Free Energy Calculation

in MD Simulation

Basic Thermodynamics

- Helmoholtz free energy

  $A = U - TS + \Sigma \mu_i N_i$

  $dA = w_{rev}$ (reversible, const N V T) eq (22.9) McQuarrie & Simon

- Gibbs free energy

  $G = U + PV - TS + \Sigma \mu_i N_i$

  $= H - TS + \Sigma \mu_i N_i$

  $dG = w_{nonPV}$ (reversible, const N P T) eq (22.16) McQuarrie & Simon

U internal energy P Pressure μ_i Chemical potential

V Volume T Temperature

S Entropy $dS = dq/T$ $S = k_B \ln W$

Implication of Free Energy

- $A \leftrightarrow B \qquad K_{eq} = [A]/[B]$

  $K_{eq} = \exp(-\Delta G_0/RT)$

  $\Delta G_0 = -RT \ln K_{eq}$

  $\Delta G = \Delta G_0 + RT \ln Q$

  $\Delta G > 0$ Unfavorable

  $\Delta G = 0$

  $\Delta G < 0$ Favorable

Statistical Mechanics

- System can be described using Hamiltonian

  $H(p_1,p_2,.....p_N, r_1, r_2,.....r_N)$
- Different ensemble (fixed system quantities)

  - Canonical ensemble (N,V,T)

  - NPT ensemble

Statistical Mechanics

$$A = -k_B T \ln Q_{NVT}$$

$$Q_{NVT} = \frac{1}{h^{3N} N!} \int \int \exp[-\frac{1}{k_B T} H(x, p_x)] dx dp_x$$

$$A = k_B T \ln \left\langle \exp\left[\frac{1}{k_B T} H(x, p_x) \right] \right\rangle$$

Common Free Energy Type

- Solvation Free Energy / Transfer Free Energy
- Binding Free energy
- Confomational Free Energy

Calculating Free Energy

- Experimentally

  - probabilities of finding the system at given states

    $\Delta G = -RT \ln (P_1/P_0)$

  - reversible work of moving the system between two states

- Computationally

  - pobability

  - reversible work → more efficient

Probability Method (Brute Force)

- Ergodic hypothesis

  time average = ensemble average

$$\bar{A} = \lim_{t \to \infty} \frac{1}{t} \int A(t)\, dt = \frac{1}{M} \sum^{M} A_i = \langle A \rangle$$

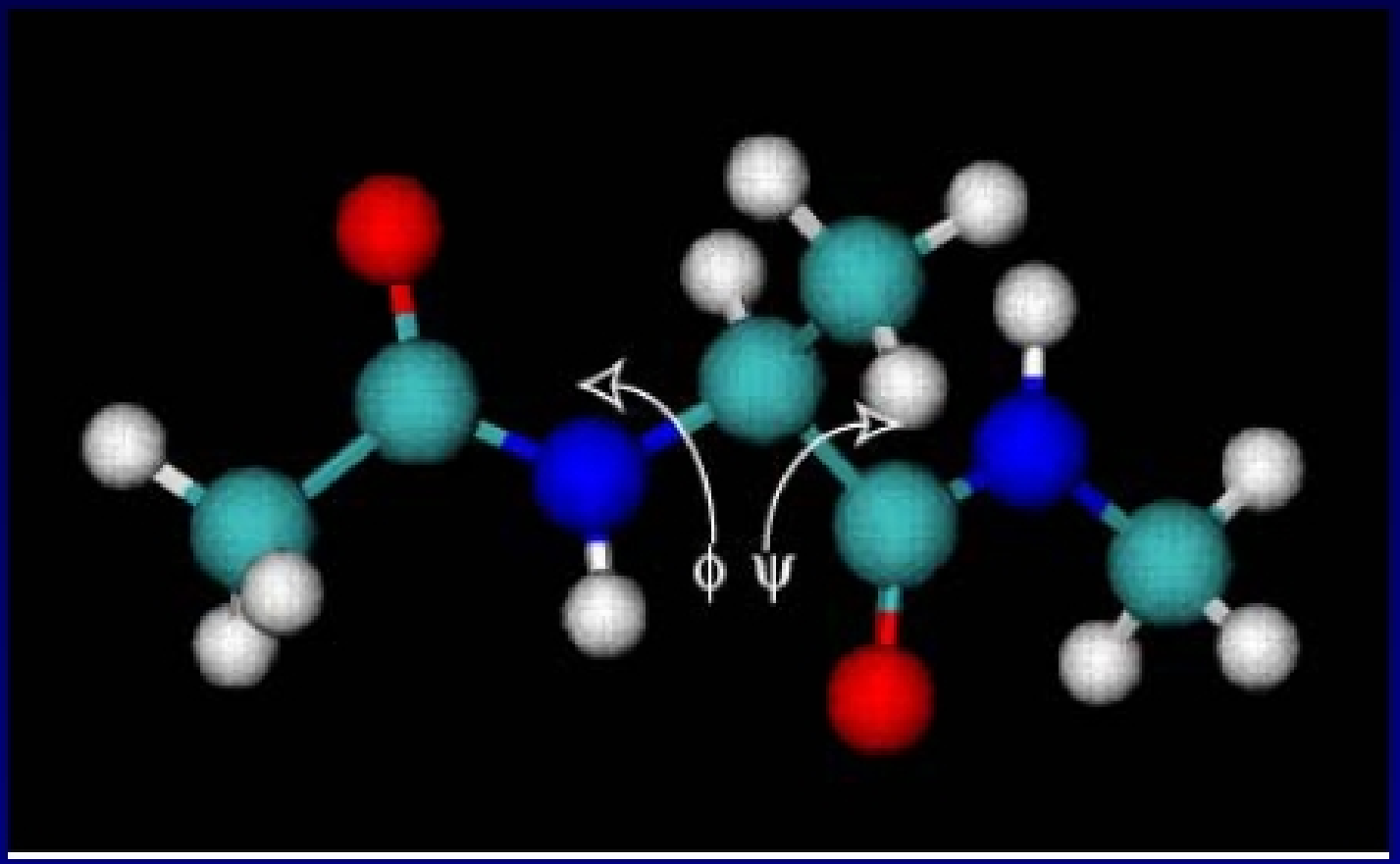
- $\Delta G = -RT \ln (P_1/P_0)$
- Is this true? In general, NO!

  - sampling time is too short computationally

  - hard to sample every state of the system

  For system state of zero sampling, $P_A = 0 \rightarrow \Delta G_{0 \rightarrow A} = \infty$

