JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Standard Binding Free Energies from Computer Simulations: What Is the Best Strategy?

James C. Gumbart,[†] Benoît Roux,*[,†,§] and Christophe Chipot*[,‡,∥]

[†]Biosciences Division, Argonne National Laboratory, Argonne, Illinois 60439, United States

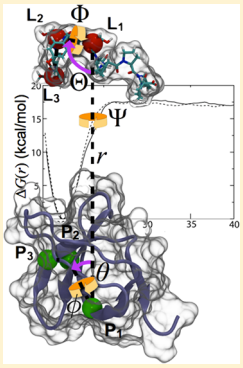
[‡]Beckman Institute, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States

[§]Department of Biochemistry and Molecular Biology and Gordon Center for Integrative Science, The University of Chicago, Chicago, Illinois 60637, United States

[∥]Équipe de Dynamique Des Assemblages Membranaires, UMR 7565, Université de Lorraine, BP 70239, 54506 Vandœuvre-lès-nancy Cedex, France

Supporting Information

ABSTRACT: Accurate prediction of standard binding free energies describing protein−ligand association remains a daunting computational endeavor. This challenge is rooted to a large extent in the considerable changes in conformational, translational, and rotational entropies underlying the binding process that atomistic simulations cannot easily sample. In spite of significant methodological advances, reflected in a continuously improving agreement with experiment, a characterization of alternate strategies aimed at measuring binding affinities, notably their respective advantages and drawbacks, is somewhat lacking. Here, two distinct avenues to determine the standard binding free energy are compared in the case of a short, proline-rich peptide associating to the Src homology domain 3 of tyrosine kinase Abl. These avenues, one relying upon alchemical transformations and the other on potentials of mean force (PMFs), invoke a series of geometrical restraints acting on collective variables designed to alleviate sampling limitations inherent to classical molecular dynamics simulations. The experimental binding free energy of $\Delta G_{\mathrm{bind}} = -7.99$ kcal/mol is well reproduced by the two strategies developed herein, with $\Delta G_{\mathrm{bind}} = -7.7$ for the alchemical route and $\Delta G_{\mathrm{bind}} = -7.8$ kcal/mol for the alternate PMF-based route. In detailing the underpinnings of these numerical strategies devised for the accurate determination of standard binding free energies, many practical elements of the proposed rigorous, conceptual framework are clarified, thereby paving way to tackle virtually any recognition and association phenomenon.



INTRODUCTION

Complete understanding of most biological processes of the cell machinery often requires, at its core, knowledge of the underlying free-energy change.[1] Such is the case, among others, of reversible protein−ligand association, which has been the object over the past decades of a plethora of theoretical endeavors aimed at predicting binding affinities within chemical accuracy.[2−21] In the realm of pharmaceutical research, determination of binding free energies from first principles has been the Holy Grail of computer-aided drug design. That being said, modelers are generally more interested in calculating relative quantities, in order to rank chemically similar lead candidates according to their affinity toward a given target, than in estimating absolute binding constants.

The choice of relative over absolute quantities can be easily rationalized by considering that, from a computational standpoint, determination of relative affinities between two analogous molecular compounds is appreciably simpler than that of standard binding free energies.[1,22,23] While calculations of relative quantities may appear to agree with experiments, such agreement could be due to a cancellation of errors; conversely, one cannot know if a discrepant result arises from the representation of the ligand, the receptor, their solubility, or something else entirely. The true, ultimate test of force fields and methodologies is the calculation of absolute binding affinities and absolute solvation free energies. Nonetheless, relative binding free energies have enjoyed widespread use, for which perturbation theory[24,25] has proven to be a powerful tool, exploiting the malleability of the potential energy function to transform between thermodynamically distinct, yet related states. On the other hand, accurate measurement of absolute binding constants constitutes an arduous theoretical problem and remains a grand computational challenge even for small, rigid ligands, let alone for large, flexible ones. One particular difficulty arises due to the incomplete knowledge of the association pathway—whereas the bound and unbound states of the ligand are well established, little is usually known about the path that connects them. As novel computer architectures are emerging, based either on graphics processing units or on task-dedicated processors, substantially longer atomistic simulations can be performed, able to capture the full sequence of events that form the association process, from the ligand lying away from the protein to its stable docking at the binding site of

Received: September 16, 2012
Published: November 7, 2012

dx.doi.org/10.1021/ct3008099 | J. Chem. Theory Comput. 2013, 9, 794−802

the latter.[19] However informative, these trajectories, by and large, only represent single, isolated events,[26] from which it would be admittedly difficult to infer meaningful thermodynamic quantities.

The lack of relevant topological information on the pathway followed by the ligand toward the target binding site of the protein rules out in principle geometric reaction-coordinate-based methods and further rationalizes why for many years thermodynamic perturbation has remained the unsurpassed and probably only available approach to tackle protein−ligand association numerically. Here, the ligand is decoupled reversibly from its environment,[1,22,27,28] in the unbound, free state and in the bound state, by scaling the corresponding nonbonded interactions. The so-called alchemical route, however, raises a number of conceptual issues, chief among which is the thermodynamic microreversibility of the transformation. In particular, as the substrate is progressively decoupled from the binding-site environment,[29] it becomes free to drift away from the latter and, thus, corresponds to an ill-defined target state of the perturbative calculation and a rather poor candidate for the reverse, coupling transformation. The resolution of this "wandering ligand" problem via a receptor−ligand restraint goes back to Hermans for gas phase,[30] Roux for bulk phase,[31] Gilson for the so-called double-decoupling method,[27] and Boresch for further application.[3] However, there is also a collateral consequence, namely an ambivalent definition of the standard state and, given the finite nature of simulation lengths, an unexpected correlation between the standard state and the size of the cell embracing the protein−ligand complex and its aqueous surroundings.[32] As will be subsequently discussed in detail, the latter weakness of the alchemical approach can be overcome via utilization of a set of appropriately chosen geometrical restraints,[2,7,13] the contributions to the affinity of which are evaluated in separate computations, some involving free-energy perturbation (FEP) simulations and some requiring only a numerical integration. Well-chosen restraining potentials can significantly improve the efficiency of sampling and, thus, greatly accelerate convergence of the simulations.[33] The review by Deng and Roux[13] provides a more complete overview of free-energy methods based on restraining potentials.

