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

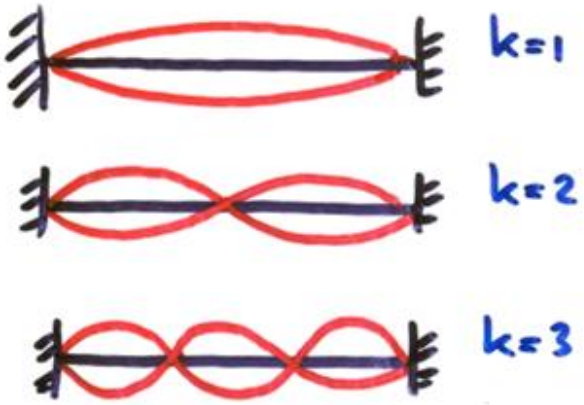
What are normal

- A string attached at both

Get a standing wave of frequency $\nu =$

Amplitude is proportional to

Each mode can be excited

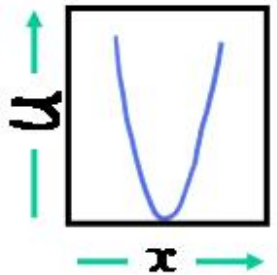


- Discrete point

C is the force constant

$U(x) = 1/2 Cx^2$. Now $F = ma = -Cx$ or $m d^2x/dt^2 = -Cx$

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



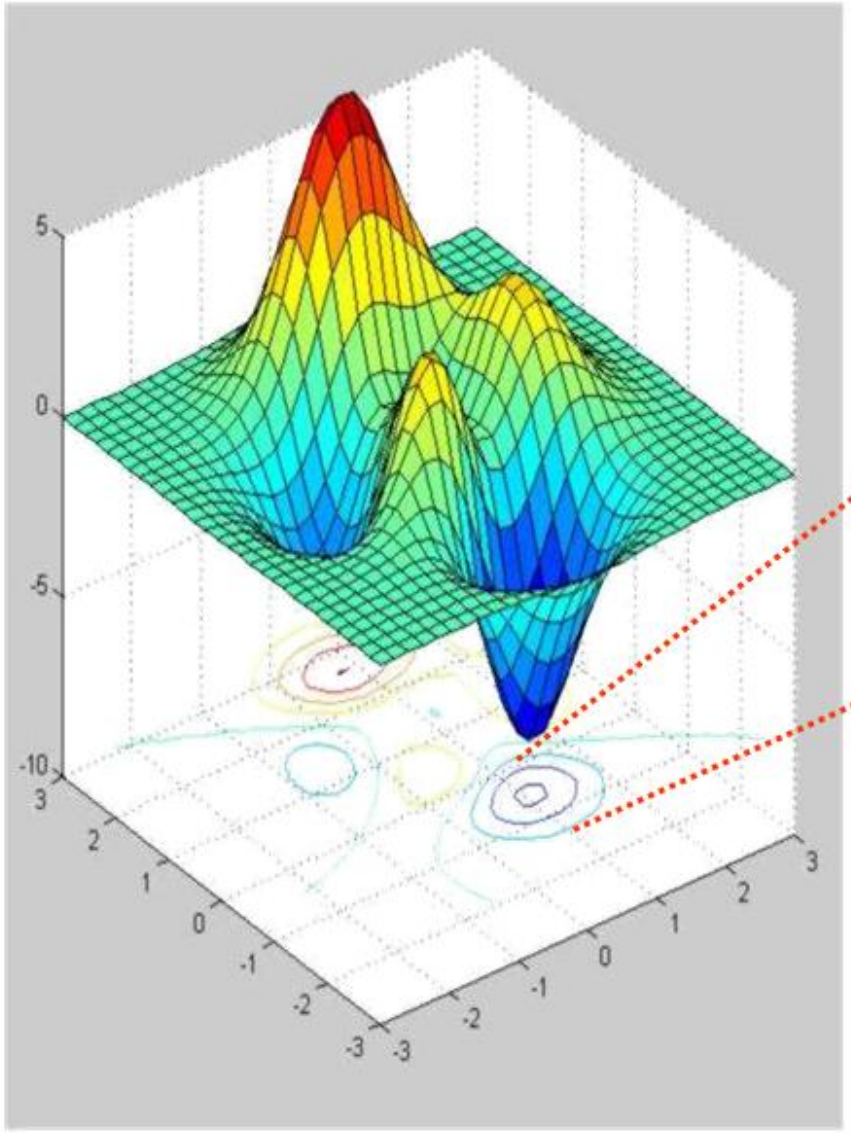
Get amplitude, a, by the equipartition

$\langle E_{potential} \rangle = 1/2 C \langle x^2 \rangle = 1/2\, k_b$

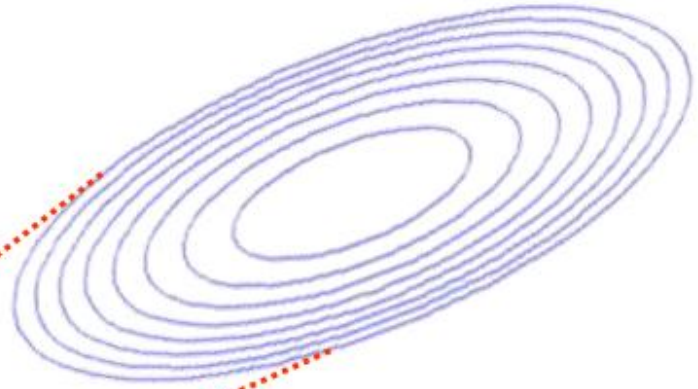
$\langle x^2 \rangle$ is the value of x^2.

Thus, $a = \text{sqrt}(2k_bT/C)$ as $\langle x^2 \rangle = 1/2\, a^2$ for a cosine wave.

NORMAL MODES IN HIGH DIMENSIONS



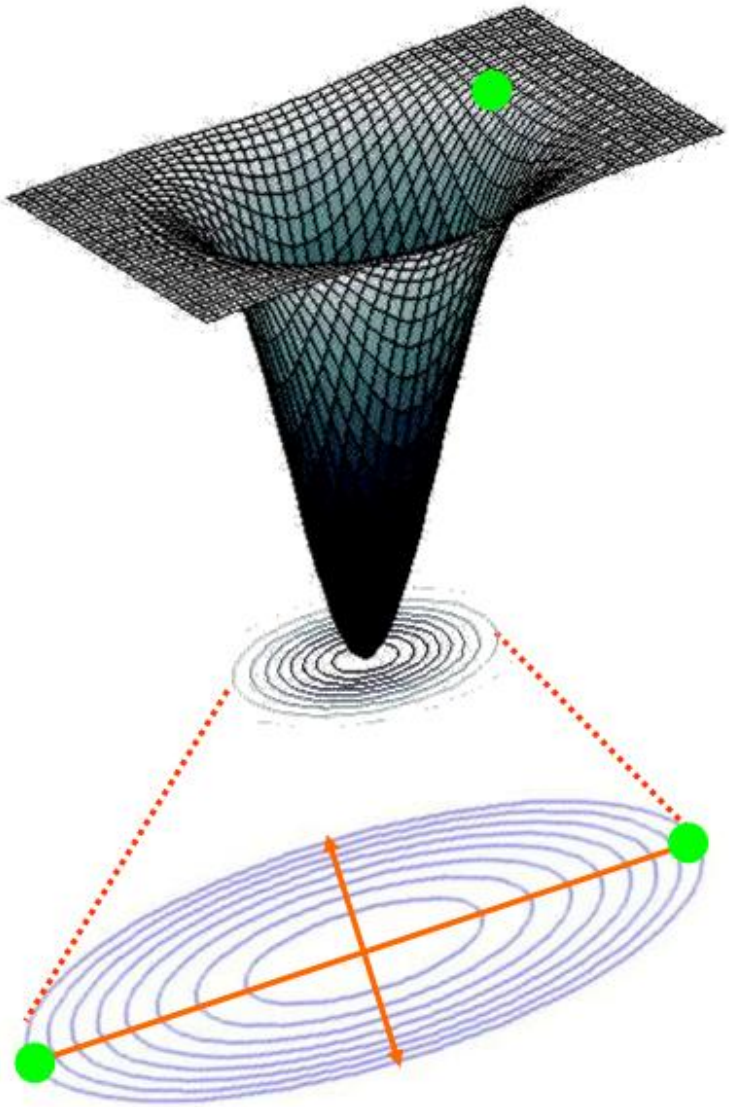
Focus on deepest energy



- Expand energy function about minimum.
- Approximate as a quadratic function

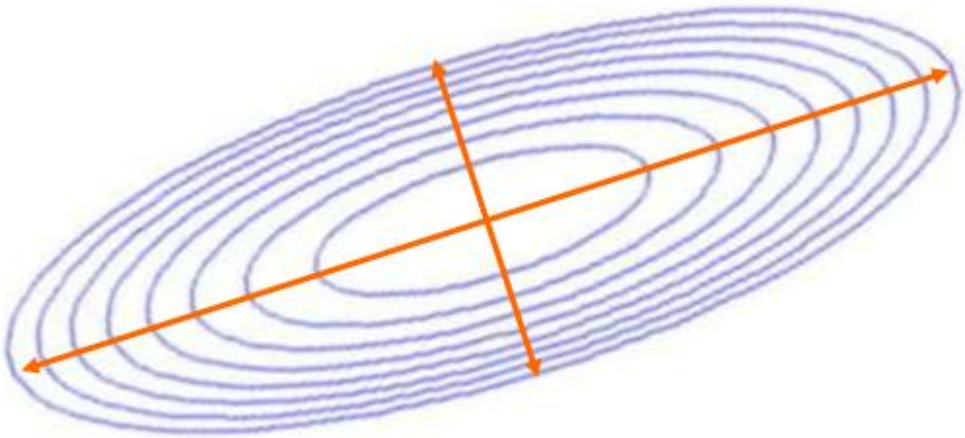
$$F(x,y) = Ax^2 + Bxy + Cy .$$

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



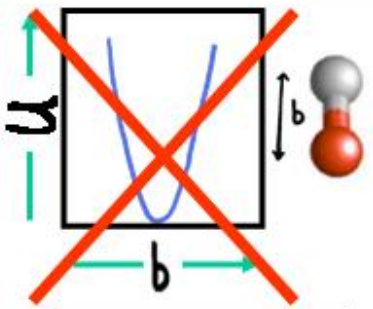
- The normal mode directions are the major and minor axes of the ellipse.
- All other motion is a linear combination of these basic motions.
- Solving for the modes requires a matrix that is N x N, where N in number of degrees of freedom.

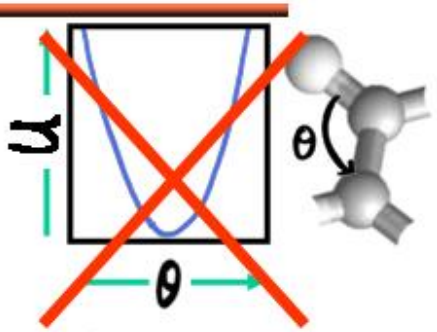
For n atoms, $N = 3n$ so it can be very big!

MOLECULAR POTENTIAL ENERGY

$$U = \sum \tfrac{1}{2} K_b (b - b_o)^2 + \sum \tfrac{1}{2} K_\theta (\theta - \theta_o)^2$$

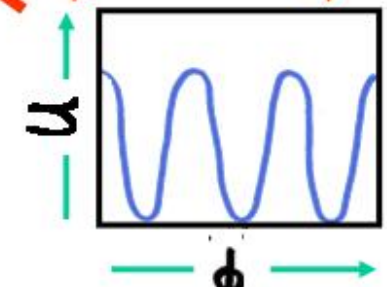
All All

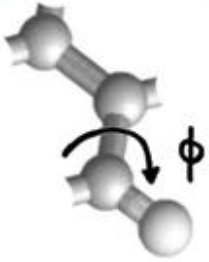




$$+ \sum K_\phi [1 - \cos(n\phi + \delta)]$$

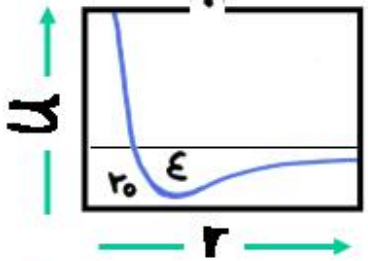
All Torsion

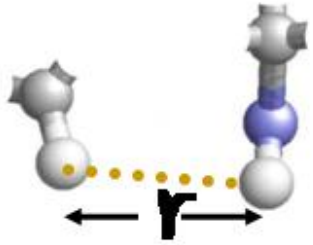




$$+ \sum \varepsilon \left[\left(r_o / r \right)^{12} - 2 \left(r_o / r \right)^{6} \right]$$

All nonbonded

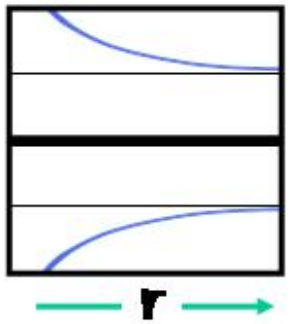


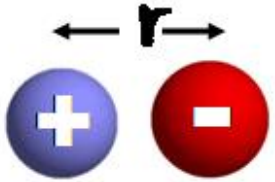


$$+ \sum 332 q_i q_j / r$$

All partial

Eliminate the strongest springs.





POTENTIAL ENERGY IN TORSION SPACE

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

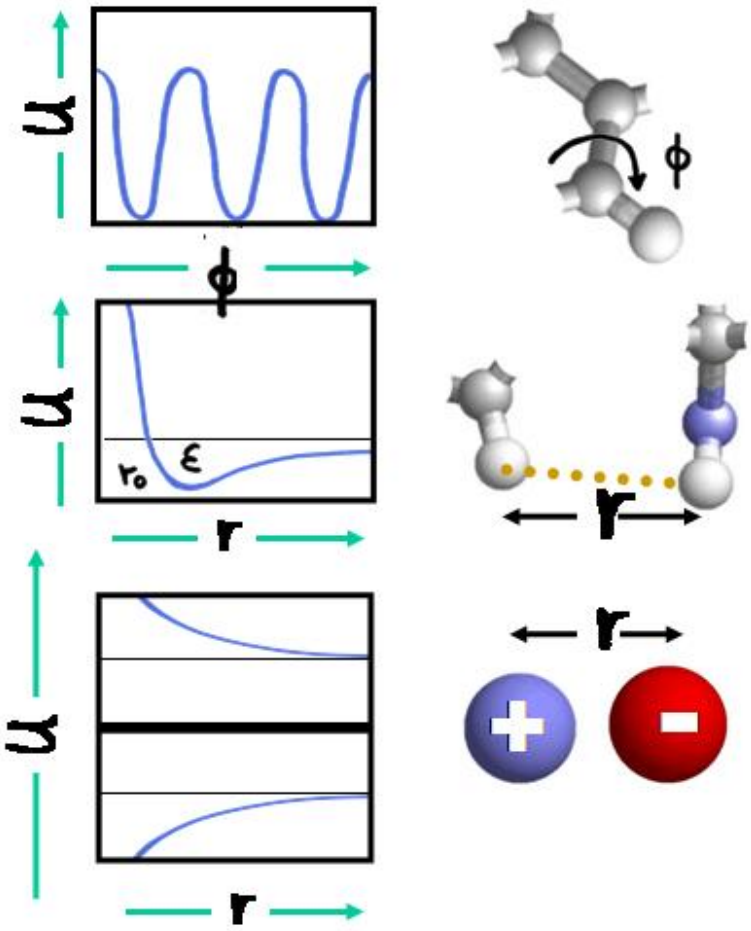
All Torsion

$$+ \sum \varepsilon \left[\left(r_0/r\right)^{12} - 2\left(r_0/r\right)^{6} \right]$$

All nonbonded

$$+ \sum 332 q_i q_j / r$$

All partial

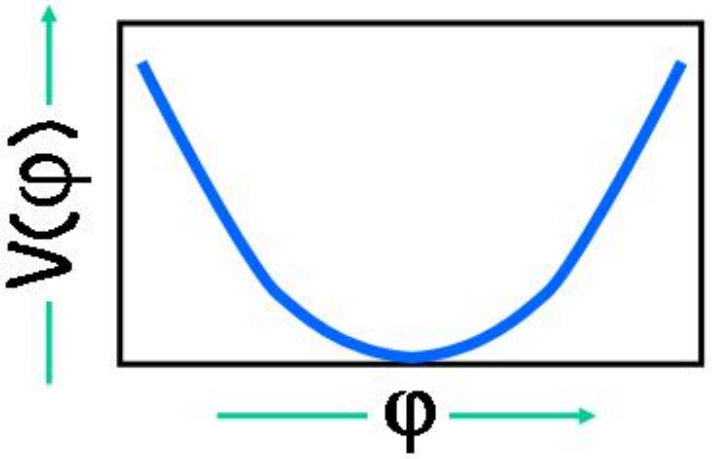


- A protein with N residues has about 4N (φ, ψ, χ) single bond torsion angles.
- The same protein has about 50N Cartesian coordinates (x, y, z).