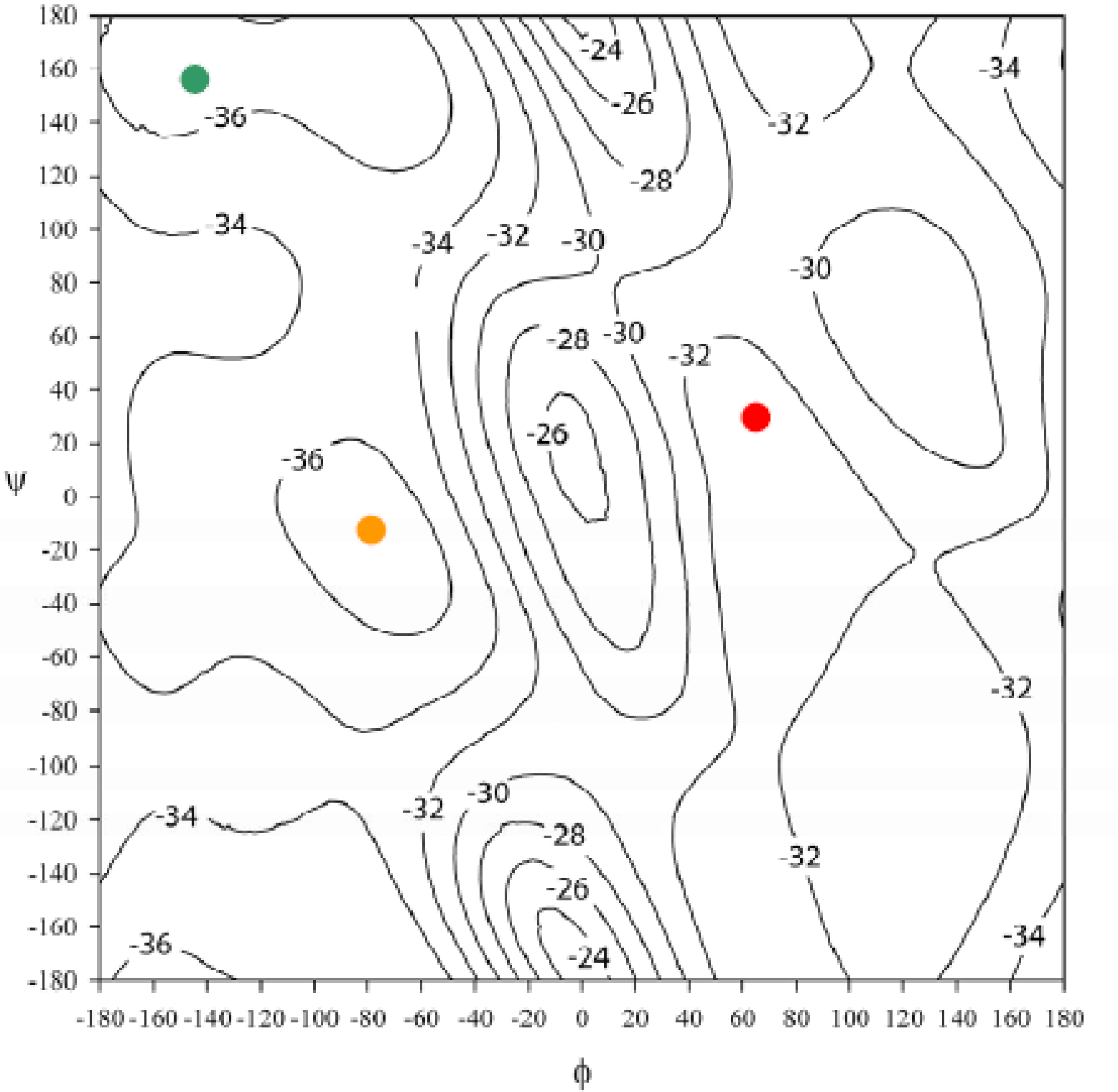
Alanine Dipeptide

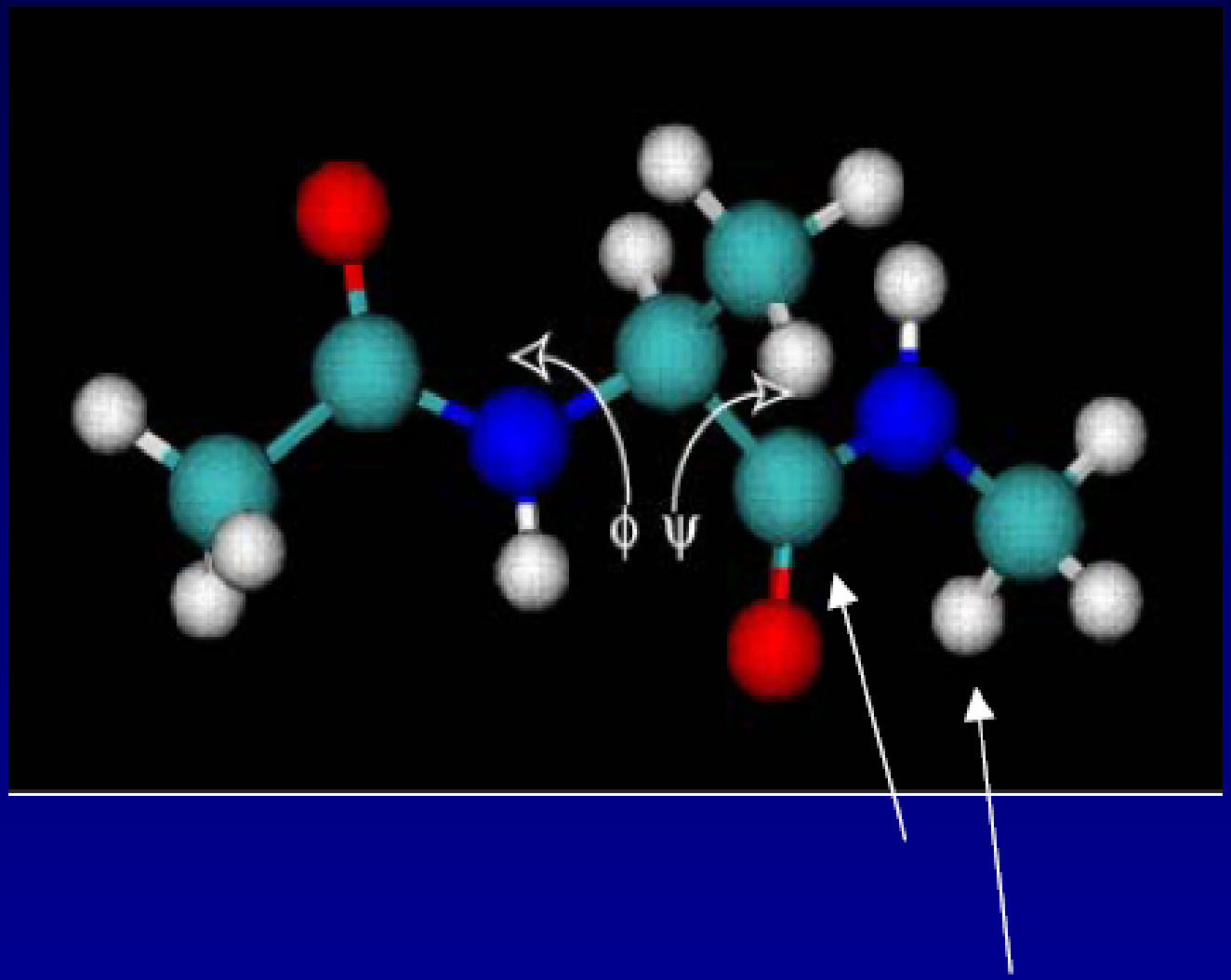


ACE (acetyl)

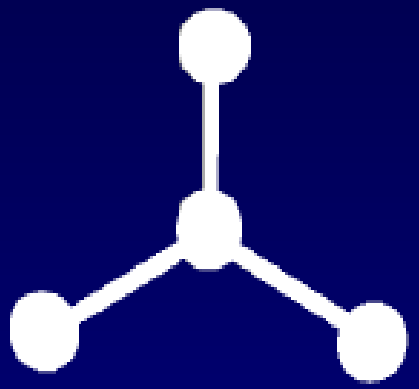
ALA

NME (N-methylamide)

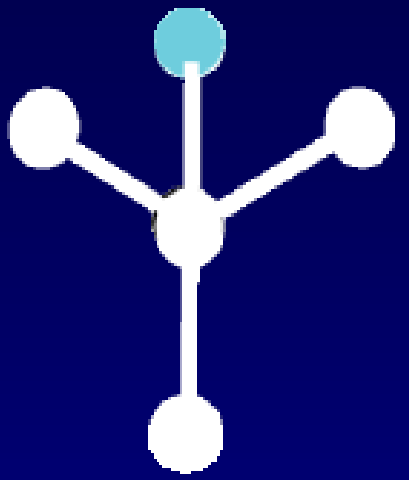

ψ
ϕ
180
160
140
120
100
80
60
40
20
0
-20
-40
-60
-80
-100
-120
-140
-160
-180
-180 -160 -140 -120 -100 -80 -60 -40 -20 0 20 40 60 80 100 120 140 160 180
-24
-26
-28
-30
-32
-34
-36



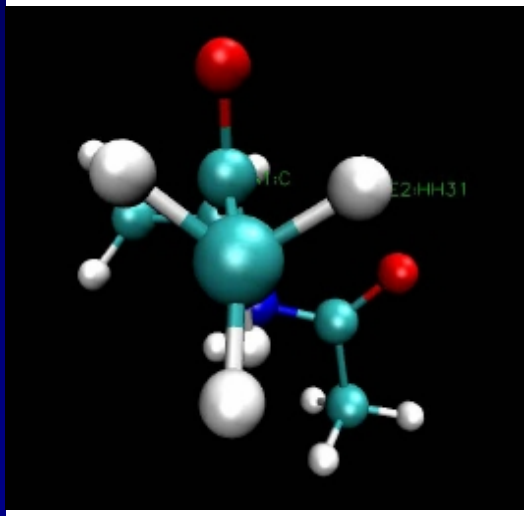
Calculate the ΔG for the dihedral angles of the terminal methyl hydrogens. The two states are eclipsed and staggered as compared to the carbon.



eclipsed



staggered



Eclipsed:

$-30° \leq$ dihedral $< 30°$

$90° \leq$ dihedral $< 150°$

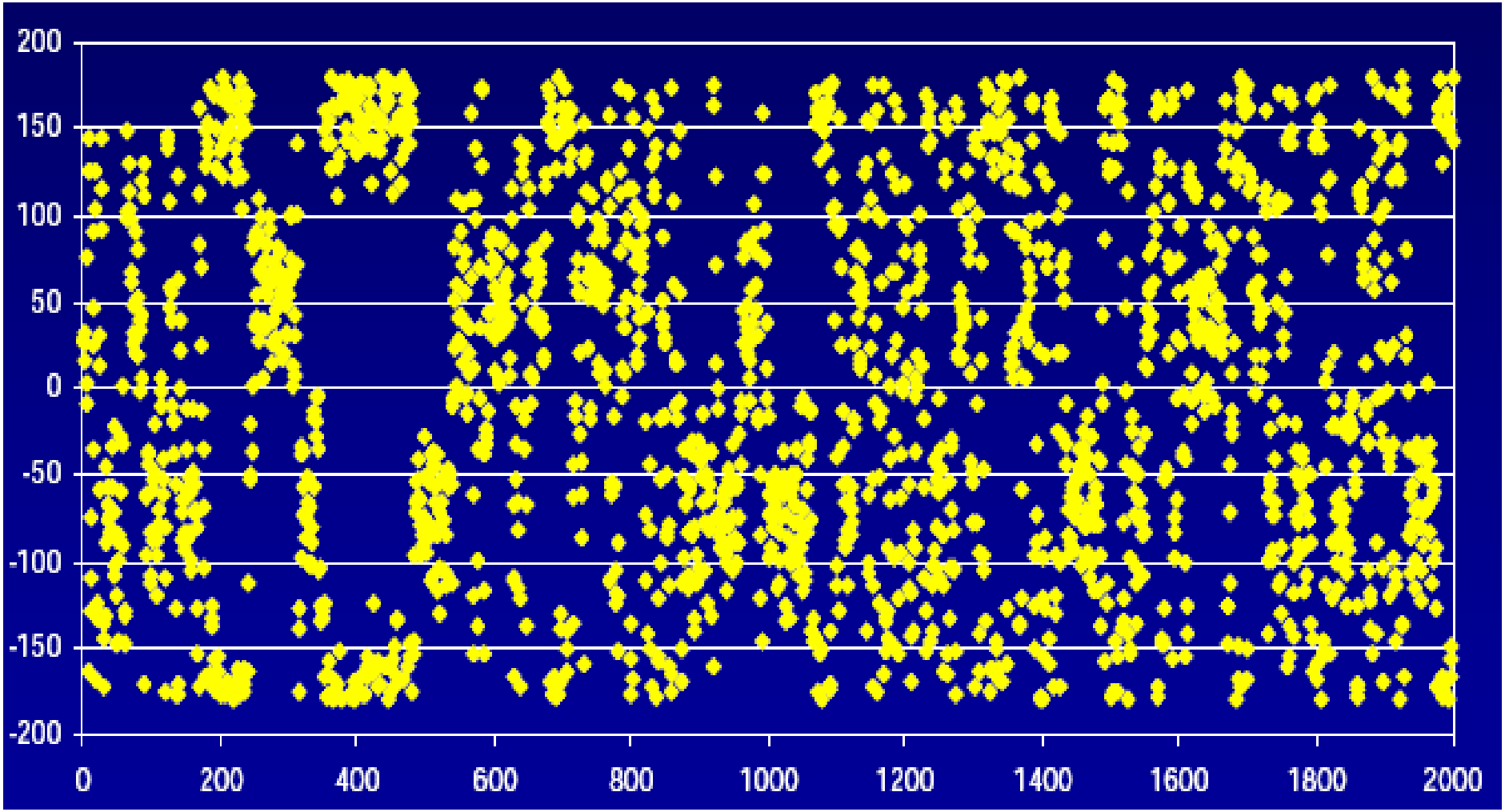
$210° \leq$ dihedral $< 270°$

Staggered:

$30° \leq$ dihedral $< 90°$

$150° \leq$ dihedral $< 210°$

$270° \leq$ dihedral $< 330°$



- Staggered:eclipsed :: 1282:718

$\Delta G = -RT \ln K = -RT \ln (1282/718)$

$= - (1.987 \text{ cal/mol/degree})(310.15 \text{ K}) \ln (1282/718)$

$= -346 \text{ cal/mol} = -0.346 \text{ kcal/mol}$

Challenge

- Accurate calculations of absolute free energy is nearly impossible due to insufficient sampling in a finite length and time scale simulation.
- Need different methods to estimate free energy.
- Common method

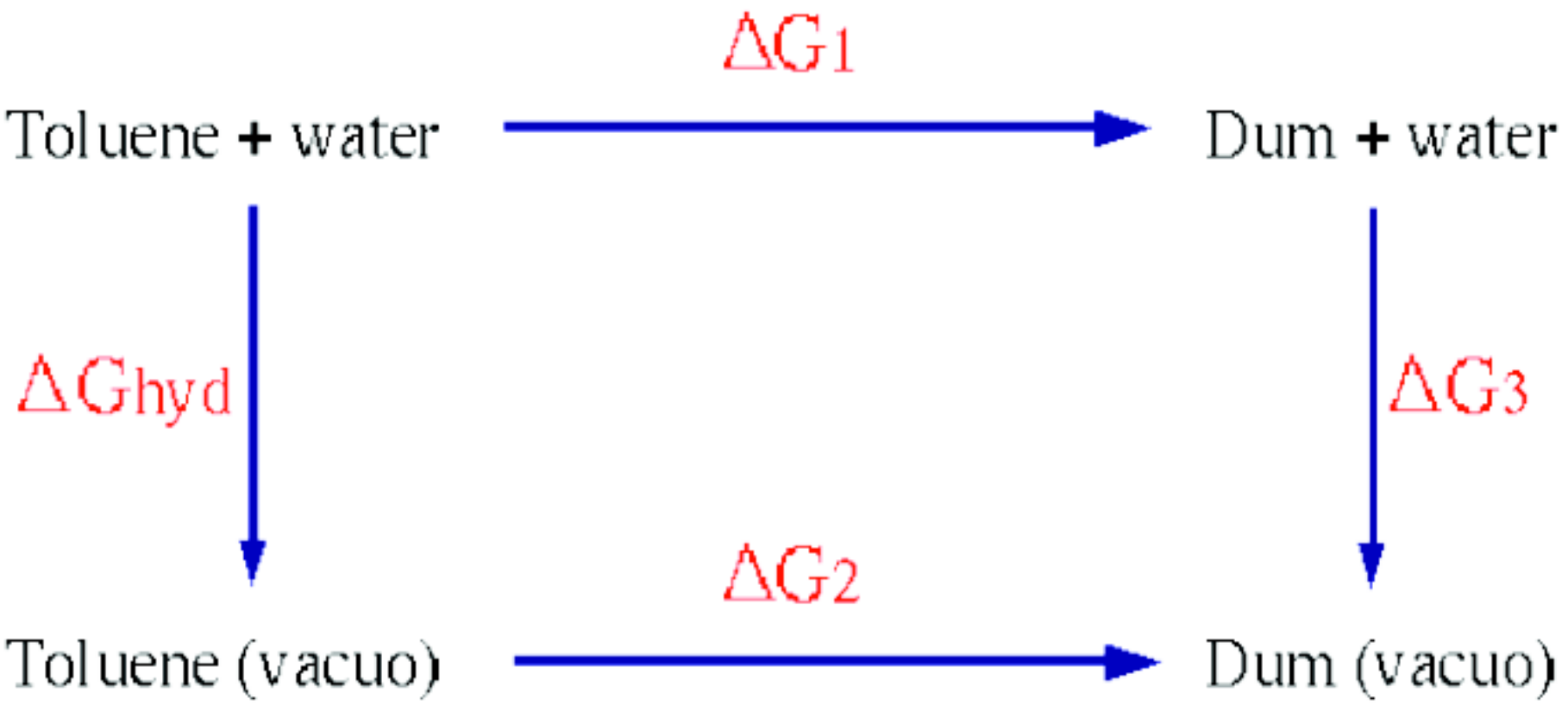
  - Thermodynamical integration

  - Free energy perturbation

  - Umbrella sampling

  - Potential of mean force

Thermodynamic Cycle



$$\Delta G_{hyd} = \Delta G_1 - \Delta G_3 - \Delta G_2 = \Delta G_1 - \Delta G_2$$

