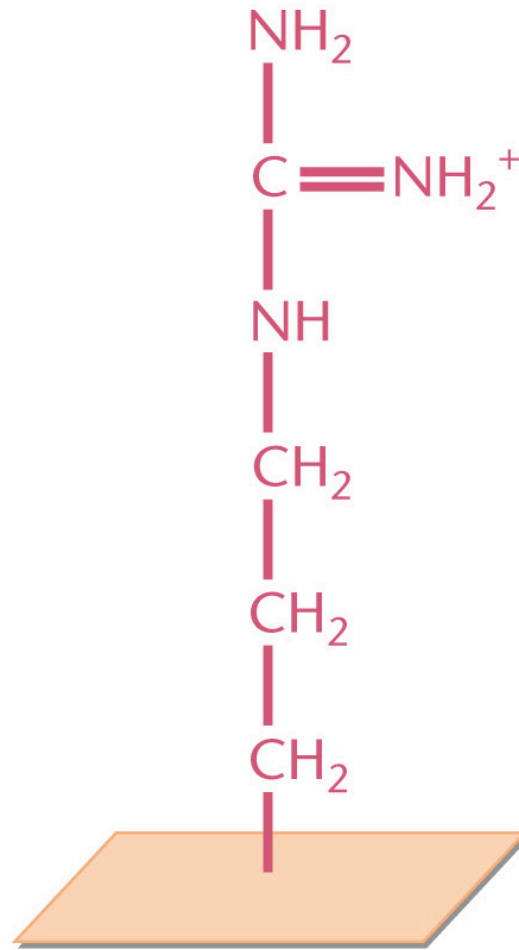
3. Polar: positively charged (basic)

Amino Acid Structures
from Klug & Cummings



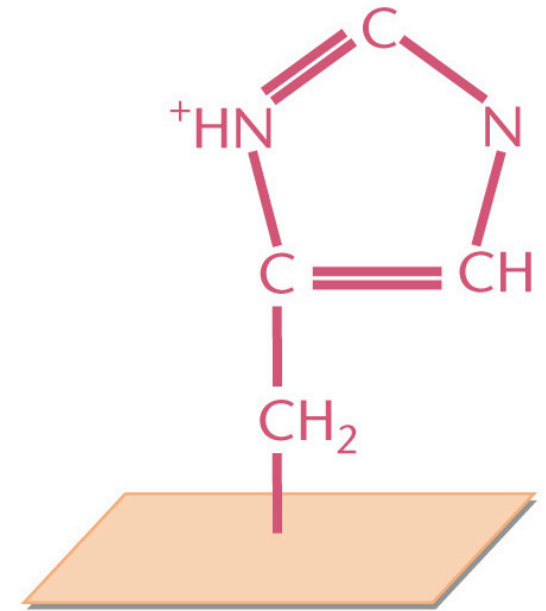
Lysine (lys-K)

3/8/2011



Arginine (arg-R)

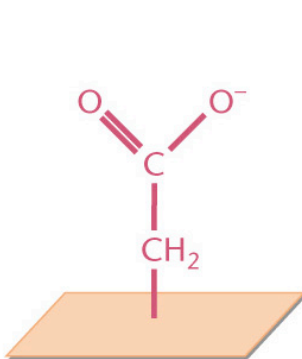
CAP5510 / CGS5166



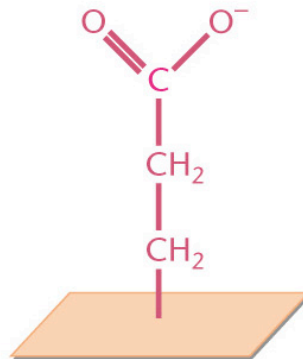
Histidine (his-H)

17

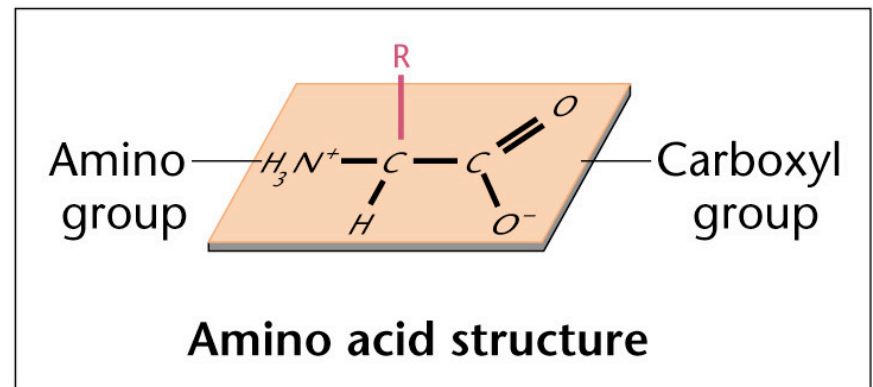
4. Polar: negatively charged (acidic)



Aspartic acid (asp-D)