THEORY OF NORMAL MODES I

- Assume Potential energy, V, is quadratic function of φ.

$$V = \frac{1}{2}\sum_{i,j} V_{ij}\,(\phi_i - \phi_i^\circ)(\phi_j - \phi_j^\circ)$$



- This means that $V_{ij} = d^2V/d\varphi_i d\varphi_j$

- Assume Kinetic energy, T, is quadratic function of $d\varphi/dt$.

$$T = \frac{1}{2}\sum_{i,j} T_{ij}\,(d\phi_i/dt)(d\phi_j/dt)$$

Note the symmetry between Potential and Kinetic energy.

- This means that $T_{ij} = d^2T/d(d\varphi_i/dt)\,d(d\varphi_j/dt)$

THEORY OF NORMAL MODES II

- Solve for $\varphi(t)$ using Lagrangian approach.

$$\sum T_{ij} (d^2\varphi_j/dt^2) = \sum V_{ij} \Delta\varphi_j$$

Analogous to $ma = F$

- Try a periodic function for $\varphi(t)$:

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

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

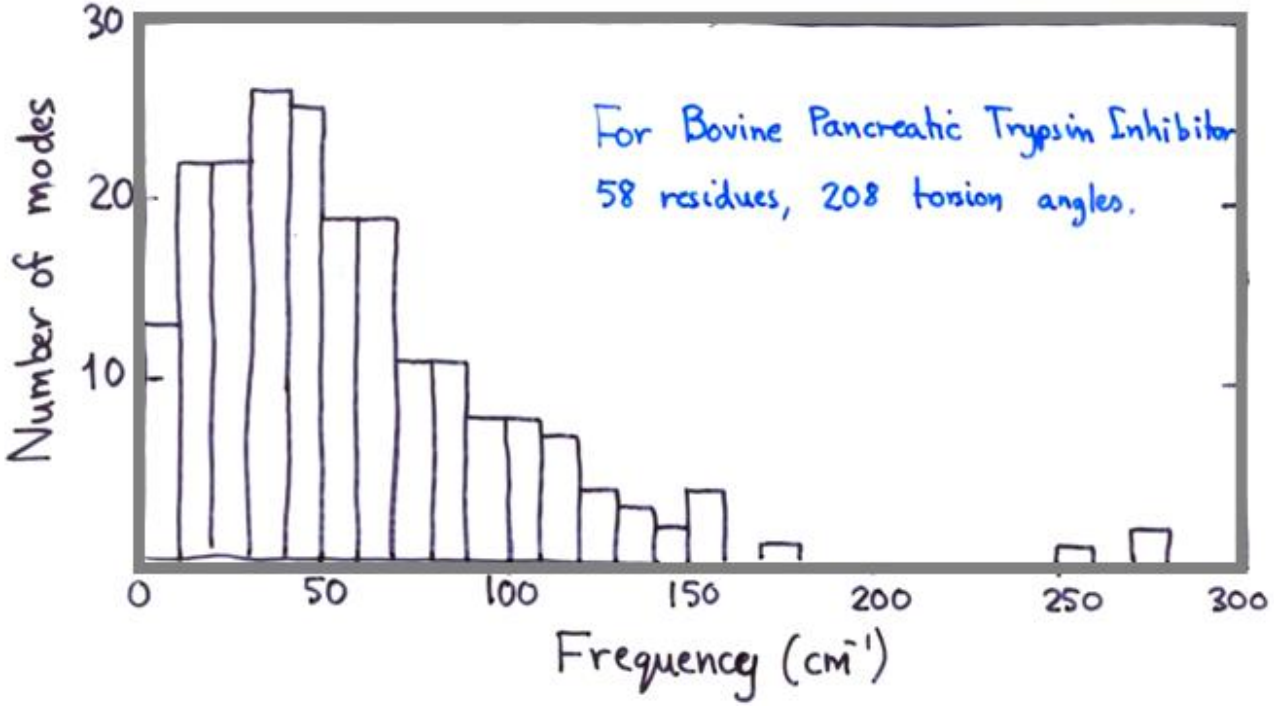
- In Matrix notation the Lagrangian equation is:

$$TA\omega^2 = VA$$

This is Eigenvalue equation that is easily solved.

Protein Normal Modes
Concept 5.2

RATES OF VIBRATION

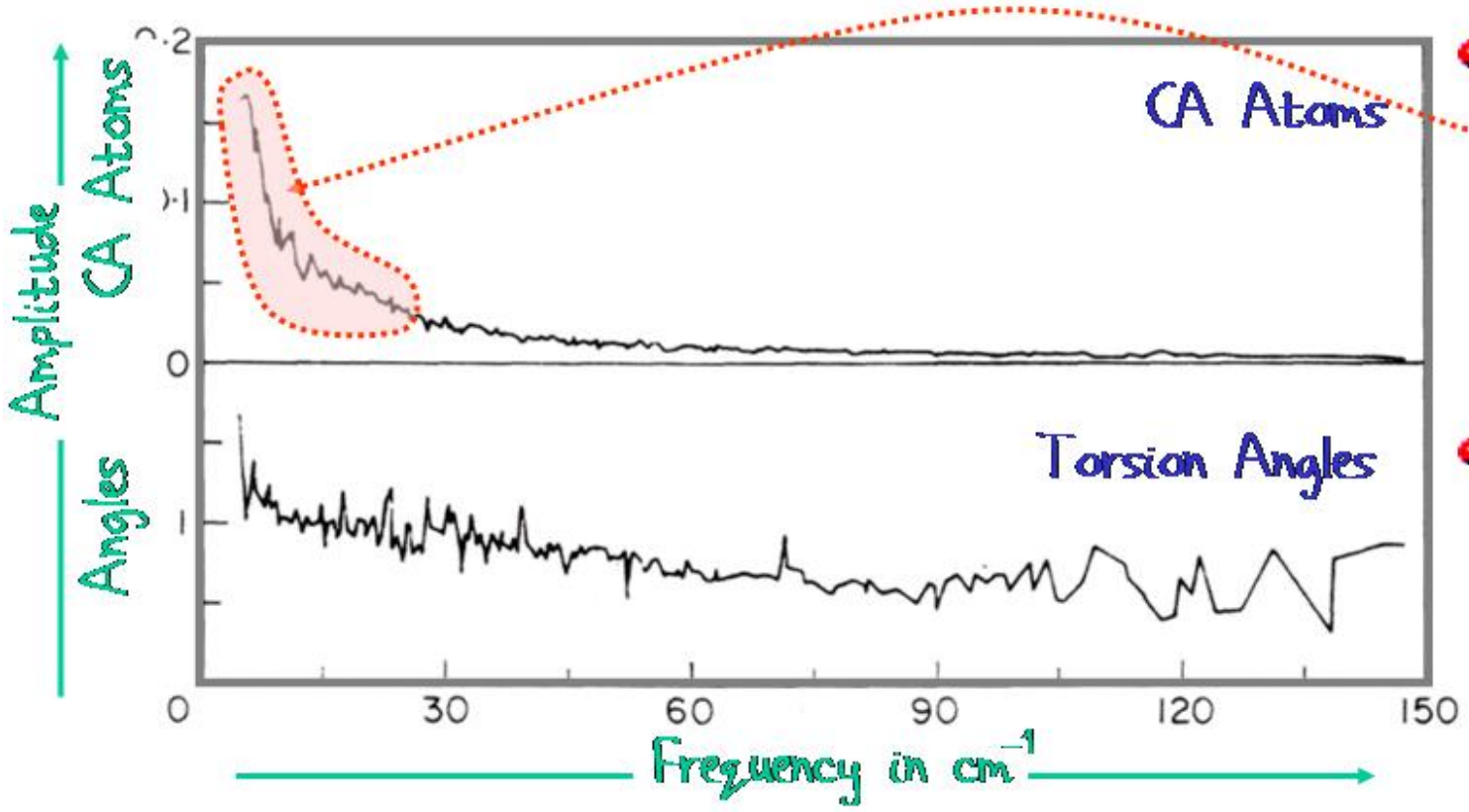


- There is a broad range of torsion angle mode frequencies.

- 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}.

For more details see: Levitt et al. J Mol. Biol. (1985).

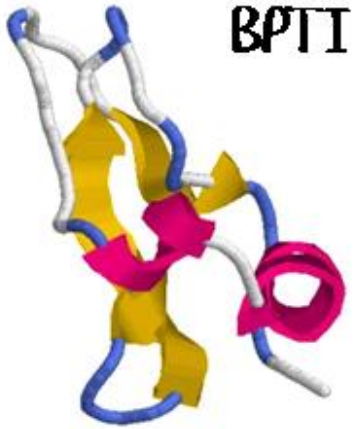
AMPLITUDES OF VIBRATION

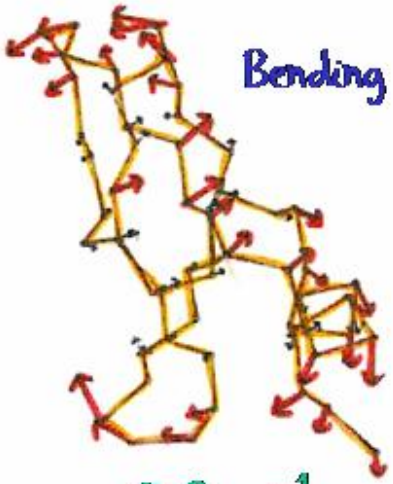


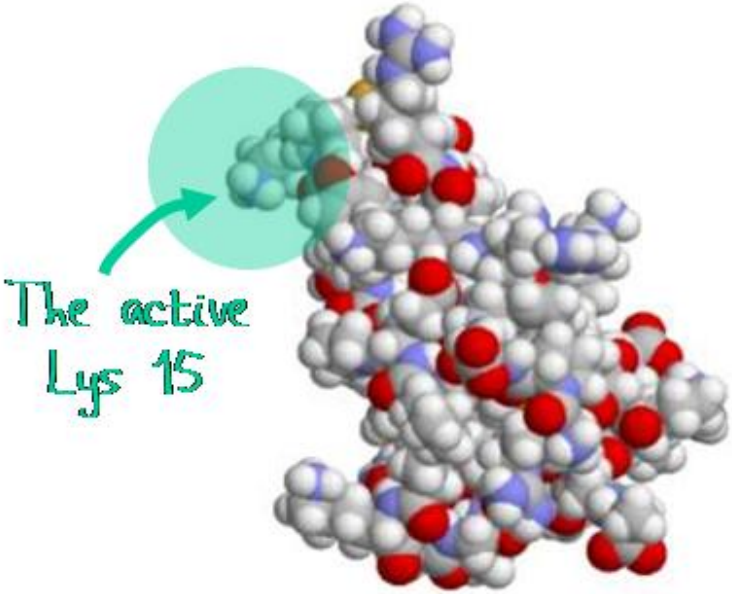
- Almost all the motion of the CA atoms comes from the lowest frequency modes.
- There is high-frequency motion of the torsion angles.

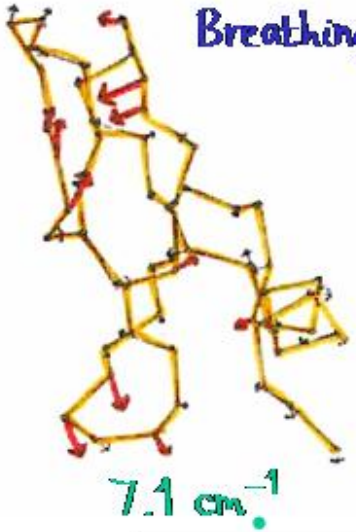
- The CA Amplitude is the RMS movement of all CA atoms as a result of activating the particular mode.
- The Torsion Angle Amplitude is the RMS movement of all torsion angles as a result of activating the particular mode.

