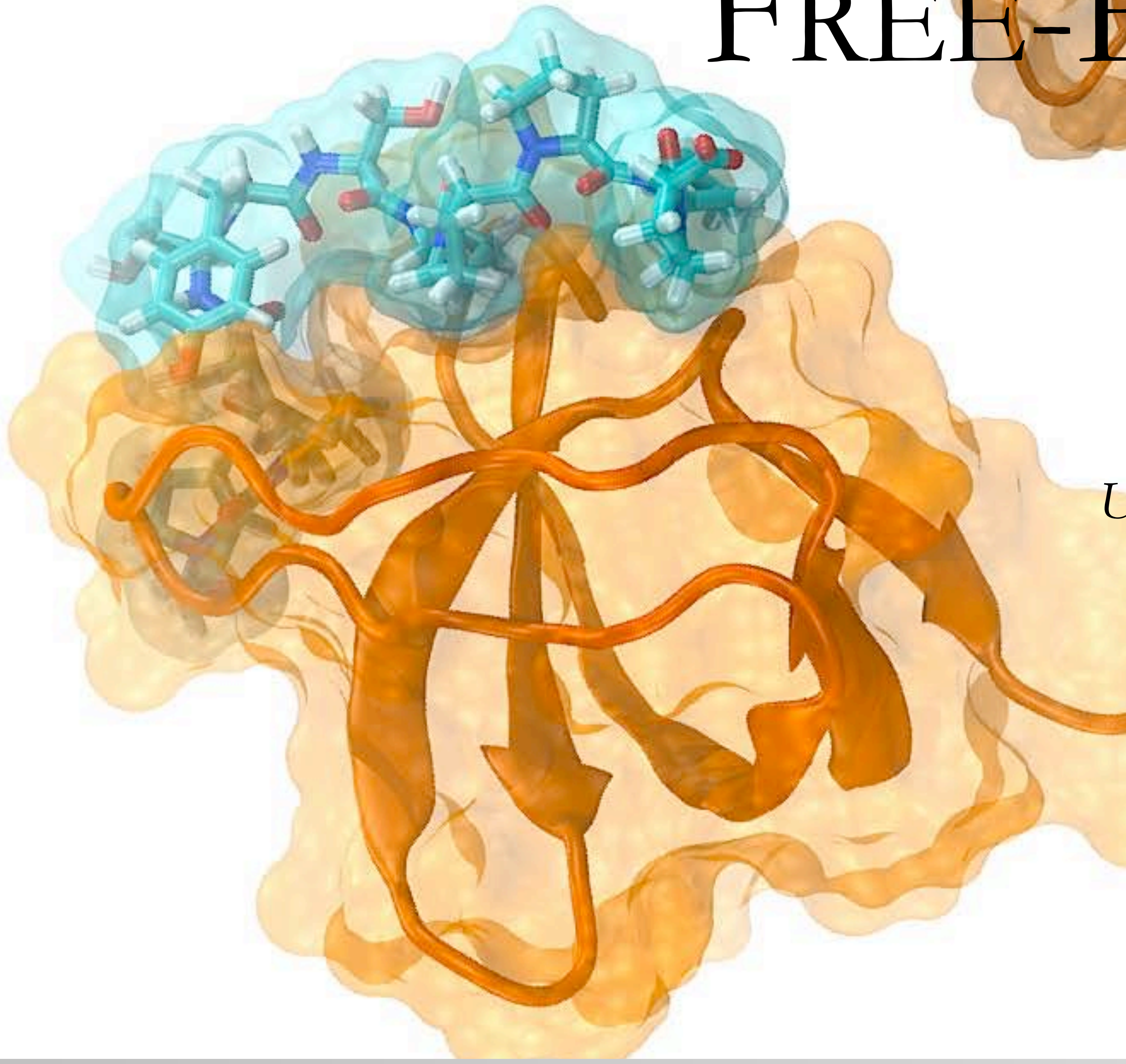
PROTEIN-LIGAND STANDARD BINDING FREE-ENERGY CALCULATIONS



Chris Chipot

Laboratoire International Associé CNRS-UIUC,
Unité Mixte de Recherche n° 7565, Université de Lorraine

Beckman Institute for Advanced Science and Technology,
Department of Physics
University of Illinois at Urbana-Champaign

THE LONG-STANDING PROTEIN-LIGAND PROBLEM

THE ALCHEMICAL ROUTE

THE GEOMETRICAL ROUTE

RELATIVE BINDING FREE ENERGIES

ALANINE SCANNING

BEYOND PROTEIN-LIGAND BINDING

The long-standing protein-ligand problem

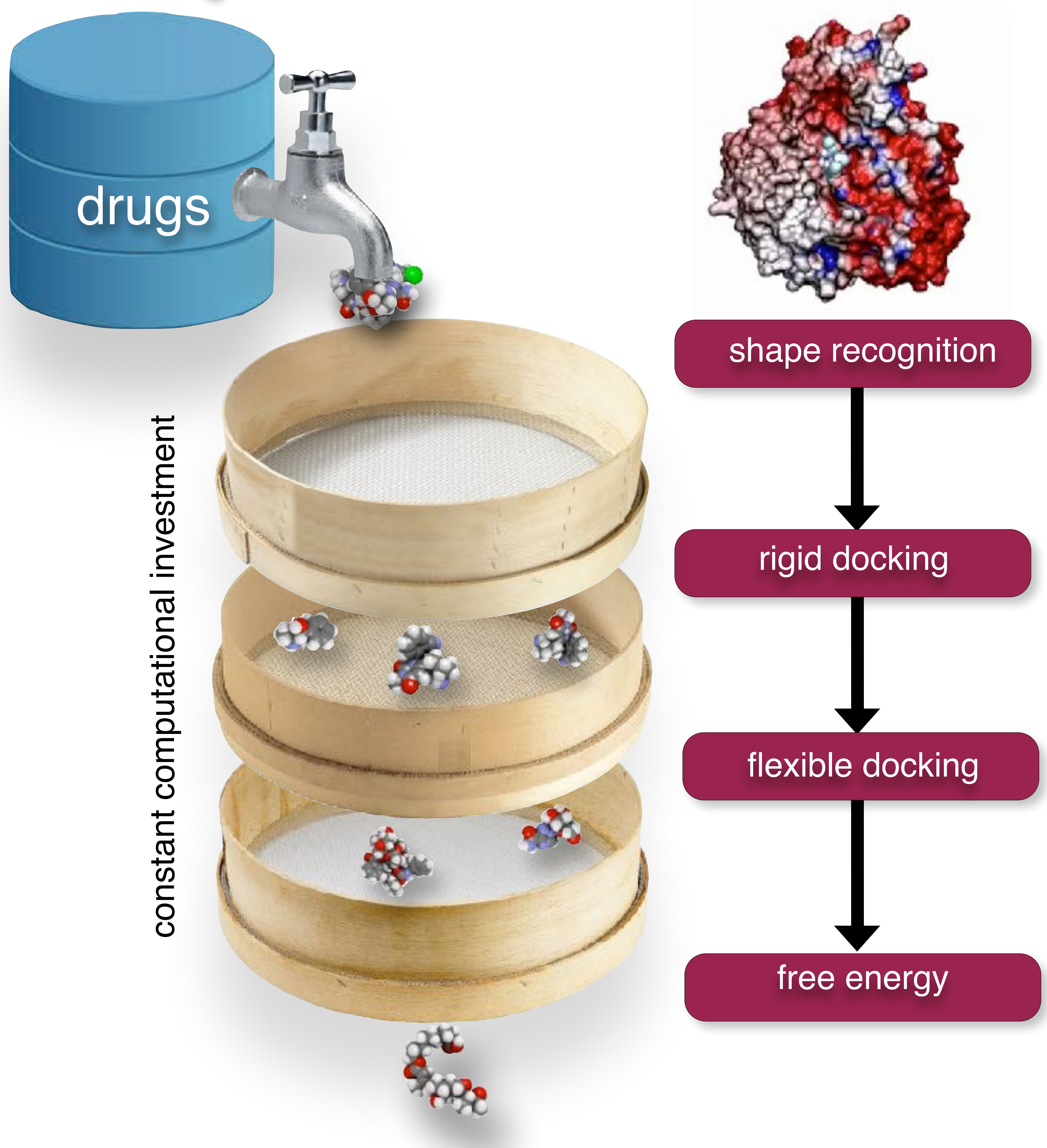
The alchemical route

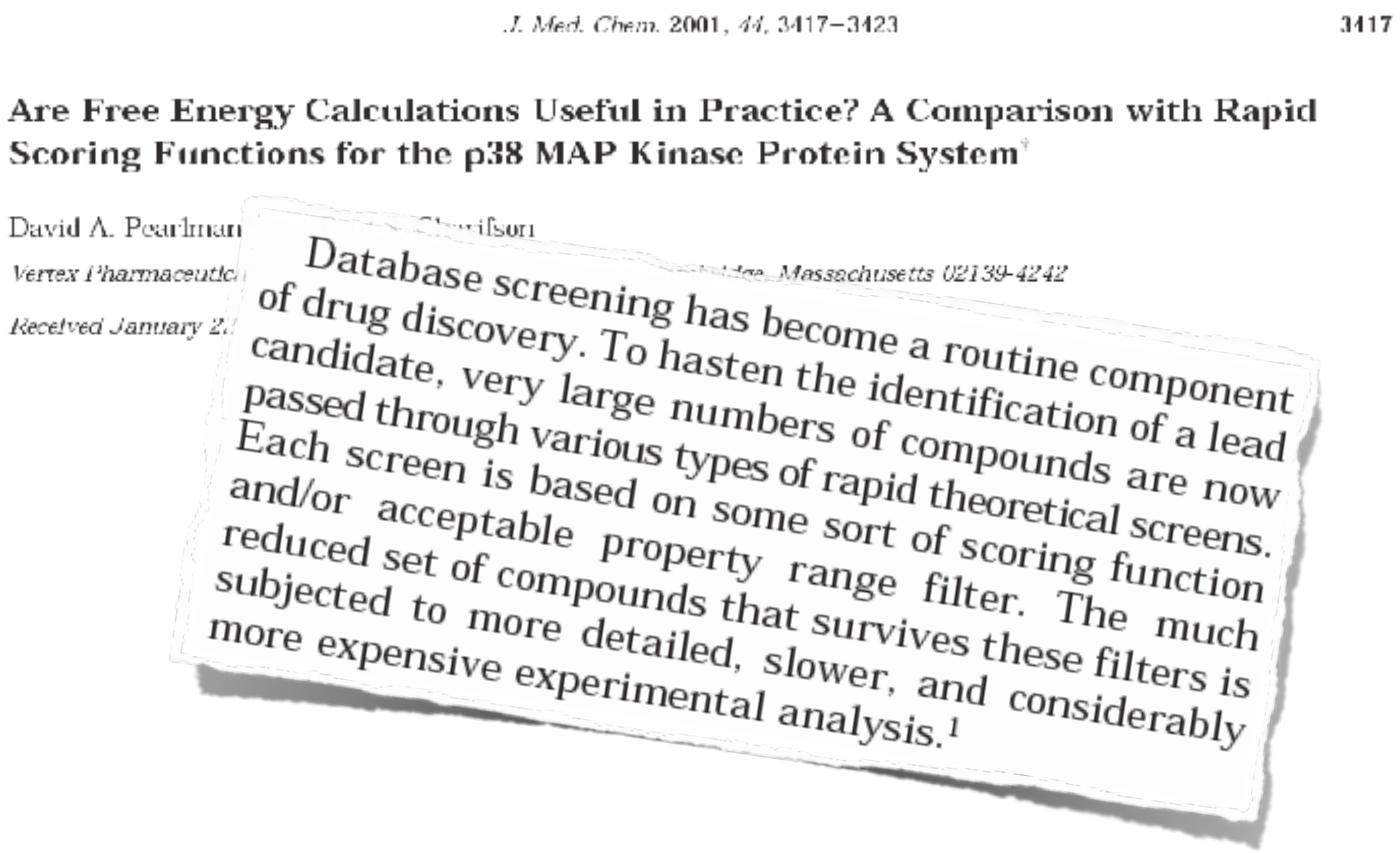
The geometrical route

Relative binding free energies

Alanine scanning

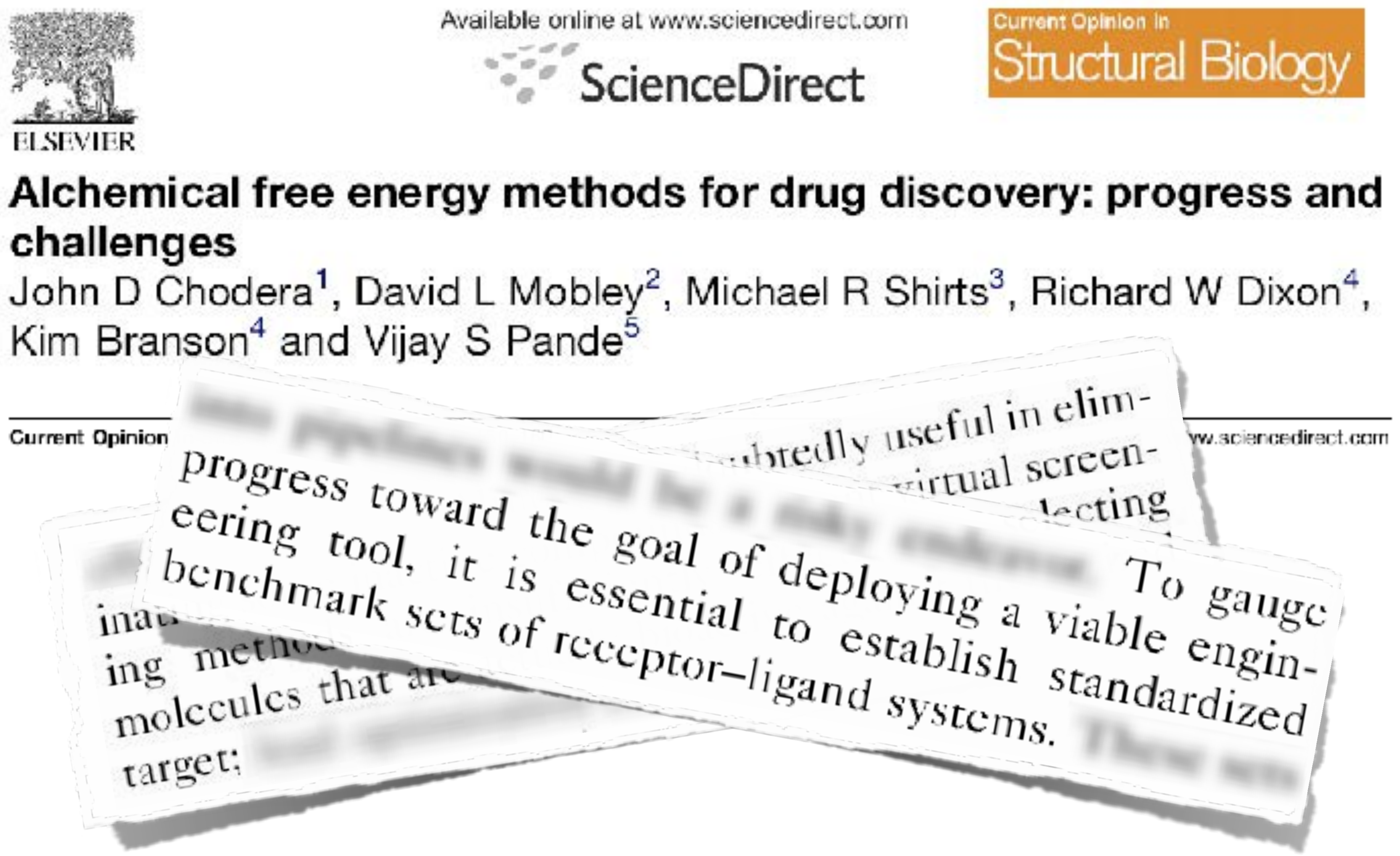
Beyond protein-ligand binding




J. Med. Chem. 2001, 44, 3417–3423 3417

Are Free Energy Calculations Useful in Practice? A Comparison with Rapid Scoring Functions for the p38 MAP Kinase Protein System

Database screening has become a routine component of drug discovery. To hasten the identification of a lead candidate, very large numbers of compounds are now passed through various types of rapid theoretical screens. Each screen is based on some sort of scoring function and/or acceptable property range filter. The much reduced set of compounds that survives these filters is subjected to more detailed, slower, and considerably more expensive experimental analysis.[1]


Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Structural Biology

Alchemical free energy methods for drug discovery: progress and challenges

John D Chodera[1], David L Mobley[2], Michael R Shirts[3], Richard W Dixon[4], Kim Branson[4] and Vijay S Pande[5]

progress toward the goal of deploying a viable engineering tool, it is essential to establish standardized benchmark sets of receptor–ligand systems.

Chipot, C.; Rozanska, X.; Dixit, S. B. *J. Comput. Aided Mol. Des.* **2005**, *19*, 765-770.
Shirts, M. R.; Mobley, D. L.; Chodera, J. D. *Annual Reports Comput. Chem.* **2007**, 3, 41-59.
Chipot, C. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 71-89.

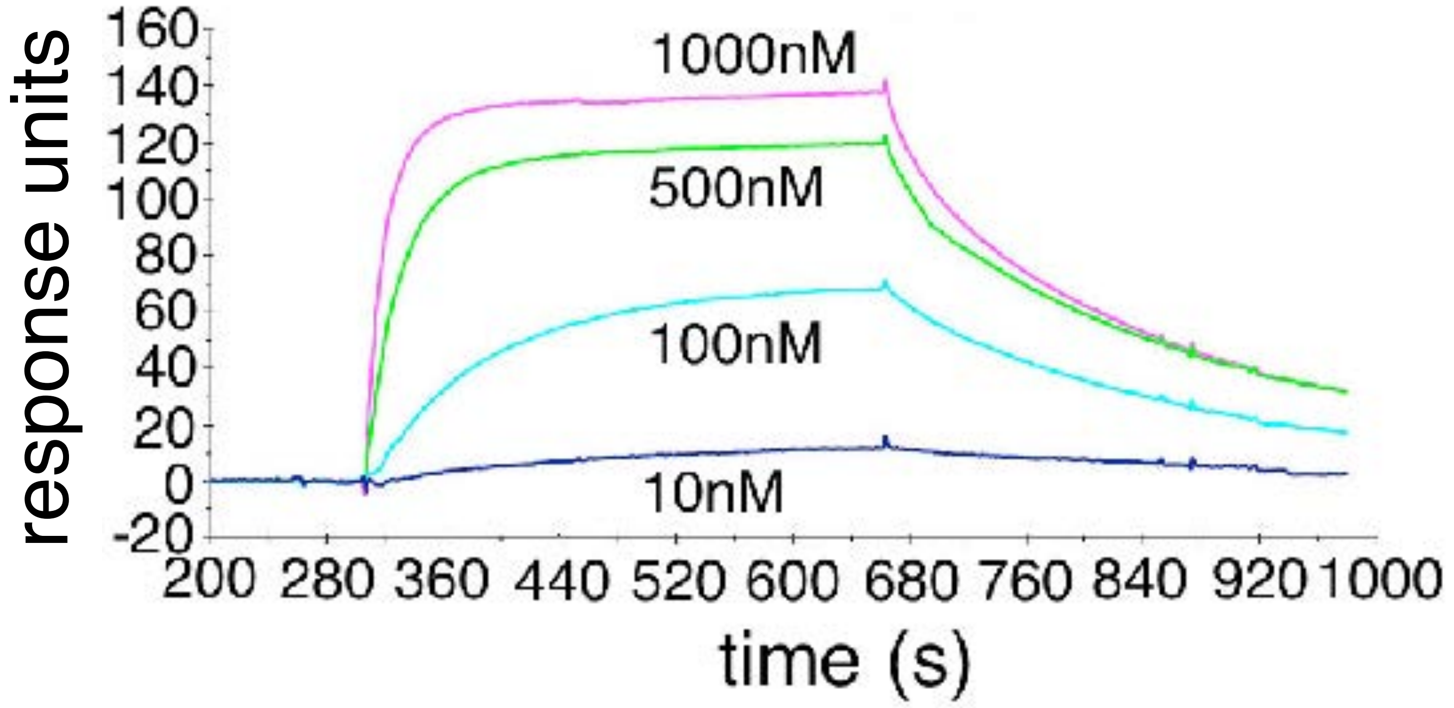
Theoretical background

$$K_{\text{eq}} = \frac{[\text{protein} : \text{ligand}]}{[\text{protein}][\text{ligand}]}$$

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

which can readily be determined by experiment:

$$K_d = \frac{k_{\text{off}}}{k_{\text{on}}}$$



— A single event is evidently not enough.