The level of intricacy of standard binding free-energy calculations is further increased by considering that the substrate may undergo conformational changes upon binding to the protein. From a computational perspective, this shortcoming is manifested in kinetic traps in conformational space and the unlikelihood of the ligand to isomerize, as it is either coupled or decoupled from its environment. Methodological enhancements of the basic perturbative approach, including boosting potentials of the slow degrees of freedom[16] and Hamiltonian hopping,[16,34] have flourished in an attempt to address the issue of incomplete sampling.

An alternative to the alchemical route consists in measuring the free-energy change along a surrogate, simplified representation of the reaction coordinate that follows the physical process of bimolecular association. Although little is known about the true reaction coordinate, except that it is evidently of high dimensionality—this is particularly true for ligands buried deep inside the protein—a handful of endeavors have been reported, whereby the protein−ligand problem is tackled by means of a naïve, one-dimensional potential of mean force (PMF).[5,15,35−39] While in the hypothetical limit of infinite sampling, a unidimensional projection should prove appropriate,[40] in general, it is expected to offer only a partial, incomplete picture of the reversible association process, failing to capture its collectivity. Of particular concern, a number of crucial degrees of freedom governing the conformation as well as the orientation of the substrate with respect to the protein are envisioned to be scarcely sampled at thermodynamic equilibrium.

To circumvent the problem of incomplete sampling, the most comprehensive approaches to determining binding free energies employ a set of explicitly defined and applied restraints,[2,5,7,13] which serve to greatly reduce the conformational entropy available to the system. Restraints on all, or at least some, of the six spatial degrees of freedom for the ligand can alleviate this sampling difficulty, while additional ones on, for example, the conformation of the ligand can further enhance convergence, provided that their contribution to the binding free energy is rigorously accounted. Although additional free-energy simulations are typically required to compute such contributions, the computational cost is immeasurably less than would be required to determine the binding free energy without them. Examples of studies exemplifying the use of comprehensive approaches include binding of peptides to SH2 domains,[5,13] protein−protein association,[41] and DNA−DNA interactions.[42]

In this paper, we explicitly detail the application of both common approaches, alchemical and PMF-based, to the same problem, computing the absolute binding free energy of a ten-residue peptide ligand to the SH3 domain of Abl kinase. The ligand, namely APSYSPPPPP, also denoted p41, was artificially designed to possess a high affinity for Abl-SH3 ($\Delta G^{\circ}_{bind} = -7.99$ kcal/mol).[43,44] Three different techniques are compared for the PMF-based approach, adaptive biasing forces (ABF), traditional umbrella sampling (US), and replica-exchange umbrella sampling (REMD-US). We carefully compare the different numerical strategies, identifying places where they are identical and where they differ. All simulations are done using NAMD 2.9, which supports a number of advanced free-energy techniques, including FEP,[24,25] REMD-US,[45−48] and ABF algorithms.[49,50] The alternate approaches are found to yield virtually identical results, in good agreement with the experimental binding free energy.[43] We conclude with an analysis of the benefits and pitfalls of each method.

METHODS

Molecular Assembly. The molecular assemblies were described by the all-atom CHARMM force field,[51,52] with the TIP3P water model.[53] In the case of the alchemical transformations, the bound state consisted of the APSYSPPPPP peptide, denoted p41,[43] bound to the SH3 domain of tyrosine kinase Abl (Protein Data Bank entry 1bbz), solvated by 10403 water molecules (32297 atoms in total), whereas the unbound state reduced to the peptide binder immersed in a bath of 3057 water molecules (9334 atoms). The dimensions of the corresponding periodic cells were, respectively, 50 × 53 × 121 Å^3 and 41 × 56 × 40 Å^3.

Molecular Dynamics Simulations. All the simulations reported herein were performed in the isobaric−isothermal ensemble with the NAMD program,[54] maintaining the temperature and the pressure at 300 K and 1 atm, respectively, using softly damped Langevin dynamics and the Langevin piston.[55] Nonbonded van der Waals interactions were smoothly switched to zero between 10 and 12 Å. The PME algorithm[56] was utilized to account for long-range electrostatic interactions. The equations of motion were integrated by means of a

dx.doi.org/10.1021/ct3008099 | J. Chem. Theory Comput. 2013, 9, 794−802

multiple-time step algorithm[57] with a time step of 2 and 4 fs for short- and long-range interactions, respectively. In alchemical transformations, singularities in the van der Waals potential,[58,59] prone to occur when interatomic distances tend toward zero upon particle creation or deletion, were circumvented by scaling and shifting the Lennard-Jones term of the force field.[59]

Alchemical Transformations. Reversible coupling of p41 to its environment, either the hydrated SH3 domain of Abl or the bulk aqueous medium, was performed using free-energy perturbation over a reaction pathway stratified in 65 windows of uneven width. Each stratum consisted of 300 000 data-collection steps (i.e., 0.6 ns) preceded by 100 000 equilibration steps (i.e., 0.2 ns), hence, a total simulation time equal to 52 ns.

For both the bound state and the unbound state, the alchemical transformation was carried out bidirectionally; that is, the peptide binder was created and annihilated in separate free-energy calculations.[27,28] The statistical data accrued in the coupling and decoupling simulations was combined by means of the Bennett acceptance ratio (BAR),[60] which provides a maximum-likelihood estimator of the free-energy change; more recently developed methods such as multistate BAR may be more efficient, however.[61]

In all alchemical transformations, p41 was restrained to its native confirmation by means of a root mean-square deviation (RMSD) collective variable. In addition, for the reversible coupling of the bound state, positional and orientational restraints were enforced, following the definitions of Figure 1. The contribution of the latter geometrical restraints was determined via thermodynamic-integration simulations, where-

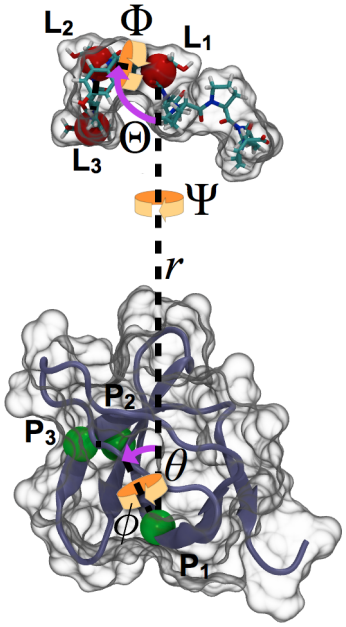


Figure 1. Definition of a frame of reference utilized to characterize the binding of peptide APSYSPPPPP (licorice-like representation colored by atom type) to the SH3 domain of the Abl kinase (secondary-structure representation). For both the protein and the ligand, three groups of atoms are arbitrarily selected, forming two triplets { P_1, P_2, P_3 } and { L_1, L_2, L_3 }, respectively. The position of APSYSPPPPP with respect to the SH3 domain is defined by means of the set of spherical coordinates $\{r, \theta, \phi\}$, where r is the P1−L1 distance, θ is the L1−P1−P2 angle, and ϕ is the P1−P1−P2−P3 angle. The relative orientation of the substrate is expressed using the three Euler angles {Θ (P1−L1−L2), Φ (P1−L1−L2−L3), Ψ (P2−P1−L1−L2)}.