TRYPSIN INHIBITOR MODES



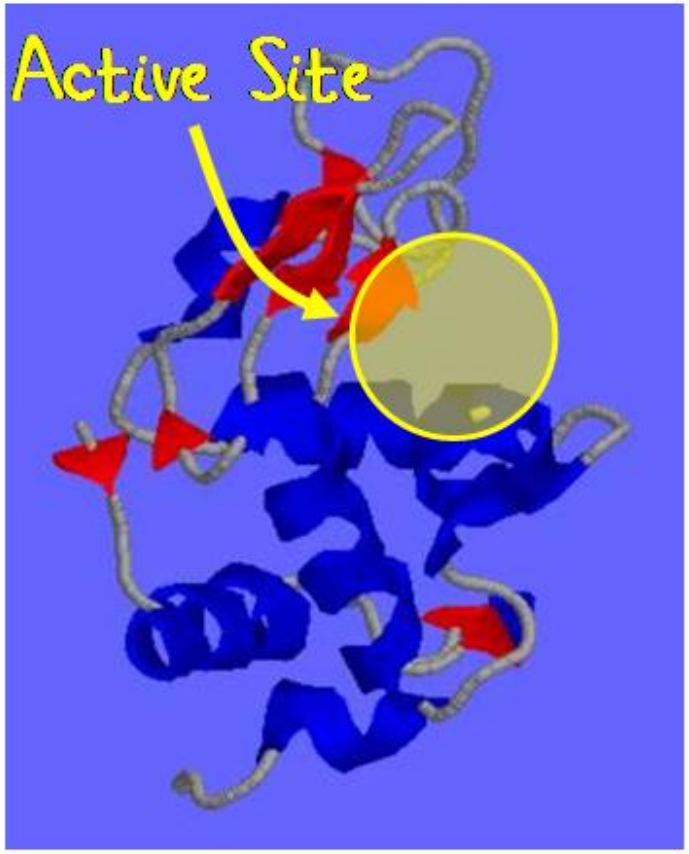


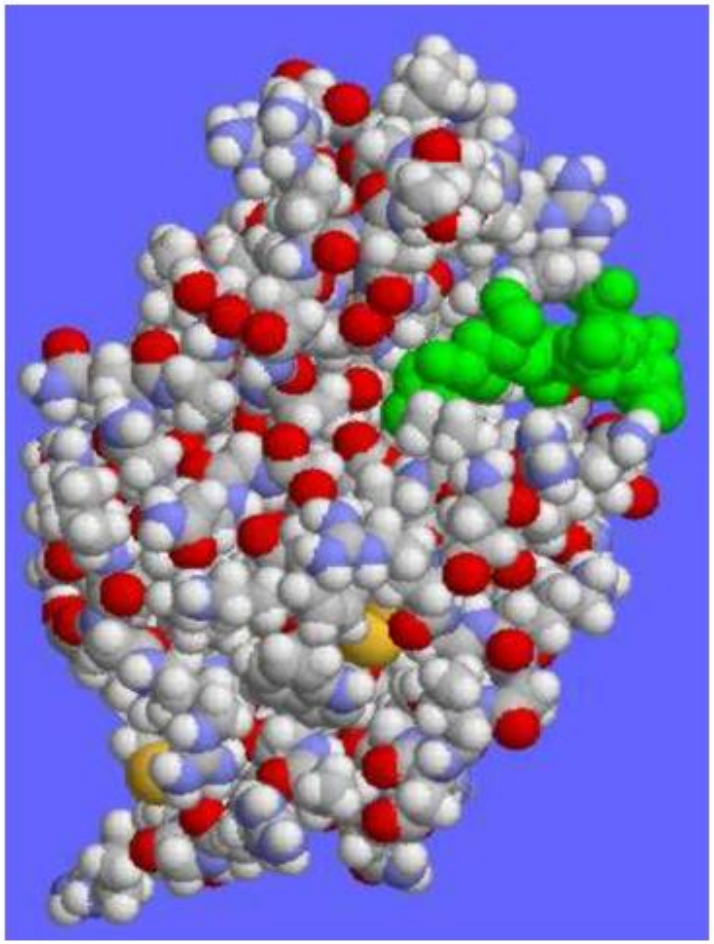




BPTI NORMAL MODES AT HIGH TEMPERATURE

LYSOZYME MODES

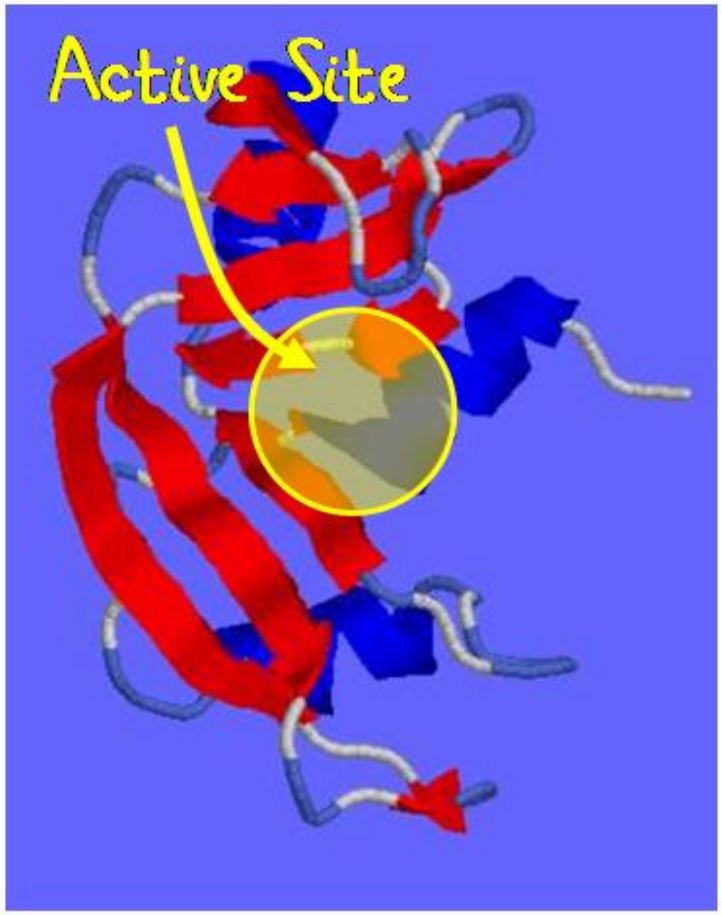


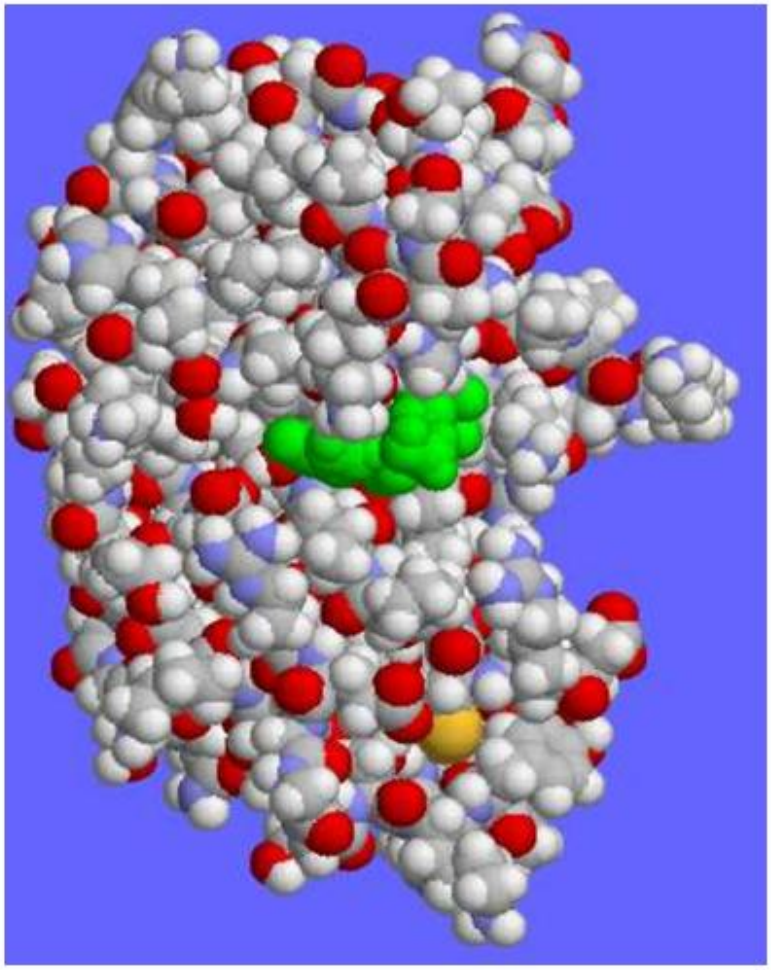


An inhibitor, which is colored in green, is bound in the active site.
The inhibitor is not included in the normal mode calculations.

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.

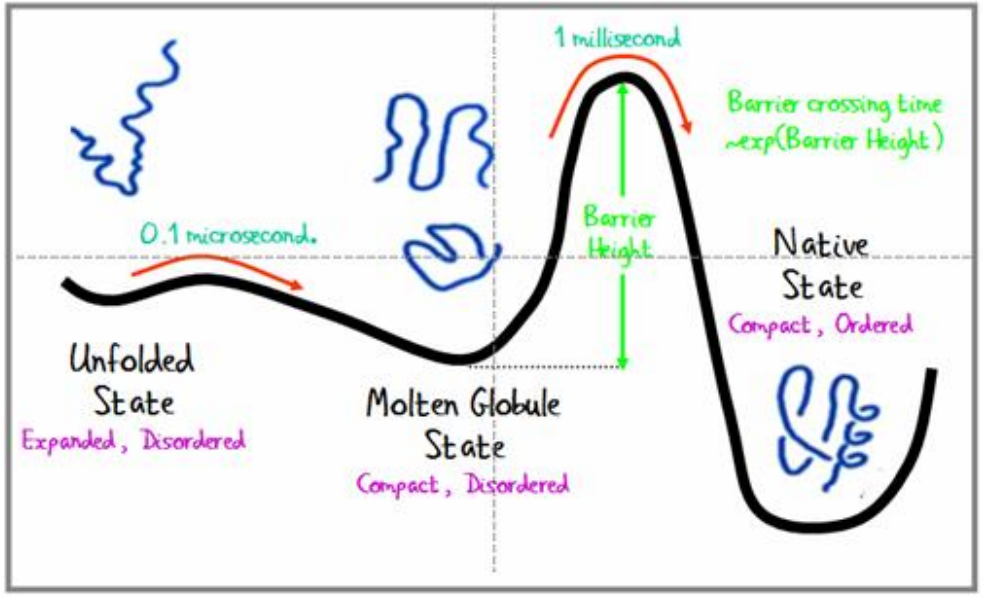
Effect of Temperature.

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

$$\text{Velocity} \propto \sqrt{\text{Temperature}}$$

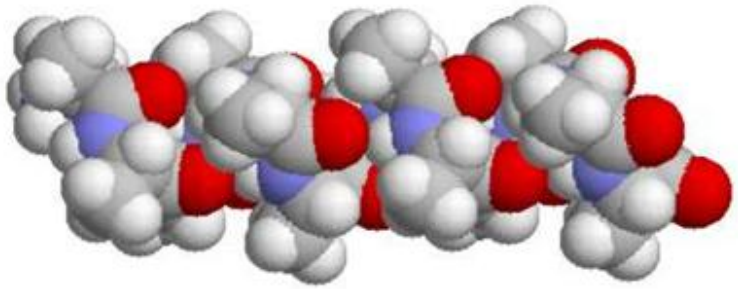
Rate of jumping

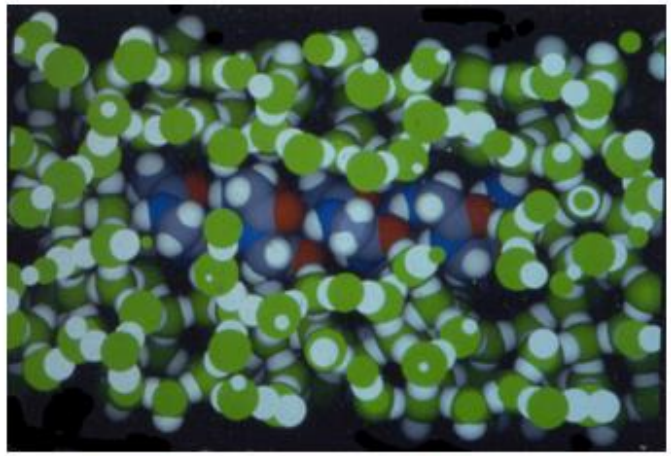
Barrier Height is ΔG

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

UNFOLD ALPHA-HELIX

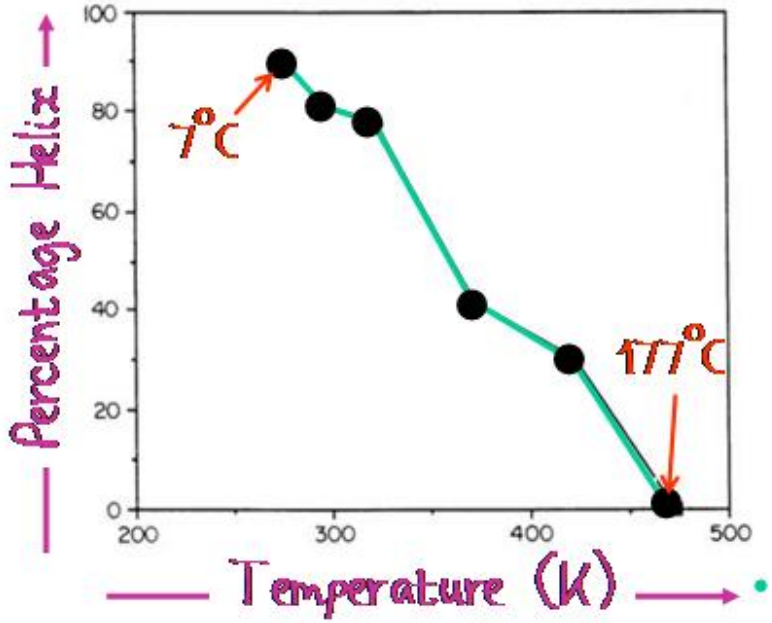
13 Alanine residues



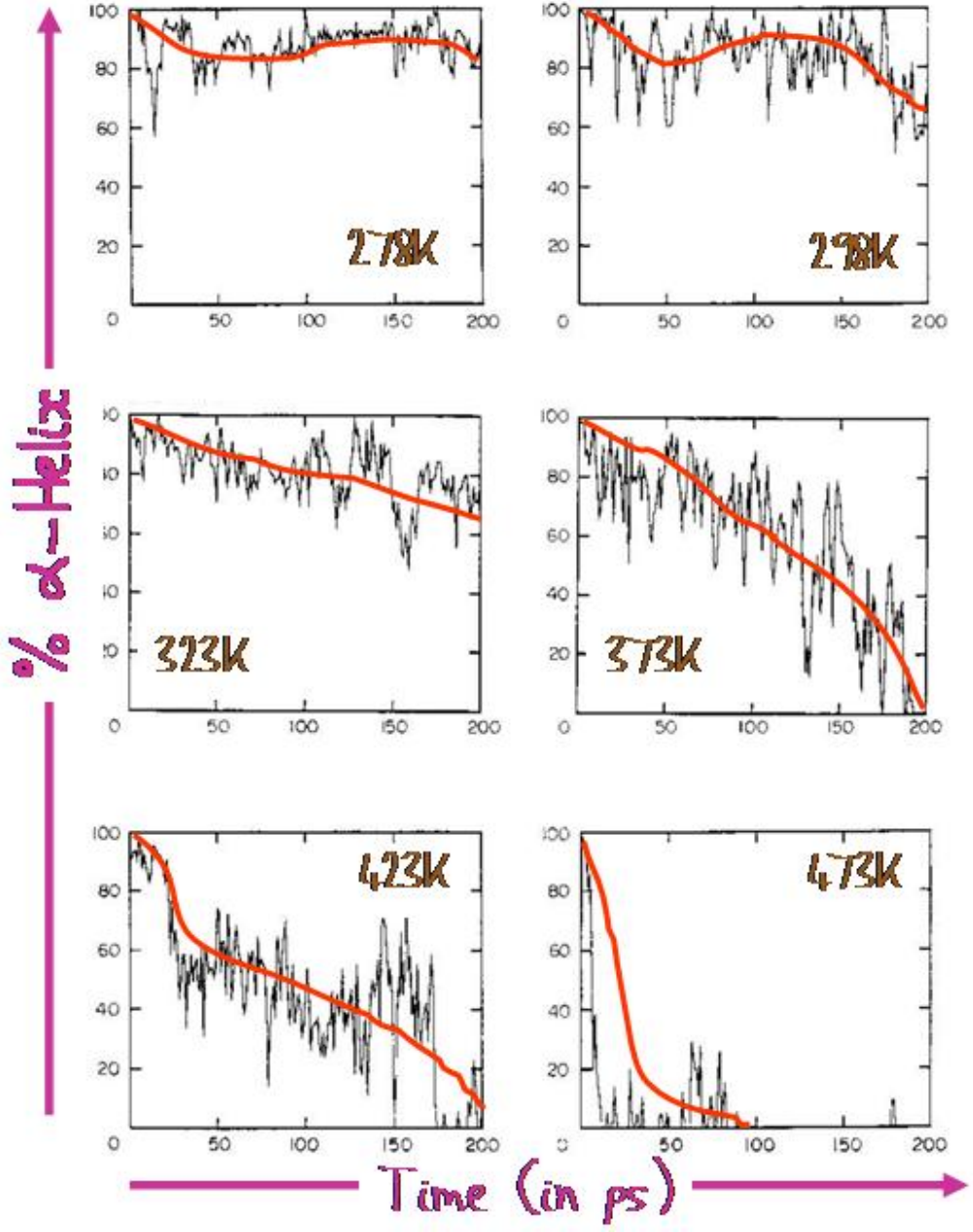


Put it in a box of water.