— Brute-force simulations are limited by k_{on} and k_{off}.

Kollman, P.A. *Chem. Rev.* **1993**, *93*, 2395-2417

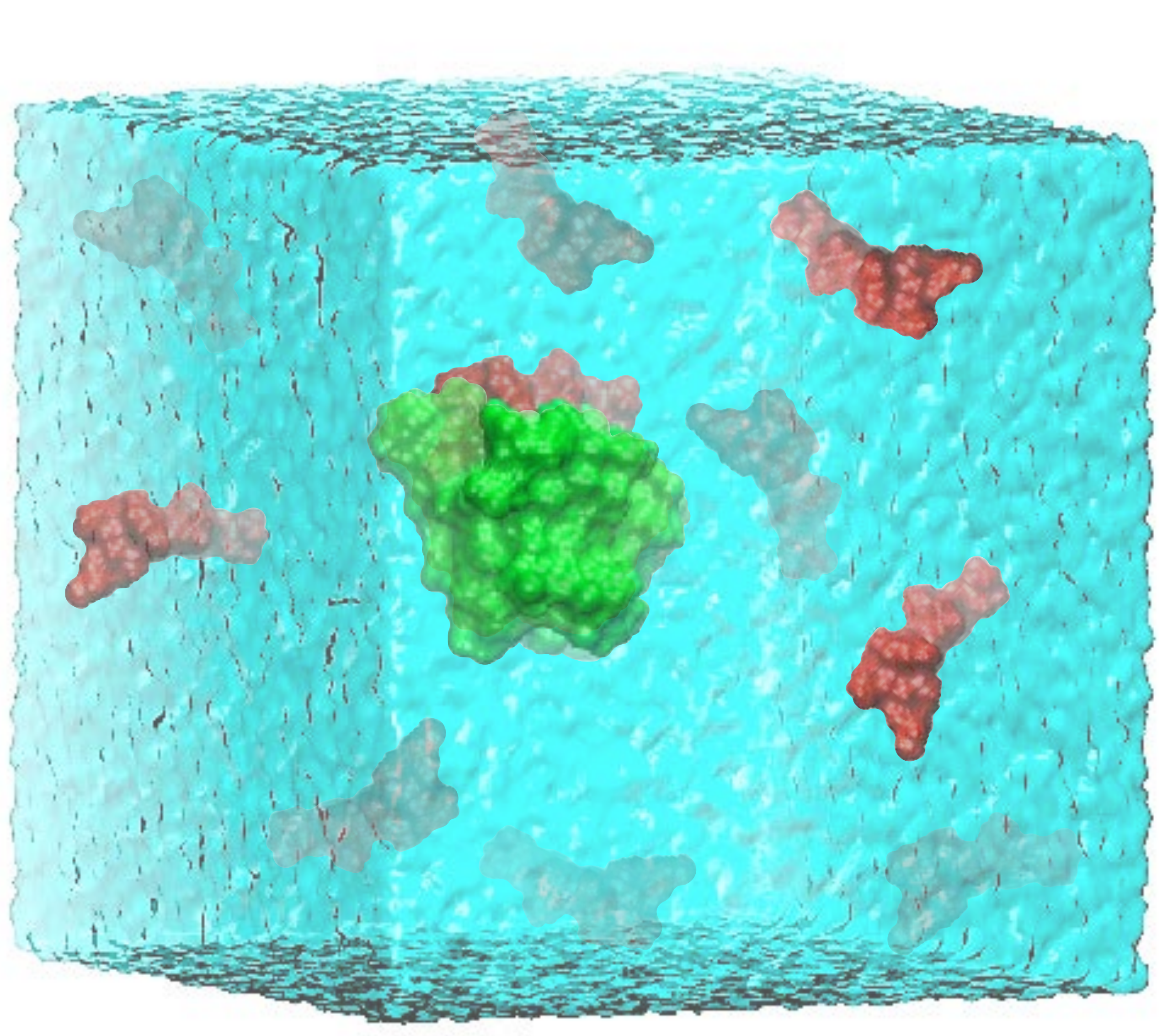
Gilson, M. K. et al. *Biophys. J.* **1997**, *72*, 1047-1069

Chipot, C.; Pohorille, A. Free-energy calculations. Springer **2007**.

Karlsson, R.; Larsson, A. *Methods Mol. Biol.* **2004**, *248*, 389-415

Buch, I.; Giorgino, T.; Fabritiis, G. D. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 10184-10189

THEORETICAL BACKGROUND



N ligands

$$\text{[protein]} = p_0 \; \text{[protein]}_{\text{tot}}$$

$$\text{[protein : ligand]} = p_1 \; \text{[protein]}_{\text{tot}}$$

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

$$K_{\text{eq}} = \frac{\text{[protein : ligand]}}{\text{[protein][ligand]}}$$

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

$$K_{\text{eq}} = \frac{1}{\text{[ligand]}} \left\{ \frac{\int_{\text{site}} \mathrm{d}\mathbf{1} \int_{\text{bulk}} \mathrm{d}\mathbf{2} \ldots \int_{\text{bulk}} \mathrm{d}\mathbf{N} \int \mathrm{d}\mathbf{x} \; \mathrm{e}^{-\beta U}}{\int_{\text{bulk}} \mathrm{d}\mathbf{1} \int_{\text{bulk}} \mathrm{d}\mathbf{2} \ldots \int_{\text{bulk}} \mathrm{d}\mathbf{N} \int \mathrm{d}\mathbf{x} \; \mathrm{e}^{-\beta U}} \right.$$

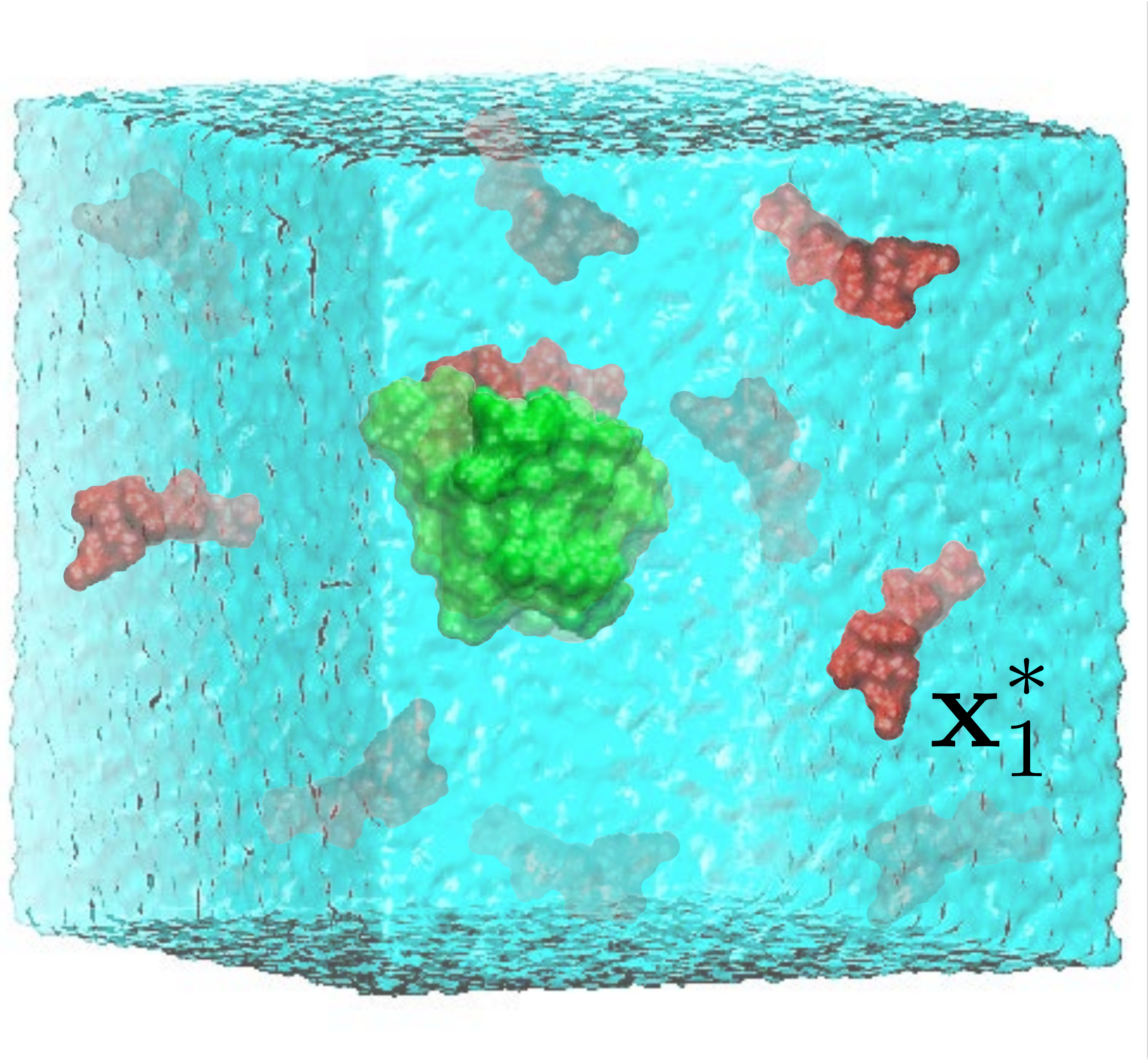
$$+ \frac{\int_{\text{bulk}} \mathrm{d}\mathbf{1} \int_{\text{site}} \mathrm{d}\mathbf{2} \ldots \int_{\text{bulk}} \mathrm{d}\mathbf{N} \int \mathrm{d}\mathbf{x} \; \mathrm{e}^{-\beta U}}{\int_{\text{bulk}} \mathrm{d}\mathbf{1} \int_{\text{bulk}} \mathrm{d}\mathbf{2} \ldots \int_{\text{bulk}} \mathrm{d}\mathbf{N} \int \mathrm{d}\mathbf{x} \; \mathrm{e}^{-\beta U}} + \cdots$$

$$\left. + \frac{\int_{\text{bulk}} \mathrm{d}\mathbf{1} \int_{\text{bulk}} \mathrm{d}\mathbf{2} \ldots \int_{\text{site}} \mathrm{d}\mathbf{N} \int \mathrm{d}\mathbf{x} \; \mathrm{e}^{-\beta U}}{\int_{\text{bulk}} \mathrm{d}\mathbf{1} \int_{\text{bulk}} \mathrm{d}\mathbf{2} \ldots \int_{\text{bulk}} \mathrm{d}\mathbf{N} \int \mathrm{d}\mathbf{x} \; \mathrm{e}^{-\beta U}} \right\}$$

Shoup, D.; Szabo, A. *Biophys. J.* **1982**, *40*, 33-39

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

THEORETICAL BACKGROUND



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

$$K_{\text{eq}} = \frac{1}{[\text{ligand}]} \frac{N \int_{\text{site}} \text{d}\mathbf{1} \int_{\text{bulk}} \text{d}\mathbf{2} \ldots \int_{\text{bulk}} \text{d}\mathbf{N} \int \text{d}\mathbf{x} \ \text{e}^{-\beta U}}{\int_{\text{bulk}} \text{d}\mathbf{1} \int_{\text{bulk}} \text{d}\mathbf{2} \ldots \int_{\text{bulk}} \text{d}\mathbf{N} \int \text{d}\mathbf{x} \ \text{e}^{-\beta U}}$$

$$= \frac{1}{[\text{ligand}]} \frac{N \int_{\text{site}} \text{d}\mathbf{1} \int \text{d}\mathbf{x} \ \text{e}^{-\beta U}}{\int_{\text{bulk}} \text{d}\mathbf{1} \int \text{d}\mathbf{x} \ \text{e}^{-\beta U}}$$

$$= \frac{1}{[\text{ligand}]} \frac{N \int_{\text{site}} \text{d}\mathbf{1} \int \text{d}\mathbf{x} \ \text{e}^{-\beta U}}{V_{\text{bulk}} \int_{\text{bulk}} \text{d}\mathbf{1} \ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \text{d}\mathbf{x} \ \text{e}^{-\beta U}}$$

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

→ alchemical route

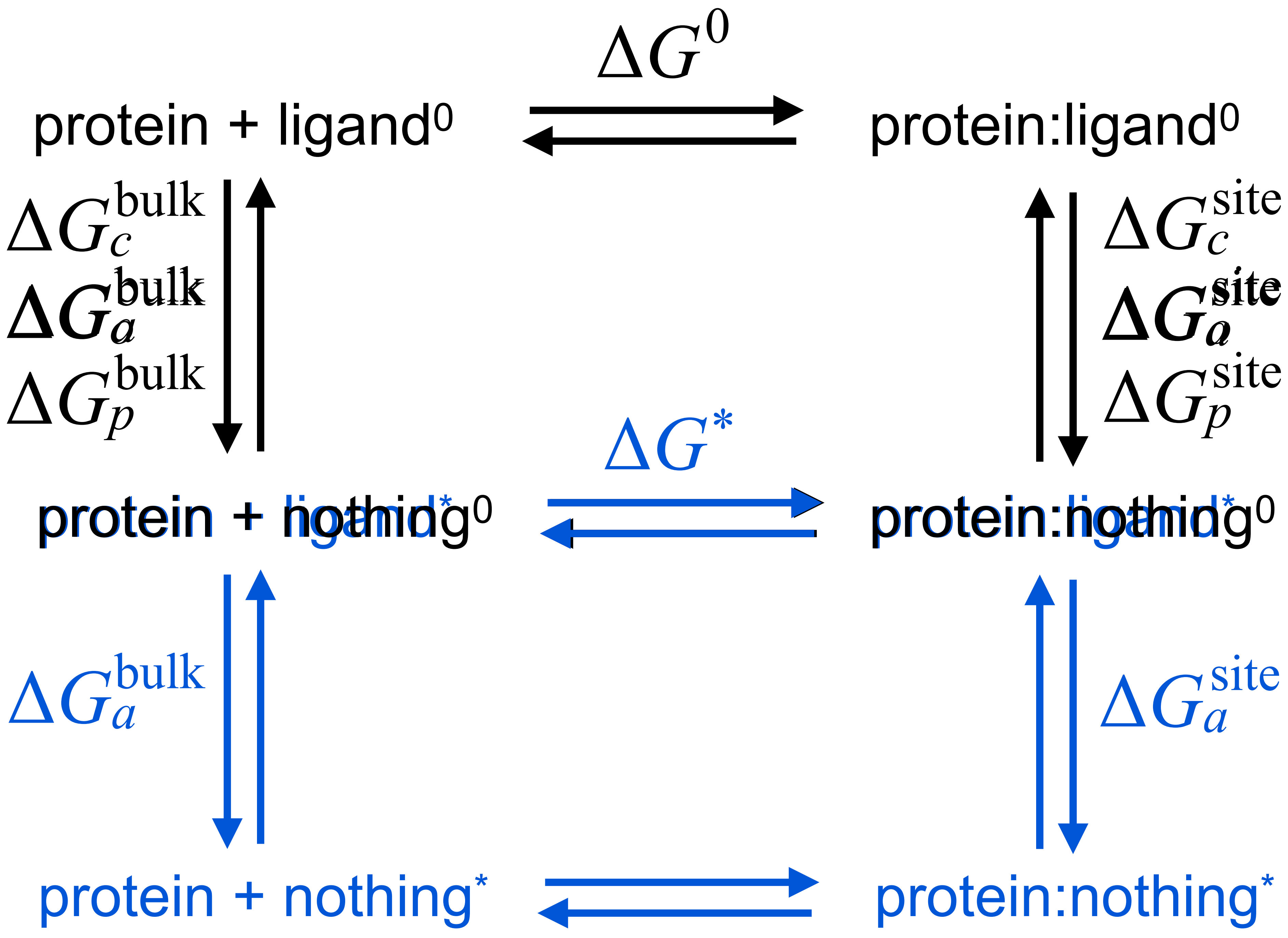
→ geometrical route

Shoup, D.; Szabo, A. *Biophys. J.* **1982**, *40*, 33-39

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

THE DOUBLE-ANNIHILATION STRATEGY

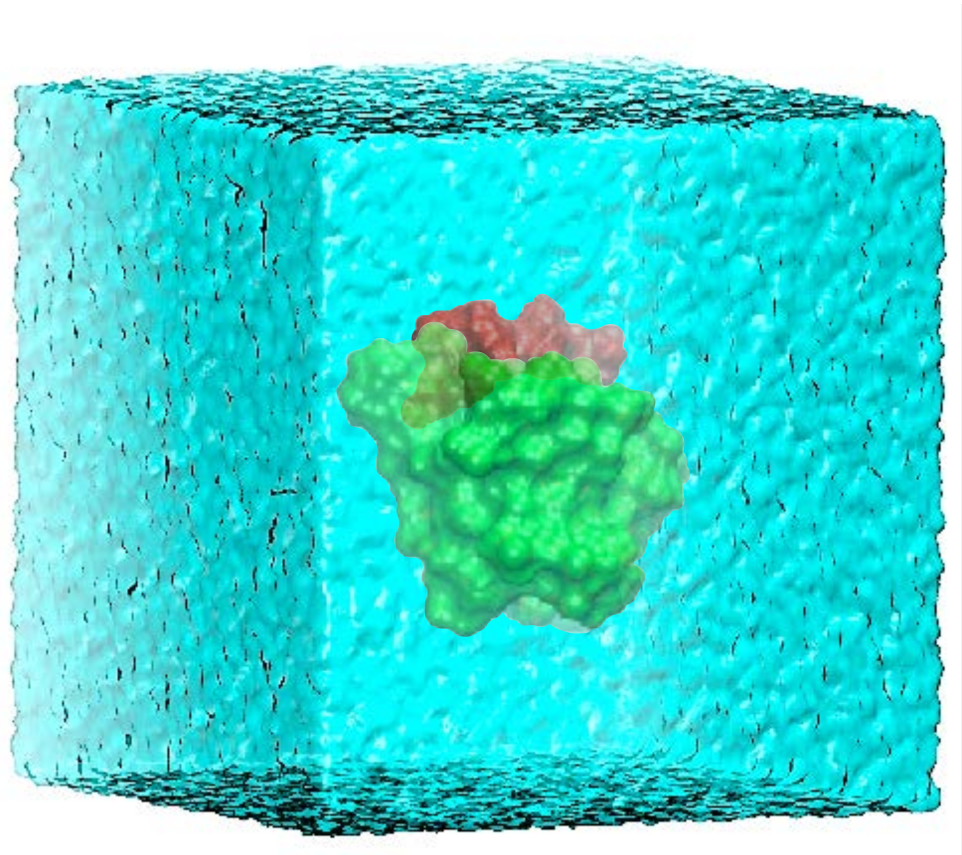
Couple reversibly the ligand to the binding site of the protein