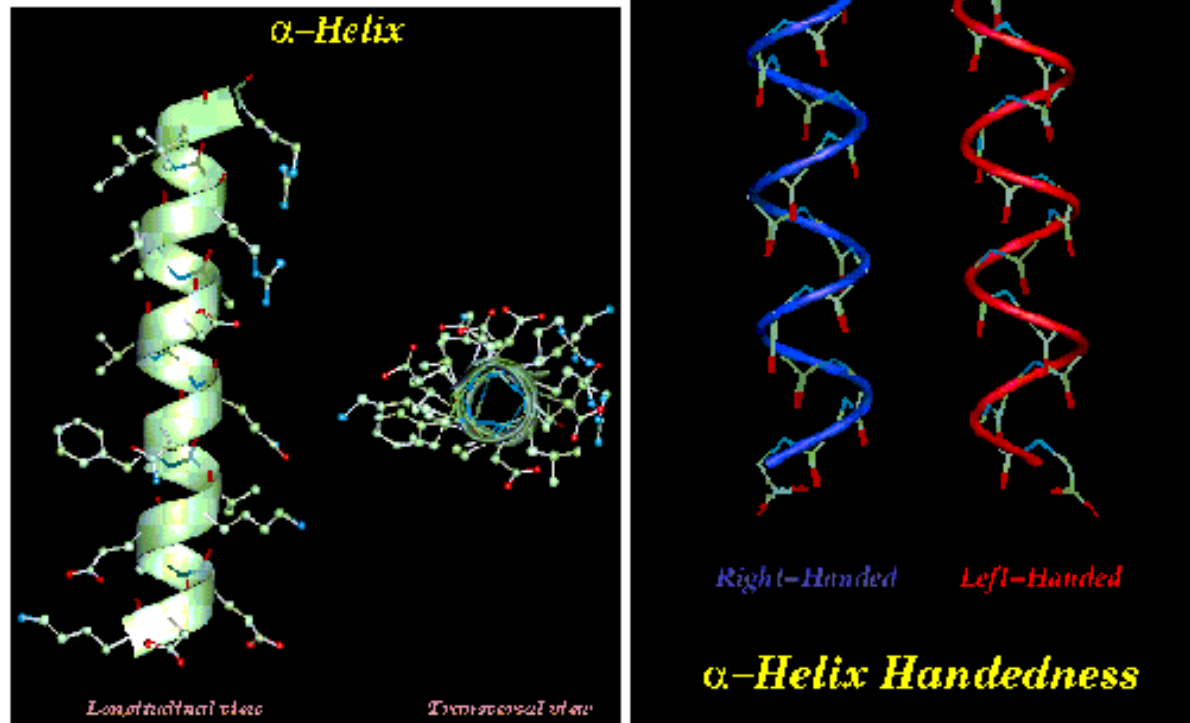


Glutamic acid (glu-E)



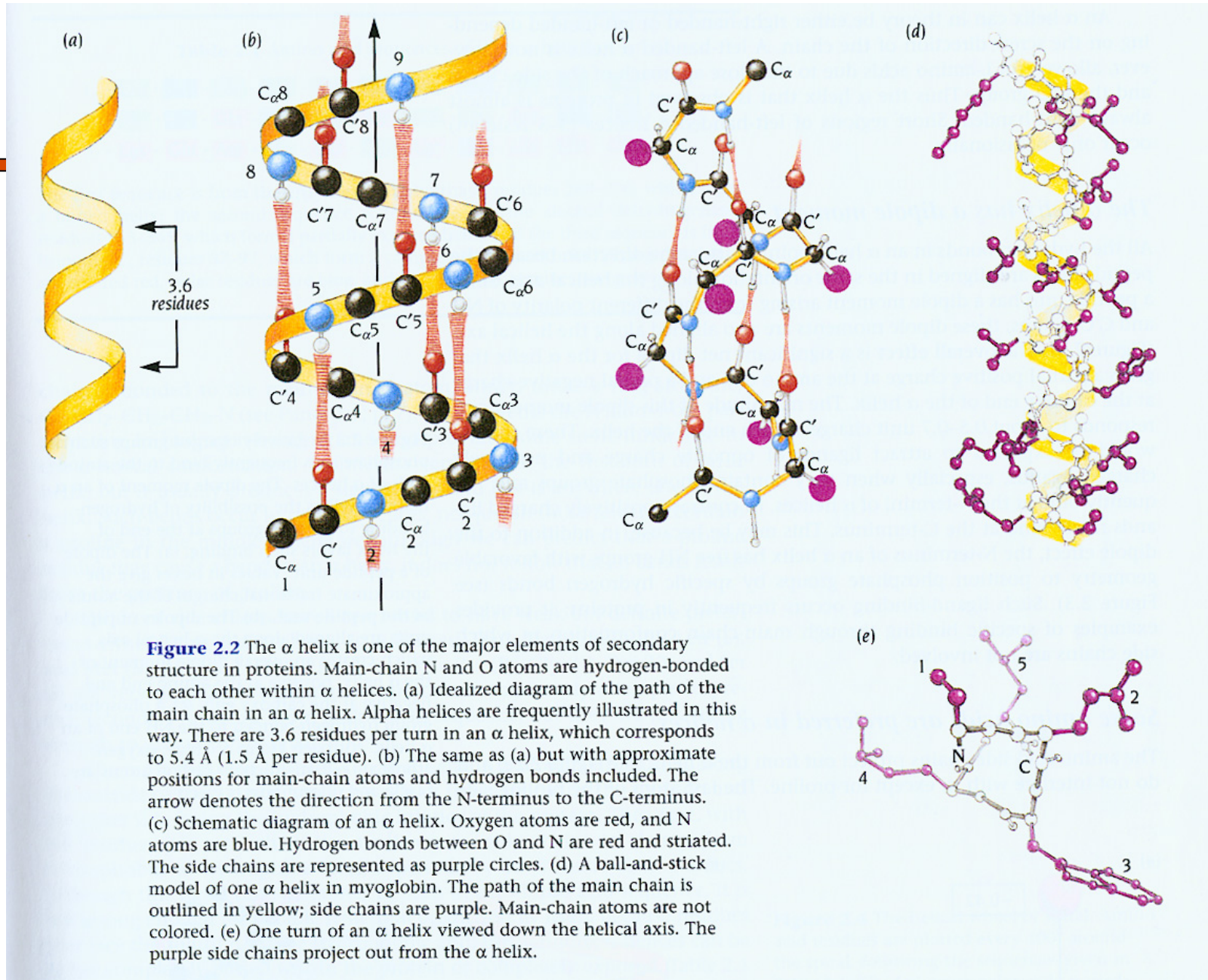
Amino Acid Structures from Klug & Cummings

