THE UNIVERSITY of TEXAS

HEALTH SCIENCE CENTER AT HOUSTON

SCHOOL of HEALTH INFORMATION SCIENCES

Free Energy Calculations

For students of HI 6327 “Biomolecular Modeling”

Willy Wriggers, Ph.D.
School of Health Information Sciences

http://biomachina.org/courses/modeling/09.html

Uses of Free Energy

Free energy is one of the most important thermodynamic quantities (reaction equilibrium, solvation, stability, and kinetics).

- Protein-protein and protein-ligand interactions (binding constants, association and disassociation)
- Mutation analysis
- Rational drug design
- Protein folding/unfolding

Difficulty (I)

Thermodynamic properties:

- Mechanical properties: related to the derivative of the partition function Q

Internal energy, pressure, and heat capacity, etc.

- Thermal properties: related to the partition function Q itself

Free energy, chemical potential, and entropy, etc.

Partition function for the canonical ensemble:

$$Q_{NVT} = \frac{1}{N!}\frac{1}{h^{3N}}\iint dp^{N} dr^{N} \exp\left[-\frac{H(p^{N}, r^{N})}{k_{B}T}\right]$$

Difficulty (II)

Internal energy E:

$$E = \frac{k_B T^2}{Q}\frac{\partial Q}{\partial T} = \iint dp^N dr^N H(p^N, r^N)\rho(p^N, r^N), \quad \rho = \frac{\exp(-H(p^N, r^N)/k_B T)}{Q}$$

High energy states have a very low probability and make an insignificant contribution to the integral, so we can get an accurate estimate of E by MD or MC.

Helmholtz free energy F:

$$F = -k_B T \ln Q = k_B T \ln\left(\iint dp^N dr^N \exp\left(+\frac{H(p^N, r^N)}{k_B T}\right)\rho(p^N, r^N)\right)$$

High energy states make a significant contribution to the integral, so the results for F will be poorly converged (inaccurate) in MD or MC.

Methods and Applications

- Free energy perturbation and thermodynamic integration
- Potential of mean force calculations
- ‘Rapid’ free energy methods

Most of the free-energy methods are based on calculation of free-energy *differences*, which may be the quantity of interest anyway. If reference is simple (such as ideal gas or harmonic crystal), its absolute free energy can be evaluated analytically.

Free Energy Perturbation and Thermodynamic Integration

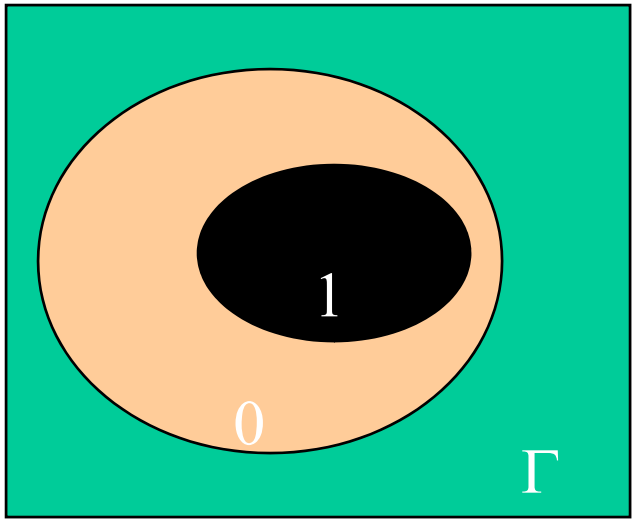
Free-Energy Perturbation

- FEP gives free-energy difference between two states
  - labeled 0, 1
- Working equation

$\Gamma=(p^N,r^N)$ $\beta=1/k_BT$

$$e^{-\beta(F_1-F_0)} = \frac{Q_1}{Q_0} = \frac{\int d\Gamma e^{-\beta H_1}}{\int d\Gamma e^{-\beta H_0}}$$

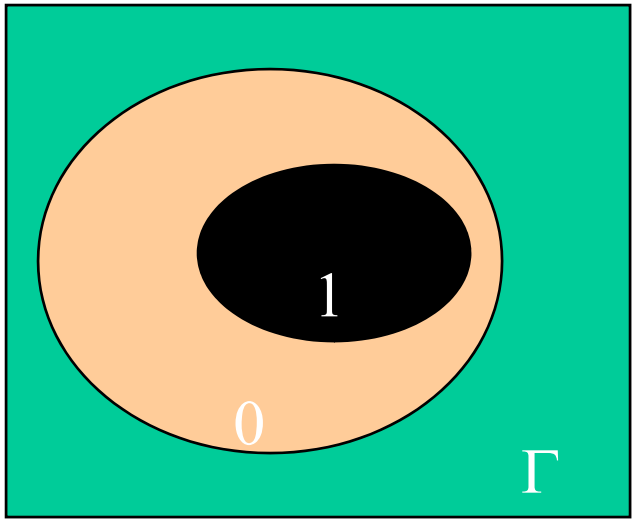
Free-energy difference is a ratio of partition functions



=

Free-Energy Perturbation

- FEP gives free-energy difference between two states
  - labeled 0, 1
- Working equation



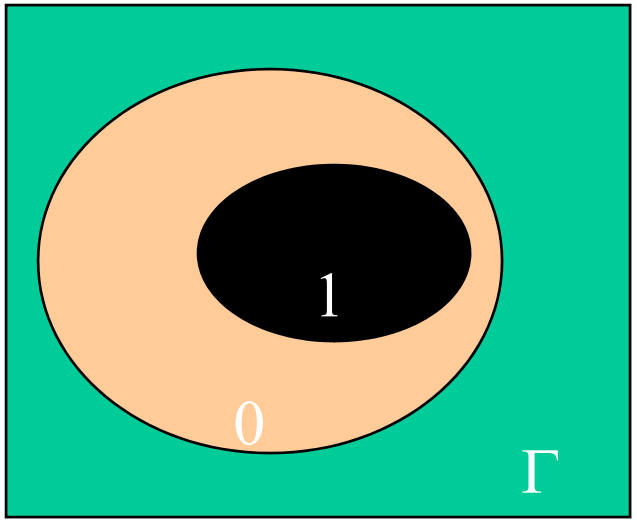
$$e^{-\beta(F_1-F_0)} = \frac{Q_1}{Q_0} = \frac{\int d\Gamma e^{-\beta H_1}}{\int d\Gamma e^{-\beta H_0}}$$

$$= \frac{\int d\Gamma e^{-\beta(H_1-H_0)} e^{-\beta H_0}}{\int d\Gamma e^{-\beta H_0}}$$

Add and subtract reference-state energy

Free-Energy Perturbation

- FEP gives free-energy difference between two states
  - labeled 0, 1
- Working equation



$$e^{-\beta(F_1-F_0)} = \frac{Q_1}{Q_0} = \frac{\int d\Gamma e^{-\beta H_1}}{\int d\Gamma e^{-\beta H_0}}$$

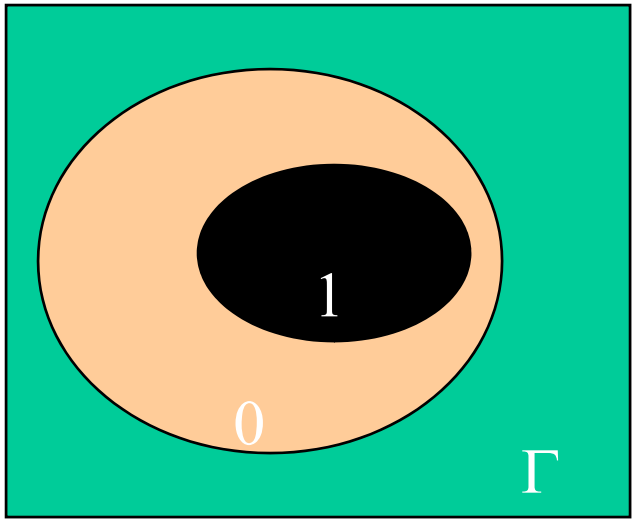
$$= \frac{\int d\Gamma e^{-\beta(H_1-H_0)} e^{-\beta H_0}}{\int d\Gamma e^{-\beta H_0}}$$

$$= \int d\Gamma e^{-\beta(H_1-H_0)} \rho_0(\Gamma)$$

Identify reference-state probability distribution

Free-Energy Perturbation

- FEP gives free-energy difference between two states
  - labeled 0, 1
- Working equation



$$e^{-\beta(F_1-F_0)} = \frac{Q_1}{Q_0} = \frac{\int d\Gamma e^{-\beta H_1}}{\int d\Gamma e^{-\beta H_0}}$$

$$= \frac{\int d\Gamma e^{-\beta(H_1-H_0)} e^{-\beta H_0}}{\int d\Gamma e^{-\beta H_0}}$$

$$= \int d\Gamma e^{-\beta(H_1-H_0)} \rho_0(\Gamma)$$

$$= \left\langle e^{-\beta(H_1-H_0)} \right\rangle_0$$

Write as reference-state ensemble average

- Sample the region important to 0 state, measure properties of 1 state

Free-Energy Perturbation

- The FEP formula may be used also with the roles of the reference and target state reversed

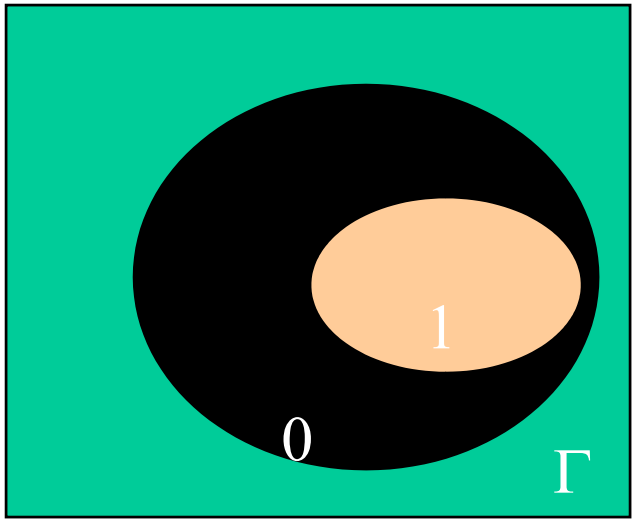
Original: $0 \rightarrow 1$ (insertion)

$$e^{-\beta(F_1-F_0)} = \left\langle e^{-\beta(H_1-H_0)} \right\rangle_0$$

Modified: $1 \rightarrow 0$ (deletion)

$$e^{+\beta(F_1-F_0)} = \left\langle e^{+\beta(H_1-H_0)} \right\rangle_1$$

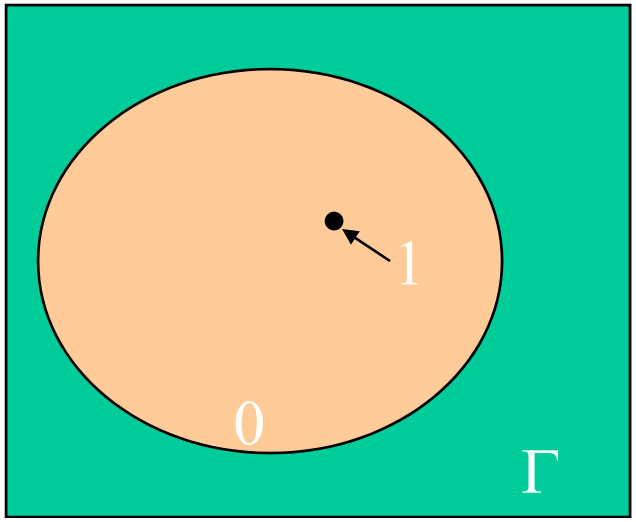
  - sample the 1 state, evaluate properties of 0 state



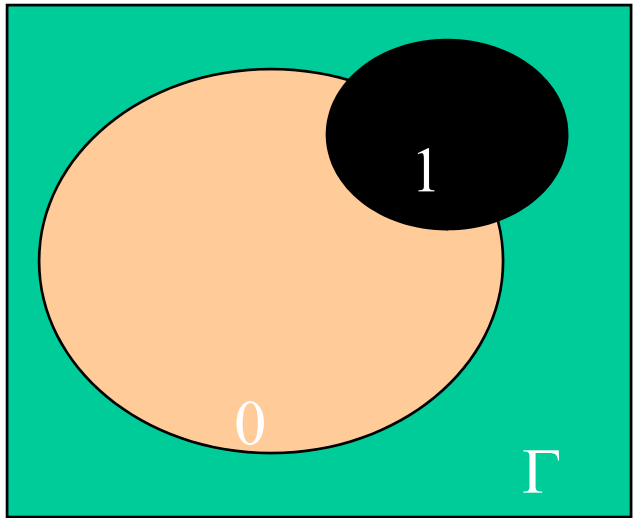
General Numerical Problems

- Sampling problems limit range of FEP calculations
- Target state configurations must be encountered when sampling reference state
- Two types of problem arise

target-state space very small



target state outside of reference

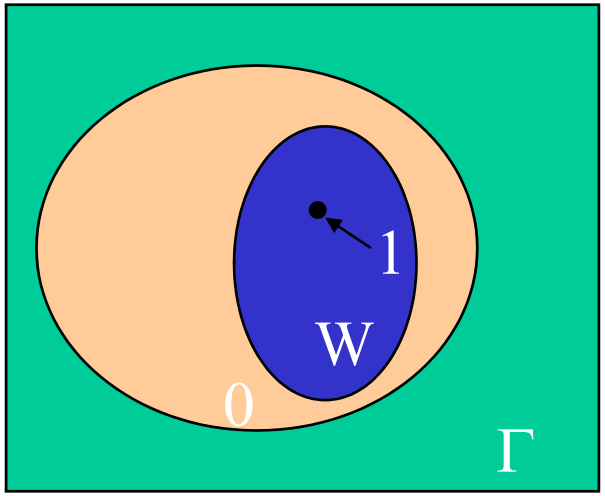


Staging Methods

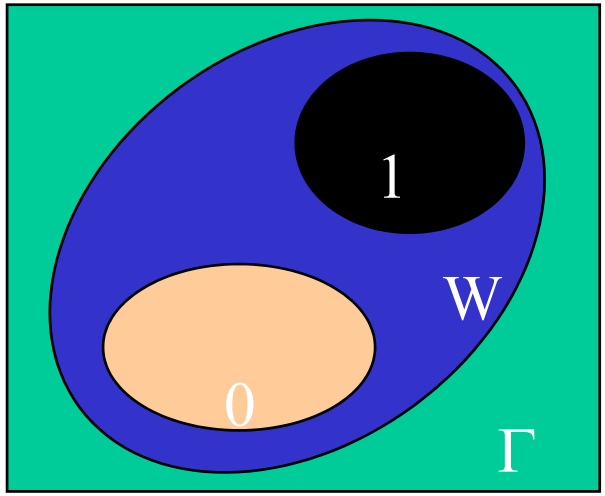
- Multistage FEP can be used to remedy the sampling problem
  - define a potential H_w intermediate between 0 and 1 states
  - evaluate total free-energy difference as $F_1 - F_0 = (F_1 - F_w) + (F_w - F_0)$
- Each stage may be sampled in either direction
  - yielding four staging schemes
  - choose to avoid deletion calculation