- Floating ligand problem.
- *Corpora non agunt nisi fixata.*

Paul Ehrlich

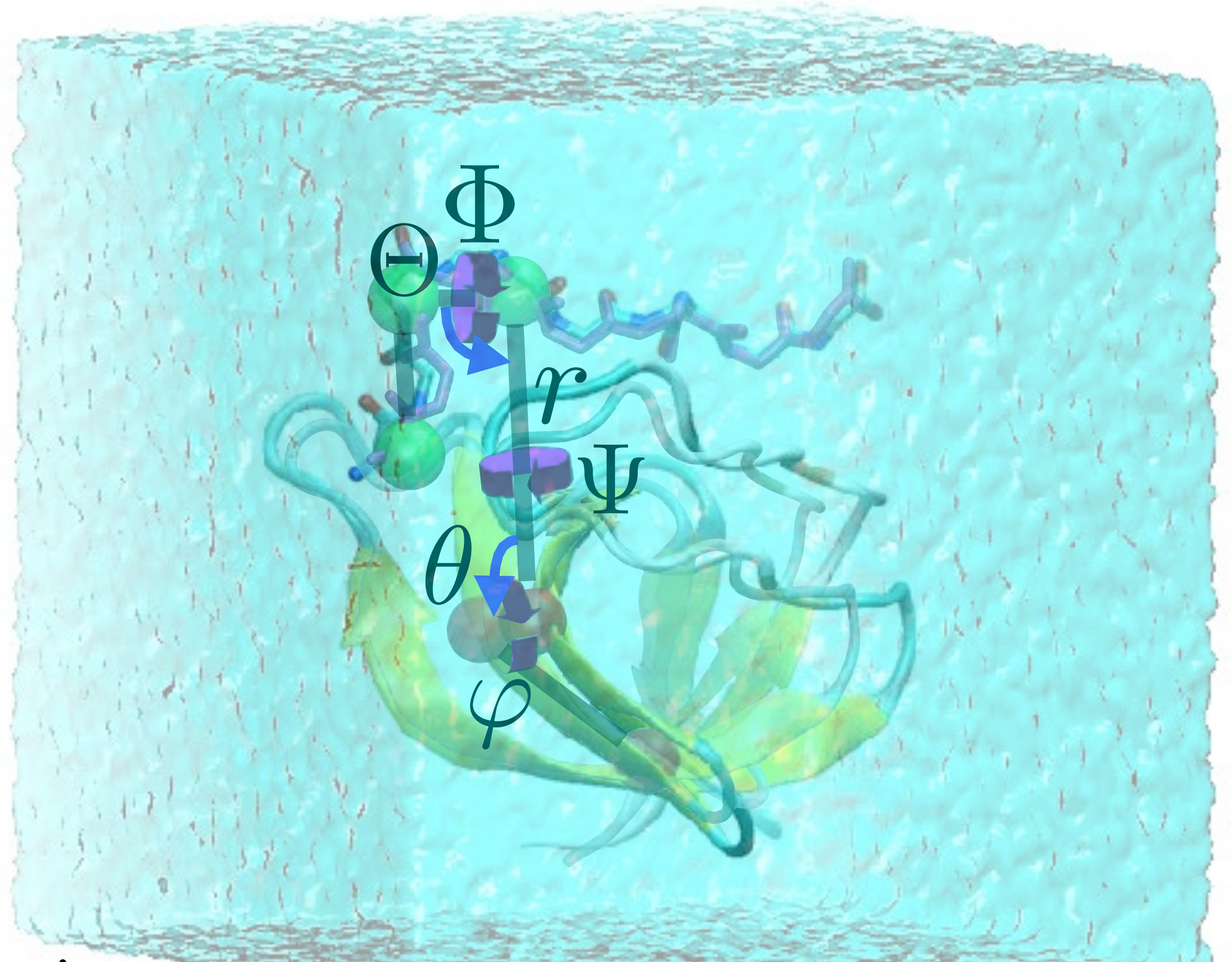
- Definition of a set of restraints.
- The loss of translational, orientational and conformational entropies contributes to the free energy.



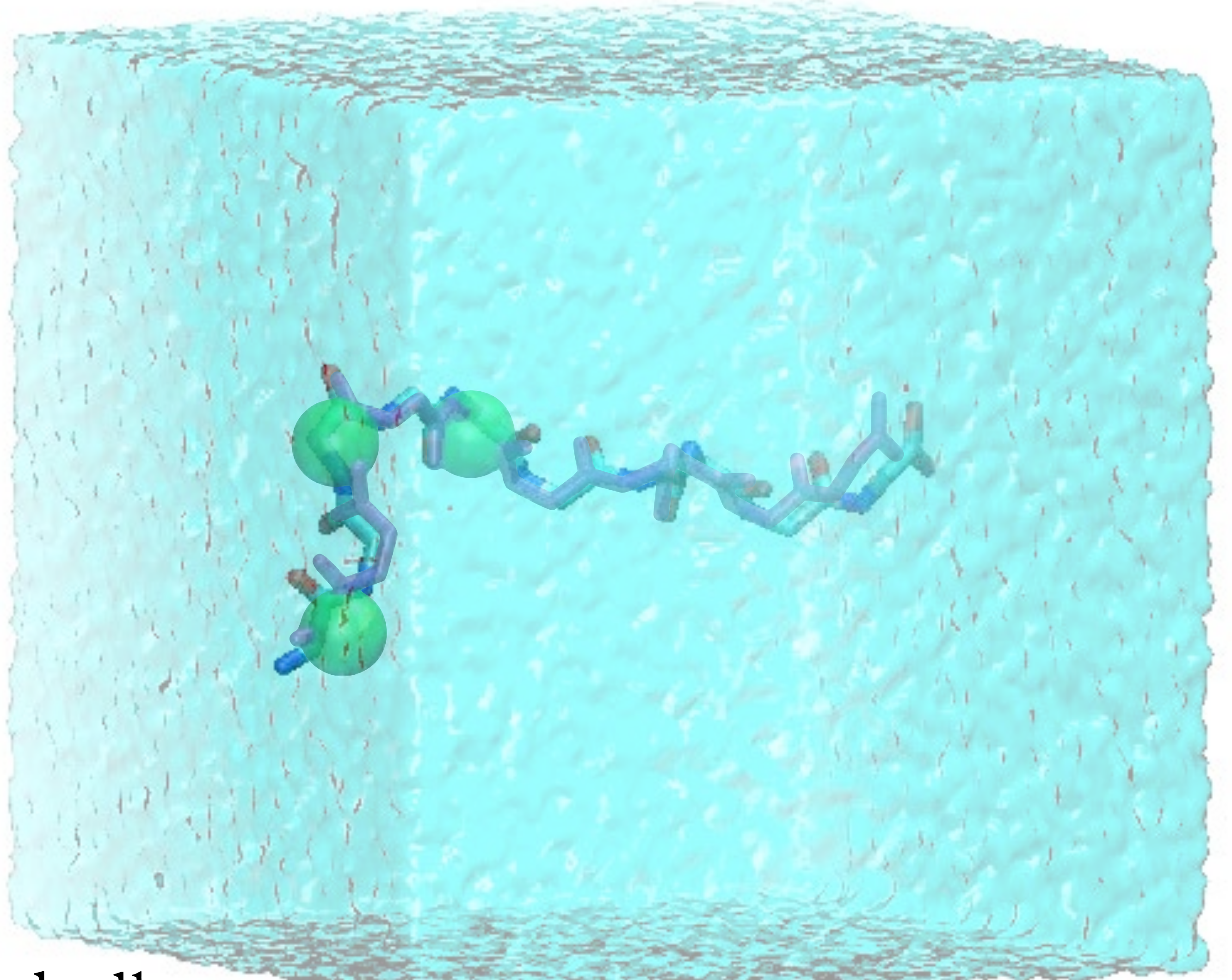
Gilson, M. K. et al. *Biophys. J.*, **1997**, *72*, 1047-1069

Hermans, J.; Wang, L. *J. Am. Chem. Soc.* **1997**, *119*, 2707-2714

THE DOUBLE-ANNIHILATION STRATEGY



site



bulk

$$K_{\rm eq} = \frac{\int_{\rm site} {\rm d}\mathbf{1} \int {\rm d}\mathbf{x}\ {\rm e}^{-\beta U_1}}{\int_{\rm site} {\rm d}\mathbf{1} \int {\rm d}\mathbf{x}\ {\rm e}^{-\beta (U_1+u_c)}} \times \frac{\int_{\rm site} {\rm d}\mathbf{1} \int {\rm d}\mathbf{x}\ {\rm e}^{-\beta (U_1+u_c)}}{\int_{\rm site} {\rm d}\mathbf{1} \int {\rm d}\mathbf{x}\ {\rm e}^{-\beta (U_1+u_c+u_o)}} \times \frac{\int_{\rm site} {\rm d}\mathbf{1} \int {\rm d}\mathbf{x}\ {\rm e}^{-\beta (U_1+u_c+u_o)}}{\int_{\rm site} {\rm d}\mathbf{1} \int {\rm d}\mathbf{x}\ {\rm e}^{-\beta (U_1+u_c+u_o+u_p)}} \times \frac{\int_{\rm site} {\rm d}\mathbf{1} \int {\rm d}\mathbf{x}\ {\rm e}^{-\beta (U_1+u_c+u_o+u_p)}}{\int_{\rm site} {\rm d}\mathbf{1} \int {\rm d}\mathbf{x}\ {\rm e}^{-\beta (U_1+u_c+u_o+u_p+u_r)}}$$

$$\times \frac{\int_{\rm site} {\rm d}\mathbf{1} \int {\rm d}\mathbf{x}\ {\rm e}^{-\beta (U_1+u_c+u_o+u_p+u_r)}}{\int_{\rm site} {\rm d}\mathbf{1} \int {\rm d}\mathbf{x}\ {\rm e}^{-\beta (U_0+u_c+u_o+u_p+u_r)}}$$

$$\times \frac{\int_{\rm bulk} {\rm d}\mathbf{1} \int {\rm d}\mathbf{x}\ {\rm e}^{-\beta (U_0+u_c+u_o+u_p+u_r)}}{\int_{\rm bulk} {\rm d}\mathbf{1}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int {\rm d}\mathbf{x}\ {\rm e}^{-\beta (U_0+u_c+u_o)}}$$

$$\times \frac{\int_{\rm bulk} {\rm d}\mathbf{1}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int {\rm d}\mathbf{x}\ {\rm e}^{-\beta (U_0+u_c+u_o)}}{\int_{\rm bulk} {\rm d}\mathbf{1}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int {\rm d}\mathbf{x}\ {\rm e}^{-\beta (U_0+u_c)}}$$

$$\times \frac{\int_{\rm bulk} {\rm d}\mathbf{1}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int {\rm d}\mathbf{x}\ {\rm e}^{-\beta (U_0+u_c)}}{\int_{\rm bulk} {\rm d}\mathbf{1}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int {\rm d}\mathbf{x}\ {\rm e}^{-\beta (U_1+u_c)}}$$

$$\times \frac{\int_{\rm bulk} {\rm d}\mathbf{1}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int {\rm d}\mathbf{x}\ {\rm e}^{-\beta (U_1+u_c)}}{\int_{\rm bulk} {\rm d}\mathbf{1}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int {\rm d}\mathbf{x}\ {\rm e}^{-\beta U_1}}$$

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

THE LONG-STANDING PROTEIN-LIGAND PROBLEM

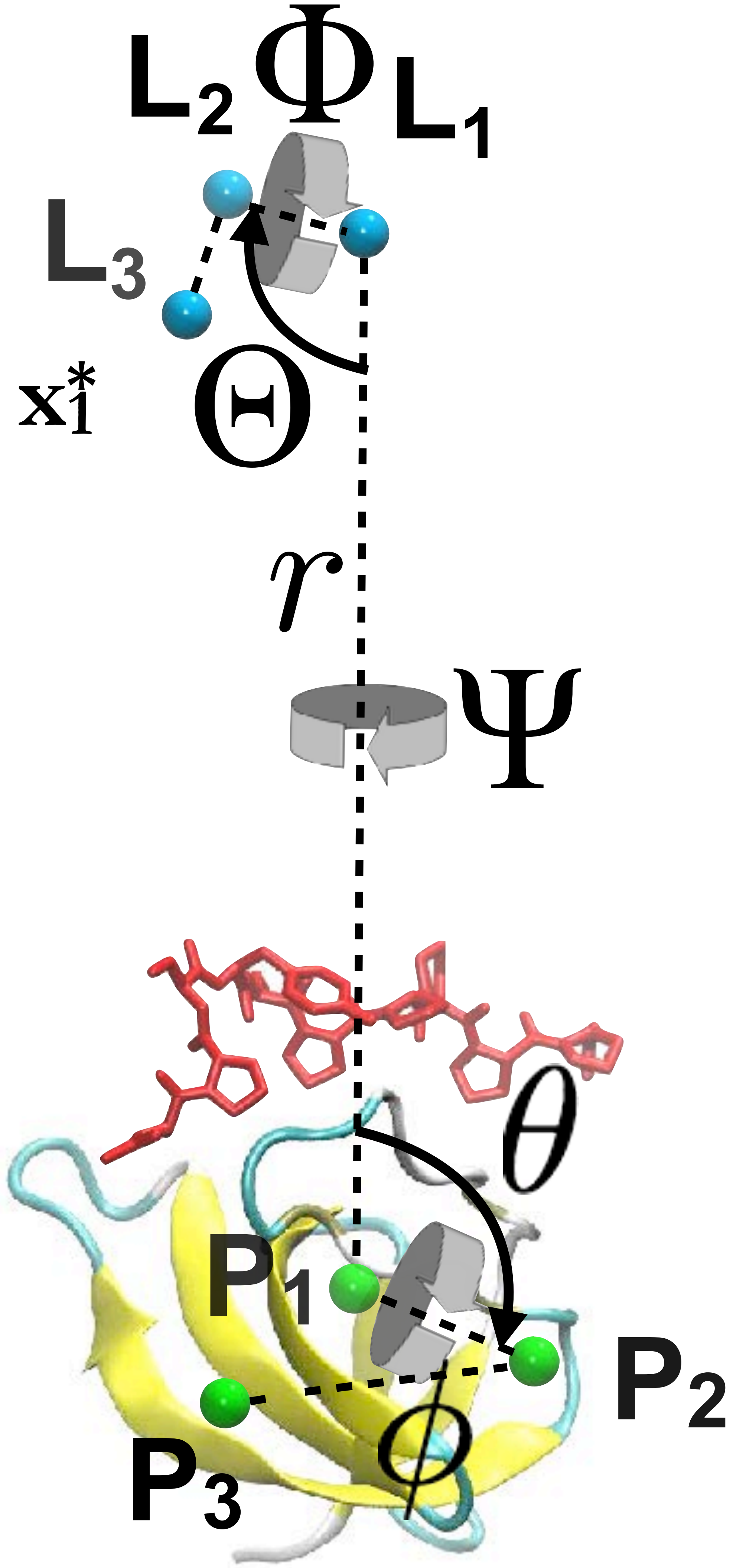
THE ALCHEMICAL ROUTE

THE GEOMETRICAL ROUTE

RELATIVE BINDING FREE ENERGIES

ALANINE SCANNING

BEYOND PROTEIN-LIGAND BINDING



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

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

$$\times \frac{\int_{\rm site} {\rm d}\mathbf{1} \int {\rm d}\mathbf{x} \ldots}{\int_{\rm site} {\rm d}\mathbf{1} \int {\rm d}\mathbf{x}\ {\rm e} \ldots}$$

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

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

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

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

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

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

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

$$\frac{\int_{\rm site} {\rm d}\mathbf{1} \int {\rm d}\mathbf{x}\ {\rm e}^{-\beta (U+u_c)}}{\int_{\rm site} {\rm d}\mathbf{1} \int {\rm d}\mathbf{x}\ {\rm e}^{-\beta u_\Theta} {\rm e}^{-\beta (U+u_c)}} = {\rm e}^{+\beta \Delta G_\Theta^{\rm site}} = \frac{{\rm d}\Theta\ {\rm e}^{-\beta w_{\rm site}(\Theta)}}{{\rm d}\Theta\ {\rm e}^{-\beta \left(w_{\rm site}(\Theta)+u_\Theta\right)}}$$

Yu, Y. B. et al. *Biophys. J.* **2001**, *81*, 1632-1642
Woo, H. J.; Roux, B. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6825-6830
Gumbart, J. C.; Roux, B.; Chipot, C. *J. Chem. Theory Comput.* **2013,** *9*, 794-802

A GEOMETRICAL ROUTE

Robust experimental data.

Sampling constitutes the primary source of error.

