COMPUTATIONAL STRUCTURAL BIOLOGY
STRUCTURE, SIMULATION, FUNCTION & PREDICTION

Lecture 5

Michael Levitt
Structural Biology, Stanford

http://csb.stanford.edu/class
MOLECULAR SIMULATION II

Normal Mode Theory.
Protein Normal Modes.
Unfolding Alpha-Helix.
Unfolding Proteins.
Folding Simple Models.
Folding Simulations.
Normal Mode Theory
Concept 5.1
NORMAL MODE DYNAMICS

• In regular Molecular Dynamics, we solve the exact equations of motions approximately.

• In Normal Mode Dynamics, we solve the approximate equations of motion exactly.

• We make a quadratic approximation to the potential energy function.
BASIC THEORY

• A string attached at both ends

Get a standing wave of frequency \( v = \) Amplitude is proportional to Each mode can be excited

• Discrete point

\[ U(x) = \frac{1}{2} C x^2. \] Now \( F = ma = -C x \) or \( m \frac{d^2 x}{dt^2} = -C x \)

Solution is \( x(t) = a \cos (\omega t + \delta) \), with \( \omega = \)

Get amplitude, \( a \), by the equipartition

\[ \langle E_{\text{potential}} \rangle = \frac{1}{2} C \langle x^2 \rangle = \frac{1}{2} k_a \]

Thus, \( a = \sqrt{\text{rt} \left( 2k_a T/C \right)} \) as \( \langle x^2 \rangle = \frac{1}{2} a^2 \) for a cosine wave.

\( \text{[Michael Levitt 04]} \)
NORMAL MODES IN HIGH DIMENSIONS

Focus on deepest energy

- Expand energy function about minimum.
- Approximate as a quadratic function
  \[ F(x,y) = Ax^2 + Bxy + Cy. \]
NORMAL MODES IN HIGH DIMENSIONS.

- Release a marble on this surface and watch the motion.
- Only if released on the orange lines will the motion be in a straight line.
- These orange lines are the normal mode directions.
NORMAL MODES IN HIGH DIMENSIONS.

- The normal mode directions are the major and minor axes of the ellipse.
- All other motion is a linear combination of these basic motions.
- Solving for the modes requires a matrix that is $N \times N$, where $N$ in number of degrees of freedom.

For $n$ atoms, $N = 3n$ so it can be very big!
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

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

All Torsion

\[ + \sum \varepsilon [(r/r_0)^2 - 2(r/r_0)^6] \]

All nonbonded

\[ + \sum 332q_i q_j/r \]

All partial

Eliminate the strongest springs.
POTENTIAL ENERGY IN TORSION SPACE

\[ U = \sum K_n [1 - \cos(n\phi + \delta)] \]

All Torsion

\[ + \sum \varepsilon [(r_{0i})^2 - 2(r_{0i})^6] \]

All nonbonded

\[ + \sum 332q_i q_j / r \]

All partial

- A protein with \( N \) residues has about \( 4N \) \((\phi, \psi, \chi)\) single bond torsion angles.
- The same protein has about \( 50N \) Cartesian coordinates \((x, y, z)\).
THEORY OF NORMAL MODES I

• Assume Potential energy, $V$, is quadratic function of $\phi$.
  
  $$V = \frac{1}{2} \sum_{ij} V_{ij} (\phi_i - \phi_i^0) (\phi_j - \phi_j^0)$$

• This means that $V_{ij} = \frac{d^2V}{d\phi_i d\phi_j}$

• Assume Kinetic energy, $T$, is quadratic function of $d\phi/dt$.
  
  $$T = \frac{1}{2} \sum_{ij} T_{ij} (d\phi_i/dt)(d\phi_j/dt)$$

• This means that $T_{ij} = \frac{d^2T}{d(d\phi_i/dt) d(d\phi_j/dt)}$

Note the symmetry between Potential and Kinetic energy.
THEORY OF NORMAL MODES II

• Solve for $\phi(t)$ using Lagrangian approach.

$$\sum T_{ij} \left( \frac{d^2 \phi_j}{dt^2} \right) = \sum V_{ij} \Delta \phi_j$$

• Try a periodic function for $\phi(t)$:

$$\Delta \phi_j(t) = \sum A_{ij} \cos(\omega_i t)$$

$$\frac{d^2 \phi_j(t)}{dt^2} = \sum A_{ij} \omega_i^2 \cos(\omega_i t)$$

• In Matrix notation the Lagrangian equation is:

$$TA \omega^2 = VA$$

This is Eigenvalue equation that is easily solved.