Alpha helices



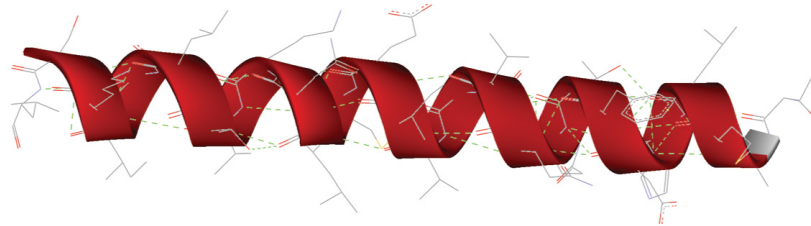
(c) David Gilbert, Aik Choon Tan, Gilleain Torrance and Malika Veeramalai 2002

16

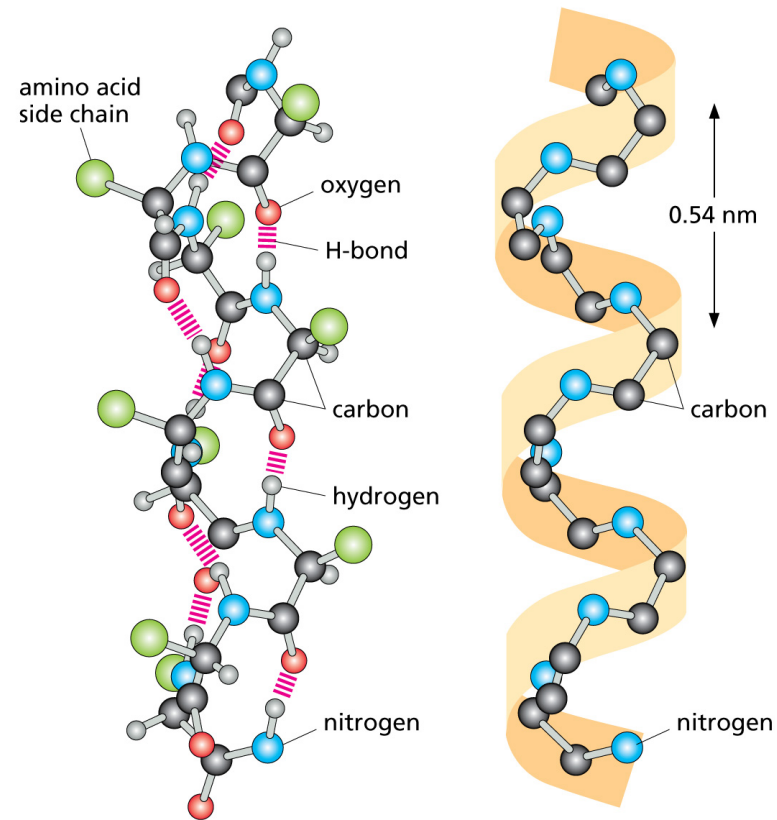


Alpha Helix

(A)

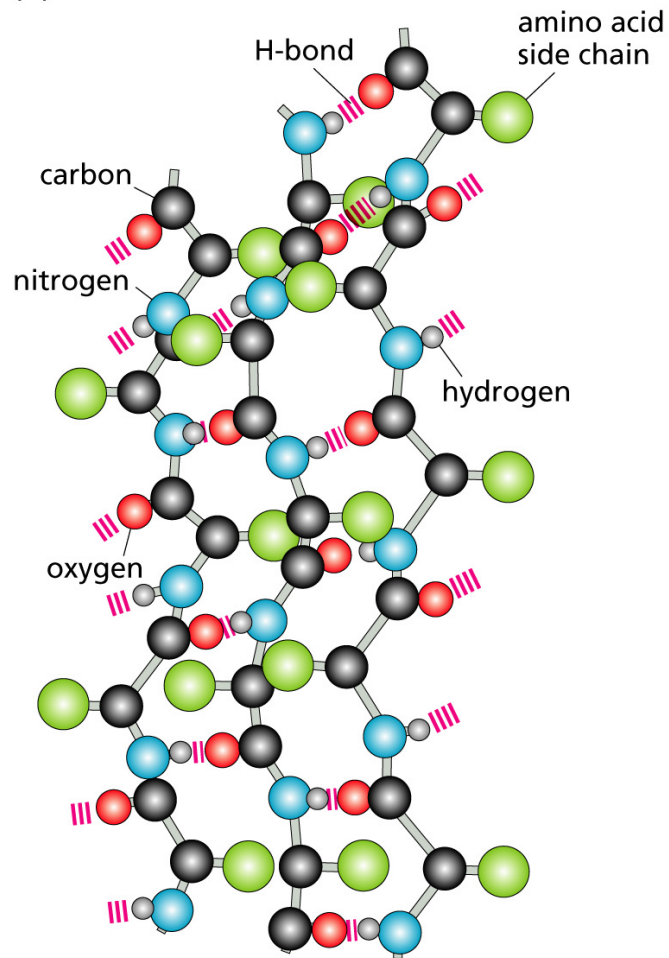


(B)

