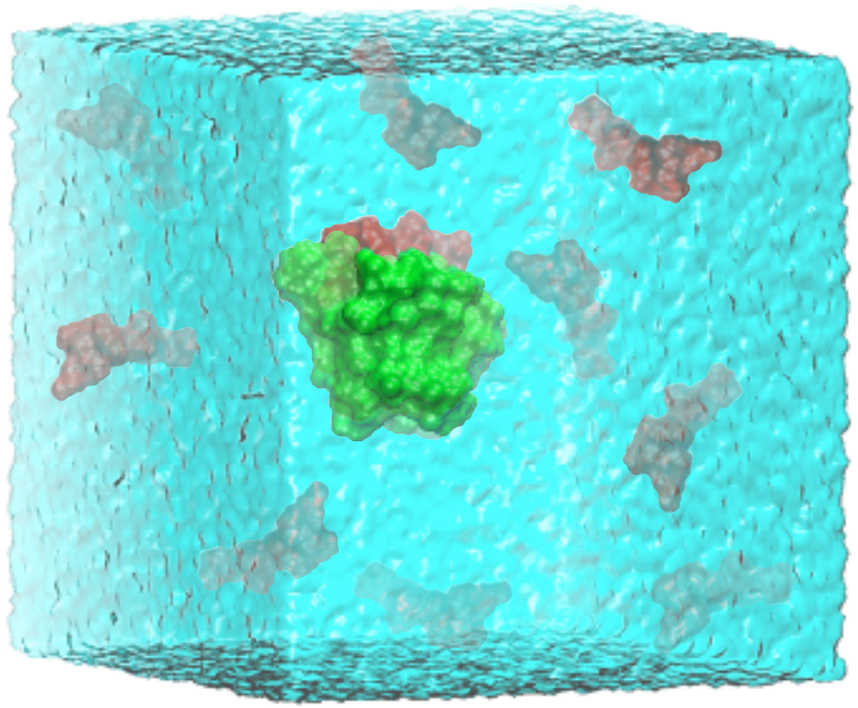
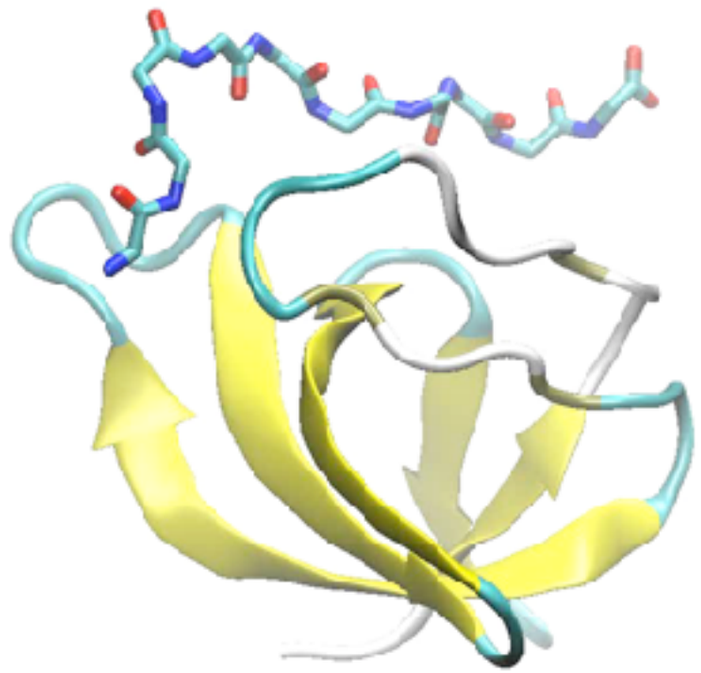
Accurate calculation of
ligand binding energy





James C. (JC) Gumbart

Georgia Institute of Technology, Atlanta

Chris Chipot

U. Illinois, Urbana & CNRS, U. Lorraine France

Computational Biophysics Workshop | DICP | July 11 2018

Outline

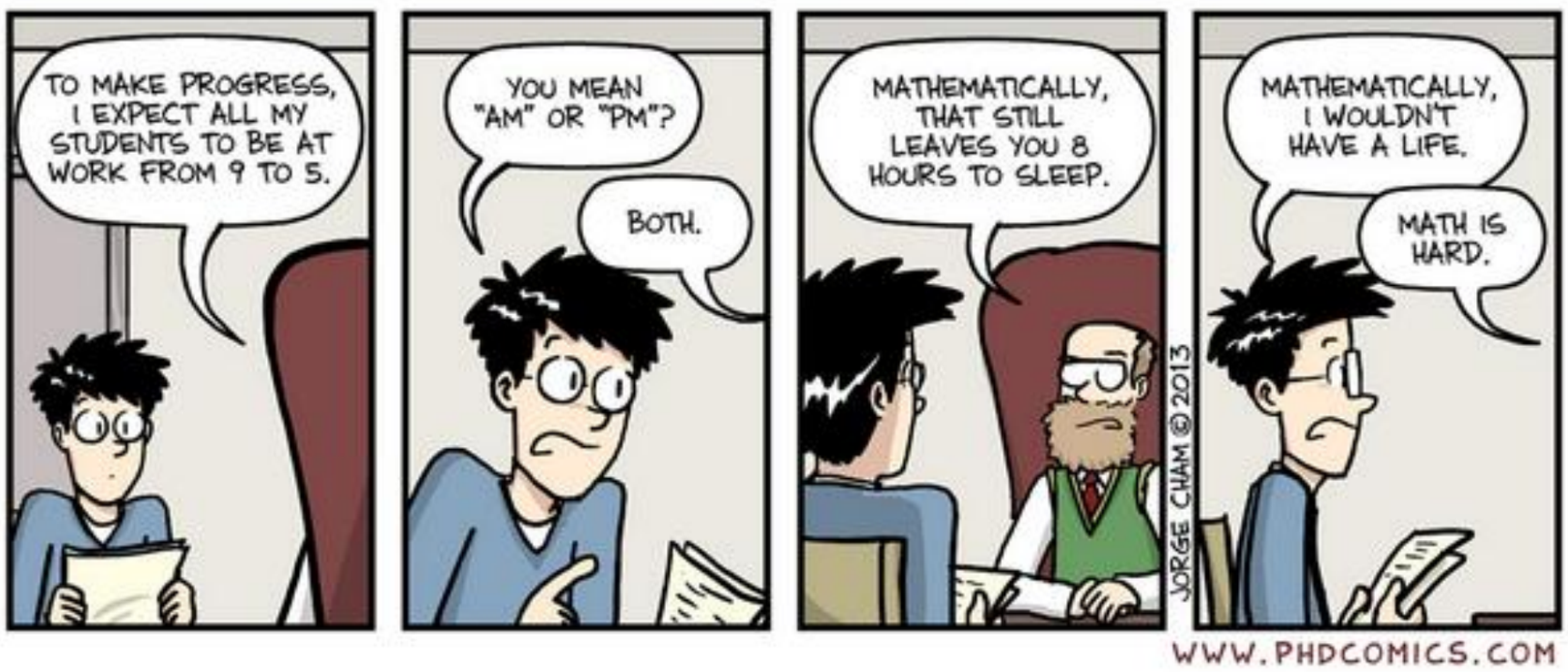
I. *What is an **absolute** binding energy?*

II. *Using restraints to reduce the sampling problem*

III. *Calculating the requisite PMFs*

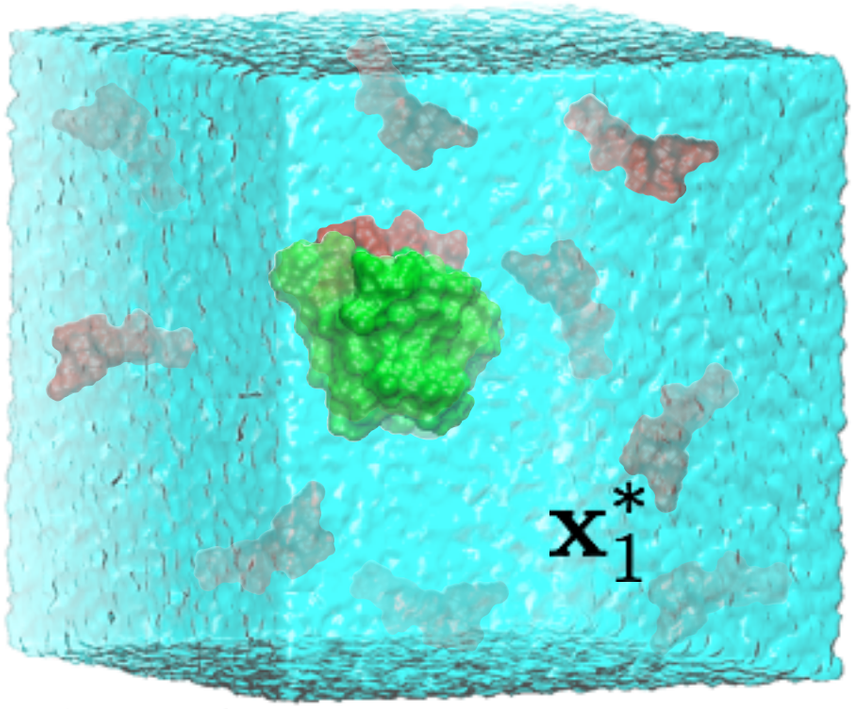
IV. *Comparing geometric with alchemical approach*

V. *Illustration with barstar-barnase binding*



http://phdcomics.com/

Challenge: Absolute binding free energies



$$\text{protein} + \text{ligand} \underset{}{\overset{K_{eq}}{\rightleftharpoons}} \text{protein : ligand}$$

$$K_{eq} = \frac{[\text{protein : ligand}]}{[\text{protein}][\text{ligand}]} = \frac{p_1[\text{protein}]_{tot}}{[\text{ligand}]\ p_0\ [\text{protein}]_{tot}} = \frac{1}{[\text{ligand}]}\ \frac{p_1}{p_0}$$

N ligands

$$K_{eq} = \frac{1}{[\text{ligand}]} \frac{N \int_{site} \mathbf{d1} \int_{bulk} \mathbf{d2} \ldots \int_{bulk} \mathbf{dN} \int \mathbf{dx}\ e^{-\beta U}}{\int_{bulk} \mathbf{d1} \int_{bulk} \mathbf{d2} \ldots \int_{bulk} \mathbf{dN} \int \mathbf{dx}\ e^{-\beta U}} = \frac{1}{[\text{ligand}]} \frac{N \int_{site} \mathbf{d1} \int \mathbf{dx}\ e^{-\beta U}}{V_{bulk} \int_{bulk} \mathbf{d1}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathbf{dx}\ e^{-\beta U}}$$

$p_i \propto Z_i$

1, 2,…,N *refer to each ligand*
1 is bound (site) in num.,
unbound (bulk) in denom.

*Energy does not depend on position of ligand when unbound (**bulk is isotropic**), so can pick out a specific point* **x**1* *and hold it there*

$$[\text{ligand}] = N/V_{bulk}$$

$$K_{eq} = \frac{\int_{site} \mathbf{d1} \int \mathbf{dx}\ e^{-\beta U}}{\int_{bulk} \mathbf{d1}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathbf{dx}\ e^{-\beta U}}$$

$$\Delta G^0 = -kT \ln(K_{eq} C^\circ) \qquad C^\circ = 1/1661 \text{Å}^3$$

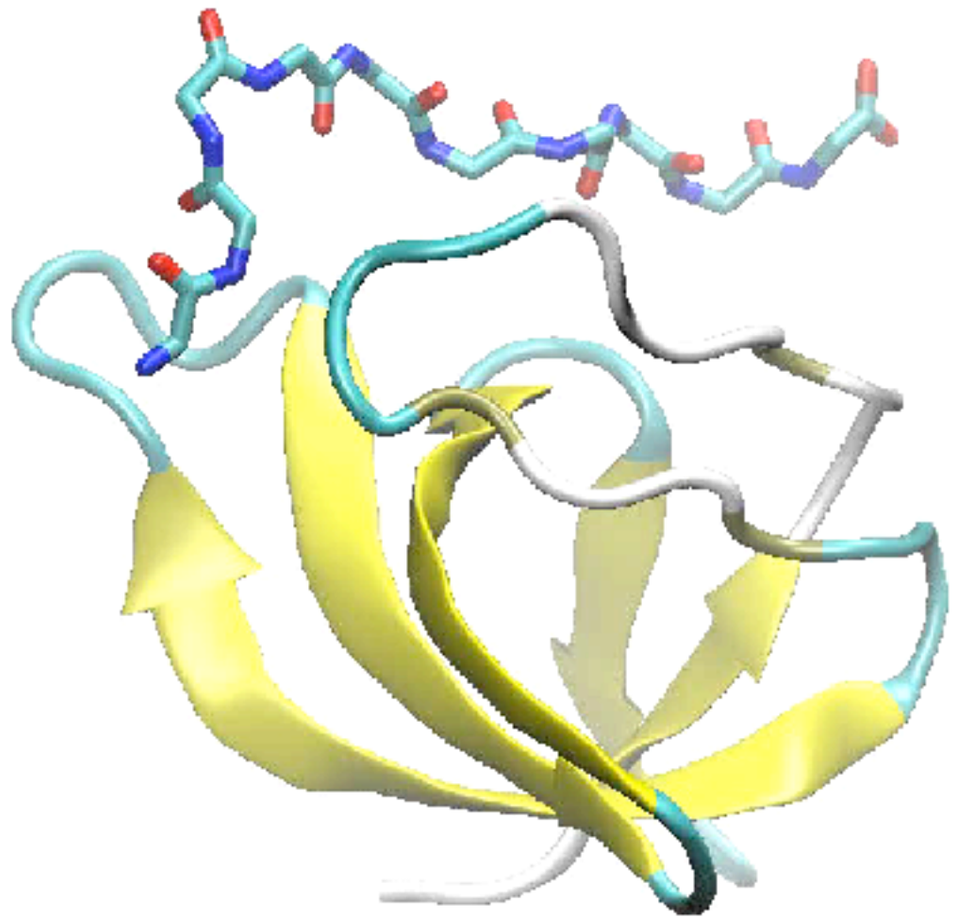
C° is the standard concentration of 1 M → **binding free energies are concentration dependent!**

Illustration using Abl SH3 domain

A well known and conserved domain of Abl kinase

Chosen ligand: APSYSPPPPP
(flexible!)

designed to bind with high affinity



peptide, so doesn't require novel parametrization

ΔG^0 = **-7.94 kcal/mol (exp)**

MM/PBSA estimate: **-2.6 kcal/mol** !

Pisabarro, M. T.; Serrano, L. *Biochemistry* **1996**, 35, 10634-10640
Hou, T. et al. *PLoS Comput. Biol*. **2006**, 2, 0046-0055

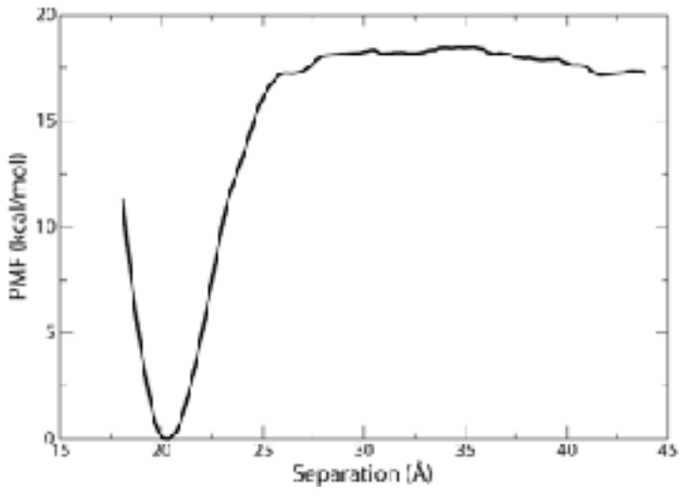
How to get K_{eq} and ΔG?

