

Predicting
Free Energy Changes
with Molecular Dynamics

David Mathews

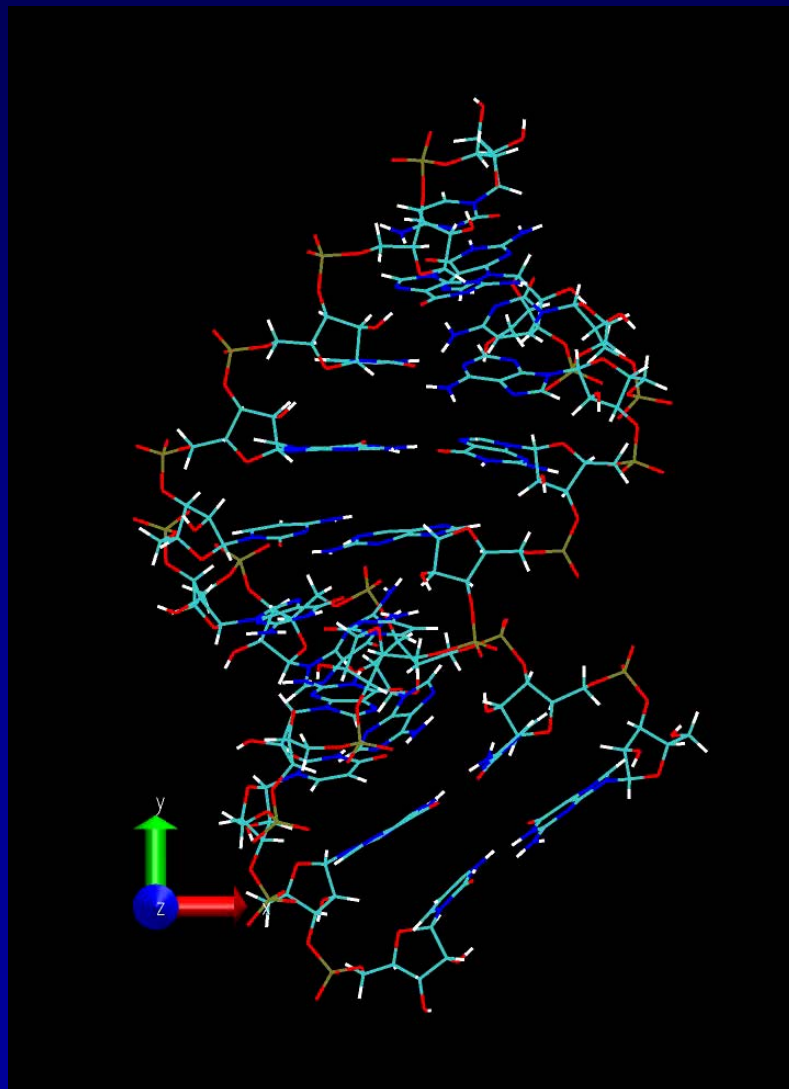
Contact Info:

- David_Mathews@urmc.rochester.edu
- Phone: x51734
- Office: 3.6830 (Medical Center)
- Web: <http://rna.urmc.rochester.edu>

Overview of Lecture:

- Review of meaning of free energy change.
- Ergodic hypothesis
- Free energy perturbation
- Umbrella sampling

Connecting MD Simulation to Thermodynamics:



Molecular Dynamics can Predict Thermodynamic Quantities:

- Conformational free energy changes.
- pKa shifts for protein side changes.
- Binding free energies.

Molecular Dynamics can Predict Thermodynamic Quantities:

- Thermodynamics uses statistical methods to understand the behavior of molecules.
- How can we predict thermodynamic behavior using MD?
- Bad Approach:
 - Start a simulation with a large number of molecules.
 - Memory use is $O(N)$ and time scales $O(N^2)$ (or $O(N \log N)$) where N is the number of atoms.

Molecular Dynamics can Predict Thermodynamic Quantities:

- Ergodic hypothesis:
 - The conformations of a molecular sampled by molecular dynamics as a function of time approximate the conformations of a molecule in a thermodynamic ensemble.
 - The ergodic hypothesis connects simulations of single molecules to macroscopic quantities.

Ergodic Hypothesis: (Ergodicity):

$$\bar{A} = \frac{1}{t} \int_{t_0}^{t_0+t} A(t) dt$$

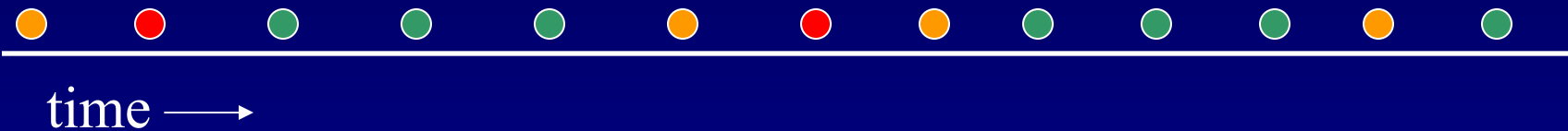
This will change depending on the starting conditions.

$$\bar{A} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_{t_0}^{t_0+t} A(t) dt = \frac{1}{M} \sum_{i=1}^M A_i = \langle A \rangle$$

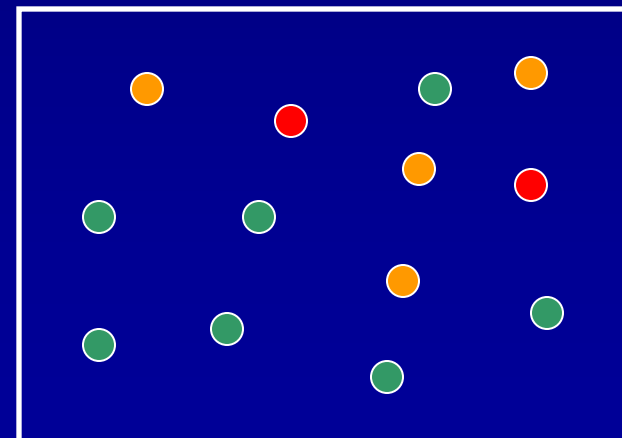
This says that an average quantity measured over long enough time will approach the ensemble average.

Ergodic Hypothesis:

- Relates the behavior of a single molecule as a function of time to the behavior of an ensemble.