- 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.



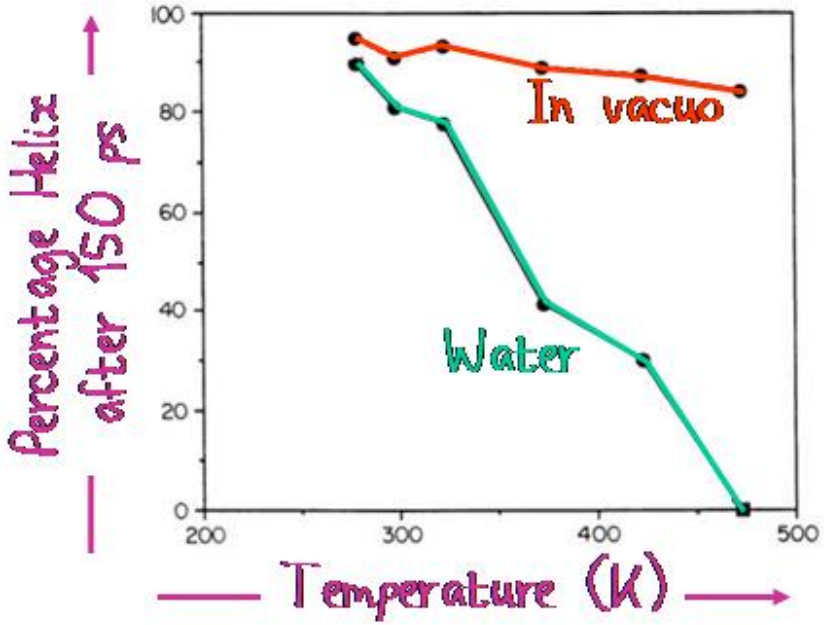
EFFECT OF TEMPERATURE



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

For more details see: Daggett & Levitt, J. Mol. Biol. 223, 1121 (1992).

HELIX LESS STABLE IN WATER

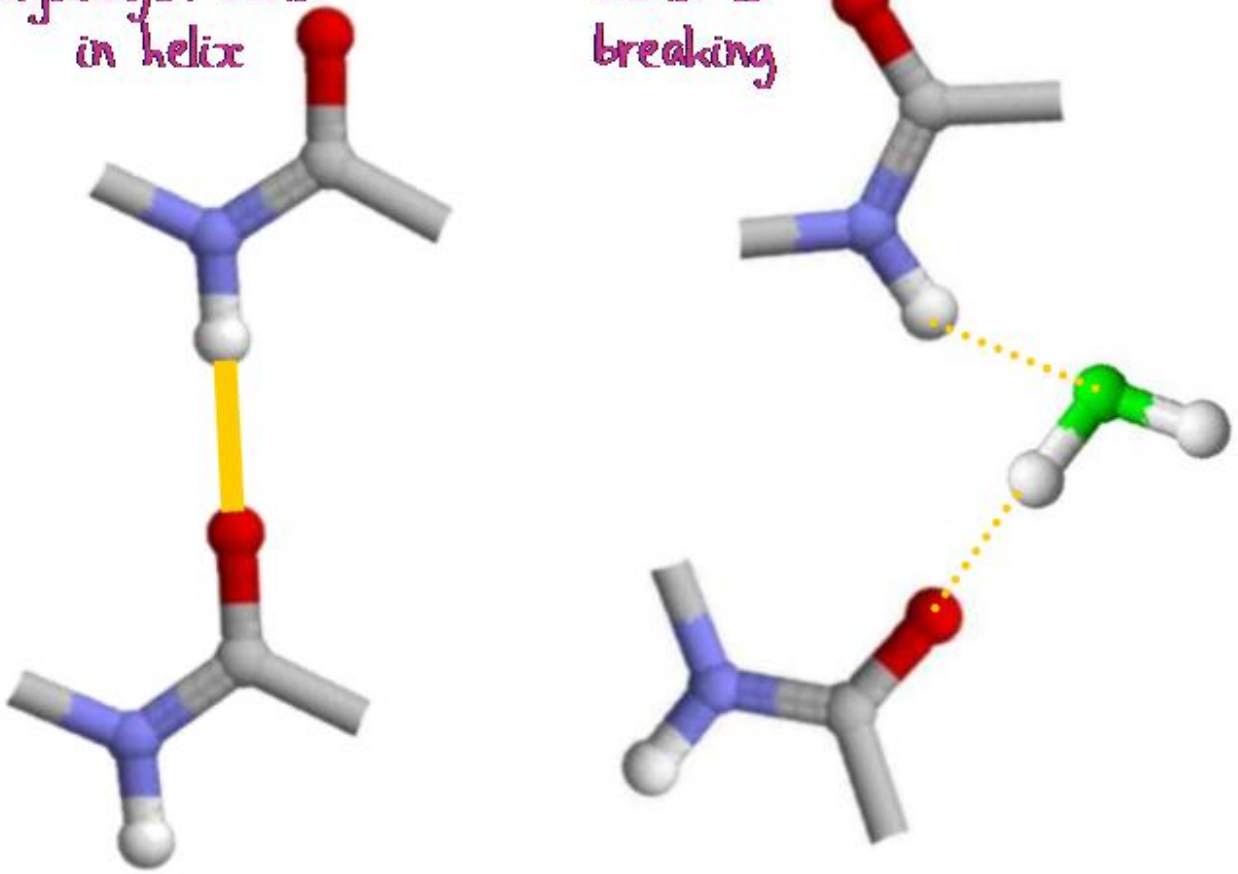


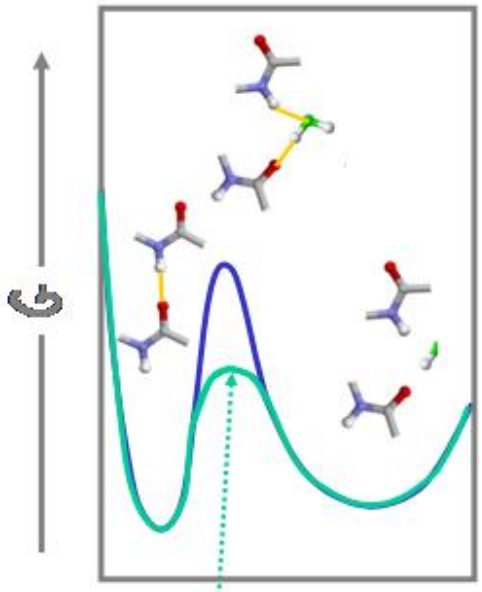
- In vacuo the helix is very stable even at high temperature.
- In water the helix is unstable at high temperature.
- The rate of melting depends on temperature.
- This happens because water molecules stabilize the transition state.

WATER ALLOWS HYDROGEN BONDS TO BREAK

Intact hydrogen bond in helix

Hydrogen bond is breaking





Free Energy barrier between states is much lower in water

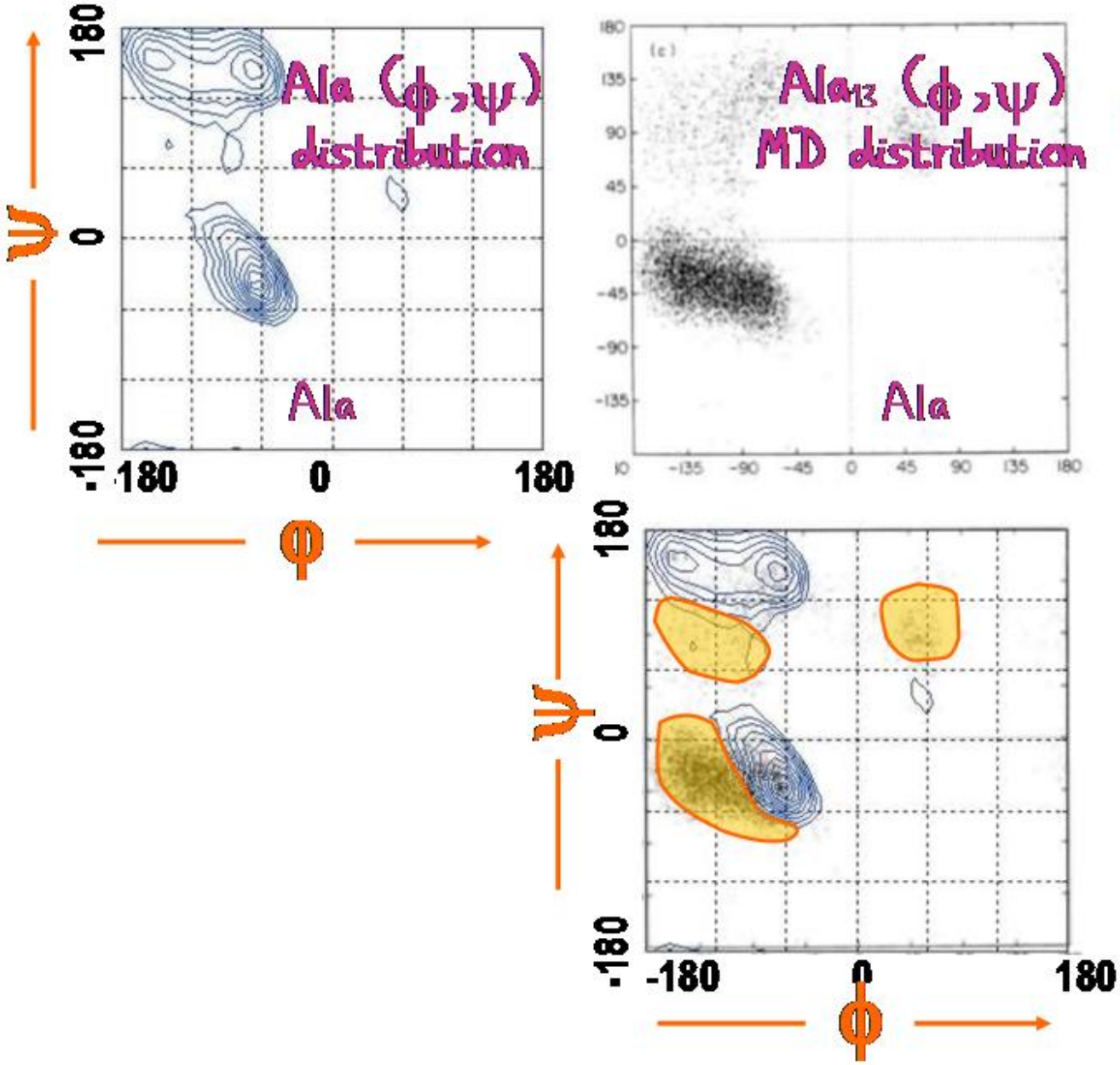
Water catalyzes the breakage of hydrogen bonds by stabilizing the transition state.

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 Ala_{12} in water at high temperature.
- There are differences shown in orange..

Unfolding Proteins
Concept 5.4

PROTEIN UNFOLDING

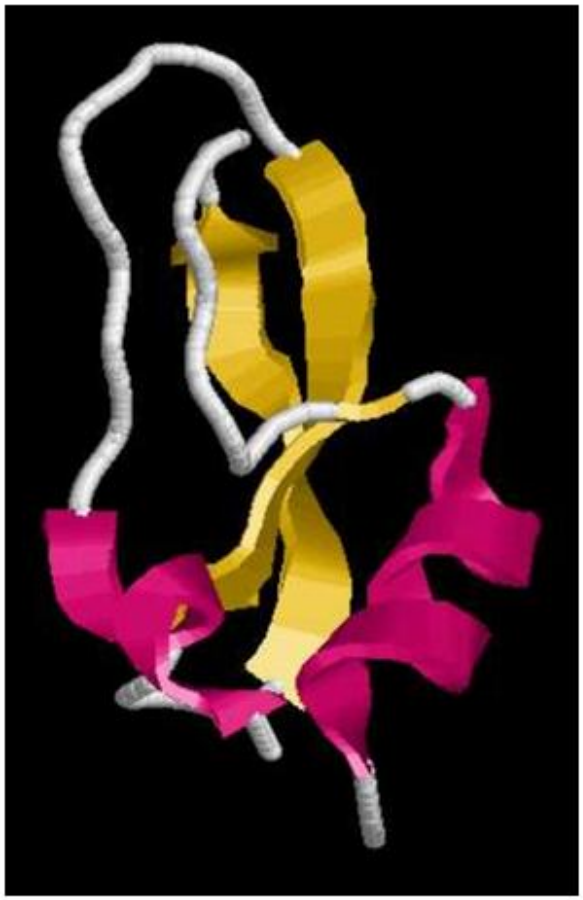
What Happens to Secondary

What Happens to Aromatic Core.

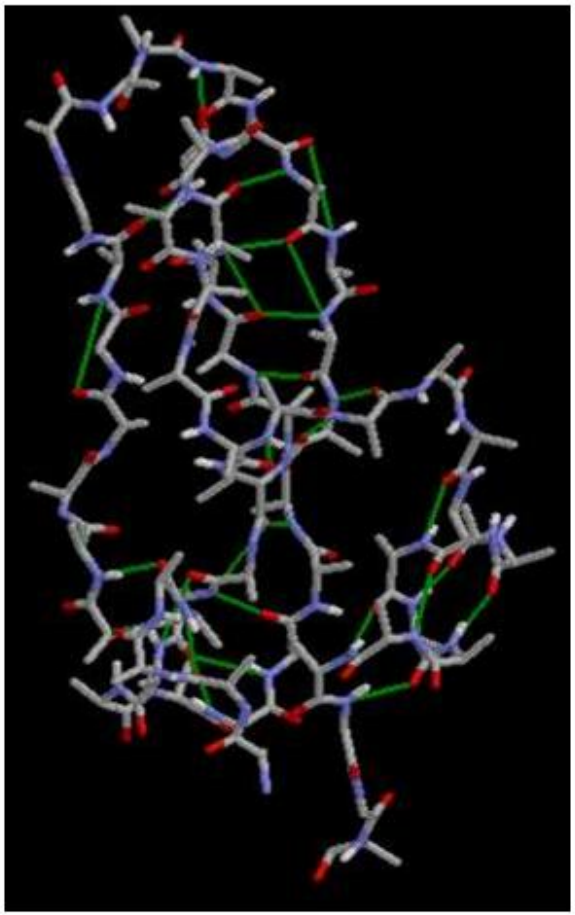
The Molten Globule?

Connection to Experiment.