$0 \leftarrow W \rightarrow 1$	Umbrella sampling
$0 \rightarrow W \leftarrow 1$	Bennett's method
$0 \leftarrow W \leftarrow 1$	Staged deletion
$0 \rightarrow W \rightarrow 1$	Staged insertion

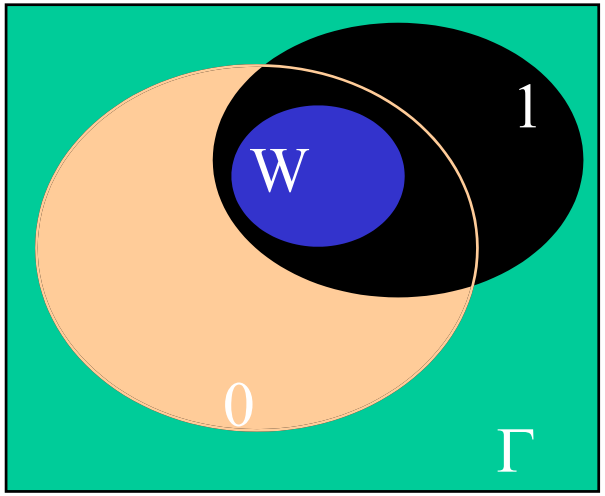
Use staged insertion



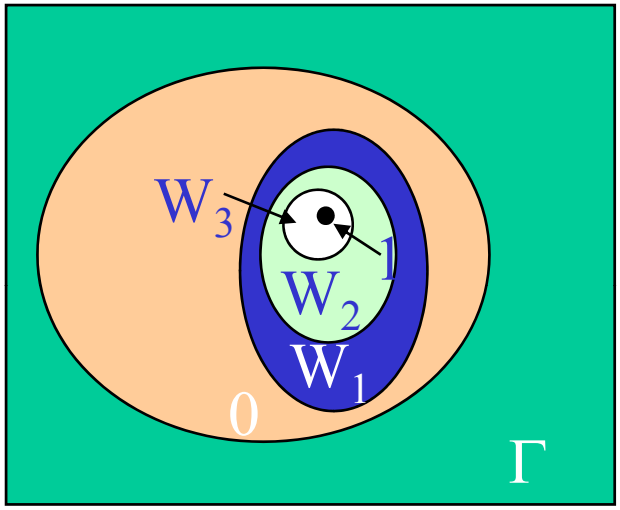
Use umbrella sampling



Use Bennett's method

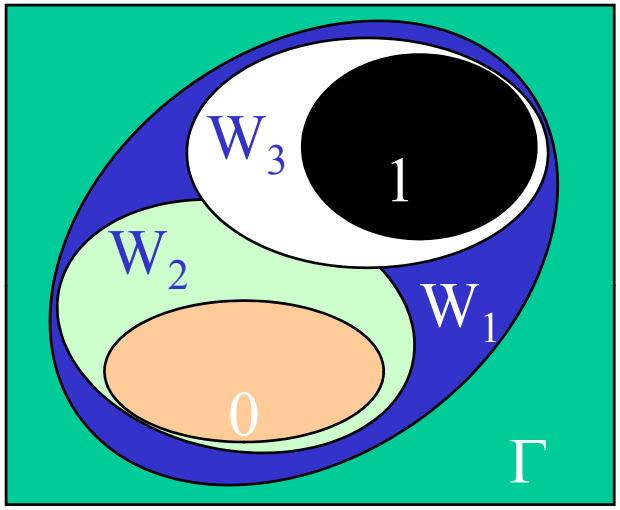


Multiple Stages



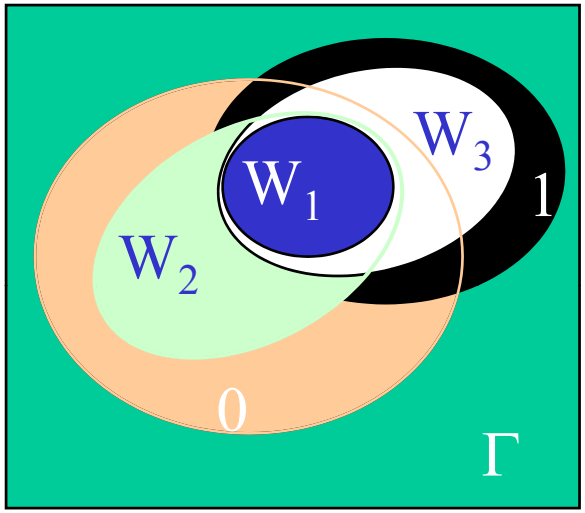
Multistage insertion

$$0 \rightarrow W_1 \rightarrow W_2 \rightarrow W_3 \rightarrow 1$$



Multistage umbrella sampling

$$0 \leftarrow W_2 \leftarrow W_1 \rightarrow W_3 \rightarrow 1$$



Multistage Bennett's method

$$0 \rightarrow W_2 \rightarrow W_1 \leftarrow W_3 \leftarrow 1$$

Thermodynamic Integration & Slow Growth

- Thermodynamic Integration

$$F(\lambda) = -k_B T \ln Q(\lambda)$$

$$\Delta F = \int_0^1 \frac{\partial F(\lambda)}{\partial(\lambda)} d\lambda$$

$$\Delta F = \int_0^1 \left\langle \frac{\partial H(p^N, r^N, \lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda$$

F is a continuous function of λ

- Slow Growth Method

From FEP expression:

$$\Delta F = -k_B T \sum_{i=0}^{N_{step}-1} \ln \left\langle \exp\left(-\left[H(\lambda_{i+1}) - H(\lambda_i)\right] / k_B T\right) \right\rangle_{NVT}$$

Taylor expansion

$$\Delta F \approx \sum_{i=0}^{N_{step}-1} \left\langle \left[H(\lambda_{i+1}) - H(\lambda_i)\right] \right\rangle_{NVT}$$

Relationship between TI and FEP

If the free energy is expressed as a Taylor series expansion in terms of λ at λ=0:

$$F(\lambda) = F(0) + F'_{\lambda=0}\lambda + \frac{1}{2!}F''_{\lambda=0}\lambda^2 + \cdots$$

$$= \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{\lambda} \lambda + \frac{1}{k_B T} \left\langle \left(\frac{\partial H}{\partial \lambda} - \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_0 \right)^2 \right\rangle_0 + \cdots$$

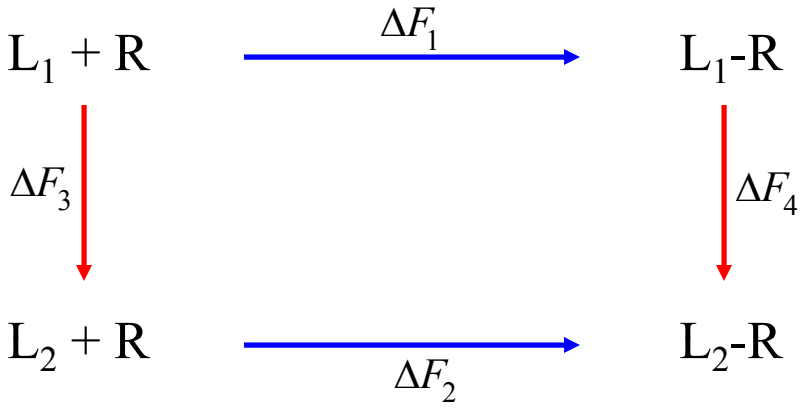
Truncating this series after the first derivative and integrating provides the basis for the TI approach. If the Taylor series expansion is continued until it converges then is equivalent to the FEP formula.

A review for FEP, TI, and SG:

Jorgensen WL: **Computation of Free Energy Changes in Solution.** In *Encyclopedia of Computational Chemistry* P. Edited by v.R. Schleyer. Wiley: New York; 1998, **2:**1061-1070.

Application (I)

Thermodynamic cycles



- In principle, the relative binding affinity: $\Delta\Delta F = \Delta F_2 - \Delta F_1$

  They are actual association process, but it would be difficult for sampling

- Free energy is a state function: $\Delta\Delta F = \Delta F_4 - \Delta F_3$

  They are non-physical processes, but quite feasible in the computer

We can perform two simulations: 'mutant' L_1 into L_2 in solution and L_1 to L_2 within the receptor.

$$H(\lambda) = \lambda H_2 + (1-\lambda) H_1$$

Application (II)

Partitioning the free energy

Which interactions contribute to the most of the overall free energy?

In Thermodynamic Integration:

$$\Delta F = \int_0^1 \left\langle \frac{\partial H(p^N, r^N, \lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \qquad \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda = \left\langle \frac{\partial H_{bonded}(\lambda)}{\partial \lambda} + \frac{\partial H_{vdw}(\lambda)}{\partial \lambda} + \frac{\partial H_{el}(\lambda)}{\partial \lambda} + \cdots \right\rangle$$

$$\Delta F = \int_0^1 \left\langle \frac{\partial H_{bonded}(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda + \int_0^1 \left\langle \frac{\partial H_{vdw}(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda + \int_0^1 \left\langle \frac{\partial H_{el}(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda + \cdots$$

$$= \Delta F_{bonded} + \Delta F_{vdw} + \Delta F_{el} + \cdots$$

In FEP, this can also be achieved by first perturbing the electrostatic and then the van-der-Waals parameters.

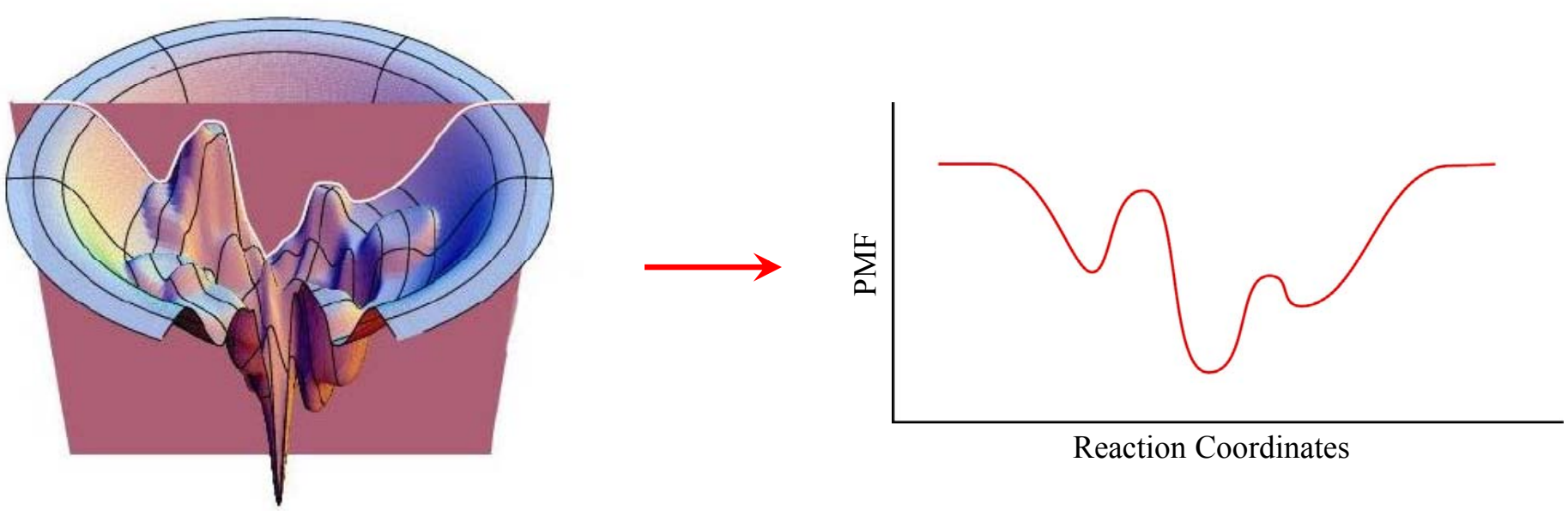
Note: only the sum of the contributions is truly meaningful, the individual contributions are not state functions.

Boresch S, Karplus M: **The meaning of component analysis: decomposition of the free energy in terms of specific interactions**. *J Mol Biol* 1995, **254:**801-807.

Potential of Mean Force Calculations

(Boltzmann Inversion)

Potential of Mean Force (PMF)



- We can identify or hypothesize one biological process to take place along some inter- or intra-molecular coordinates, called reaction coordinates (RC).
- PMF is basically the free energy profile along the reaction coordinates, and all the other degrees of freedom will be averaged out.

Potential of Mean Force (PMF)

A simple example, We select the distance between two atoms as RC, the PMF is the free energy change as the separation (r) between the atoms is changed. The distribution of r can be described by the radial distribution function *g(r)*, so:

$$F(r) = -k_B T \ln g(r)$$

For a general RC q: $F(q) = -k_B T \ln g(q)$

For multi-dimension cases, (q,s) :

$$F(q,s) = -k_B T \ln g(q,s)$$

It is often difficult to find suitable RC for detailed biological processes.