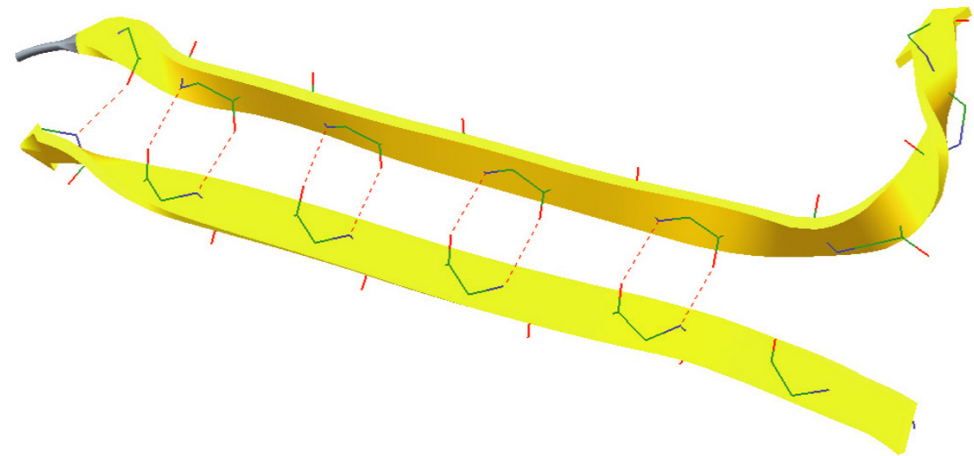


Beta Strands and Sheets

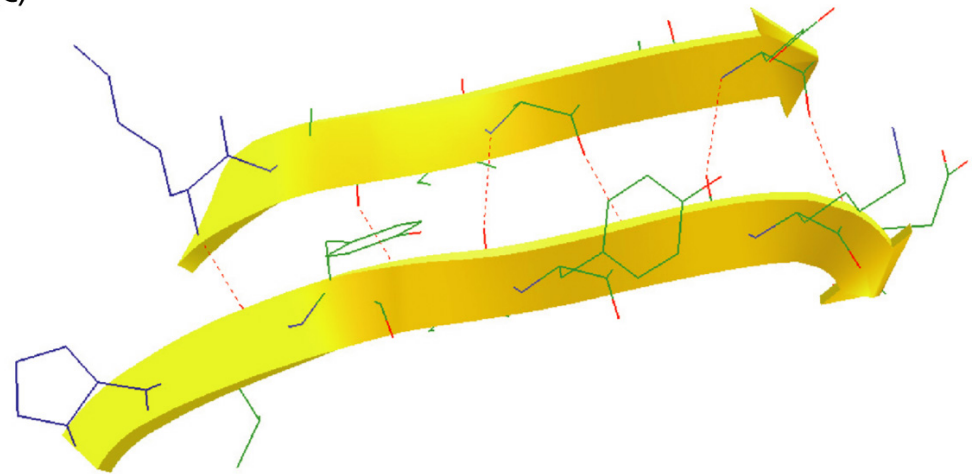
(A)



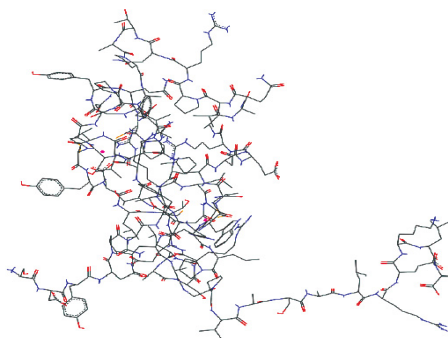
(B)



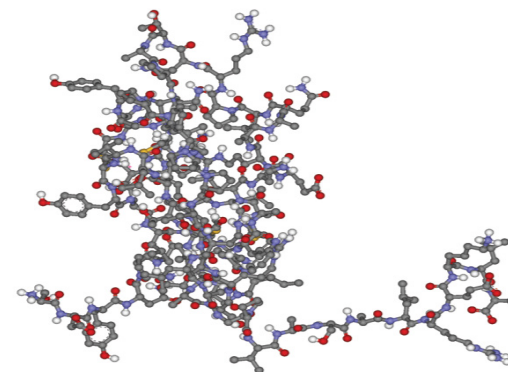
(C)