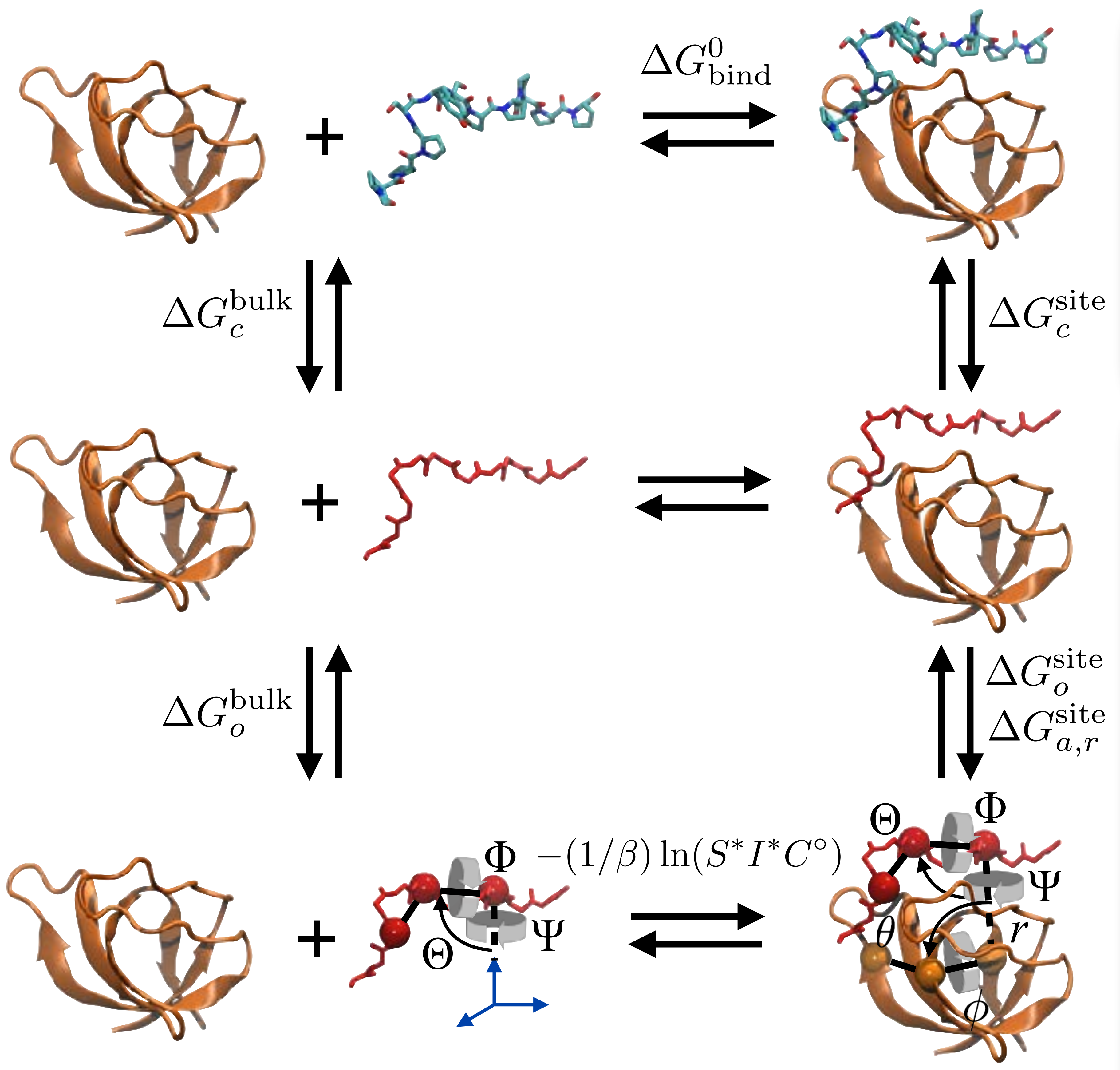
The protein: Abl Src homology domain 3.

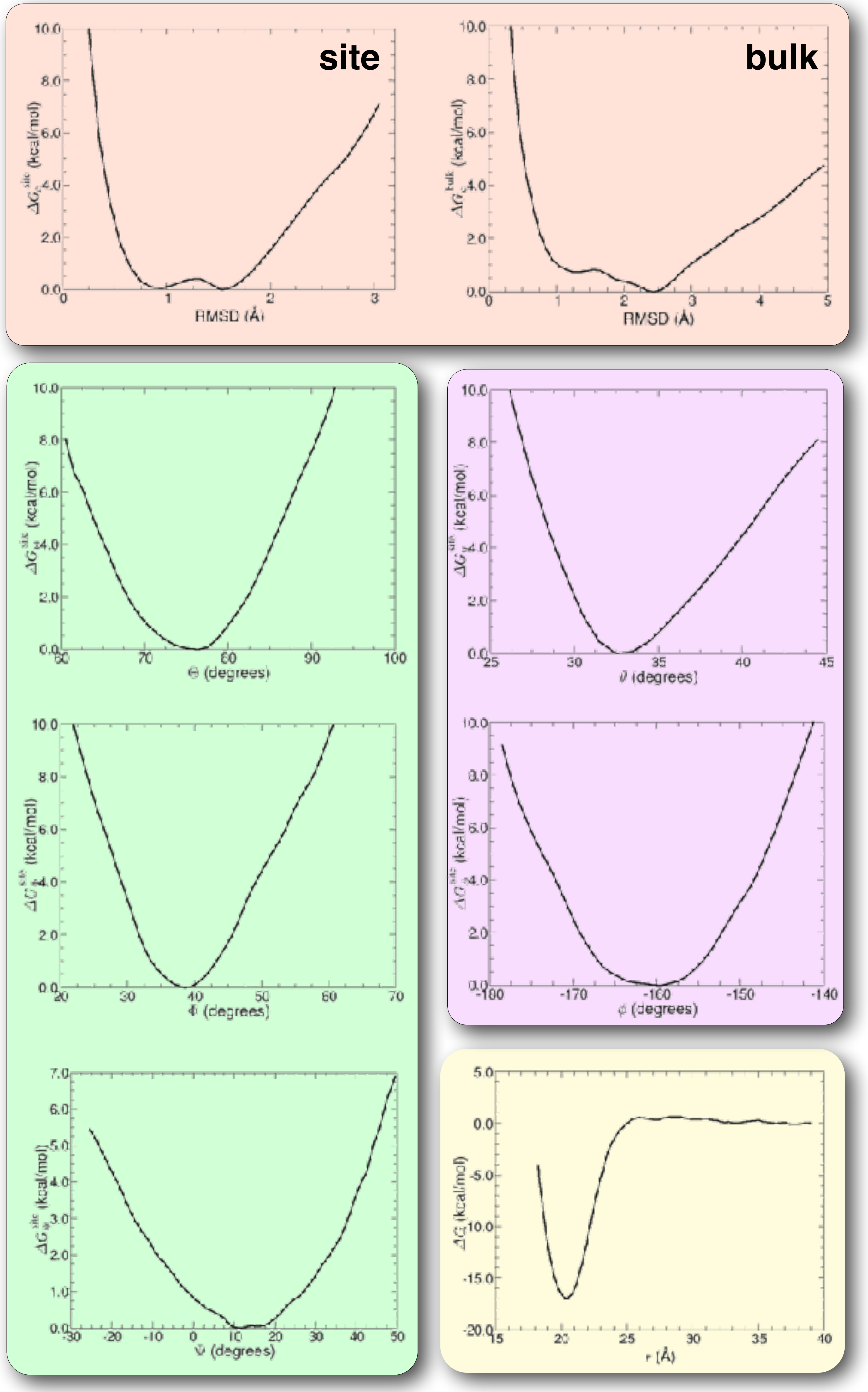
The binder: APSYSPPPPP (p41).

ΔG^0 = -7.94 kcal/mol (experimental).

Fully geometrical route: **-7.8 ± 0.9 kcal/mol**

Fully alchemical route: **-7.7 ± 1.0 kcal/mol**





Pisabarro, M. T.; Serrano, L. *Biochemistry* **1996**, *35*, 10634-10640

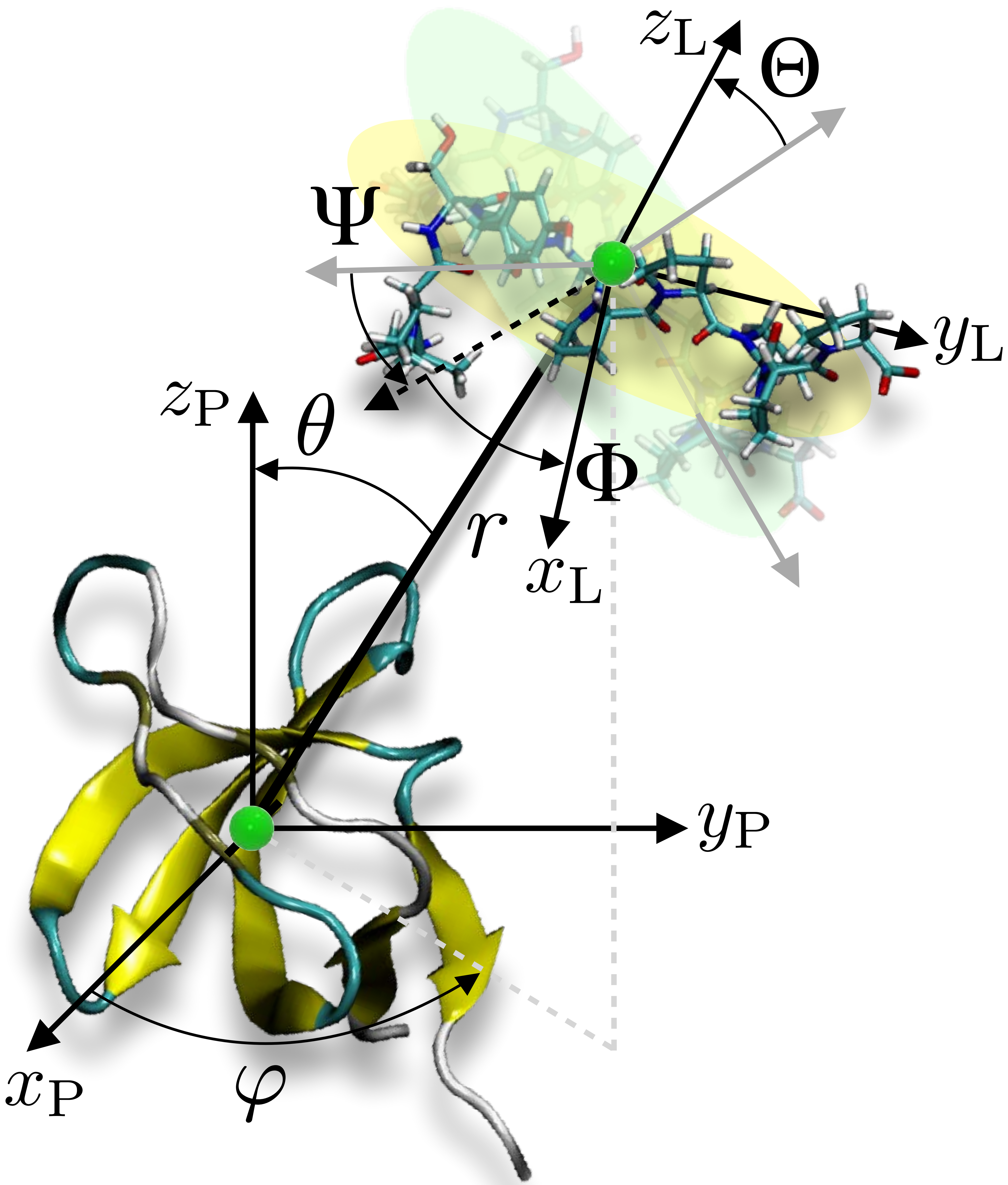
Pisabarro, M. T.; Serrano, L.; Wilmanns, M. *J. Mol. Biol.* **1998**, *281*, 513-521

Gumbart, J. C.; Roux, B.; Chipot, C. *J. Chem. Theory Comput.* **2013,** *9*, 794-802

Fu, H.; Cai, W.; Hénin, J.; Roux, B.; Chipot, C. *J. Chem. Theory Comput.* **2017**, *13*, 5173-5178

Fu, H. and Gumbart, J. C. and Chen, H. and Shao, X. and Cai, W. and Chipot, C. *J. Chem. Inf. Model.* **2018**, *18*, 556-560

NEW COLLECTIVE VARIABLES



Euler angles:

$$\begin{cases} \Phi = \mathrm{atan2}\left(2(q_0q_1 + q_2q_3), 1 - 2(q_1^2 + q_2^2)\right) \\ \Theta = \mathrm{asin}\left(2(q_0q_2 - q_3q_1)\right) \\ \Psi = \mathrm{atan2}\left(2(q_0q_3 + q_1q_2), 1 - 2(q_2^2 + q_3^2)\right) \end{cases}$$

Polar angles:

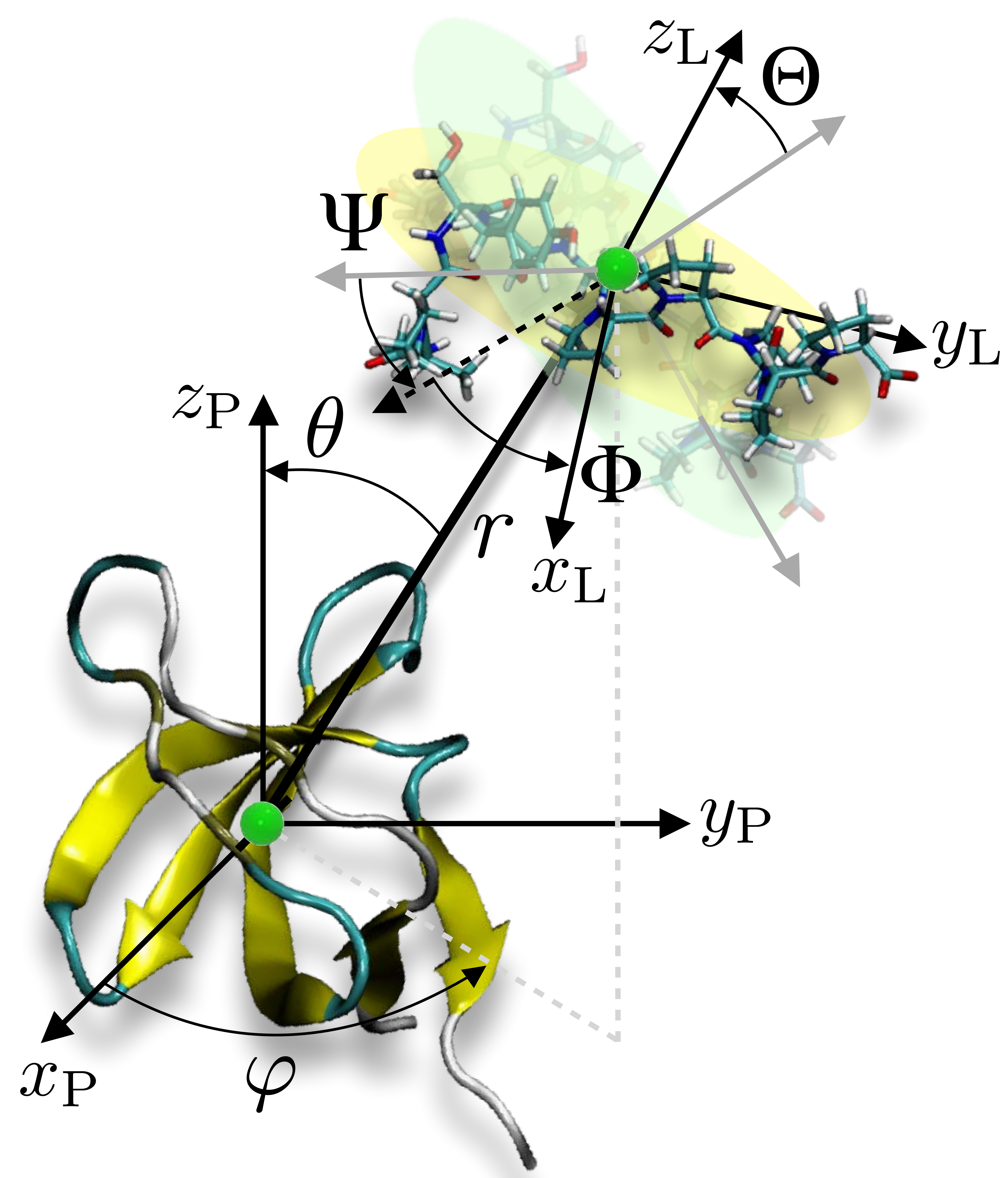
$$\begin{cases} \theta = \mathrm{acos}(z) \\ \varphi = \mathrm{atan2}(y, x) \end{cases}$$