Woo, H. J.; Roux, B. Proc. Natl. Acad. Sci. USA, **2005**, 102, 6825-6830

$$K_{\mathrm{eq}} = \frac{\displaystyle\int_{\mathrm{site}} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\ \mathrm{e}^{-\beta U}}{\displaystyle\int_{\mathrm{bulk}} \mathrm{d}\mathbf{1}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{d}\mathbf{x}\ \mathrm{e}^{-\beta U}}$$

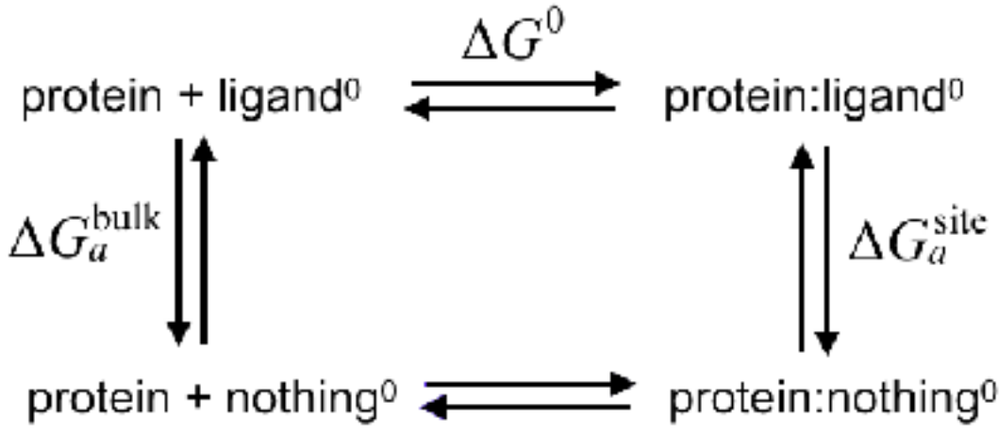
geometrical route

Forcibly separate the ligand from the protein and calculate a PMF



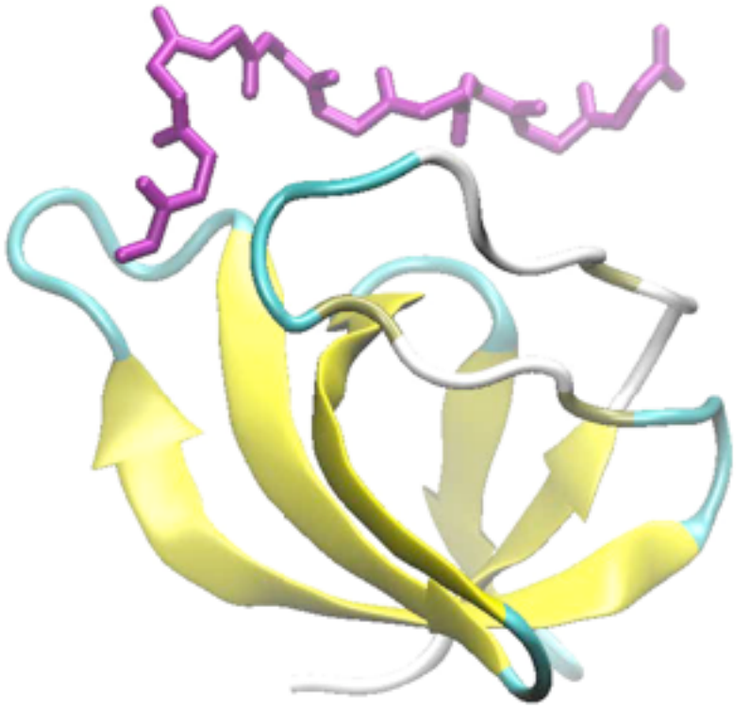
alchemical route

Make the ligand vanish from the binding site and from bulk water

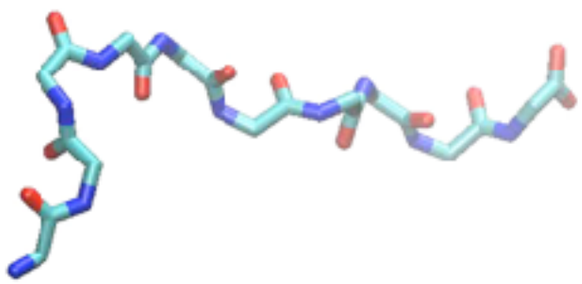


Both approaches suffer major sampling deficiencies when used on their own!!!

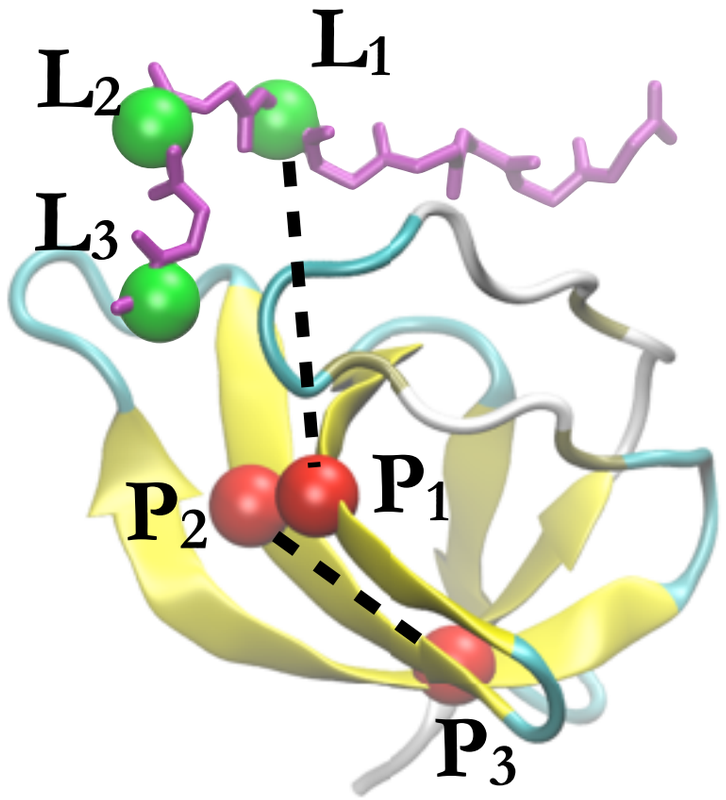
Overcoming sampling issues with restraints



Bound state RMSD restrained



Free state RMSD restrained



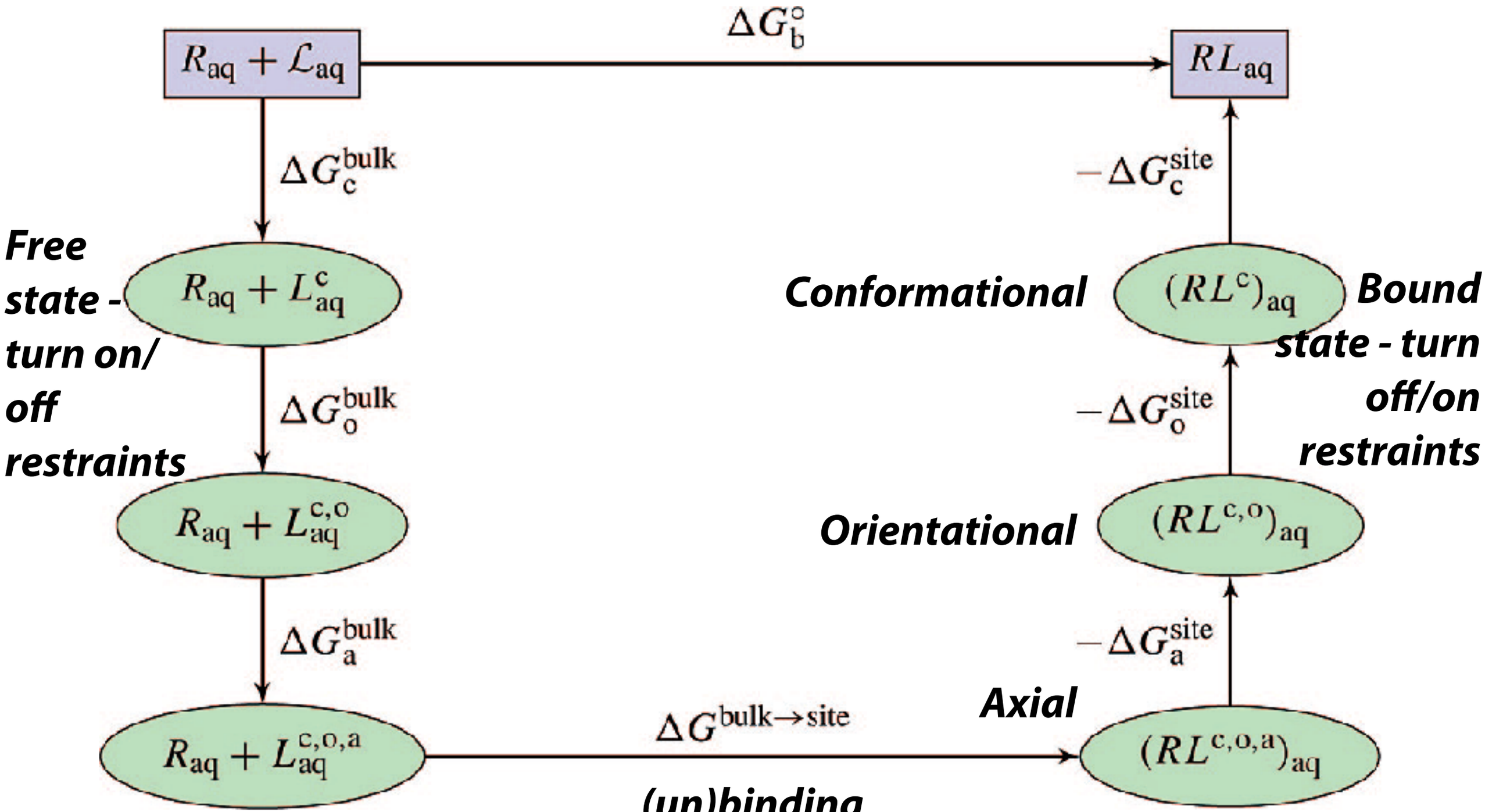
Assorted spatial/rotational restraints

-Design set of restraints to reduce conformational space needed to be sampled

-Contributions of each restraint to free energy need to be rigorously computed

Remember! Biasing is okay as long as we can unbias

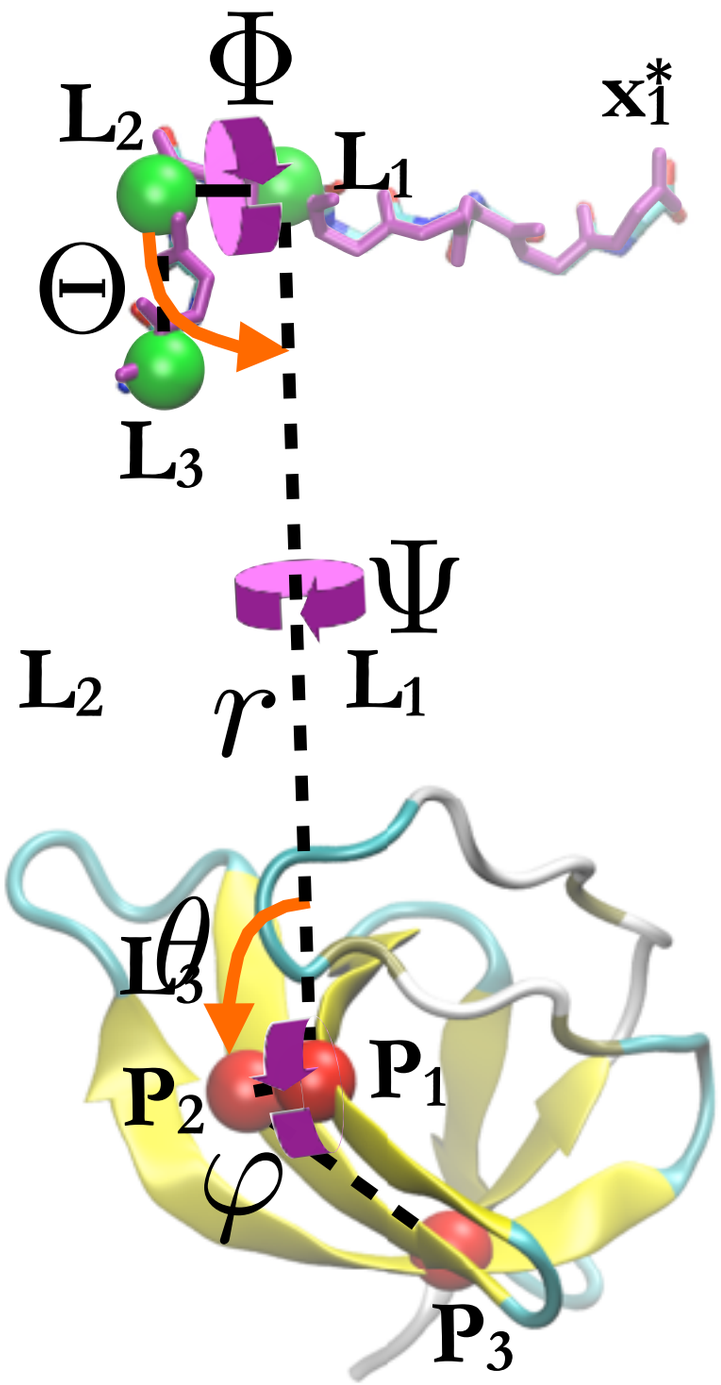
Overcoming sampling issues with restraints



Schematic of process

From: Deng and Roux. (2009)
J. Phys. Chem. **113**: 2234-2246.

Binding free energy (geometrical route)



$$K_{eq} = \frac{\int_{site} d\mathbf{1} \int d\mathbf{x}\, e^{-\beta U}}{\int_{site} d\mathbf{1} \int d\mathbf{x}\, e^{-\beta(U+u_c)}}$$

$$\times \frac{\int_{site} d\mathbf{1} \int d\mathbf{x}\, e^{-\beta(U+u_c)}}{\int_{site} d\mathbf{1} \int d\mathbf{x}\, e^{-\beta(U+u_c+u_\Theta)}}$$

$$\times \frac{\int_{site} d\mathbf{1} \int d\mathbf{x}\, e^{-\beta(U+u_c+u_\Theta)}}{\int_{site} d\mathbf{1} \int d\mathbf{x}\, e^{-\beta(U+u_c+u_\Theta+u_\Phi)}}$$

$$\times \frac{\int_{site} d\mathbf{1} \int d\mathbf{x}\, e^{-\beta(U+u_c+u_\Theta+u_\Phi)}}{\int_{site} d\mathbf{1} \int d\mathbf{x}\, e^{-\beta(U+u_c+u_\Theta+u_\Phi+u_\Psi)}}$$

$$\times \frac{\int_{site} d\mathbf{1} \int d\mathbf{x}\, e^{-\beta(U+u_c+u_o)}}{\int_{site} d\mathbf{1} \int d\mathbf{x}\, e^{-\beta(U+u_c+u_o+u_\theta)}}$$

$$\times \frac{\int_{site} d\mathbf{1} \int d\mathbf{x}\, e^{-\beta(U+u_c+u_o+u_\theta)}}{\int_{site} d\mathbf{1} \int d\mathbf{x}\, e^{-\beta(U+u_c+u_o+u_\theta+u_\varphi)}}$$

$$\times \frac{\int_{site} d\mathbf{1} \int d\mathbf{x}\, e^{-\beta(U+u_c+u_o+u_p)}}{\int_{bulk} d\mathbf{1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int d\mathbf{x}\, e^{-\beta(U+u_c+u_o)}}$$

$$\times \frac{\int_{bulk} d\mathbf{1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int d\mathbf{x}\, e^{-\beta(U+u_c+u_\Theta+u_\Phi+u_\Psi)}}{\int_{bulk} d\mathbf{1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int d\mathbf{x}\, e^{-\beta(U+u_c+u_\Theta+u_\Phi)}}$$

$$\times \frac{\int_{bulk} d\mathbf{1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int d\mathbf{x}\, e^{-\beta(U+u_c+u_\Theta+u_\Phi)}}{\int_{bulk} d\mathbf{1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int d\mathbf{x}\, e^{-\beta(U+u_c+u_\Theta)}}$$

$$\times \frac{\int_{bulk} d\mathbf{1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int d\mathbf{x}\, e^{-\beta(U+u_c+u_\Theta)}}{\int_{bulk} d\mathbf{1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int d\mathbf{x}\, e^{-\beta(U+u_c)}}$$

$$\times \frac{\int_{bulk} d\mathbf{1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int d\mathbf{x}\, e^{-\beta(U+u_c)}}{\int_{bulk} d\mathbf{1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int d\mathbf{x}\, e^{-\beta U}}$$

Yu, Y. B. et al. (2001) *Biophys. J.*, **81**:1632-1642.

