Protein Images
A protein is a solid object.
SEE THE BONDED ATOMS

Draw the bonds as sticks. It is hard to see what is important in the structure.
SEE THE BONDED ATOMS

Draw the bonds as sticks. It is hard to see what is important in the structure.
TRY TO SEE THE MAIN CHAIN 1

It is still not clear. Where are the ends?
TRY TO SEE THE MAIN CHAIN 3

Can now find the ends.

Start or Finish?

Emphasize the secondary structure (alpha-helices and beta-sheets).
TRY TO SEE THE CHAIN PATH

Can now find the ends.

Start or Finish?

Use color to see the chain direction. Start (N) is blue (positive). End (C) is red (negative).
Proteins are stable in water. Water molecules and proteins are both just atoms.

Now you see the protein is interacting with the water.
Underlying Physics
BOLTZMANN’S DISTRIBUTION

Two minima corresponding to states A and B

Area gives probability of state A or B

- Probability of system being at position $x$ is
  $$P(x) = \frac{\exp(-U(x)/kT)}{Q}.$$  
  $U(x)$ is Potential Energy at position $x$.

- Find $Q$, the "Partition Function", so total probability is 1.
  $$Q = \sum \exp(-U(x)/kT).$$
Imagine \( a(t) \) is some value that changes with time.

This time-course is characterized by:

- Amplitude calculated as (just like Standard Deviation)
  \[
  \text{Amp} = \sqrt{\frac{1}{n} \sum a^2(t) - \left[ \frac{1}{n} \sum a(t) \right]^2}
  \]

- Rate or time constant, \( \tau \) (this is a "typical time").
TYPES OF MOTION

Periodic or Harmonic motion

Distribution of Value

1/cos Distribution

Random or Stochastic motion

Distribution of Value

Normal Distribution
The actual transition from State A to B is very quick (a few picoseconds).

What takes time is the waiting. The average wait before going from A to B is:

$$\tau_{A\rightarrow B} = \left( \frac{h}{k_b T} \right) \exp \left[ +\frac{\Delta G}{k_b T} \right]$$

where $$\Delta G = (G_T - G_A)$$. 

$$\left( \frac{h}{k_b T} \right) \sim 0.016 \text{ picoseconds at } T = 300^\circ K (27^\circ C)$$. 

$h$ is Planck’s constant, $k_b$ is Boltzmann’s constant.
## Rates of Motion

<table>
<thead>
<tr>
<th>PROCESS</th>
<th>TIME (ps)</th>
<th>FREQUENCY (cm⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond Stretching</td>
<td>0.01</td>
<td>3000</td>
</tr>
<tr>
<td>Angle Bending</td>
<td>0.1</td>
<td>300</td>
</tr>
<tr>
<td>Methyl Rotation</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>Water Tumbling</td>
<td>20</td>
<td>1.5</td>
</tr>
<tr>
<td>Protein Tumbling</td>
<td>20,000</td>
<td>0.0015</td>
</tr>
<tr>
<td>Enzyme Reaction</td>
<td>10⁷ to 10⁹</td>
<td>10⁻⁵ to 10⁻⁸</td>
</tr>
<tr>
<td>Protein Folding</td>
<td>10⁹ to 10¹²</td>
<td>10⁻⁸ to 10⁻¹¹</td>
</tr>
</tbody>
</table>
Simulation Methods
MOVING OVER ENERGY SURFACE

- Energy Minimization drops into local minimum.

- Molecular Dynamics uses thermal energy to move smoothly over surface.

- Monte Carlo Moves are random. Accept with probability $\exp\left(-\frac{\Delta U}{kT}\right)$. 

