Lecture 2: Protein Structure from Secondary to Tertiary

1. Protein 3D Structure Types
2. Factors that Influence Protein Folding
3. The Anfinson Experiment
4. Protein Sequence Chemical Analysis
5. X-ray and How to access Protein 3D structures

Relevant Reading: Chapter 2 (pp. 46-59); Chapter 3 (pp. 87-92)

Linus Pauling (1901-1994) began his studies at Caltech in 1922 and was directed by Arthur Amos Noyes to study a new technique – X-ray crystallography. From his research, he published “The Nature of the Chemical Bond” in 1939. His research also elucidated the alpha-helix and beta-sheet repeating systems central to protein structure and determined the molecular basis of sickle-cell anemia in 1949. During 1958-1963, Pauling fought to halt nuclear testing and illuminated the dangers of radiation to the American public.

Read more about Pauling and more scientists at: Profiles in Science, The National Library of Medicine, profiles.nlm.nih.gov
The 20 Natural Amino Acids (How to memorize them)

Glycine, Gly (G)  
R = H  
• most flexible linkage

Valine, Val (V)  
R = iso-propyl  
• a “beta-branched” aa  
• staggered conf. shown

Isoleucine, Ile (I)  
R = sec-butyl  
• it’s “beta-branched”  
• the sidechain is (S)-

Serine, Ser (S)  
R = CH₂OH  
• good for nucleophilic catalysis

Cysteine, Cys (C)  
R = CH₂SH  
• L-Cys is (R)- because S has “priority” over O

Alanine, Ala (A)  
R = CH₃

Leucine, Leu (L)  
R = “iso-butyl”

Proline, Pro (P)  
R = -CH₂CH₂CH₂-  
• the only cyclic aa  
• called an “imino acid”

Threonine, Thr (T)  
R = CH(Me)OH  
• it’s “beta-branched”  
• the sidechain is (R)-

Methionine, Met (M)  
R = -CH₂CH₂SMe

• All the amino acids on this page have the one-letter code = first letter of the name.  
• All have the same “L-“ stereochemistry. 19 are (S)- and one is (R)-  
• Ala, Val, Leu, Ile, Pro, Cys and Met are all considered “hydrophobic” aa’s.  
(Gly is not considered hydrophobic because it has no sidechain “protecting” the polar groups.)
The 20 Natural Amino Acids (continued)

Phenylalanine, Phe (F)
R = benzyl
- hydrophobic aa

Histidine, His (H)
R = -CH₂-imidazole
- weakly basic (pKa~6.0)

Lysine, Lys (K)
R = -[CH₂]₄-NH₂
- basic (pKa~10.8)
  amine form shown
  ammonium (protonated)

Tyrosine, Tyr (Y)
R = p-hydroxybenzyl
- phenolic pKa~10.9

Tryptophan, Trp (W)
R = -CH₂-indole

Arginine, Arg (R)
-CH₂CONH[=NH]NH₂
- basic (pKa~12.5)
  guanidine form shown
  guanidinium (protonated)

Aspartate, Asp (D)
R = -CH₂CO₂H
- acidic (pKa~4.1)

Glutamate, Glu (E)
R = -CH₂CH₂CO₂H
- acidic (pKa~4.1)

Asparagine, Asn (N)
R = -CH₂CONH₂

Glutamine, Gln (Q)
R = -CH₂CH₂CONH₂

- One-letter code for all except His (H) have to be memorized. Three “basic” aa’s + Two “acidic” aa’s.
From Secondary Structure to Tertiary Structure

Antiparallel β-sheets frequently reverse direction by including a “turn sequence” of two or more amino acids. The most common turn motifs are the Type I and Type II β-turns which account for ~40-60% of turns found in proteins. Type I and Type II β-turns are related by “amide plane inversion” and display all four sidechains (i to i+4) on the same side of the β-sheet.