- Make the Hamiltonian a function of a coupling parameter λ

$$H(x, p_x; \lambda_a) = H(x, p_x; \lambda = 0)$$

$$H(x, p_x; \lambda_b) = H(x, p_x; \lambda = 1)$$

$$H(x, p_x, \lambda) = H_0(x, p_x) + \lambda H_b(x, p_x) + (1-\lambda) H_a(x, p_x)$$

Thermodynamic Integration

$$\Delta A_{a\to b} = A(\lambda_b) - A(\lambda_a) = \int_{\lambda_a}^{\lambda_b} \frac{dA(\lambda)}{d\lambda} d\lambda$$

$$\frac{dA(\lambda)}{d\lambda} = \frac{\int \frac{\partial H(x, p_x; \lambda)}{d\lambda} \exp - \frac{1}{k_B T} H(x, p_x; \lambda) dx dp_x}{\int \exp - \frac{1}{k_B T} H(x, p_x; \lambda) dx dp_x}$$

$$\Delta A_{a\to b} = \int_{\lambda_a}^{\lambda_b} \left\langle \frac{\partial H(x, p_x; \lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

Thermodynamic Integration

- the value of <dA/dλ> is accurately determined for a number of intermediate values of λ, the total free energy is determined with numerical integration methods based on these values

$$\Delta A_{a\to b} = \int_{\lambda_a}^{\lambda_b} \left\langle \frac{\partial H(x, p_x; \lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

- A(λ) is smooth enough and converged
- Intermediate values:

  fluctuation of ∂H/∂λ for each value of dA/dλ

Free Energy Perturbation

$$\Delta A_{a\to b} = A(\lambda_b) - A(\lambda_a) = -k_B T \ln \frac{Q_{NVT}(\lambda_b)}{Q_{NVT}(\lambda_a)}$$

$$\Delta A_{a\to b} = -k_B T \ln \left\langle \exp\left\{ -\frac{1}{k_B T}\left[H(x, p_x; \lambda_b) - H(x, p_x, \lambda_a) \right] \right\} \right\rangle_{\lambda_a}$$

- The perturbation formula only holds for small changes between the states
- Reaction pathway often broken up into intermediate states, such that the configuration sampled in state A also have a high probability in state B which is the criterion for the ensemble average to converge

$$\Delta A_{a\to b} = -k_B T \sum_{k=1}^{N-1} \ln\left\langle \exp\left\{ -\frac{1}{k_B T}\left[H(x, p_x; \lambda_b) - H(x, p_x, \lambda_a) \right] \right\} \right\rangle_{\lambda_k}$$

Error Estimation

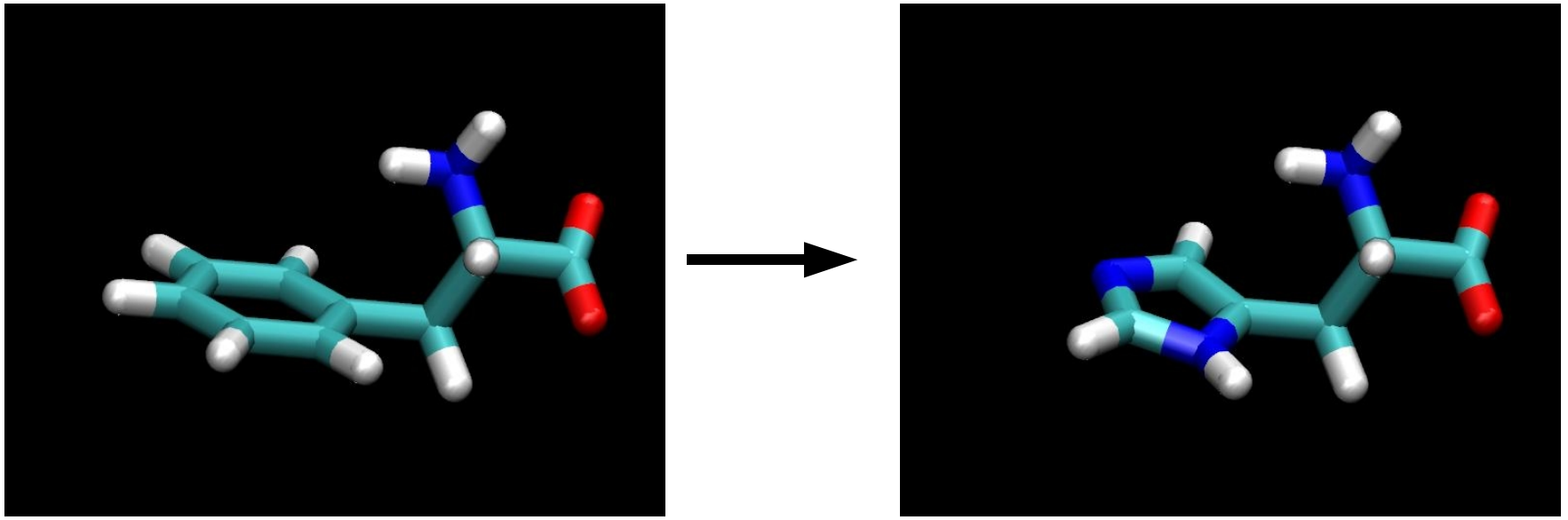
Chris Chipot @ NAMD mailing list

- In FEP, convergence may be probed by monitoring the time-evolution of the ensemble average. This is, however, a necessary, but not sufficient condition for convergence, because apparent plateaus of the ensemble average often conceal anomalous overlap of the density of states characterizing the initial and the final states. The latter should be the key-criterion to ascertain the local convergence of the simulation for those degrees of freedom that are effectively sampled.
- Statistical errors in FEP calculations may be estimated by means of a first-order expansion of the free energy, which involves an estimation of the sampling ratio of the latter of the calculation (Straatsma, 1986).

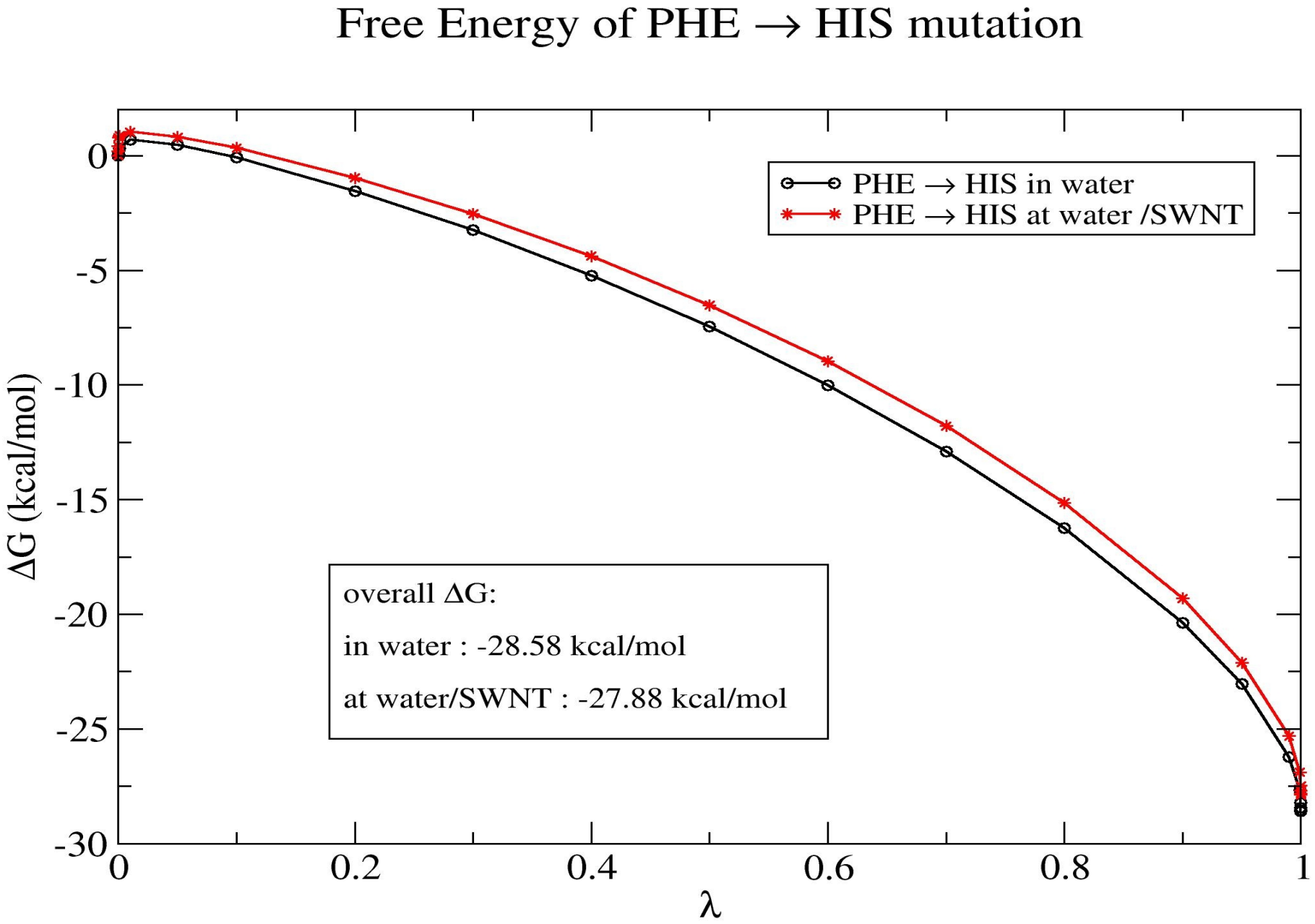
http://www.cirm.univ-rs.fr/videos/2006/exposes/02_LeBris/Chipot.pdf

