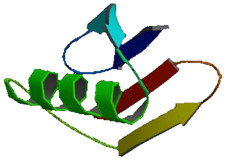
2010 K-State GROW Summer Workshop

How to fold a protein?

Jianhan Chen

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

Department of Biochemistry
Kansas State University

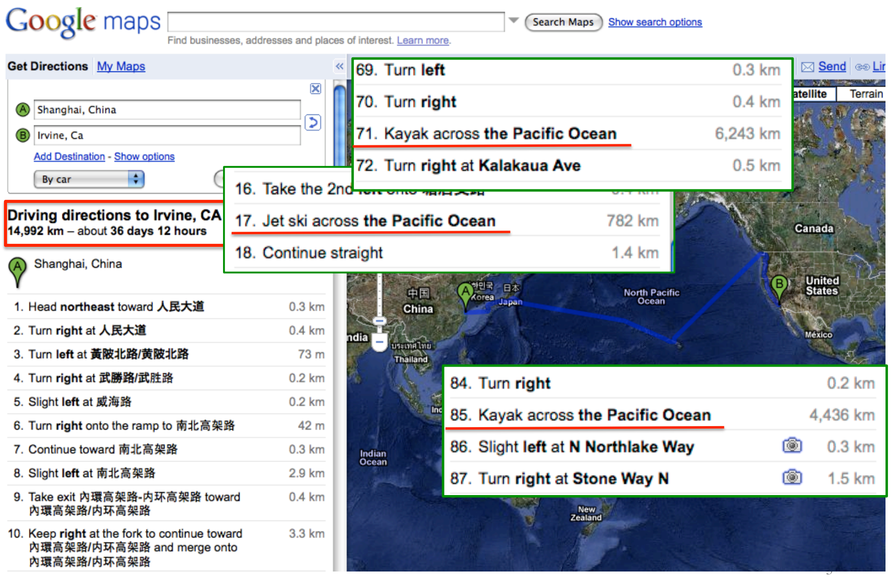


Outline

- 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

---

Who am I: a computational biophysicist from Biochemistry Department



A: Shanghai, China
B: Irvine, Ca

Driving directions to Irvine, CA
14,992 km – about **36 days 12 hours**

Shanghai, China

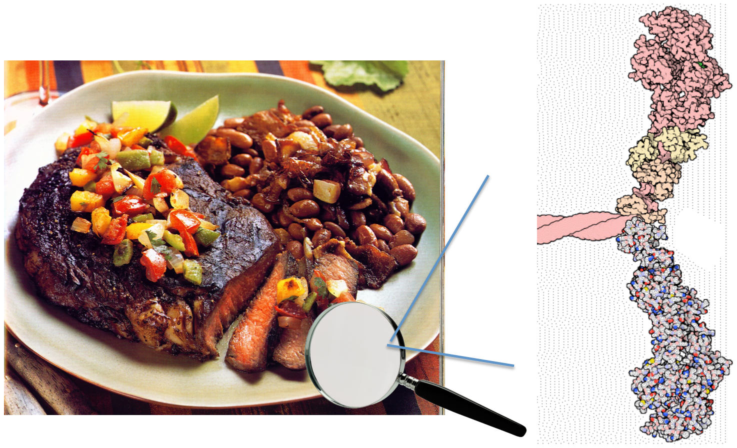
1. Head **northeast** toward **人民大道** 0.3 km
2. Turn **right** at **人民大道** 0.4 km
3. Turn **left** at **黃陂北路/黄陂北路** 73 m
4. Turn **right** at **武勝路/武胜路** 0.2 km
5. Slight **left** at **威海路** 0.2 km
6. Turn **right** onto the ramp to **南北高架路** 42 m
7. Continue toward **南北高架路** 0.3 km
8. Slight **left** at **南北高架路** 2.9 km
9. Take exit **內環高架路-内环高架路** toward **內環高架路/内环高架路** 0.4 km
10. Keep **right** at the fork to continue toward **內環高架路/内环高架路** and merge onto **內環高架路/内环高架路** 3.3 km

16. Take the 2nd
17. Jet ski across **the Pacific Ocean** 782 km
18. Continue straight 1.4 km

69. Turn **left** 0.3 km
70. Turn **right** 0.4 km
71. Kayak across **the Pacific Ocean** 6,243 km
72. Turn **right** at **Kalakaua Ave** 0.5 km