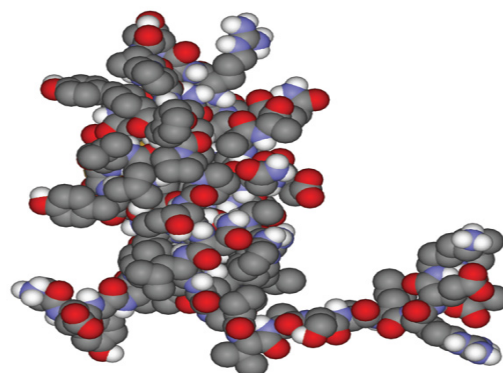
Molecular Representations



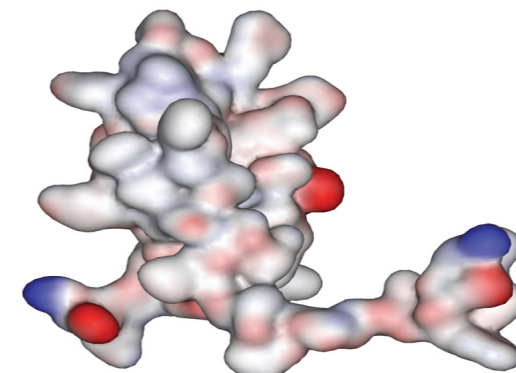
wire-frame



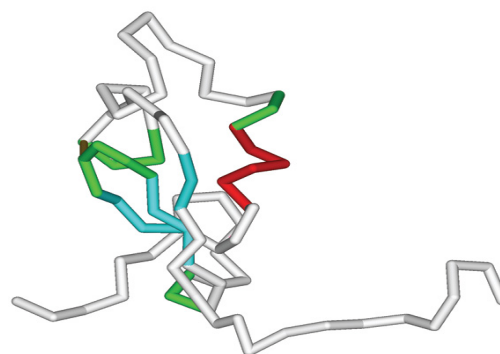
ball and stick



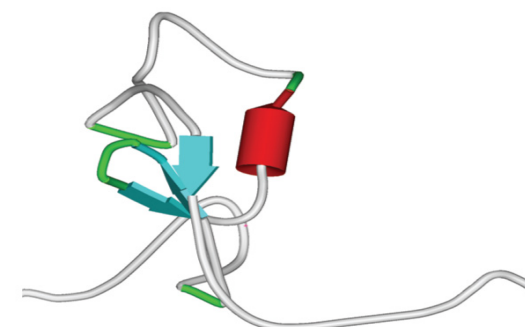
space-filling



surface



C α representation



α/β schematic

3/8/2011

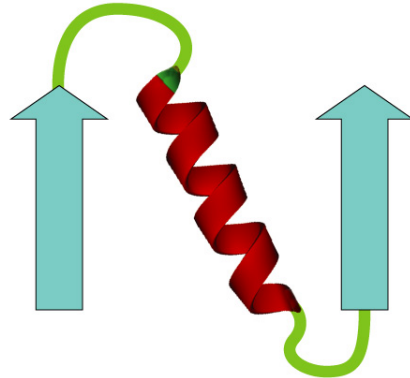
CAP5510 / CGS5166

24

Supersecondary structures

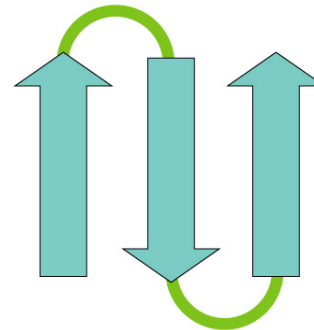
(A)

$\beta\alpha\beta$ repeat



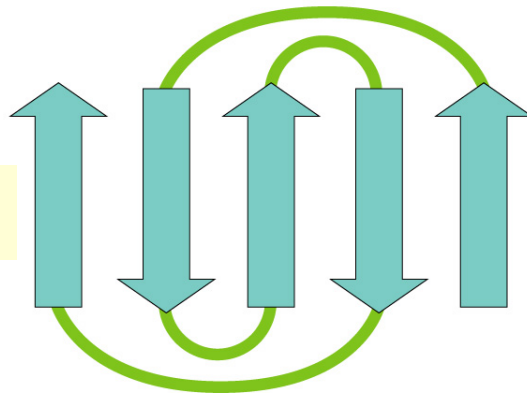
(B)

$\beta\alpha\beta$ -meander



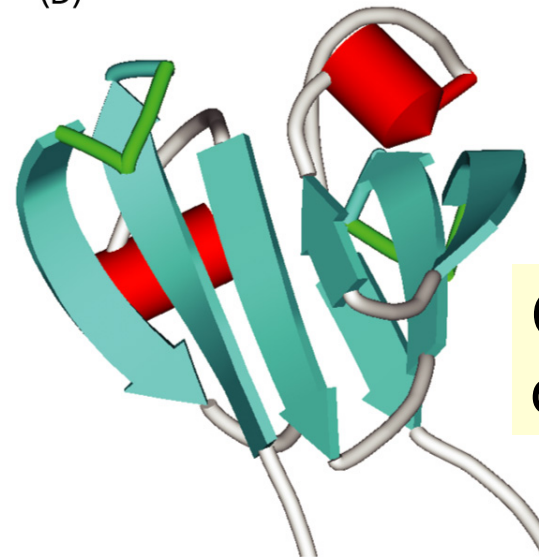
(C)

Greek Key



(D)