WHAT HAPPENS TO SECONDARY STRUCTURE



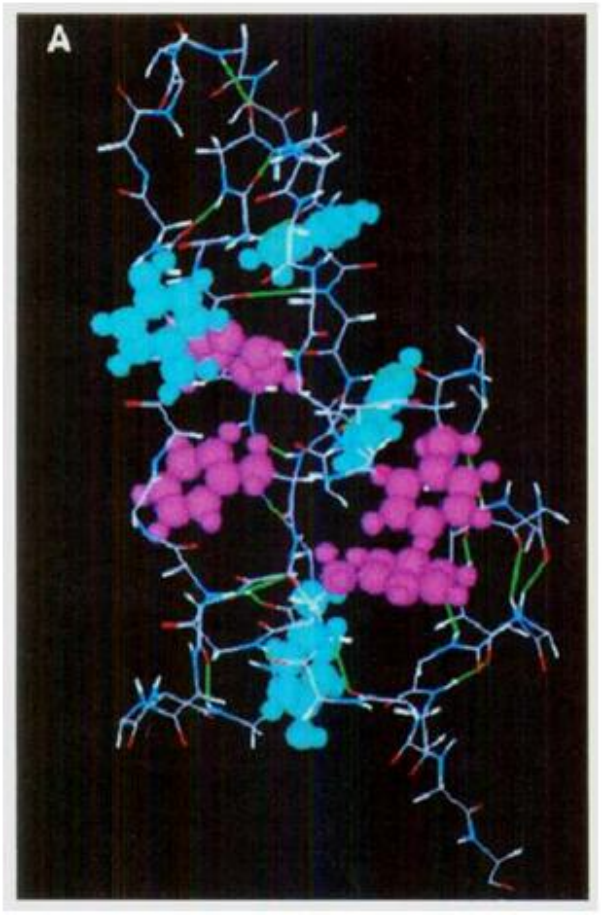
Secondary



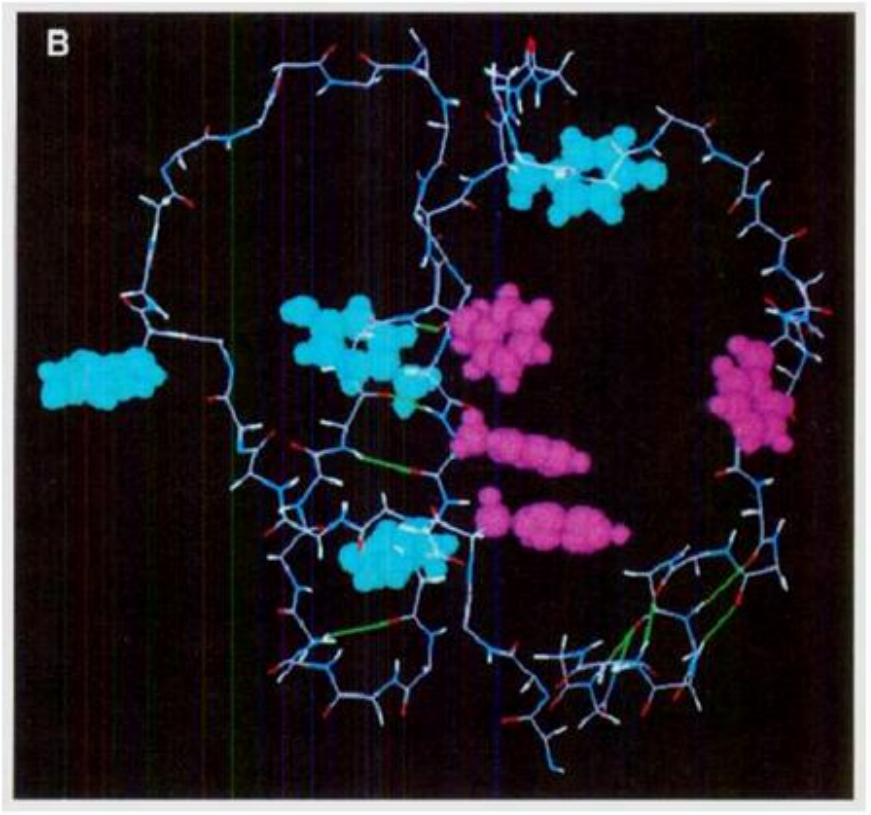
Hydrogen

PROTEIN UNFOLDING, IN WATER AT HIGH TEMPERATURE

WHAT HAPPENS TO AROMATIC SIDECHAINS



Native

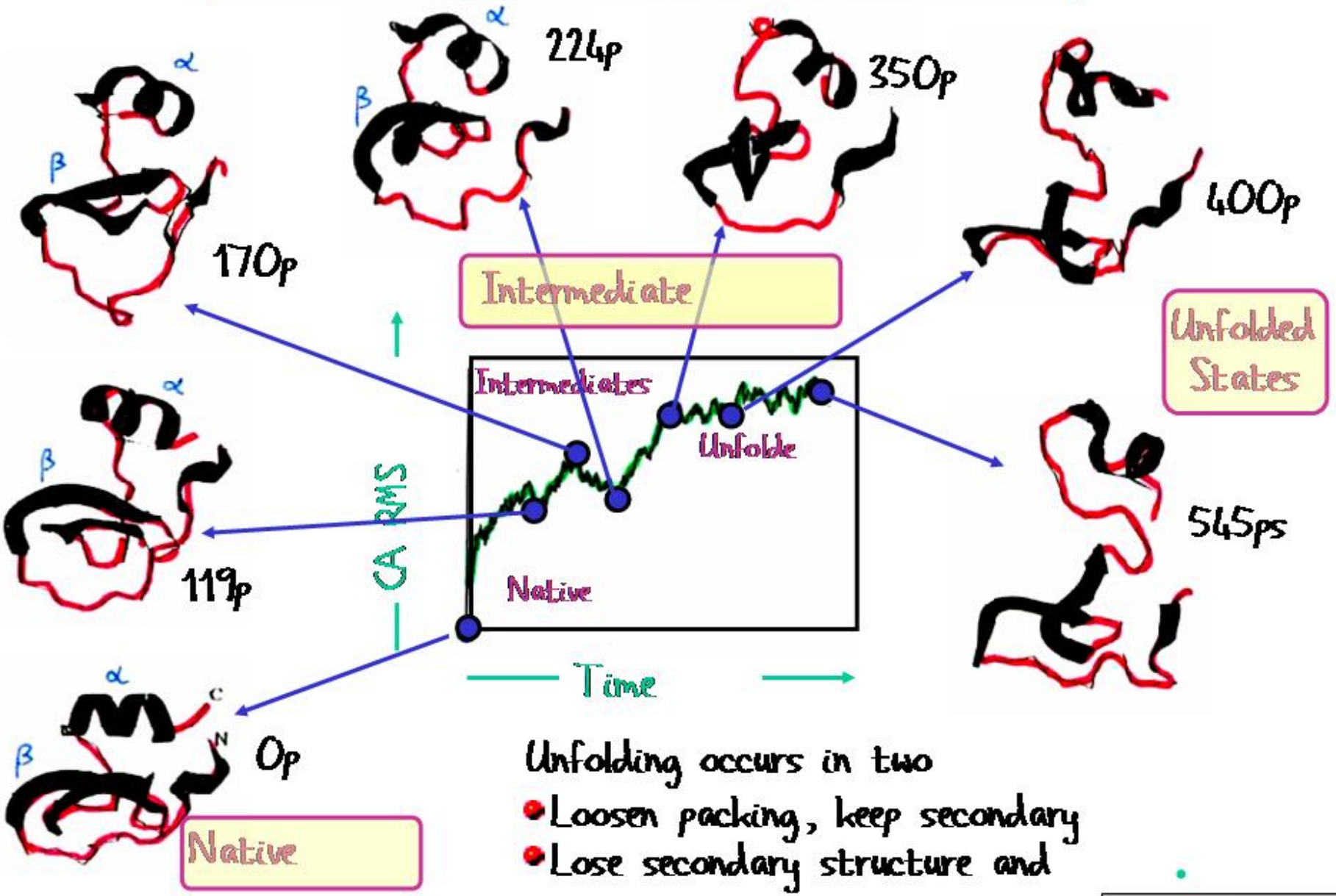


?

PROTEIN UNFOLDING

IN WATER AT HIGH TEMPERATURE

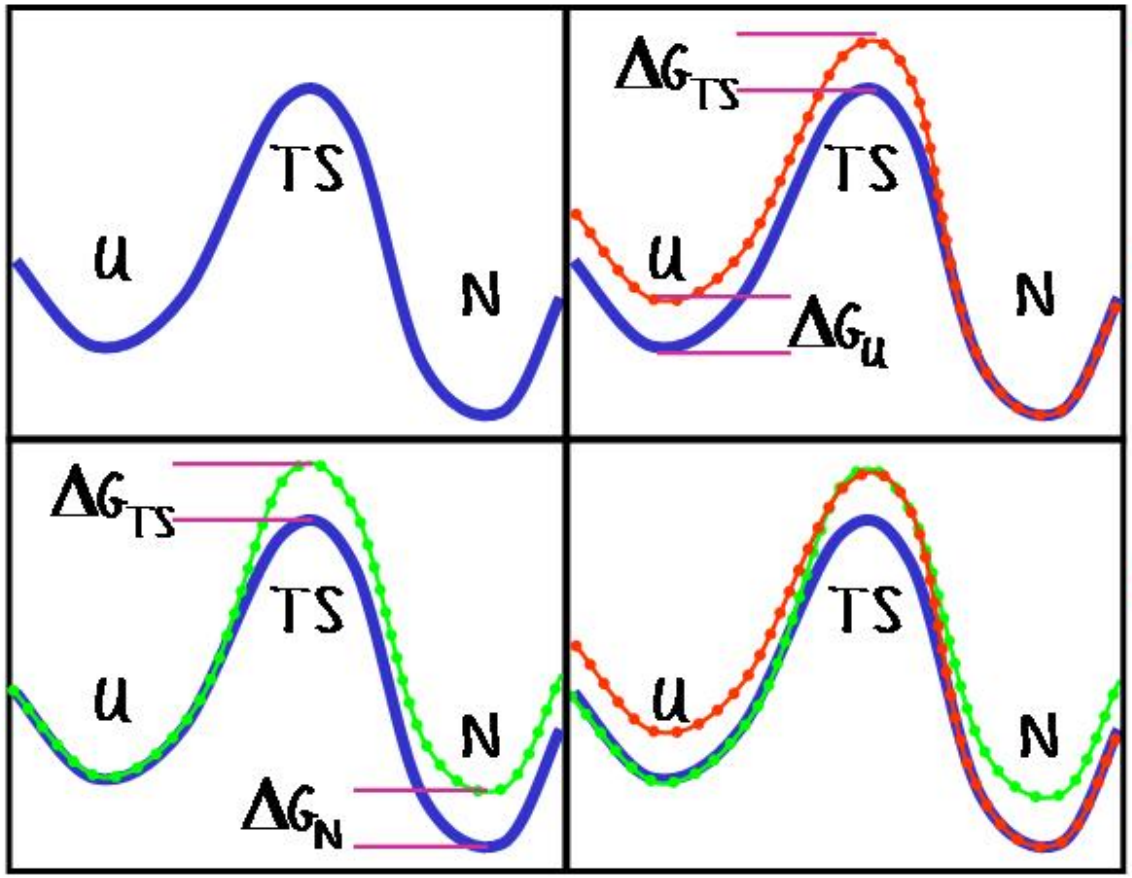
UNFOLDING PATHWAY



Unfolding occurs in two

- Loosen packing, keep secondary
- Lose secondary structure and

CONNECTION TO EXPERIMENT



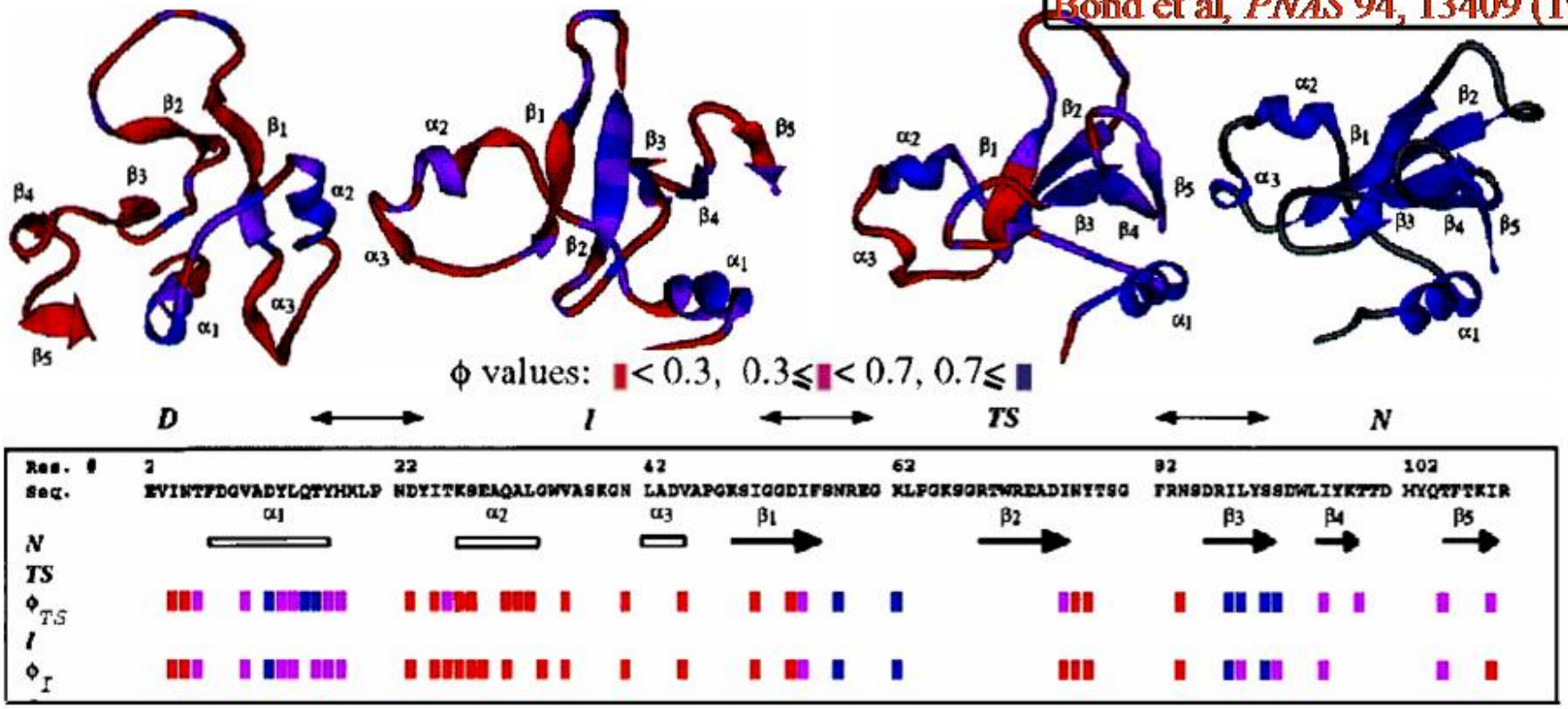
- $\varphi_{TS} = (\Delta G_{TS} - \Delta G_U)/(\Delta G_N - \Delta G_U)$
- $\Delta G_{TS} = \Delta G_U$ means that TS is like U. $\varphi_{TS} = 0$.
- $\Delta G_{TS} = \Delta G_N$ means that TS is like N. $\varphi_{TS} = 1$.

φ_{TS} measures the relative effect of changing a particular residue to Ala.