84. Turn **right** 0.2 km
85. Kayak across **the Pacific Ocean** 4,436 km
86. Slight **left** at **N Northlake Way** 0.3 km
87. Turn **right** at **Stone Way N** 1.5 km

Proteins



Myosin:
(molecular muscle)

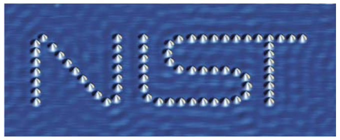
100 million X

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 ~ 10^{-27} kg with a radius ~ 10^{-10} m (1 Å)

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

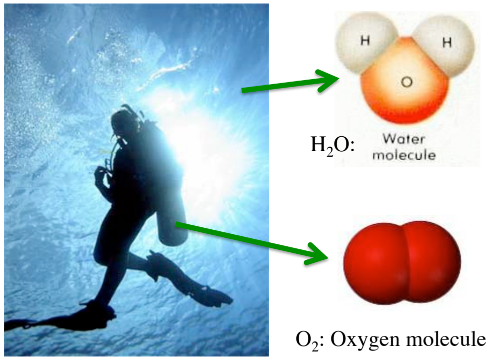
If divided among all people on earth (6.8 billion), each person will receive about a 100 billion atoms!



STM image of a 40-nm logo of NIST made with cobalt atoms, Science, 2004

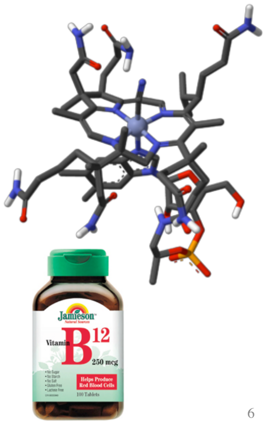
Molecules

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



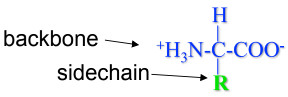
H_2O:

O_2: Oxygen molecule



Protein Molecules

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

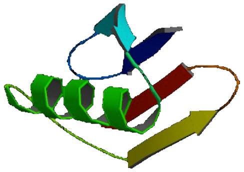


$$\mathrm{{}^{+}H_3N\text{-}\overset{\displaystyle H}{\overset{|}{C}}\text{-}COO^{-}}\quad \underset{\displaystyle R}{\,}$$

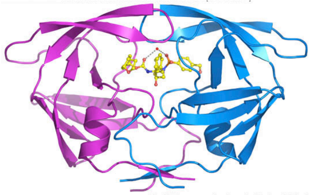
- Most abundant organic chemicals in the body
  - 50% of body's dry weight and 15% of cell content
- Critical in building and maintenance of the body
  - enzymes, motion, transport, structure, regulation, protection ….

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
Radius ~ 2.5 nm

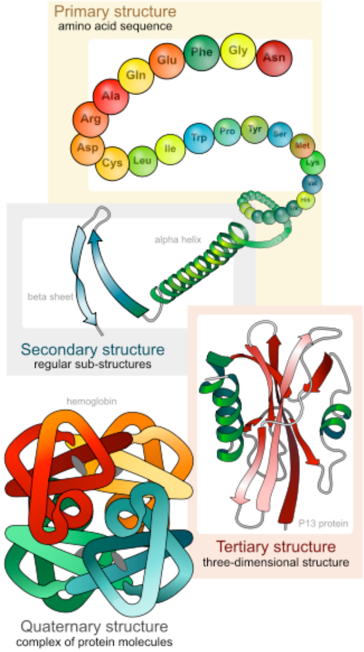


http://dbb.urmc.rochester.edu/labs/wedekind/

HIV-1 Protease (PDB: 2HB3)
198 residues, 1582 atoms
Radius ~ 5 nm

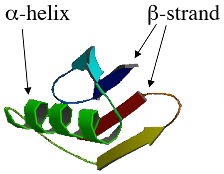
Protein is Structured

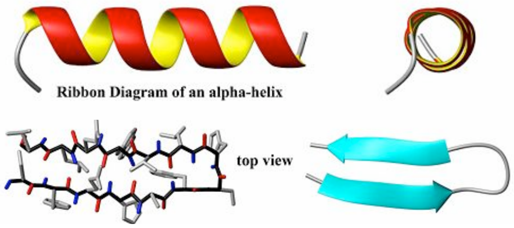
- Hierarchical organizations
  - Primary, secondary, tertiary, quaternary structures

