How to fold a protein?

Jianhan Chen
Helpers: Dr. Debabani Ganguly (Wednesday)
Chester McDowell and Weihong Zhang (Thursday)

Department of Biochemistry
Kansas State University

Outline

• Introduction:
  – Basic workshop setup
  – Who we are

• Background:
  – What is a protein molecule?
  – Protein structure: why it is necessary for function?
  – Why are motions of proteins important?
  – What is computer modeling?
  – How to fold (or unfold) a protein (to its native structure)?

• Hand-on activities:
  – Visualization and analysis of protein structures
  – Visual analysis of protein dynamics (motions)
  – Folding of a small protein: visual analysis of folding mechanism
**Atoms**

- "Atom" means "unable to be divided"
- First proposed by an ancient Greek philosopher, Demokritos (~ B.C. 500).
- Smallest unit of an element that retains its chemical properties.
- **How small:** weight \(\approx 10^{-27}\) kg with a radius \(\approx 10^{-10}\) m (1 Å)

![STM image of a 40-nm logo of NIST made with cobalt atoms, Science, 2004](image1)

Each droplet of water contains about \(10^{21}\) atoms!

If divided among all people on earth (6.8 billion), each person will receive about a 100 billion atoms!

**Molecules**

- Smallest units of pure substance
- Consist of atoms joint together by chemical bonds
- Come in all sizes and shapes!

![Oxygen molecule](image2)

**Protein Molecules**

- Large biological molecules than mainly consist of hydrogens, carbons and nitrogens
- Polymers of amino acids

![Protein structure](image3)

- Most abundant organic chemicals in the body
  - 50% of body's dry weight and 15% of cell content
- Critical in building and maintenance of the body
  - enzymes, motion, transport, structure, regulation, protection ....

**How big are proteins?**

- About 300 amino acids on average
  - Titin (connectin): largest protein with 34,350 residues
- Consist of hundreds of up to > million atoms
- About several nanometers on average (nano particles!)

![Protein G B1 (PDB: 1PGB) 56 residues, 436 atoms](image4)

Radius ~ 2.5 nm

![HIV-1 Protease (PDB: 2HB3) 198 residues, 1582 atoms](image5)

Radius ~ 5 nm
Protein is Structured

- Hierarchical organizations
  - Primary, secondary, tertiary, quaternary structures

α-Helix and β-Sheet

- Two of the basic secondary structures of proteins

Protein Structure and Function

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Folded Structure</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTYKLILNGK TLKGETTTEA VDAATAEKVF KQYANDNGV DGEWYTDDA TKTFTVTE ...</td>
<td>catalysis</td>
<td>transport</td>
</tr>
<tr>
<td></td>
<td></td>
<td>signaling</td>
</tr>
<tr>
<td></td>
<td></td>
<td>motors ...</td>
</tr>
</tbody>
</table>

- Structural Genomics (NMR, X-ray)
- Protein structure predictions
- Structure-based (rational) drug discovery
Exercise 1: Visualization using VMD

- **Software**: VMD (Visual Molecular Dynamics)
- **Two protein molecules**
  - B1 domain of protein G (1pgb.pdb)
  - HIV-1 Protease (2hb3.pdb)

Detailed Instruction (1pgb.pdb)

1. Double click to open software VMD (three windows will pop up)
2. Click “File” on the VMD Main window, and select “New Molecule” from the pull-down menu (a Molecule File Browser window will pop up)
3. Click Browse button on the Molecular File Browser (a Windows file explorer window will pop up)
4. Navigate to find the file “1PGB.pdb” (located within folder “exercise1” in the “folding” folder on your Desktop); Select “1PGB.pdb” and open it by clicking on “Open” button (file explorer window will disappear)
5. Back to the Molecular File Browser window, click on “Load” button (a protein molecule will show up in your VMD 1.8.6 OpenGL Display window). You should be able to drag to rotate the molecule.
6. To change the drawing method, click “Graphics” and select “Representations” on the pull-down menu (a Graphical Representation window will pop up)
7. Within Graphical Representation window, click the triangle button below the “Drawing Method” and select the desired method (such as “New Cartoon”) from the pull-down menu to change drawing method
8. One can also change the coloring method from the pull-down menu by clicking on the triangle button below “Coloring Method” (such as “Secondary Structure”)

Q: Can you count how many helices and beta strands 1PGB have?