Fersht et al, *Nature* 342, 122 (1989)

CONNECTION TO EXPERIMENT

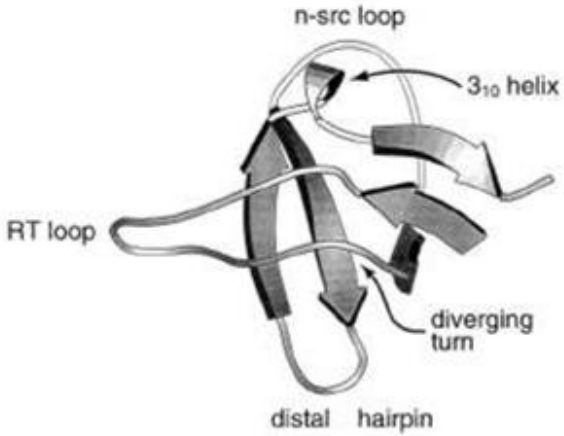
Bond et al, *PNAS* 94, 13409 (1997)



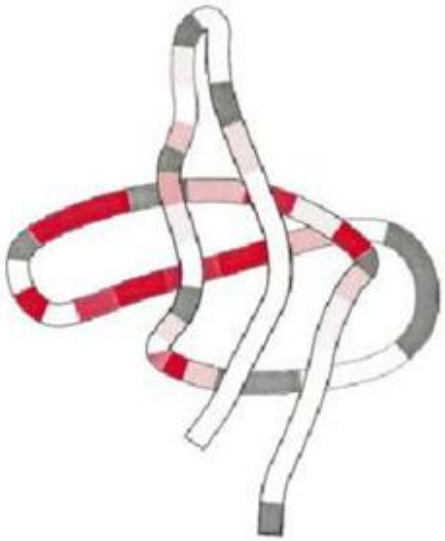
- φ_{TS} measures the relative effect on the folding rate of changing a particular residue to Ala. $\varphi_{TS}=0$ implies a denatured-like transition state structure and $\varphi_{TS}=1$ implies a native-like transition state structure at that position.
- Residues found experimentally to have high φ_{TS} values are found to interact in the simulated transition state.

CONNECTION TO EXPERIMENT

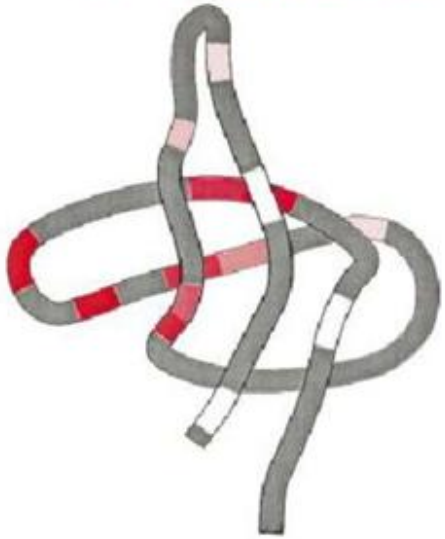
SH3



Crucial residues from simulation



Crucial residues from experiment



- Residues marked in red remain native-like in SH3 unfolding simulations. Experiment implicates residues from the same region of the structure.

Tsai et al. *J.Mol.Biol.*291, 215 (1999)

Experimentalists and theoreticians are interacting to study protein unfolding.

Folding Simplified Chains Concept 5.5

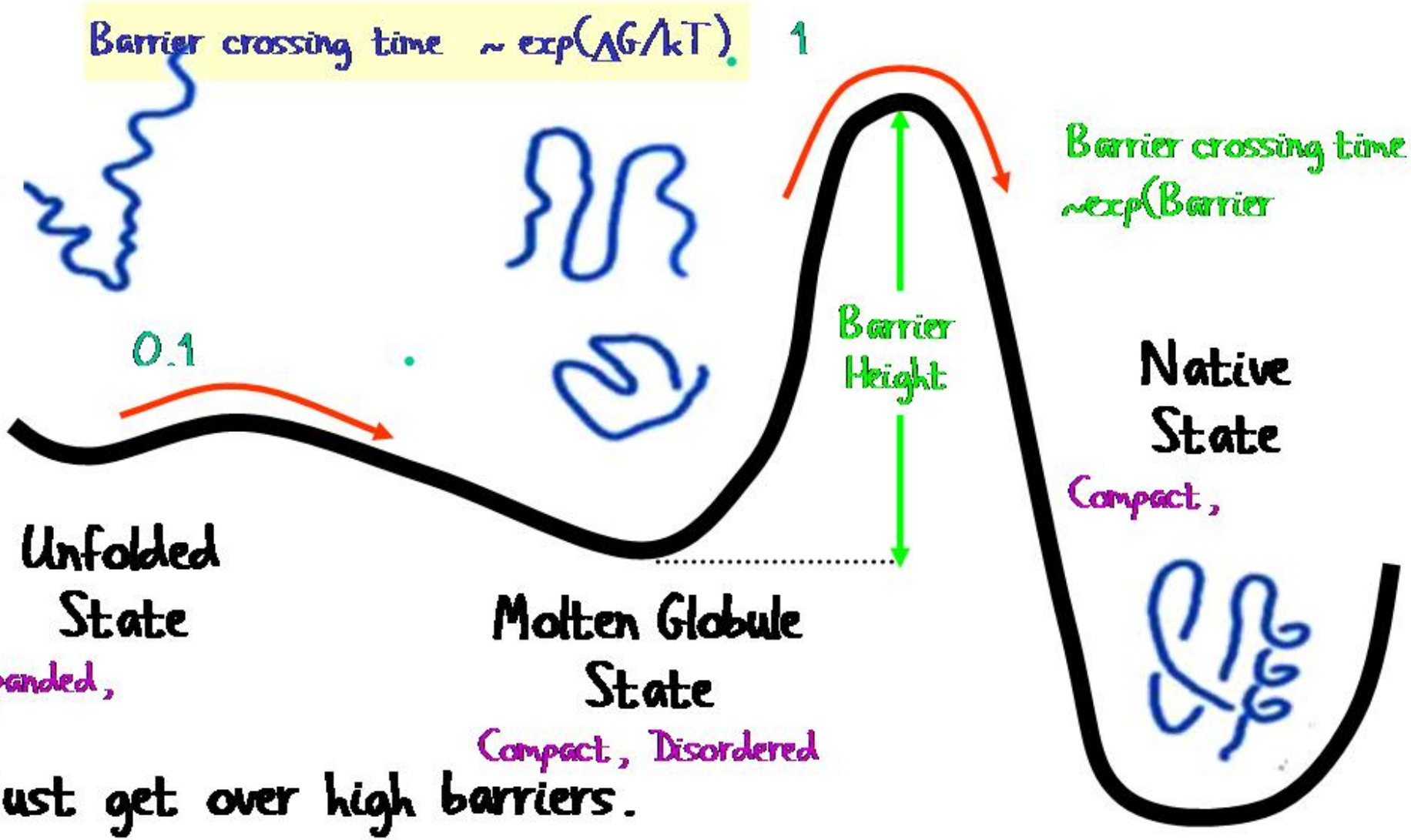
FOLD SIMPLIFIED PROTEINS

Folding With Minimization.

Why Folding is so

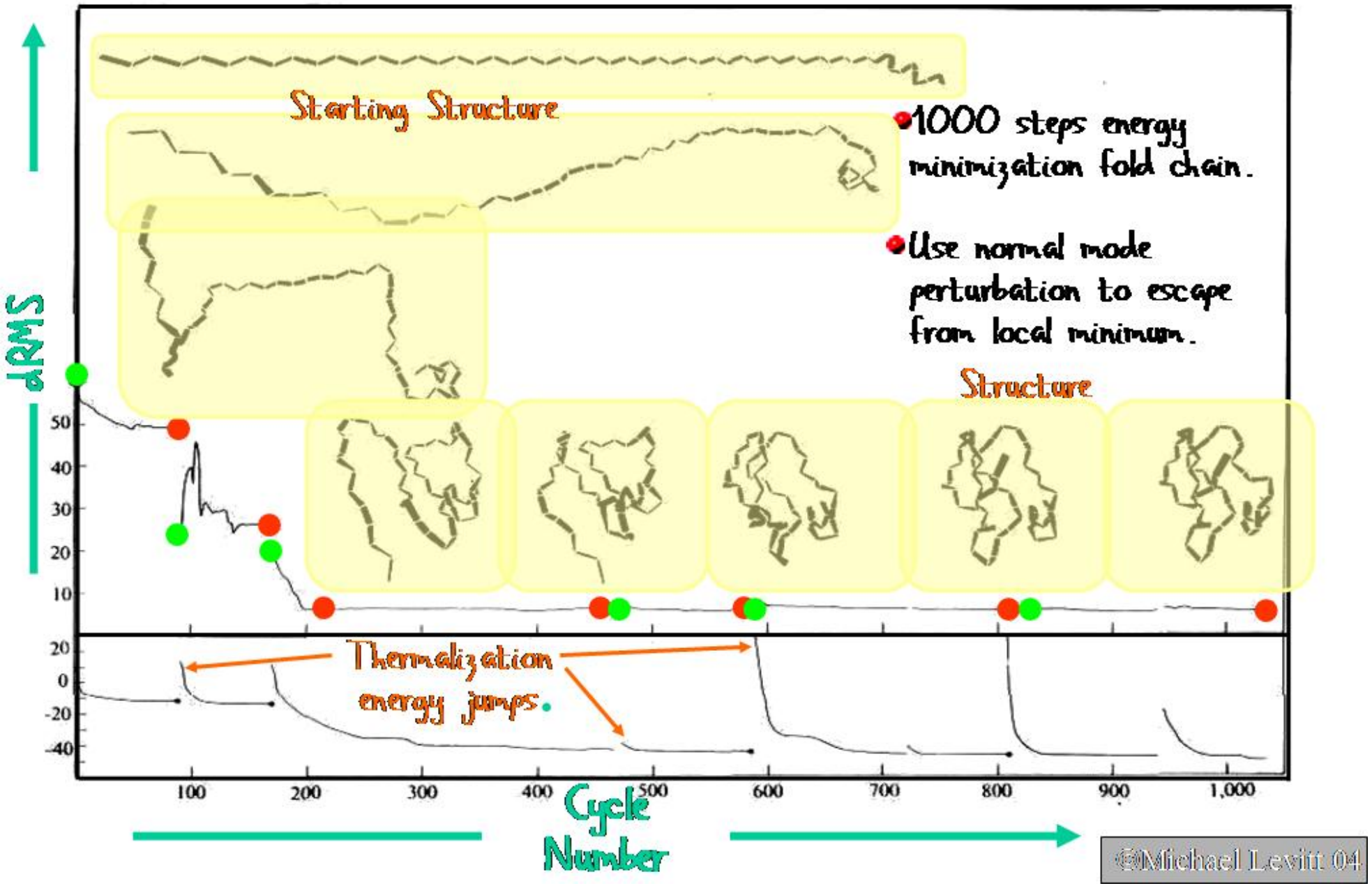
Lattice Model Folding.

WHY IS FOLDING SO DIFFICULT

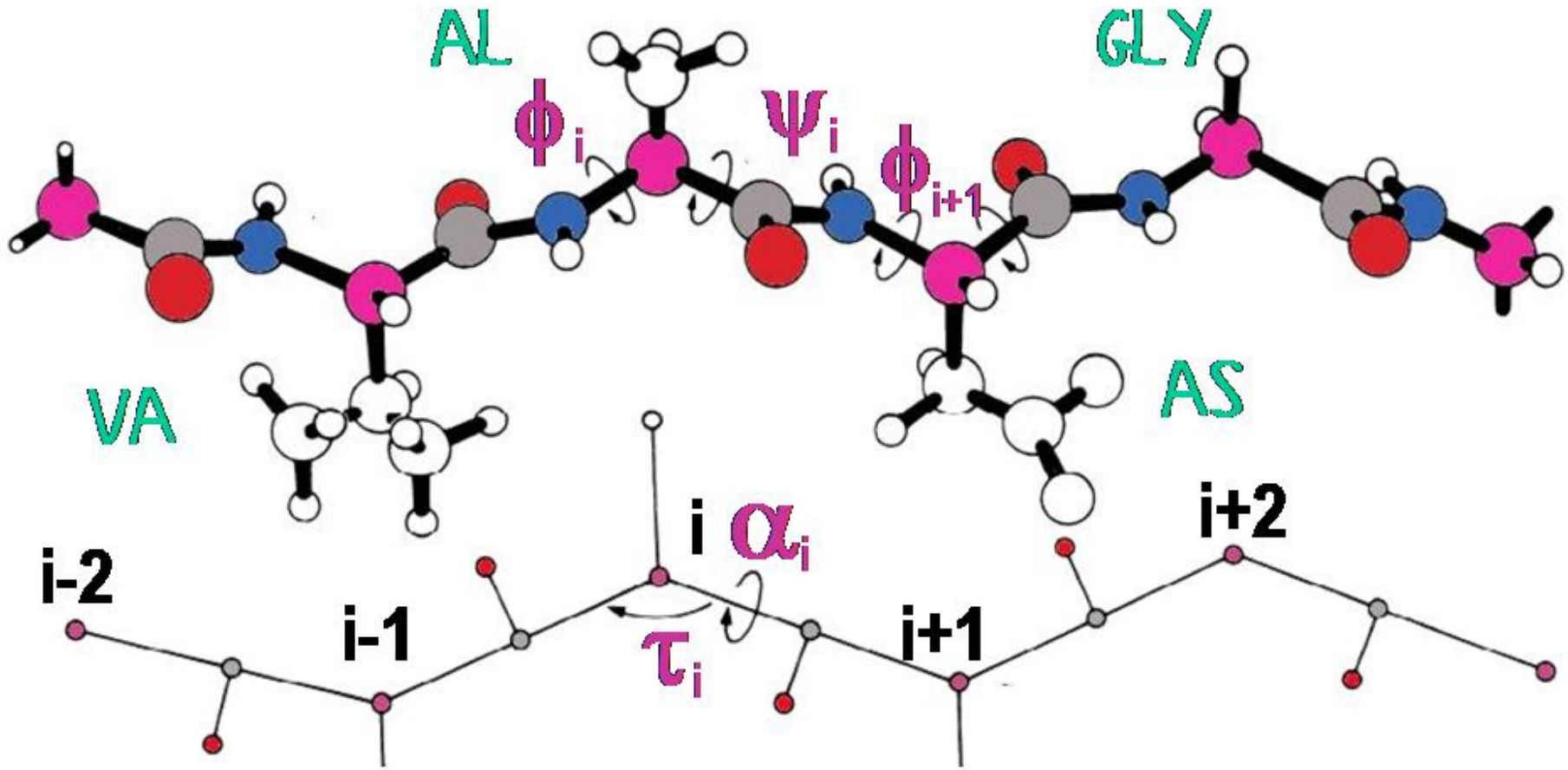


- Must get over high barriers.
- Many degrees of freedom: huge set of possible structures.