$\text{Michael Levitt 04}$
Protein Normal Modes
Concept 5.2
RATES OF VIBRATION

For Bovine Pancreatic Trypsin Inhibitor
58 residues, 208 torsion angles.

- There is a broad range of torsion angle mode frequencies.

- Peak near 30 cm$^{-1}$, which is a period of 1 ps.

- Lowest frequency is at 3 cm$^{-1}$ or 10 ps.

- There are 12 modes below 10 cm$^{-1}$.

AMPLITUDES OF VIBRATION

- Almost all the motion of the CA atoms comes from the lowest frequency modes.
- There is high-frequency motion of the torsion angles.

- The CA Amplitude is the RMS movement of all CA atoms as a result of activating the particular mode.
- The Torsion Angle Amplitude is the RMS movement of all torsion angles as a result of activating the particular mode.
TRYPsin INHIBITOR MODES

BPTI

The active Lys 15

Top Twist
4.5 cm⁻¹

Bending
5.9 cm⁻¹

Sliding Feet
6.4 cm⁻¹

Breathing
7.1 cm⁻¹
BPTI NORMAL MODES

AT HIGH TEMPERATURE
LYSOZYME MODES

An inhibitor, which is colored in green, is bound in the active site. The inhibitor is not included in the normal mode calculations.
LYSOZYME NORMAL MODES AT HIGH TEMPERATURE
RIBONUCLEASE MODES

An inhibitor, which is colored in green, is bound in the active site. The inhibitor is not included in the normal mode calculations.
RIBONUCLEASE
NORMAL MODES
AT HIGH TEMPERATURE
Unfolding Alpha Helix
Concept 5.3
ALPHA-HELIX UNFOLDING

Why Simulate Unfolding?
Unfold Alpha-Helix.
Effect of Temperature.
Effect of Environment.
Hydrogen Bond Breaking.
($\phi,\psi$) Distributions.
WHY SIMULATE UNFOLDING

- At 200°C, move 25% faster than at 25°C.
- At 200°C, can get over a barrier 1,000,000 to 1,000,000,000 times faster than at 25°C.

Rate of motion

Velocity \( \propto \sqrt{\text{Temperature}} \)

Rate of jumping

Barrier Height is \( \Delta G \)

Time = \( 10^{-13} \exp[\Delta G/RT] \)

©Michael Levitt 04
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.
EFFECT OF TEMPERATURE

- At higher temperature, the helix breaks down more rapidly.

HELIX LESS STABLE IN WATER

- 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

Water catalyzes the breakage of hydrogen bonds by stabilizing the transition state.

Free Energy barrier between states is much lower in water.

Michael Levitt 04
HELIX UNFOLDING
IN WATER AT HIGH TEMPERATURE
(φ, ψ) Distributions in Solution

- Distribution of Ala residues in globular proteins is similar to the distribution found for Alaβ in water at high temperature.

- There are differences shown in orange.
Unfolding Proteins
Concept 5.4
PROTEIN UNFOLDING

What Happens to Secondary
What Happens to Aromatic Core.
The Molten Globule?
Connection to Experiment.
WHAT HAPPENS TO SECONDARY STRUCTURE

Secondary

Hydrogen
PROTEIN UNFOLDING,
IN WATER AT HIGH TEMPERATURE
WHAT HAPPENS TO AROMATIC SIDECHAINs

Native

?
PROTEIN UNFOLDING
IN WATER AT HIGH TEMPERATURE
UNFOLDING PATHWAY

Unfolding occurs in two stages:
1. Loosen packing, keep secondary structure and functional groups intact.
2. Lose secondary structure and gain random coil.

Time

Native

Intermediates

Intermediate

Unfolded States

Native

Op

119p

170p

224p

350p

400p

545ps

CA RMS
**CONNECTION TO EXPERIMENT.**

- \( \Phi_{TS} = \frac{(\Delta G_{TS} - \Delta G_u)}{(\Delta G_N - \Delta G_u)} \)

- \( \Delta G_{TS} = \Delta G_u \) means that TS is like U.
  \( \Phi_{TS} = 0 \).

- \( \Delta G_{TS} = \Delta G_N \) means that TS is like N.
  \( \Phi_{TS} = 1 \).