Example

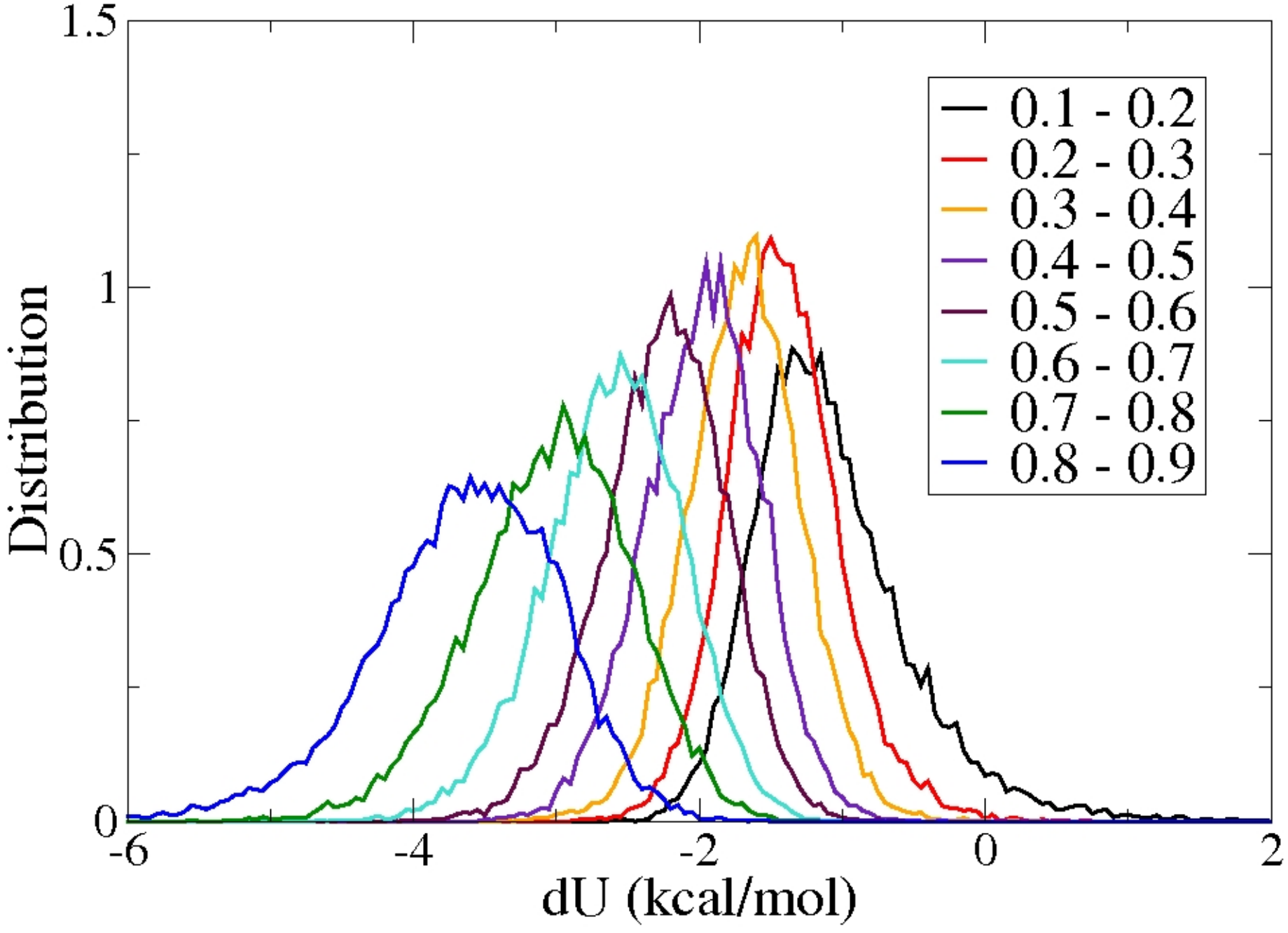
- Mutate PHE to HIS



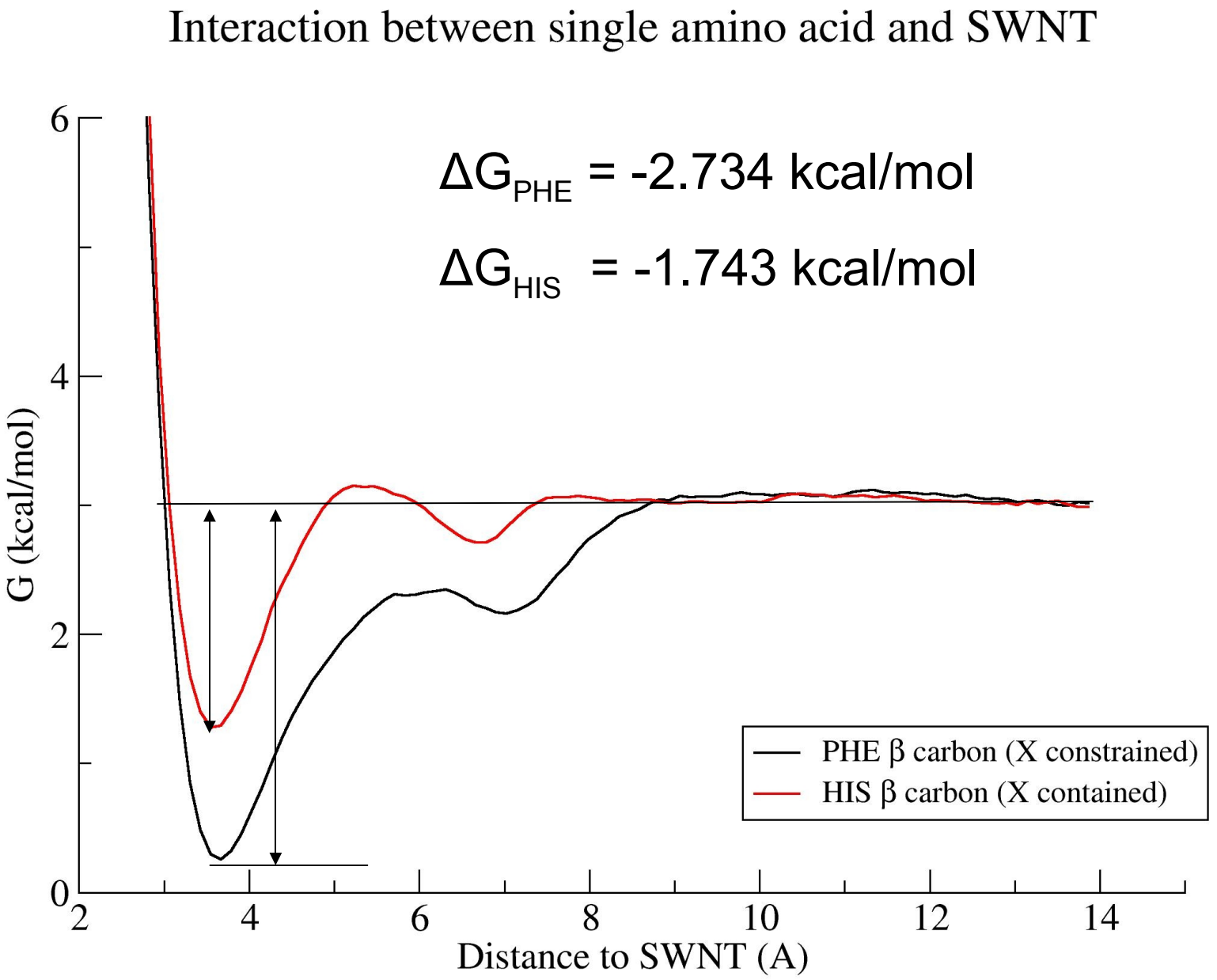
- Phenylalanine → Histidine

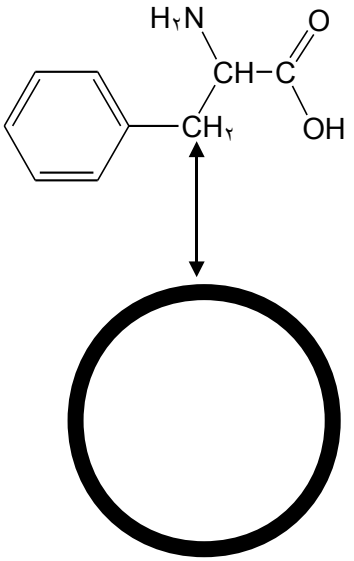


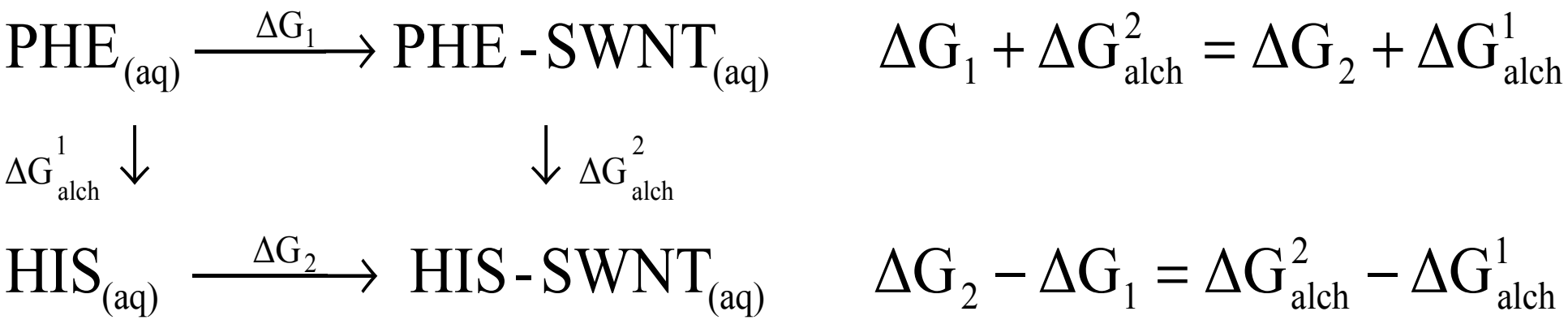
F to H SWNT dU


1.5
1
0.5
0
Distribution
-6
-4
-2
0
2
dU (kcal/mol)
0.1 - 0.2
0.2 - 0.3
0.3 - 0.4
0.4 - 0.5
0.5 - 0.6
0.6 - 0.7
0.7 - 0.8
0.8 - 0.9

Phenylalanine vs. Histidine







- From amino acid – SWNT free energy profile, we can estimate ΔG_1 and ΔG_2 :

  ΔG_1 = -2.734 kcal/mol ΔG_2 = -1.743 kcal/mol

  $\Delta\Delta G = \Delta G_2 - \Delta G_1$ = ***+0.991 kcal/mol***

- From alchemical transformation:

  ΔG^1_{alch} = -28.58 kcal/mol ΔG^2_{alch} = -27.88 kcal/mol

  $\Delta\Delta G' = \Delta G^2_{alch} - \Delta G^1_{alch}$ = ***+0.7 kcal/mol***

- To be continued.....