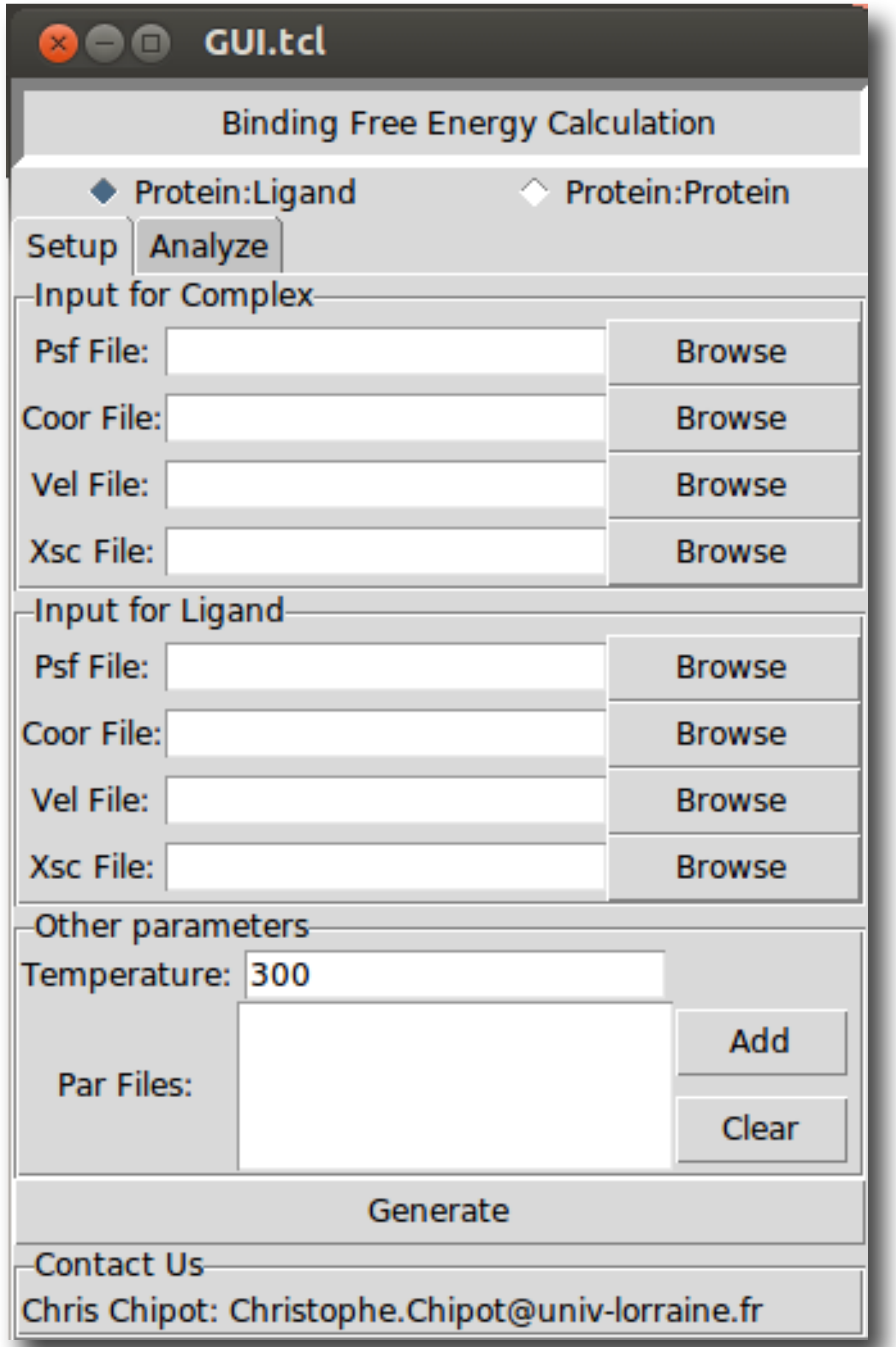
- Colvars scripted functions

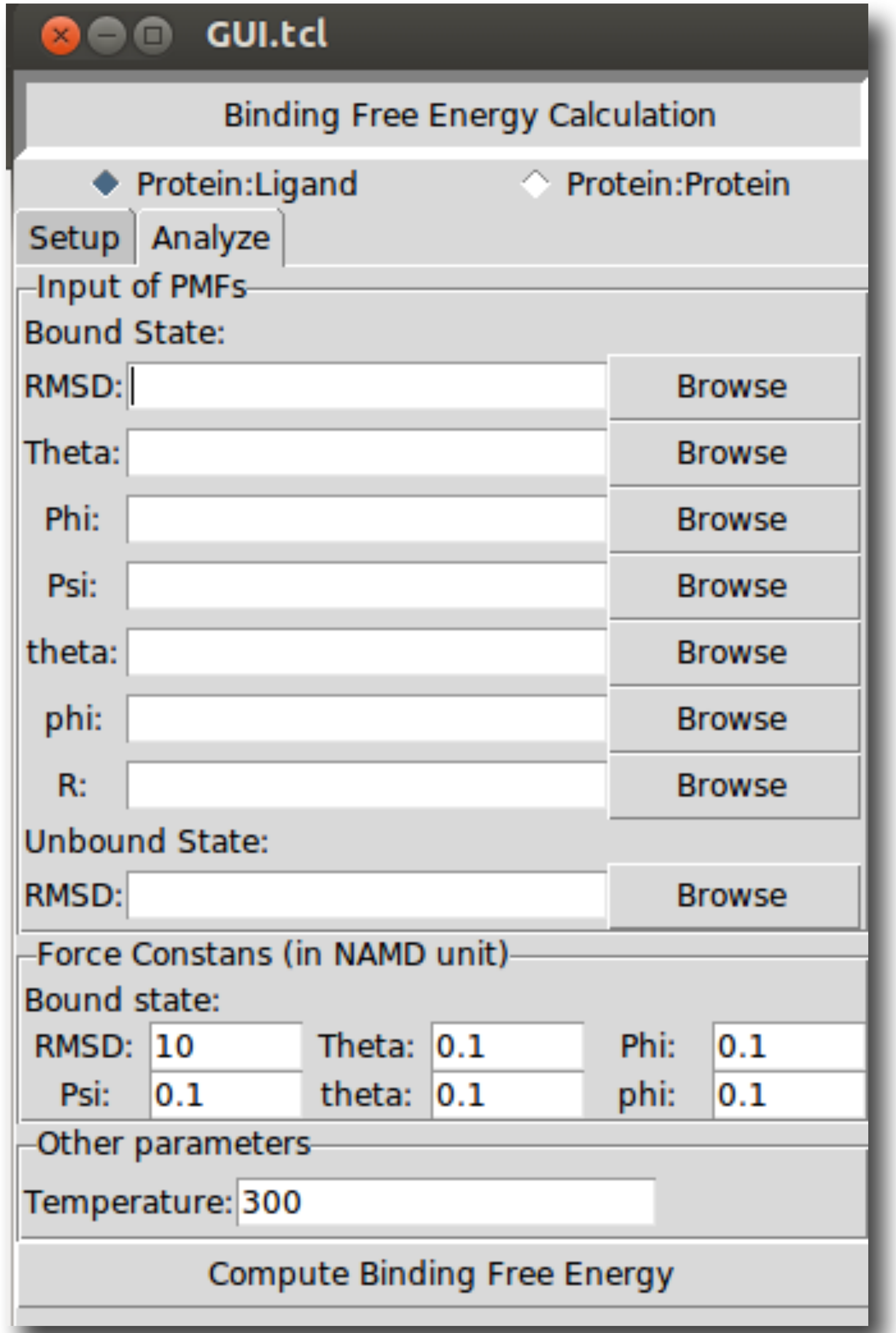
```
# Euler angles
# Phi
namespace eval eulerPhi { }
proc calc_eulerPhi { args } {

    global eulerPhi::q0
    global eulerPhi::q1
    global eulerPhi::q2
    global eulerPhi::q3

    set q0 [ lindex [ lindex $args 0 ] 0 ]
    set q1 [ lindex [ lindex $args 0 ] 1 ]
    set q2 [ lindex [ lindex $args 0 ] 2 ]
    set q3 [ lindex [ lindex $args 0 ] 3 ]

    set f  [ expr 180 / 3.1415926 * atan2(2 * ($q0 * $q1 + $q2 * $q3), 1 - 2 * ($q1 * $q1 + $q2 * $q2)) ]

    return $f
}
```

- Handled by the extended adaptive biasing force algorithm (eABF)

NEW COLLECTIVE VARIABLES







- The long-standing protein-ligand problem
- The alchemical route
- The geometrical route
- **Relative binding free energies**
- Alanine scanning
- Beyond protein-ligand binding

THE LONG-STANDING PROTEIN-LIGAND PROBLEM

Relative binding affinity — alternate guests

$$\text{protein} + \text{ligand}^0_{\text{A}} \underset{}{\overset{\Delta G^0_{\text{A}}}{\rightleftharpoons}} \text{protein:ligand}^0_{\text{A}}$$

$$\Delta G_a^{\text{bulk}} \qquad \Delta G_a^{\text{site}}$$

$$\text{protein} + \text{ligand}^0_{\text{B}} \underset{}{\overset{\Delta G^0_{\text{B}}}{\rightleftharpoons}} \text{protein:ligand}^0_{\text{B}}$$

Relative binding affinity — alternate hosts

$$\text{protein}_{\text{A}} + \text{ligand}^0 \underset{}{\overset{\Delta G^0_{\text{A}}}{\rightleftharpoons}} \text{protein}_{\text{A}}\text{:ligand}^0$$

$$\Delta G_a^{\text{bulk}} \qquad \Delta G_a^{\text{site}}$$

$$\text{protein}_{\text{B}} + \text{ligand}^0 \underset{}{\overset{\Delta G^0_{\text{B}}}{\rightleftharpoons}} \text{protein}_{\text{B}}\text{:ligand}^0$$

$$\Delta\Delta G^0 = \Delta G^0_{\text{B}} - \Delta G^0_{\text{A}} = \Delta G_a^{\text{site}} - \Delta G_a^{\text{bulk}}$$

- Easier to carry out than standard binding free-energy calculations
- Cheaper than standard binding free-energy calculations
- Well-suited to series of congeneric compounds
- May require the introduction of geometric restraints

Chipot, C.; Pohorille, A. Free-energy calculations. Springer 2007.

EVALUATING THE CONTRIBUTIONS (P5)

contribution	PMF(kcal/mol)
ΔG_c^{site} (1)	-6.6
ΔG_Θ^{site} (2)	-0.4
ΔG_Φ^{site} (2)	-0.3
ΔG_Ψ^{site} (2)	-0.3
ΔG_θ^{site} (3)	-0.2
ΔG_ϕ^{site} (3)	-0.2
ΔG_r^{site} (4)	
$-(1/\beta)\ \ln(S^* I^* C^\circ)$ (5)	-15.6
$\Delta G_{decouple}^{site}$ (6)	
$\Delta G_r^{bulk} + \Delta G_a^{bulk}$ (7)	
ΔG_o^{bulk} (8)	11.5
ΔG_{couple}^{bulk} (9)	
ΔG_c^{bulk} (10)	6.6
ΔG_{bind}	-5.5

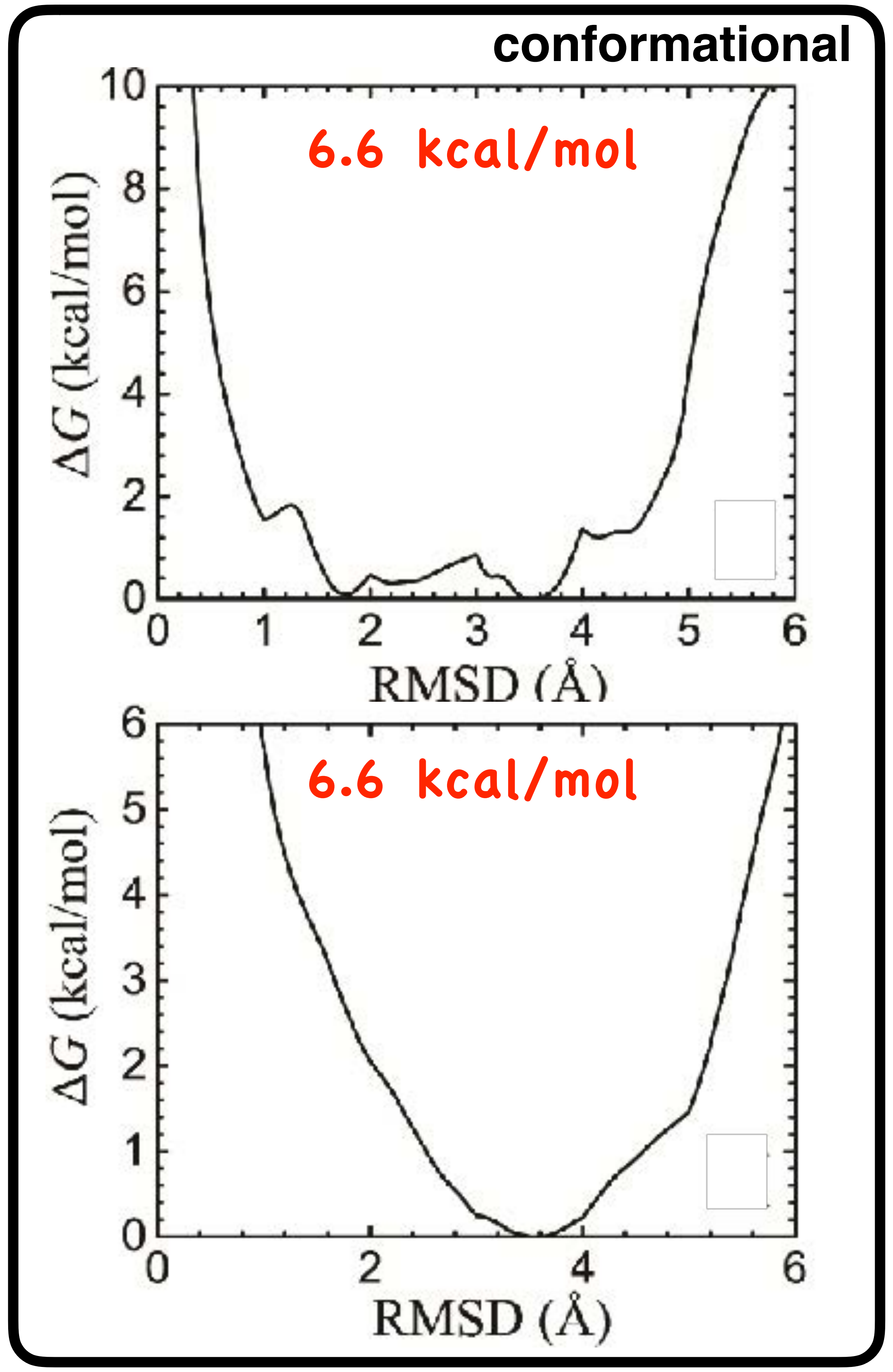
p5: APTYPPPLNP

$$\text{Abl-SH3} + \text{p41} \underset{}{\overset{\Delta G^\circ_{p41}}{\rightleftharpoons}} \text{Abl-SH3:p41}$$

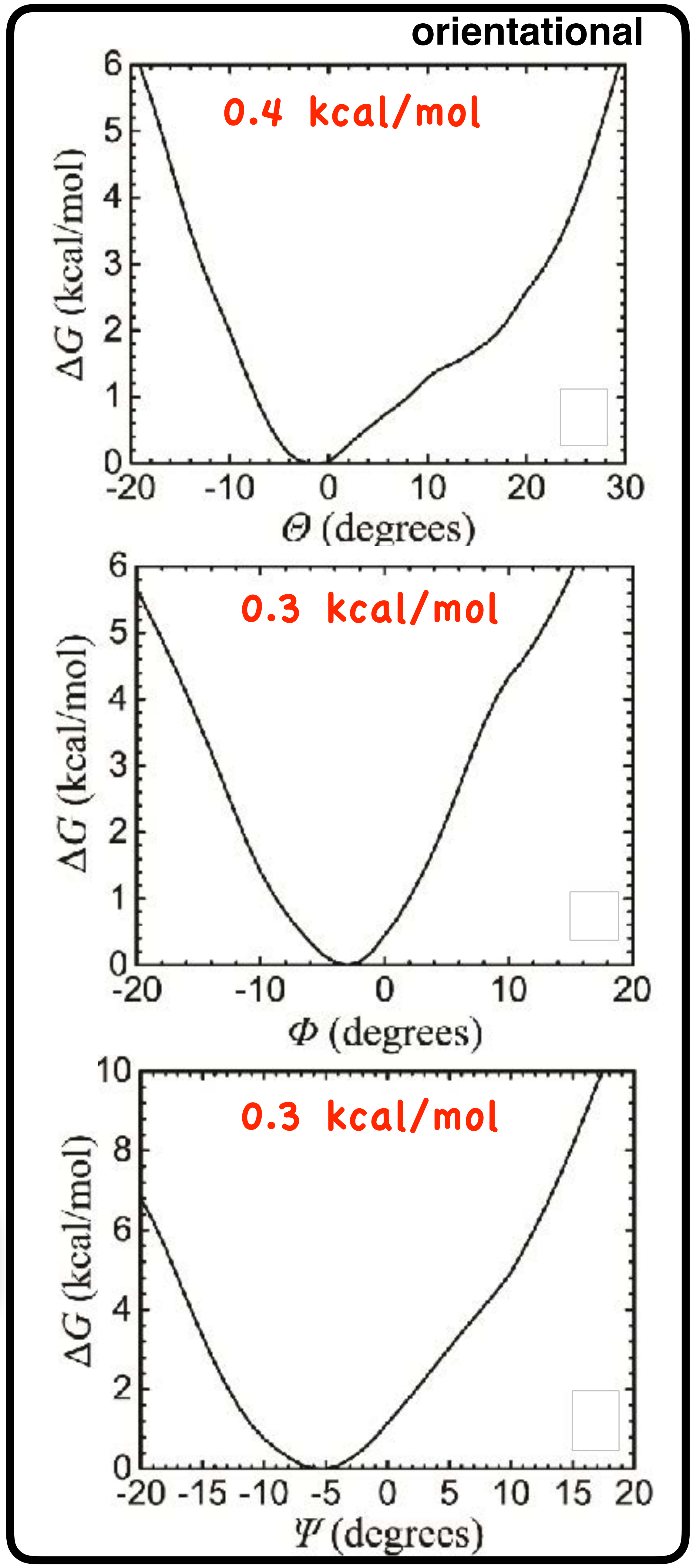
$\Delta G_{unbound}$ (Abl-SH3 + p41 ⇅ Abl-SH3 + p5) $\qquad$ ΔG_{bound} (Abl-SH3:p41 ⇅ Abl-SH3:p5)

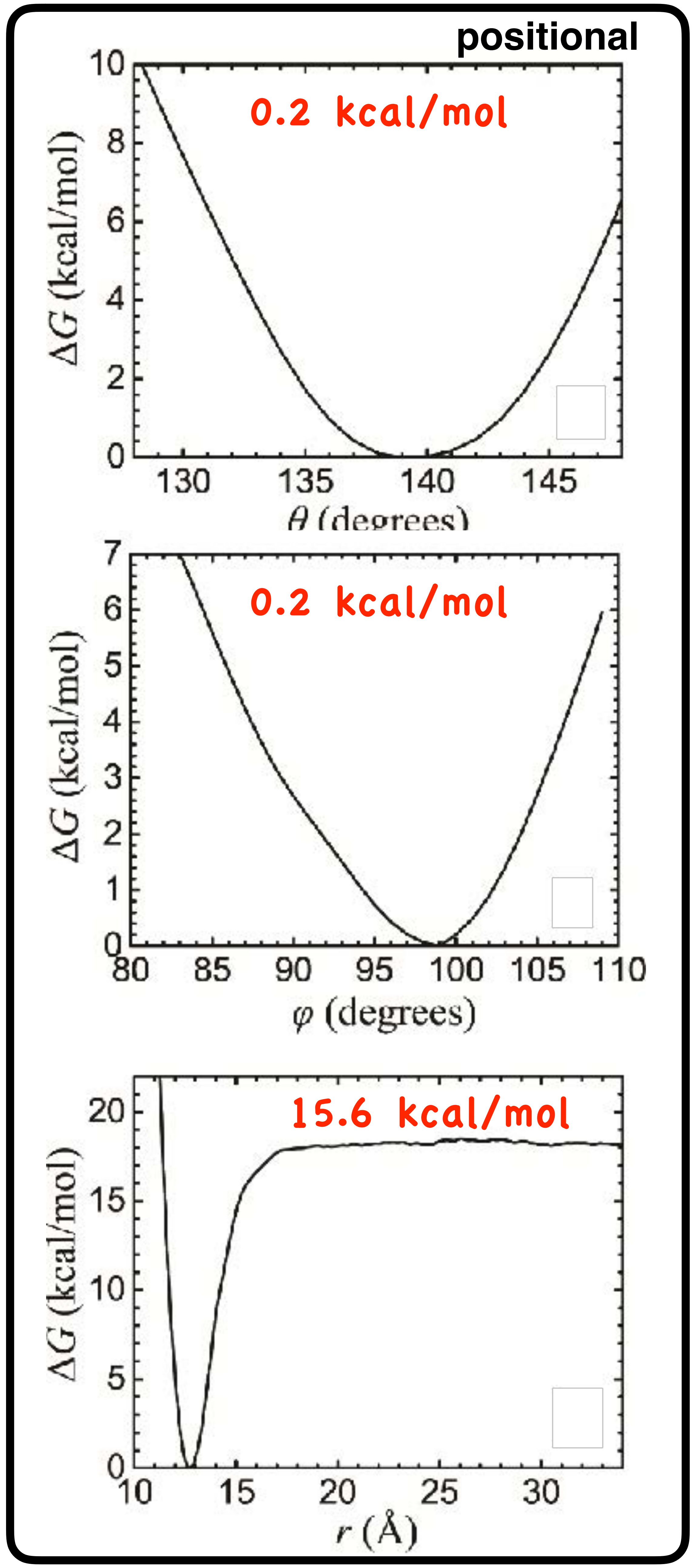
$$\text{Abl-SH3} + \text{p5} \underset{\Delta G^\circ_{p5}}{\rightleftharpoons} \text{Abl-SH3:p5}$$