Potential of Mean Force (PMF)

Example: membrane channel. There is a “natural” RC

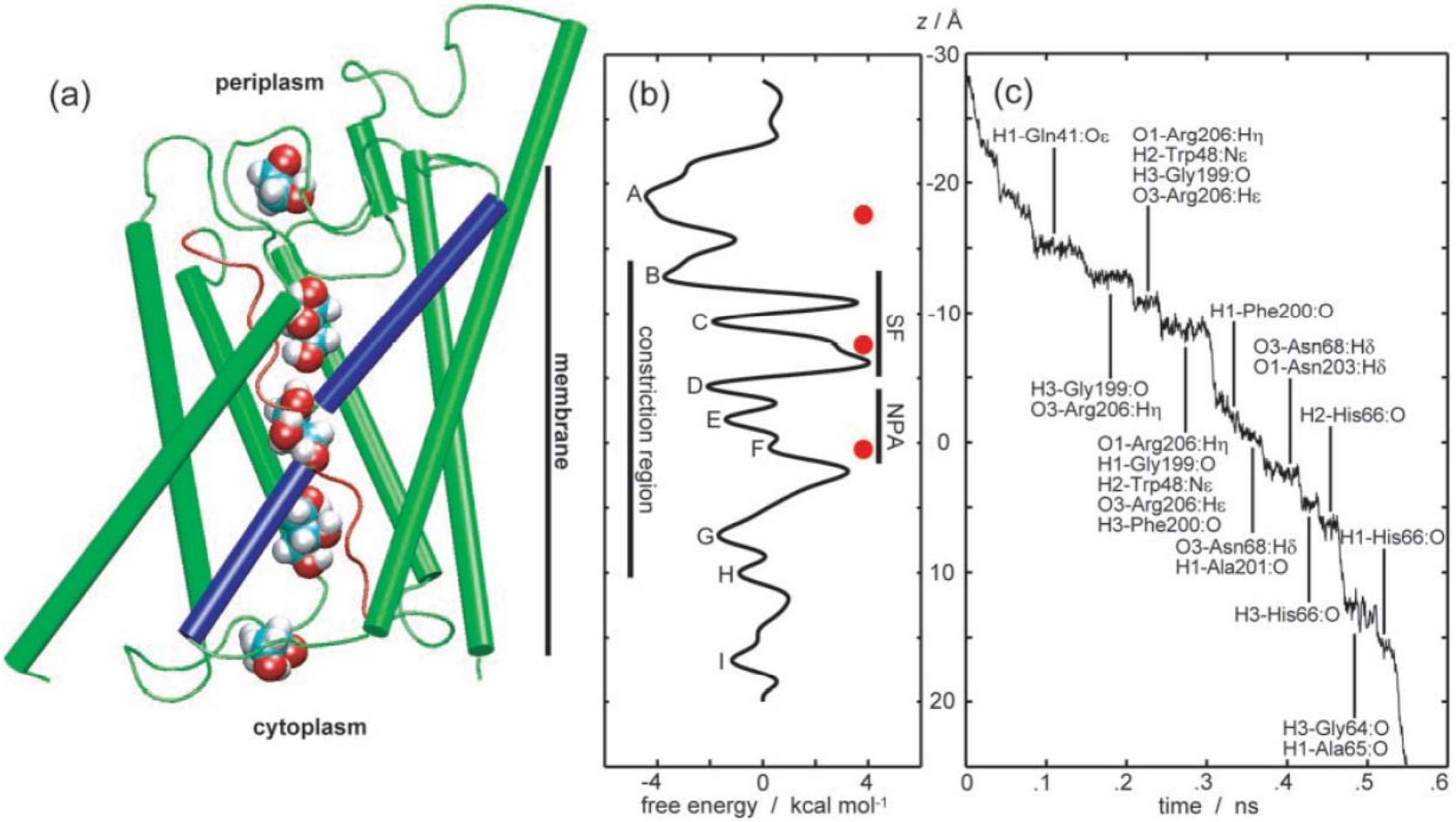
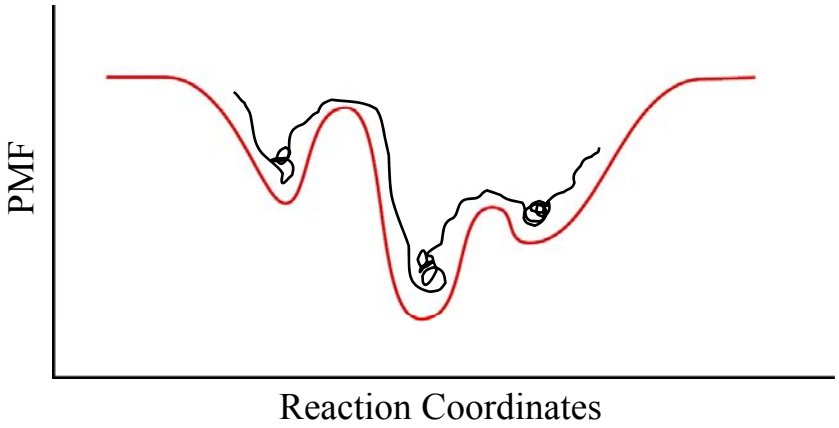


Fig. 4. Energetics of glycerol conduction through GlpF.

Jensen MØ, Park S, Tajkhorshid E, Schulten K: **Energetics of glycerol conduction through aquaglyceroporin GlpF.** *Proc Natl Acad Sci USA* 2003, **99:**6731-6736.

Potential of Mean Force (PMF)

The logarithmic relationship between the PMF and *g(q)* means that a small change in the free energy may correspond to *g(q)* changing by an order of magnitude or more from its most likely value. Standard MC or MD methods do not adequately sample regions where *g(q)* differs drastically from the most likely value, leading to inaccurate values for the PMF.



We can calculate the PMF using the FEP method. But FEP is commonly used to study 'mutations', which are often along non-physical pathways. We want to calculate PMF for a physically achievable process, so we can get the transition states and derive kinetic quantities such as rate constants. The traditional way to avoid the sampling problem is Umbrella Sampling.

Umbrella Sampling

Umbrella sampling attempts to overcome the sampling problem by modifying the Hamiltonian so that the unfavorable states are sampled sufficiently. The modification can be written as a perturbation.

$$H = H^0 + U(q)$$

U(q) is a weighting function, often takes a quadratic form:

$$U(q) = k(q - q^0)^2$$

For configurations that are far from the equilibrium state q^0, the weighting function will be large and so the simulation will be biased along the relevant RC toward the region of q^0. This technique is called “umbrella sampling”, since the resulting distributions are broader than the Boltzmann distribution.

Umbrella Sampling

Umbrella distribution *g'(q)* is non-Boltzmann. The Boltzmann distribution can subsequently be recovered from the biased one [Torrie and Valleau 1977].

$$g(q) = const \times g'(q) \exp\left[U(q)/k_B T\right]$$

Free energy:

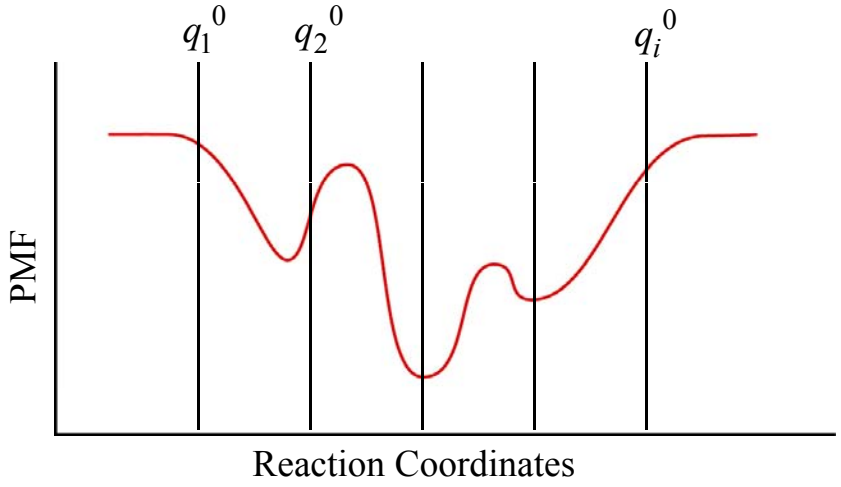
$$F(q) = -k_B T \ln g'(q) - U(q) + F'$$

The free energy shift *F'* is not obvious and depends on *U'(q)*

Limitation: larger *U(q)* will bring numerical uncertainties, and the computed averages will be dominated by only a few terms.

Umbrella Sampling

A succession of simulations using q_1^0, q_2^0, q_i^0, …., can be devised to overcome this problem.



A limit is set for each sampling window

Problem: How to recombine the results from umbrella sampling in different windows / simulations? Different simulations have different free energy offsets.

WHAM: Weighted Histogram Analysis Method

The WHAM equations express the optimal estimate for the unbiased distribution function as a q-dependent *weighted* sum over the N_w sampling windows / simulations.

$$\langle \rho(q) \rangle = \sum_{i=1}^{N_w} n_i \langle \rho(q) \rangle_{(i)} \Big/ \sum_{j=1}^{N_w} n_j \exp\left[-\left(U_j(q) - F_j \right) / k_B T \right]$$

$$F_j = -k_B T \ln\left(\int dq \exp\left[-U_j(q) / k_B T \right] \langle \rho(q) \rangle \right)$$

- n_i = number of counts in histogram bin of simulation i
- U_j, F_j = biasing potential and free energy shift in simulation j

Solve by iteration to self consistency for unknown F_j and $\langle \rho(q) \rangle$
Then $F = -k_B T \ln \langle \rho(q) \rangle$

Kumar S, Bouzida D, Swendsen RH, Kollmann PA, Rosenberg JM: **The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method.** *J Comp Chem 1992*, **13**: 1011-1021.

Roux B: **The calculation of the potential of mean force using computer simulations.** *Comp Phys Commu 1995*, **91**: 275-282.

Adaptive Umbrella Sampling

Instead of a quadratic form of umbrella potential:

$$U(q) = k(q - q^0)^2$$

We can get it from PMF. A simulation with a potential:

$$H = H^0 + U(q)$$

Its probability density is proportional to:

$$\rho(q) \propto \rho^0(q) \exp\left[-U(q)/k_B T\right]$$

So uniform sampling can be obtained by setting:

$$U(q) = k_B T \ln \rho^0(q)$$

Which is just the negative of the PMF.

$\rho^0(q)$ is not known at the beginning, an iterative procedure is used to obtain successively improved approximations to the ideal umbrella potential.

AUS Example

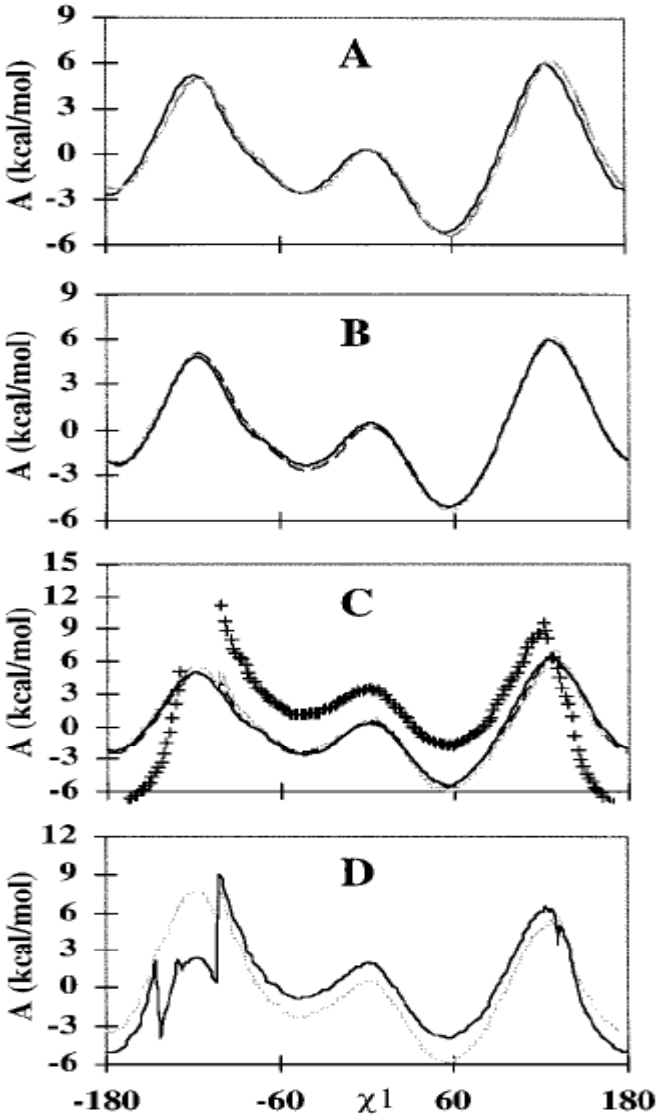


FIGURE 1. Potentials of mean force for the dihedral angle $\chi 1$ in the threonine dipeptide. (A) Comparison of the potentials obtained with different techniques: gray line, adaptive umbrella sampling technique; black line, internal coordinate constraints combined with thermodynamic perturbation. (B) Reproducibility of results from adaptive umbrella sampling. The results are shown from three runs using different random seeds for the assignment of the initial velocities. (C) Convergence of the adaptive umbrella sampling technique: (+) Result after 10 updates of the umbrella potential (100 ps of simulation time); no symbols are plotted in the region where no sampling occurred. Gray line, dashed black line, and continuous black line are the results after 20, 50, and 100 updates of the umbrella potential (200, 500, and 1000 ps of simulation time), respectively. (D) Effect of discarding the data of the fourth simulation: gray and black lines show the results after 14 updates with and without the data from the fourth simulation, respectively.

Bartels C, Karplus M: **Multidimensional adaptive umbrella sampling: applications to main chain and side chain peptide conformations.** *J Comp Chem* 1997, **18**: 1450-1462.

Multi-Canonical Algorithm

In the canonical ensemble, probability distribution is:

$$\rho_B(E,T) \propto n(E)e^{-E/k_BT}$$

At a lower T, $\rho_B(E, T)$ is small in the high-energy region, and at a higher T, $\rho_B(E,T)$ is small in the low-energy region. How to obtain the precise distribution at room T? By multi-canonical algorithm:

$$\rho_{mu}(E,T) \propto n(E)W_{mu}(E) \equiv const$$

The multi-canonical weight factor satisfies:

$$W_{mu}(E) \propto n^{-1}(E)$$

Which can be determined by a few iterations of simulations.