Thermodynamic Integration

$$H(x, p_x; \lambda_a) = H(x, p_x; \lambda = 0)$$

$$H(x, p_x; \lambda_b) = H(x, p_x; \lambda = 1)$$

$$\Delta A_{a \to b} = A(\lambda_b) - A(\lambda_a) = \int_{\lambda_a}^{\lambda_b} \frac{dA(\lambda)}{d\lambda} d\lambda$$

$$\frac{dA(\lambda)}{d\lambda} = \frac{\int \frac{\partial H(x, p_x; \lambda)}{d\lambda} \exp - \frac{1}{k_B T} H(x, p_x; \lambda) dx dp_x}{\int \exp - \frac{1}{k_B T} H(x, p_x; \lambda) dx dp_x}$$

$$\Delta A_{a \to b} = \int_{\lambda_a}^{\lambda_b} \left\langle \frac{\partial H(x, p_x; \lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

$$H(p_1, p_2, \ldots\ldots p_N, r_1, r_2, \ldots\ldots r_N)$$

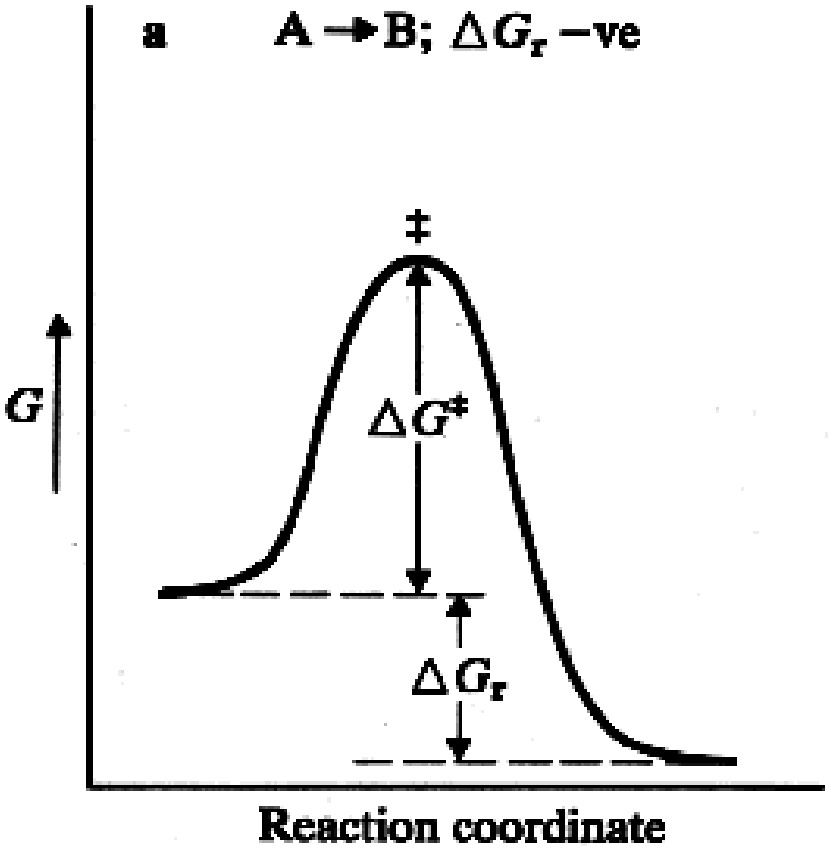
$$\Delta A_{a\to b} = A(\xi_b) - A(\xi_a) = \int \frac{dA(\xi)}{d\xi} d\xi$$

$$\frac{dA(\xi)}{d\xi} = \frac{\int \frac{\partial H(r,p)}{\partial \xi} \exp -\beta H(r,p)\, dr\, dp}{\int \exp -\beta H(r,p)\, dr\, dp}$$

$$\Delta A_{a\to b} = \int \left\langle \frac{\partial H(r,p)}{\partial \xi} \right\rangle d\xi$$

$$\left\langle \frac{\partial H(r,p)}{\partial \xi} \right\rangle = -\langle F_\xi \rangle = \textit{constraint force}$$

Reaction Coordinate ξ

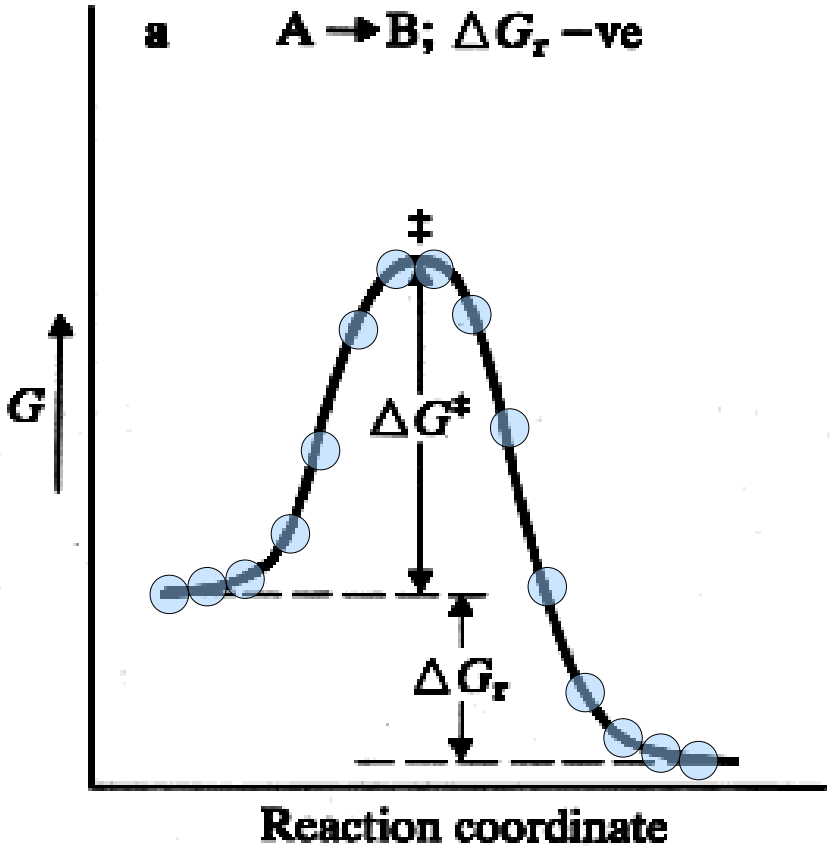


Potential of mean force

- According to the concept of PMF, if a force depending on some reaction coordinate can be extracted, then

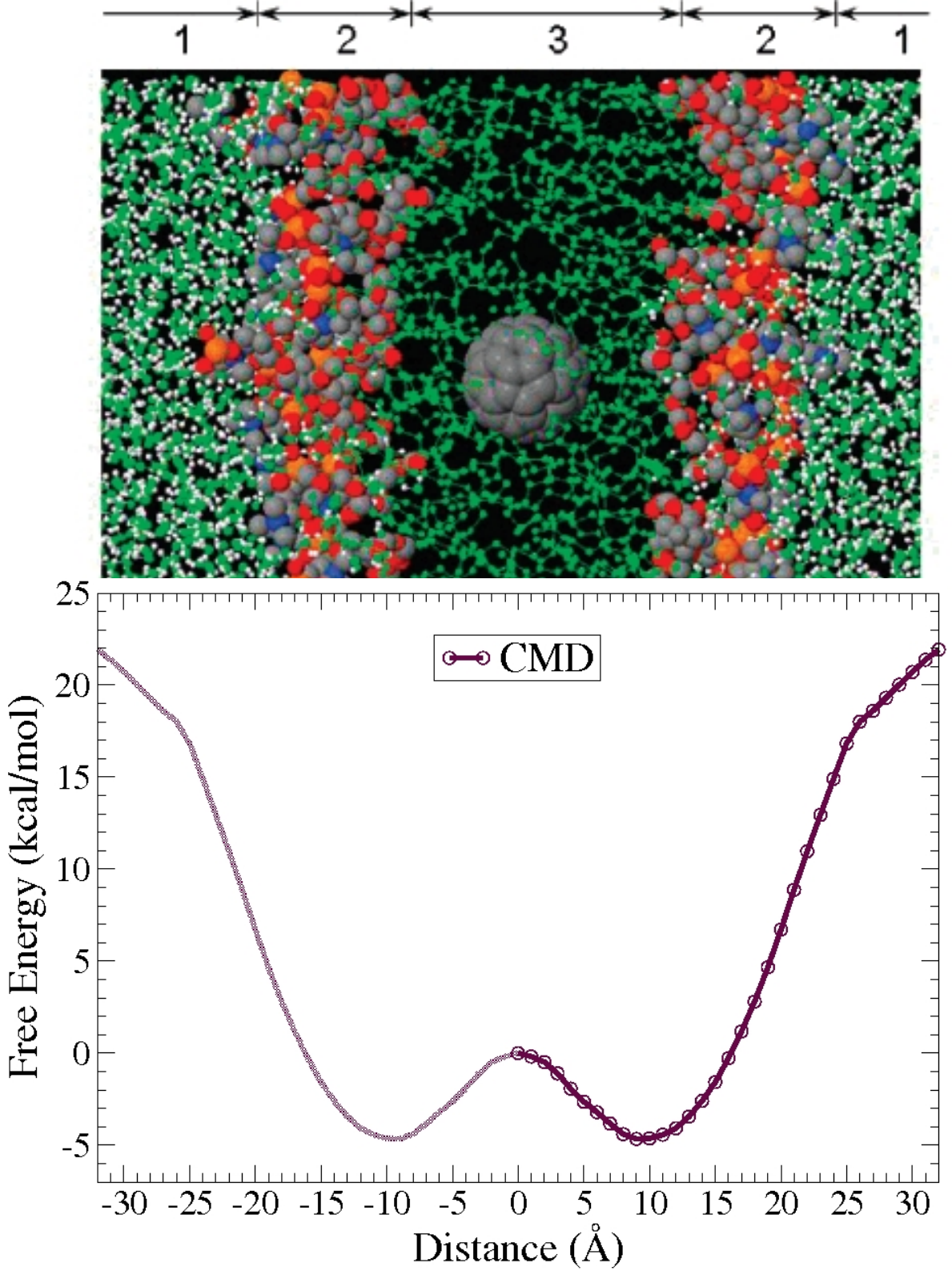
$$\frac{\partial}{\partial \xi}\Delta A_{a\to b} = -\left\langle F_{\xi} \right\rangle_{\xi}$$

- Constraint force : force required to constrain the system at a fix reaction coordinate ξ



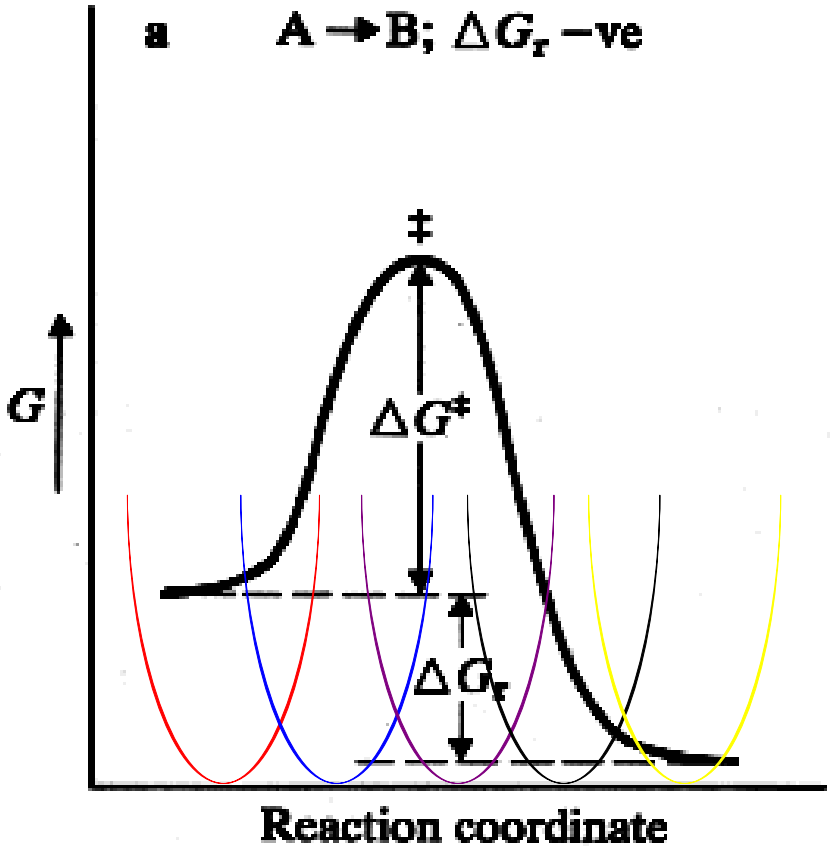
Example

- Translocation of C60 into DPPC bilayer
- Reaction coordinate : Distance between the COM of C60 and COM of DPPC membrane
- 33 constrained points : 0 – 32 Å, 1 Å spacing
- Harmonic constraint force
- k = 100 kcal/mol/$Å^2$


1
2
3
2
1
CMD
Free Energy (kcal/mol)
25
20
15
10
5
0
-5
-30
-25
-20
-15
-10
-5
0
5
10
15
20
25
30
Distance (Å)

Umbrella Sampling

- Sample with umbrella potential U'(x)
- Compute biased probability P'(x)
- Estimate unbiased free energy
- $A(x) = -k_B T \ln P'(x) - U'(x) + F$
- F is undetermined but irrelevar