Woo, H. J.; Roux, B. (2005) *Proc. Natl. Acad. Sci. USA*, **102**:6825-6830.

Maffeo, C., Luan, B., Aksimentiev, A. (2012) *Nucl. Acids Res.* **40**:3812-3821.

How to evaluate all of these integrals?

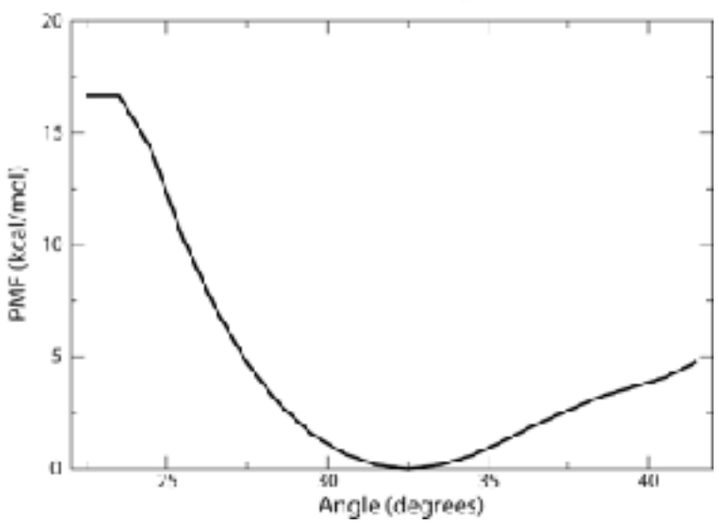
$$\frac{\int_{\text{site}} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\ \mathrm{e}^{-\beta(U+u_c+u_\Theta+u_\Phi)}}{\int_{\text{site}} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\ \mathrm{e}^{-\beta(U+u_c+u_\Theta+u_\Phi+u_\Psi)}} =$$

ratio of integrals can be related to a free energy

$$\frac{\int_{\text{site}} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\ \mathrm{e}^{-\beta(U+u_c+u_\Theta+u_\Phi)}}{\int_{\text{site}} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\ (e^{-\beta u_\Psi})\mathrm{e}^{-\beta(U+u_c+u_\Theta+u_\Phi)}} = \frac{1}{\mathrm{e}^{-\beta\Delta G_\Psi^{\text{site}}}} = \mathrm{e}^{+\beta\Delta G_\Psi^{\text{site}}}$$

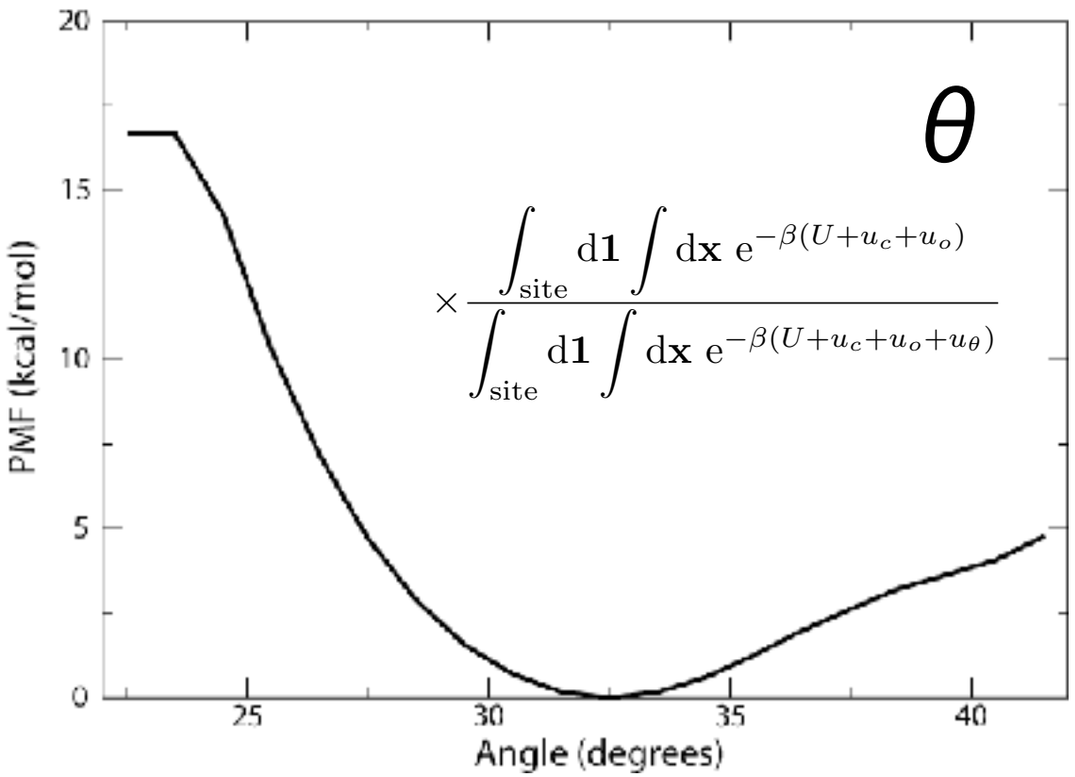
$$\mathrm{e}^{+\beta\Delta G_\Psi^{\text{site}}} = \frac{\int \mathrm{d}\Psi \mathrm{e}^{-\beta[w_{\text{site}}(\Psi)]}}{\int \mathrm{d}\Psi \mathrm{e}^{-\beta[w_{\text{site}}(\Psi)+u_\Psi]}}$$

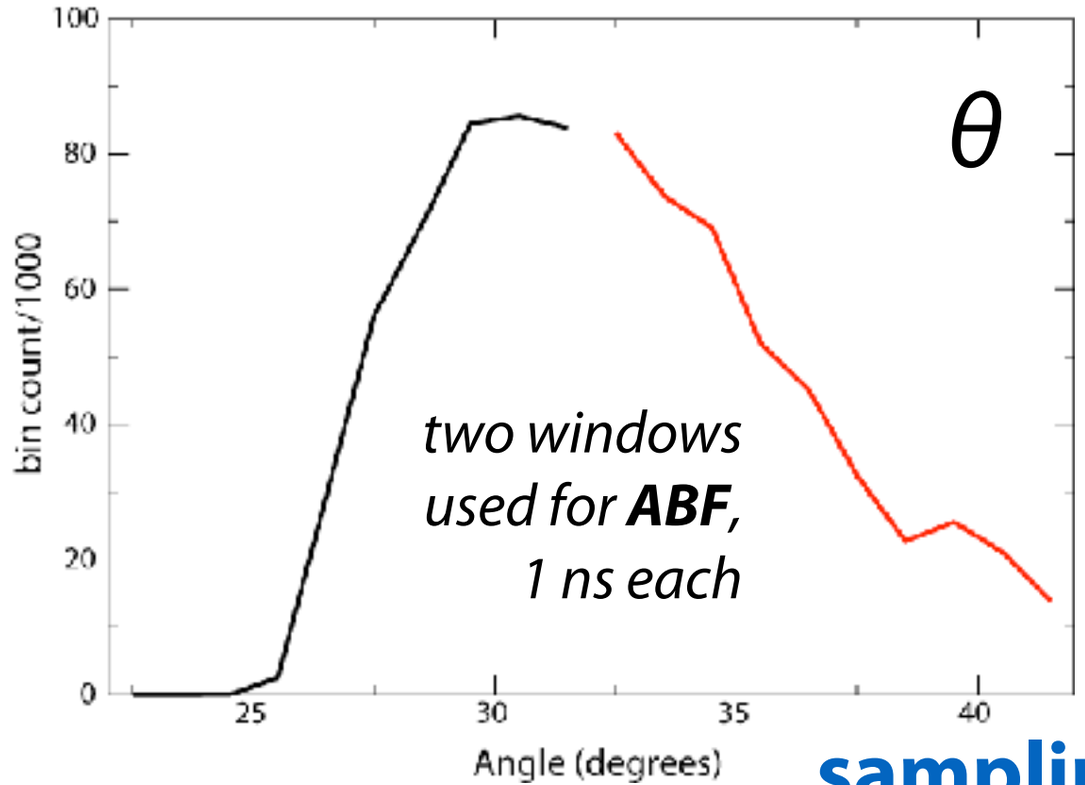
Potential of mean force, $w_{\text{site}}(\Psi)$, encapsulates all degrees of freedom



In practice, one determines the PMFs successively and then integrates them as prescribed above