The canonical distribution for wide range of temperatures can be obtained by the use of the re-weighting techniques:

$$\rho_B(E,T) = \frac{\rho_{mu}(E)W_{mu}^{-1}(E)e^{-E/k_BT}}{\int dE'\,\rho_{mu}(E')W_{mu}^{-1}(E')e^{-E'/k_BT}}$$

Example: Multi-Canonical Algorithm

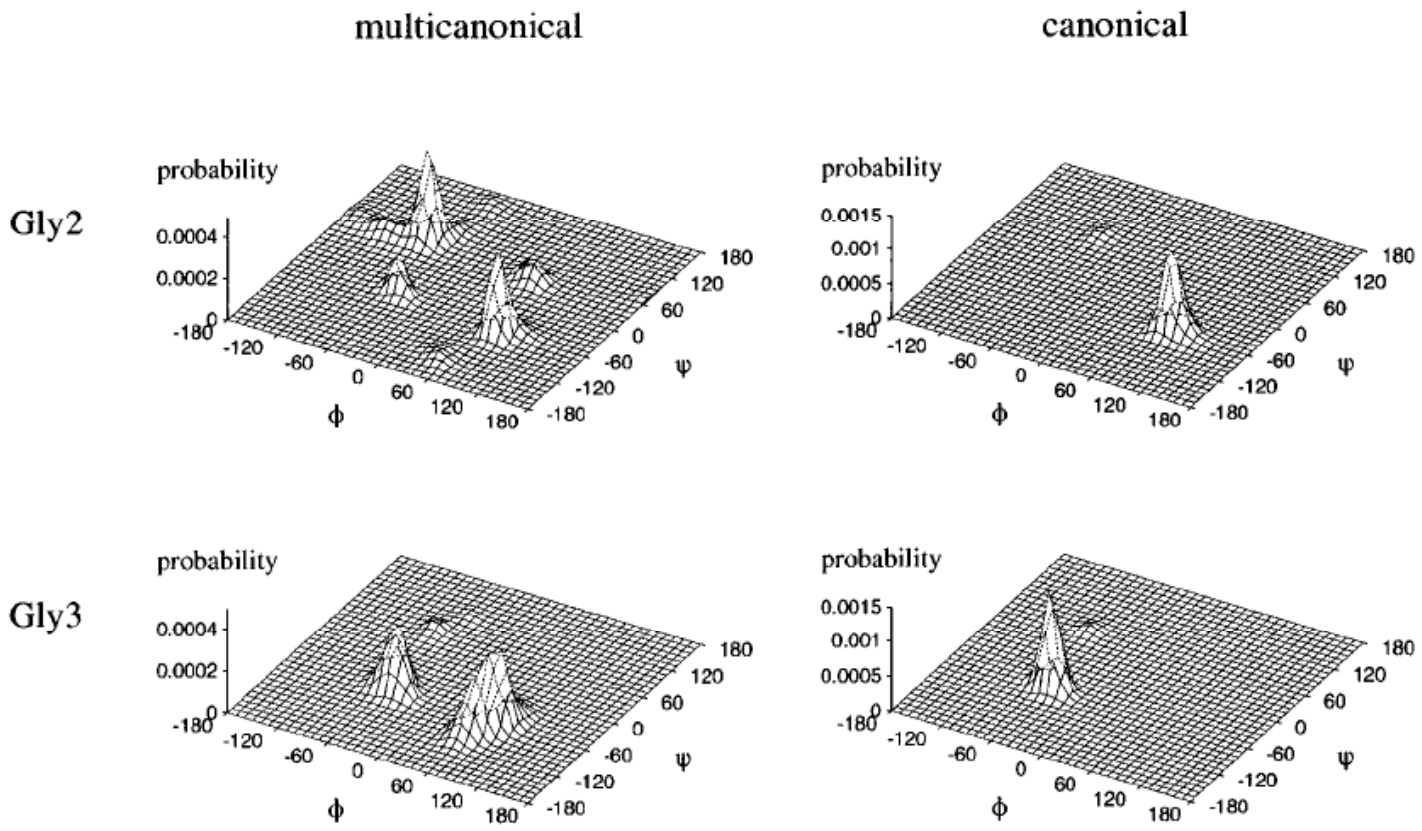
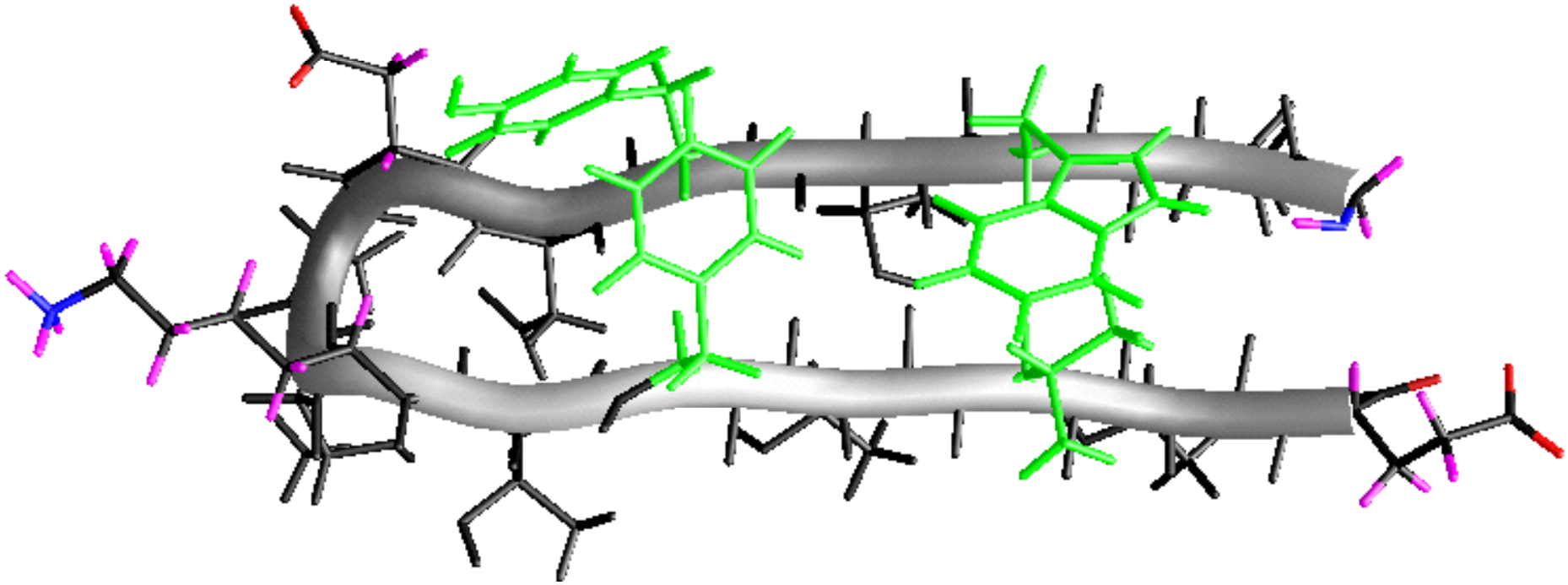


Figure 5. Conformational distributions of the torsion angles, ϕ and ψ, of Gly2 and Gly3 of Met-enkephalin. The figures on the left (indicated as multicanonical) are the canonical distribution at 300 K obtained by reweighting the results of the multicanonical MD. The figures on the right (indicated as canonical) are those of the results of the canonical MD at 300 K.

Nakajima N, Nakamura H, Kidera A: **Multicanonical ensemble generated by molecular dynamics simulations for enhanced conformational sampling of peptides.** *J Phys Chem B* 1997, **101**: 817-824.

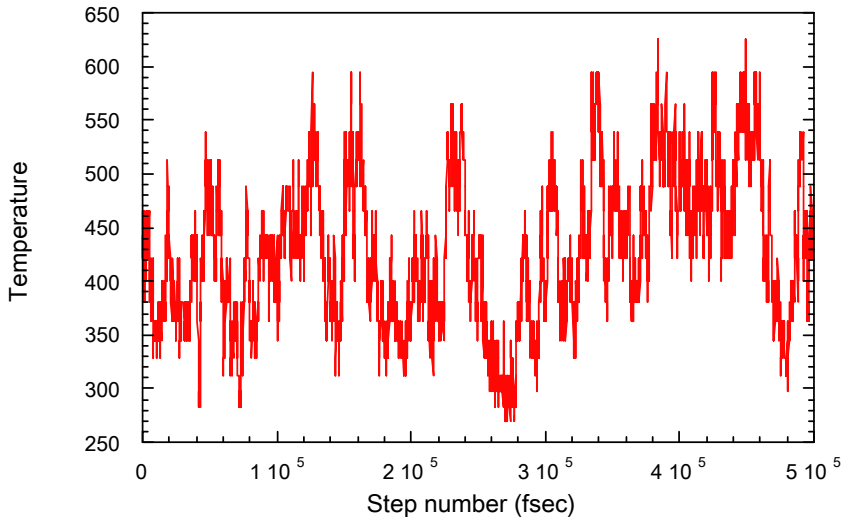
The β-Hairpin of B1 Domain of Protein G



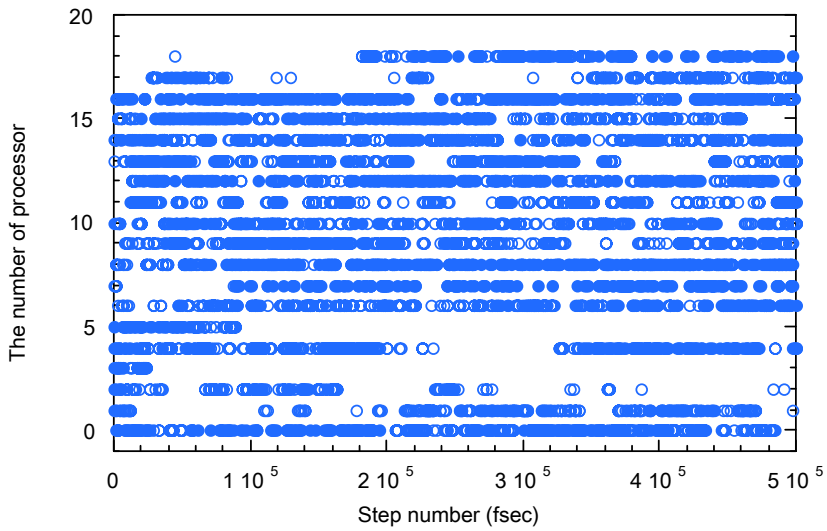
The hydrophobic sidechains are in green.

Replica Exchange Sampling: β-hairpin Folding

- Replica exchange sampling* is a method to effectively sample rough energy landscapes which have high dimensionality - the β -hairpin has 768 degrees of freedom
- ~20 MD simulations of the β-hairpin run in parallel over the temperature range 270 K -690 K.
- Every 50 MD steps MC replica exchange moves are attempted
- Total sampling time: 20 processors $\times 4 \times 10^6$ step/processor = 80×10^6 steps



Time series of the temperature for one replica



Time series of the replicas for one Temp., T = 442 K

Sugita Y, Okamoto Y: **Replica-exchange molecular dynamics method for protein folding.** *Chem Phys Lett* 1999, **314**: 141-151.

T-WHAM

• A way to combine data from simulations at various temperatures to obtain properties at one given temperature.

Energy distribution:

$$\rho(E;T_i) = \frac{Q(T_0)}{Q(T_i)} e^{-(\beta_i-\beta_0)E} \rho(E;T_0)$$

$$\rho(E;T_0) = \frac{n(E)\, e^{-\beta_0 E}}{Q(T_0)}$$

Solve for *n(E)* and insert into expression for $\rho(E;T_i)$.

- Given $\rho(E_j;T_0)$ can predict histogram of energies at any temperature.
- Select $\rho(E_j;T_0)$ that best reproduces observed histograms (maximum likelihood solution assuming multinomial-distributed counts).

WHAM equations:

$$\begin{cases} \rho(E_j;T_0) = \dfrac{\sum_i n_i \rho(E_j;T_i)}{\sum_i n_i\, e^{F_i} e^{-(\beta_i-\beta_0)E_j}} \\ e^{-F_i} \equiv \dfrac{Q(T_i)}{Q(T_0)} = \sum_j \rho(E_j;T_0)\, e^{-(\beta_i-\beta_0)E_j} \end{cases}$$

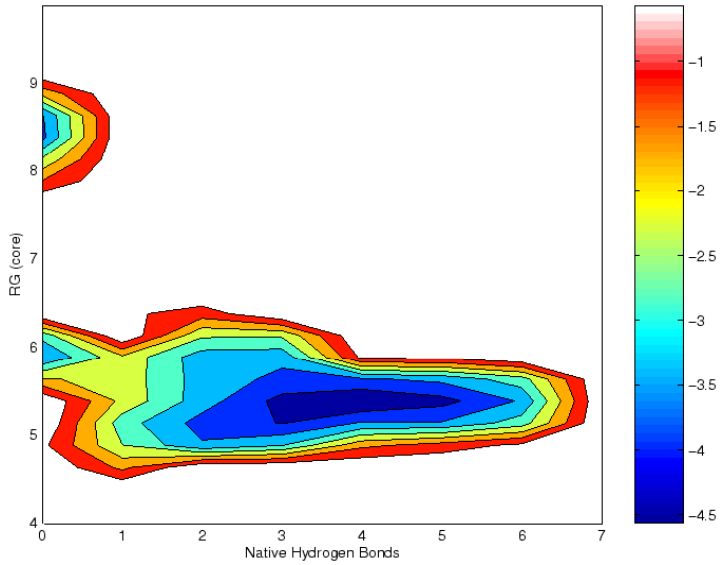
Same derivation for joint probability $\rho(x,E;T)$.

Gallicchio E, Andrec M, Felts AK, Levy RM: **T-WHAM, replica exchange, and transition paths.** *J. Phys. Chem. B, **109** (14), 6722 -6731, 2005*