$\Delta G^\circ_{p41} - \Delta G^\circ_{p5} = \Delta G_{bound} - \Delta G_{unbound}$ =125.7-125.7=0



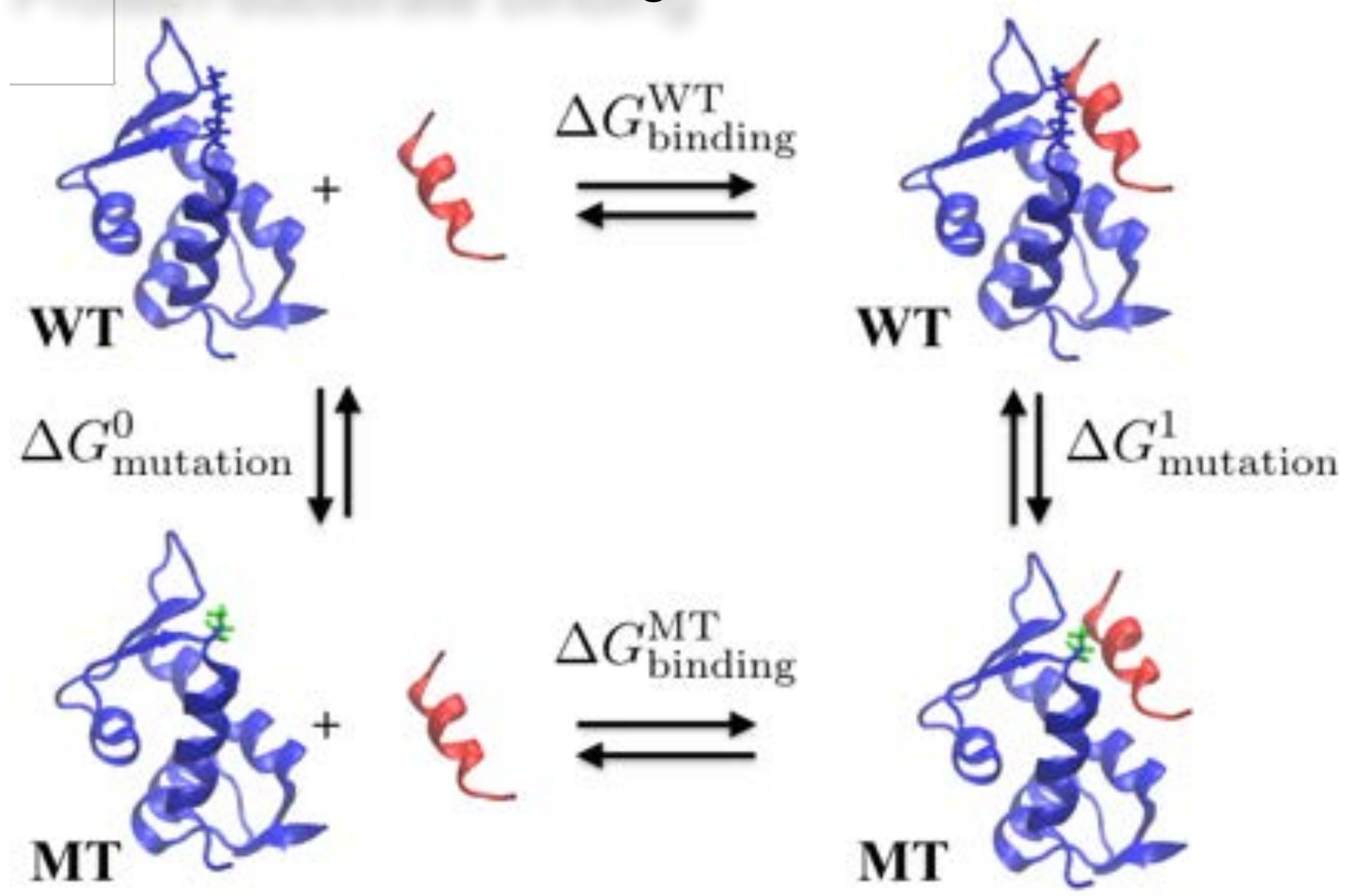
$\Delta G^\circ_{bind}(\text{exp}) = -5.52$ kcal/mol