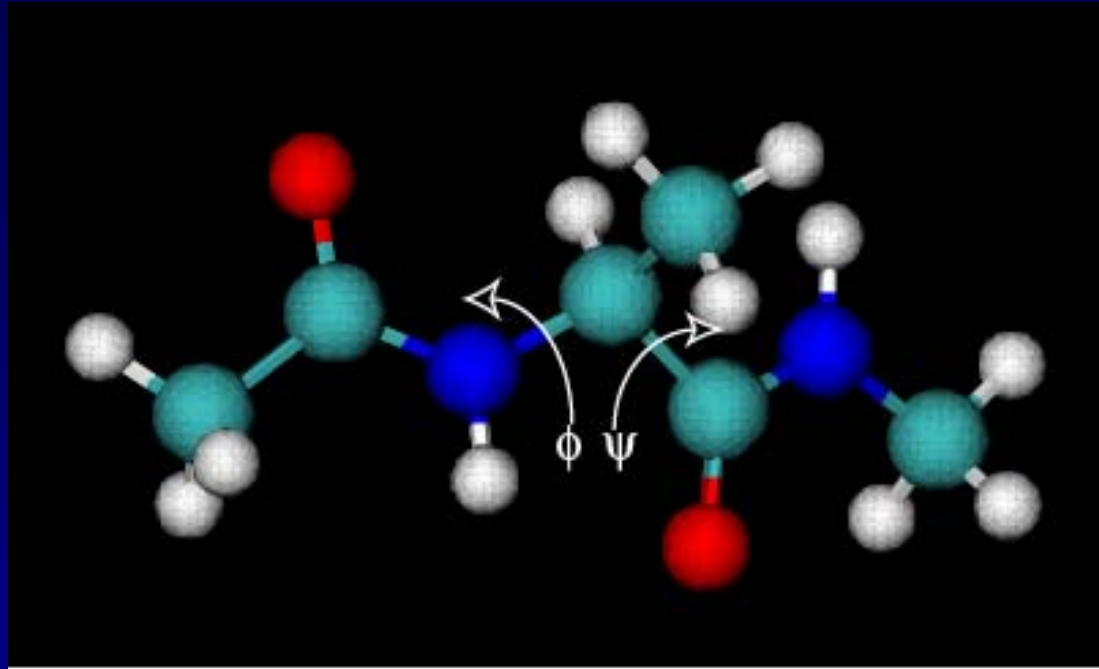
Green: Orange: Red
7:4:2



Is the Ergodic Hypothesis True in General?

- No.
 - Because sampling time is too short.
- The ergodic hypothesis is true for many special cases.
 - Where barriers between states are small (on the order of RT), we can model the thermodynamic behavior of the two states in relation to each other.

Alanine Dipeptide:



ACE

ALA

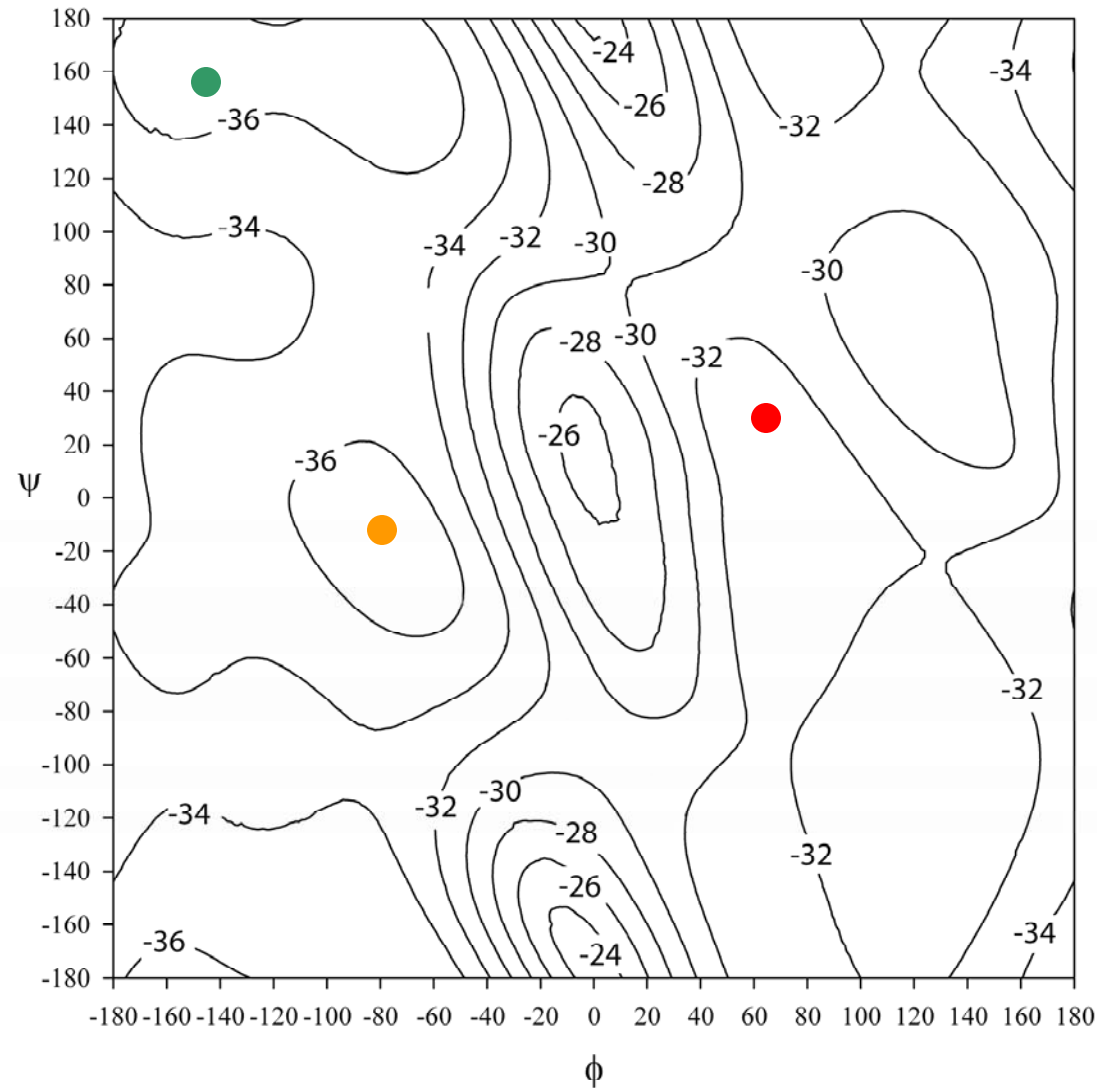
NME

(acetyl)

(N-methylamide)

- Alanine dipeptide used as a model molecule because the energy surface can be displayed topographically.

An Example with Alanine Dipeptide:



Gibb's Free Energy (ΔG°):



$$K = \frac{[B]}{[A]} = e^{-\Delta G_i^\circ / RT}$$

where R is the gas constant and T is the absolute temperature.

ΔG° quantifies an equilibrium between two states.

About Free Energy:

- Free energies in the NVT ensemble are Helmholtz Free Energies (A).
- Free energies in the NPT ensemble are Gibb's Free Energies (G).

Absolute Free Energy:

In a canonical ensemble :

$$Q = \int \int e^{-V/kT} dqdp$$

$$A = -kT \ln Q$$

$$= kT \ln(1/Q)$$

$$= kT \ln \frac{\int \int e^{V/kT} e^{-V/kT} dqdp}{\int \int e^{-V/kT} dqdp}$$

$$= kT \ln \langle e^{E/kT} \rangle$$

- This is self-defeating because the probable states ($e^{-V/kt}$) with low energy are the states that contribute little to the integral ($e^{V/kT}$) and *vice versa*.

Free Energies Changes can be Predicted:

- $\Delta G = \Delta H - T\Delta S$
- Conformational potential energy changes provide the enthalpy changes.
- The entropy change is derived from understanding the ensemble.
 $S = \ln(\Omega)$.