T-WHAM

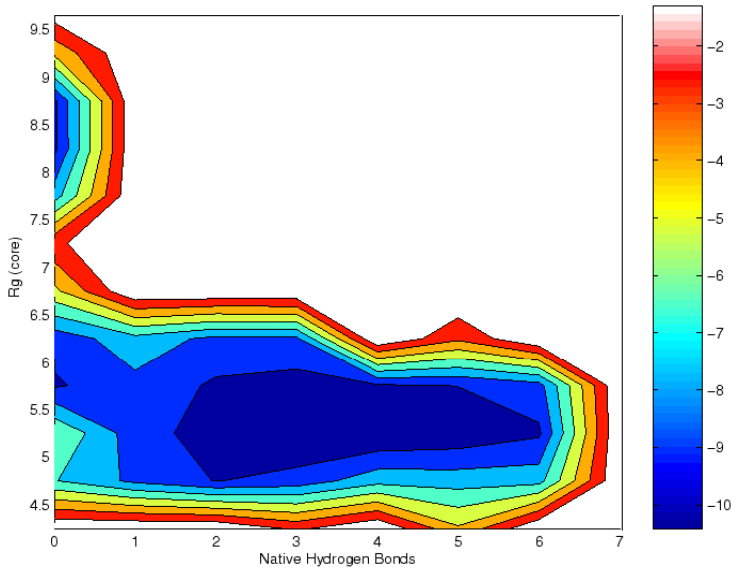
Two-Dimensional PMF

No WHAM



ΔG_{max}=5 kcal/mol

T-WHAM

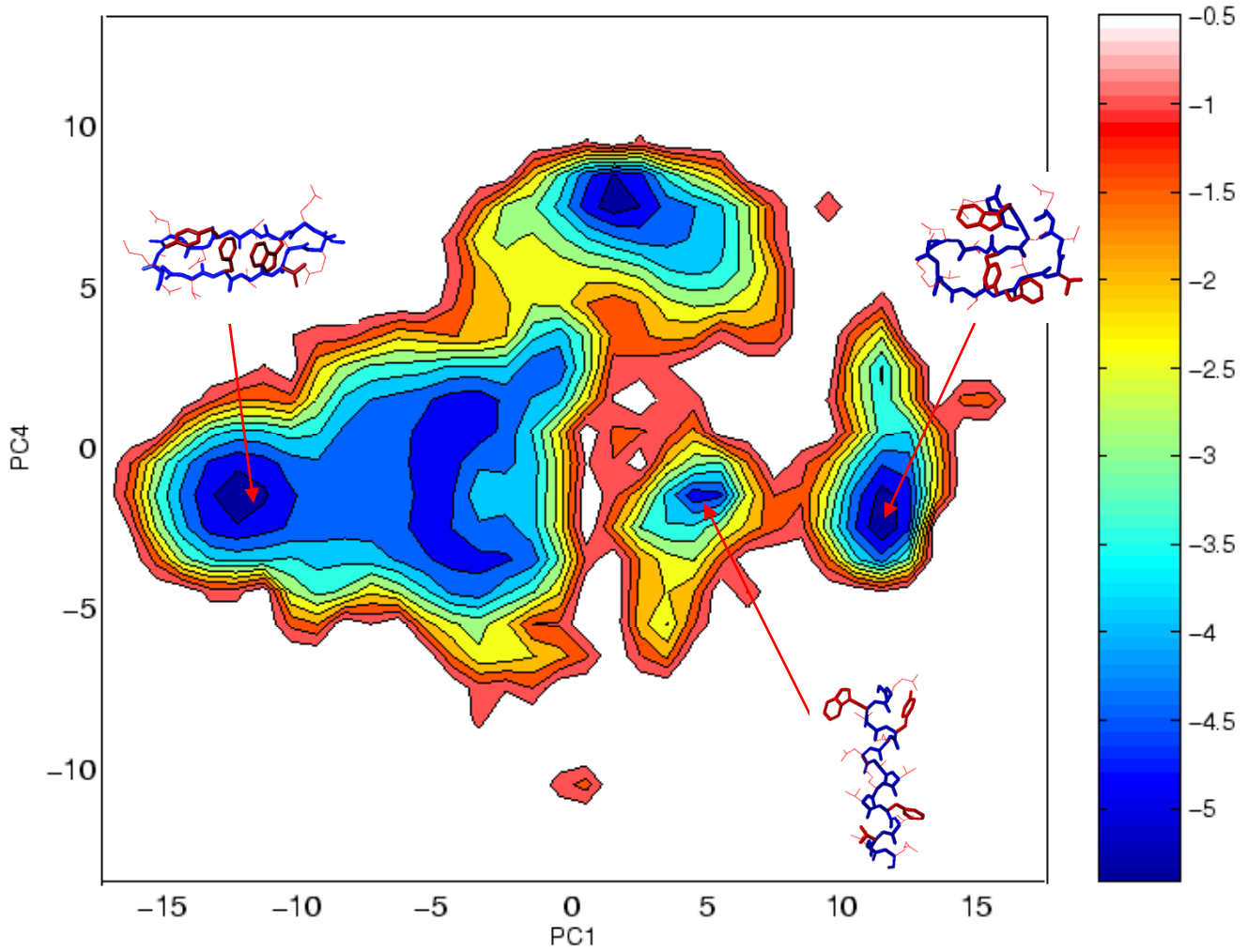


ΔG_{max}=10 kcal/mol

Gibbs free energy in this example

T-WHAM

Free Energy Surface of the Protein G β-Hairpin
With Respect to the (1,4) Principle Components

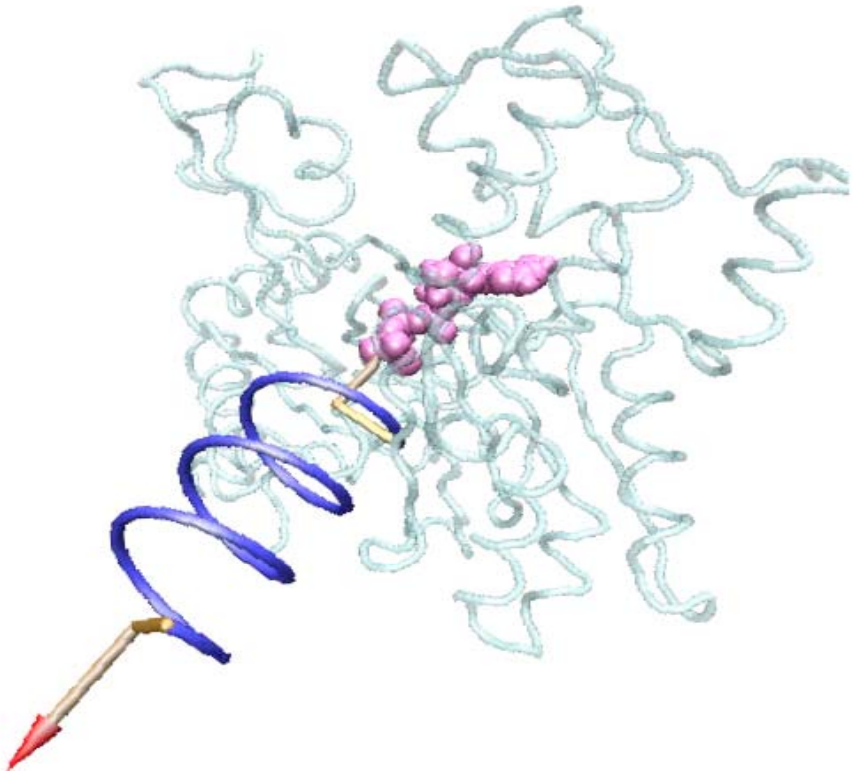


Steered Molecular Dynamics

In SMD, time-dependent external forces are applied to a system, which induce unbinding of ligands and conformational changes in biomolecules on time scales accessible to MD simulations.

Assuming a reaction coordinate x, we add an external force along the path, a simple way is by a harmonic spring:

$$f = k\left(x_0 + vt - x\right)$$



Similar to experiments by Atomic Force Microscopy, a "spring" of stiffness k is attached to the ligand and a constant pulling rate is applied to measure the adhesion forces while the ligand detaches from the protein.

PMF Calculation

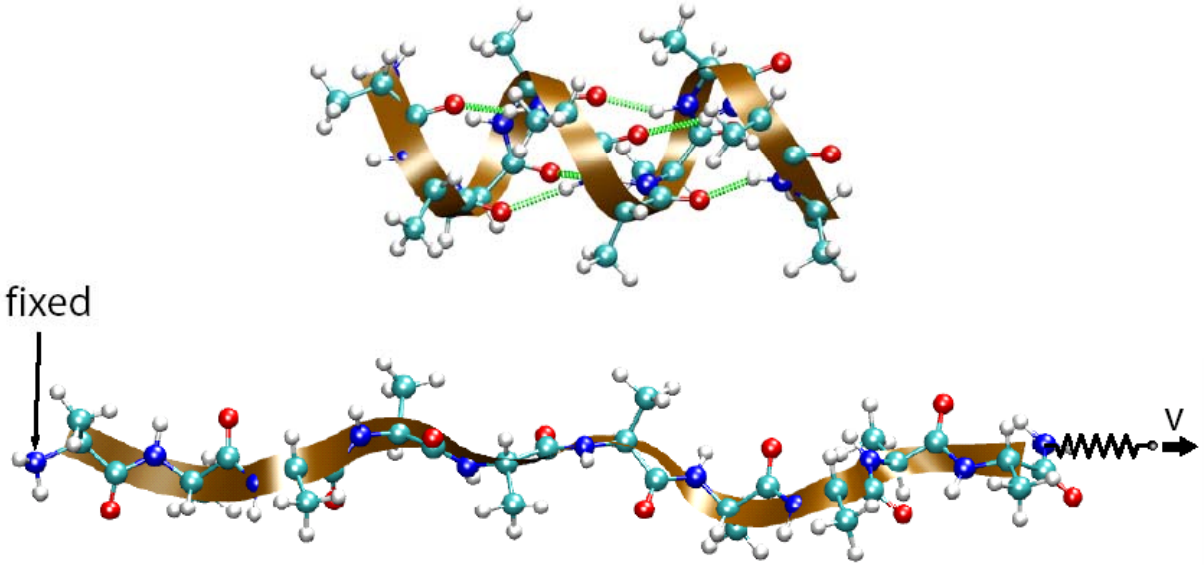
- From equilibrium MD simulations using "umbrella sampling" and the weighted histogram analysis method (WHAM)
- From non-equilibrium SMD simulations using the Jarzynski equality (many samples) or local diffusion processes (single sample).

Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K: **Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality.** *J Chem Phys* 2003, **119:**3559-3566.

Park S, Schulten K: **Calculating potential of mean force from steered molecular dynamics simulations.** *J Chem Phys* 2004, **120:**5946-5961.

Calderon CP: **On the Use of Local Diffusion Models for Path Ensemble Averaging in Potential of Mean Force Computations.** *J Chem Phys* 2007, 126:84106-84111.

Helix-Coil Transition of Deca-Alanine in Vacuum



Goal:

Systematic study of the methodology of free energy calculation

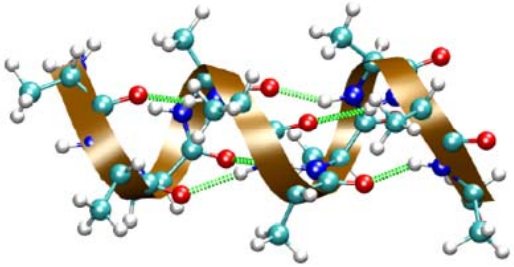
- Which averaging scheme works best

  with small number (~10) of trajectories ?

Why deca-alanine in vacuum?

- small, but not too small: 104 atoms
- short relaxation time → reversible pulling → exact free energy

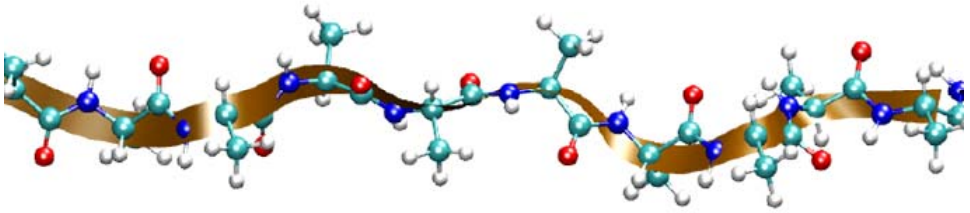
2nd Law of Thermodynamics and Jarzynski's Equality



x = x(t)

heat Q

work W



Stretching Deca-Alanine…

x = reaction coordinate (e.g., end-to-end distance, position of substrate along a channel)

2nd law of thermodynamics: $\langle W \rangle \geq \Delta F = F(x) - F(x_0)$

Jarzynski (1997): $\langle \exp(-\beta W) \rangle = \exp(-\beta \Delta F)$

Statistical average, dominated by small work values that arise only rarely: difficult to estimate!

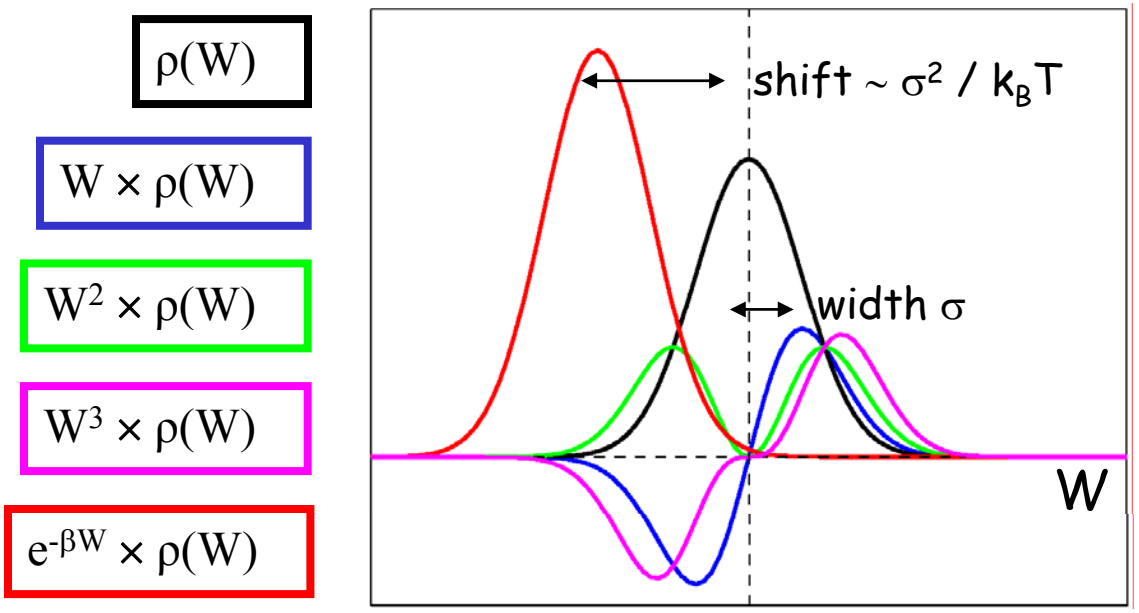
Cumulant Expansion of Jarzynski's Equality

$$\Delta F = -(1/\beta) \log \langle e^{-\beta W} \rangle$$

$$= \langle W \rangle - (\beta/2)\left(\langle W^2 \rangle - \langle W \rangle^2 \right)$$

$$+ (\beta^2/6)\left(\langle W^3 \rangle - 3\langle W^2 \rangle \langle W \rangle + 2\langle W \rangle^3 \right) + \bullet\bullet\bullet$$