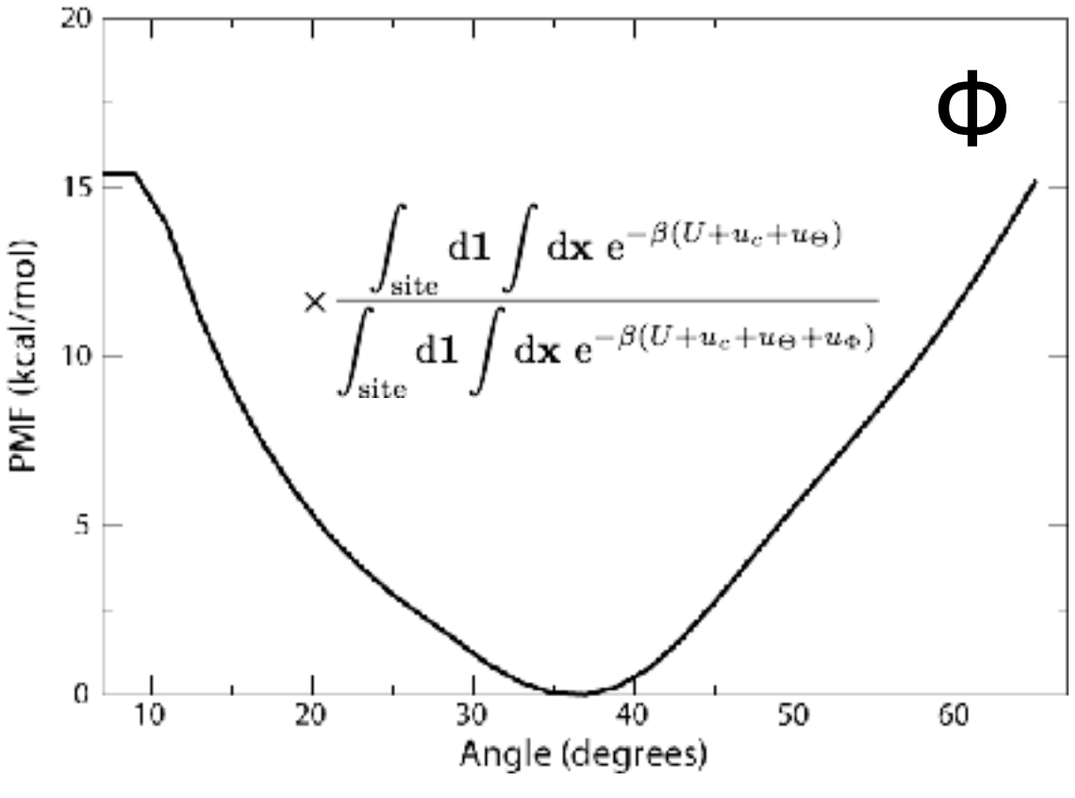
Many PMFs are very straightforward

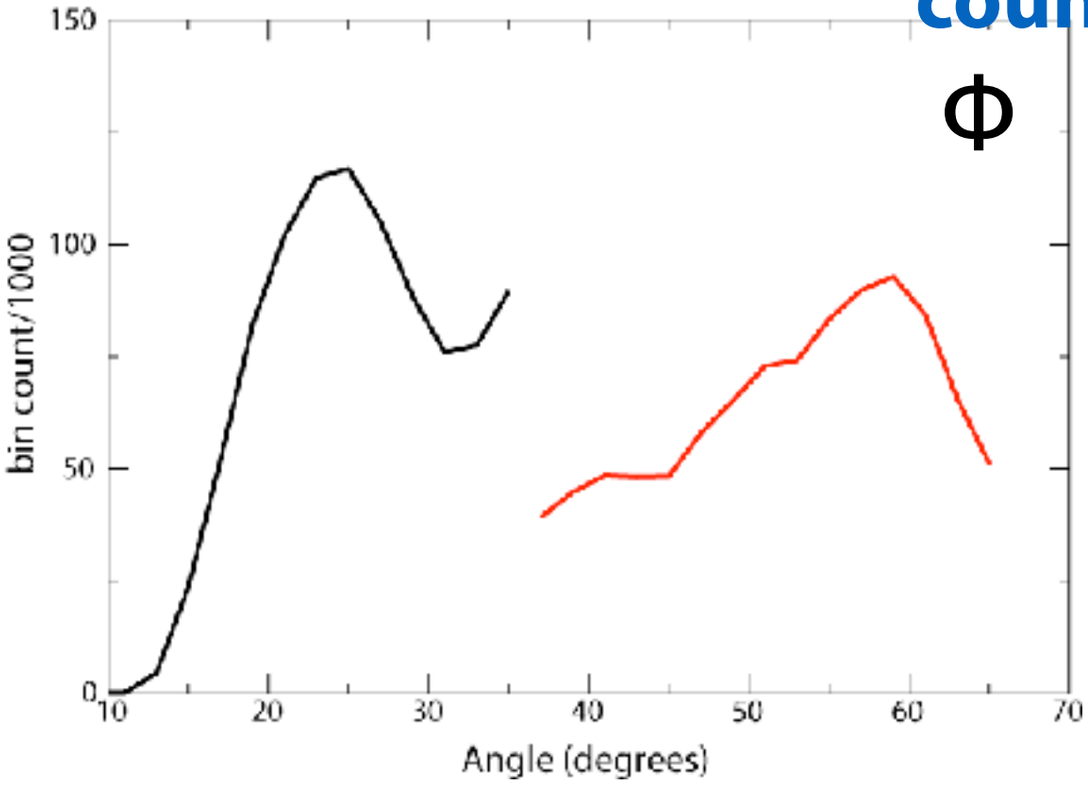




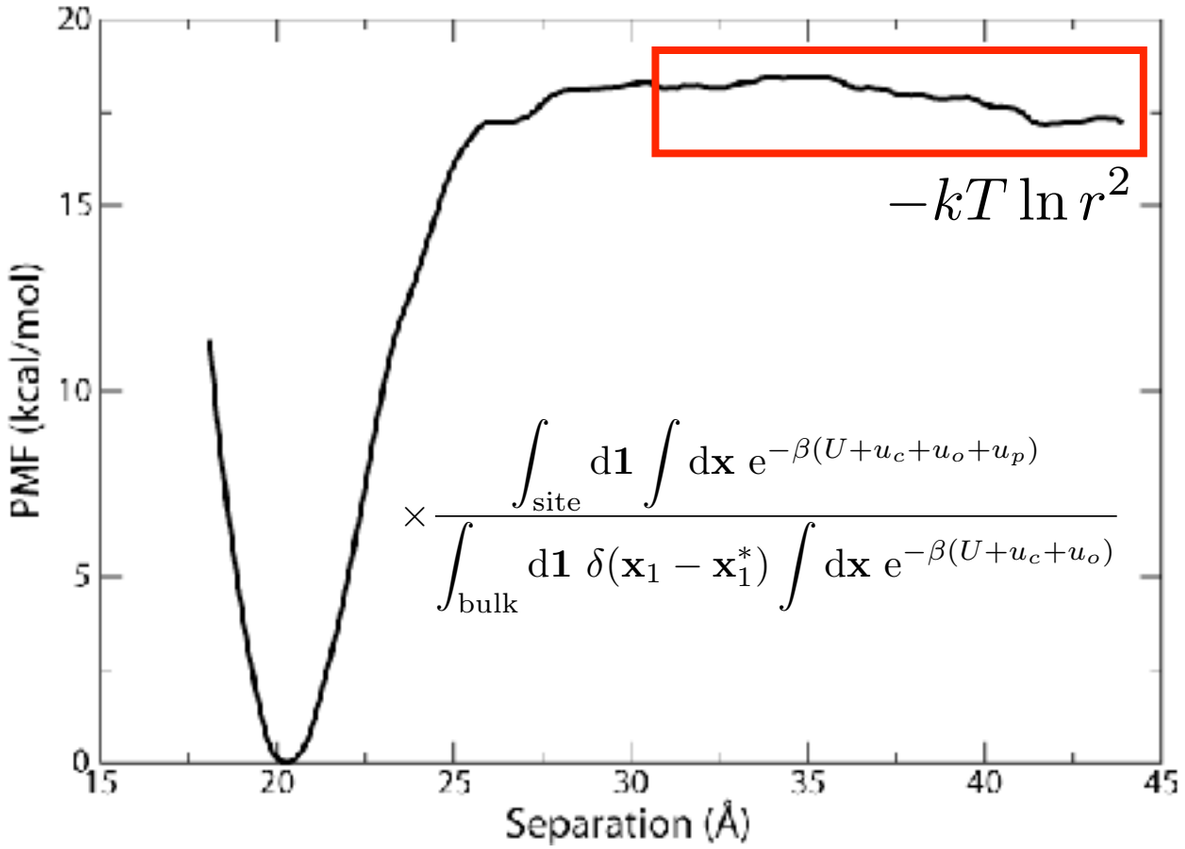
PMFs

sampling counts

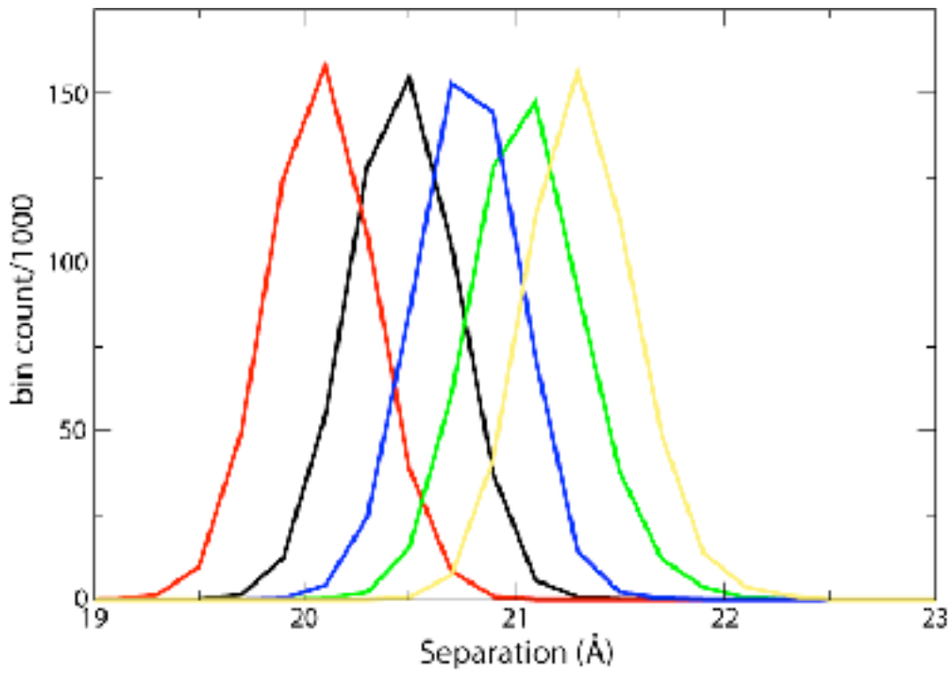




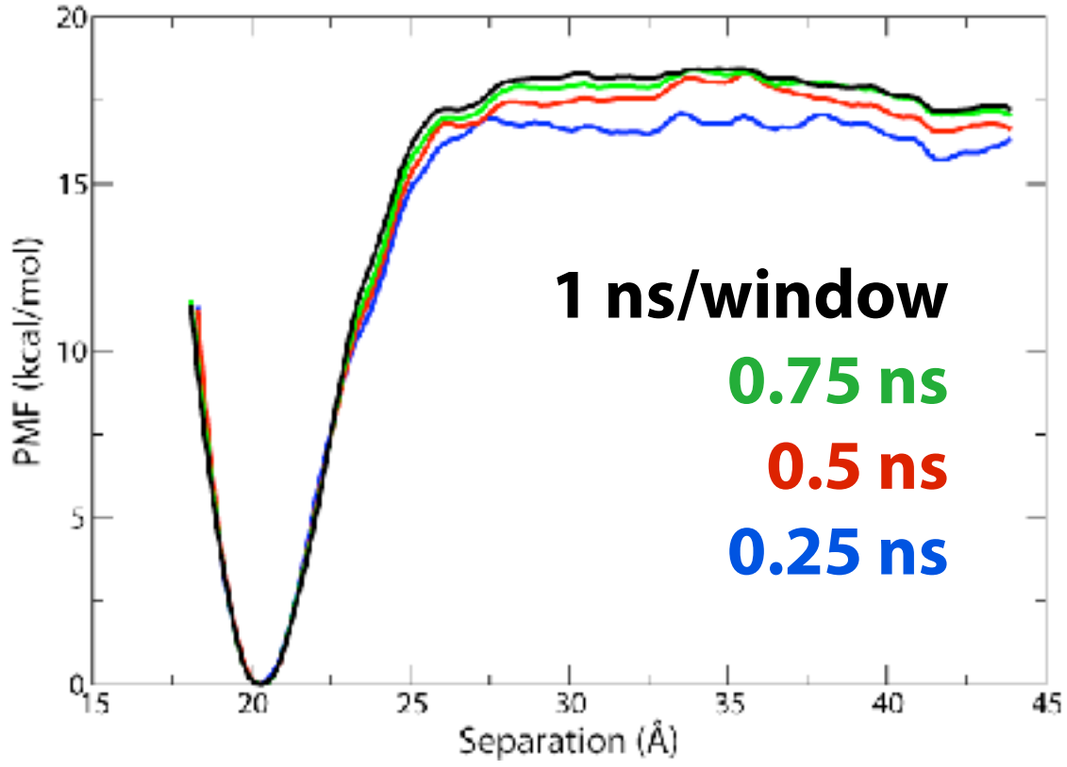
Separation PMF from umbrella sampling



entropic decay despite no interactions

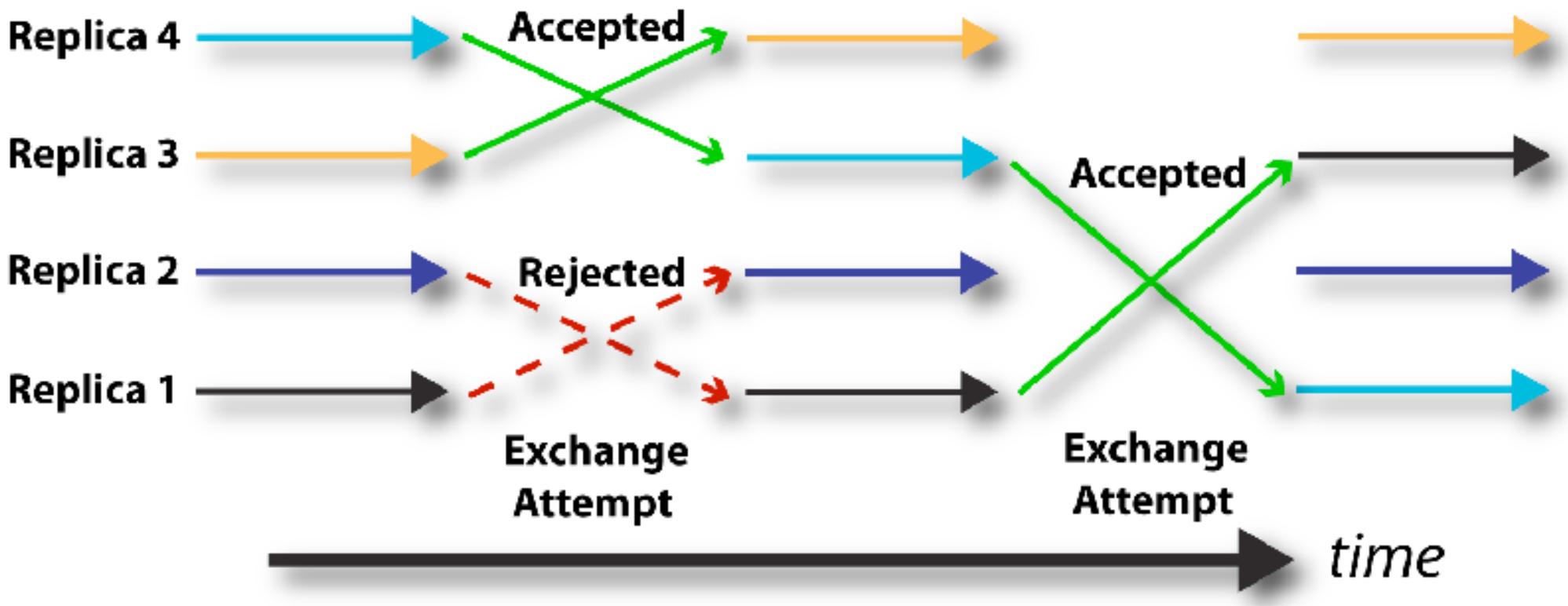


37 windows used, spaced
0.5 - 1 Å apart
-histograms are overlapping



PMF was already converged
within ~20 ns

Replica-exchange umbrella sampling (REUS)



-helps to circumvent limitations in **US** by exchanging coordinates periodically between different windows

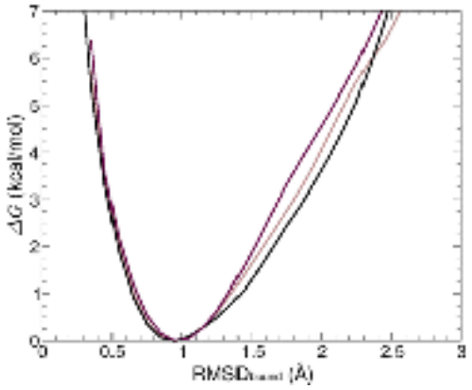
-exchanges accepted with some probability: $\min(1, e^{-\Delta E/kT})$

where $\Delta E = (w_i(\xi_j) - w_i(\xi_i)) + (w_j(\xi_i) - w_j(\xi_j))$

(swapped) (original) (swapped) (original)

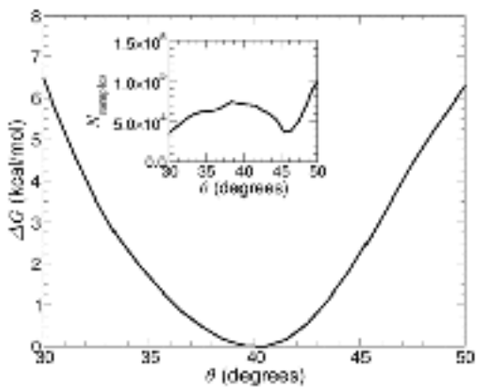
See tutorial [Methods for Calculating Potentials of Mean Force]

What you get in the end (a big mess!)



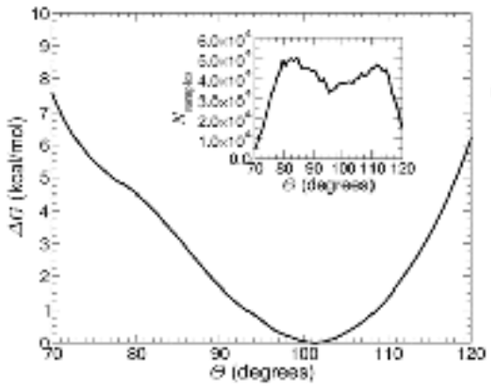
$$e^{-\beta\Delta G_c^{\text{site}}} = \frac{\int_{\text{site}} d\mathbf{l} \int d\mathbf{x}\, e^{-\beta(U+u_c)}}{\int_{\text{site}} d\mathbf{l} \int d\mathbf{x}\, e^{-\beta U}} = \langle e^{-\beta u_c} \rangle_{\text{site},U}$$

$$\Delta G_c^{\text{site}} = +3.5 \text{ kcal/mol}$$



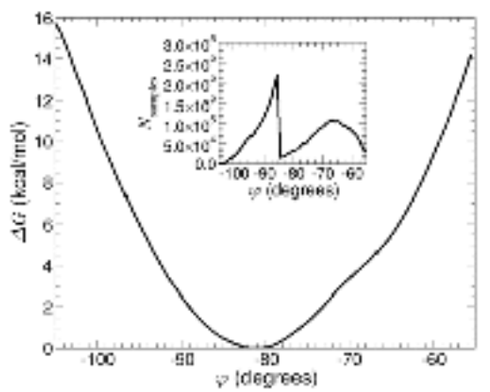
$$e^{-\beta\Delta G_\theta^{\text{site}}} = \frac{\int_{\text{site}} d\mathbf{l} \int d\mathbf{x}\, e^{-\beta(U+u_c+u_o+u_\theta)}}{\int_{\text{site}} d\mathbf{l} \int d\mathbf{x}\, e^{-\beta(U+u_c+u_o)}} = \langle e^{-\beta u_\theta} \rangle_{\text{site},U+u_c+u_o}$$

$$\Delta G_\theta^{\text{site}} = +0.1 \text{ kcal/mol}$$



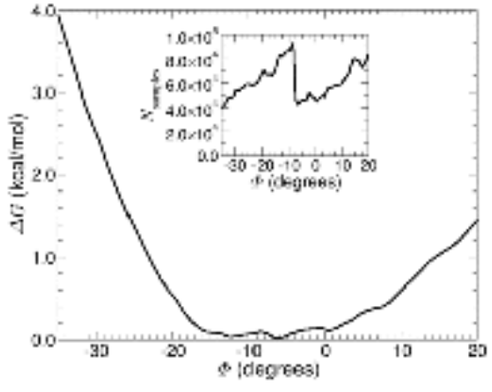
$$e^{-\beta\Delta G_\Theta^{\text{site}}} = \frac{\int_{\text{site}} d\mathbf{l} \int d\mathbf{x}\, e^{-\beta(U+u_c+u_\Theta)}}{\int_{\text{site}} d\mathbf{l} \int d\mathbf{x}\, e^{-\beta(U+u_c)}} = \langle e^{-\beta u_\Theta} \rangle_{\text{site},U+u_c}$$

$$\Delta G_\Theta^{\text{site}} = +0.1 \text{ kcal/mol}$$



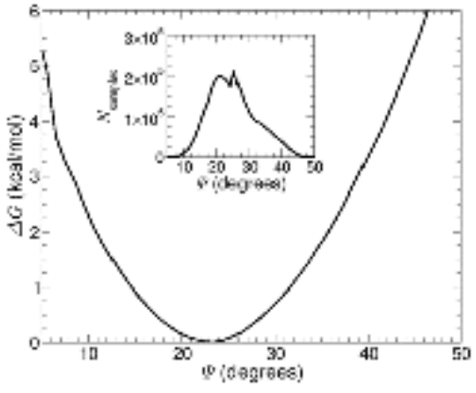
$$e^{-\beta\Delta G_\varphi^{\text{site}}} = \frac{\int_{\text{site}} d\mathbf{l} \int d\mathbf{x}\, e^{-\beta(U+u_c+u_o+u_\theta+u_\varphi)}}{\int_{\text{site}} d\mathbf{l} \int d\mathbf{x}\, e^{-\beta(U+u_c+u_o+u_\theta)}} = \langle e^{-\beta u_\varphi} \rangle_{\text{site},U+u_c+u_o+u_\theta}$$

$$\Delta G_\varphi^{\text{site}} = +0.2 \text{ kcal/mol}$$



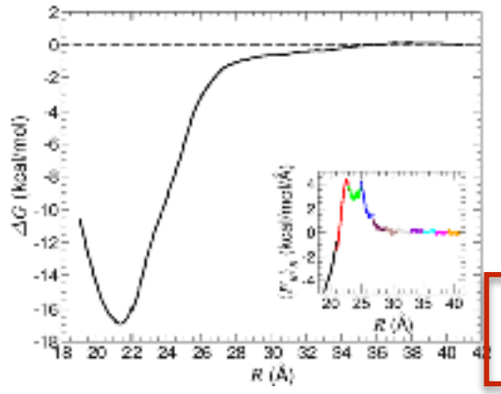
$$e^{-\beta\Delta G_\Phi^{\text{site}}} = \frac{\int_{\text{site}} d\mathbf{l} \int d\mathbf{x}\, e^{-\beta(U+u_c+u_\Theta+u_\Phi)}}{\int_{\text{site}} d\mathbf{l} \int d\mathbf{x}\, e^{-\beta(U+u_c+u_\Theta)}} = \langle e^{-\beta u_\Phi} \rangle_{\text{site},U+u_c+u_\Theta}$$

$$\Delta G_\Phi^{\text{site}} = +0.1 \text{ kcal/mol}$$



$$e^{-\beta\Delta G_\Psi^{\text{site}}} = \frac{\int_{\text{site}} d\mathbf{l} \int d\mathbf{x}\, e^{-\beta(U+u_c+u_\Theta+u_\Phi+u_\Psi)}}{\int_{\text{site}} d\mathbf{l} \int d\mathbf{x}\, e^{-\beta(U+u_c+u_\Theta+u_\Phi)}} = \langle e^{-\beta u_\Phi} \rangle_{\text{site},U+u_c+u_\Theta+u_\Phi}$$