Basic Free Energies Changes:

- Change from configuration 1 to configuration 2.
- $\Delta G = -RT \ln K$
- $\Delta G = -RT \ln P_2/P_1$

where P_2 is the probability of being in configuration 2 and P_1 is the probability of being in configuration 1.

This method works when the energy barrier is very low, $\sim RT$ (~ 1 kcal/mol).

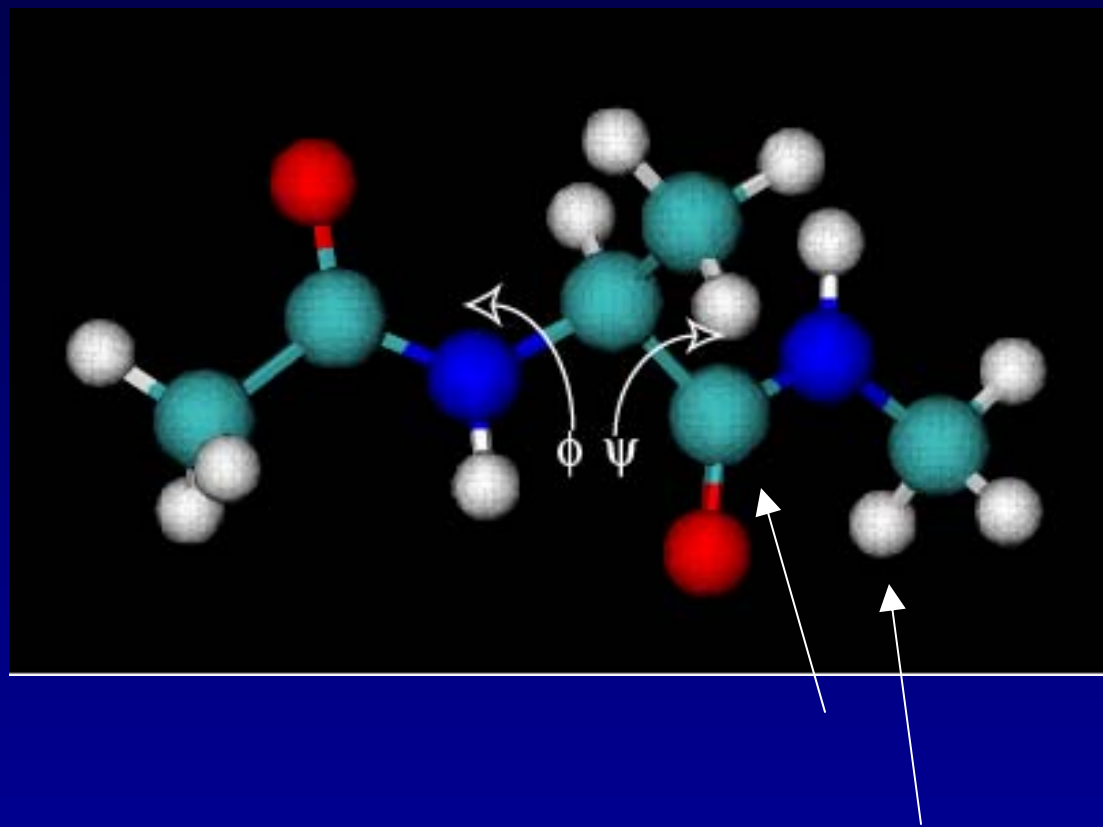
Example with Green, Orange, Red:

- Green: Orange: Red :: 7:4:2



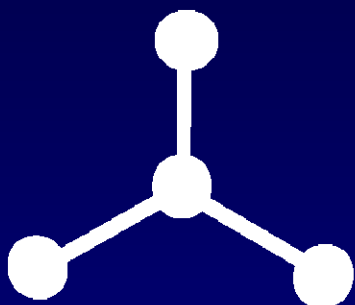
$$\begin{aligned}\Delta G_{\text{red-green}} &= -RT \ln K = -RT \ln (7/2) \\ &= - (1.987 \text{ cal/mol/degree})(310.15 \text{ K}) \ln (7/2) \\ &= -772 \text{ cal/mol} = -0.772 \text{ kcal/mol}\end{aligned}$$

An Example with Alanine Dipeptide:

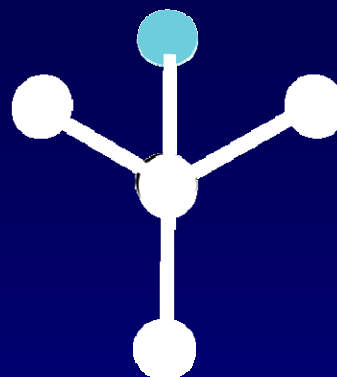


Calculate the ΔG for the dihedral angles of the terminal methyl hydrogens. The two states are eclipsed and staggered as compared to the carbon.

Dihedrals:



eclipsed



staggered

Eclipsed:

$$-30^\circ \leq \text{dihedral} < 30^\circ$$

$$90^\circ \leq \text{dihedral} < 150^\circ$$

$$210^\circ \leq \text{dihedral} < 270^\circ$$

Staggered:

$$30^\circ \leq \text{dihedral} < 90^\circ$$

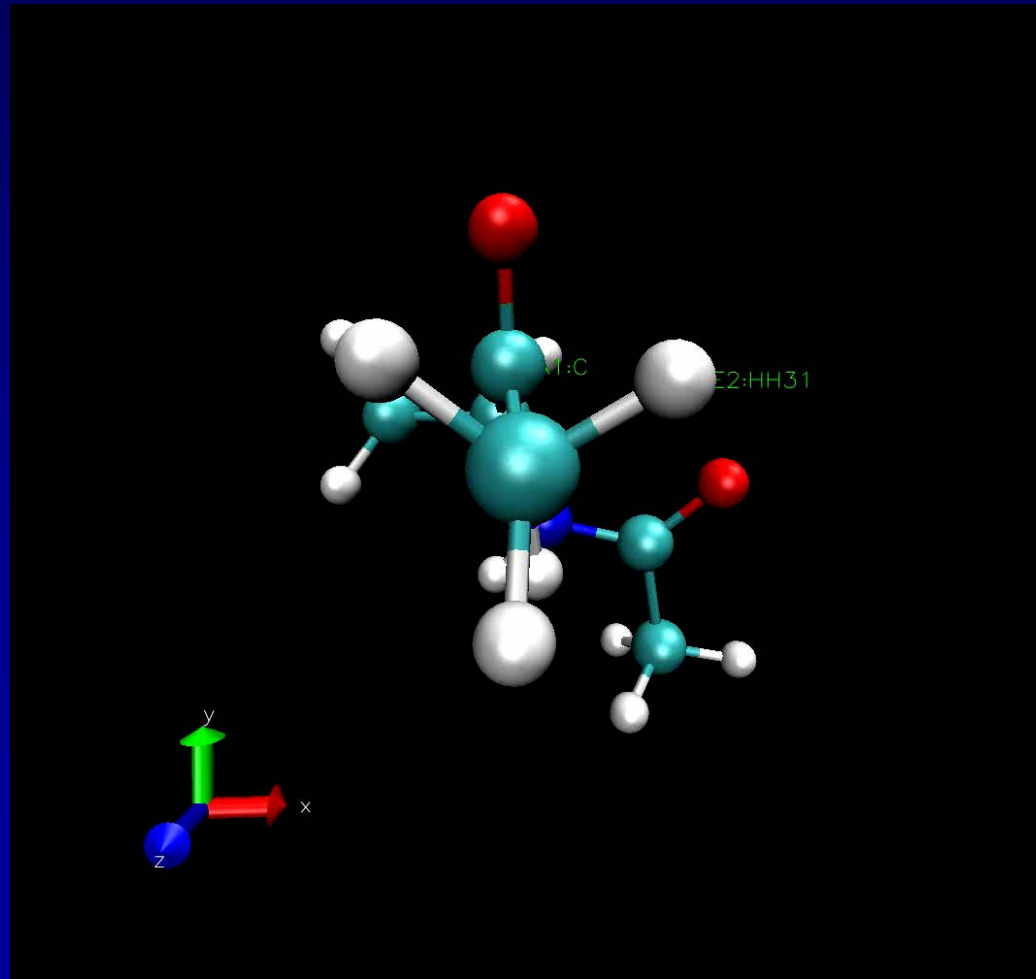
$$150^\circ \leq \text{dihedral} < 210^\circ$$

