Introduction to Proteins

Lecture 4
Module I: Molecular Structure & Metabolism
Molecular Cell Biology Core Course (GSND5200)
Matthew Neiditch - Room E450U ICPH
matthew.neiditch@umdnj.edu
What is a protein?

Met-Asp-Leu-Tyr-His-Val-Phe-Ala-Ile-Asn-Lys-Pro-Gly

primary sequence

Sequence determines 3D structure (shape)

Structure determines function
Proteins constitute most of a cell’s dry mass and have diverse functions

<table>
<thead>
<tr>
<th>Protein</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hemoglobin</td>
<td>Storage and transport</td>
</tr>
<tr>
<td>Insulin receptor</td>
<td>Signal reception</td>
</tr>
<tr>
<td>Insulin</td>
<td>Secreted messenger</td>
</tr>
<tr>
<td>IgA and IgG</td>
<td>Antigen recognition</td>
</tr>
<tr>
<td>K and Na channels</td>
<td>Ion conductance through membrane</td>
</tr>
<tr>
<td>Kinesin</td>
<td>Motor</td>
</tr>
<tr>
<td>Actin, Tubulin, Keratin</td>
<td>Cytoskeleton</td>
</tr>
<tr>
<td>Collagen</td>
<td>Connective tissue</td>
</tr>
<tr>
<td>Aspartate Transcarbamoylase</td>
<td>Enzyme</td>
</tr>
</tbody>
</table>
Amino acid
Amino acids are linked by peptide bonds

methionine (Met)

aspartic acid (Asp)

H₂O

polypeptide backbone

amino terminus or N-terminus

peptide bond

carboxyl terminus or C-terminus
Proteins are polypeptides
Polypeptides can rotate in some places and not in others.
Rotation around phi and psi enable the chain to bend into shapes called secondary structures.

alpha helix

beta sheet

Cartoon models of polypeptide backbones
Pymol demonstration
alpha helix
Why is proline not commonly found in alpha helices?
Rotation around phi and psi enable the chain to bend into shapes called **secondary structures**

- **alpha helix**
- **beta sheet**

Cartoon models of polypeptide backbone
Beta sheets

antiparallel

parallel
Pymol demonstration
beta sheet
Secondary structures are joined by loops and turns forming tertiary structures.

- All alpha helical (4-helix bundle)
- All beta sheet (barrel)
- Mixed alpha-beta
Structure determines function
Interaction between protein surface and ligand
Structure determines function

ligand binding induces a conformational change (animation)
Some proteins are the right shape to bind to a chemical reaction’s transition state.

These proteins are called enzymes.
Weak non-covalent forces that stabilize protein tertiary structure
Why do proteins fold?

\[ \Delta G = \Delta H - T\Delta S \]

Recall, for a process to be thermodynamically favorable \( \Delta G \) must be negative.

The protein is becoming more ordered (2° and 3° structure), which decreases entropy (inc \( \Delta G \)).

The protein is making lots of H-bonds to itself, but no more than it made to the solvent (no effect on \( \Delta G \)).

Where is the energy for folding coming from? Still need negative \( \Delta H \) or positive \( \Delta S \) from somewhere to fold.
Proteins bury hydrophobic residues in their core

unfolded polypeptide  folded conformation in aqueous environment
The hydrophobic effect

Ordered water around surface exposed hydrophobic amino acid

low entropy

high entropy

Burying hydrophobic amino acids releases caged water (increases $\Delta S$) $\Delta G = \Delta H - T \Delta S$
Disulphide bonds stabilize protein tertiary structure primarily in extracellular proteins.
Protein folding

Many proteins fold quickly (sec to min)

Proteins cannot possibly fold by sampling all possible conformations

Consider a protein with 100 amino acids
Assigning even only 2 conformations to each amino acid

total possible conformations = $2^{100} = 1.3 \times 10^{30}$
even sampling one conformation every 0.1 picoseconds = $10^{-13}$ sec
$(10^{-13})(1.3 \times 10^{30}) = 1.3 \times 10^{17}$ sec = $4 \times 10^9$ years per protein

Age of universe = $13.73 \times 10^9$ years

Proteins sample a limited number of conformations to reach a thermodynamically stable folded (native) conformation

Cyrus Levinthal’s Paradox
Nucleation/condensation model of protein folding

1) Nucleation/condensation
   - Hydrophobic collapse
   - Local 2° structure formation

2) Intermediate, partly folded states
   - Compact, but more expanded than native structure
   - Significant 2° structure formation
   - No stable 3° structure (molten globule)

3) Final rearrangements
   - Final docking of 2° structure to form 3° structure
   - Proline isomerization
   - Disulphide bond formation

some proteins cannot fold on their own so they require chaperones
Critical Assessment of Techniques for Protein Structure Prediction (CASP)

Competition held every 2 years since 1994 to test the state of the art algorithms for protein structure prediction given only the primary sequence

Structure prediction categories

1) Template-free modeling
Ab initio approaches using programs that specify attractive and repulsive folding forces to search for the structure with the lowest energy state

2) Template-based modeling
Use existing proteins with very similar sequences (homology models)

Structure prediction is a computationally exhaustive process but you can help:

Distributed computing
Folding@home  Vijay Pande (Stanford)
Rosetta@home  David Baker (Univ of Washington)
Misfolded proteins can cause neurodegenerative diseases
Structural motifs

Simple combinations of secondary structure elements with a specific geometric arrangement

Motif sequence can be used to predict a protein’s function

- coiled-coil
- helix-loop-helix
- zinc finger
Coiled-coils

collapsed-coil

GAL4 transcription factor

collagen triple helix
Helix-loop-helix (EF Hand)

12 residues in Ca$^{2+}$-coordinating loop
1,3,5,7,9,12 - preference for oxygen-containing side chains (Glu and Asp)

12 - always Glu or Asp

6 - always Gly
Most proteins are composed of a series of **domains**

**Domains** are polypeptide modules (composed of motifs and other secondary structure elements) that fold independently into their tertiary structure

Chromosomal recombination creates multidomain proteins

- EGF
- Chymotrypsin
- Urokinase
- Factor IX
- Fibronectin
Quaternary structure
deoxy hemoglobin
Allostery

Ligand binding at one site is affected by ligand binding at another site

oxy hemoglobin  oxy hemoglobin

Binding increases affinity - positive cooperativity
Summary

amino acid

primary Met-Asp-Leu-Tyr-His-Val-Phe-Ala-Ile-Asn-Lys-Pro-Gly

secondary

tertiary

quaternary

Sequence determines structure

Structure determines function
Protein Dynamics in Health and Disease
(BIOC5030) 3 credits
Carolyn Suzuki (Biochem and Molecular Bio)
Spring
Prerequisites: Fall semester of the Core Course or Molecular & Genetic Medicine (MGM).

Practical Approaches for Studying Protein Function
(CBMM5002) # credits
Maha Abdellatif (Cell Bio)
Spring
Prerequisites: GSND 5200 A & B