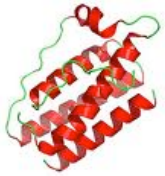


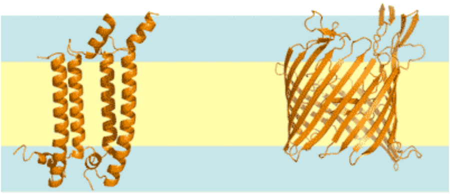
α-Helix and β-Sheet

- Two of the basic secondary structures of proteins









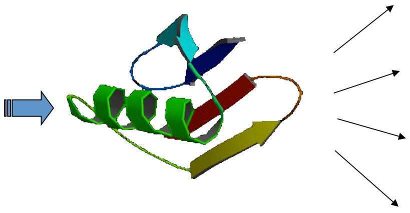
Protein Structure and Function

Sequence

Folded Structure

Functions

MTYKLILNGK
TLKGETTTEA
VDAATAEKVF
KQYANDNGV
DGEWTYDDA
TKTFTVTE …



catalysis

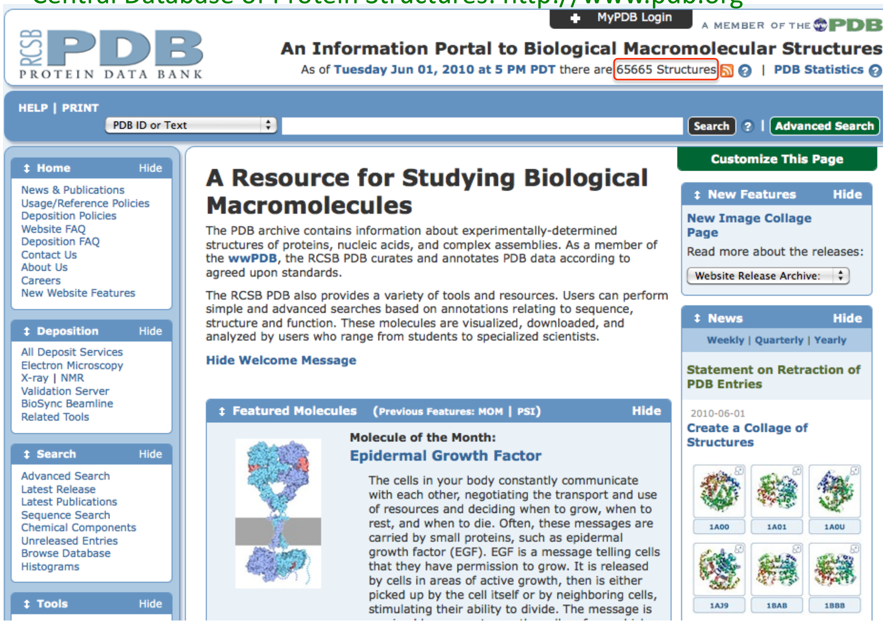
transport

signaling

motors …

- Structural Genomics (NMR, X-ray)
- Protein structure predictions
- Structure-based (rational) drug discovery

Central Database of Protein Structures: http://www.pdb.org



An Information Portal to Biological Macromolecular Structures

As of Tuesday Jun 01, 2010 at 5 PM PDT there are 65665 Structures | PDB Statistics

A Resource for Studying Biological Macromolecules

The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the wwPDB, the RCSB PDB curates and annotates PDB data according to agreed upon standards.

The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists.

Hide Welcome Message

Featured Molecules (Previous Features: MOM | PSI) Hide

Molecule of the Month:

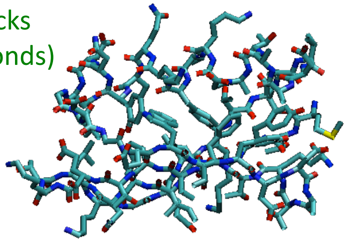
Epidermal Growth Factor

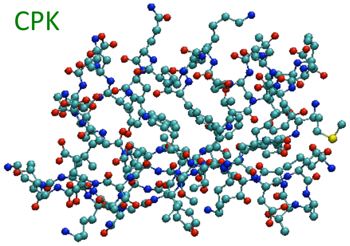
The cells in your body constantly communicate with each other, negotiating the transport and use of resources and deciding when to grow, when to rest, and when to die. Often, these messages are carried by small proteins, such as epidermal growth factor (EGF). EGF is a message telling cells that they have permission to grow. It is released by cells in areas of active growth, then is either picked up by the cell itself or by neighboring cells, stimulating their ability to divide. The message is