in the force constant of the corresponding harmonic potentials was progressively scaled by means of a general-extent parameter. A stratification strategy of 12 individual points, at which the gradient of the potential energy with respect to the collective variable of interest was measured. A total of 400 000 data-collection steps were generated at each point (i.e., 0.8 ns) after a period of equilibration representing 100 000 molecular-dynamics steps (i.e., 0.2 ns), yielding a total simulation time of 12 ns. Scaling of the force constants was performed in bidirectional simulations, from which the maximum-likelihood estimator of the free energy was inferred.

Potential-of-Mean-Force Calculations. The free-energy change along the collective variables utilized to define the reversible binding of p41 to SH3 was determined using both the adaptive biasing force method and umbrella sampling. The first approach connects the derivative of the free energy measured at discrete value of the model reaction pathway to the average force exerted along the latter.[49,50] Incorporation of the average force in the equations of motion results in an apparent erasure of the ruggedness of the free-energy landscape, to the extent that progression of the system along the collective variable is primarily dictated by its diffusion properties.

For numerical efficiency, the reaction pathway was stratified into a series intermediate windows, ranging from two for the polar and Euler angles to nine for the more intricate separation of the ligand from the protein. Instantaneous values of the force were accrued in bins of width equal to 1°, 0.05 Å, and 0.1 Å, for the angular, RMSD, and separation PMFs, respectively. It ought to be underlined that while the ABF algorithm appreciably improves sampling along the collective variables, it does not necessarily do so in the slow manifolds orthogonal to it, thus, explaining the disparity in not only the number of utilized strata but also the amount of sampling per stratum required to obtain converged gradients, continuous across the entire reaction pathway.

A stringent, yet undocumented limitation of the ABF algorithm as implemented in NAMD[50,62] is the inability to incorporate to the free-energy calculation additional restraints, in particular when the latter are coupled to the model reaction coordinate along which the instantaneous force is measured. This inability arises because ABF in NAMD cannot discriminate between thermodynamic forces arising from the potential energy function and forces due to restraining harmonic potentials; under these premises, the time-dependent bias cancels the ruggedness of an apparent free-energy landscape that embraces the restraints. It can, however, be circumvented by coupling the collective variable of interest to a fictitious particle through a harmonic spring and, thus, forming an extended, generalized coordinate. As a result, the biasing force now acts on the fictitious particle rather than on the atoms involved in the model reaction coordinate and the free-energy change along the latter is recovered by means of a deconvolution of the harmonic-spring contribution from that of the force field. For the coupling, a stiff spring was used, with force constants of k = 6−15 kcal/mol·deg^2 for the angular PMFs, k = 250−1000 kcal/mol·Å^2 for the RMSD PMF, and k = 60−250 kcal/mol·Å^2 for the separation PMF. The fictitious particle was subject to Langevin dynamics with a damping coefficient of 1 ps^{-1}.

For replica-exchange umbrella sampling (REMD-US) simulations,[45,48] 37 windows were used, covering a distance between the centers of p41 and SH3 of 17−43 Å. The windows were spaced by 0.5 Å in the region from 20 to 30 Å and by 1 Å

dx.doi.org/10.1021/ct3008099 | J. Chem. Theory Comput. 2013, 9, 794−802

Table 1. Final Results for Each Contribution in Eqs 1−10 for Both the PMF-Based and Alchemical Approaches[a]

contribution	PMF(kcal/mol)	PMF (ns)	alch.(kcal/mol)	alch. (ns)
ΔG_c^{site} (1)	−3.6 ± 0.5	16	−4.2 ± 0.4	
ΔG_Θ^{site} (2)	−0.1 ± 0.2	4	−0.4 ± 0.0	
ΔG_Φ^{site} (2)	−0.1 ± 0.2	4	−0.8 ± 0.0	
ΔG_Ψ^{site}(2)	−0.4 ± 0.2	4	−0.5 ± 0.0	
ΔG_θ^{site} (3)	−0.0 ± 0.2	4	−0.3 ± 0.0	
ΔG_ϕ^{site} (3)	−0.6 ± 0.2	4	−0.3 ± 0.0	
ΔG_r^{site} (4)			−0.4 ± 0.0	24
$-(1/\beta)\ln(S^*I^*C^\circ)$ (5)	−14.5/14.4 ± 0.4	19/59		
$\Delta G_{decouple}^{site}$ (6)			+35.9 ± 0.7	104
$\Delta G_r^{bulk}+\Delta G_a^{bulk}$ (7)			+4.0	
ΔG_o^{bulk} (8)	+5.8		+6.6	
ΔG_{couple}^{bulk} (9)			−53.3 ± 0.7	104
ΔG_c^{bulk} (10)	+5.8 ± 0.5	60	+6.1 ± 0.4	48
ΔG_{bind}	−7.8/7.7 ± 0.9	115/155	−7.7 ± 1.0	280

[a]Note that distinct definitions of the reference coordinates were used for defining the angles in each of the PMF and alchemical routes, and as such, the free energies for each individual stage are not directly comparable. The two numbers given for eq 5 and the final result for the PMF approach embody the REMD-US and ABF separation PMFs, respectively. Calculation of the free energies associated to eqs 1−4 were carried out simultaneously in the alchemical route. Additionally, the reported time for each alchemical simulation encompasses both forward and backward runs to permit use of the BAR estimator, although, in principle, only one direction is needed.

everywhere else. Harmonic restraints were used for each window with a force constant k = 2.5 kcal/mol·Å^2. Exchanges between neighboring windows were attempted every 1000 time steps and were accepted or rejected according to a Metropolis energy criterion. Each window was run for 0.5 ns, giving a total time of 18.5 ns. After sorting of the resulting trajectories into the specified windows, the weighted histogram analysis method (WHAM) was used to reconstruct the PMF,[63,64] and the error in the PMF was determined using the method of Zhu and Hummer.[65]

