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: AKIGLFYGTQVTQITAESIQEFGGESIVDLNIDIANADA...

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

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

- ALIPHATIC
- HYDROPHOBIC
- AROMATIC
- CHARGED
- POSITIVE
- SMALL
- TINY
All 3 figures are cartoons of an amino acid residue.

**Structure of a single amino acid**

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 $\phi$ and $\psi$ 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 $\phi$ and $\psi$. 
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

3/8/2011  CAP5510 / CGS5166  16
3. Polar: positively charged (basic)

Lysine (lys-K)  Arginine (arg-R)  Histidine (his-H)

Amino Acid Structures from Klug & Cummings
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, Gillean Torrance and Mallika Veeramalsi 2002

3/8/2011
Ramachandran Plot

(A) 180
150
120
90
60
30
0
-30
-60
-90
-120
-150
-180
0 60 120 180

(B) 180
135
90
45
0
-45
-90
-135
-180
0 45 90 135 180

3/8/2011
CAP5510 / CGS5166
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.
Alpha Helix
Beta Strands and Sheets

(A) H-bond
amino acid side chain

(B) carbon
nitrogen
hydrogen

(C) oxygen
Molecular Representations

- wire-frame
- ball and stick
- space-filling
- surface
- Cα representation
- α/β schematic
Supersecondary structures

(A) βαβ repeat

(B) βαβ-meander

(C) Greek Key

(D) Gamma β crystallin
Secondary Structure Prediction Software

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

Figure 11.3 Comparison of secondary structure predictions by various methods. The sequence of flavodoxin, an αβ 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).
Chou & Fasman Propensities

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>helix</th>
<th></th>
<th>strand</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Designation</td>
<td>$P$</td>
<td>Designation</td>
</tr>
<tr>
<td>Ala</td>
<td>F</td>
<td>1.42</td>
<td>b</td>
</tr>
<tr>
<td>Cys</td>
<td>l</td>
<td>0.70</td>
<td>f</td>
</tr>
<tr>
<td>Asp</td>
<td>l</td>
<td>1.01</td>
<td>B</td>
</tr>
<tr>
<td>Glu</td>
<td>F</td>
<td>1.51</td>
<td>B</td>
</tr>
<tr>
<td>Phe</td>
<td>f</td>
<td>1.13</td>
<td>f</td>
</tr>
<tr>
<td>Gly</td>
<td>B</td>
<td>0.61</td>
<td>b</td>
</tr>
<tr>
<td>His</td>
<td>f</td>
<td>1.00</td>
<td>f</td>
</tr>
<tr>
<td>Ile</td>
<td>f</td>
<td>1.08</td>
<td>F</td>
</tr>
<tr>
<td>Lys</td>
<td>f</td>
<td>1.16</td>
<td>b</td>
</tr>
<tr>
<td>Leu</td>
<td>F</td>
<td>1.21</td>
<td>f</td>
</tr>
<tr>
<td>Met</td>
<td>F</td>
<td>1.45</td>
<td>f</td>
</tr>
<tr>
<td>Asn</td>
<td>b</td>
<td>0.67</td>
<td>b</td>
</tr>
<tr>
<td>Pro</td>
<td>B</td>
<td><strong>0.57</strong></td>
<td>B</td>
</tr>
<tr>
<td>Gln</td>
<td>f</td>
<td>1.11</td>
<td>h</td>
</tr>
<tr>
<td>Arg</td>
<td>l</td>
<td>0.98</td>
<td>l</td>
</tr>
<tr>
<td>Ser</td>
<td>l</td>
<td>0.77</td>
<td>b</td>
</tr>
<tr>
<td>Thr</td>
<td>l</td>
<td>0.83</td>
<td>f</td>
</tr>
<tr>
<td>Val</td>
<td>f</td>
<td>1.06</td>
<td>F</td>
</tr>
<tr>
<td>Trp</td>
<td>f</td>
<td>1.08</td>
<td>f</td>
</tr>
<tr>
<td>Tyr</td>
<td>b</td>
<td>0.69</td>
<td>F</td>
</tr>
</tbody>
</table>
GOR IV prediction for 1bbc

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

- **input layer**
- **hidden layer(s)**
- **output layer**

Two-layered NN (perceptron)
Neural Network Prediction of SS

N-terminal...THIS IS A HIDDEN MESSAGE...C-terminal

input layer

hidden layer

output layer

firing result

prediction result
PDB: Protein Data Bank

- Database of protein tertiary and quaternary structures and protein complexes. [http://www.rcsb.org/pdb/](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

An Information Portal to Biological Macromolecular Structures

1. 1X62
   - Solution structure of the LIM domain of carboxyl terminal LIM domain protein 1
   - 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.

2. 1X4K
   - Solution structure of LIM domain in LIM-protein 3
   - 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,

3. 1X4L
   - Solution structure of LIM domain in Four and a half LIM domains protein 2
   - 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,
Protein Folding

- Unfolded
  - Rapid (< 1s)

- Molten Globule State
  - Slow (1 - 1000 s)

- Folded Native State

- How to find minimum energy configuration?