©Michael Levitt 07
Molecular Dynamics
\[ U = \sum \frac{1}{2} K_b (b-b_0)^2 - \sum \frac{1}{2} K_\theta (\theta-\theta_0)^2 \]

- All Bonds
- All Angles

\[ + \sum K_\Phi [1-\cos(n\phi+\delta)] \]

- All Torsion Angles

\[ + \sum \varepsilon \left[ \left( \frac{r}{r_0} \right)^6 - 2 \left( \frac{r}{r_0} \right)^2 \right] \]

- All nonbonded pairs

\[ + \sum \frac{332q_i q_j}{r} \]

- All partial charges

Simple sum over many terms

©Michael Levitt 07
Molecular Dynamics Theory

- All atoms move together.
- Forces between atoms change with time.
- Analytical solution to give $x(t)$ and $v(t)$ is impossible.
- Numerical solution is trivial.

\[
x(t + \Delta t) = x(t) + v(t) \Delta t + \left[ 4a(t) - a(t - \Delta t) \right] \Delta t^2 / 6
\]

New position  Old position  Old velocity  Acceleration

\[
v(t + \Delta t) = v(t) + \left[ 2a(t + \Delta t) + 5a(t) - a(t - \Delta t) \right] \Delta t / 6
\]

New velocity  Old velocity  Acceleration

\[
U_{\text{kinetic}} = \frac{1}{2} \sum m_i v_i(t)^2 = \frac{1}{2} n k_B T
\]

Kinetic energy  Atomic masses, velocities  Boltzmann's Constant  Number of coordinates (not atoms)  Temperature

Total energy ($U_{\text{potential}} + U_{\text{kinetic}}$) must not change with time.

Time step, $\Delta t$, must be very small at $10^{-15}$ seconds or 0.001 ps.
Simulating Solutions
WATER IS A VERY SIMPLE MOLECULE

The water molecule is one of the simplest.

The properties of liquid water are not simple.

Water and solutions have emergent properties.

Many simple objects lead to complexity.

Simple geometry

Simple electrostatics
SIMULATING LIQUIDS

- Periodic box with 216 water molecules. Simulate motion for 100 ps.

- Calculate key experimental properties:
  - Heat of vaporization.
  - Structure.
  - Internal pressure.
  - Diffusion constant.

- Compare with experiment and calibrate.

Rahman & Stillinger 1970
LIQUIDS: ARGON AND WATER

Argon is like a collection of hard spheres. Each Argon has 12 to 14 neighbors.

Water has an open structure. Due to tetrahedral geometry, each water has 4 to 5 neighbors.
PURE WATER
DYNAMICS
IN WATER AT ROOM TEMPERATURE
A Box of Water Molecules 47.6 ps
HYDROPHOBIC EFFECT

- Hydrophobic molecules cluster in water (not soluble).
- The energy is proportional to the surface area buried in cluster.
- This is not a pairwise additive force.

Box with periodic boundaries.
• Make a fence between A and B so that all points closer to A are on left, all points closer to B on right. Easy.

• Now add third point C.

• How about adding points D and E?
VORONOI ANALYSIS OF CONTACTS

- Measure cluster formation by Voronoi construction.
- Contact measured by area of face common to A and B.
- Distance (AB) same as distance (BC), but only A and B touch though a shared face.
BENZENE SOLUTION
DYNAMICS
IN WATER AT ROOM TEMPERATURE
Slope of line is about 40 cal/mol per Å² which is close to the experimental value.
Simulating Alpha-Helix Unfolding
DYNAMICS OF THE ALPHA-HELIX

- Put α-helix in a box or water.
- Run molecular dynamics.
- Focus attention on hydrogen bonds.
WHY SIMULATE UNFOLDING

Rate of motion
Velocity $\propto \sqrt{\text{Temperature}}$

Rate of jumping
Barrier Height is $\Delta G$

Time $= 10^{-13} \exp \left[ \frac{\Delta G}{RT} \right]$

- At $200^\circ C$, move 25% faster than at $25^\circ C$.
- At $200^\circ C$, can get over a barrier 1,000,000 to 1,000,000,000 times faster than at $25^\circ C$. 
UNFOLD ALPHA-HELIX

13 Alanine residues

- Start as an ideal α-helix. In a box of water.

- Run 200 ps (100,000 time steps) of molecular dynamics at six different temperatures.

- Record percentage α-helix formed for last 50 ps.

- See temperature-induced melting on picosecond time-scale.

Put it in a box of water.