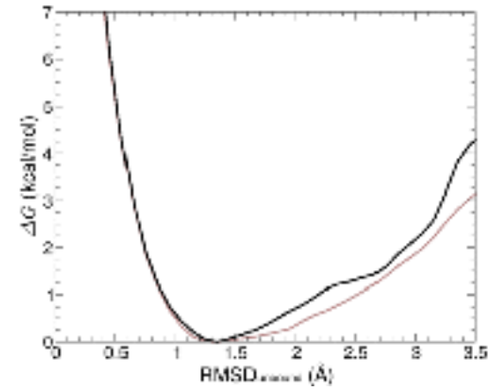
$$\Delta G_\Psi^{\text{site}} = +0.5 \text{ kcal/mol}$$



$$S^* = r^{*2} \int_0^{\pi} d\theta_1 \sin\theta_1 \int_0^{2\pi} d\phi_1\, e^{-\beta u_p} = +12.6 \text{ Å}^2$$

$$I^* = \int_{\text{site}} d\mathbf{x}_1\, e^{-\beta[w(\mathbf{x}_1)-w(\mathbf{x}_1^*)]} = +6.1 \times 10^{12} \text{ Å}$$

$$\Delta G_r = -\frac{1}{\beta} \ln(S^* I^*) = -14.6 \text{ kcal/mol}$$



$$e^{-\beta\Delta G_c^{\text{bulk}}} = \frac{\int_{\text{bulk}} d\mathbf{l}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int d\mathbf{x}\, e^{-\beta(U+u_c)}}{\int_{\text{bulk}} d\mathbf{l}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int d\mathbf{x}\, e^{-\beta U}} = \langle e^{-\beta u_c} \rangle_{\text{bulk},U}$$

$$\Delta G_c^{\text{bulk}} = +5.4 \text{ kcal/mol}$$

Back to the Abl kinase story...

$$K_{\text{eq}} = \frac{\int_{\text{site}} \mathrm{d1} \int \mathrm{dx}\ \mathrm{e}^{-\beta U}}{\int_{\text{site}} \mathrm{d1} \int \mathrm{dx}\ \mathrm{e}^{-\beta(U+u_c)}}$$

$$\times \frac{\int_{\text{site}} \mathrm{d1} \int \mathrm{dx}\ \mathrm{e}^{-\beta(U+u_c)}}{\int_{\text{site}} \mathrm{d1} \int \mathrm{dx}\ \mathrm{e}^{-\beta(U+u_c+u_\Theta)}}$$

$$\times \frac{\int_{\text{site}} \mathrm{d1} \int \mathrm{dx}\ \mathrm{e}^{-\beta(U+u_c+u_\Theta)}}{\int_{\text{site}} \mathrm{d1} \int \mathrm{dx}\ \mathrm{e}^{-\beta(U+u_c+u_\Theta+u_\Phi)}}$$

$$\times \frac{\int_{\text{site}} \mathrm{d1} \int \mathrm{dx}\ \mathrm{e}^{-\beta(U+u_c+u_\Theta+u_\Phi)}}{\int_{\text{site}} \mathrm{d1} \int \mathrm{dx}\ \mathrm{e}^{-\beta(U+u_c+u_\Theta+u_\Phi+u_\Psi)}}$$

$$\times \frac{\int_{\text{site}} \mathrm{d1} \int \mathrm{dx}\ \mathrm{e}^{-\beta(U+u_c+u_o)}}{\int_{\text{site}} \mathrm{d1} \int \mathrm{dx}\ \mathrm{e}^{-\beta(U+u_c+u_o+u_\theta)}}$$

$$\times \frac{\int_{\text{site}} \mathrm{d1} \int \mathrm{dx}\ \mathrm{e}^{-\beta(U+u_c+u_o+u_\theta)}}{\int_{\text{site}} \mathrm{d1} \int \mathrm{dx}\ \mathrm{e}^{-\beta(U+u_c+u_o+u_\theta+u_\varphi)}}$$

$$\times \frac{\int_{\text{site}} \mathrm{d1} \int \mathrm{dx}\ \mathrm{e}^{-\beta(U+u_c+u_o+u_p)}}{\int_{\text{bulk}} \mathrm{d1}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\ \mathrm{e}^{-\beta(U+u_c+u_o)}}$$

$$\times \frac{\int_{\text{bulk}} \mathrm{d1}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\ \mathrm{e}^{-\beta(U+u_c+u_\Theta+u_\Phi+u_\Psi)}}{\int_{\text{bulk}} \mathrm{d1}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\ \mathrm{e}^{-\beta(U+u_c+u_\Theta+u_\Phi)}}$$

$$\times \frac{\int_{\text{bulk}} \mathrm{d1}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\ \mathrm{e}^{-\beta(U+u_c+u_\Theta+u_\Phi)}}{\int_{\text{bulk}} \mathrm{d1}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\ \mathrm{e}^{-\beta(U+u_c+u_\Theta)}}$$

$$\times \frac{\int_{\text{bulk}} \mathrm{d1}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\ \mathrm{e}^{-\beta(U+u_c+u_\Theta)}}{\int_{\text{bulk}} \mathrm{d1}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\ \mathrm{e}^{-\beta(U+u_c)}}$$

$$\times \frac{\int_{\text{bulk}} \mathrm{d1}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\ \mathrm{e}^{-\beta(U+u_c)}}{\int_{\text{bulk}} \mathrm{d1}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\ \mathrm{e}^{-\beta U}}$$

~30 ns $\Delta G_c^{\text{site}} = 3.52$ kcal/mol

6 ns $\Delta G_o^{\text{site}} = 0.71$ kcal/mol

4 ns $\Delta G_a^{\text{site}} = 0.20$ kcal/mol

37 ns $\Delta G_r^{\text{sep}} = -14.47$ kcal/mol

(analytical) $\Delta G_o^{\text{bulk}} = 5.77$ kcal/mol

30 ns $\Delta G_c^{\text{bulk}} = 5.43$ kcal/mol

$$\Delta G^o = (\Delta G_c^{\text{bulk}} - \Delta G_c^{\text{site}}) + (\Delta G_o^{\text{bulk}} - \Delta G_o^{\text{site}})$$

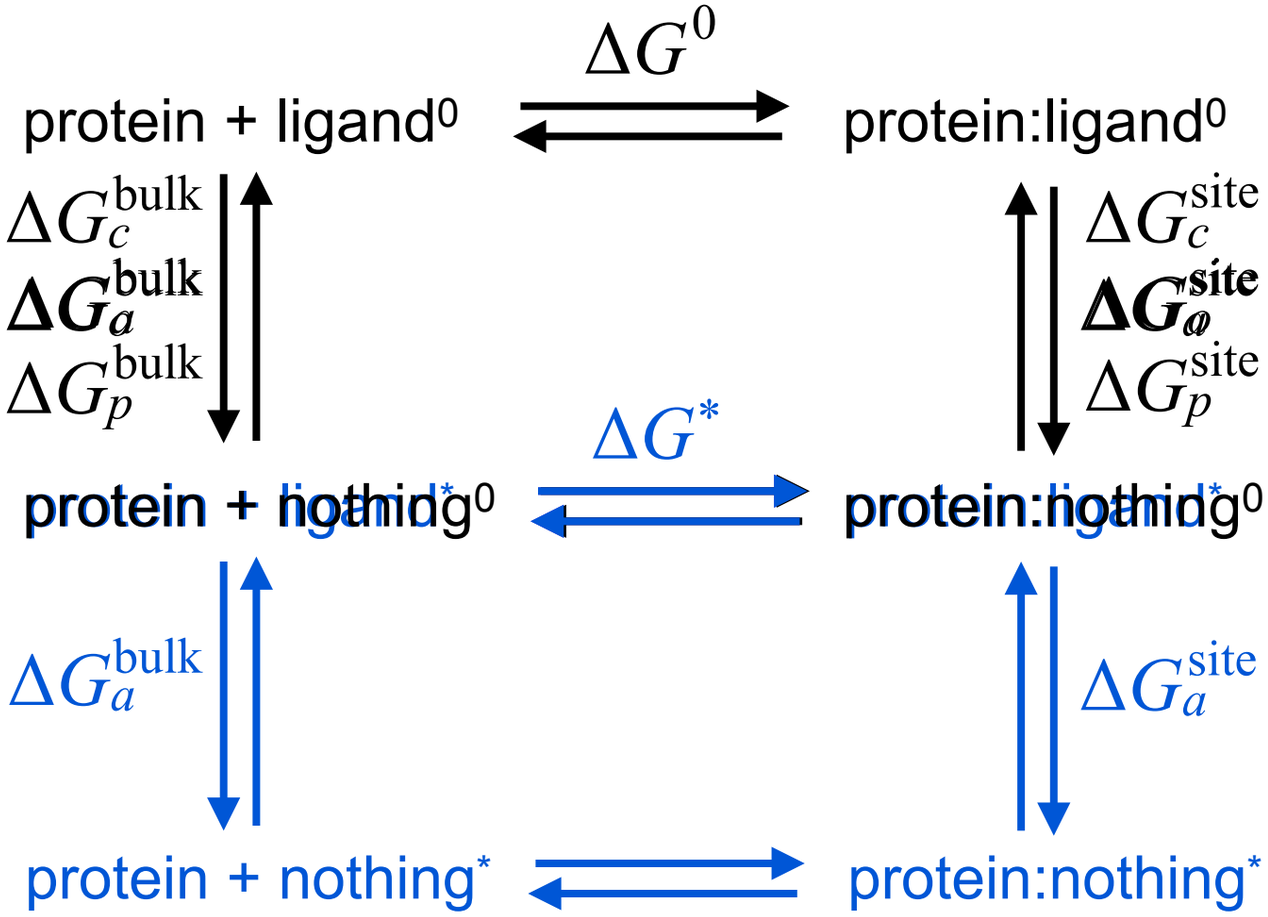
$$+\Delta G_r^{\text{sep}} - \Delta G_a^{\text{site}} = -7.7 \text{ kcal/mol}$$

~ 120 ns required

ΔG^0 = -7.94 kcal/mol (**exp**) **Agreement within 0.25 kcal/mol!**

There's more than one way to…

Can use **FEP** to (de)couple the ligand to the binding site of the protein



"**Floating ligand**" problem

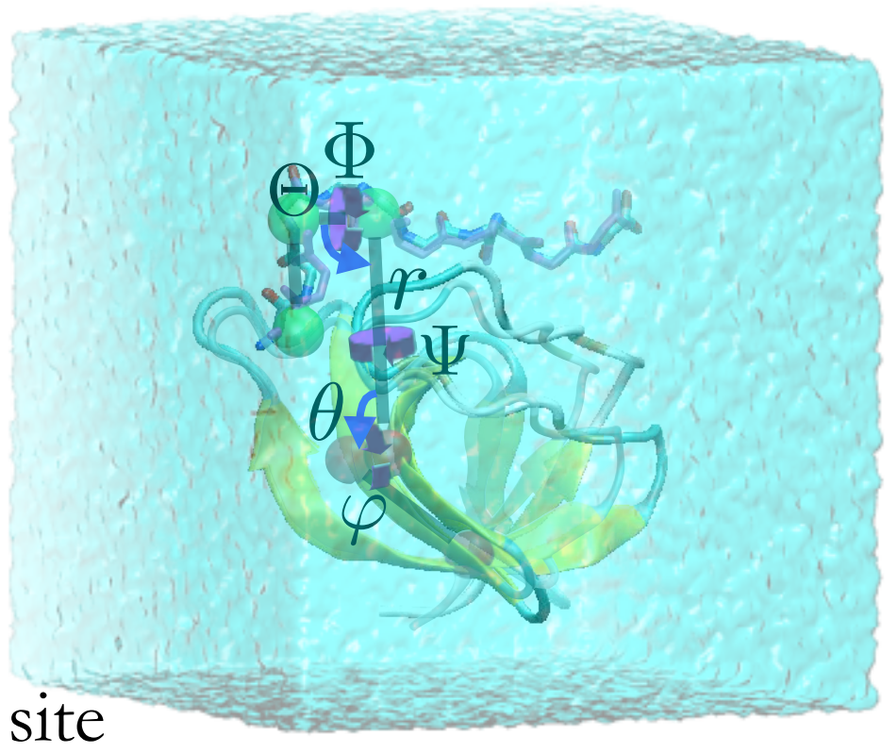
Avoided through definition of a set of **restraints**

Follow a formalism akin to the reaction-coordinate (geometric) route

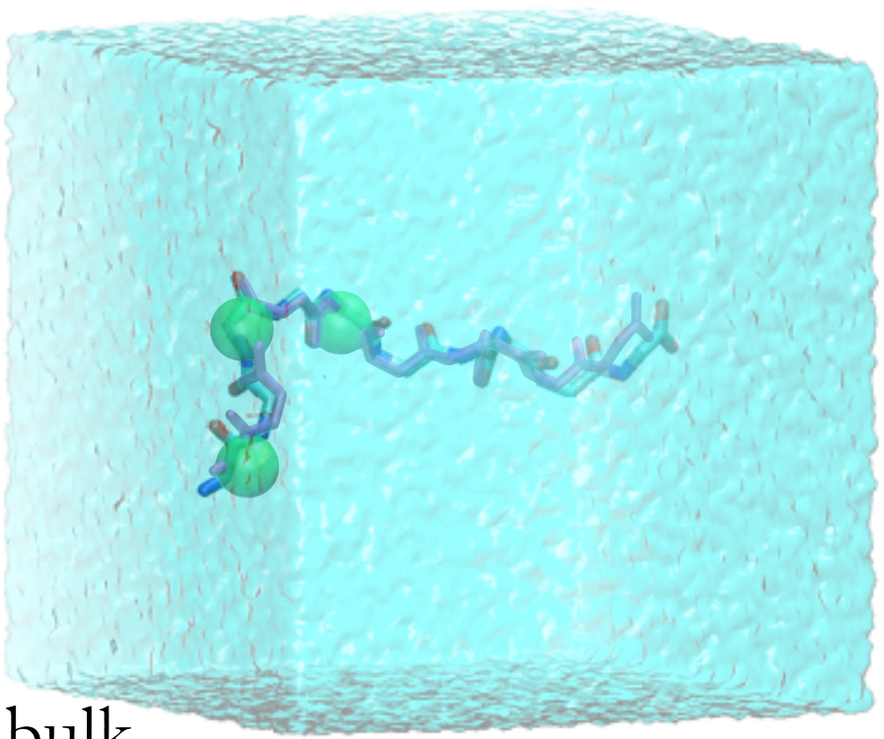
- Alchemical transformations performed bidirectionally using **FEP**
- Bennett acceptance ratio (BAR) estimator
- Free-energy contributions due to restraints measured using TI
- Most appropriate for buried ligands (no extraction pathway)

Bennett, C. H. *J. Comp. Phys.* , **1976**, *22*, 245-268 Gilson, M. K. et al. *Biophys. J.*, **1997**, *72*, 1047-1069