Both parallel and antiparallel beta sheets can adopt structures displaying twists, barrels and complex combinations of parallel and anti-parallel strands. The arrow shows the direction of the peptide sequence going N-terminal to C-terminal. These lead to a diverse region in the Ramachandran plot. Below, α-chymotrypsin shows an antiparallel β-barrel, carbonic anhydrase shows a β-saddle and triosephosphate isomerase shows a parallel “beta-alpha-beta barrel”.

α-Chymotrypsin  Carbonic Anhydrase (1CA2)  Triosephosphate isomerase (β)  Triosephosphate isomerase (1TIM)
The Protein Folding Problem – Levinthal’s Paradox

Consider a small protein (polypeptide) comprised of 100 aa:

• If each amino acid can assume 2 conformations, then the polypeptide can adopt $2^{100} \approx 10^{30}$ conformations.
• Assume that conformations interconvert with a time constant of $10^{-12}$ s.

The time required to sample all conformations is roughly:

$(10^{30} \text{ conformations}) \times (10^{-12} \text{ s/conformation}) = 10^{18} \text{ s} \approx 10^{10} \text{ y}$

⇒ Proteins do not fold by a random search of conformation space.

The energetics of protein folding to the native state (N) are more like the bottom diagram:

• Secondary Structural features are sampled independently.
• path towards the native structure is energetically favorable.
• some proteins require assistance from prolyl cis/trans isomerases in order to fold correctly.
How do Biological Polypeptides Achieve Complex 3D Structures?

1. Spontaneous Folding based on Sequence
2. “Assisted” Folding by the Living Organism

Examples of Both

1950’s experiments by Christian Anfinson demonstrated that a “denatured” protein possessed the information in its sequence to refold into active enzyme.

Addition of excess 2-Mercaptoethanol reduces disulfides. Removal of 2-Mercaptoethanol allows air oxidation of cysteines to form cystine linkages.

“Catalytic” 2-Mercaptoethanol allows disulfide exchange.

Christian Anfinson, Nobel Prize 1972

Bovine Pancreatic Ribonuclease (7RSA)
Protein Folding Chaperones – Mediate Protein Folding for Some Key Enzymes and Structural Proteins
Protein Misfolding and Disease

Ribosome → Nascent polypeptide chain

Genetics & Environment → Misfolding → Aggregation → Disease
Neurodegenerative Diseases Characterized by Protein/Peptide Deposits

Plaques: β-amyloid, non-β-amyloid component (NAC) EXTRACELLULAR tangles: phosphorylated tau INTRACYTOPLASMIC

Huntingtin with expanded polyglutamine tracts INTRANUCLEAR

Driving Forces in Protein Folding

Unfolded

- hydrophobic groups stabilized by water requires ordered structure
- Polar groups stabilized by water
- ionic groups in water

Native (folded)

- hydrophobic groups stabilized by packing in interior of protein
- polar groups hydrogen bonding
- ionic groups $\rightarrow$ salt bridges inside protein

Water plays a large part in the process of protein folding:

Just like for water stabilizing methane, an ordered structure is required to stabilize hydrophobic parts of the polypeptide chain.

- Protein folding releases “ordered” water: $\Delta S$ is positive

Spontaneous protein folding is aided by release of entropy:

$$\Delta G = \Delta H - T\Delta S$$

- Typical $\Delta G$ values for protein folding are $\approx -10$ to $-20$ kcal/mole (unfolded$\rightarrow$folded).
Determining Protein Folding Using “Bio-Robots”

"Predicting protein structures with a multiplayer online game", with Seth Cooper, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker and Foldit players, in *Nature* 446 p. 756-760 (05 August 2010). [pdf](project)
X-Ray Diffraction

Data in reciprocal space

Data in reciprocal space

Fourier Transform
inner circle of data

Fourier Transform
inner circle of data

Low resolution structure

Fourier Transform
full data set

Low resolution structure

Fourier Transform
full data set

High resolution structure

Myoglobin X-ray Exposure

Myoglobin X-ray Exposure

“My Plexiglas Structure”

Myoglobin α-carbon trace
How To Download Protein Structures from the PDB

The Protein Data Bank was started in 1971 as a repository for biological structural data. Today, the PDB contains over 100,000 structures of everything from proteins to DNA to whole virions. Although originally intended for X-ray coordinates, the PDB also contains NMR structures and EM data files.