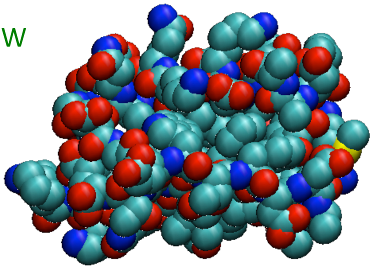
Read more about the releases:

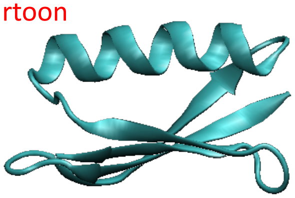
2010-06-01

Visualizing Protein Structure


Sticks (bonds)

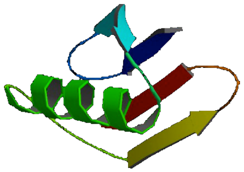

CPK

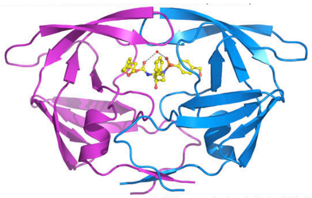

VDW


Cartoon

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)


Protein G B1 (PDB: 1PGB)
56 residues, 436 atoms
Radius ~ 2.5 nm


http://dbb.urmc.rochester.edu/labs/wedekind/

HIV-1 Protease (PDB: 2HB3)
198 residues, 1582 atoms
Radius ~ 5 nm

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 Brower 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?

Detailed Instruction (cont.)

1. Exit VMD and re-open VMD
2. Repeat the operations to open “2HB3.pdb”
3. Repeat the operations to try different drawing and coloring methods

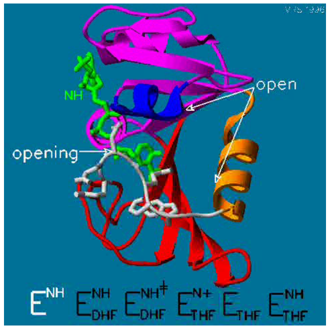
Q: Can you count how many helices and beta strands 2HB3 have?

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

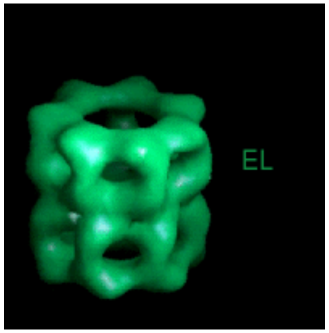
1. Exit VMD and re-open VMD
2. Click “File” on the VMD Main window, and select “Load State” from the pull-down menu (a Windows File explorer window will pop up)
3. Navigate to find the file “view_2hb3.pdb” (located within folder “exercise1” in the “folding” folder on your Desktop); Select the file and 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

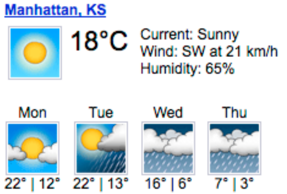


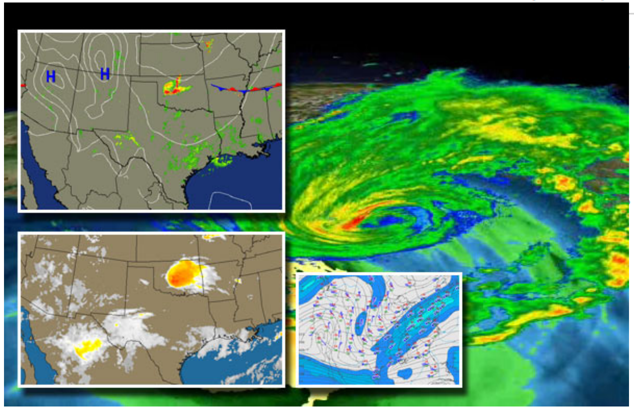
Movie Credits:
DHFR: http://chem-faculty.ucsd.edu/kraut/dhfr.html
GroEL: http://people.cryst.bbk.ac.uk/~ubcg16z/cpn/elmovies.html