RESULTS

In each of the alchemical and PMF-based strategies to computing the binding free energy, a number of restraints on collective variables[62] are used. A full, stepwise description of the two approaches is presented in eqs 1−10, with steps that are identical given on the same line and those distinct to each approach on a separate line. First, the conformational space available to the ligand itself is limited via restraints on its root mean-square deviation (RMSD, denoted u_c in eq 1) compared to a reference structure. The reference structure can be, for example, the crystallographic structure or an equilibrated structure of the complex, the former being used in the alchemical approach and the latter in the PMF approach. Additionally, restraints defining the orientation of the ligand relative to the protein require three Euler angles, denoted Θ, Φ, and Ψ (their collective restraint given by u_o in eqs 2 and 11), while those for the position of the ligand require an additional two, denoted ϕ and θ (see u_a in eqs 3 and 12 and Figure 1). How one defines these angles is not uniquely prescribed, as long as they effectively restrict all rotational degrees of freedom between the protein and ligand. To make our point more cogent, two different sets of definitions were used for the two approaches. After separation via forced extraction (eq 5) or alchemical decoupling (eqs 6 and 9), both representing the most expensive steps in their respective approaches, the restraints are successively removed. Boxed expressions denote those contributions in the free state that do not require actual simulations but only numerical integrals.

$$K_{eq}^{PMF} = \frac{\int_{site} d\mathbf{1}\int d\mathbf{X} e^{-\beta U}}{\int_{site} d\mathbf{1}\int d\mathbf{X} e^{-\beta[U+u_c]}} \quad (1)$$

$$\times \frac{\int_{site} d\mathbf{1}\int d\mathbf{X} e^{-\beta[U+u_c]}}{\int_{site} d\mathbf{1}\int d\mathbf{X} e^{-\beta[U+u_c+u_o]}} \quad (2)$$

$$\times \frac{\int_{site} d\mathbf{1}\int d\mathbf{X} e^{-\beta[U+u_c+u_o]}}{\int_{site} d\mathbf{1}\int d\mathbf{X} e^{-\beta[U+u_c+u_o+u_a]}} \quad (3)$$

$$\times \frac{\int_{site} d\mathbf{1}\int d\mathbf{X} e^{-\beta[U+u_c+u_o+u_a]}}{\int_{bulk} d\mathbf{1}\delta(\mathbf{r}_1-\mathbf{r}_1^*)\int d\mathbf{X} e^{-\beta[U+u_c+u_o]}} \quad (5)$$

$$\times \boxed{\frac{\int_{bulk} d\mathbf{1}\delta(\mathbf{r}_1-\mathbf{r}_1^*)\int d\mathbf{X} e^{-\beta[U+u_c+u_o]}}{\int_{bulk} d\mathbf{1}\delta(\mathbf{r}_1-\mathbf{r}_1^*)\int d\mathbf{X} e^{-\beta[U+u_c]}}} \quad (8)$$

$$\times \frac{\int_{bulk} d\mathbf{1}\delta(\mathbf{r}_1-\mathbf{r}_1^*)\int d\mathbf{X} e^{-\beta[U+u_c]}}{\int_{bulk} d\mathbf{1}\delta(\mathbf{r}_1-\mathbf{r}_1^*)\int d\mathbf{X} e^{-\beta U}} \quad (10)$$

$$K_{eq}^{FEP} = \frac{\int_{site} d\mathbf{1}\int d\mathbf{X} e^{-\beta U_1}}{\int_{site} d\mathbf{1}\int d\mathbf{X} e^{-\beta[U_1+u_c]}} \quad (1)$$

$$\times \frac{\int_{site} d\mathbf{1}\int d\mathbf{X} e^{-\beta[U_1+u_c]}}{\int_{site} d\mathbf{1}\int d\mathbf{X} e^{-\beta[U_1+u_c+u_o]}} \quad (2)$$

$$\times \frac{\int_{site} d\mathbf{1}\int d\mathbf{X} e^{-\beta[U_1+u_c+u_o]}}{\int_{site} d\mathbf{1}\int d\mathbf{X} e^{-\beta[U_1+u_c+u_o+u_a]}} \quad (3)$$

$$\times \frac{\int_{site} d\mathbf{1}\int d\mathbf{X} e^{-\beta[U_1+u_c+u_o+u_a]}}{\int_{site} d\mathbf{1}\int d\mathbf{X} e^{-\beta[U_1+u_c+u_o+u_a+u_r]}} \quad (4)$$

$$\times \frac{\int_{site} d\mathbf{1}\int d\mathbf{X} e^{-\beta[U_1+u_c+u_o+u_a+u_r]}}{\int_{site} d\mathbf{1}\int d\mathbf{X} e^{-\beta[U_0+u_c+u_o+u_a+u_r]}} \quad (6)$$

$$\times \boxed{\frac{\int_{bulk} d\mathbf{1}\int d\mathbf{X} e^{-\beta[U_0+u_c+u_o+u_a+u_r]}}{\int_{bulk} d\mathbf{1}\delta(\mathbf{r}_1-\mathbf{r}_1^*)\int d\mathbf{X} e^{-\beta[U_0+u_c+u_o]}}} \quad (7)$$

$$\times \boxed{\frac{\int_{bulk} d\mathbf{1}\delta(\mathbf{r}_1-\mathbf{r}_1^*)\int d\mathbf{X} e^{-\beta[U_0+u_c+u_o]}}{\int_{bulk} d\mathbf{1}\delta(\mathbf{r}_1-\mathbf{r}_1^*)\int d\mathbf{X} e^{-\beta[U_0+u_c]}}} \quad (8)$$

$$\times \frac{\int_{bulk} d\mathbf{1}\delta(\mathbf{r}_1-\mathbf{r}_1^*)\int d\mathbf{X} e^{-\beta[U_0+u_c]}}{\int_{bulk} d\mathbf{1}\delta(\mathbf{r}_1-\mathbf{r}_1^*)\int d\mathbf{X} e^{-\beta[U_1+u_c]}} \quad (9)$$

$$\times \frac{\int_{bulk} d\mathbf{1}\delta(\mathbf{r}_1-\mathbf{r}_1^*)\int d\mathbf{X} e^{-\beta[U_1+u_c]}}{\int_{bulk} d\mathbf{1}\delta(\mathbf{r}_1-\mathbf{r}_1^*)\int d\mathbf{X} e^{-\beta U_1}} \quad (10)$$