Procedure:
1. Log onto Google and type “RCSB” — You’re in RCSB.
2. Log onto Google and Type:
   a. Chimera UCSF
   b. Avogadro
   c. PyMol

   Modeling

3. You can download the modeling software of your choice. They are free!
4. Type in the name of the enzyme or the 4-character “pdb code” e.g., 1mbn or the name of the protein (myoglobin).
Some Features of Protein Structures and Function

Proteins have more surprises than just α-helices and β-sheets:

1. other secondary structures – Collagen triple helix
2. recurring tertiary structural motifs: β-barrel, coiled coil, SH2 domain of oncogenes
3. multiple chains & subunits – Quaternary Structure
4. disulfide linkages intra- and inter-chain. (usually for excreted proteins)
5. disordered loops and special types of turns.
6. coordinated metal ions for catalysis (Mg$^{2+}$, Zn$^{2+}$, etc.) and structural stability
7. covalently or non-covalently bound heme groups for complexing Fe, Cu ions.
8. methods of getting special amino acids – selenocysteine, hydroxyproline
9. post-translational modified amino acids – methylation, phosphorylation, glycosylation

These allow proteins to serve a diverse array of cellular functions:

1. Enzyme Catalysis -
   - simple: CO$_2$ + H$_2$O -> H$_2$CO$_3$
   - intricate: DNA replication
   - enormous catalytic power - $\geq 10^6$
2. Transport and Storage -
   - O$_2$, nutrients like Fe$^{2+}$
3. Coordinated Motion -
   - muscle, motor proteins, cilia/flagella, mitosis
4. Mechanical Support -
   - collagen, tubulin, cytoskeleton
5. Immune Protection -
   - antibodies
6. Nerve Impulses -
   - ion transport across membranes
7. Control of growth and differentiation -
   - gene expression and promoters
Protein Sequence and Structure Analysis – N-Terminal aa Identification

Figure 2.27
Determination of the amino-terminal residue of a peptide. Fluorodinitrobenzene (Sanger’s reagent) is used to label the peptide, which is then hydrolyzed. The DNP-amino acid (DNP-alanine in this example) is identified by its chromatographic characteristics.
Figure 2-28
The Edman degradation. The labeled amino-terminal residue (PTH-alanine in the first round) can be released without hydrolyzing the rest of the peptide. Hence, the amino-terminal residue of the shortened peptide (Gly-Asp-Phe-Arg-Gly) can be determined in the second round. Three more rounds of the Edman degradation reveal the complete sequence of the original peptide.
Protein Sequence and Structure Analysis – Chemical And Enzymic Cleavage

**Figure 2.29**
Cyanogen bromide cleaves polypeptides on the carboxyl side of methionine residues.

**Figure 2.30**
Trypsin hydrolyzes polypeptides on the carboxyl side of arginine and lysine residues.

**Table 2.2**
Specific cleavage of polypeptides

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Cleavage site</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chemical cleavage</strong></td>
<td></td>
</tr>
<tr>
<td>Cyanogen bromide</td>
<td>Carboxyl side of methionine residues</td>
</tr>
<tr>
<td>Hydroxylamine</td>
<td>Asparagine–glycine bonds</td>
</tr>
<tr>
<td>2-Nitro-5-thiocyanobenzoate</td>
<td>Amino side of cysteine residues</td>
</tr>
<tr>
<td><strong>Enzymatic cleavage</strong></td>
<td></td>
</tr>
<tr>
<td>Trypsin</td>
<td>Carboxyl side of lysine and arginine residues</td>
</tr>
<tr>
<td>Clostripain</td>
<td>Carboxyl side of arginine residues</td>
</tr>
<tr>
<td>Staphylococcal protease</td>
<td>Carboxyl side of aspartate and glutamate residues</td>
</tr>
<tr>
<td></td>
<td>(glutamate only under certain conditions)</td>
</tr>
</tbody>
</table>