Temperature (K)

Percentage Helix

©Michael Levitt 07
ALPHA-HELIX DYNAMICS IN WATER AT HIGH TEMPERATURE
Alpha-Helices Unfolding in Solution
EFFECT OF TEMPERATURE

At higher temperature, the helix breaks down more rapidly.

In vacuo the helix is very stable even at high temperature.

In water the helix is unstable at high temperature.

The rate of melting depends on temperature.

This happens because water molecules stabilize the transition state.
WATER ALLOWS HYDROGEN BONDS TO BREAK

Intact hydrogen bond in helix.

Hydrogen bond is breaking.

Free Energy barrier between states is much lower in water

Water catalyzes the breakage of hydrogen bonds by stabilizing the transition state.
Simulating Folded Protein
Native BPT1 in Water at 298 K  1.2 ps
Unfolding Simulations
BPTI UNFOLDING AT HIGH TEMPERATURE
Reduced BPTI in Water at 498 K 310.8 ps
Folding Simulations
The native state is believed to have the lowest free-energy. Run down to it.

Run down to it.

Sounds Easy.
LEVINTHAL'S PARADOX

- If the energy surface is flat except for one deep minimum (golf-course) landscape, then proteins would never fold. Too many possible states.

- If each of 100 amino acids has 5 states, then the entire chain has $5^{100}$ states. Too big!

Free Energy, $\Delta G$

Extent of Folding

©Michael Levitt 07
Why should the free-energy surface be flat? The same energy terms that stabilize the native state should stabilize progressively near-native states.

This is a folding funnel.

With it, it does not matter how large the space of possible states is.
FOLDING@HOME

Fold proteins on 250,000 computers using the program as a Screen Saver!

Most powerful resource in the world.

Like SETI@home
History of Simulation
The Birth of the Monte Carlo Method.

When any sufficiently large nuclear explosion occurs within a container, unless the radioactive material is properly contained and the timing of triggering explosions perfect, neutrons stream out of one side of the container. This leak causes an asymmetrical, much weaker, and more unpredictable blast. In order to make the most potent blast possible, a series of complex events must be modeled so that the radioactive material explodes symmetrically. This research appears under the hygienic guise of solving the "neutron diffusion problem." Until 1943, when von Neumann and Stanley Ulam worked on the neutron diffusion problem, there were essentially only two sorts of modeling employed by scientists and mathematicians to describe complex events: deterministic methods (which are essentially applied mathematics) and variations on stochastic techniques (which were known simply as simulation).

To get around the apparently inevitable incorporation of the random, von Neumann devised a third kind of simulation called the "Monte Carlo" in homage to the games of luck he enjoyed in the gambling capital of Europe. He held that random elements in simulations were unacceptable, a form of contamination tantamount to cheating at cards. Indeed, his aversion to stochastic modeling and his appreciation of rule-based games is at the heart of his epistemology. In the Monte Carlo simulation, Von Neumann devised a non-stochastic formula for approximating the stochastic operators in non-trivial simulations. Essentially, he had found a deterministic way to model random events. At the same time, he had rigged the game in the house's favor. When the Monte Carlo simulation worked, it suggested not only that we could describe nature without relying on randomness or chance, but that nature itself was deterministic.
Molecular Model-building by Computer

In which biochemists observe models of giant molecules as they are displayed on a screen by a computer and try to fold them into the shapes that they assume in nature

by Cyrus Levinthal

Many problems of modern biology are concerned with the detailed relation between biological function and molecular structure. Some of the questions currently being asked will be completely answered only when we have an understanding of the structure of all the molecular components of a biological system and a knowledge of how they interact. There are, of course, a large number of problems in biology which biologists have some insight but concerning which they cannot yet ask suitable questions in terms of molecular structure. As they are such problems more clearly, however, they inevitably find an increasing need for structural information. In our laboratory at the Massachusetts Institute of Technology we have recently started using a computer to help gain such information about the structure of large biological molecules.