Gamma β
crystallin



Secondary Structure Prediction Software

254

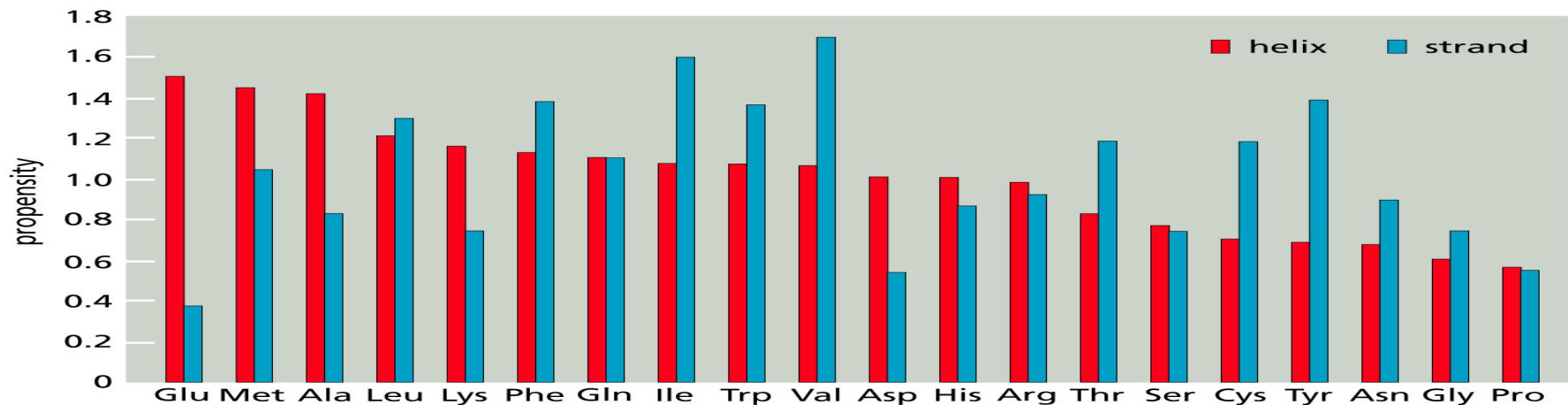


Figure 11.3 Comparison of secondary structure predictions by various methods. The sequence of flavodoxin, an α/β protein, was used as the query and is shown on the first line of the alignment. For each prediction, H denotes an α helix, E a β strand, T a β turn; all other positions are assumed to be random 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 (10FV, Smith et al., 1983).

Recent Ones:
 GOR V
 PREDATOR
 Zpred
 PROF
 NNSSP
 PHD
 PSIPRED
 Jnet

Chou & Fasman Propensities

Amino Acid	helix		strand	
	Designation	<i>P</i>	Designation	<i>P</i>
Ala	F	1.42	b	0.83
Cys	l	0.70	f	1.19
Asp	l	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	l	0.98	l	0.93
Ser	l	0.77	b	0.75
Thr	l	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
CCCCCCHHHHHHHHHHHHHHHCCCCCHHHHEEECCCCCHHHHHHHHHHH
AQDKSGFIEEDELKLFQNFKADARALTDGETKTFLKAGDSDGDGKIGVD
HHCCCCCHHHHHHHHHHHHHHHHHCCCCCEEEEECCCCCCCCCEEECC
DVTALVKA
CEEEEEEC

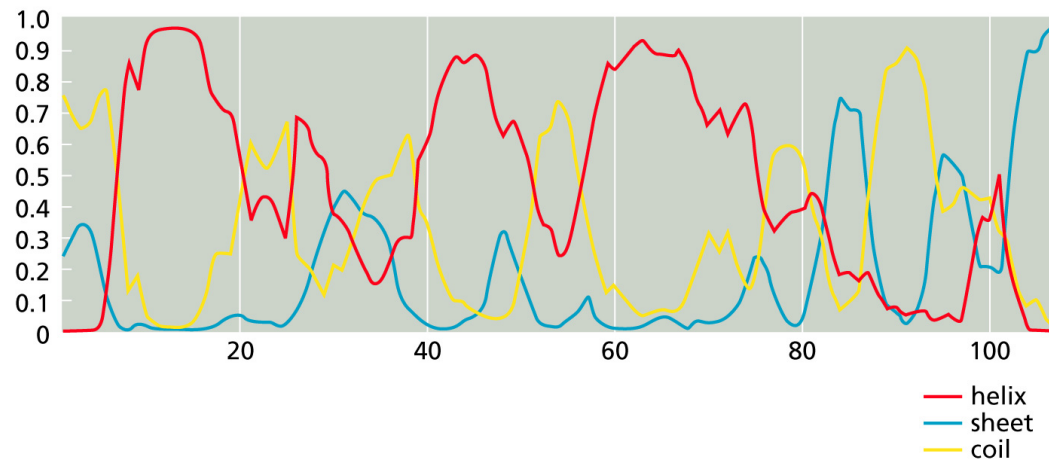
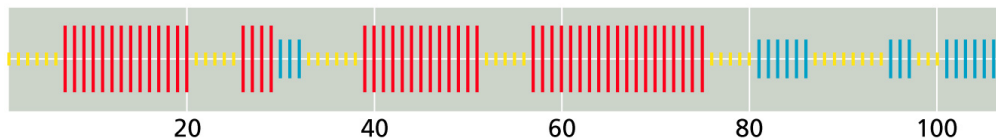
sequence length: 108

GOR IV:

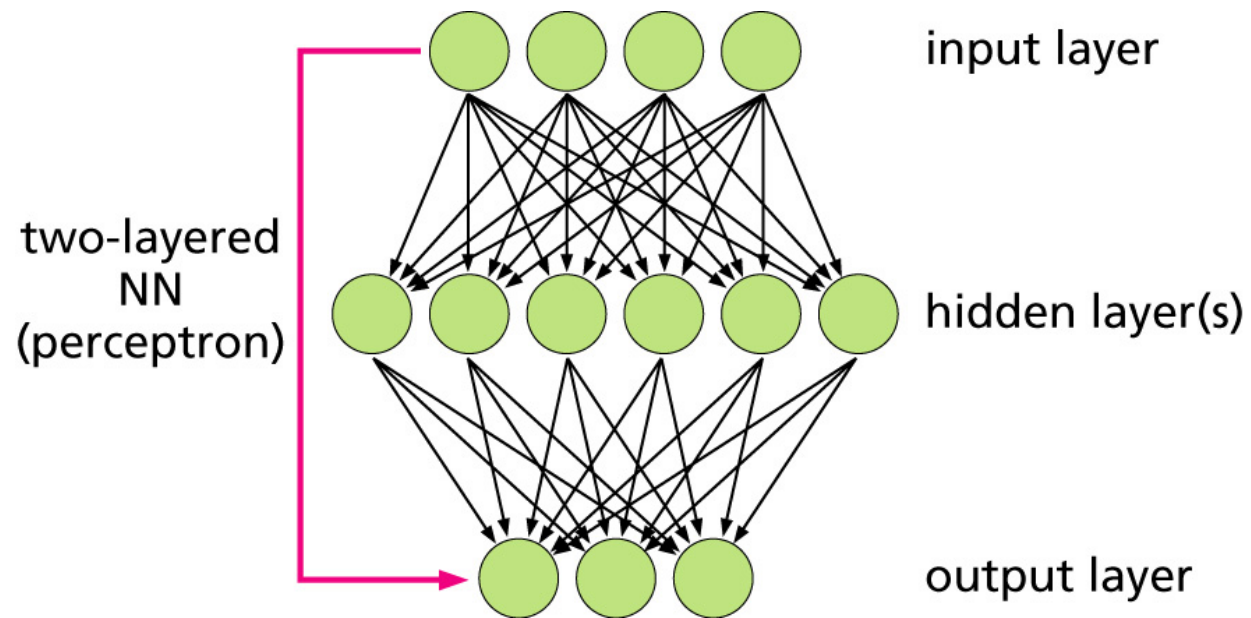
alpha helix (Hh) : 50 is 46.30%

beta sheet (Ee) : 18 is 16.67%

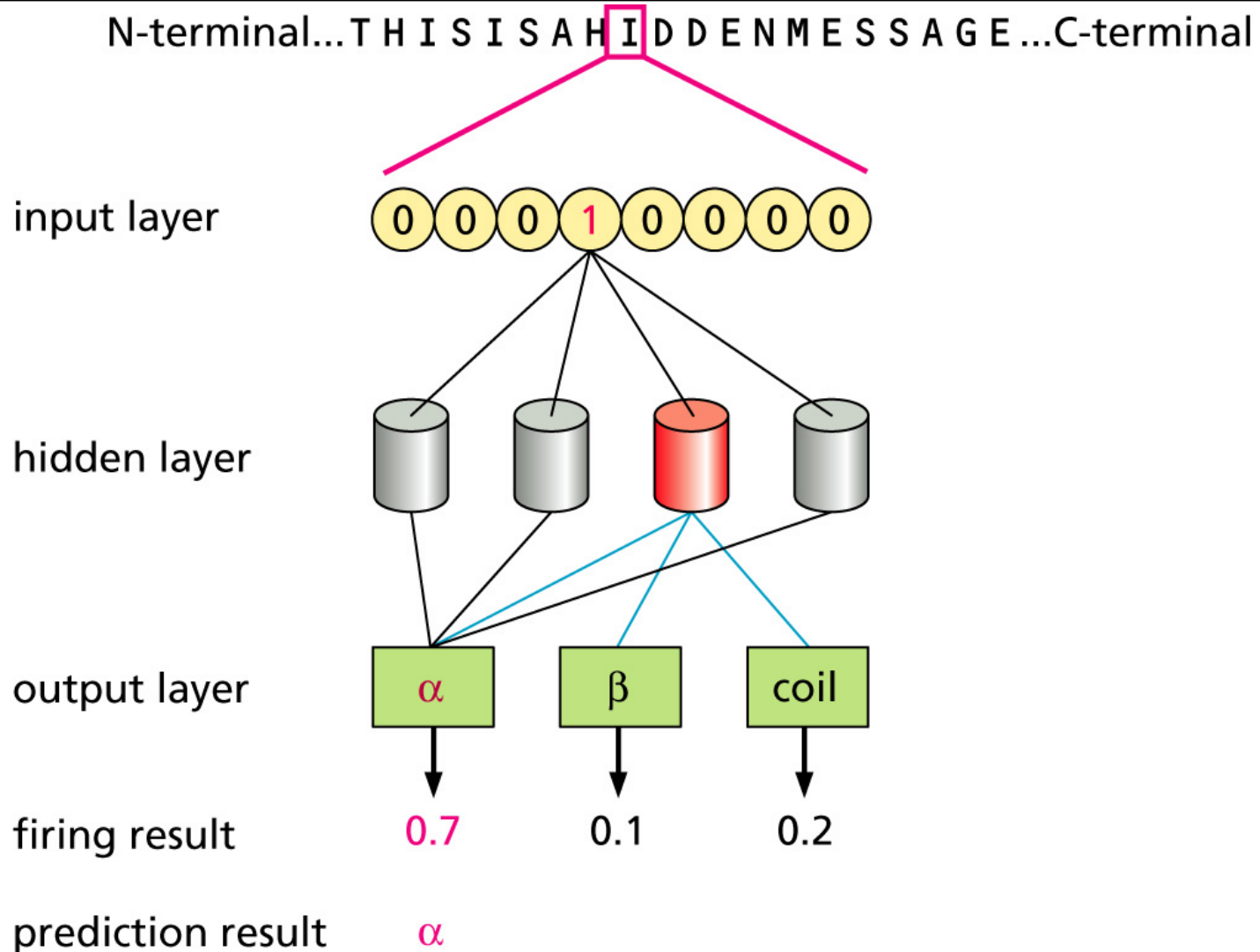
random coil (Cc) : 40 is 37.04%



Neural Networks



Neural Network Prediction of SS



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

RCSB **PDB** PROTEIN DATA BANK

A MEMBER OF THE **PDB**

An Information Portal to Biological Macromolecular Structures


Contact Us | Help | Print Page


PDB ID or keyword Author **SEARCH** | Advanced Search


Home Search **Results** Queries

91 Structure Hits 127 Web Page Hits 1 Unreleased Structure

1 2 3 4 5 .. 10 ↩

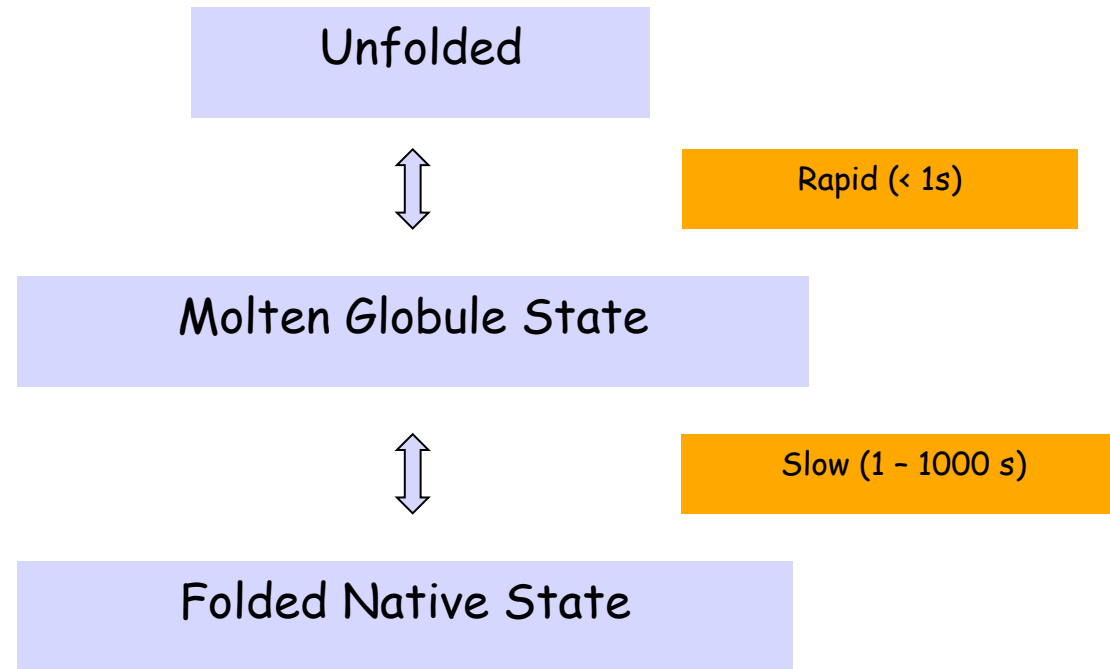
1X62  **Characteristics** **Classification** **Compound** **Authors** **Release Date:** 17-Nov-2005 **Exp. Method:** NMR 20 Structures **Structural Protein** **Mol. Id:** 1 **Molecule:** C Terminal Lim Domain Protein 1 **Fragment:** Lim Domain **Authors:** Qin, X.R., Nagashima, T., Hayashi, F., Yokoyama, S.

1X4K  **Characteristics** **Classification** **Compound** **Authors** **Release Date:** 14-Nov-2005 **Exp. Method:** NMR 20 Structures **Metal Binding Protein** **Mol. Id:** 1 **Molecule:** Skeletal Muscle Lim Protein 3 **Fragment:** Lim Domain **Authors:** He, F., Muto, Y., Inoue, M., Kigawa, T., Shirouzu, M., Terada, T., Yokoyama,

1X4L  **Characteristics** **Classification** **Compound** **Authors** **Release Date:** 14-Nov-2005 **Exp. Method:** NMR 20 Structures **Metal Binding Protein** **Mol. Id:** 1 **Molecule:** Skeletal Muscle Lim Protein 3 **Fragment:** Lim Domain **Authors:** He, F., Muto, Y., Inoue, M., Kigawa, T., Shirouzu, M., Terada, T., Yokoyama,

- Results (1-10 of 91)
- Results ID List
- Refine this Search
- 1 Structures Awaiting Release
- Select All
- Deselect All
- Download Selected
- Tabulate
- Narrow Query
- Sort Results
- Results per Page
- Show Query Details
- Results Help

Protein Folding



□ How to find minimum energy configuration?