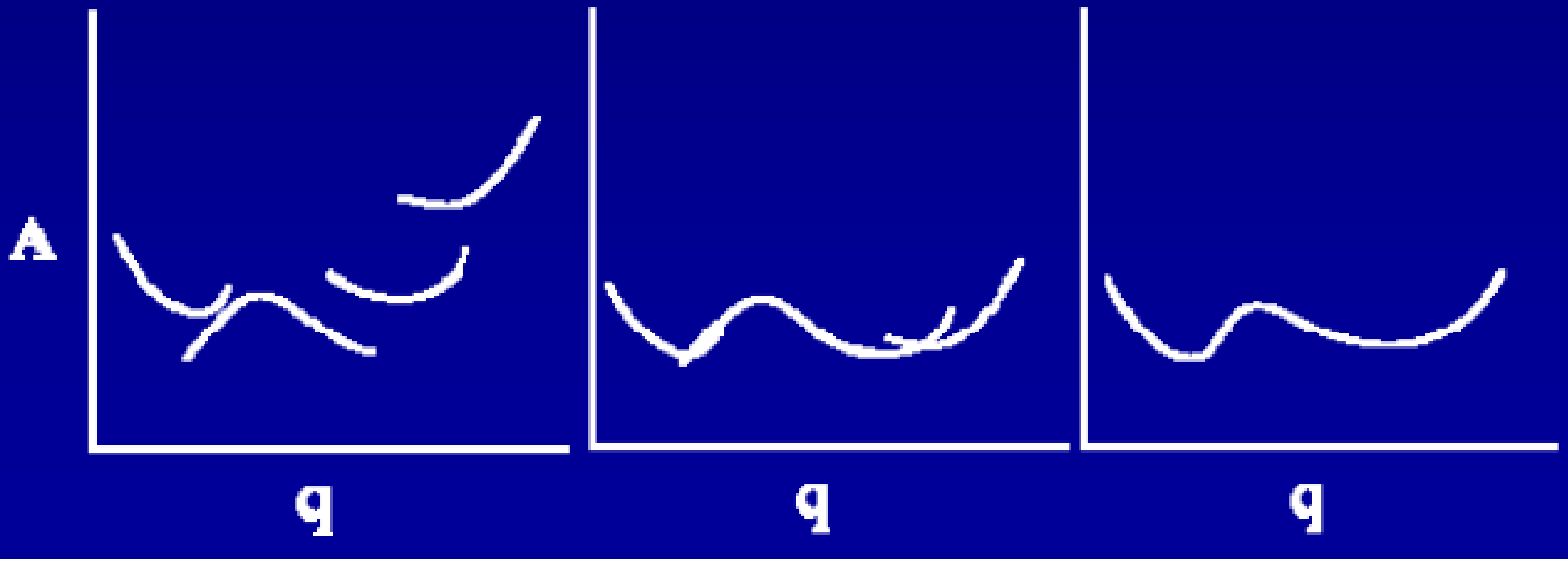


- Complex surfaces have multiple barriers
- Need to know the free energy surface to know an efficient bias
- Harmonic biasing function
- Multiple simulations
- Put the minimum of the bias in a different place for each simulation (sampling windows)
- Estimate P'(x) for each simulation
- Combine results from all simulations
- From one simulation

  $$A(x) = -k_B T \ln P'(x) - U'(x) + F$$

- F depends on U'(x)
- Different simulations have different offsets
- Not obvious how to combine different windows

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



WHAM

- Weighted Histogram Analysis Method
- Determines optimal F values for combining simulations
- Kumar, et al. J Comput Chem, 13, 1011-1021, 1992
- Generalizations
  - Multidimension reaction coordinates
  - Multiple temperatures

$$P(x) = \frac{\sum_{i=1}^{N_{sims}} n_i(x)}{\sum_{i=1}^{N_{sims}} N_i \exp([F_i - U_{bias,i}(x)]/k_B T)}$$

$$F_i = -k_B T \ln\left\{\sum_{x_{bins}} P(x) \exp[-U_{bias,i}(x)/k_B T]\right\}$$

- N_{sims} = number of simulations
- $n_i(x)$= number of counts in histogram bin associated with x
- $U_{bias,i}$, F_i = biasing potential and free energy shift from simulation i
- P(x) = best estimate of unbiased probability distribution
- F_i and P(x) are unknowns
- Solve by iteration to self consistency

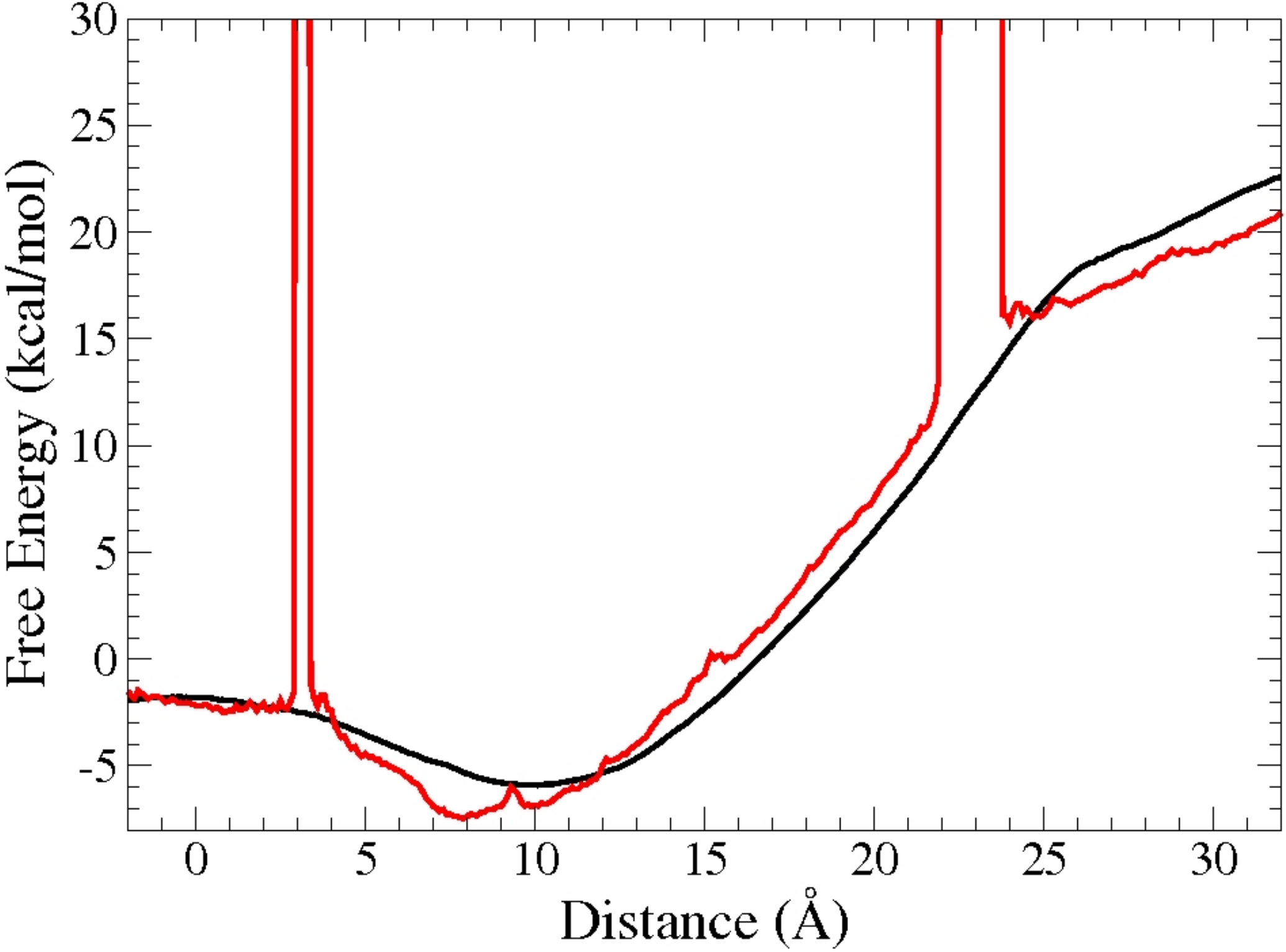
Running US simulation

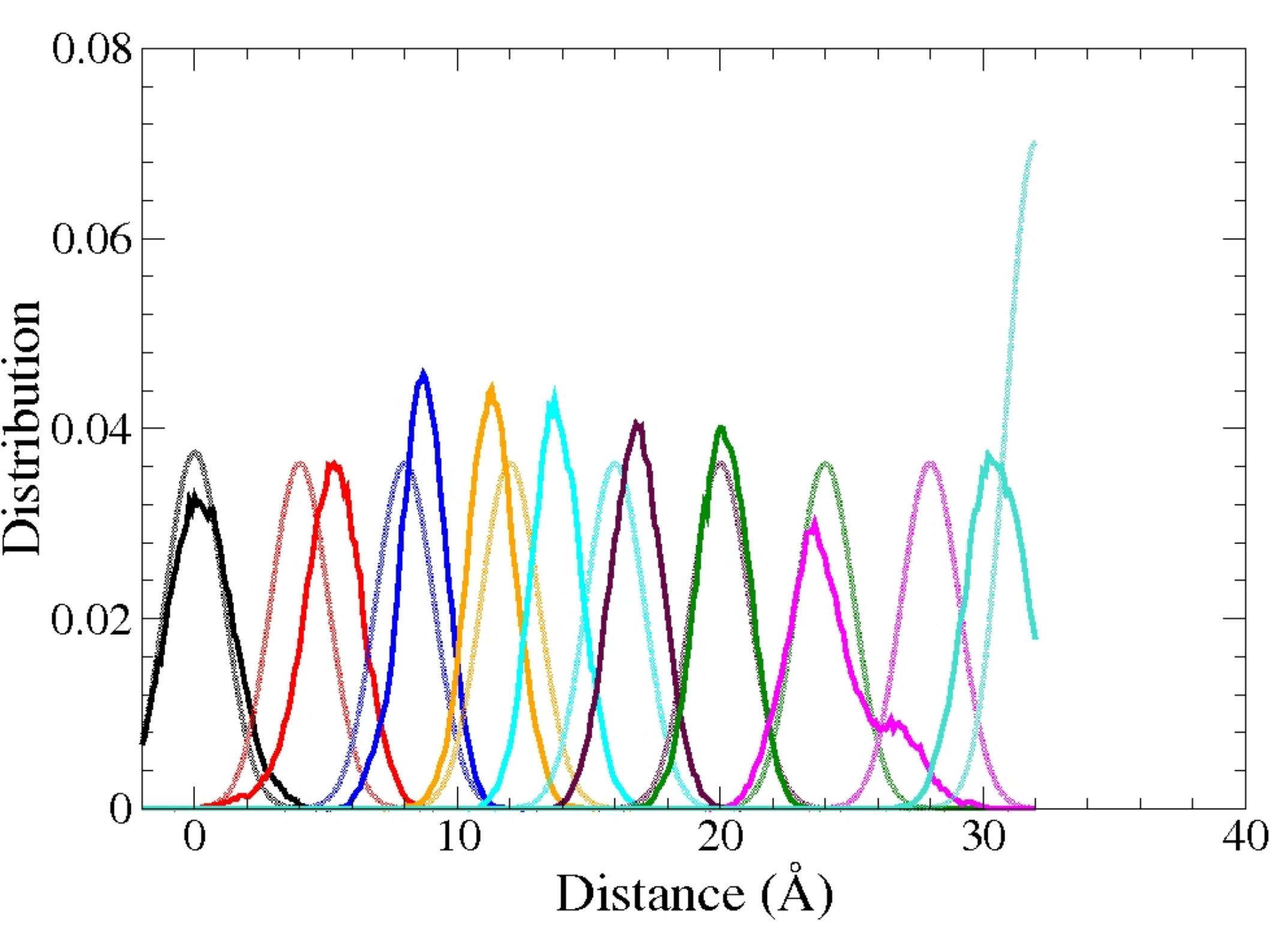
- Choose the reaction coordinate
- Choose the number of windows and the biasing potential
- Run the simulations
- Compute time series for the value of the reaction coordinate (histograms)
- Apply the WHAM equations