SIMPLIFIED MODELS FOR FOLDING

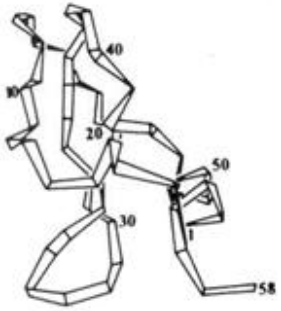


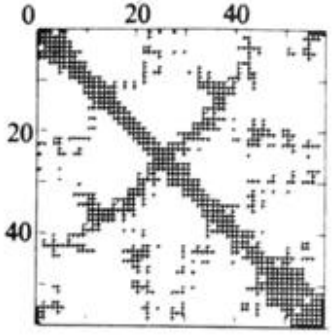
VIRTUAL BONDS



- α_i is defined by CA_{i-1} - CA_i - CA_{i+1} - $_{i+2}$
- α_i is approximately $\psi_i + \phi_{i+1} + 180$

SIMPLIFIED MODELS FOR FOLDING

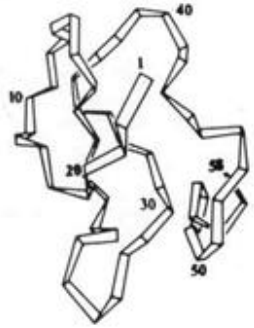


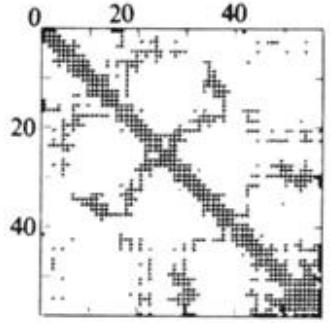


Native
X-Ray
Structure

This seemed significant in 1975.

The overall chain path is similar.





Folded.
dRMS=5.5Å
cRMS=8.1Å

It was not really significant.

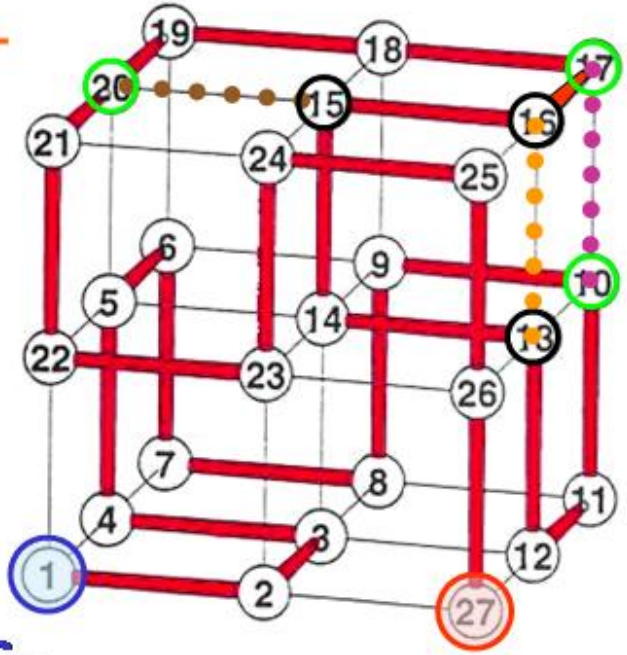
- Simplify the atomic structure to one center per residue (CA).
- Use effective hydrophobic forces with one degree of freedom per residue.
- Fold by energy minimization.
- Use Normal mode thermalization to escape local minima.
  (pump energy into low-frequency modes).

Levitt & Warshel. Computer Simulation of Protein Folding. Nature, 253: 694-698 (1975).

CARTOON FOLDING AND UNFOLDING OF BPTI

3x3x3 CUBE PROTEIN FOLDING

The

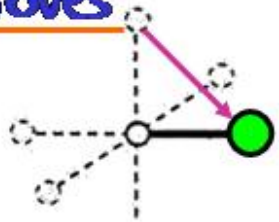


- Model protein as 27 centers of 3 by 3 by 3 cube. This is a lattice.
- Have a simple pair-wise energy.
- Have a simple move set.
- Fold with Monte Carlo:
  accept new arrangement if

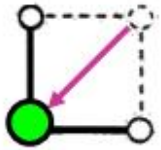
$$\exp(-\Delta U/kT) > \text{Ran}(1),$$

  where Ran(1) is a random number between 0 and 1.
- Repeat about 1,000,000 times to get native cube state.

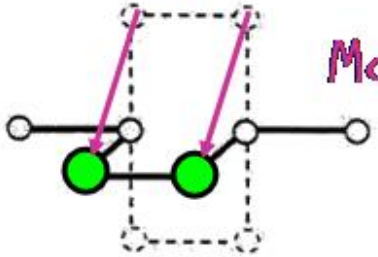
The Moves



Waggle



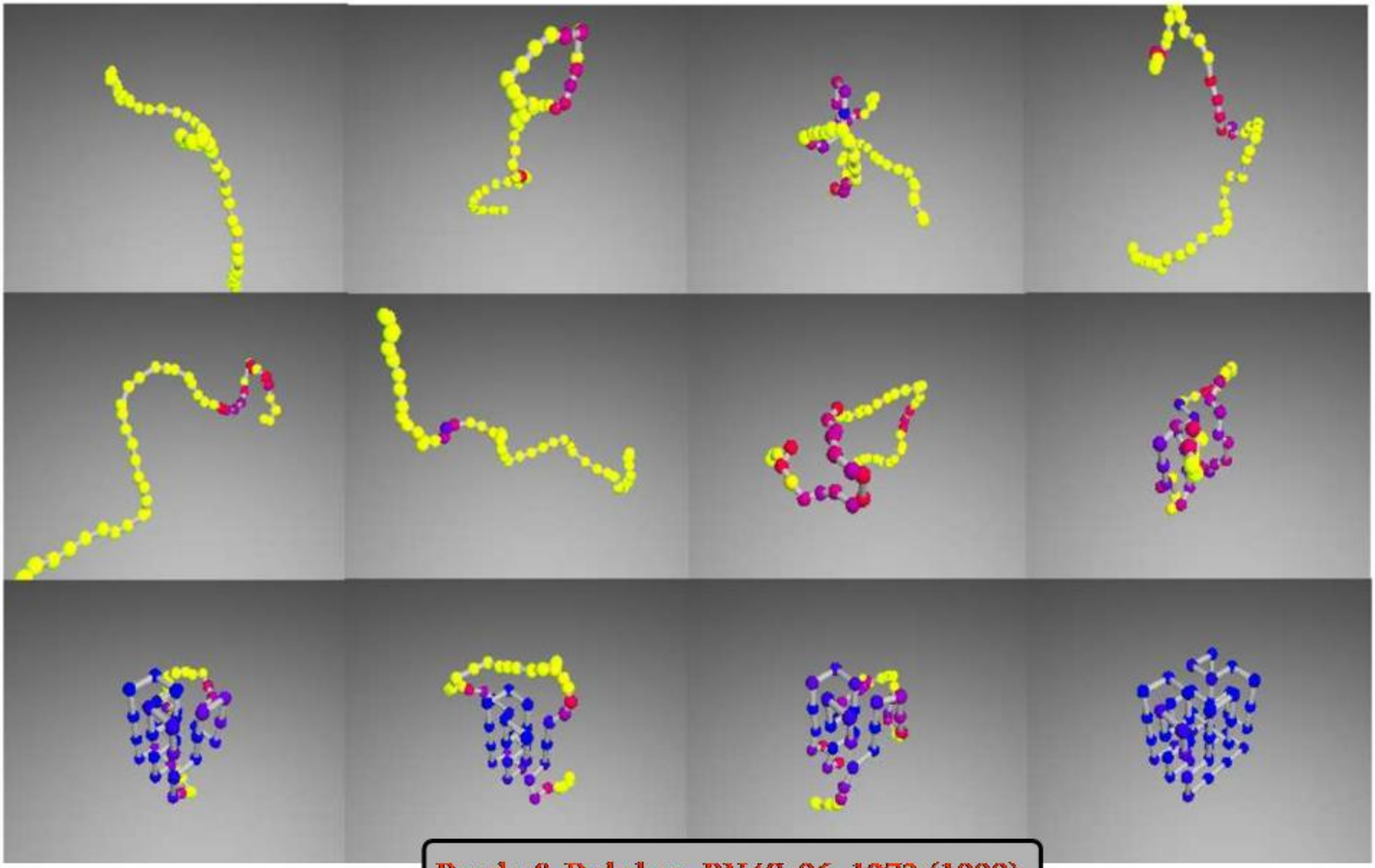
Move



Move as

Sali et a. *J.Mol.Biol.* 235, 1614 (1994)

LATTICE MODEL OF FOLDING



Pande & Rokshar. *PNAS*, 96, 1273 (1999).

Folding Simulations Concept 5.6

FOLDING SIMULATIONS

Need Massive Computational

Villin Folding (A Small Proteins).

Blue Gene Project.

Folding@Home.

Other Simulations.

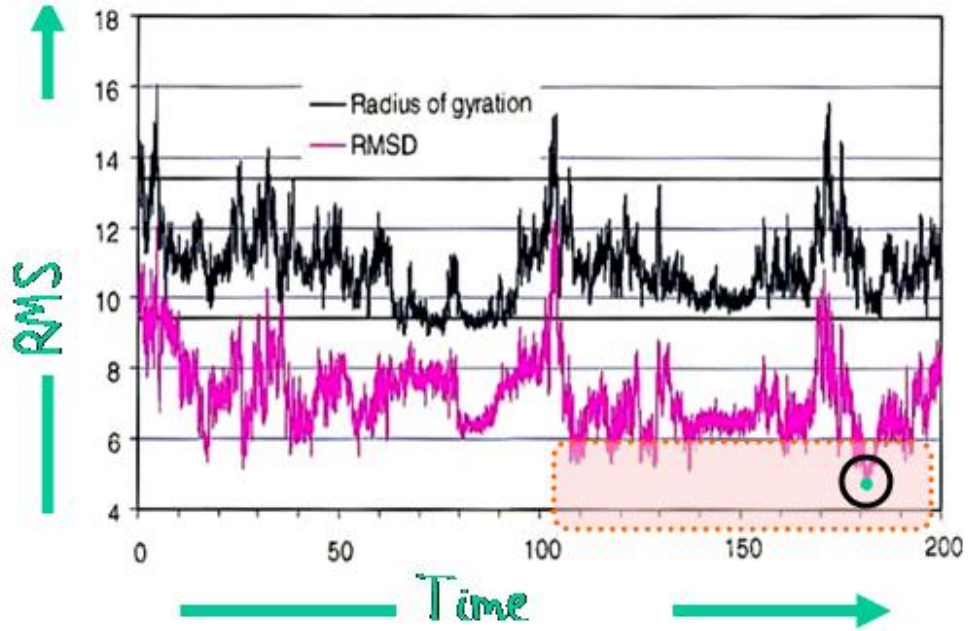
NEED MASSIVE COMPUTATIONAL RESOURCES

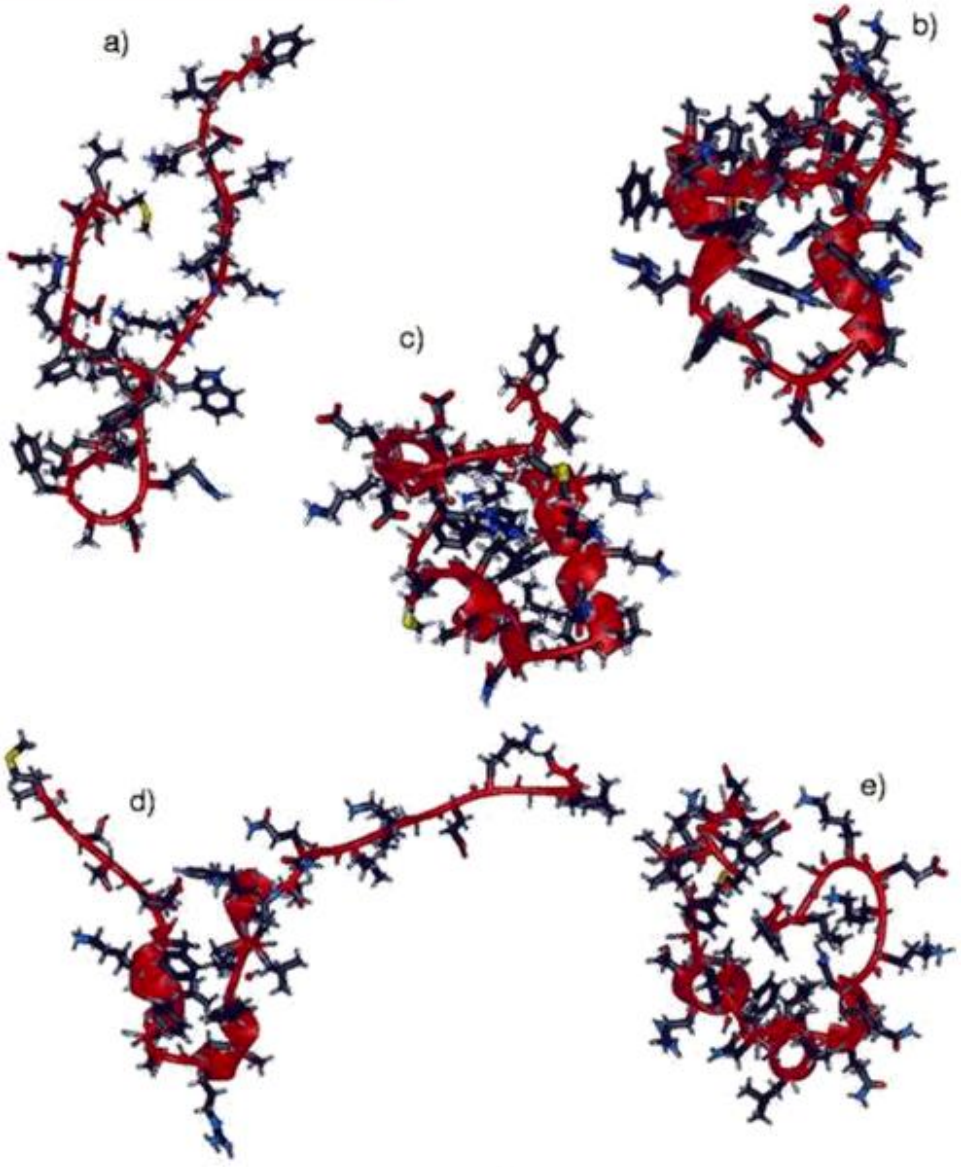
Empty Supercomputers.

Blue Gene (IBM).

Folding@home (Vijay Pande).