statistical error & truncation error

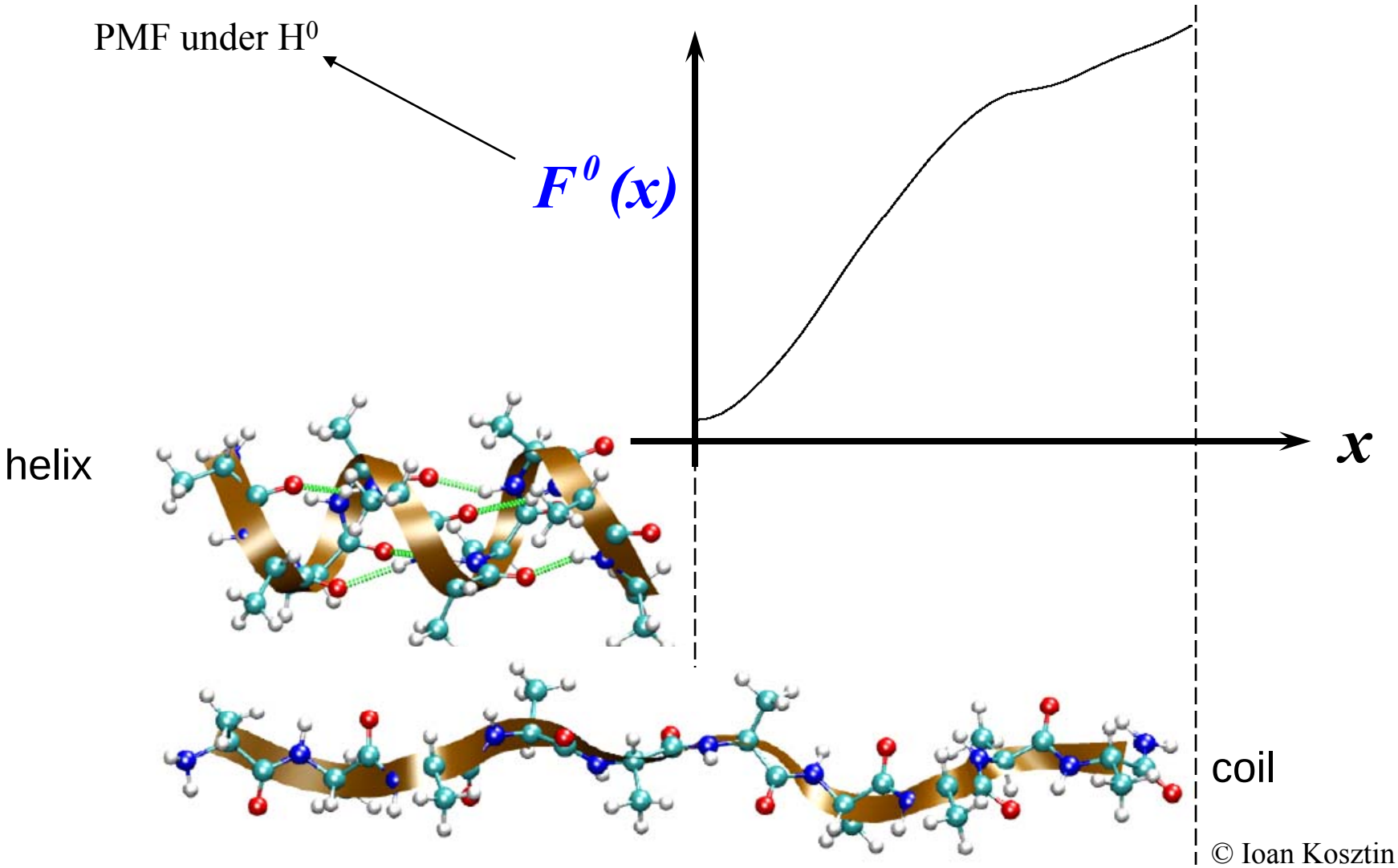


shift/width ~ $\sigma / k_B T$

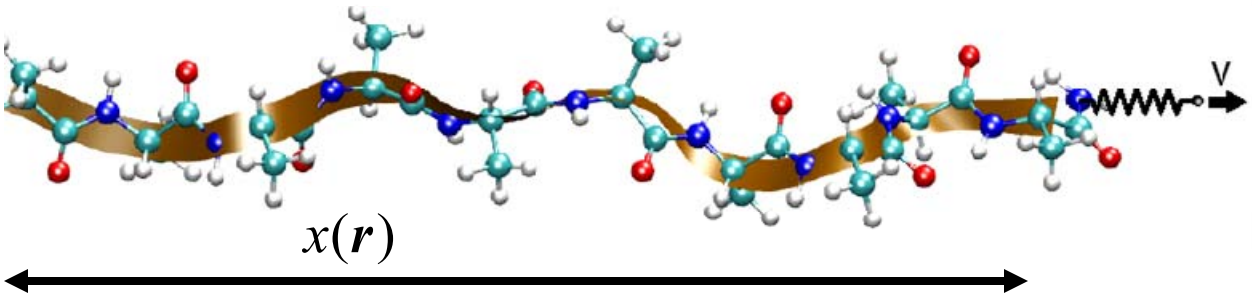
large in strong nonequilibrium

Potential of Mean Force (PMF)

$$e^{-\beta F^0(x')} \propto \int e^{-\beta H^0(\boldsymbol{r},\boldsymbol{p})} \delta\left[x(\boldsymbol{r}) - x'\right] d\boldsymbol{r}d\boldsymbol{p}$$



PMF from Jarzynski's Equality



Introduce an external parameter λ, which is correlated with the RC x:

Guiding potential: $h(\mathbf{r},\lambda)=\frac{k}{2}\left[x(r)-\lambda\right]^2 \quad H(r,p,\lambda)=H^0(r,p)+h(r,\lambda)$

$$\lambda \equiv \lambda(t) = \lambda_0 + vt$$

External Work: $$W(\lambda) = -kv\int_0^t dt'\left[x(t') - \lambda_0 - vt'\right]$$

By Jarzynski's equality:

$$\exp\left\{-\beta\left[F(\lambda_t) - F(\lambda_0)\right]\right\} = \left\langle \exp(-\beta W_{0\to t})\right\rangle$$

What we get is PMF under H: F(λ), how can we get F^0(x)?

Stiff Spring Approximation

$$\exp\left[-\beta F(\lambda)\right] = \int dr\, dp\, \exp\left[-\beta H(r,p,\lambda)\right]$$

$$= \int dr\, dp\, \exp\left\{-\beta H^0(r,p) - \frac{\beta k}{2}\left[x(r)-\lambda\right]^2\right\}$$

$$= \int dr\, dp \int dx'\, \delta\left(x(r)-x'\right) \exp\left\{-\beta H^0(r,p) - \frac{\beta k}{2}\left[x(r)-\lambda\right]^2\right\}$$

$$= \int dx' \exp\left[-\beta F^0(x') - \frac{\beta k}{2}(x'-\lambda)^2\right]$$

When *k* is large, most contribution to the integral comes from the region around: $x' = \lambda$

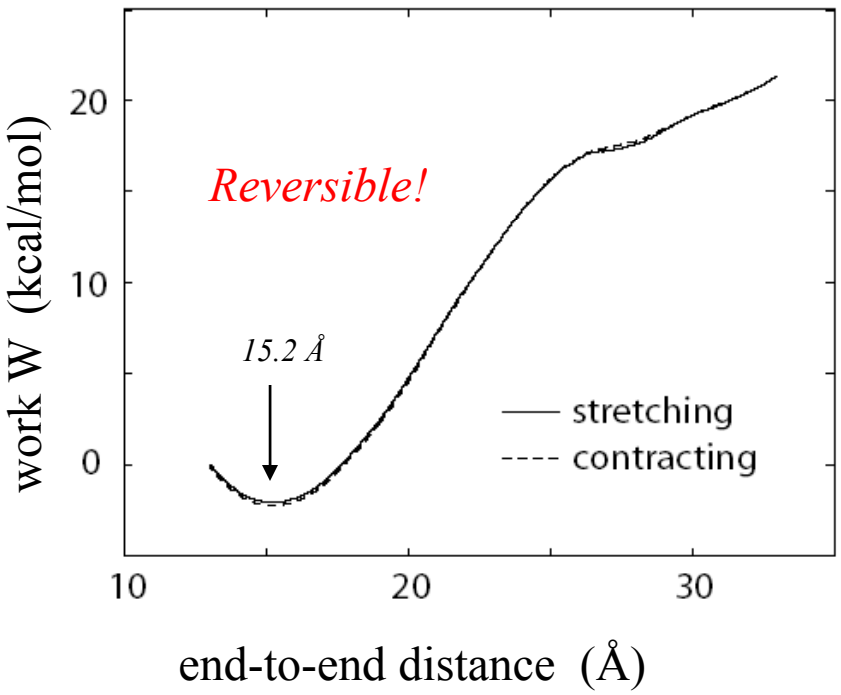
PMF (stiff spring approximation):

$$F(\lambda) \approx F^0(\lambda)$$

$$F^0(\lambda_t) = F(\lambda_0) - \frac{1}{\beta}\log\left\langle e^{-\beta W_{0\to t}} \right\rangle$$

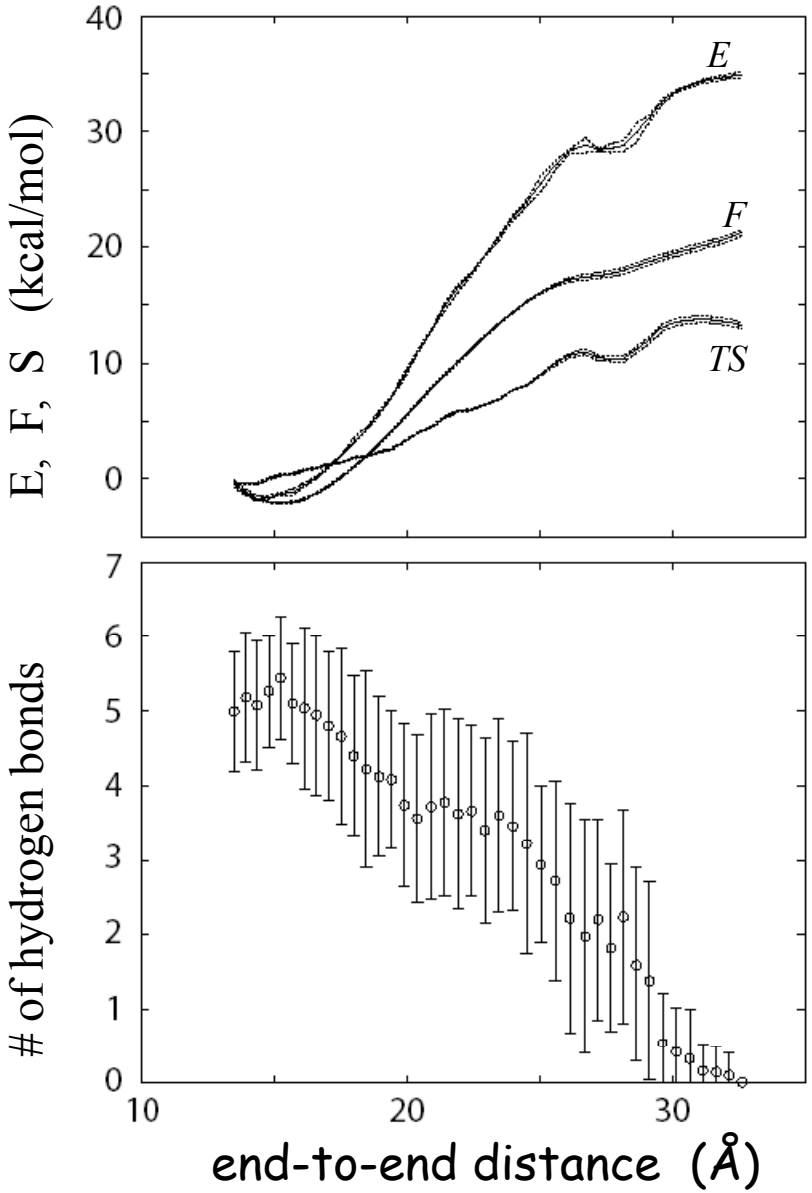
k is large enough, so that the RC closely follows the constraint center λ.

SMD: Reversible Pulling (v = 0.1 Å/ns)