The alchemical (FEP) route



site



bulk

$$K_{\text{eq}} = \frac{\int_{\text{site}} \mathbf{d1} \int \mathbf{dx}\, \mathrm{e}^{-\beta U_1}}{\int_{\text{site}} \mathbf{d1} \int \mathbf{dx}\, \mathrm{e}^{-\beta(U_1+u_c)}}$$

$$\times \frac{\int_{\text{site}} \mathbf{d1} \int \mathbf{dx}\, \mathrm{e}^{-\beta(U_1+u_c)}}{\int_{\text{site}} \mathbf{d1} \int \mathbf{dx}\, \mathrm{e}^{-\beta(U_1+u_c+u_o)}}$$

$$\times \frac{\int_{\text{site}} \mathbf{d1} \int \mathbf{dx}\, \mathrm{e}^{-\beta(U_1+u_c+u_o)}}{\int_{\text{site}} \mathbf{d1} \int \mathbf{dx}\, \mathrm{e}^{-\beta(U_1+u_c+u_o+u_p)}}$$

$$\times \frac{\int_{\text{site}} \mathbf{d1} \int \mathbf{dx}\, \mathrm{e}^{-\beta(U_1+u_c+u_o+u_p)}}{\int_{\text{site}} \mathbf{d1} \int \mathbf{dx}\, \mathrm{e}^{-\beta(U_1+u_c+u_o+u_p+u_r)}}$$

$$\times \frac{\int_{\text{site}} \mathbf{d1} \int \mathbf{dx}\, \mathrm{e}^{-\beta(U_1+u_c+u_o+u_p+u_r)}}{\int_{\text{site}} \mathbf{d1} \int \mathbf{dx}\, \mathrm{e}^{-\beta(U_0+u_c+u_o+u_p+u_r)}}$$

$$\times \frac{\int_{\text{bulk}} \mathbf{d1} \int \mathbf{dx}\, \mathrm{e}^{-\beta(U_0+u_c+u_o+u_p+u_r)}}{\int_{\text{bulk}} \mathbf{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathbf{dx}\, \mathrm{e}^{-\beta(U_0+u_c+u_o)}}$$

$$\times \frac{\int_{\text{bulk}} \mathbf{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathbf{dx}\, \mathrm{e}^{-\beta(U_0+u_c+u_o)}}{\int_{\text{bulk}} \mathbf{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathbf{dx}\, \mathrm{e}^{-\beta(U_0+u_c)}}$$

$$\times \frac{\int_{\text{bulk}} \mathbf{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathbf{dx}\, \mathrm{e}^{-\beta(U_0+u_c)}}{\int_{\text{bulk}} \mathbf{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathbf{dx}\, \mathrm{e}^{-\beta(U_1+u_c)}}$$

$$\times \frac{\int_{\text{bulk}} \mathbf{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathbf{dx}\, \mathrm{e}^{-\beta(U_1+u_c)}}{\int_{\text{bulk}} \mathbf{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathbf{dx}\, \mathrm{e}^{-\beta U_1}}$$

Deng, Y.; Roux, B. *J. Phys. Chem. B* **2009**, *113*, 2234-2246

Comparison of alchemical and geometric routes

$$K_{\mathrm{eq}}^{\mathrm{PMF}} = \frac{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta U}}{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c)}} \times \frac{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c)}}{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c+u_o)}} \times \frac{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c+u_o)}}{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c+u_o+u_p)}}$$

RMSD PMF
16 ns

Θ, Φ, Ψ PMFs
12 ns

θ, ϕ PMFs
8 ns

$$\times \frac{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c+u_o+u_p)}}{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c+u_o)}}$$

r PMF
20 ns

$$\times \frac{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c+u_o)}}{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c)}}$$

RMSD PMF
60 ns

$$\times \frac{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c)}}{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta U}}$$

$$K_{\mathrm{eq}}^{\mathrm{FEP}} = \frac{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta U_1}}{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_1+u_c)}} \times \frac{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_1+u_c)}}{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_1+u_c+u_o)}} \times \frac{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_1+u_c+u_o)}}{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_1+u_c+u_o+u_p)}}$$

RMSD

Θ, Φ, Ψ

θ, ϕ

24 ns

$$\times \frac{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_1+u_c+u_o+u_p)}}{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_1+u_c+u_o+u_p+u_r)}}$$

r

$$\times \frac{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_1+u_c+u_o+u_p+u_r)}}{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_0+u_c+u_o+u_p+u_r)}}$$

decoupling
104 ns

$$\times \frac{\int_{\mathrm{bulk}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_0+u_c+u_o+u_p+u_r)}}{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_0+u_c+u_o)}}$$

$$\times \frac{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c+u_o)}}{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c)}}$$

$$\times \frac{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_0+u_c)}}{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_1+u_c)}}$$

coupling
104 ns

$$\times \frac{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_1+u_c)}}{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta U_1}}$$

RMSD
48 ns

Comparison of alchemical and geometric routes

$$K_{\mathrm{eq}}^{\mathrm{PMF}} = \frac{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta U}}{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c)}} \times \frac{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c)}}{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c+u_o)}} \times \frac{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c+u_o)}}{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c+u_o+u_p)}} \times \frac{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c+u_o+u_p)}}{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c+u_o)}} \times \frac{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c+u_o)}}{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c)}} \times \frac{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c)}}{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta U}}$$

$\Delta G_c^{\mathrm{site}} = +3.5$ kcal/mol

$\Delta G_o^{\mathrm{site}} = +0.7$ kcal/mol

$\Delta G_p^{\mathrm{site}} = +0.3$ kcal/mol

$\Delta G_r = -14.6$ kcal/mol

$\Delta G_o^{\mathrm{bulk}} = +5.8$ kcal/mol

$\Delta G_c^{\mathrm{bulk}} = +5.4$ kcal/mol

$\Delta G^0 = -7.8$ kcal/mol

$$K_{\mathrm{eq}}^{\mathrm{FEP}} = \frac{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta U_1}}{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_1+u_c)}} \times \frac{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_1+u_c)}}{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_1+u_c+u_o)}} \times \frac{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_1+u_c+u_o)}}{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_1+u_c+u_o+u_p)}} \times \frac{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_1+u_c+u_o+u_p)}}{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_1+u_c+u_o+u_p+u_r)}} \times \frac{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_1+u_c+u_o+u_p+u_r)}}{\int_{\mathrm{site}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_0+u_c+u_o+u_p+u_r)}} \times \frac{\int_{\mathrm{bulk}} \mathrm{d1} \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_0+u_c+u_o+u_p+u_r)}}{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_0+u_c+u_o)}} \times \frac{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c+u_o)}}{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta(U+u_c)}} \times \frac{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_0+u_c)}}{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_1+u_c)}} \times \frac{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta(U_1+u_c)}}{\int_{\mathrm{bulk}} \mathrm{d1}\, \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \mathrm{dx}\, \mathrm{e}^{-\beta U_1}}$$

$\Delta G_c^{\mathrm{site}} = +4.2$ kcal/mol

$\Delta G_o^{\mathrm{site}} = +1.7$ kcal/mol

$\Delta G_p^{\mathrm{site}} = +0.6$ kcal/mol

$\Delta G_r^{\mathrm{site}} = +0.4$ kcal/mol

$\Delta G_a^{\mathrm{site}} = -35.9$ kcal/mol

$\Delta G_t^{\mathrm{bulk}} = +4.0$ kcal/mol

$\Delta G_o^{\mathrm{bulk}} = +6.6$ kcal/mol

$\Delta G_a^{\mathrm{bulk}} = -53.3$ kcal/mol

$\Delta G_c^{\mathrm{bulk}} = +6.1$ kcal/mol

$\Delta G^0 = -7.7$ kcal/mol

Error analysis

Geometrical route

$$\Delta A \simeq -\Delta\xi \sum_{i=1}^{p} \langle F_\xi \rangle_{\xi_i}$$

$$\begin{cases} \mu &= \langle F_\xi \rangle_\xi = \dfrac{1}{p} \sum_{i=1}^{p} \langle F_\xi \rangle_{\xi_i} = \dfrac{1}{p} \sum_{i=1}^{p} \mu_i \\ \sigma^2 &= \dfrac{1}{p} \sum_{i=1}^{p} \sigma_i^2 + (\mu - \mu_i)^2 \end{cases}$$

$$\langle |\Delta A - \overline{\Delta A}|^2 \rangle = (\Delta\xi)^2 \sum_{i,j=1}^{p} \sum_{k=1}^{n_i} \sum_{l=1}^{n_j} \frac{1}{n_i n_j} \langle (F^i_{\xi,l} - \mu)(F^i_{\xi,k} - \mu) \rangle$$

$$\simeq (\Delta\xi)^2 \left\{ \frac{p}{N} \sum_{i=1}^{p} [\sigma_i^2 + (\mu - \mu_i)^2] + 2\sigma^2 \frac{p^2}{N} \kappa \right\}$$

$$\sigma_{\Delta A} \simeq \Delta\xi \, \frac{\sigma}{N^{1/2}} (1 + 2\kappa)^{1/2}$$

RMSD	±0.5 kcal/mol
$\Theta, \Phi, \Psi, \theta, \phi$	±0.2 kcal/mol
r	±0.4 kcal/mol
	±0.9 kcal/mol

Hahn, A. M.; Then, H. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. **2009**, 80, 031111

Pohorille, A. et al. J. Phys. Chem. B, **2010**, 114, 10235-10253

Alchemical route

$$\sigma^2_{\Delta\widehat{A}} = \frac{(1+2\kappa_0)}{n_0\beta^2} \left(\frac{\langle f^2(x)\rangle_0}{\langle f(x)\rangle_0^2} - 1 \right) + \frac{(1+2\kappa_1)}{n_1\beta^2} \left(\frac{\langle f^2(-x)\rangle_1}{\langle f(-x)\rangle_1^2} - 1 \right)$$

$$f(x) = 1/(1+e^x) \quad \text{with} \quad x = \beta(\Delta U - C)$$

- Low statistical errors
- Estimates burdened by systematic error

RMSD	±0.4 kcal/mol
$\Theta, \Phi, \Psi, \theta, \phi$	±0.0 kcal/mol
r	±0.0 kcal/mol
alchemy	±0.7 kcal/mol
	±1.0 kcal/mol

often very tedious but you should still do it! (*reviewers will often request it anyway!*)

Rodriguez-Gomez, D. et al. J. Chem. Phys., **2004**, 120, 3563-3578

Hénin, J.; Chipot, C. *J. Chem. Phys.* **2004**, *121*, 2904-2914

Error analysis

Geometrical route

Advantages

- Rigorous, formally correct framework

- Reasonably inexpensive

- Access to the statistical error for all terms

- In principle, applicable to protein:protein dimers

Shortcomings

- Cumbersome

- Convergence of RMSD term; Degeneracy

- Convergence of separation term; ⊥ DoF's ?

- Limited to interfacial binding sites

Alchemical route

Advantages

- Rigorous, formally correct framework

- Reasonably inexpensive

- Access to the statistical error for all terms

- Embarrassingly parallelizable

Shortcomings

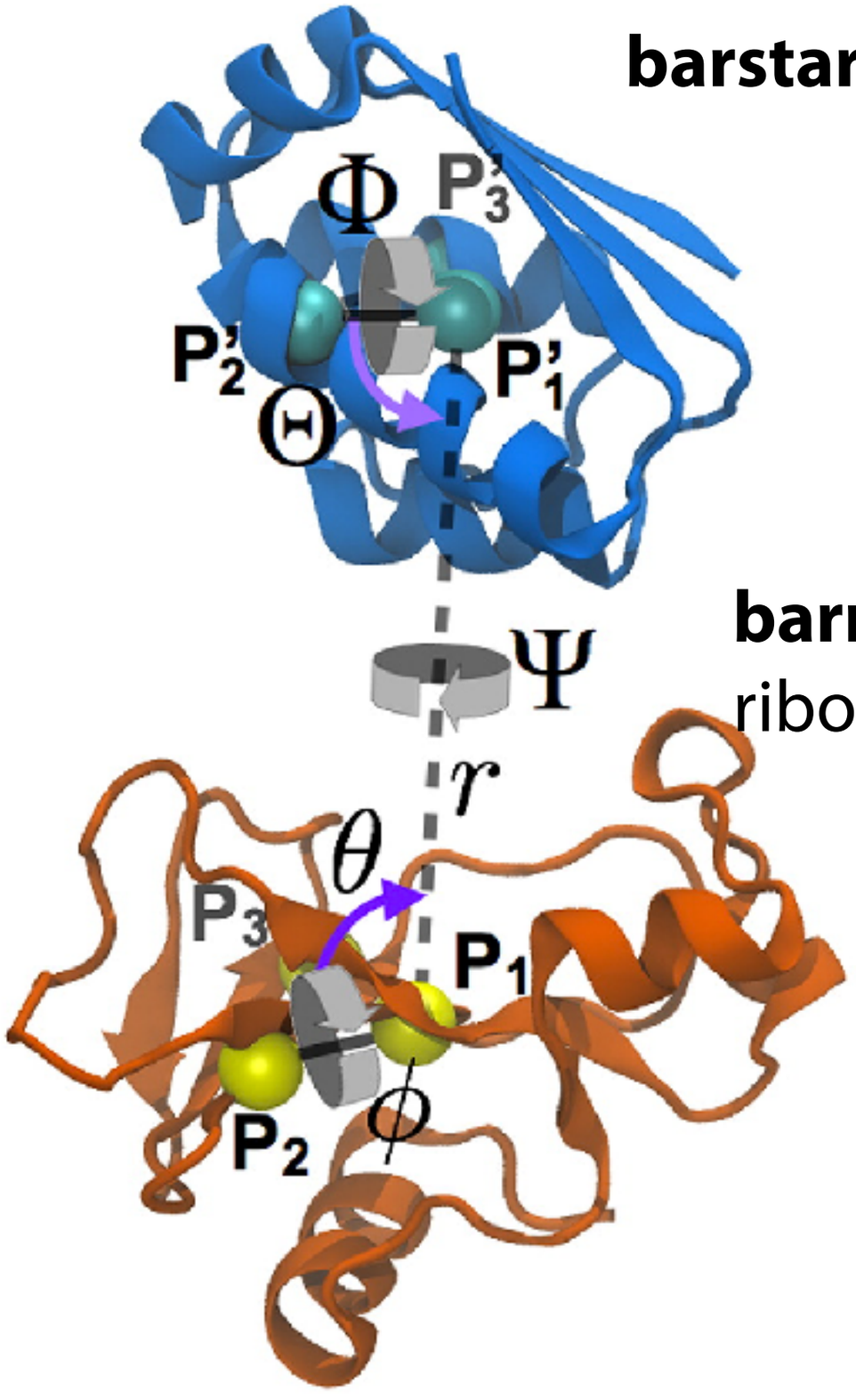
- Cumbersome

- Convergence of alchemical transformation

- Convergence of restraint term

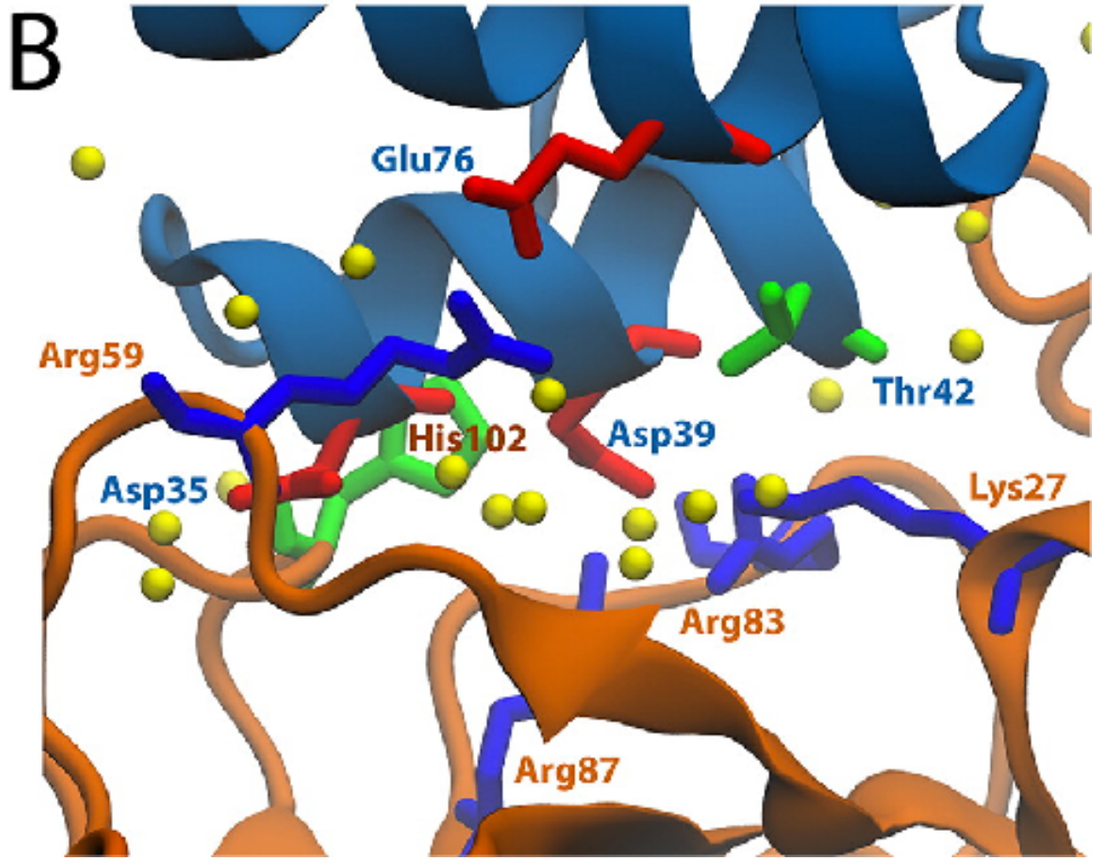
- In principle, limited to small ligands