Loading a previously saved VMD state file (e.g., view_2hb3.vmd):

1. Exit VMD and re-open VMD
2. Click “File” on the VMD Main window, and select “Load State” from the pull-down menu (a Windows File explorer window will pop up)
3. Navigate to find the file “view_2hb3.pdb” (located within folder “exercise1” in the “folding” folder on your Desktop); Select the file and open it by clicking on “Open” button (file explorer window will disappear, and VMD start loading and drawing the protein using the options specified in the state file)
4. Does any of the two state files show the proteins in ways that you have not already explored earlier during the exercise?
**Protein Dynamics and Function**

DHFR Catalysis Cycle  
GroEL ATPase Cycle

**Movie Credits:**  
DHFR: [http://chem-faculty.ucsd.edu/kraut/dhfr.html](http://chem-faculty.ucsd.edu/kraut/dhfr.html)  
GroEL: [http://people.cryst.bbk.ac.uk/~ubcg16z/cpn/elmovies.html](http://people.cryst.bbk.ac.uk/~ubcg16z/cpn/elmovies.html)

---

**Molecular Dynamics Simulations**

Channel-forming peptides in a fully solvated membrane bilayer;  
Channel: 1795 atoms; All: 26254 atoms

Ab initio folding of a designed mini-protein Trp-Cage in implicit solvent  
(System size: 304 atoms)

---

**Exercise 2: Visual analysis of protein dynamics**

- See directory “exercise2”
  - Open VMD and load state “play_md.vmd”
  - 1 nanosecond simulation of Protein G B1 at 300 K
  - A total 500,000 molecular dynamics steps
  - Snapshots taken every 1000 steps (500 frames @ 1 frame/2 ps)
- **Assignment:** can you tell which parts of the protein are more flexible?

---

**Weather Forecast**

- **Data collection:** observation stations, radar, aircraft, balloon, satellites ...
- **Computer simulation:** physics of fluid dynamics
- **Output analysis:** statistics, interpretation

---

(c) Jianhan Chen
How are proteins made in our body?

- Sequence (of amino acids) encoded in DNA (replication)
- DNA transcribed into RNA (transcription)
- Ribosome reads RNA and synthesizes the peptide (translation)

Protein Folding

The newly synthesized peptide chain needs to fold into specific 3D structures before finally becoming functional proteins!

It is a spontaneous process!

We need to understand how it works!

The Protein Folding Problem

- How does the primary Sequence specify the native fold? (Afiński, Science, 1973)

- Levinthal’s Paradox: how protein folds this fast?
  \[ \tau_{\text{fold}} < 1 \text{ second} \]
  10 conformations per residue, \(10^{-11}\) (10 ps) per state
  \[ 10^{46} \text{ states for a 60-residue protein, } >10^{46} \text{ years for random search!} \]

Protein can misfold too!

- Our body has many built-in mechanisms to prevent misfolding, such as simply by removing misfolded proteins
- If out of control, the consequences can be severe (protein misfolding diseases)
**Amyloidosis**

- Abnormal accumulation of insoluble fibrous protein aggregates in various organs
- Implicated in various neurodegenerative diseases such as Alzheimer’s diseases, type II diabetes and over a dozen others

**Alzheimer's Disease (AD)**

- Incurable neurodegenerative disease
- Generally diagnosed for people >65 years old
  - >20 million people affected worldwide
  - 1 out 14 of age 65-70 and ¼ among 85+
  - Predicted to affect 1 in 85 people by 2050
- Aggregation of Abeta (Aβ) peptide
  - 39-43 residues long with unknown function

**Exercise 3: Folding a protein**

- See directory “exercise3”
- Open VMD and load state “play_folding.vmd”
  - 1 nanosecond unfolding simulation of Protein G B1
  - 500 frames @ 1 frame / 2 ps

Q: Can you determine the folding mechanism of protein G B1?

Specifically: assign the sequence of folding for the following elements:

- Helix
  - N-terminal hairpin (residue 1-20)
  - C-terminal hairpin (residue 36-56)

**In Summary**

- Proteins are very large bio-molecules that carry out all kinds of critical functions.
- Proteins adopt specific 3D structures for function.
- Proteins are dynamic living molecules.

Sequence  | Folded Structure  | Functions
-----------|-------------------|----------
catalysis  | transport         | signaling
transport | signaling         | motors ...