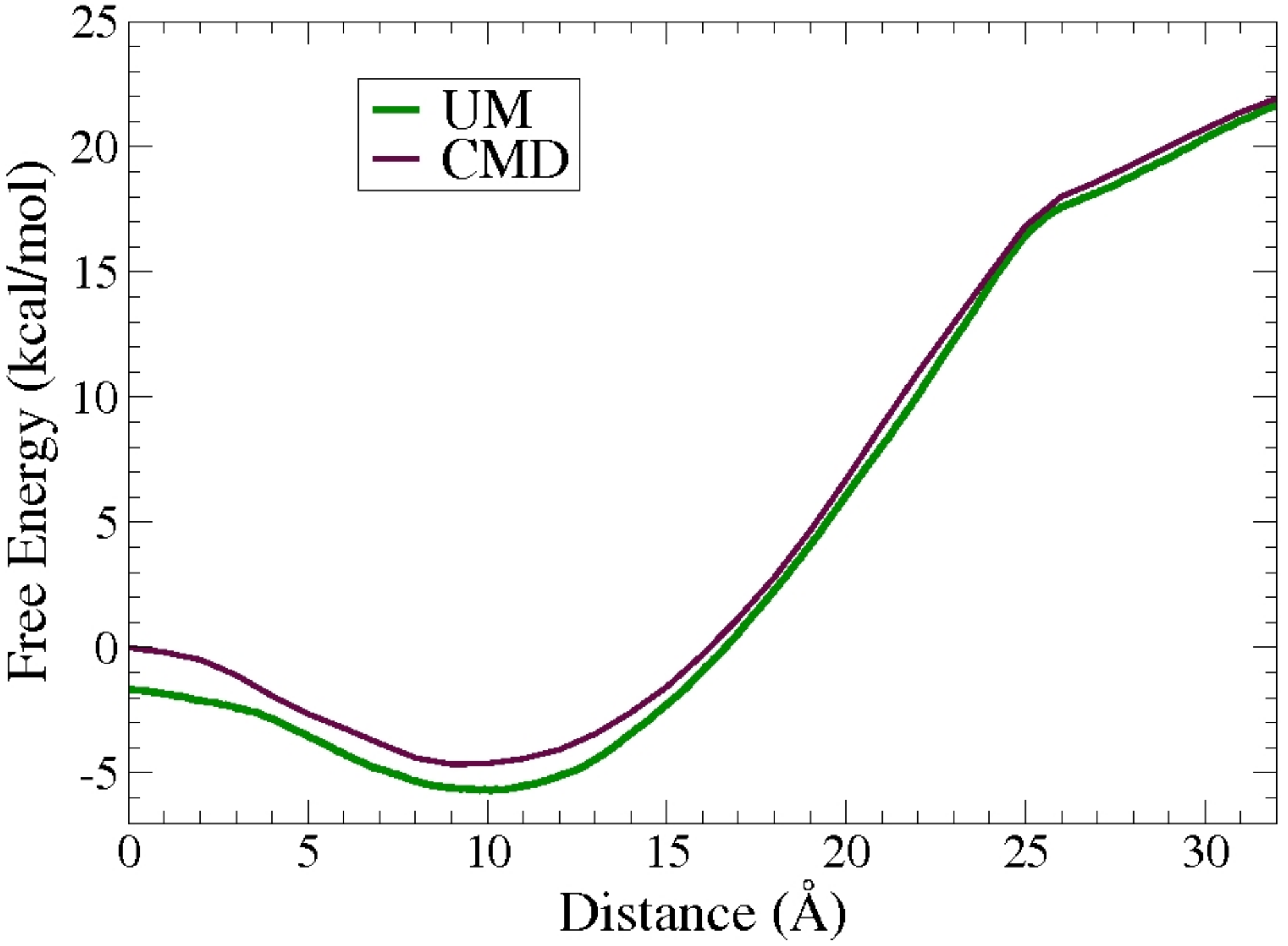
- Convergence is probed by two criteria:
  - Convergence of individual windows. The statistical error can be measured through block-averaging over sub-runs
  - Appropriate overlap of free energy profiles between adjacent windows
- The approaches to estimate errors for the different methods based on a single simulation only reflect the statistical precision of the method
- Statistical accuracy can be derived from an ensemble of simulations starting from different regions in phase space

Example

- Translocation of C60 into DPPC bilayer
- Reaction coordinate : Distance between the COM of C60 and COM of DPPC membrane
- 9 windows : 0 – 32 Å, 4 Å spacing
- Harmonic constraint force
- k = 0.5 kcal/mol/Å^2


Free Energy (kcal/mol)
30
25
20
15
10
5
0
-5
0
5
10
15
20
25
30
Distance (Å)


0.08
0.06
0.04
0.02
0
Distribution
0
10
20
30
40
Distance (Å)


UM
CMD
Free Energy (kcal/mol)
Distance (Å)
25
20
15
10
5
0
-5
0
5
10
15
20
25
30

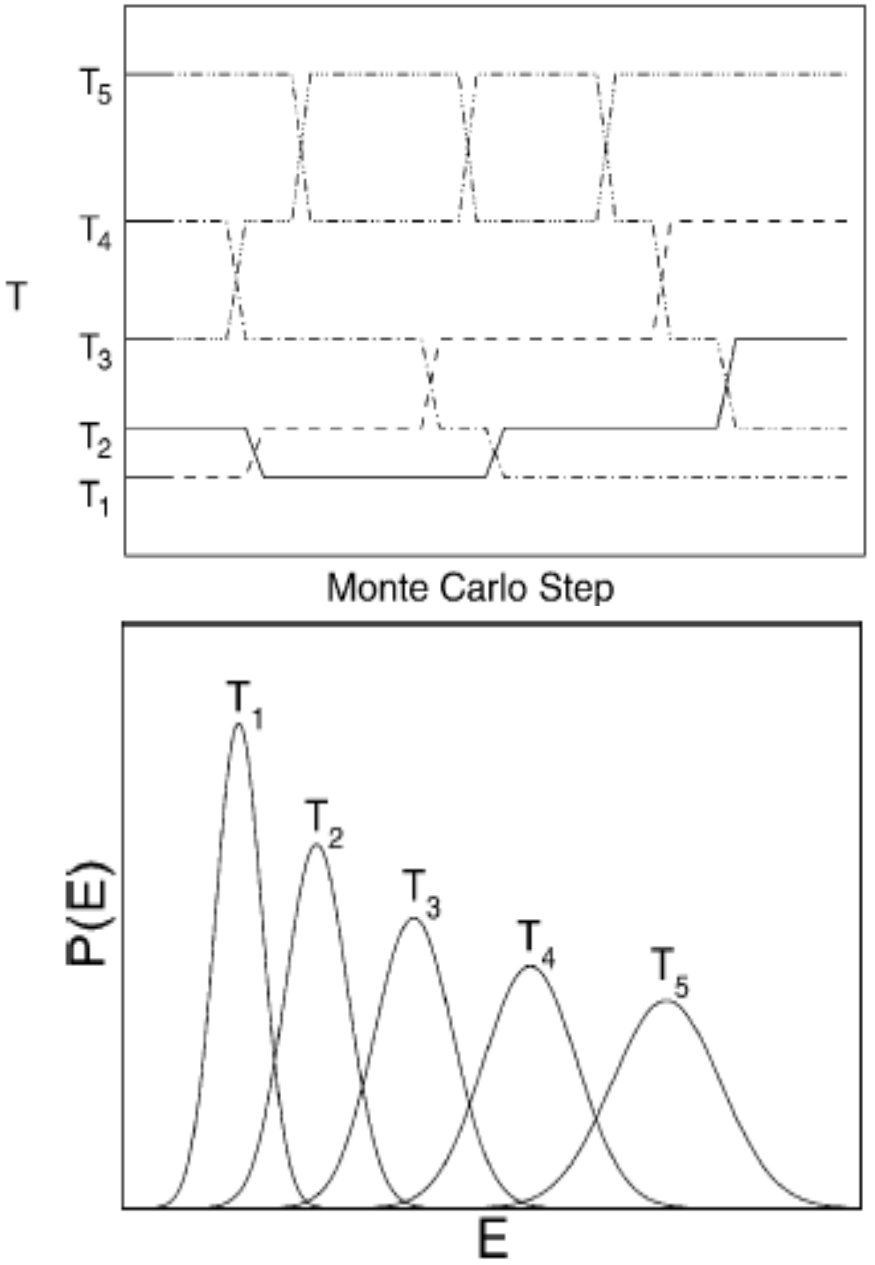
Other Methods

- Steered MD
- Adaptive Biased Force Method (ABF)
- Parallel Tempering (Replica exchange)
  - overcome free energy barriers
- etc.

Replica Exchange Overcoming Free Energy Barrier

- Non-directed method

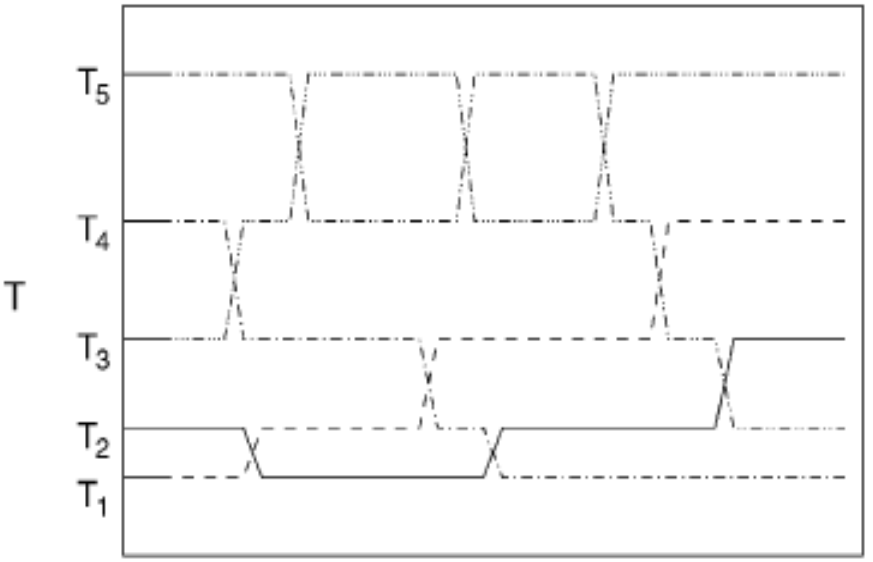
  (no reaction coordinate)
- How to sample unfavorable states?
- At high T, barriers are easier to overcome.
- Heat and cool the system to push it over barriers to sample new configurations

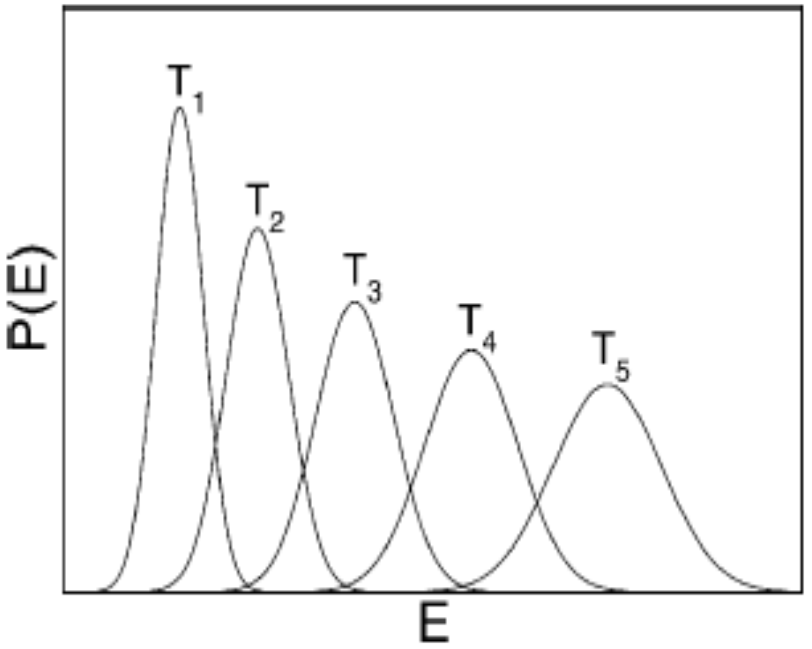


Replica Exchange Overcoming Free Energy Barrier

- Launch simulations at different temperatures
- Swap configurations based on a monte carlo criterion
- This criterion guarantees that the lowest (target) temperature "trajectory" samples from the Boltzmann distribution.
- Swapping configurations effectively improves sampling