$$270^\circ \leq \text{dihedral} < 330^\circ$$

Note that $-30^\circ = 330^\circ$

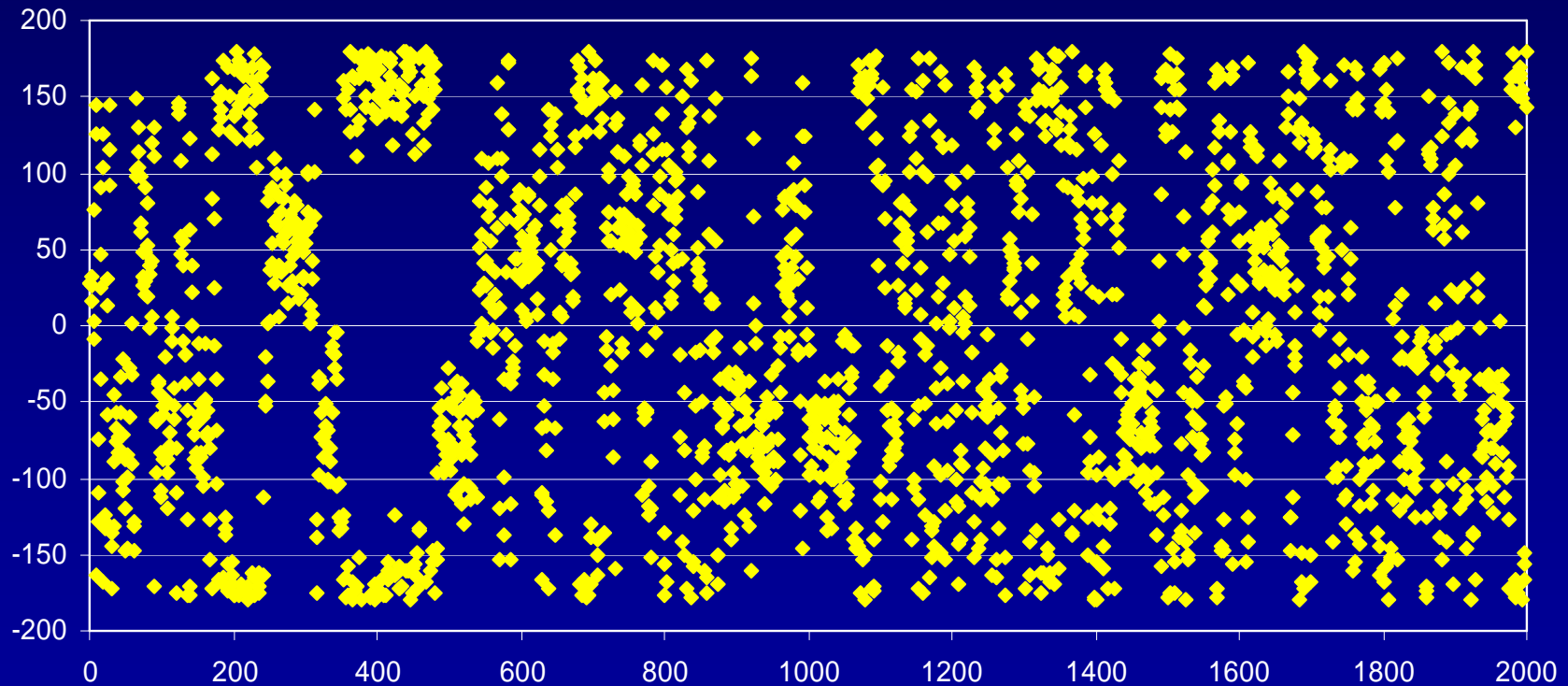
Molecular Dynamics:

- Run 100 ps of MD, sampling dihedrals every 50 fs for 2000 data points.



Molecular Dynamics:

Run 100 ps of MD, sampling dihedrals every 50 fs
for 2000 data points.



Staggered = 1282, Eclipsed = 718

Free Energy Change:

- Staggered:eclipsed :: 1282:718

$$\begin{aligned}\Delta G &= -RT \ln K = -RT \ln (1282/718) \\ &= - (1.987 \text{ cal/mol/degree})(310.15 \text{ K}) \\ &\quad \ln (1282/718) \\ &= -346 \text{ cal/mol} = -0.346 \text{ kcal/mol}\end{aligned}$$

Free Energy Differences are Tractable:

- Calculate the free energy difference between A and B.
- $V_B = V_A + V_{BA}$

$$\Delta A = A_B - A_A = -kT \ln Q_B/Q_A$$

$$\begin{aligned}\Delta A &= -kT \ln \frac{\int \int e^{-V_B/kT} dqdp}{\int \int e^{-V_A/kT} dqdp} \\ &= -kT \ln \frac{\int \int (e^{-V_A/kT} e^{+V_A/kT}) e^{-V_B/kT} dqdp}{\int \int e^{-V_A/kT} dqdp} \\ &= -kT \ln \frac{\int \int e^{-V_A/kT} e^{-V_{BA}/kT} dqdp}{\int \int e^{-V_A/kT} dqdp} \\ &= -kT \ln \langle e^{-V_{BA}/kT} \rangle_A\end{aligned}$$