\( \Phi_{TS} \) measures the relative effect of changing a particular residue to Ala.

*Fersht et al, Nature 342, 122 (1989)*
**Connection to Experiment,**


- $\Phi_{TS}$ measures the relative effect on the folding rate of changing a particular residue to Ala. $\Phi_{TS}=0$ implies a denatured-like transition state structure and $\Phi_{TS}=1$ implies a native-like transition state structure at that position.

- Residues found experimentally to have high $\Phi_{TS}$ values are found to interact in the simulated transition state.
RESIDUES marked in red remain native-like in SH3 unfolding simulations. Experiment implicates residues from the same region of the structure.


Experimentalists and theoreticians are interacting to study protein unfolding.

Michael Levitt 04
Folding Simplified Chains Concept 5.5
FOLD SIMPLIFIED PROTEINS

Folding With Minimization.
Why Folding is so
Lattice Model Folding.
WHY IS FOLDING SO DIFFICULT

Barrier crossing time $\sim \exp(\Delta G/kT)$.

- Must get over high barriers.
- Many degrees of freedom: huge set of possible structures.

Native State
Compact,

Unfolded State
Expanded,

Molten Globule State
Compact, Disordered

Barrier Height

Barrier crossing time
$\sim \exp(\text{Barrier})$
SIMPLIFIED MODELS FOR FOLDING

- Starting Structure
- 1000 steps energy minimization fold chain.
- Use normal mode perturbation to escape from local minimum.

Thermalization energy jumps.
Virtual Bonds

- $\alpha_i$ is defined by $\text{CA}_{i-1} - \text{CA}_i - \text{CA}_{i+1} - \text{CA}_{i+2}$
- $\alpha_i$ is approximately $\psi_i + \phi_{i+1} + 180$
SIMPLIFIED MODELS FOR FOLDING.

- Simplify the atomic structure to one center per residue (CA).
- Use effective hydrophobic forces with one degree of freedom per residue.
- Fold by energy minimization.
- Use Normal mode thermalization to escape local minima.

       Native
       X-Ray Structure

       Folded.

       dRMS=5.5Å
       cRMS=8.1Å

This seemed significant in 1975.
The overall chain path is similar.
It was not really significant.

Levitt & Warshel.
CARTOON FOLDING AND UNFOLDING OF BPTI
3x3x3 CUBE PROTEIN FOLDING

- Model protein as 27 centers of 3 by 3 by 3 cube. This is a lattice.
- Have a simple pair-wise energy.
- Have a simple move set.
- Fold with Monte Carlo: accept new arrangement if
  \[ \exp(-\Delta U/kT) > \text{Ran}(1), \]
  where Ran(1) is a random number between 0 and 1.
- Repeat about 1,000,000 times to get native cube state.

LATTICE MODEL OF FOLDING

Folding Simulations Concept 5.6
FOLDING SIMULATIONS

Need Massive Computational
Villin Folding (A Small Proteins).
Blue Gene Project.
Folding@Home.
Other Simulations.
NEED MASSIVE COMPUTATIONAL RESOURCES

Empty Supercomputers.
Blue Gene (IBM).
Folding@home (Vijay Pande).
VILLIN FOLDING

- Use explicit water molecules with 36-residue villin headpiece.
- Have between 3,000 and 6,500 water molecules.
- Run for 200 ns (tour de force).
- Get to within 4.7 Å RMS.

Duan, et al. PNAS, 95, 9897 (1998)
IBM BLUE GENE PROJECT

IBM BLUE GENE DESIGN


- 5 years?
- $100
FOLDING AT HOME

http://www.stanford.edu/group/pandegroup/folding/education/

- Fold proteins on 100,000 computers using the program as a Screen Saver!
- Most Powerful resource in the world
FOLDING AT HOME HELIX FOLDING

- Run with effective solvent (pseudo vacuum) using Tinker, Jay Ponder's molecular dynamics simulation program.

- Reproducibly fold helix in 10's of nanoseconds (10,000,000 Δt steps).
HELIX FOLDING IN
IMPLICIT WATER
AT ROOM
TEMPERATURE
VILLIN FOLDING IN IMPLICIT WATER AT ROOM TEMPERATURE
FOLDING AT HOME RATES

- Can look at rates from 1 to 50,000 nanoseconds (50 microseconds).
- Get good fit to experiment over the entire range.