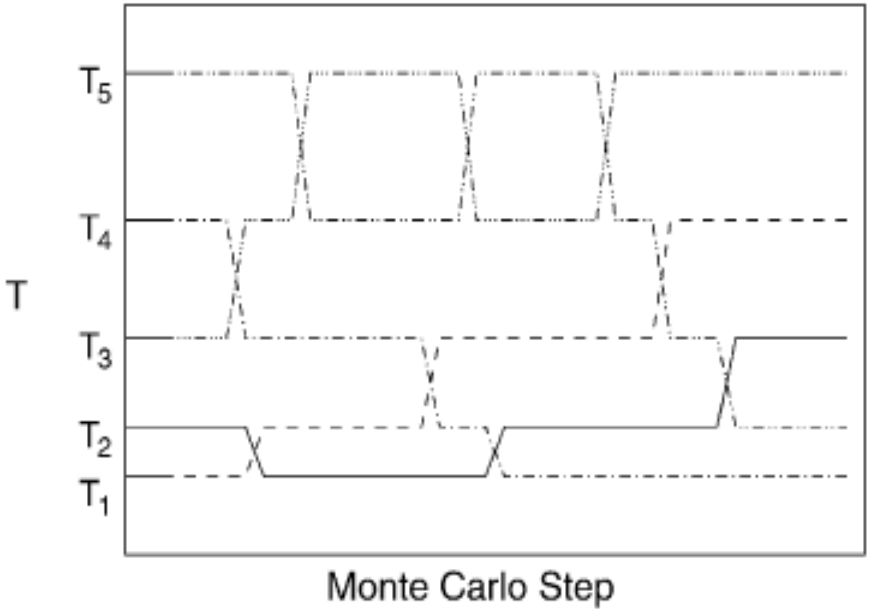
$$A = \min\{1, \exp[+(\beta_i - \beta_j)(U(\mathbf{r}_i^N) - \mathrm{U}(\mathbf{r}_j^N))]\}.$$

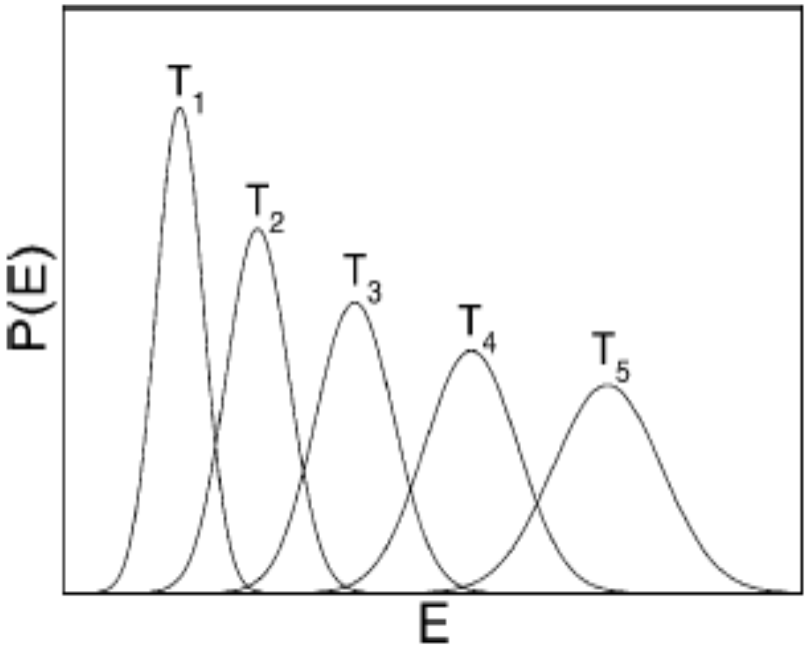




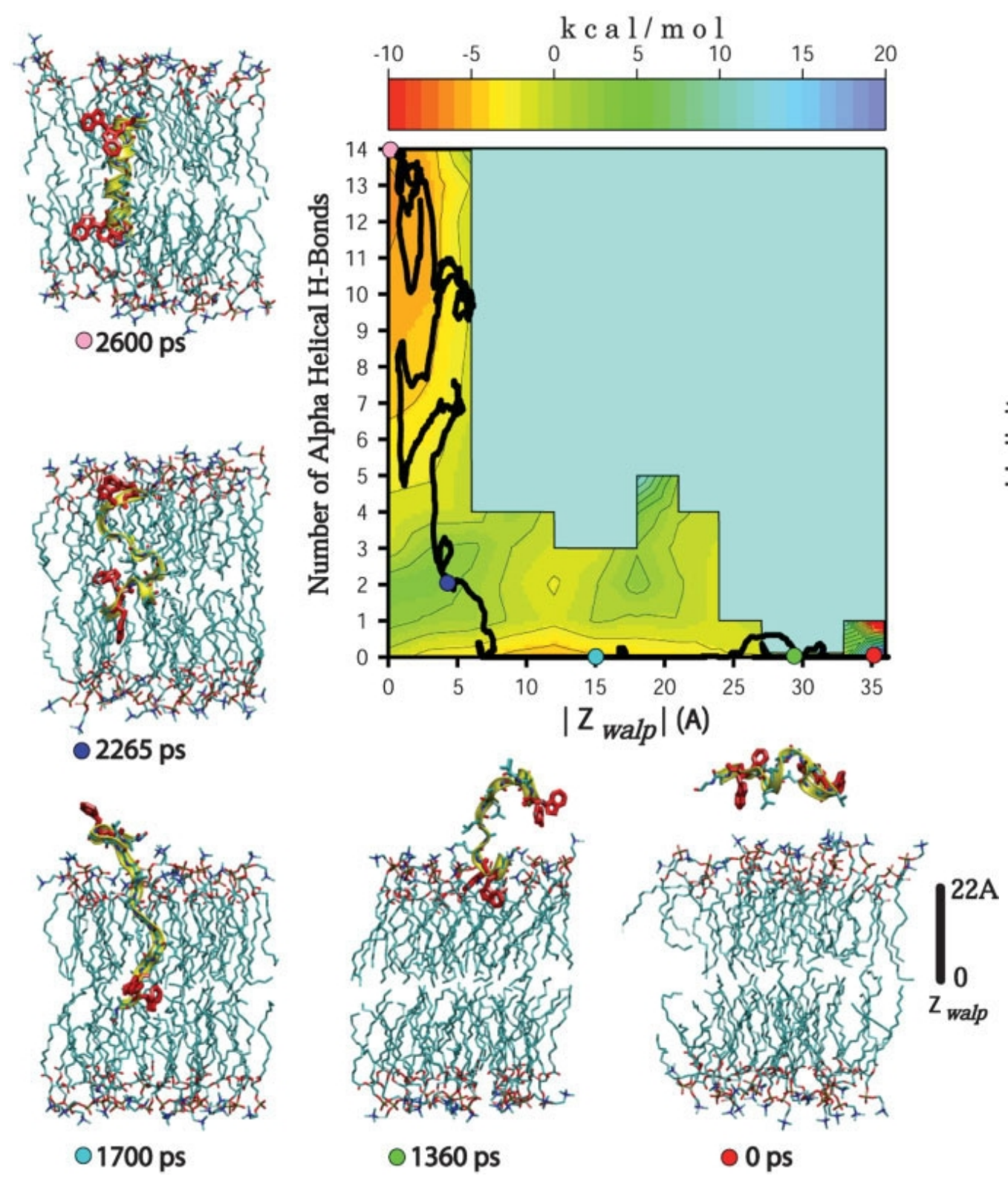
Overcoming Free Energy Barrier

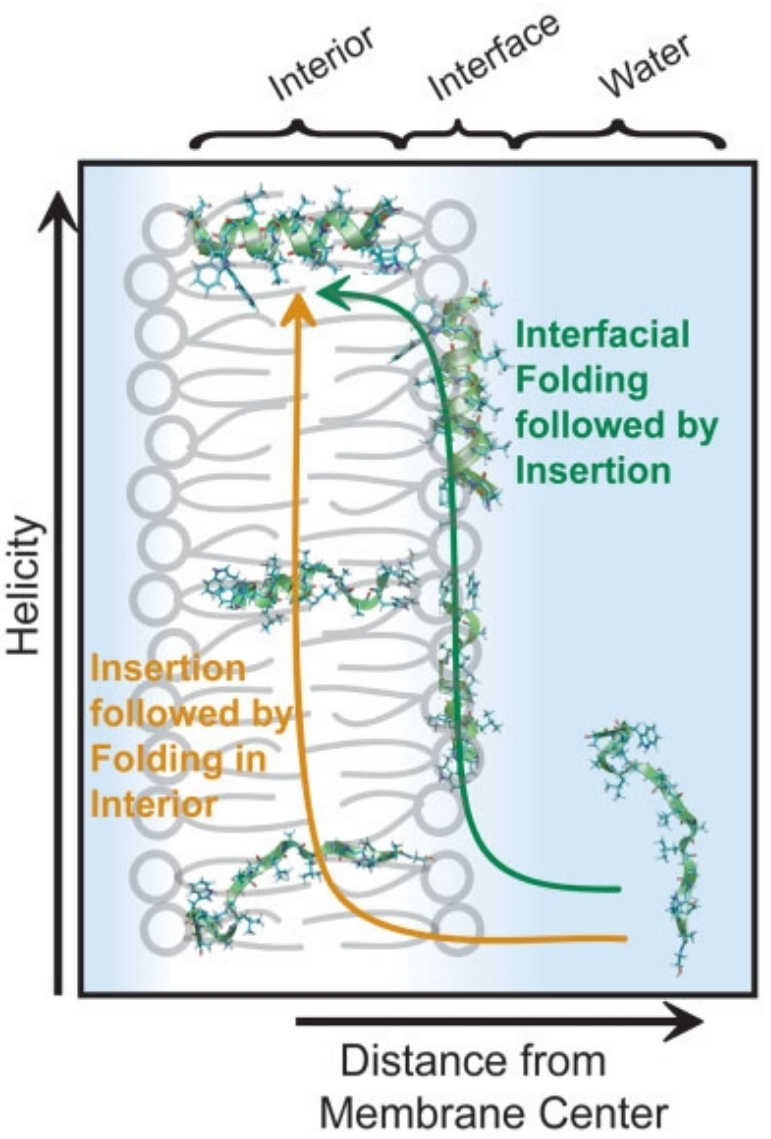
- When doing parallel tempering molecular dynamics, one must take care in the interpretation of the results. A parallel tempering exchange is an ‘unphysical’ move, and so one cannot draw conclusions about dynamics. That is, when using parallel tempering molecular dynamics, one is only really doing a form of sampling and not ‘true’ molecular dynamics.





Deem et al. Phys . Chem. Chem. Phys . 2 0 0 5 , 7 , 3 9 1 0 – 3 9 1 6





Garcia et al.
PROTEINS: Structure, Function, and Bioinformatics
59:783–790 (2005)

Reference

- Anna Johansson
  xray.bmc.uu.se/~calle/md_phd/free_energ.pdf
- David Mathews

  http://rna.urmc.rochester.edu/teaching.html
- Chris Chipot

  http://www.cirm.univ-rs.fr/videos/2006/exposes/02_LeBris/Chipot.pdf