with

$$u_o = u_\Theta + u_\Phi + u_\Psi \quad (11)$$

$$u_a = u_\theta + u_\phi \quad (12)$$

$$u_r = \frac{1}{2}k(r - r_0)^2 \quad (13)$$

dx.doi.org/10.1021/ct3008099 | J. Chem. Theory Comput. 2013, 9, 794−802

Common Geometrical Restraints. The resulting contribution of each of the restraints common to both approaches (i.e., eqs 1−3, 8, and 10) is given in Table 1. These restraints include those on the RMSD of the ligand in the bound (subscript: site) and free (subscript: bulk) states as well as those on the orientation and position of the ligand. For both the bound and the unbound states, the free-energy change due to the restraint on the RMSD is calculated from simulations. The orientational and positional restraints are calculated from simulation when the ligand is bound. To account for some deviation in the position of the ligand relative to Abl-SH3, which may occur naturally during their separation, a very soft (k = 0.03 kcal/mol·deg^2) restraint on each angle is used. Thus, the contribution of these restraints is quite small—typically just a small fraction of a kcal/mol. Their contribution in the bulk state of the ligand can be calculated through numerical integration of an analytical expression, due to the isotropy of space.[5,66] The sum of the resulting free-energy changes for Θ, Φ, and Ψ is found to be an order of magnitude larger in the bulk than in the binding site (see Table 1). While the free-energy cost for removing the positional restraints on θ and ϕ in the bulk is explicitly calculated in the alchemical formulation (eq 7), in the PMF approach, their removal is accounted for indirectly in the integrals in eq 5. The complete calculation for each component of the binding free energy is given in the Supporting Information.

PMF-Based Approach. In the PMF-based route, the contribution of each restraint to the binding free energy is determined by relating each ratio of integrals in eqs 1−10 to an expectation value in the given ensemble. This relationship is achieved through use of the PMF of the relevant coordinate, $W(\xi)$. For example, the contribution due to the conformational (RMSD) restraint in the bound state (i.e., eq 1) is given by

$$\frac{\int_{\text{site}} d\mathbf{1} \int d\mathbf{X}\, e^{-\beta U}}{\int_{\text{site}} d\mathbf{1} \int d\mathbf{X}\, e^{-\beta[U+u_c]}} = \frac{\int_{\text{site}} d\xi e^{-\beta W(\xi)}}{\int_{\text{site}} d\xi\, e^{-\beta W(\xi)}\, e^{-\beta u_c(\xi)}} = \langle e^{\beta u_c}\rangle_{(\text{site},U)} = e^{\beta G_c^{\text{site}}} \quad (14)$$

where U is the net potential energy bereft of the restraint. Most of the PMFs, in particular, those for the various angular terms, present as simple harmonic-like functions with a single minimum around the bound state (see Figure S1 in the SI). Similarly, the RMSD PMF in the binding site exhibits two close minima, peaking sharply around them. On the contrary, while the RMSD in the bulk also has two apparent local minima, it is much broader, clearly illustrating that binding restricts the conformational freedom of the ligand (see Figure 2A). All PMFs for the angular and conformational terms were computed using ABF.

Alchemical Approach. The alchemical route consists of a reversible double annihilation,[27] wherein the ligand is decoupled and recoupled in separate simulations to the protein, both in the bound and in the unbound state. The statistical data accrued from these bidirectional FEP calculations is subsequently employed to measure the maximum-likelihood BAR estimator[60] of the binding free energy (see the Methods section). To circumvent the "wandering-ligand" problem common to FEP alchemical decoupling schemes and, hence, to satisfy thermodynamic microreversibility, a translational restraint tethering the substrate to its bound position on Abl-SH3 was introduced. This restraint, denoted u_r in eqs 4 and 13, is added to the geometric restraints also used in the PMF approach (see Figure 1). The contribution of these restraints to the net binding free energy was evaluated alchemically by nullifying reversibly the force constant associated to the individual harmonic potentials. Plots of the free energy as a function of the reaction coordinate λ are provided in Figure S2 in the SI.

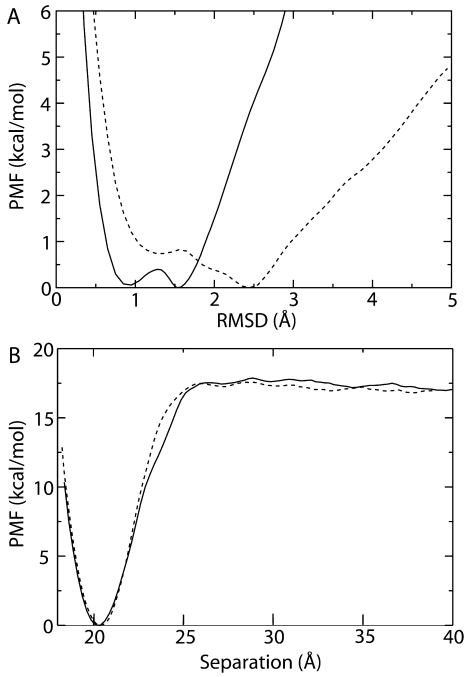


Figure 2. Potentials of mean force for (A) RMSD restraints and (B) extraction of the ligand from the binding site. (A) PMF for RMSD of p41 in the bound state (solid line) and in the unbound state (dashed line). (B) Separation PMF, $W(r)$, calculated from REMD-US (solid line) and ABF (dashed line).

Calculation of S^* and I^* in the PMF-Based Approach. The only term specific to the PMF approach represents the separation of the ligand from the binding site and into the bulk along the axis defined by the positional restraints on θ and ϕ (eq 5). This ratio of integrals can be recast in the form of two terms, denoted S^* and I^*:

$$S^* = r^{*2} \int_0^{\pi} d\theta \sin\theta \int_0^{2\pi} d\phi\, e^{-\beta u_a(\theta,\phi)} \quad (15)$$

and

$$I^* = \int_{\text{site}} dr\, e^{-\beta[W(r)-W(r^*)]} \quad (16)$$

with $W(r)$ being the PMF delineating the separation along an axis in the presence of the specified restraints and r^* a reference point far from the protein (see Woo and Roux[5] for a complete derivation). The term S^*, which has units of the square of a distance, effectively describes the small fraction of surface area on the sphere of radius r^* centered on the binding site that is accessible to the restrained ligand;[5,7,13] again, noting the isotropy for θ and ϕ in the bulk, S^* can be calculated numerically, similar to eq 7 in the alchemical approach.