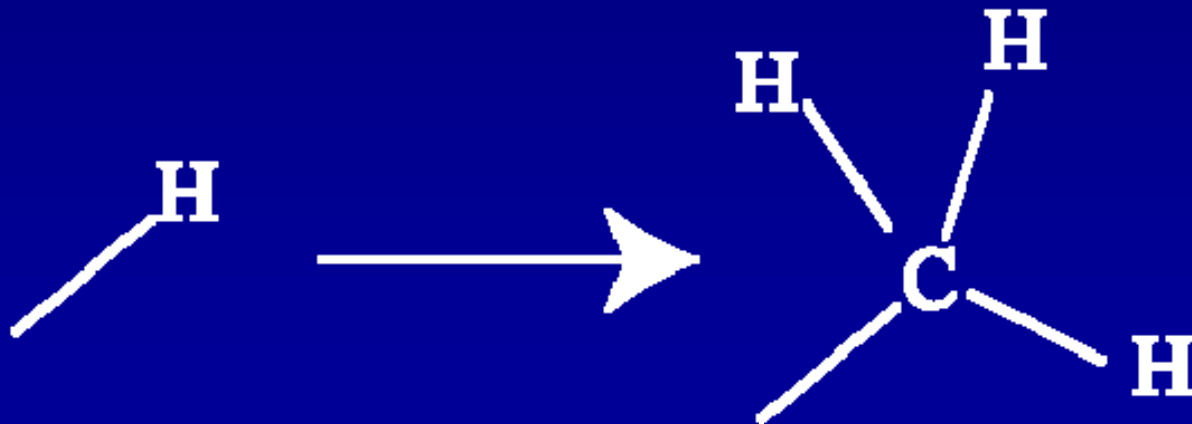
Free Energy Perturbation:

$$\Delta A = -kT \ln \langle e^{V_{BA}/kT} \rangle_A$$

- The free energy difference is a function of the potential difference between state A and state B as an ensemble average over states in A.
- Unlike perturbation theory (Quantum mechanics) this is an exact equation.
- Examples:
 - Predict the effect on stability of changes in amino acid sequence.
 - Predict pKa's for titratable side chains

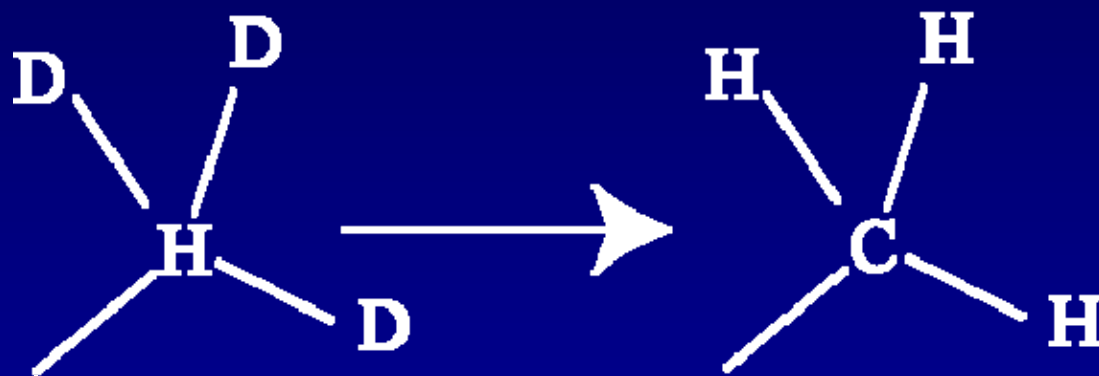
What About Atom Numbers:

- We stated that we are using the NVT ensemble.
- Interesting perturbations require changes in atom number.
- Consider the change from glycine to alanine:



Dummy Atoms:

- Add dummy atoms that do not interact with the system:



Practical Scheme:

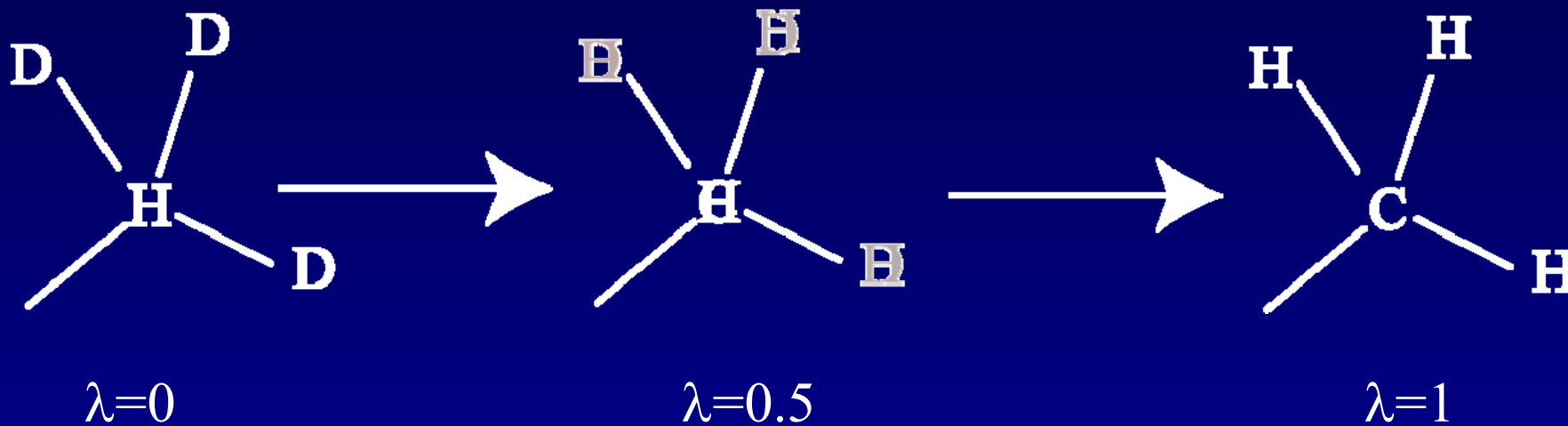
$$\Delta A = -kT \ln \langle e^{V_{BA}/kT} \rangle_A$$

- Errors are smaller if the perturbation term, V_{BA} is smaller.

$$\begin{aligned} V(\lambda) &= (1-\lambda)V_A + \lambda V_B \\ \Delta A &= -kT \int_0^1 \frac{\partial \ln \langle e^{-V(\lambda)} \rangle_\lambda}{\partial \lambda} \partial \lambda \\ &= -kT \int_0^1 \frac{\partial \ln \langle 1 + \frac{-V(\lambda)}{kT} \rangle_\lambda}{\partial \lambda} \partial \lambda \\ &= -kT \int_0^1 \frac{\partial \langle \frac{-V(\lambda)}{kT} \rangle_\lambda}{\partial \lambda} \partial \lambda \\ &= \int_0^1 \frac{\partial \langle -V(\lambda) \rangle_\lambda}{\partial \lambda} \partial \lambda \end{aligned}$$

- Calculation can be divided into a number of calculations from $\lambda = 0$ to $\lambda = 1$.