Protein-protein binding free energy



barstar - an inhibitor

barnase - a ribonuclease



interface is highly solvated

ΔG^0 = **-19.0 kcal/mol (exp)**

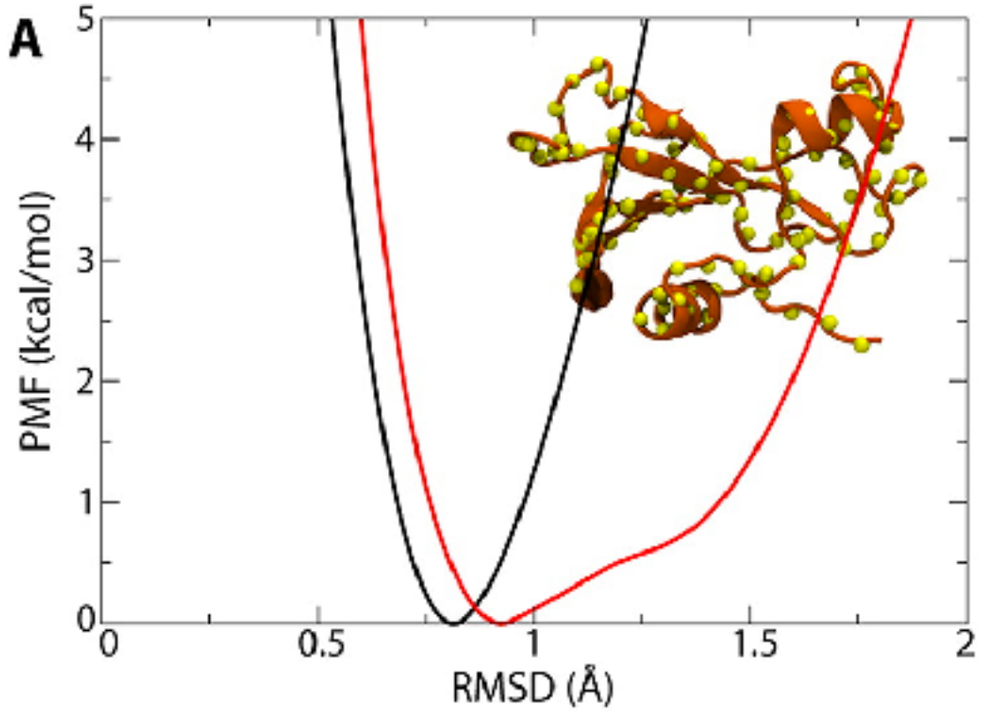
Schreiber & Fersht. *JMB*, **248**:478-486. 1995.

Gumbart, Roux, Chipot. Efficient Determination of Protein–Protein Standard Binding Free Energies from First Principles. *JCTC* **9**:3789-3798. 2013.

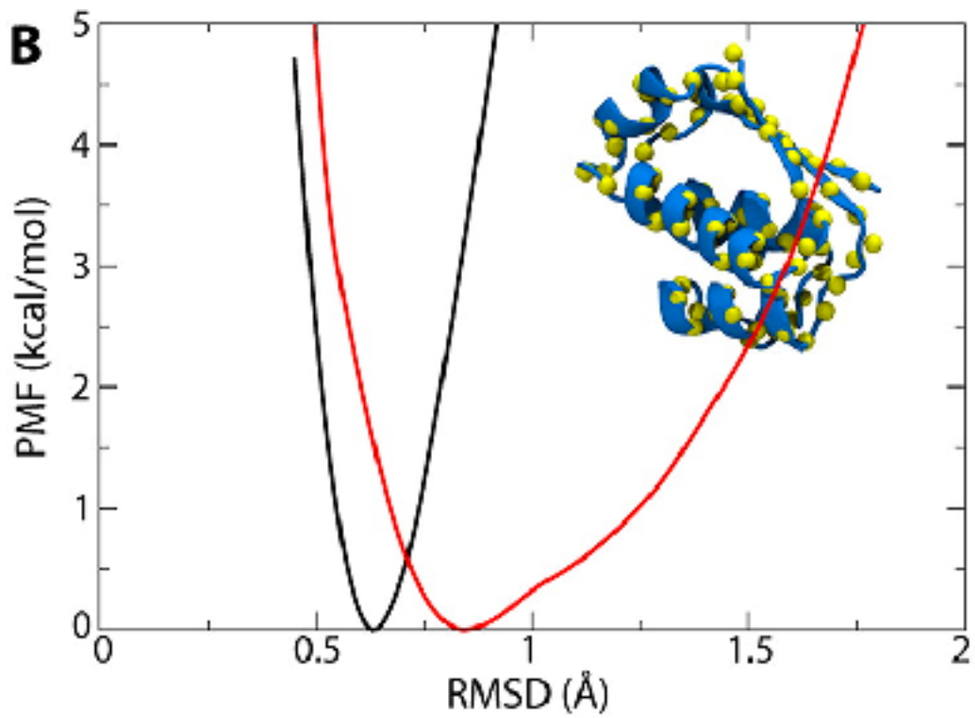
Numerous restraints needed

Gumbart, Roux, Chipot. *JCTC* **9**:3789-3798. 2013.