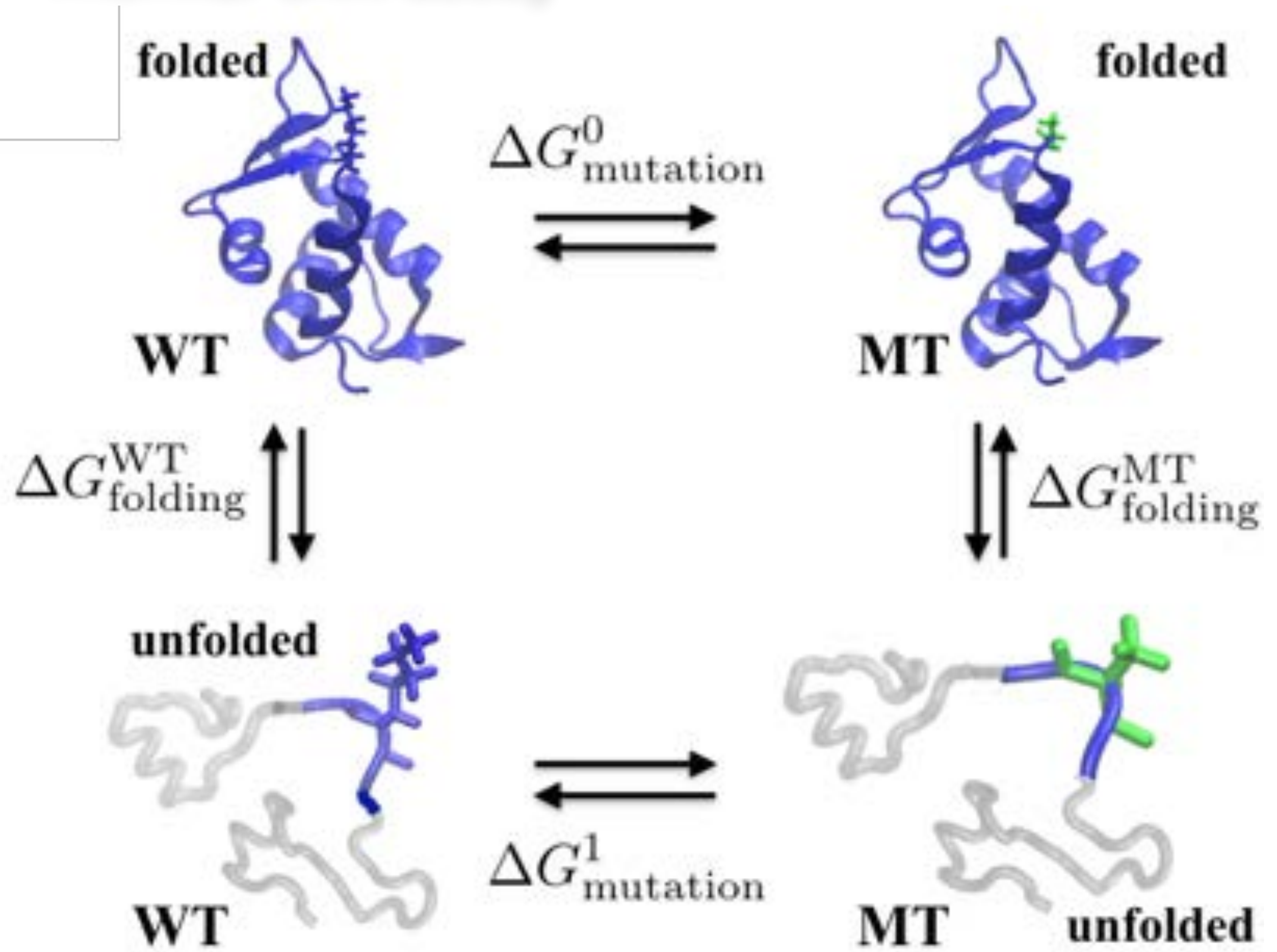


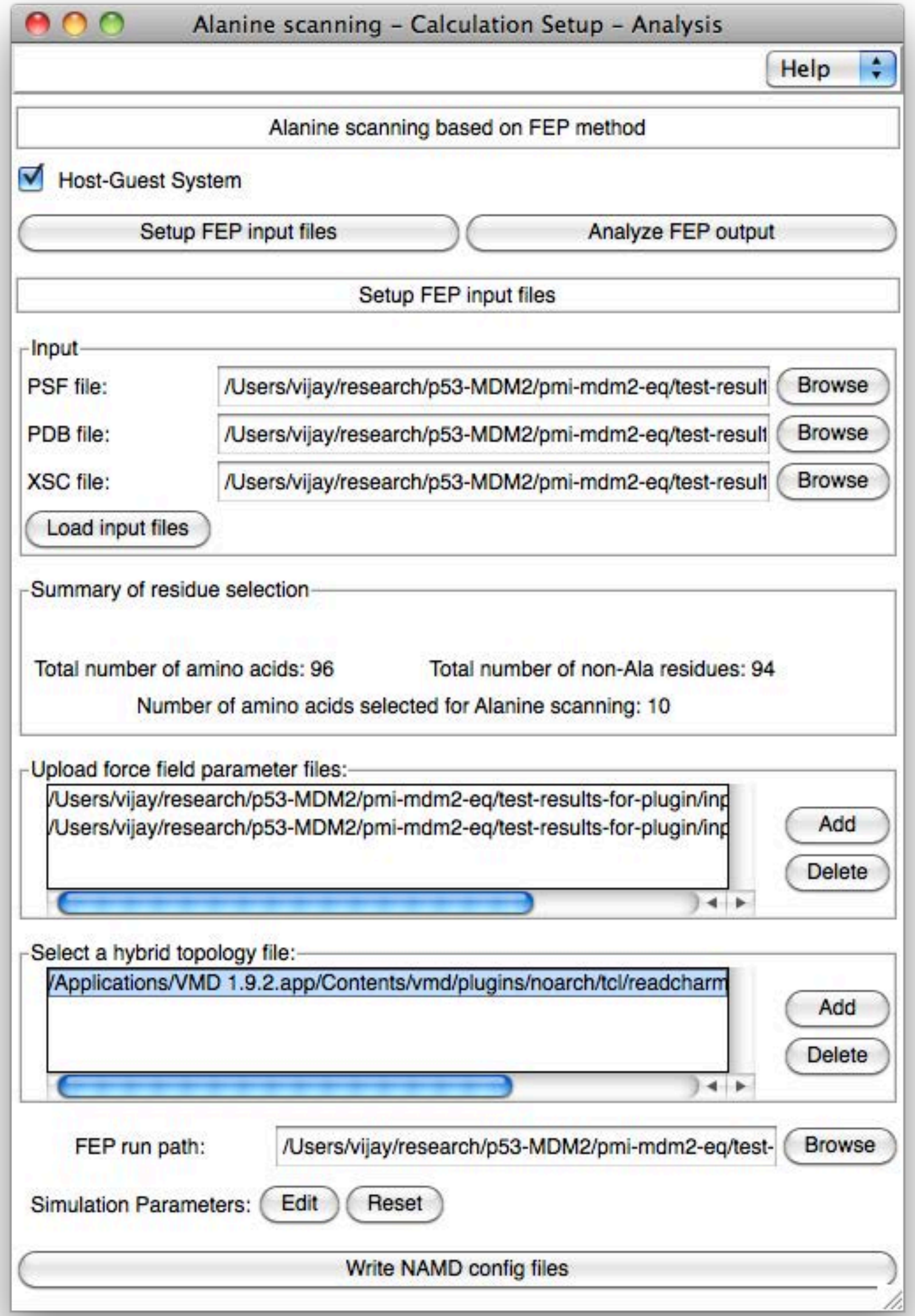
ALANINE-SCANNING EXPERIMENTS

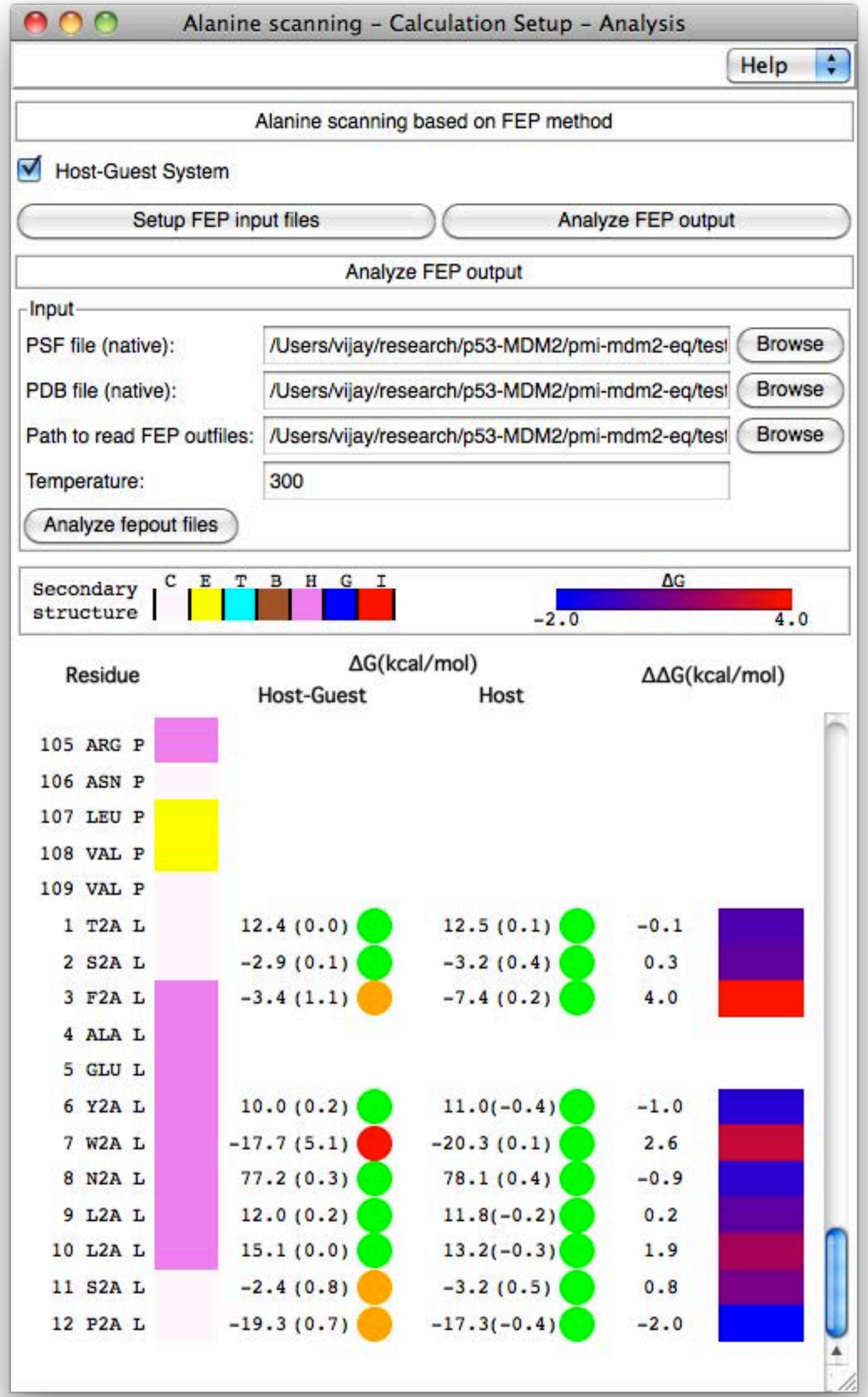
Protein-substrate binding



Thermal-shift assay





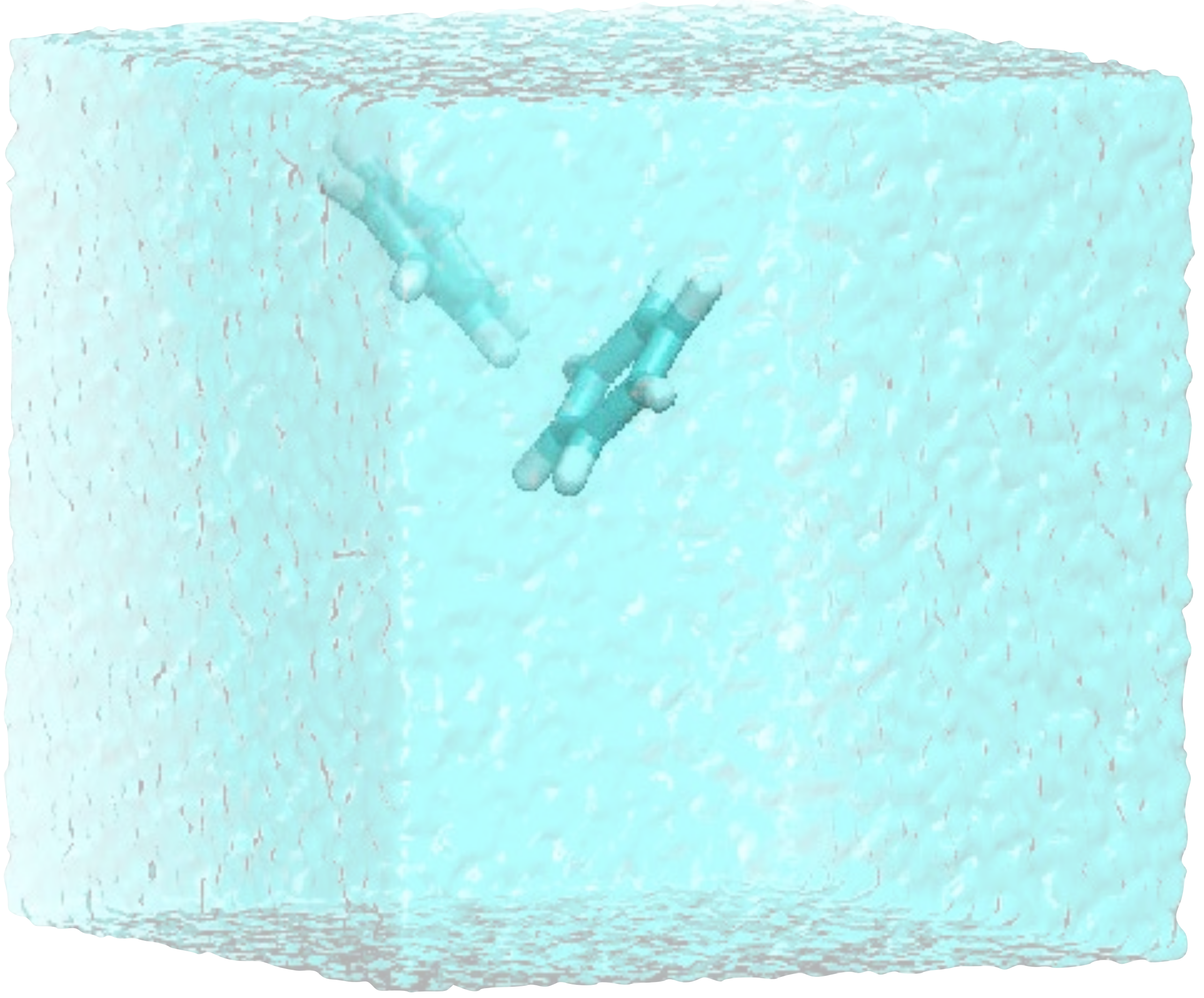


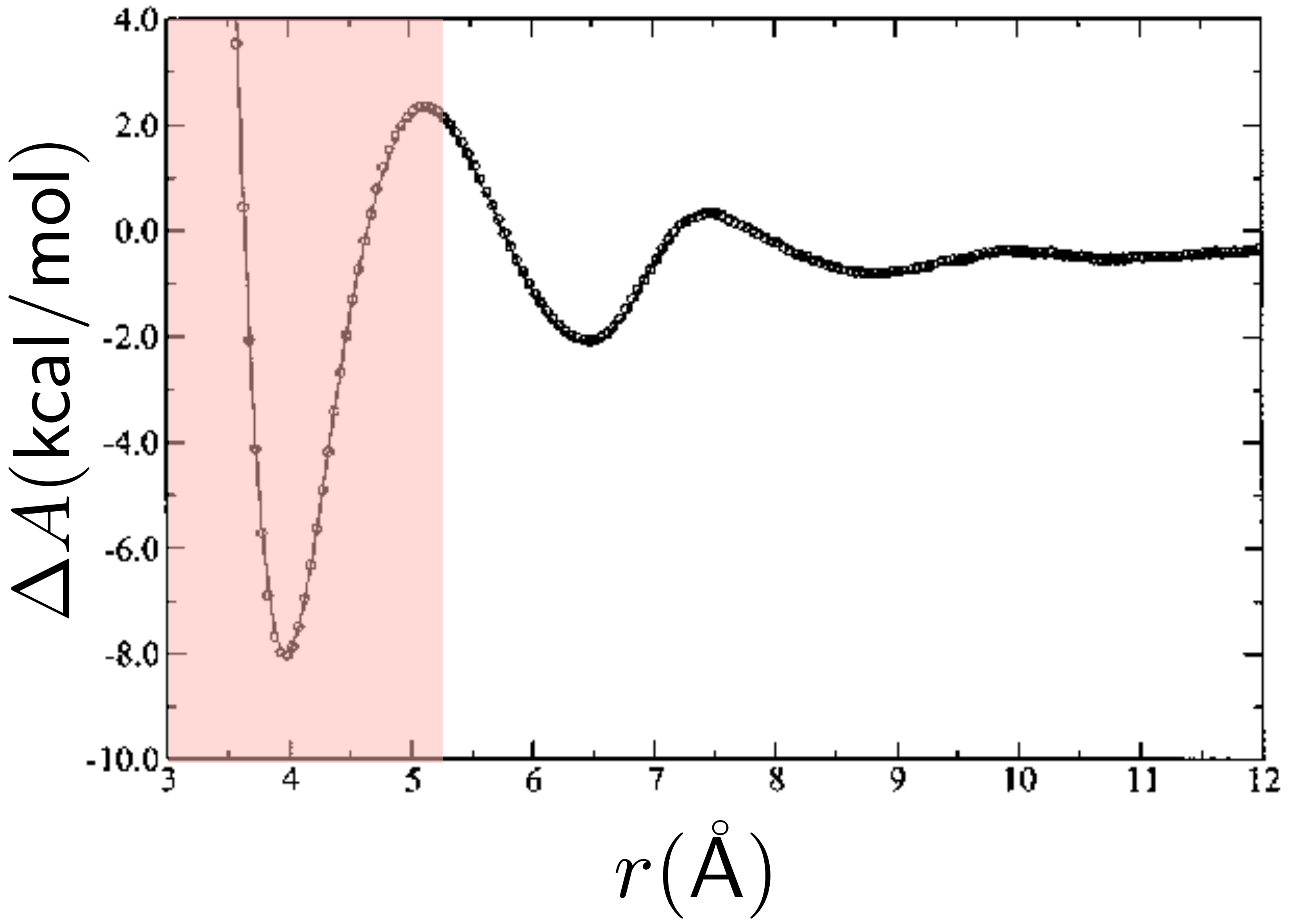
Ramadoss, V.; Dehez, F. ; Chipot, C. *J. Chem. Info. Model.* **2016**, *56*, 1122-1126

PROTEIN-PROTEIN ASSOCIATION

Measuring binding constants from one-dimensional separation potentials of mean force is justified in the limit of all other degrees of freedom being appropriately sampled. This is true for small, fast-relaxing molecular species.

$$K_a = 4\pi \int_0^{R_c} \mathrm{d}r \; r^2 \; \exp[-\beta \Delta A(r)]$$





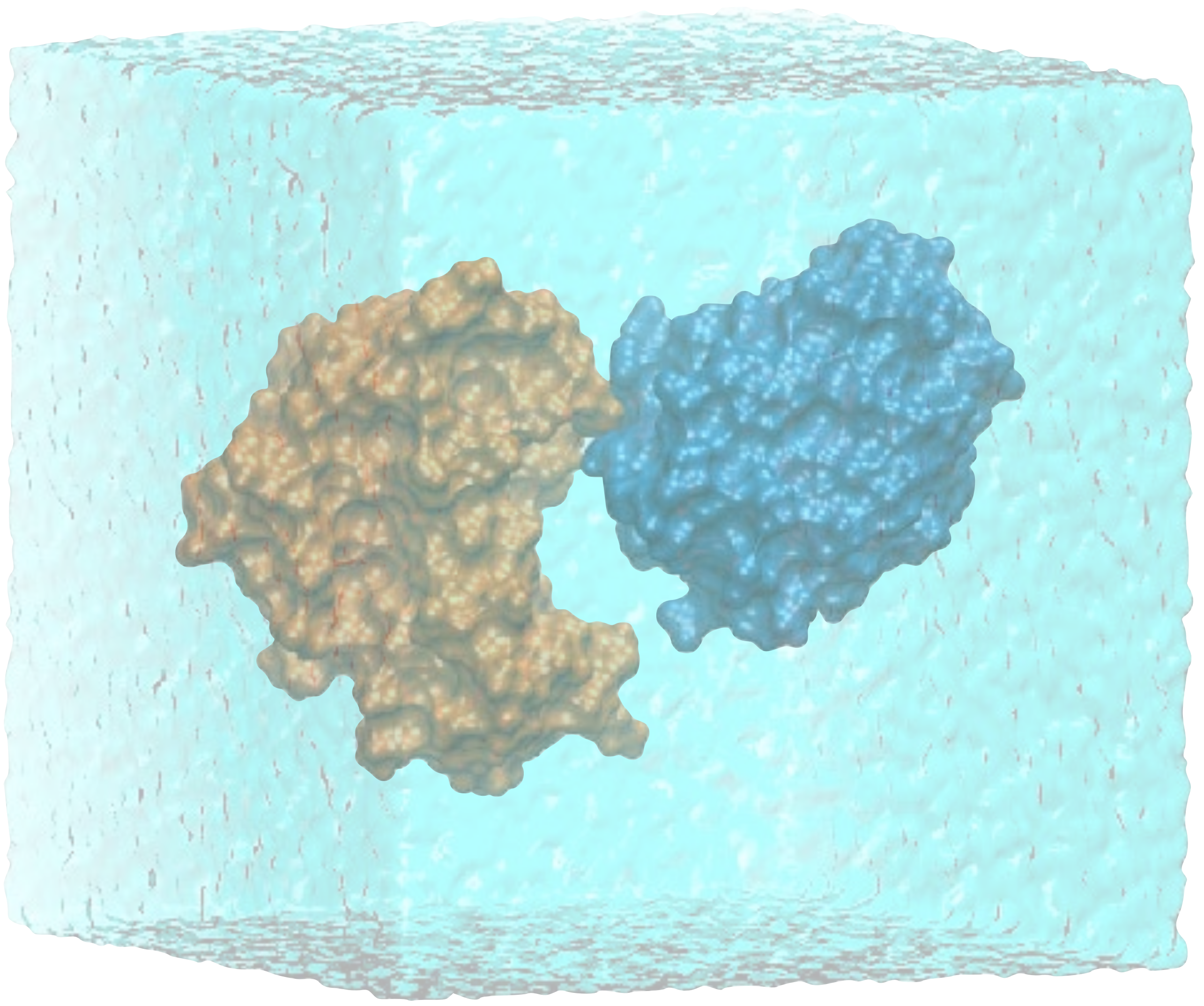
Shoup, D.; Szabo, A. *Biophys. J.* **1982**, *40*, 33-39

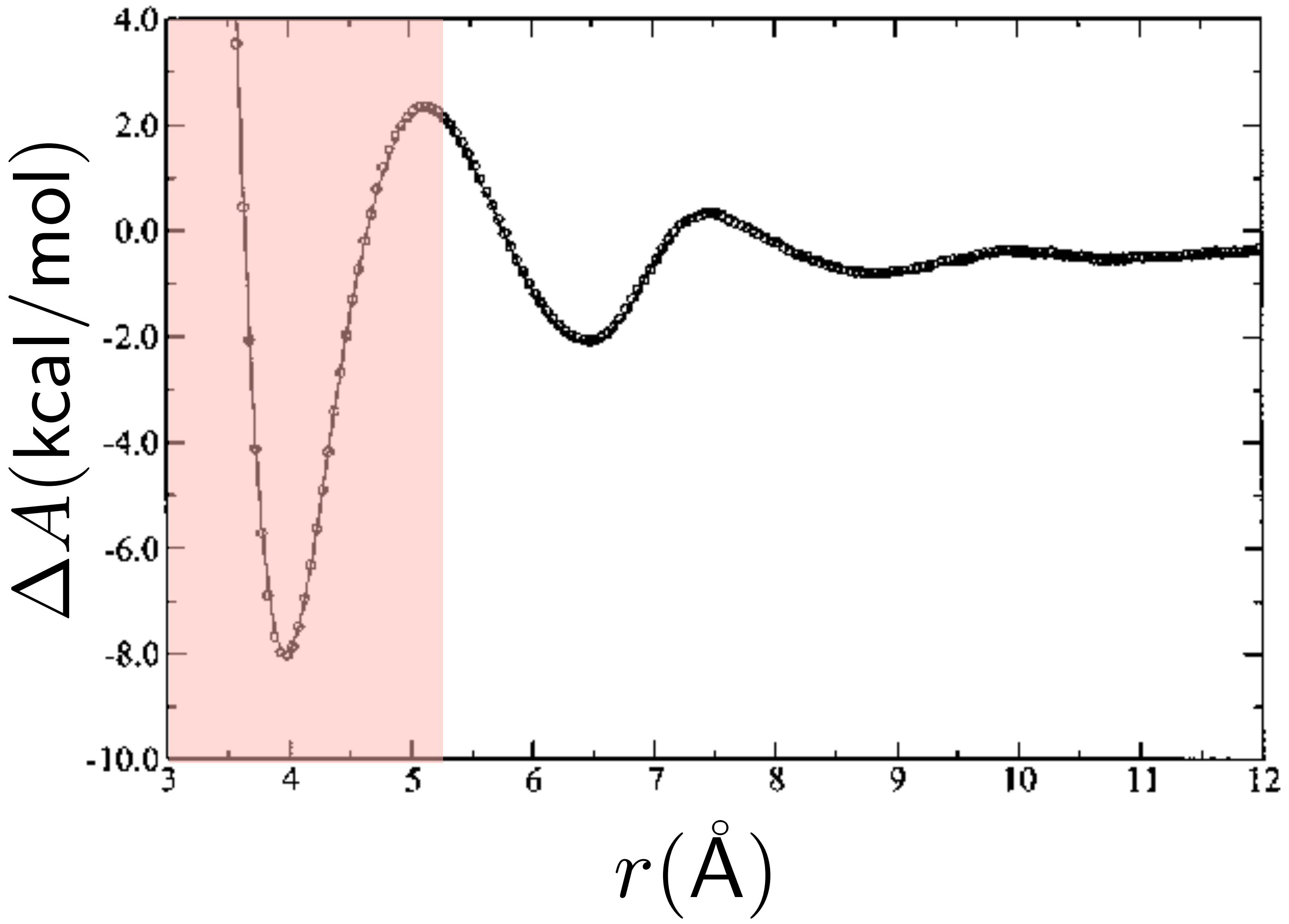
Woo, H. J.; Roux, B. *Proc. Natl. Acad. Sci. USA* **2005**, 102, 6825-6830

PROTEIN-PROTEIN ASSOCIATION

Measuring binding constants from one-dimensional separation potentials of mean force is justified in the limit of all other degrees of freedom being appropriately sampled. This is true for small, fast-relaxing molecular species.

$$K_a = 4\pi \int_0^{R_c} \mathrm{d}r \; r^2 \; \exp[-\beta \Delta A(r)]$$



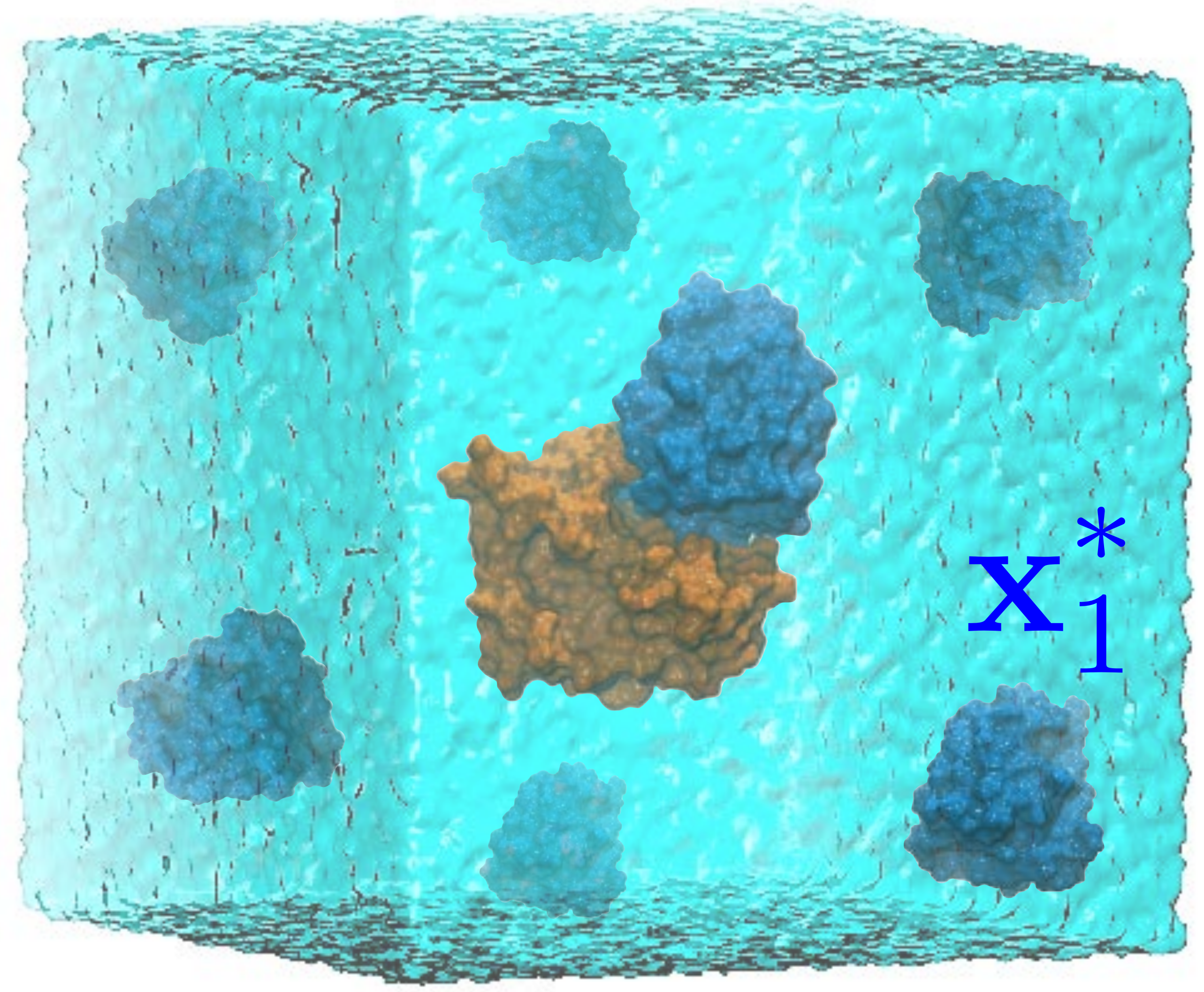


In more complex molecular assemblies, e.g., protein-ligand complexes, the partners acquire upon separation additional configurational - i.e., conformational, positional and orientational entropy, not easily captured over timescales amenable to molecular dynamics.