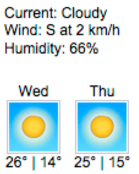
---

Weather Forecast

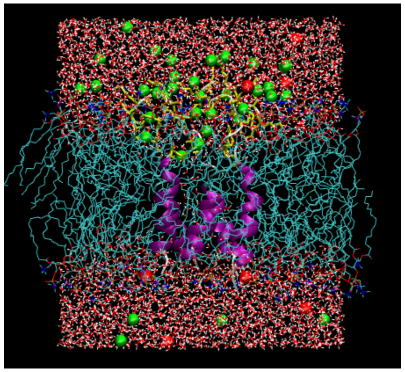
- Data collection: observation stations, radar, aircraft, balloon, satellites …
- Computer simulation: physics of fluid dynamics
- Output analysis: statistics, interpretation



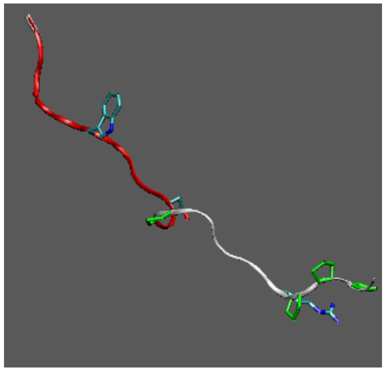




Molecular Dynamics Simulations



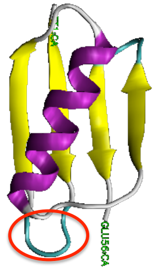
Channel-forming peptides
in a fully solvated membrane bilayer;
Channel: 1795 atoms; All: 26254 atoms

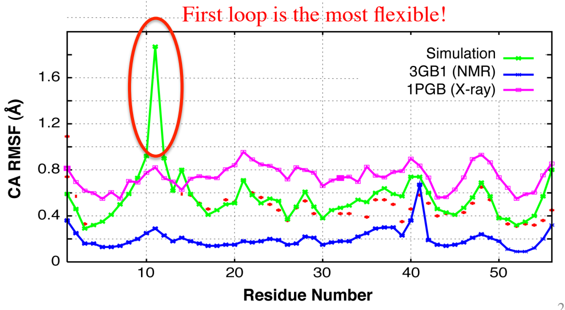


Ab initio folding of a designed mini-protein Trp-Cage in implicit solvent
(System size: 304 atoms)

Exercise 2: Visual analysis of protein dynamics

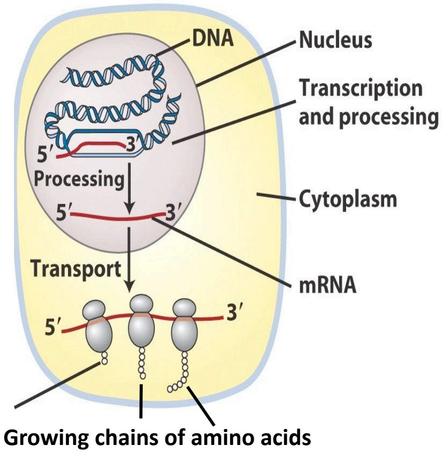
- See directory “exercise2”
  - Open VMD and load state “play_md.vmd”
  - 1 nanosecond simulation of Protein G B1 at 300 K
  - A total 500,000 molecular dynamics steps
  - Snapshots taken every 1000 steps (500 frames @ 1 frame/2 ps)
- Assignment: can you tell which parts of the protein are more flexible?





How are proteins made in our body?