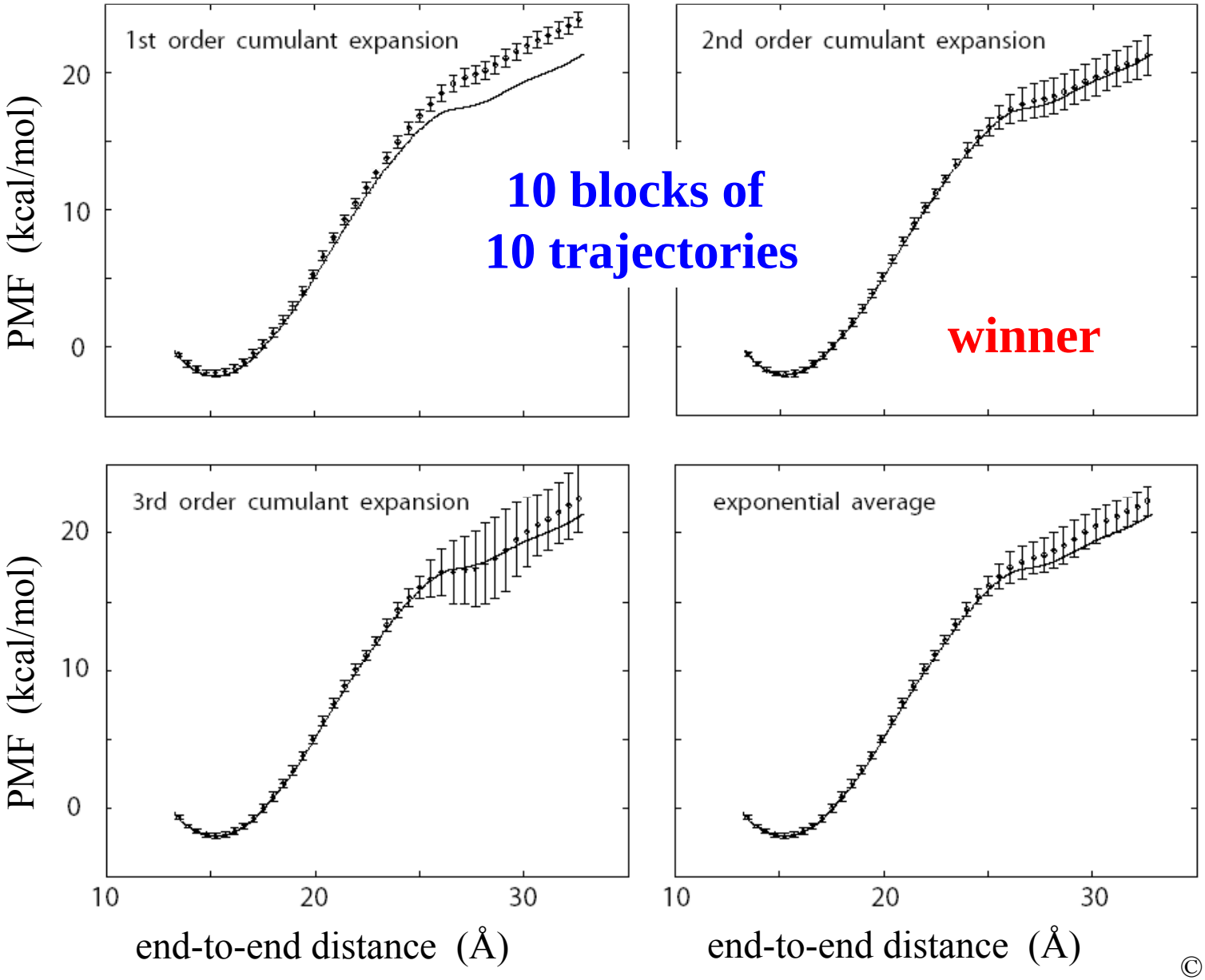


$$\Delta F = \langle W \rangle$$

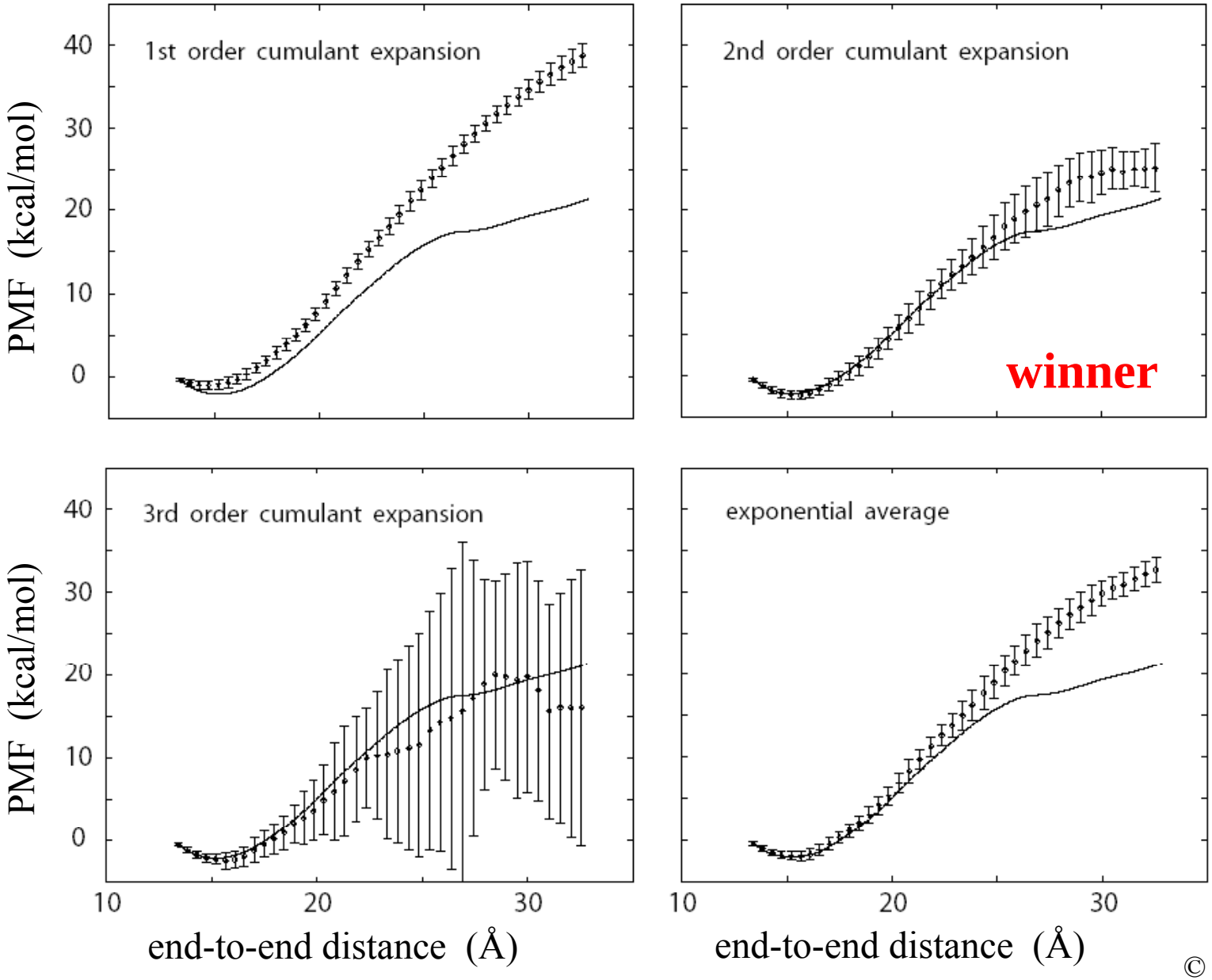
$$TS = E - F$$



SMD: Irreversible Pulling (v = 10 Å/ns)



SMD: Irreversible Pulling (v = 100 Å/ns)



PMF from Umbrella Sampling

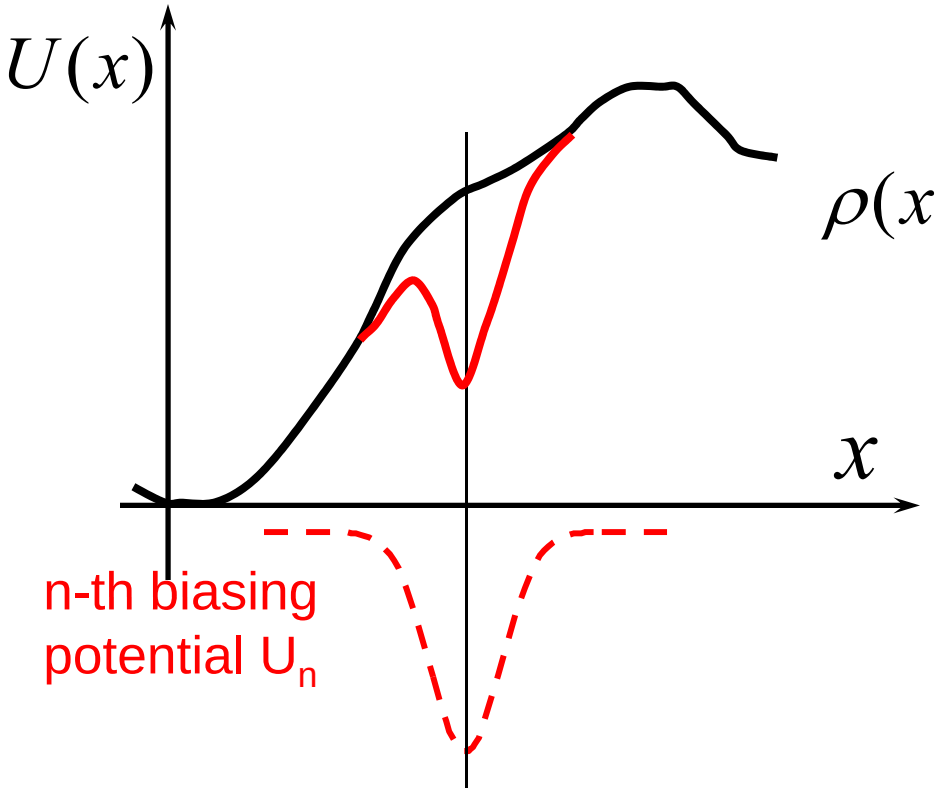
$$\rho(x') = e^{-\beta U(x)} = \left\langle \delta\left[x(\boldsymbol{r}) - x'\right]\right\rangle$$

histogram built from equilibrium MD simulation

probability distribution (density)

$$= \frac{\int d\boldsymbol{r}\, e^{-\beta U(\boldsymbol{r})}\delta\left[x(\boldsymbol{r}) - x'\right]}{\int d\boldsymbol{r}\, e^{-\beta U(\boldsymbol{r})}}$$

partition function (Q)

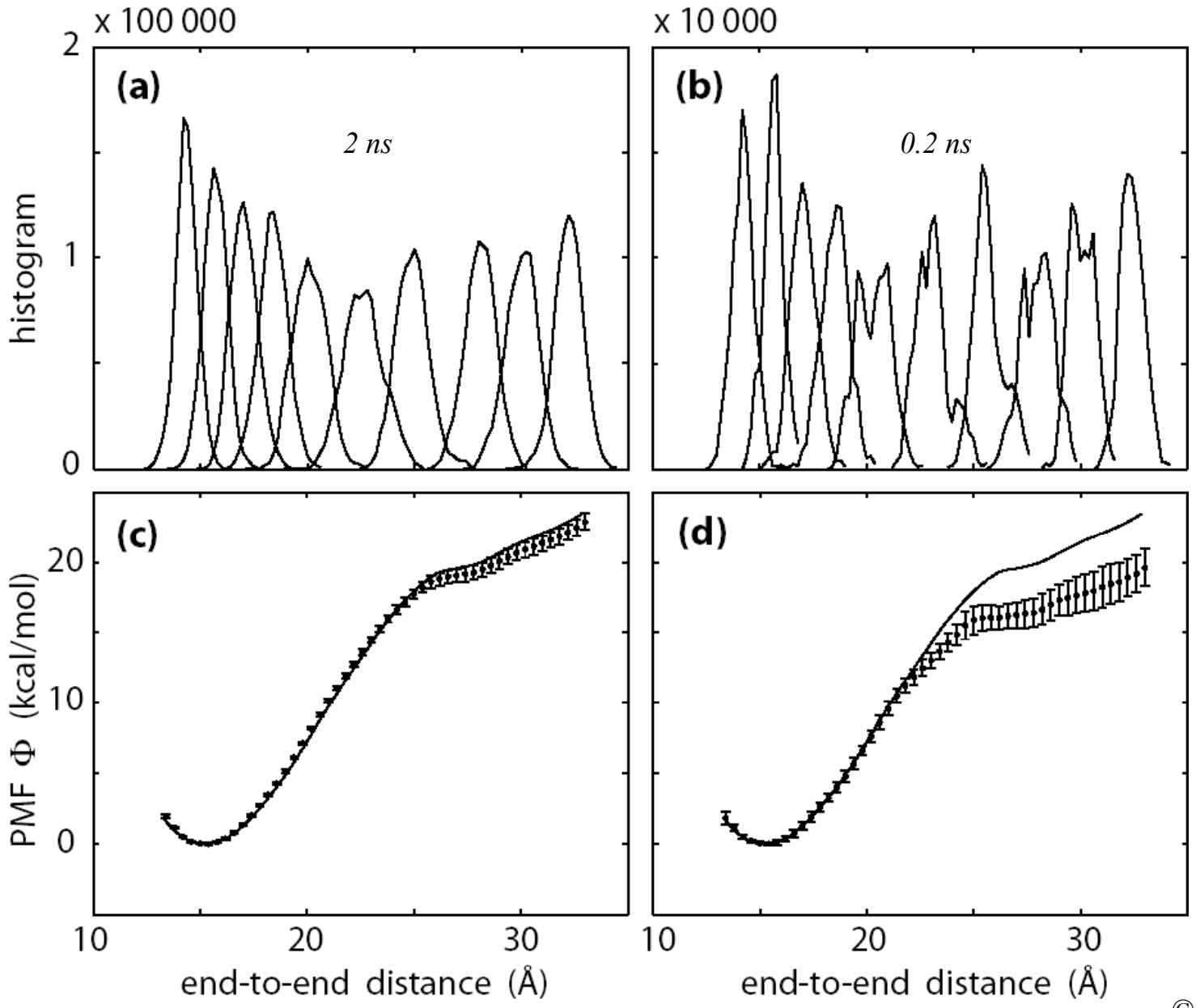


$$\rho(x') = e^{-\beta U(x')} = \left\langle \delta\left[x(\boldsymbol{r}) - x'\right]\right\rangle$$

$$= \frac{Q_n}{Q}\left\langle e^{-\beta U_n(\boldsymbol{x}')}\delta\left[x(\boldsymbol{r}) - x'\right]\right\rangle_n$$

$$= \frac{Q_n}{Q} e^{-\beta U_n(x')}\left\langle \delta\left[x(\boldsymbol{r}) - x'\right]\right\rangle_n$$

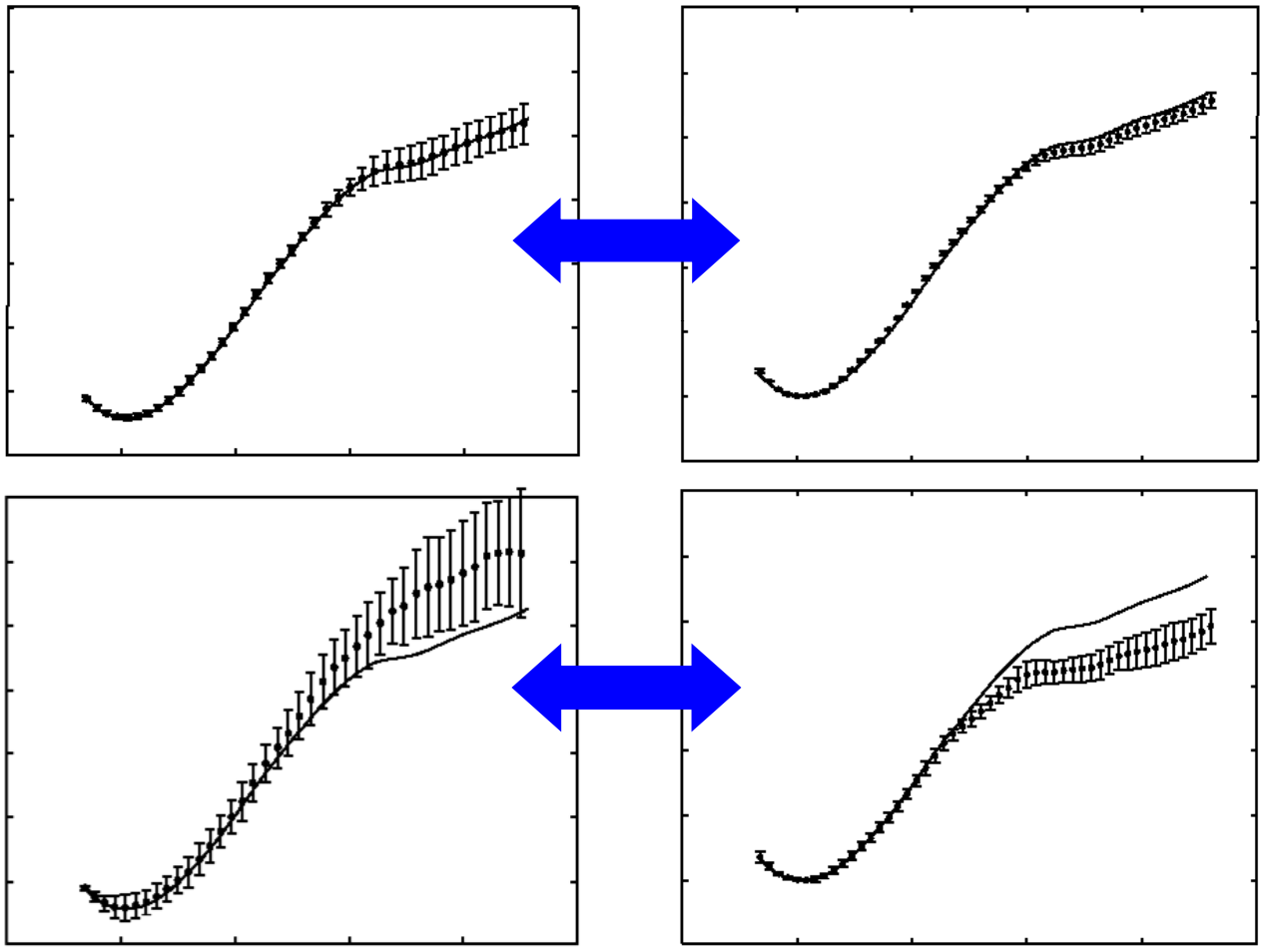
Umbrella Sampling w/ WHAM



Comparison

SMD-Jarzynski ⟷ Umbrella Sampling

(same amount of simulation time)



simple analysis

uniform sampling of the RC

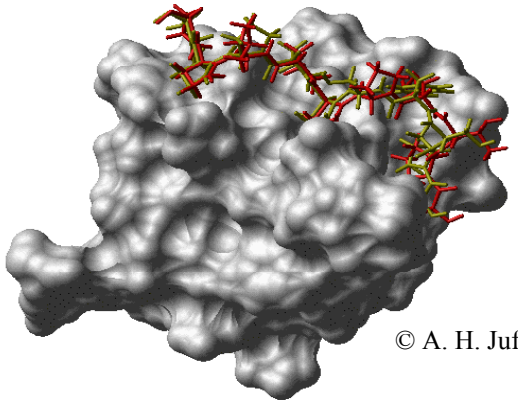
coupled nonlinear equations (WHAM)

nonuniform sampling of the RC

‘Rapid’ Free Energy Methods

Motivation

Free energy calculations are very important in computer-aided drug design. However, if the calculations takes longer to perform than a candidate drug molecule can be synthesized and tested, then there is little practical benefit from attempting the calculation.