Glycine to Alanine:



$$V(\lambda) = (1-\lambda)V_A + \lambda V_B$$

Solving the Integral:

- Use a Gaussian Quadrature to solve the integral:

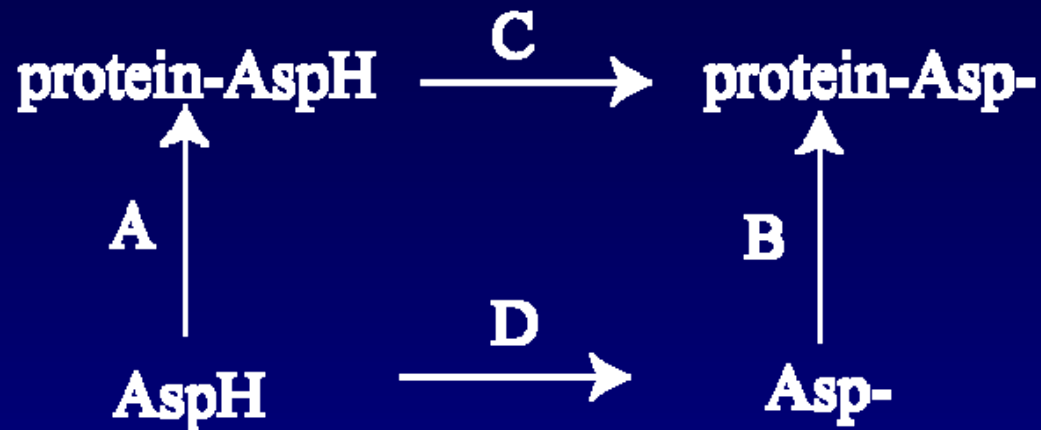
$$\Delta A = \int_0^1 \frac{\partial \langle -V(\lambda) \rangle_\lambda}{\partial \lambda} d\lambda = \sum_{i=1}^n w_i \frac{\partial \langle -V(\lambda) \rangle_{\lambda_i}}{\partial \lambda}$$

- Press *et al.* *Numerical Recipes in {C, Fortran, C++}*. Cambridge University Press.

n	λ_i	w_i
5	.04691	.11846
	.23076	.23931
	.5	.28444
	.76923	.23931
	.95308	.11846
9	.01592	.04064
	.08198	.09032
	.19331	.13031
	.33787	.15617
	.5	.16512
	.66213	.15617
	.80669	.13031
	.91802	.09032
	.98408	.04064

Application: Predicting pKa shifts:

- Simonson, Carlsson, Case. 2004. *JACS* 126:4167.



$$\Delta G = -RT \ln K_a$$

$$\text{p}K_a = -\log_{10} K_a = [1/(2.303kT)]\Delta G$$

$$\text{We want: } \Delta\Delta G = \Delta G_B - \Delta G_A$$

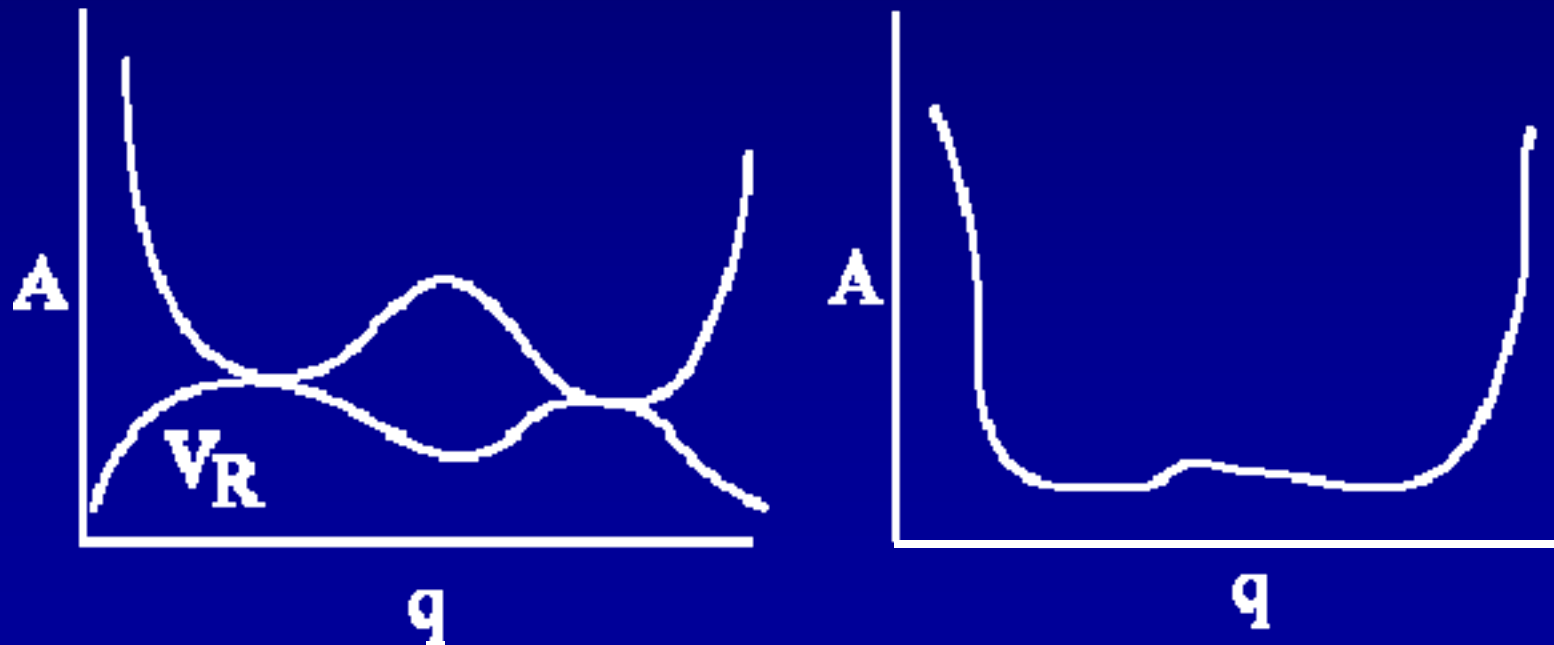
$$\text{Know: } -\Delta G_A + \Delta G_D + \Delta G_B - \Delta G_C = 0$$

$$\text{rearrange: } \Delta\Delta G = \Delta G_B - \Delta G_A = \Delta G_C - \Delta G_D$$

$$\text{p}K_a (\text{C}) = \text{p}K_a (\text{D}) + [1/(2.303kT)]\Delta\Delta G$$

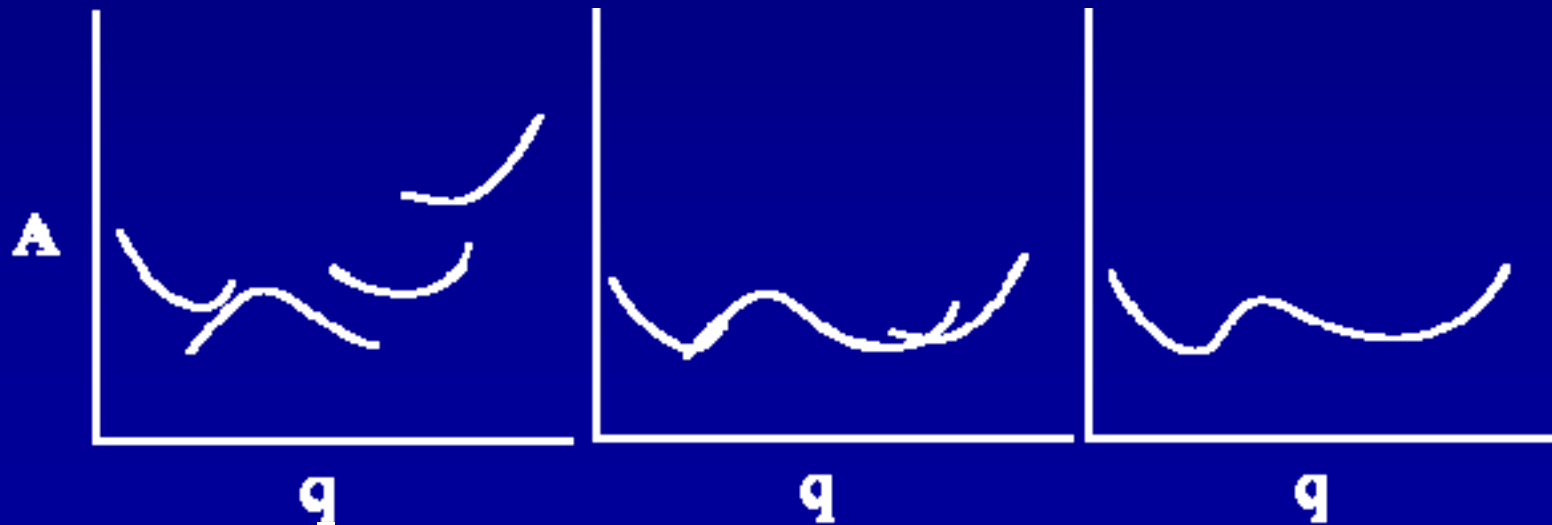
Potential of Mean Force Calculation:

- $A(q) = -kT \ln P(q)$, where q is the reaction coordinate, is the potential of mean force.
- Reduce the barrier height with a restraint potential (called Umbrella Sampling):

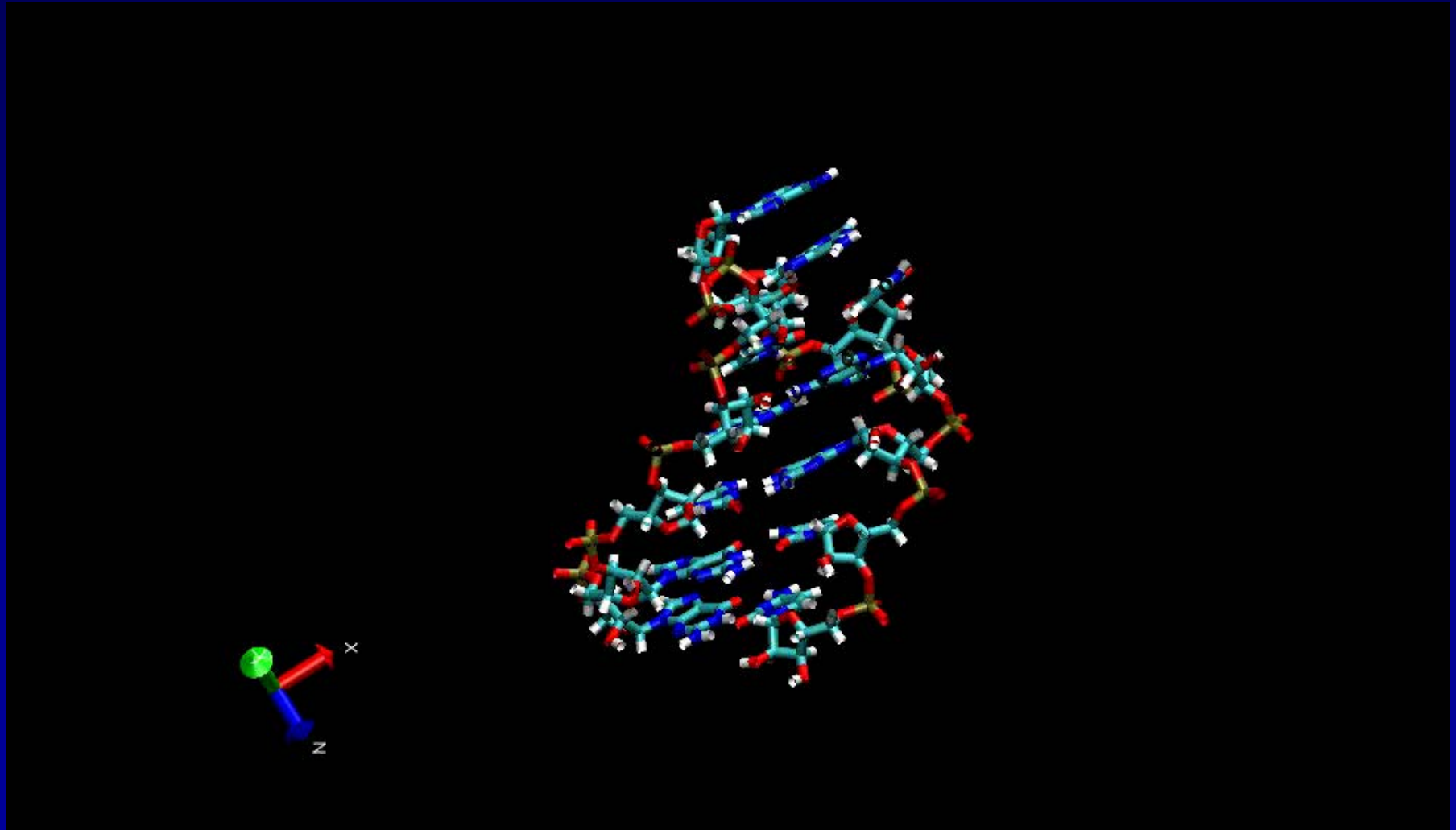


Potential of Mean Force:

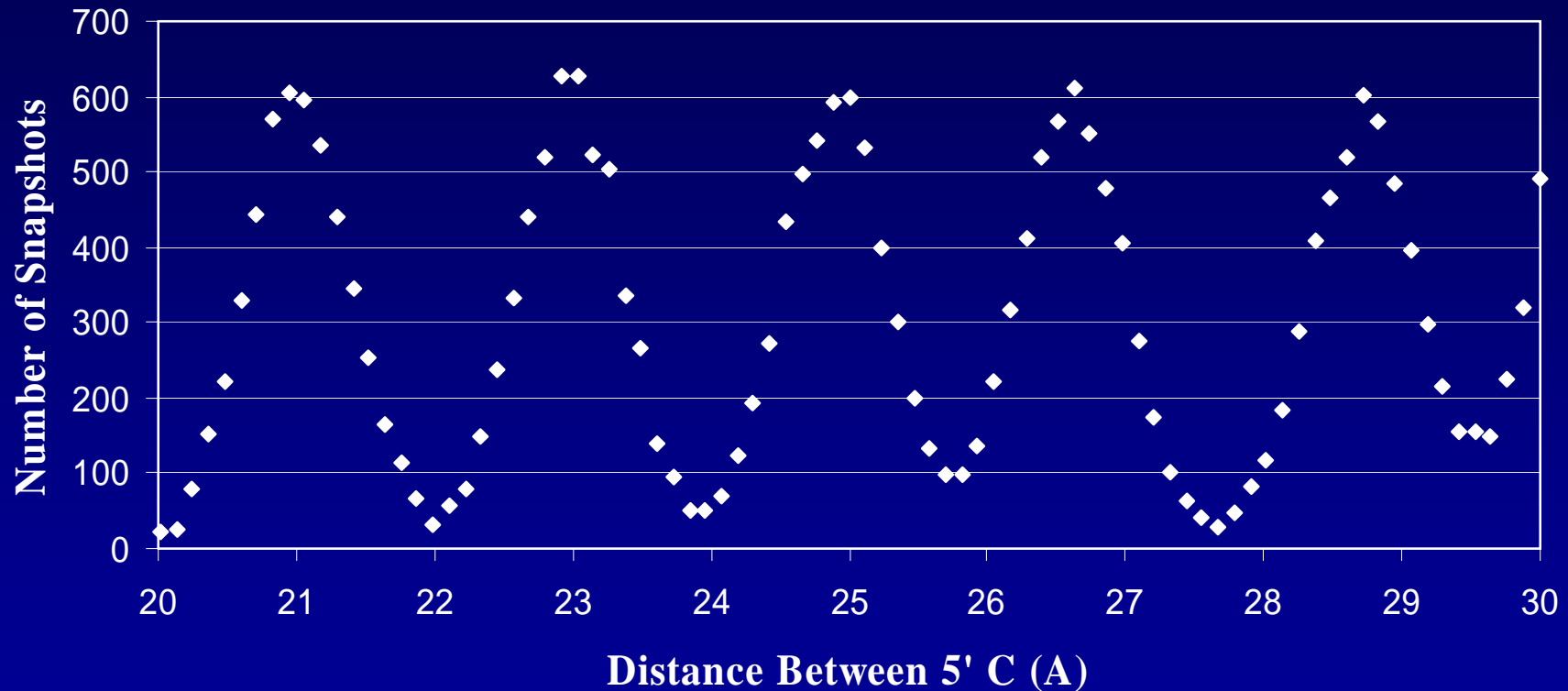
- $A(q) = -kT \ln P(q)$
- $P(q) = P_R(q)[e^{V_R(q)/kT}/C]$
- $-kT \ln (P(q)) = -kT \ln (P_R(q)[e^{V_R(q)/kT}/C])$
- $A(q) = A_R(q) - V_R(q) + C'$
- WHAM (Weighted Histogram Analysis Method):



Example – Denaturation of RNA Hairpin:

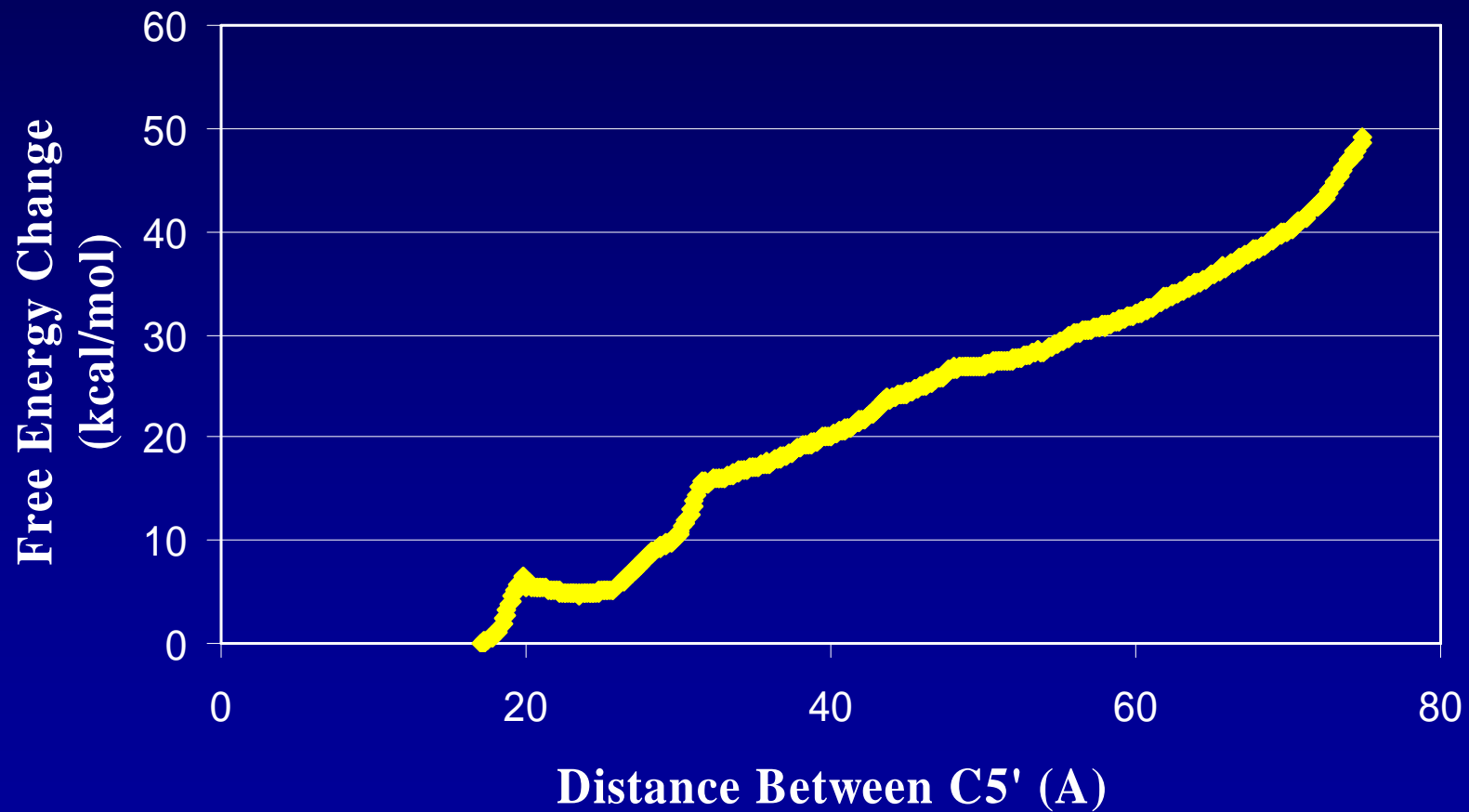


Population Plots Along Denaturation Distance:



Summary of results for simulations centered at 21 Å, 23 Å, 25 Å, 27 Å, and 29 Å,

Stitched Together Free Energy Change:



Convergence:

- What if we want to know the ensemble average for Φ ?
- We would run the simulation for some period of time (1 ns).
- We would calculate Φ (Φ_1).
- Then we would run another 1 ns and recalculate Φ (Φ_2) using the whole time.
 - If $\Phi_1 = \Phi_2$, there is evidence that the simulation converged.
 - If $\Phi_1 \neq \Phi_2$, the calculation has not converged and should be run longer.
- Alternatively, we could start a new simulation from a different starting conformation and calculate Φ .

Free Energy Calculations:

- These basic methods work for many other types of calculations:
 - Binding free energy for small molecules to macromolecules.
 - pKa shift calculations (binding of a proton to a side chain atom)

Further Reading:

- Free energy perturbation (as applied to predicting pKa shifts in protein side chains):
 - Simonson, Carlsson, & Case. 2004. Proton Binding to Proteins: pKa Calculations with Explicit and Implicit Solvent Models. *Journal of the American Chemical Society*. 126: 4167-4180.
- Free Energy of Binding:
 - Gohlke, Kiel, & Case. 2003. Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *J. Mol. Biol.* 330: 891-913.