The term I^*, which contains the separation PMF, $W(r)$, is by far the largest contribution to the final binding free energy. While there is no strict equivalent, this term replaces the alchemical decoupling and recoupling steps in the FEP approach. Even when using multiple restraining potentials to reduce the conformational, orientational, and spatial freedom of

dx.doi.org/10.1021/ct3008099 | J. Chem. Theory Comput. 2013, 9, 794−802

the ligand, sampling of, e.g., side-chain rotamers or specific interactions between pairs of residues can severely hamper convergence of the separation PMF. One solution is to introduce a second variable covering the slowly evolving degree of freedom and calculate a two-dimensional PMF, from which the unidimensional separation PMF can be recovered;[5] however, this solution can become computationally expensive and is particularly challenging if the degrees of freedom to target are not obvious. Another innovative approach is to carry out a random walk in orthogonal space, which does not require identification at the outset of the intransigent degrees of freedom.[67] Alternatively, enhanced sampling techniques can be used, provided they either do not alter the resulting PMF or that the unbiased PMF can be recovered. One example is replica-exchange MD umbrella sampling (REMD-US), in which unique conformations (more precisely, all the coordinates) can be exchanged between neighboring windows according to a Metropolis energy criterion.[48] Thus, conformations reached easily in some windows but comparatively inaccessible in others can quickly be communicated across all windows. To explore its potential for the problem of p41 binding to Abl-SH3, REMD-US was employed (see the Methods section). The PMF, shown in Figure 2B displays a deep well in the binding site at r = 20.3 Å of −17.5 kcal/mol. While this energy is much greater than the final binding free energy, it will be counter-balanced by contributions from the applied restraints. Convergence of the PMF was fast, as indicated by its progression over time, as shown in Figure S3A in the SI. For comparison, we determined the PMF with standard umbrella sampling (no exchange between windows) as well. The favorable comparison between the two PMFs, given in Figure S3B in the SI, suggests that advanced methods such as REMD-US are not required to determine the PMF for separation of p41 from Abl-SH3, although this could not have been known *a priori*.

The separation PMF $W(r)$ was also determined using a third method, namely ABF with an extended coordinate, as described (see the Methods section). Although it ultimately yielded a similar result to that from REMD-US (see Figure 2B), convergence was much more difficult to achieve. Besides requiring a longer production simulation time of nearly 60 ns, some windows needed to be rerun multiple times to ensure continuity of the gradients across window boundaries. It was found that individual ABF runs are very sensitive to the initial conditions, particularly for systems in which the reaction coordinate is strongly coupled to other slow degrees of freedom, as is the case for p41 binding to SH3. Therefore, if the gradient for a given window was very incongruous compared to the expected final result, convergence was exceedingly slow. Restarting the window with different initial conditions increased the probability of converging within a limited simulation time. The simulation time of 59 ns should, therefore, be considered only as indicative, since it does not include the duration of failed attempts.

Because the choice of r^* (i.e., the point in space far from the binding site in eqs 15 and 16) is arbitrary, the final binding free energy should be independent of it. Thus, because S^* depends on r^{*2}, I^* should vary as $1/r^{*2}$, making their product free of r^*. The dependence on r^* in I^* comes from the factor $\exp[\beta W(r^*)]$, which can be directly extracted from the integral in eq 16. For the product S^*I^* to be independent of r^*, the relationship $W(r^*) = -kT \ln(\alpha r^{*2})$, where α is a constant, must hold. This form of $W(r)$ at large values of r (because r^* should be far enough that the protein and the ligand no longer interact) is equivalent to the known Jacobian correction to the PMF[68,69] and fits the PMFs in Figure 2B reasonably well for values r > 28 Å (correlation coefficient of 0.91 for ABF and 0.87 for REMD-US). Because the fit is not perfect, however, we computed the product S^*I^* for multiple values of r^*, finding it to vary by 0.3 kcal/mol at most (see the SI). We use the central value at r^* = 35 Å for determination of the binding free energy. The value of I^* is also dependent on the range of integration, which should only encompass the binding site, an ill-defined region. However, because of the sharp rise in $W(r)$ away from the minimum, the exponential term in eq 16 becomes negligible very quickly. The difference in I^* between integration carried out over the full range of $W(r)$ and that over ±2 Å about the minimum is only $\sim 6.7 \times 10^{-7}$ Å.

Final Calculation of $\Delta G^\circ_{\text{bind}}$. The equilibrium binding constant, K_{eq} is given in the two approaches by

$$K_{\text{eq}}^{\text{PMF}} = S^*I^* \, e^{-\beta[G_c^{\text{bulk}} + G_o^{\text{bulk}} - G_a^{\text{site}} - G_o^{\text{site}} - G_c^{\text{site}}]} \tag{17}$$

$$K_{\text{eq}}^{\text{FEP}} = e^{-\beta[G_c^{\text{bulk}} - G_{\text{couple}}^{\text{bulk}} + G_o^{\text{bulk}} + G_r^{\text{bulk}} + G_a^{\text{bulk}} + G_{\text{decouple}}^{\text{site}} - G_a^{\text{site}} - G_r^{\text{site}} - G_o^{\text{site}} - G_c^{\text{site}}]} \tag{18}$$

and the resulting binding free energy is

$$\Delta G^\circ_{\text{bind}} = -kT \log(K_{\text{eq}} C^\circ) \tag{19}$$

where C° = 1/1661 Å^3 is the standard concentration.[7] With all contributions to the binding free energy determined and listed in Table 1, ΔG_{bind} can be computed for each method. For the PMF approach, $\Delta G_{\text{bind}}^{\text{PMF}}$ = −7.8 kcal/mol if $W(r)$ is calculated with REMD-US (identical for traditional US) and −7.7 kcal/mol if $W(r)$ is calculated with ABF. Similarly, from the alchemical approach, $\Delta G_{\text{bind}}^{\text{alch.}}$ = −7.7 kcal/mol. Despite the distinct routes to the final answer, both results are very close to one another and, more importantly, to the experimental binding free energy, $\Delta G^\circ_{\text{bind}}$ = −7.99 kcal/mol.[43]

The striking agreement between computation and experiment may partly reflect some fortuitous cancellation of errors in the protracted approaches, although the fact that they independently arrived at the same conclusion provides some cause for confidence. Nonetheless, statistical uncertainties are present in free-energy calculations, as in any simulation method.