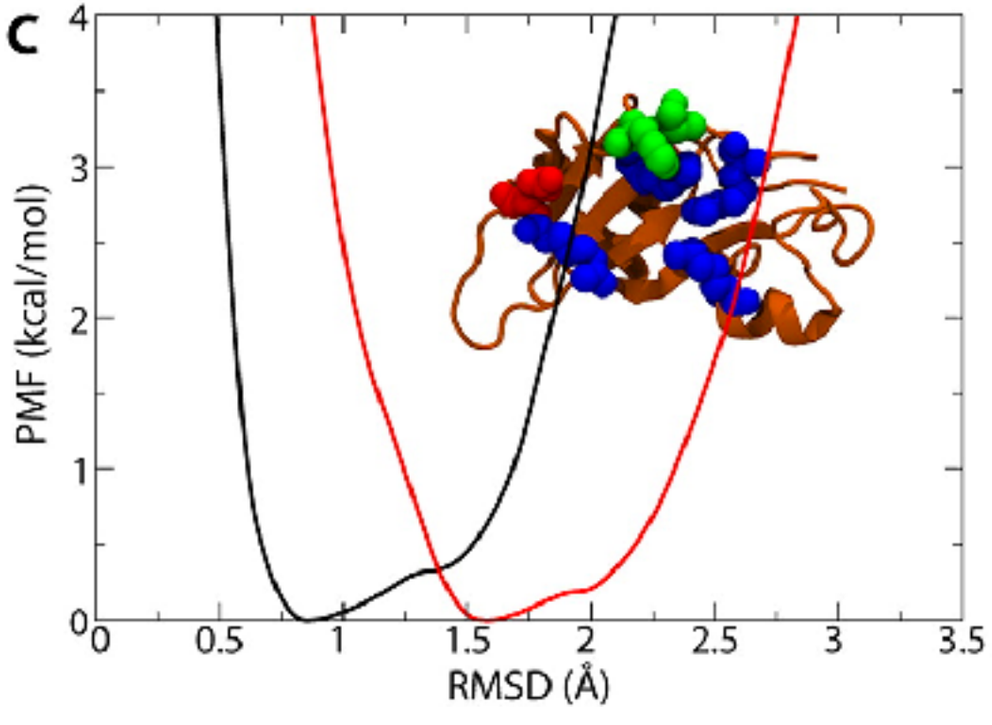
RMSD on **barnase backbone**



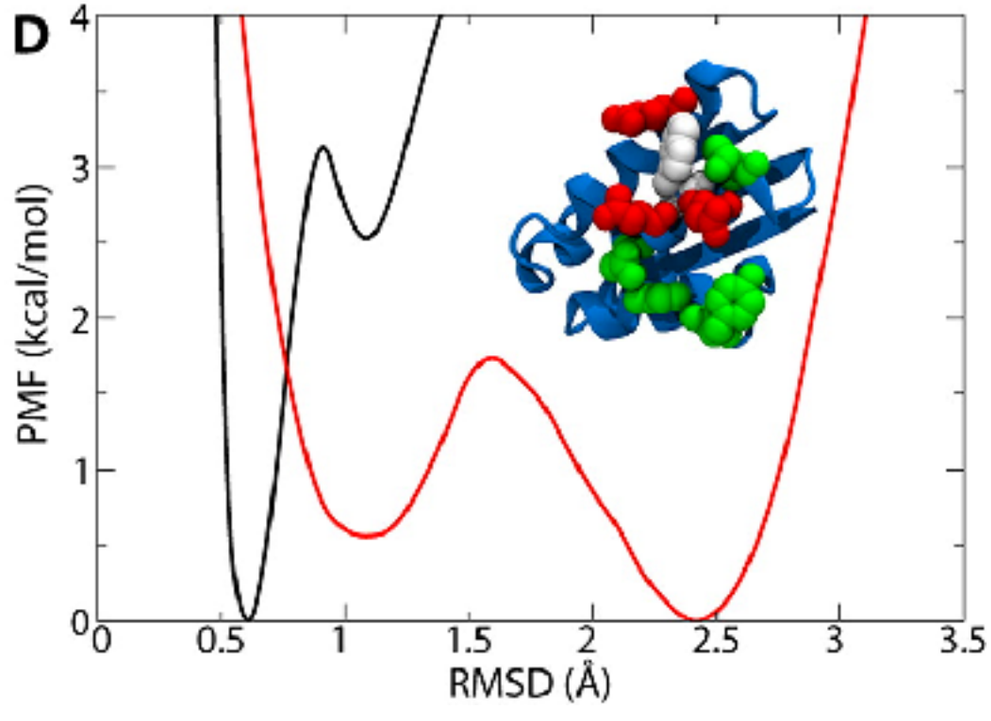
RMSD on **barstar backbone**



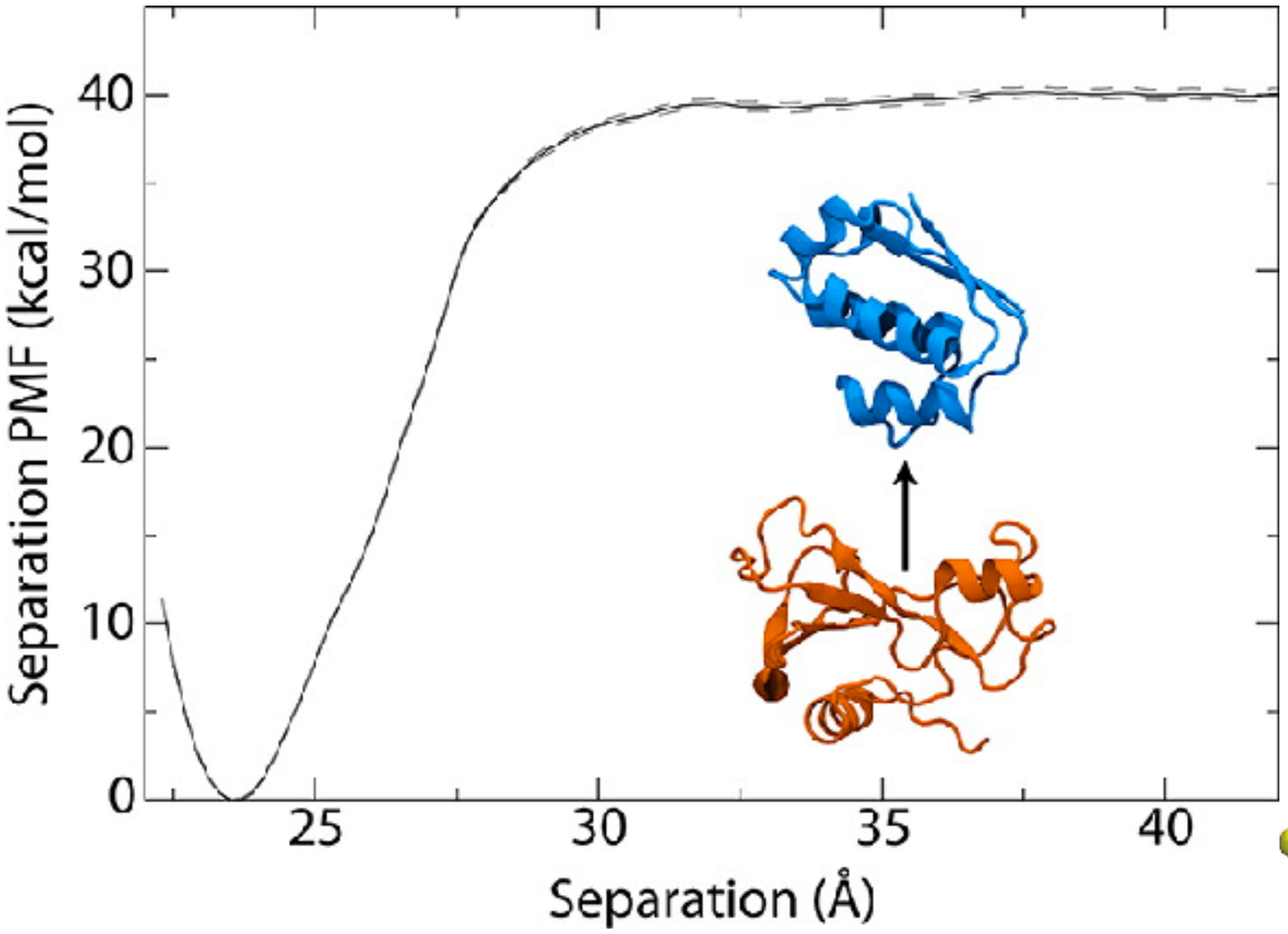
RMSD on **barnase side chains**



RMSD on **barstar side chains**



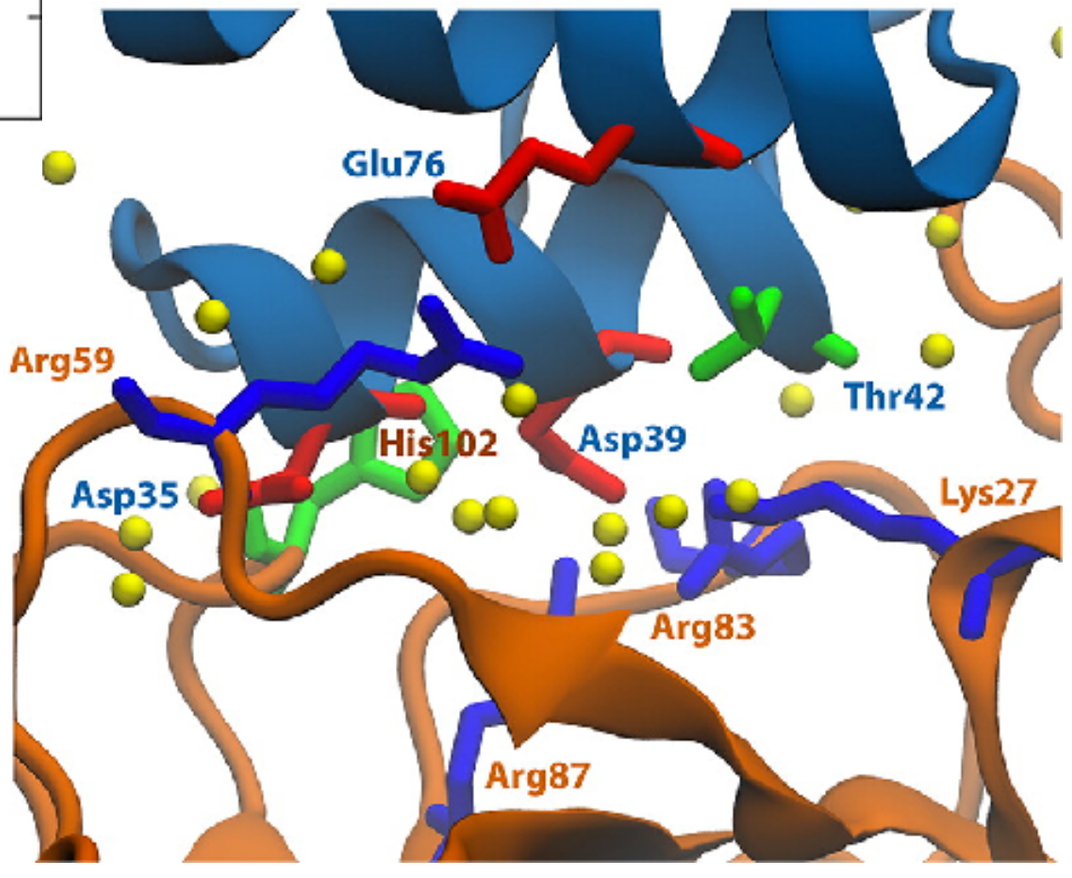
Separating the proteins



PMF took over 50 windows spaced by 0.5 Å and ~200 ns to fully converge

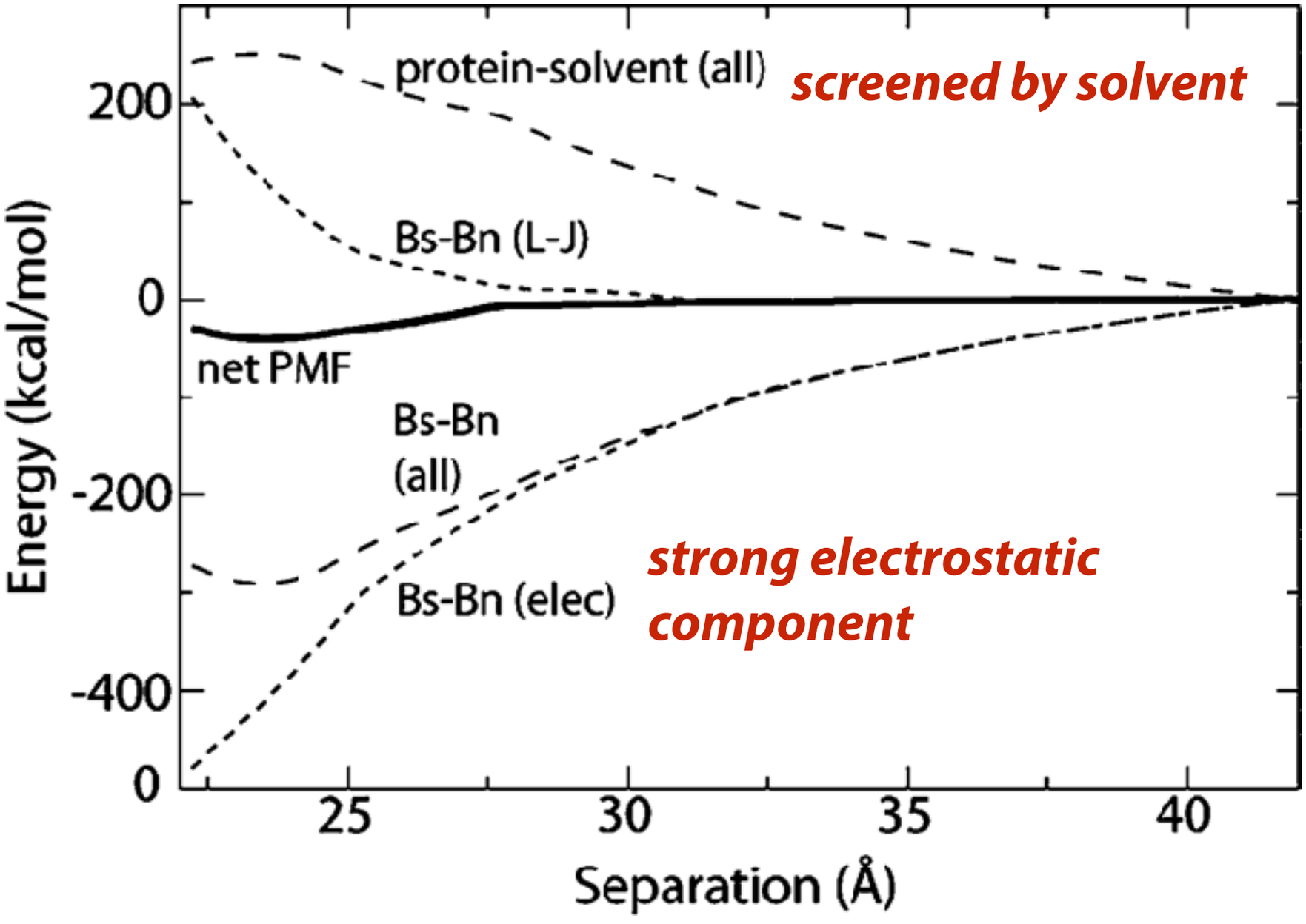
without side-chain restraints, PMF did not converge ***even in 400 ns***

the appropriate choice of restraints is problem dependent!!!



Gumbart, Roux, Chipot. *JCTC* **9**:3789-3798. 2013.

Decomposing the PMF



Force decomposition reveals key contributions to the PMF

Gumbart, Roux, Chipot. *JCTC* **9**:3789-3798. 2013.

And fourteen separate calculations later…

contribution	PMF (kcal/mol)	time (ns)
$\Delta G^{site}_{Bs,c}$	-1.98 ± 0.31	6
$\Delta G^{site}_{Bn,c}$	-3.13 ± 0.06	12
$\Delta G^{site}_{Bs,res}$	-1.87 ± 0.75	12
$\Delta G^{site}_{Bn,res}$	-3.45 ± 0.63	24
ΔG^{site}_{Θ}	-0.09 ± 0.42	8
ΔG^{site}_{Φ}	-0.35 ± 0.08	4
ΔG^{site}_{ψ}	-0.24 ± 0.09	8
ΔG^{site}_{θ}	-0.13 ± 0.33	4
ΔG^{site}_{ϕ}	-0.05 ± 0.12	4
$-(1/\beta)\ln(S^*I^*C^\circ)$	-37.10 ± 0.29	212
ΔG^{bulk}_{o}	+6.61	
$\Delta G^{bulk}_{Bn,res}$	$+8.05 \pm 0.31$	21
$\Delta G^{bulk}_{Bs,res}$	$+5.15 \pm 0.42$	15
$\Delta G^{bulk}_{Bn,c}$	$+4.15 \pm 0.53$	18
$\Delta G^{bulk}_{Bs,c}$	$+3.39 \pm 0.17$	24
ΔG°_{bind}	-21.04 ± 1.43	372

ΔG^0 = **-19.0 kcal/mol (exp)**

Within 2 kcal/mol!!!

Gumbart, Roux, Chipot. *JCTC* **9**:3789-3798. 2013.

Forget everything you just saw: BFEE plugin

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

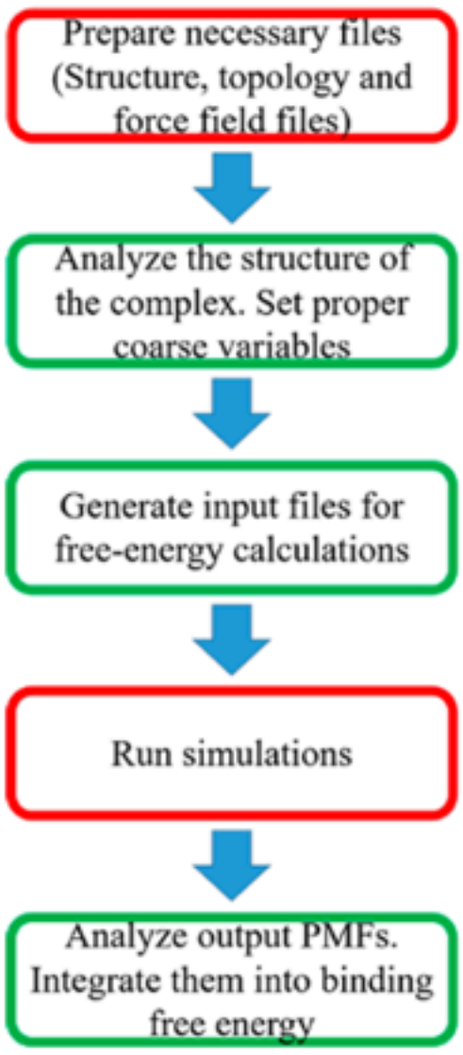
Application Note

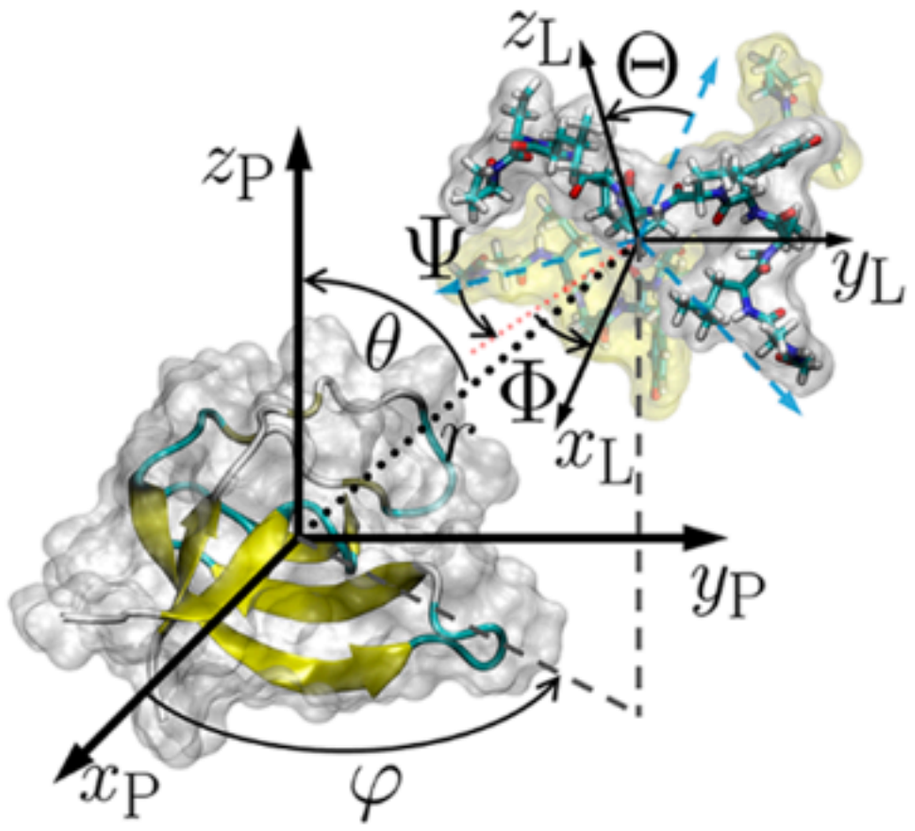
pubs.acs.org/jcim

BFEE: A User-Friendly Graphical Interface Facilitating Absolute Binding Free-Energy Calculations

Haohao Fu,[†] James C. Gumbart,[∥] Haochuan Chen,[†] Xueguang Shao,[†,‡,#] Wensheng Cai,*[,†,‡] and Christophe Chipot*[,⊥,$,§]

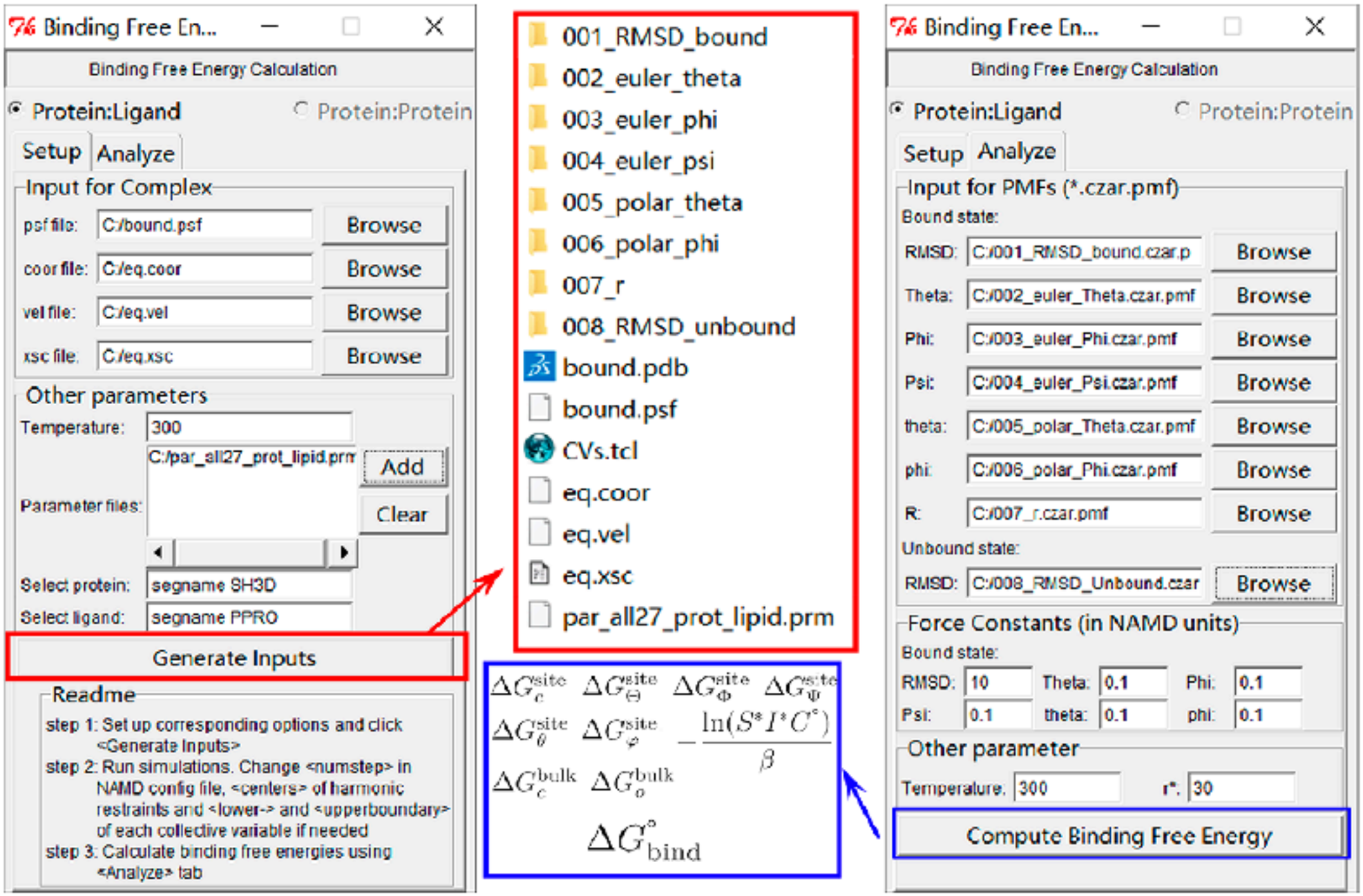
A VMD plugin that aids setup and analysis of all the steps to calculate an absolute binding free energy





Fu et al. BFEE: A user-friendly graphical interface facilitating absolute binding free-energy calculations. *J. Chem. Inf. Model*. 2018, **58**, 556-560.

Forget everything you just saw: BFEE plugin



Will be widely available in next released version of VMD 1.9.4

To install now, obtain from supplement of published paper

Fu et al. *J. Chem. Inf. Model*. 2018, **58**, 556-560.