Free energy calculations are time-consuming. It is necessary to develop some alternative methods, which still being based upon ‘exact’ statistical mechanics, are intended to provide free energy with less computational effort than a full free energy calculation.

Linear Interaction Energy (LIE)

A semi-empirical method for estimating absolute binding free energies of ligands binding to proteins. The interaction between the ligand and protein or solvent is broken down into the electrostatic and van der Waals contributions.

$$\Delta F = \alpha\left(\left\langle E_{l-p}^{el}\right\rangle - \left\langle E_{l-s}^{el}\right\rangle\right) + \beta\left(\left\langle E_{l-p}^{vdw}\right\rangle - \left\langle E_{l-s}^{vdw}\right\rangle\right)$$

To determine ΔF one thus needs to perform just two simulations, one of the ligand in the solvent and the other of the ligand bound to the protein.

Linear Interaction Energy (LIE)

$$\Delta F = \alpha\left(\left\langle E_{l-p}^{el}\right\rangle - \left\langle E_{l-s}^{el}\right\rangle\right) + \beta\left(\left\langle E_{l-p}^{vdw}\right\rangle - \left\langle E_{l-s}^{vdw}\right\rangle\right)$$

What remains is to determine values of the parameters α and β. By some analytical theories, the parameter α related to the electrostatic contribution is around 1/2.

$$\alpha \approx 0.5$$

For the van-der-Waals component no such analytical theory exists. β depends on different force field, and the nature of the binding sites (different distributions of polar and non-polar groups in different binding sites).

Example: LIE

Binding free energies of different compounds binding to Avidin

Wang W, Wang J, Kollman PA: **What determines the van der Waals coefficient *β* in the LIE (Linear Interaction Energy) method to estimate binding free energies using molecular dynamics simulations?** *Proteins Struct Funct Genet 1999*, **34:**395-402.

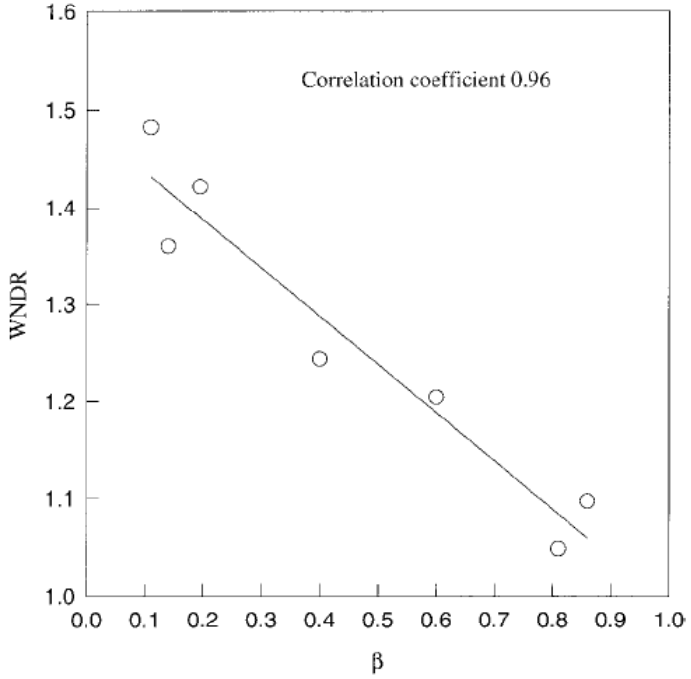


Fig. 2. β value versus weighted non-polar desolvation ratio (WNDR) for the seven calibration systems.

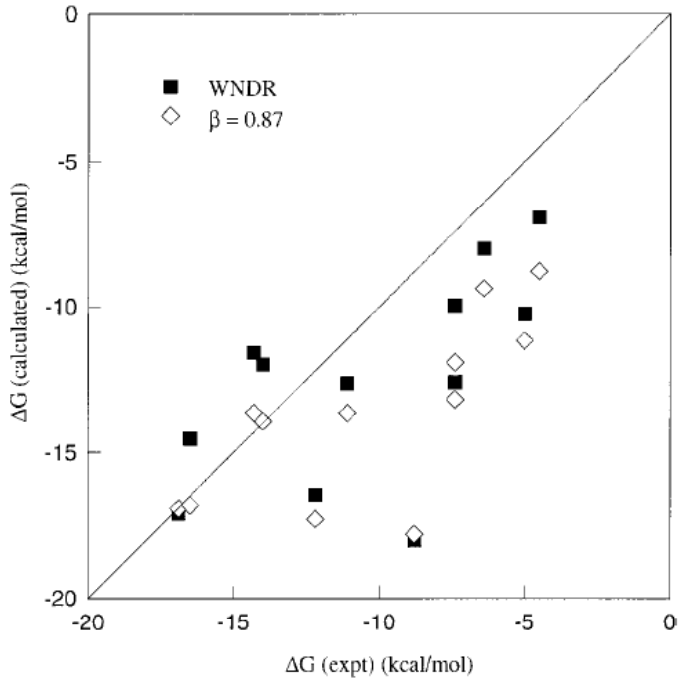


Fig. 4. Observed versus calculated binding free energies for the 12 compounds binding to avidin using β predicted from the correlation obtained from Figure 2 and β = 0.87, respectively.

The weighted non-polar desolvation ratio (WNDR) is the ratio of all non-polar groups' weighted desolvation SAS to total weighted desolvation SAS.

It is generally more accurate to calibrate *β* if experimental binding data for similar ligands is available. Choose a value based on the WNDR could give better results.

MM/PBSA

(Molecular Mechanics Poisson-Boltzmann Surface Area) Method

The MM/PBSA approach represents the post-processing method to evaluate free energies of binding or to calculate absolute free energies of molecules in solution, which combines the molecular mechanical energies with the continuum solvent approaches. In this method, we usually carry out a MD simulation with explicit water and counterions. The we post-process these structures, removing any solvent and counterion molecules, and calculate the Gibbs free energy.

Calculated average free energy $\longrightarrow$ $\overline{G} = \overline{E}_{MM} + \overline{G}_{PBSA} - TS_{MM}$

Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case D, Cheatham TE, III: **Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models.** *Acc Chem Res* 2000, **33:**889-897.

MM/PBSA

The components in MM/PBSA equation: $\overline{G} = \overline{E}_{MM} + \overline{G}_{PBSA} - TS_{MM}$

$\overline{E}_{MM} = \overline{E}_{bond} + \overline{E}_{angle} + \overline{E}_{torsion} + \overline{E}_{vdw} + \overline{E}_{elec}$ → average molecular mechanical energy

$\overline{G}_{PBSA} = \overline{G}_{elec} + \overline{G}_{nonpolar}$ → Solvation free energy

$\overline{G}_{elec} = \overline{G}_{PB}$ → Numerical solution of Poisson-Boltzmann equation or

$\overline{G}_{elec} = \overline{G}_{GB} = -166.0(1-\frac{1}{\varepsilon})\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{(r_{ij}^2 + a_{ij}^2 e^{-D_{ij}})^{0.5}}$ → Generalized Born model

$\overline{G}_{nonpolar} = \gamma SA + b$ → Solvent-accessible surface area

$-TS_{MM}$ → Solute entropy, which is likely to be much smaller then other terms. It can be estimated by harmonic analysis or normal mode analysis.

Example: MM/PBSA (I)

Binding free energy of protein-ligand

$$\Delta G = \overline{G}_{complex} - \overline{G}_{protein} - \overline{G}_{ligand}$$

Two methods: (a) separate simulations of complex, protein, and ligand or (b) evaluate all three terms using just the snapshots from a complex simulation.

(b) is a good approximation in cases that, there are no large conformational changes of protein and ligand before and after their association.

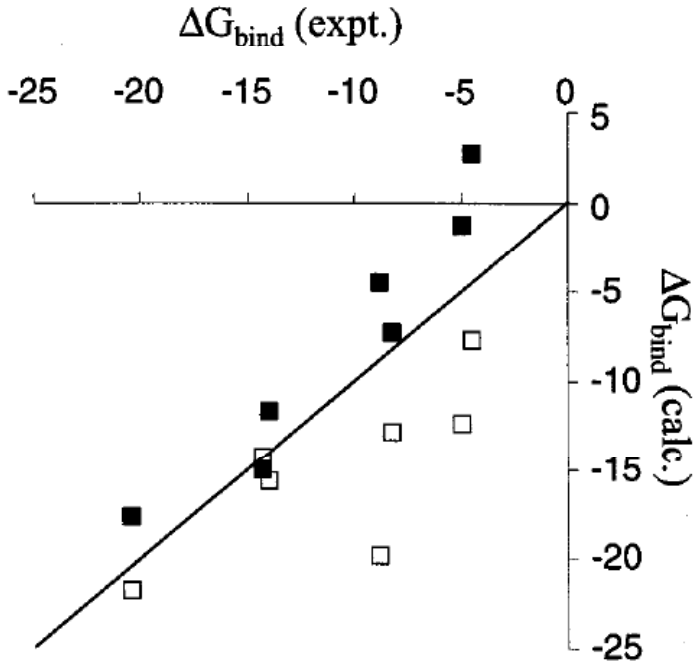


FIGURE 2. Correlation between calculated and experimental protein−ligand binding free energies for avidin and several biotin analogues. Black squares denote MM−PBSA calculated free energies,[30] and white squares refer to LIA calculations.[31] The solid line indicates perfect correlation ($r^2 = 1$).

Kuhn B, Kollman PA: **Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models.** *J Med Chem* 2000, **43:**3786-3791.

Example: MM/PBSA (II)

Binding free energy of protein-RNA

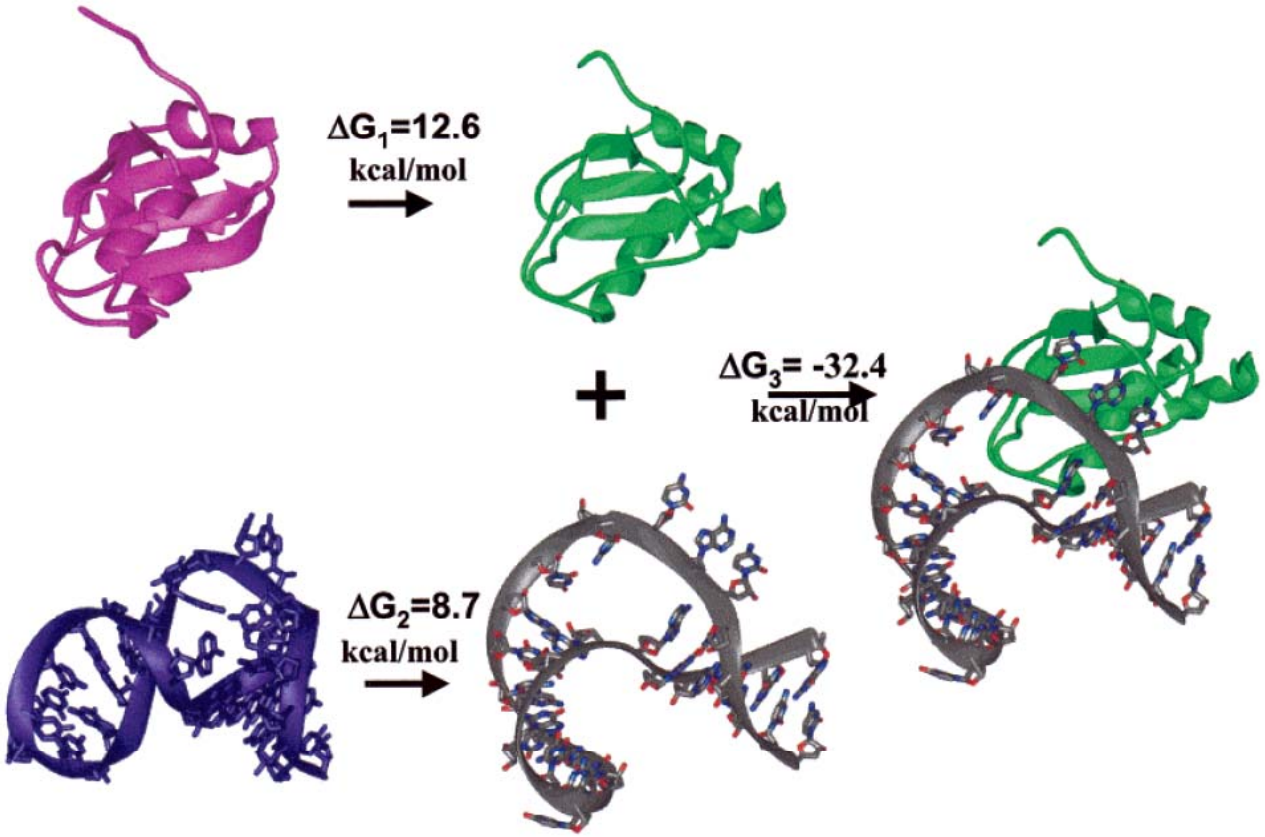


FIGURE 4. Conformational change upon binding of U1A protein and internal loop (IL) RNA. ΔG_1 and ΔG_2 are the MM–PBSA free energy differences between free and bound protein and RNA, respectively. ΔG_3 is the free energy of association of protein and RNA in their bound structures.

Reyes C, Kollman PA: **Structure and thermodynamics of RNA-protein binding: using molecular dynamics and free energy analysis to calculating both the free energies of binding and conformational change.** *J Mol Biol* 2000, **297:**1145-1158.

MM/PBSA

• By using a continuum model, integrating out all the solvent coordinates;

• Calculating the absolute free energy directly instead of the relative free energy along a RC;

• large errors, but we can often calculate ΔG in respectable agreement with experiment;

• Rate-limiting step is MD, but a hierarchy of techniques can be used.

Resources and Further Reading

Acknowledgement:
Zhiyong Zhang, University of Utah

David A. Kofke, SUNY Buffalo

Dr. Ronald M. Levy, Rutgers University

Dr. Ioan Kosztin, University of Missouri-Columbia

Textbooks:

D. Frenkel and B. Smit, “Understanding Molecular Simulation: From Algorithm to Applications”.

Leach, “Molecular Modeling: Principles and Applications”.