As discussed in greater detail in the SI, the precision and accuracy of free-energy estimates are of different natures and sources,[70] thereby making their measure somewhat cumbersome and difficult to interpret. In the case of the alchemical route, differences in the forward and backward values for the various transformations, which represent an upper bound on the statistical error estimate, give a net standard deviation of ±1.0 kcal/mol. For the PMF-based route, error estimates for ABF calculations, that is, those for all angular and conformational restraints and the ABF-based separation PMF, are derived from discontinuities in the gradients at the boundaries between windows. As for the separation PMF determined with REMD-US, the method of Zhu and Hummer[65] was utilized to compute the variance of the PMF (see Figure S3C in the SI). This variance was then tracked through the integration and subsequent calculations that, when combined with the errors from the restraints, gives an error associated to the net binding free energy of ±0.9 kcal/mol (see the SI for additional details). While the errors in both the alchemical and PMF-based approaches are apparently large, it should be noted that they

likely correspond to maximal error estimates, prone to accumulation over the multiple steps in each approach.

DISCUSSION

Here, we have presented two distinct, formal routes to determining the standard binding free energy of a ligand, the ten-residue oligopeptide p41, associated to the SH3 domain of Abl kinase. The first route is based on PMF calculations and requires forced extraction of the ligand from the binding pocket. The second route utilizes alchemical FEP calculations to decouple the ligand from the system while located in the binding pocket. To achieve convergence on accessible time scales, both routes take advantage of a number of geometrical restraints, which limit the freedom of the ligand to move or contort during the critical separation (PMF) or reversible decoupling (alchemical) simulations. These restraints were chosen specifically to reduce the conformational, orientational, and positional entropies of the substrate, although any set of restraints are, in principle, allowable as long as their contribution to the binding free energy is removed in the final calculation. Indeed, for each route, a different definition of the restrained angles was used, as well as different reference states for the RMSD calculations. Nonetheless, both routes led to virtually identical results, namely a standard binding free energy in very close agreement with the experimental value. The necessity of these extended routes for determining absolute free energies is underscored by the notable difference from previous molecular mechanics/Poisson−Boltzmann solvent area calculations ($\Delta G_{bind} = -2.6$ kcal/mol).[71] However, it should be emphasized that only a relative ranking of different binders to Abl-SH3 was sought in that study, and thus the significant contributions to the free energy that result in the unbound state were purposefully neglected.[71]

At this stage of the study, it is legitimate to examine the merits and the drawbacks of the two alternative routes for the determination of standard binding affinities—in other words, should one approach be preferred over the other? It can be asserted from the onset that both the alchemical and the PMF routes share noteworthy common advantages, namely, (i) they rely on a sound, rigorous, and formal theoretical framework, (ii) they are equally reliable, and (iii) they correspond to a roughly comparable computational cost, in the limit of the alchemical transformations being performed unidirectionally − albeit FEP can be up to 1/3 slower due to additional energy evaluations required. As has been discussed repeatedly,[70] however, combining forward and backward simulations, employing for instance the BAR estimator, is recommended to improve the reliability and the efficiency of the free-energy calculations. The computational burden incurred in performing independent calculations in both directions can be somewhat reduced using numerical schemes such as the double-wide sampling algorithm[60,72] wherein the free-energy changes between λ and $\lambda - \Delta\lambda$, and between λ and $\lambda + \Delta\lambda$ are determined concomitantly. An interesting benefit of the alchemical route lies in the possibility to perform the different steps in parallel, while, in principle, the PMF route supposes that the free-energy profiles are computed in a sequential fashion, albeit this does not necessarily need to be so.

Quite unfortunately, the two approaches are also affected by a number of common drawbacks, which, while not insurmountable, ought to be mentioned here. Specifically, (i) their complexity necessitates careful bookkeeping, and (ii) they are plagued by convergence issues, notably for the RMSD terms of the PMF route on account of conformational degeneracy and for the coupling and decoupling transformations of the alchemical route. The propensity of the bound ligand, particularly a flexible peptide, to become trapped in metastable conformations greatly impedes sampling of other states, although advanced methods such as REMD-US help to alleviate such bottlenecks. Both traditional US and REMD-US converged relatively quickly in determination of the PMF for separation of the ligand and protein, whereas ABF proved more difficult, being especially sensitive to initial conditions in this case.

Arguably the most pronounced weakness of the alchemical route is its inherent limitation to small ligands, because convergence of the FEP calculation is stringently dictated by the magnitude of the perturbation. Additionally, due to their large solvation free energies, highly charged ligands, for example, phosphotyrosine peptides, are problematic. In sharp contrast with the alchemical formulation, the PMF route cannot handle deeply buried substrates, particularly in crooked internal binding pockets for which a simple, one-dimensional reaction coordinate would undoubtedly fail to describe the protein−ligand separation with uniform sampling. However, the PMF-based method could possibly be extended to allow a curvilinear coordinate system to describe the separation process using, for example, the string method.[73] The current shortcoming is evidently irrelevant for coupling and decoupling transformations, which in large measure ignore the topology of the binding site of the protein. On the other hand, the limitation of the PMF route to interfacial binding also constitutes its strength—this route is envisioned to be the only viable option to describe protein−protein association,[41] which, by and large, is interfacial in nature. Additionally, the PMF-based approach can be used directly with implicit solvent methods, such as generalized Born,[74] whereas FEP with implicit solvent would require further considerations.

Free-energy calculations have admittedly reached a level of maturity such that one can now discriminate between methodological and sampling issues, choosing the method most appropriate to the problem at hand. However, while the above discussion has revolved around purely mechanical aspects of the calculations, a potentially significant component of the overall error on the free-energy estimate still remains intimately related to the choice of force-field parameters. Here, we have made the implicit, yet not completely justified, assumption that the CHARMM force field is appropriate for proteins and small peptides (e.g., binding of p41 to SH3). Even for proteins, though, this force field is under constant revision, most recently with the 2012 CHARMM36 iteration.[75] With a nonpeptide ligand, it would be evidently more complicated to discriminate between methodological/sampling and force-field issues, and thus, care must be taken in the choice and validation of the force-field parameters. Nonetheless, the robust approaches for determining standard binding free energies presented here minimize the former issues, allowing one to be relatively assured of reaching a result within the limits of the chemical accuracy amenable to the potential energy function utilized.