Shoup, D.; Szabo, A. *Biophys. J.* **1982**, *40*, 33-39

Gumbart, J. C.; Roux, B.; Chipot, C. *J. Chem. Theor. Comput.* **2013**, *9*, 3789-3798

PROTEIN-PROTEIN ASSOCIATION



$$[\text{barstar}] = \frac{N}{V_{\text{bulk}}}$$

$$K_{\text{eq}} = \frac{1}{[\text{barstar}]} \frac{N \int_{\text{site}} \text{d}\mathbf{1} \int_{\text{bulk}} \text{d}\mathbf{2} \ldots \int_{\text{bulk}} \text{d}\mathbf{N} \int \text{d}\mathbf{x}\ \text{e}^{-\beta U}}{\int_{\text{bulk}} \text{d}\mathbf{1} \int_{\text{bulk}} \text{d}\mathbf{2} \ldots \int_{\text{bulk}} \text{d}\mathbf{N} \int \text{d}\mathbf{x}\ \text{e}^{-\beta U}}$$

$$= \frac{1}{[\text{barstar}]} \frac{N \int_{\text{site}} \text{d}\mathbf{1} \int \text{d}\mathbf{x}\ \text{e}^{-\beta U}}{\int_{\text{bulk}} \text{d}\mathbf{1} \int \text{d}\mathbf{x}\ \text{e}^{-\beta U}}$$

$$K_{\text{eq}} = \frac{1}{[\text{barstar}]} \frac{N}{V_{\text{bulk}}} \frac{\int_{\text{site}} \text{d}\mathbf{1} \int \text{d}\mathbf{x}\ \text{e}^{-\beta U}}{\int_{\text{bulk}} \text{d}\mathbf{1}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \text{d}\mathbf{x}\ \text{e}^{-\beta U}}$$

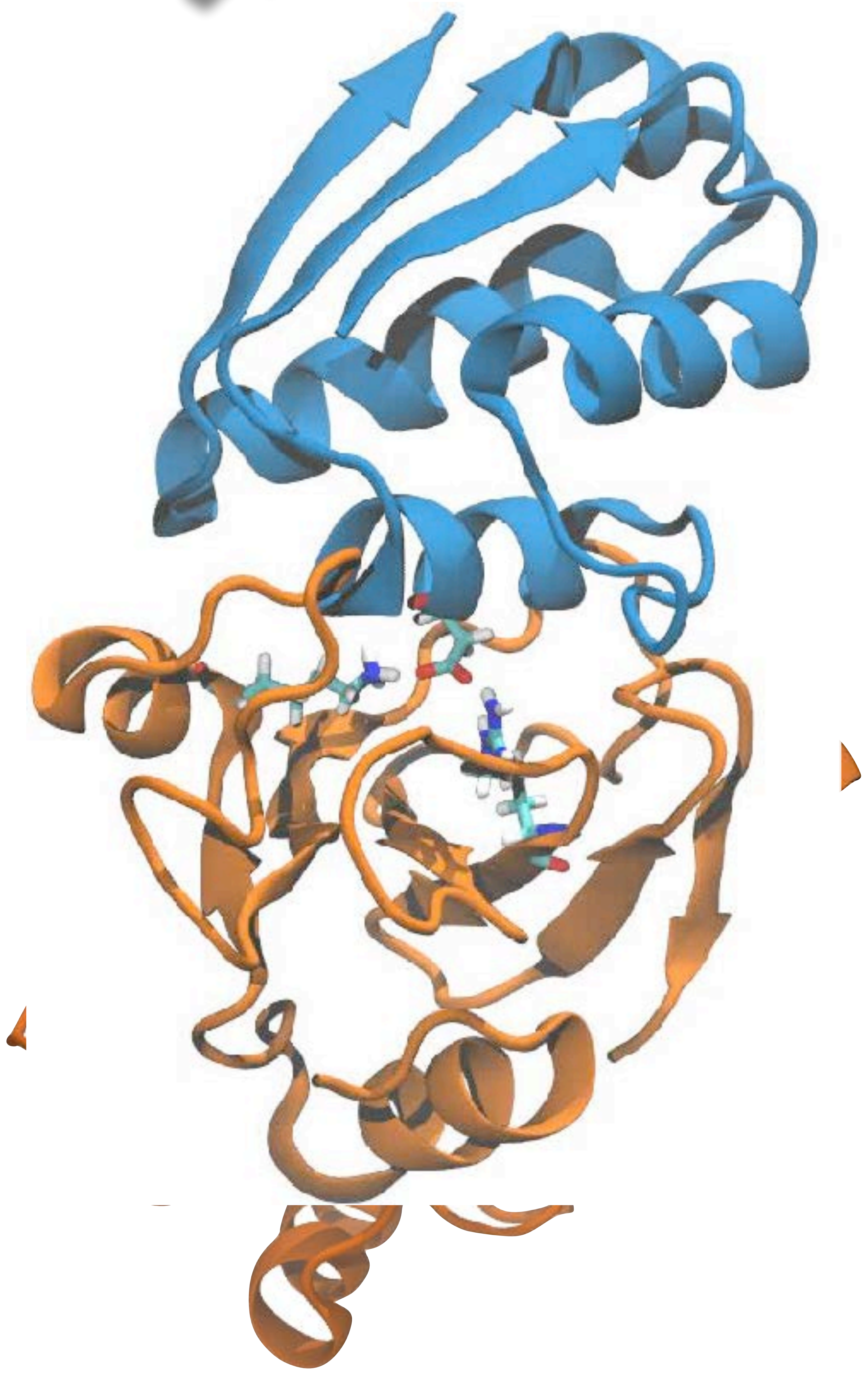
$$K_{\text{eq}} = \frac{\int_{\text{site}} \text{d}\mathbf{1} \int \text{d}\mathbf{x}\ \text{e}^{-\beta U}}{\int_{\text{bulk}} \text{d}\mathbf{1}\ \delta(\mathbf{x}_1 - \mathbf{x}_1^*) \int \text{d}\mathbf{x}\ \text{e}^{-\beta U}}$$

~~alchemical route~~

geometrical route

Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, 72, 1047-1069
Woo, H. J.; Roux, B. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6825-6830

BEYOND PROTEIN-LIGAND BINDING



$$K_{\rm eq} = \frac{\int_{\rm site} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta U}}{\int_{\rm site} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta(U+u_{\rm BS,c})}} \times \frac{\int_{\rm site} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta(U+u_{\rm BS,c})}}{\int_{\rm site} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta(U+u_{\rm BS,c}+u_{\rm BN,c})}} \times \frac{\int_{\rm site} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta(U+u_{\rm c})}}{\int_{\rm site} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta(U+u_{\rm c}+u_{\rm BS,res})}} \times \frac{\int_{\rm site} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta(U+u_{\rm c}+u_{\rm BS,res})}}{\int_{\rm site} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta(U+u_{\rm c}+u_{\rm BS,res}+u_{\rm BN,res})}}$$

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

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

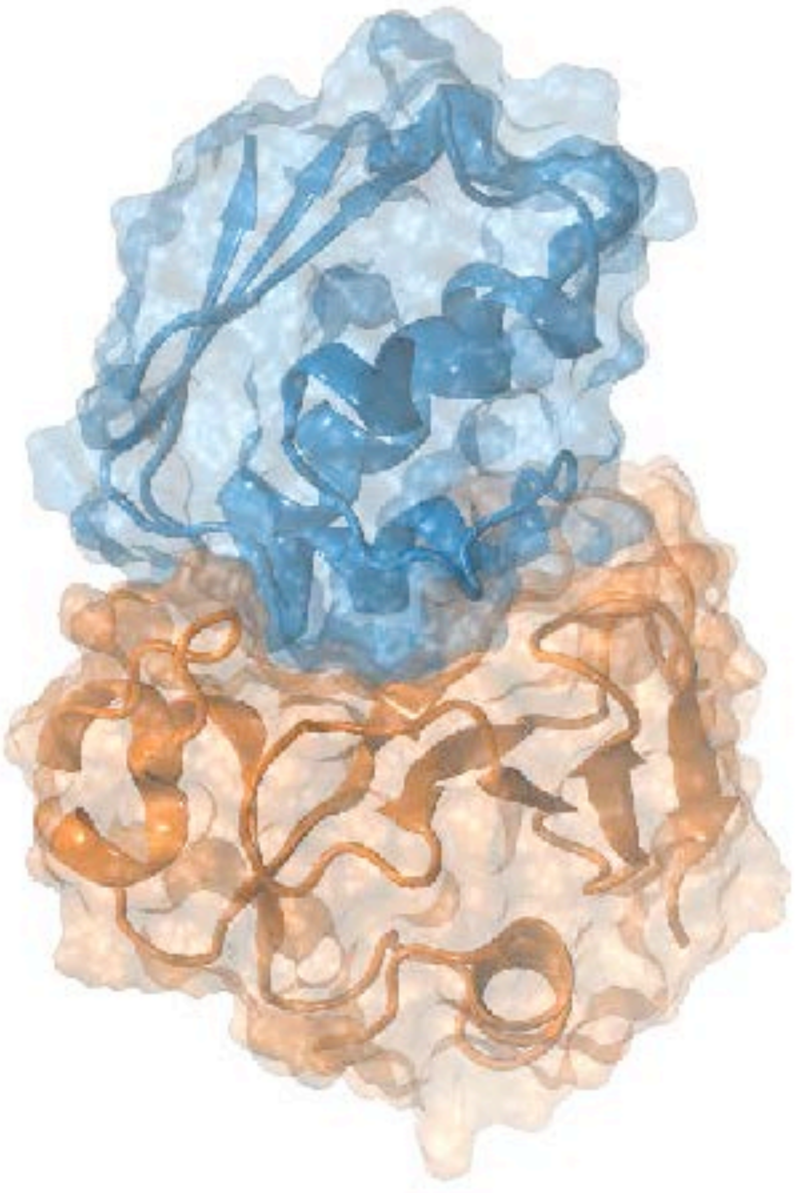
$$\times \frac{\int_{\rm site} \mathrm{d}\mathbf{1} \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta(U+u_{\rm c}+u_o+u_p)}}{\int_{\rm bulk} \mathrm{d}\mathbf{1}\, \delta(\mathbf{x}_1-\mathbf{x}_1^*) \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta(U+u_{\rm c}+u_o)}}$$

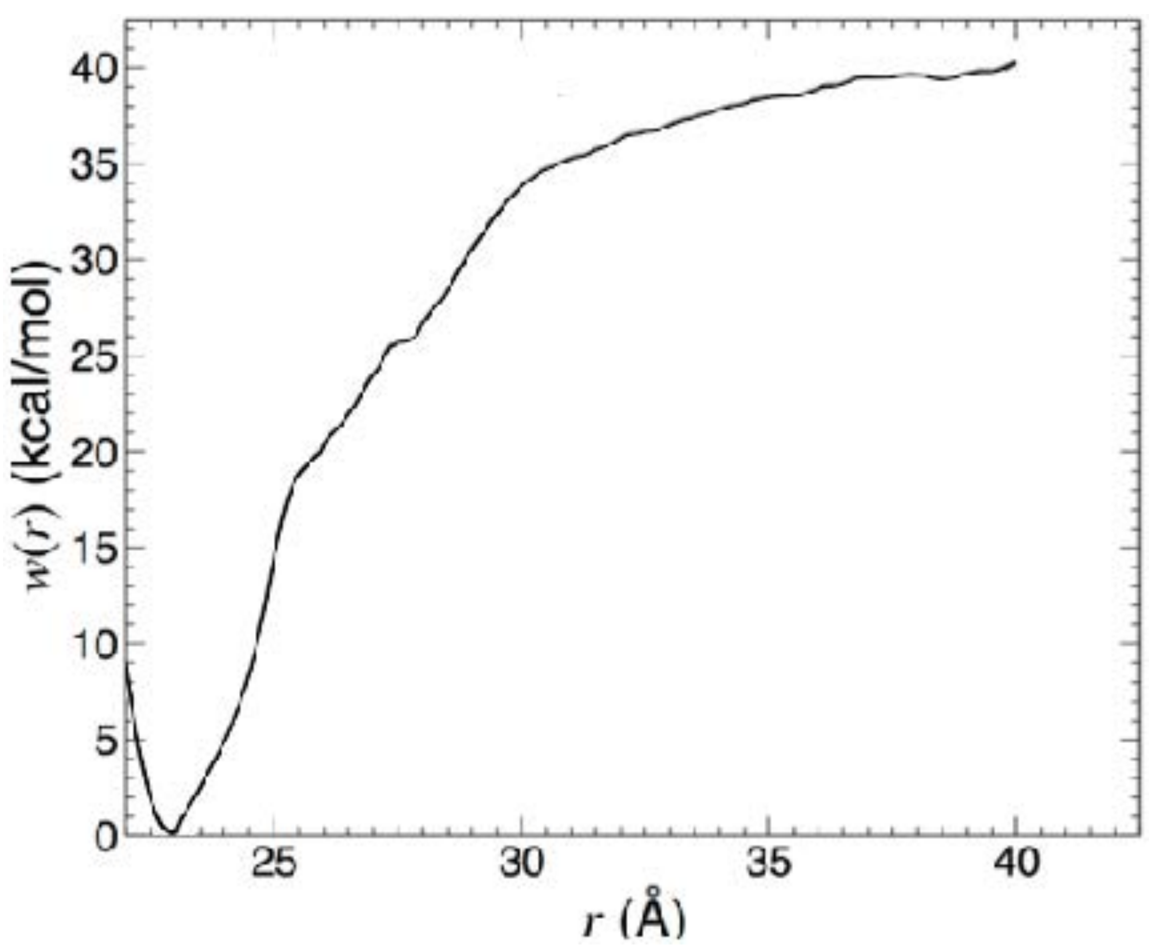
$$\times \frac{\int_{\rm bulk} \mathrm{d}\mathbf{1}\, \delta(\mathbf{x}_1-\mathbf{x}_1^*) \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta(U+u_{\rm c}+u_\Theta+u_\Phi+u_\Psi)}}{\int_{\rm bulk} \mathrm{d}\mathbf{1}\, \delta(\mathbf{x}_1-\mathbf{x}_1^*) \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta(U+u_{\rm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

$$\times \frac{\int_{\rm bulk} \mathrm{d}\mathbf{1}\, \delta(\mathbf{x}_1-\mathbf{x}_1^*) \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta(U+u_{\rm c}+u_{\rm BS,res}+u_{\rm BN,res})}}{\int_{\rm bulk} \mathrm{d}\mathbf{1}\, \delta(\mathbf{x}_1-\mathbf{x}_1^*) \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta(U+u_{\rm c}+u_{\rm BS,res})}} \times \frac{\int_{\rm bulk} \mathrm{d}\mathbf{1}\, \delta(\mathbf{x}_1-\mathbf{x}_1^*) \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta(U+u_{\rm c}+u_{\rm BS,res})}}{\int_{\rm bulk} \mathrm{d}\mathbf{1}\, \delta(\mathbf{x}_1-\mathbf{x}_1^*) \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta(U+u_{\rm c})}} \times \frac{\int_{\rm bulk} \mathrm{d}\mathbf{1}\, \delta(\mathbf{x}_1-\mathbf{x}_1^*) \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta(U+u_{\rm BS,c}+u_{\rm BN,c})}}{\int_{\rm bulk} \mathrm{d}\mathbf{1}\, \delta(\mathbf{x}_1-\mathbf{x}_1^*) \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta(U+u_{\rm BS,c})}} \times \frac{\int_{\rm bulk} \mathrm{d}\mathbf{1}\, \delta(\mathbf{x}_1-\mathbf{x}_1^*) \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta(U+u_{\rm BS,c})}}{\int_{\rm bulk} \mathrm{d}\mathbf{1}\, \delta(\mathbf{x}_1-\mathbf{x}_1^*) \int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta U}}$$

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

Gumbart, J. C.; Roux, B.; Chipot, C. *J. Chem. Theory Comput.* **2013,** *9*, 794-802





component	free energy (kcal/mol)	time (ns)
$\Delta G^{\text{site}}_{\text{BS},c}$	-2.0±0.3	6
$\Delta G^{\text{site}}_{\text{BN},c}$	-3.1±0.1	12
$\Delta G^{\text{site}}_{\text{BS,res}}$	-1.9±0.8	12
$\Delta G^{\text{site}}_{\text{BN,res}}$	-3.5±0.6	24
$\Delta G^{\text{site}}_{\Theta}$	-0.1±0.4	8
$\Delta G^{\text{site}}_{\Phi}$	-0.4±0.1	4
$\Delta G^{\text{site}}_{\Psi}$	-0.2±0.1	8
$\Delta G^{\text{site}}_{\theta}$	-0.1±0.3	4
$\Delta G^{\text{site}}_{\phi}$	-0.1±0.1	4
$-1/\beta \ln(S^* I^* c_0)$	-37.1±0.3	212
$\Delta G^{\text{bulk}}_{o}$	+6.6	
$\Delta G^{\text{bulk}}_{\text{BN,res}}$	+8.1±0.3	21
$\Delta G^{\text{bulk}}_{\text{BS,res}}$	+5.2±0.4	15
$\Delta G^{\text{bulk}}_{\text{BN},c}$	+4.2±0.5	18
$\Delta G^{\text{bulk}}_{\text{BS},c}$	+3.2±0.2	24
$\Delta G^{0}_{\text{bind}}$	**-21.0±1.4**	**372**

backbone: +2.5 kcal/mol

interface: +8 kcal/mol

orientation: +5.8 kcal/mol

ΔV_{eff} = 12.8 Å^3

$\Delta G°$ = -19.0 kcal/mol

Gumbart, J. C.; Roux, B.; Chipot, C. *J. Chem. Theory Comput.* **2013**, *9*, 3789–3798 standard binding free energies

geometric free-energy calculations

alchemical free-energy calculations

advanced tutorial

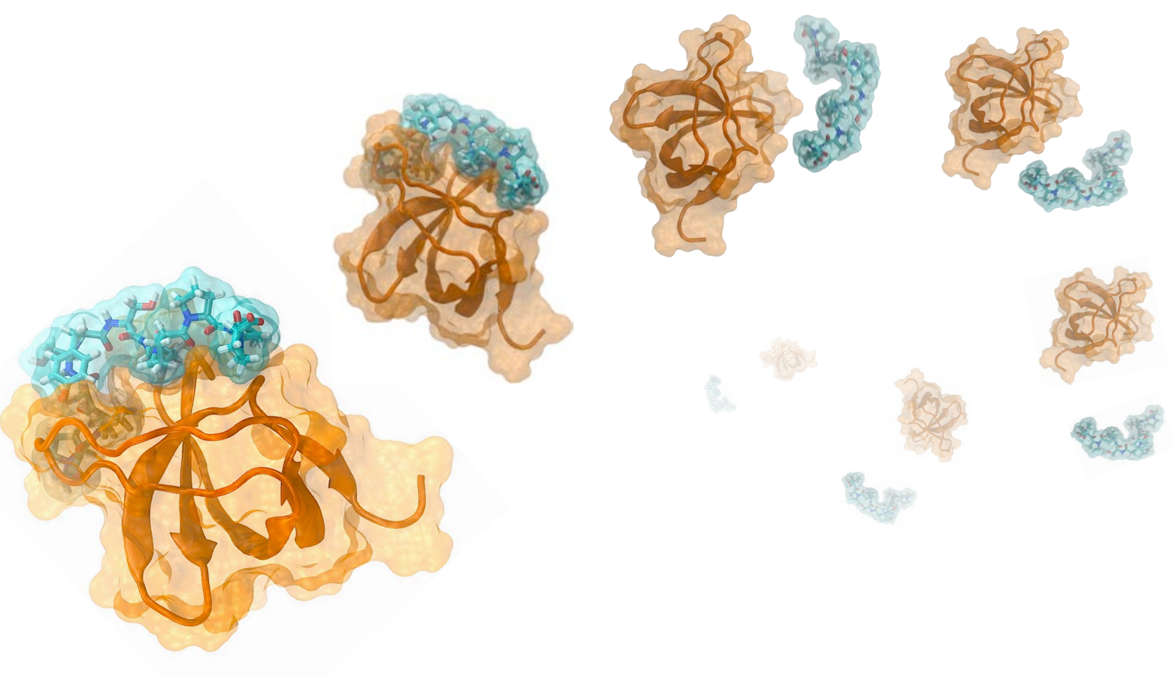
School of Physics
Georgia Institute of Technology

Department of Biochemistry and Molecular Biology
Gordon Center for Integrative Science
The University of Chicago

Centre National de la Recherche Scientifique
Laboratoire International Associé CNRS-UIUC
Université de Lorraine

University of Illinois at Urbana-Champaign
Beckman Institute for Advanced Science and Technology
Theoretical and Computational Biophysics Group

Protein:ligand standard binding free energies:
A tutorial for alchemical and geometrical transformations



James Gumbart
Benoît Roux
Christophe Chipot

July 4, 2013

Please visit www.ks.uiuc.edu/Training/Tutorials/ to get the latest version of this tutorial, to obtain more tutorials like this one, or to join the `tutorial-l@ks.uiuc.edu` mailing list for additional help.

Contributors: Gumbart, J. C.; Hénin, J.; Fajer, M.; Roux, B.; Chipot, C.