For the first half of this century the metabolic and structural relations among the small molecules of the living cell were the principal concern of biochemistry. The chemical reactions these molecules undergo have been studied extensively. Such reactions are catalyzed by the large protein molecules called enzymes, many of which have now been purified and also studied. It is only within the past few years, however, that X-ray diffraction techniques have made it possible to determine the molecular structure of such protein molecules. These giant molecules, which contain from a thousand to tens of thousands of atoms, constitute more than half of the dry weight of cells. Protein molecules not only act as enzymes but also provide many of the cell's structural components. Further class of giant molecules, the nucleic acids, determine what kind of protein the cell can produce, but most of the physiological behavior of a cell is determined by the properties of its proteins.

The X-ray diffraction methods for investigating the three-dimensional structure of protein molecules are difficult and time-consuming. So far the structure of only three proteins have been worked out: myoglobin, hemoglobin, and lysozyme [see "The Three-dimensional Structure of a Protein Molecule," by John C. Kendrew, Scientific American, December, 1964], and the Hens-globin Molecules," by M. R. Perutz, November, 1964]. In their studies of the hemoglobin molecule M. R. Perutz and his associates at the Laboratory of Molecular Biology in Cambridge, England, have observed that the structure of the molecule changes slightly when
50 YEARS OF SIMULATION

- We have 10,000,000 times more resources.
- Systems have become larger (100 times).
- Runs have become longer (100,000 times).
- Energy functions have become simpler.
- Fit reality well. Nothing bad has happened!

1955 Argon 1970 Water 1988 Protein in Water

©Michael Levitt 07
MOLECULAR POTENTIAL ENERGY

\[ U = \sum \frac{1}{2} K_b (b-b_0)^2 - \sum \frac{1}{2} K_\theta (\theta-\theta_0)^2 \]

All Bonds  All Angles

\[ + \sum K_\phi [1 - \cos(n\phi + \delta)] \]

All Torsion Angles

\[ + \sum \epsilon \left[ \left( \frac{r_0}{r} \right)^{12} - 2 \left( \frac{r_0}{r} \right)^6 \right] \]

All nonbonded pairs

\[ + \sum \frac{\epsilon q_i q_j}{r} \]

All partial charges

Simple sum over many terms
\[
U^{TOTAL} = \min_t \left\{ \sum_{ab} \left( U^{ES}_{ab}(t_a, t_b; R_a, R_b) + U^{EX}_{ab}(r_{ab}) \right) + \sum_a U^{IN}_a(t_a; R) \right\} + \sum_{ab} U^{DS}_{ab}(R_{ab})
\]

where \( R_{ab} = |R_a - R_b| \) and \( r_{ab} = |(R_a + t_a) - (R_b + t_b)| \)

\[
U^{ES}_{ab}(t_a, t_b; R_a, R_b) = \tilde{Z}_a \tilde{Z}_b \varphi(R_{ab}; 0, 0) + \tilde{Q}_a \tilde{Q}_b \varphi(r_{ab}; \tilde{w}_a, \tilde{w}_b)
\]
\[
+ \tilde{Q}_a \tilde{Z}_b \varphi(R_a + t_a - R_b; \tilde{w}_a, 0)
\]
\[
+ \tilde{Z}_a \tilde{Q}_b \varphi(R_b + t_b - R_a; 0, \tilde{w}_b)
\]

\[
U^{EX}_{ab}(r_{ab}) = \tilde{C}_a \tilde{C}_b \left[ 1 + \left( \frac{2r_{ab}}{\tilde{w}_a + \tilde{w}_b} \right)^2 \right] \exp \left( -\frac{2r_{ab}}{\tilde{w}_a + \tilde{w}_b} \right)
\]

\[
U^{IN}_a(t_a; R) = \frac{\tilde{Q}_a t_a}{\tilde{\alpha}_a} \left( 1 - \sqrt{1 - \frac{1}{\tilde{t}_a^2} (t_a - t_a^0)^2} \right)
\]
where \( t_a^0 = -\sum_b \tilde{r}_{ab} \vec{n}_{ab} / \tilde{w}_0 \)

\[
U^{DS}_{ab}(R_{ab}) = -\frac{2 \vec{E}_a \vec{E}_b}{\vec{E}_a + \vec{E}_b} \left( \frac{\vec{R}_0}{R_{ab} + \min(\vec{R}_a, \vec{R}_b)} \right)^6
\]
THE END
of Lecture II