VILLIN FOLDING

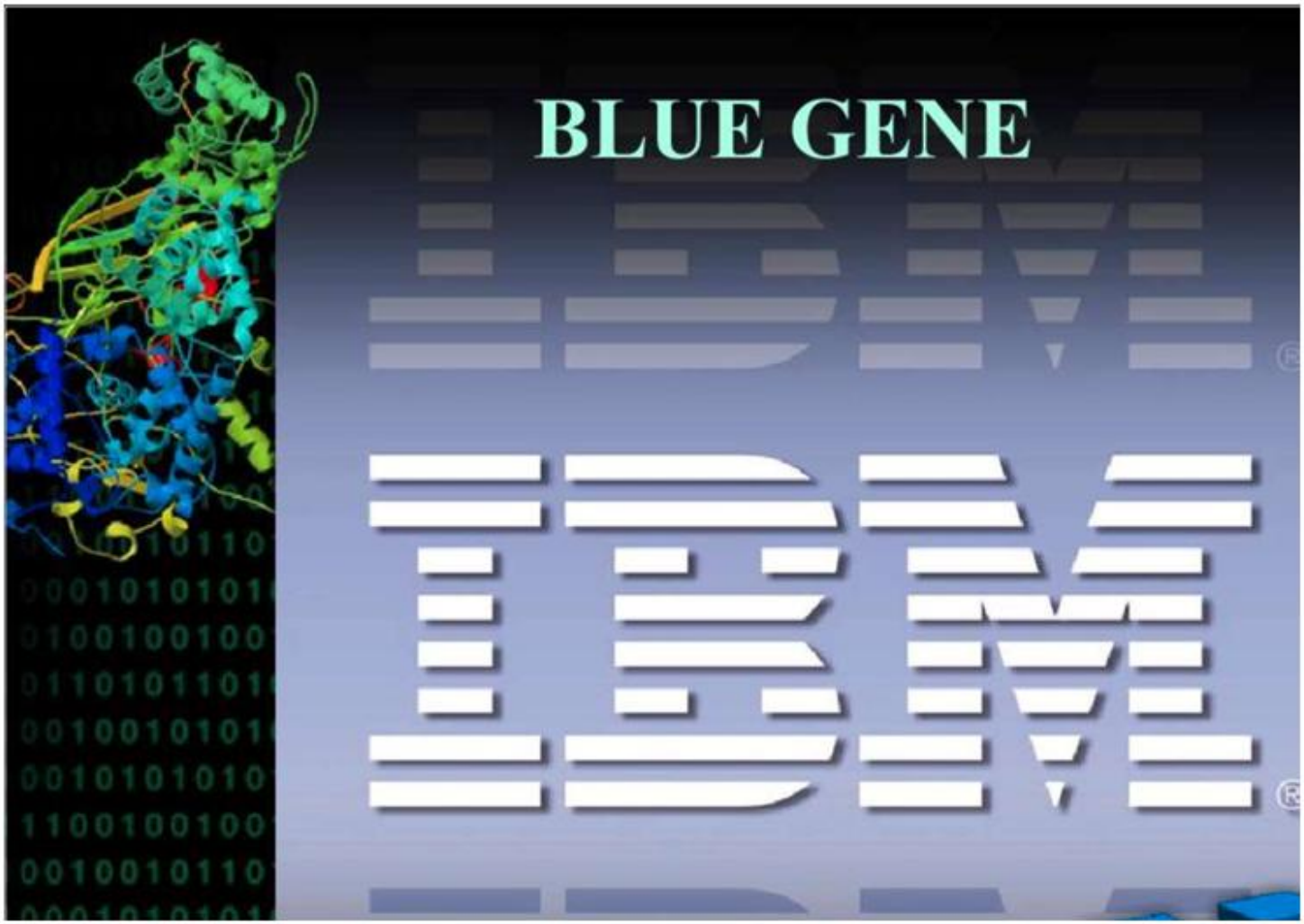




- Use explicit water molecules with 36-residue villin headpiece.
- Have between 3,000 and 6,500 water molecules.
- Run for 200 ns (tour de force).
- Get to within 4.7 Å RMS.

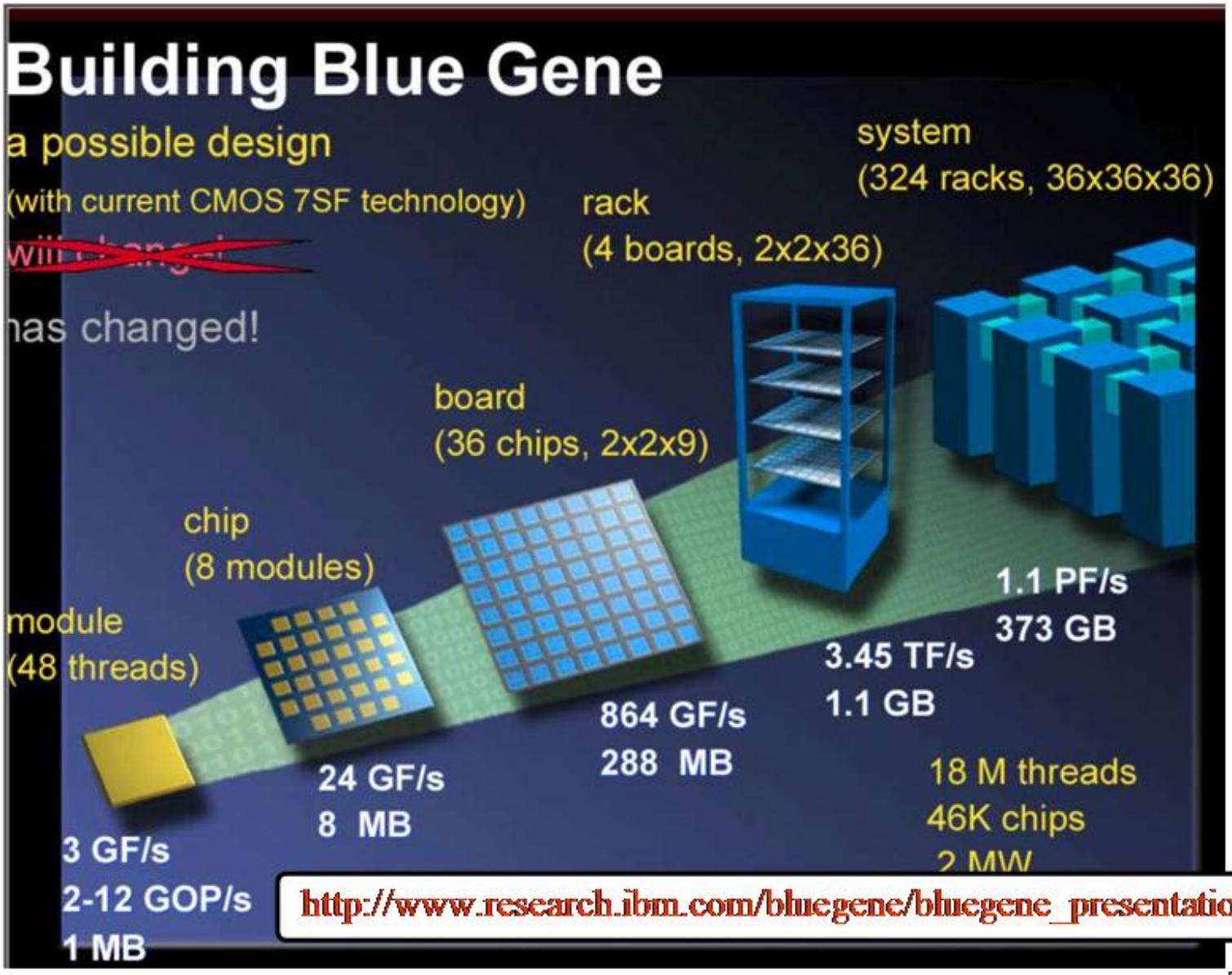
Duan, et al. *PNAS*, 95, 9897 (1998)

IBM BLUE GENE PROJECT



http://www.research.ibm.com/bluegene/bluegene_presentation.pdf

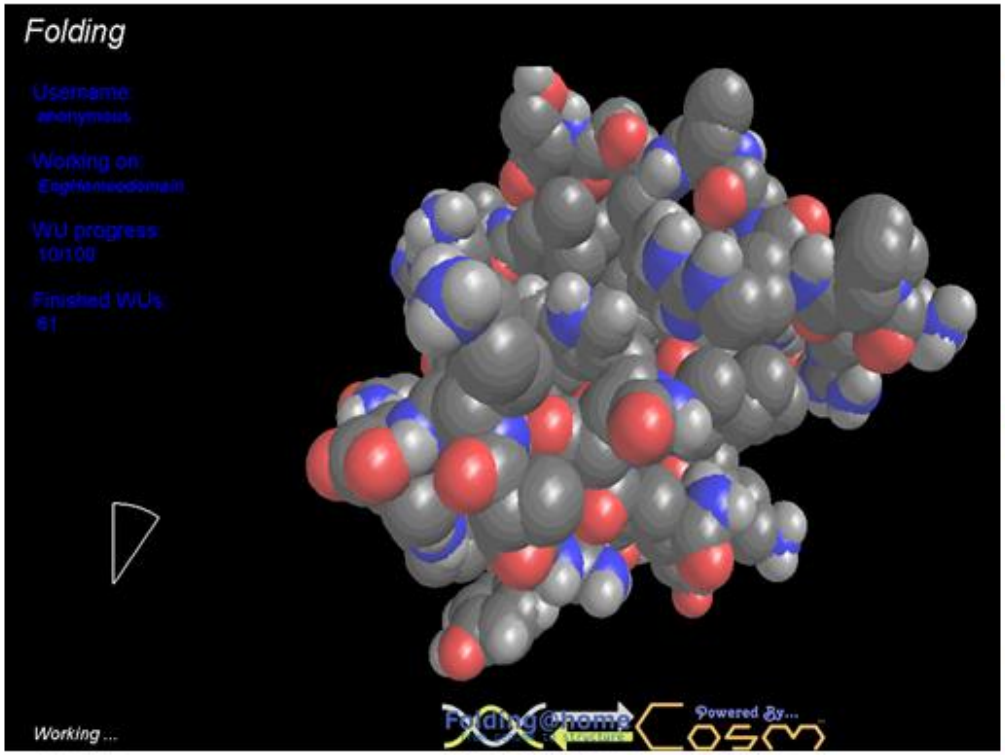
IBM BLUE GENE DESIGN



- 5 years?
- $100

FOLDING AT HOME

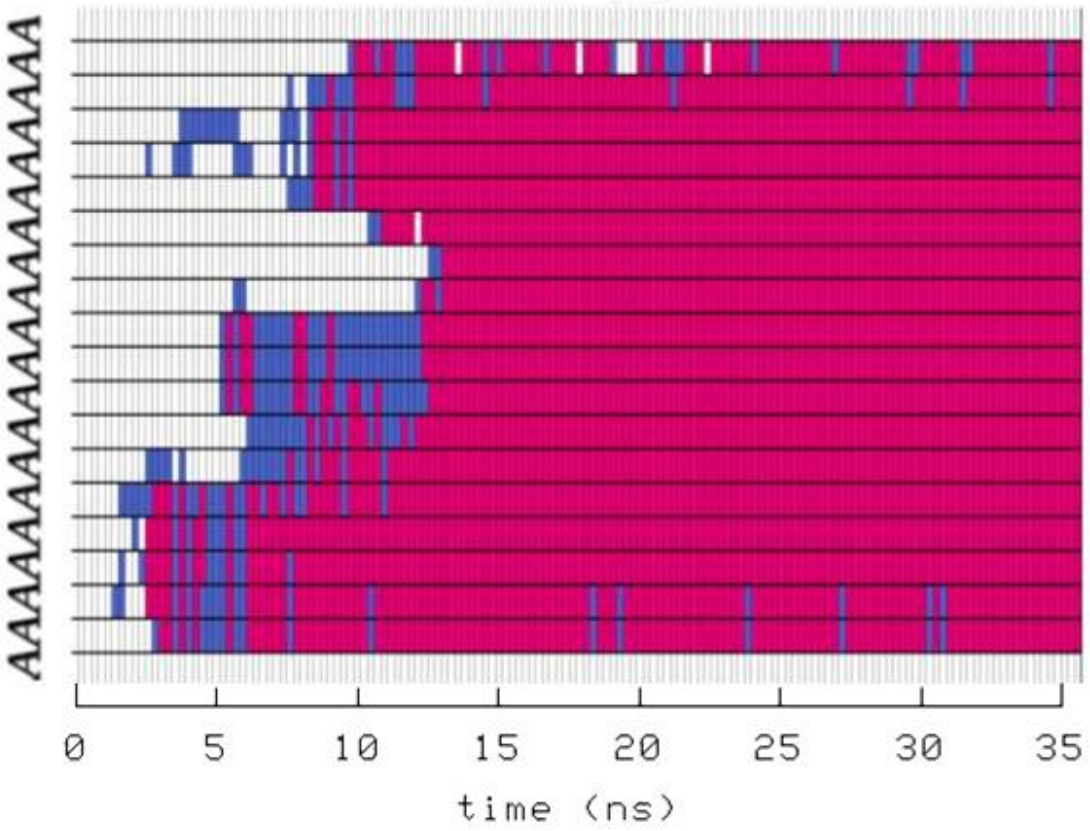
http://www.stanford.edu/group/pandegroup/folding/education/



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

Like

FOLDING AT HOME HELIX FOLDING

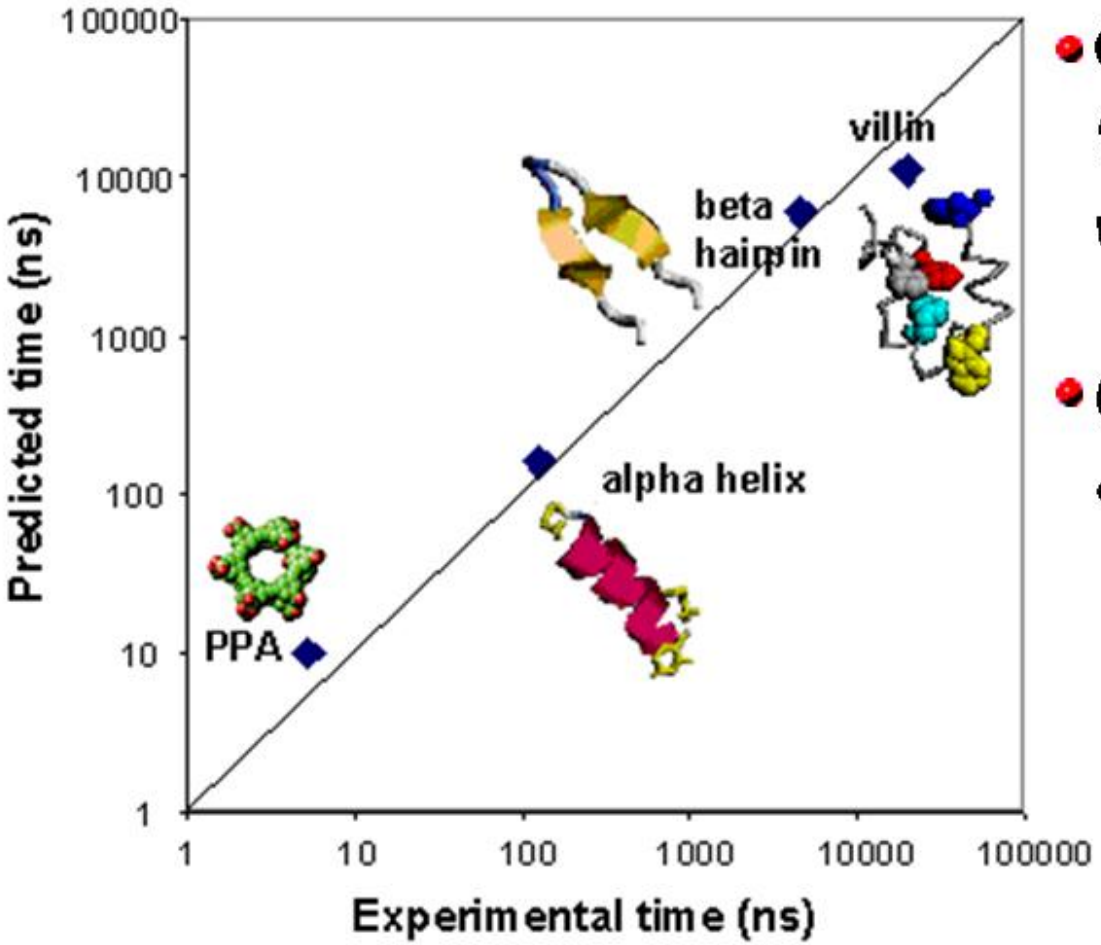


- Run with effective solvent (pseudo vacuum) using Tinker, Jay Ponder's molecular dynamics simulation program.
- Reproducibly fold helix in 10's of nanoseconds (10,000,000 Δt steps).

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.