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

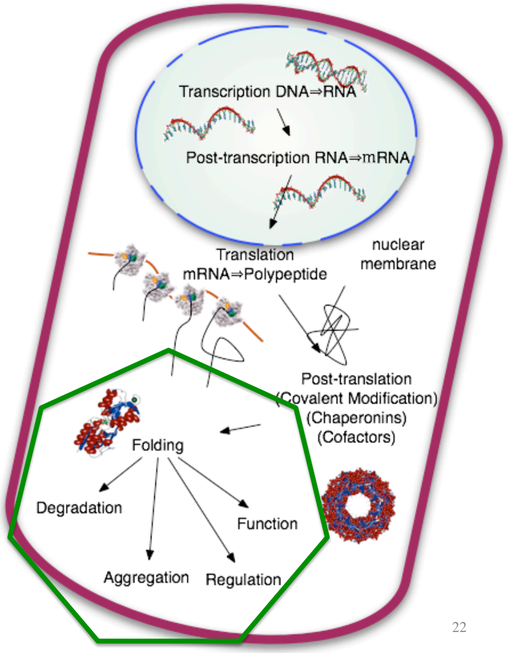


Protein Folding

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

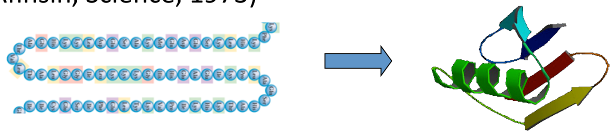
It is a spontaneous process!

We need to understand how it works!



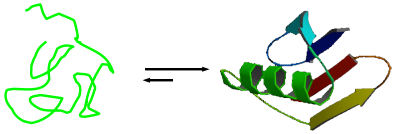
The Protein Folding Problem

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



- Levinthal's Paradox: how protein folds this fast?

$\tau_{\text{fold}} < 1$ second



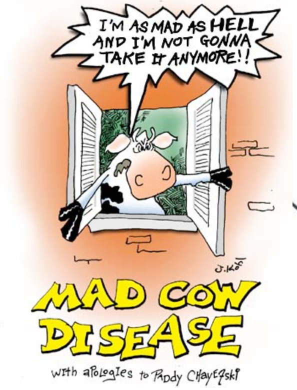
10 conformations per residue,
10^{-11} (10 ps) per state

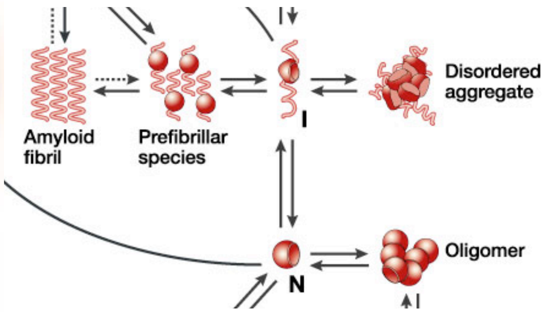
10^{60} states for a 60-residue protein,
$>10^{40}$ years for random search !

(c) Jianhan Chen

Protein can misfold too!

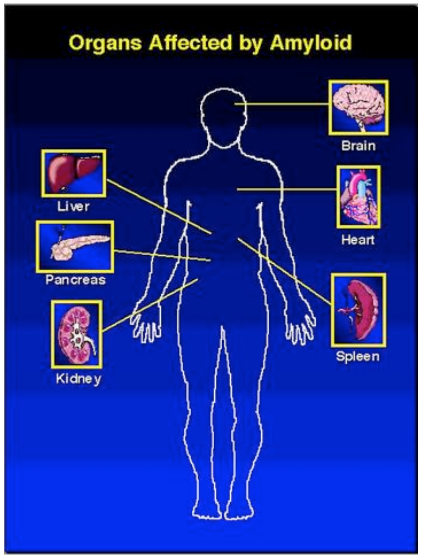
- Our body has many built-in mechanisms to prevent misfolding, such as simply by removing misfolded protiens
- If out of control, the consequences can be severe (protein misfolding diseases)





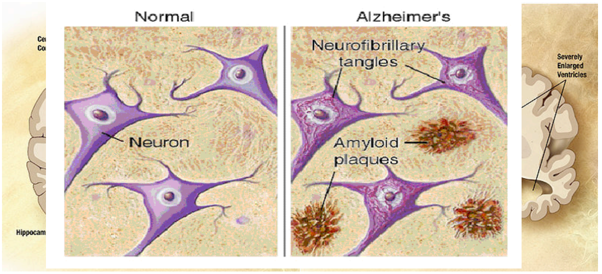
Amyloidosis

- 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 world wide
  - 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?

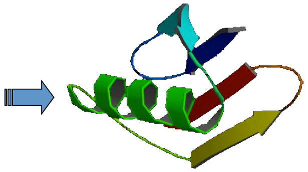
Specifically: assign the sequence of folding for the following elements:

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.

MTYKLILNGK
TLKGETTTEA
VDAATAEKVF
KQYANDNGV
DGEWTYDDA
TKTFTVTE …



catalysis
transport
signaling
motors …

Sequence   Folded Structure   Functions