ASSOCIATED CONTENT

Supporting Information

Three figures, discussion of sources of error, and expanded calculation of the binding free energy. This information is available free of charge via the Internet at http://pubs.acs.org/.

dx.doi.org/10.1021/ct3008099 | J. Chem. Theory Comput. 2013, 9, 794−802

AUTHOR INFORMATION

Corresponding Author

*E-mail: roux@uchicago.edu; chipot@ks.uiuc.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors are grateful to the France and Chicago Collaborating in the Sciences (FACCTS) Center for their support. The research is funded by grant MCB-0920261 from the National Science Foundation (NSF) and by grant R01-CA093577 from the National Institute of Health (B.R.). Simulations were carried out using the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by NSF grant number OCI-1053575. The Centre Informatique National de l'Enseignement Supérieur, Montpellier, is also gratefully acknowledged for generous provision of computer time. J.C.G. is supported by a Director's Postdoctoral Fellowship from Argonne National Laboratory.

REFERENCES

(1) *Free Energy Calculations. Theory and Applications in Chemistry and Biology*; Chipot, C.; Pohorille, A., Eds.; Springer Verlag: Weinheim, Germany, 2007.

(2) Dixit, S. B.; Chipot, C. Can absolute free energies of association be estimated from molecular mechanical simulations? The biotin–streptavidin system revisited. *J. Phys. Chem. A* **2001**, *105*, 9795–9799.

(3) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute binding free energies: A quantitative approach to their calculation. *J. Phys. Chem. B* **2003**, *107*, 9535–9551.

(4) Huang, D.; Caflisch, A. Efficient evaluation of binding free energy using continuum electrostatics solvation. *J. Med. Chem.* **2004**, *47*, 5791–5797.

(5) Woo, H. J.; Roux, B. Calculation of absolute protein–ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6825–6830.

(6) Rodinger, T.; Howell, P. L.; Pomès, R. Absolute free energy calculations by thermodynamic integration in four spatial dimensions. *J. Chem. Phys.* **2005**, *123*, 34104.

(7) Deng, Y.; Roux, B. Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput.* **2006**, *2*, 1255–1273.

(8) Ytreberg, F. M.; Zuckerman, D. M. Simple estimation of absolute free energies for biomolecules. *J. Chem. Phys.* **2006**, *124*, 104105.

(9) Wang, J.; Deng, Y.; Roux, B. Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys. J.* **2006**, *91*, 2798–2814.

(10) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. Predicting absolute ligand binding free energies to a simple model site. *J. Mol. Biol.* **2007**, *371*, 1118–1134.

(11) J., D.; Golubkov, P. A.; Darden, T. A.; Ren, P. Calculation of protein–ligand binding free energy by using a polarizable potential. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6290–6295.

(12) Rodinger, T.; Howell, P. L.; Pomès, R. Calculation of absolute protein–ligand binding free energy using distributed replica sampling. *J. Chem. Phys.* **2008**, *129*, 155102.

(13) Deng, Y.; Roux, B. Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem. B* **2009**, *113*, 2234–2246.

(14) Boyce, S. E.; Mobley, D. L.; Rocklin, G. J.; Graves, A. P.; Dill, K. A.; Shoichet, B. K. Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site. *J. Mol. Biol.* **2009**, *394*, 747–763.

(15) Doudou, S.; Burton, N. A.; Henchman, R. H. Standard free energy of binding from a one-dimensional potential of mean force. *J. Chem. Theory Comput.* **2009**, *5*, 909–918.

(16) Jiang, W.; Roux, B. Free energy perturbation Hamiltonian replica-exchange molecular dynamics (FEP/H-REMD) for absolute ligand binding free energy calculations. *J. Chem. Theory Comput.* **2010**, *6*, 2559–2565.

(17) Singh, N.; Warshel, A. Absolute binding free energy calculations: On the accuracy of computational scoring of protein–ligand interactions. *Proteins* **2010**, *78*, 1705–1723.

(18) General, I. J.; Dragomirova, R.; Meirovitch, H. Calculation of the absolute free energy of binding and related entropies with the HSMD-TI method: The FKBP12-L8 complex. *J. Chem. Theory Comput.* **2011**, *7*, 4196–4207.

(19) Buch, I.; Giorgino, T.; Fabritiis, G. D. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10184–10189.

(20) General, I. J.; Dragomirova, R.; Meirovitch, H. New method for calculating the absolute free energy of binding: The effect of a mobile loop on the avidin/biotin complex. *J. Phys. Chem. B* **2011**, *115*, 168–175.

(21) General, I. J.; Dragomirova, R.; Meirovitch, H. Absolute free energy of binding of avidin/biotin, revisited. *J. Phys. Chem. B* **2012**, *116*, 6628–6636.

(22) Kollman, P. A. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.

(23) Miyamoto, S.; Kollman, P. A. What determines the strength of noncovalent association of ligands to proteins in aqueous solution? *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8402–8406.

(24) Landau, L. D. *Statistical Physics*; The Clarendon Press: Oxford, 1938.

(25) Zwanzig, R. W. High–temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.* **1954**, *22*, 1420–1426.

(26) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 13118–13123.

(27) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The statistical–thermodynamic basis for computation of binding affinities: A critical review. *Biophys. J.* **1997**, *72*, 1047–1069.

(28) Hermans, J.; Wang, L. Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4 lysozyme. *J. Am. Chem. Soc.* **1997**, *119*, 2707–2714.

(29) Jorgensen, W. L. Free-energy calculations: A breakthrough for modeling organic chemistry in solutions. *Acc. Chem. Res.* **1989**, *22*, 184–189.

(30) Hermans, J.; Shankar, S. The free energy of xenon binding to myoglobin from molecular dynamics simulation. *Isr. J. Chem.* **1986**, *27*, 225–227.

(31) Roux, B.; Nina, M.; Pomès, R.; Smith, J. C. Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: A molecular dynamics free energy perturbation study. *Biophys. J.* **1996**, *71*, 670–681.

(32) Fujitani, H.; Tanida, Y.; Ito, M.; Jayachandran, G.; Snow, C. D.; Shirts, M. R.; Sorin, E. J.; Pande, V. S. Direct calculation of the binding free energies of FKBP ligands. *J. Chem. Phys.* **2005**, *123*, 084108.

(33) van Duijneveldt, S.; Frenkel, D. Computer-simulation study of free-energy barriers in crystal nucleation. *J. Chem. Phys.* **1992**, *96*, 4655–4668.

(34) Woods, C. J.; Essex, J. W.; King, M. A. The development of replica-exchange-based free-energy methods. *J. Phys. Chem. B* **2003**, *107*, 13703–13710.

(35) Jorgensen, W. L. Interactions between amides in solution and the thermodynamics of weak binding. *J. Am. Chem. Soc.* **1989**, *111*, 3770–3771.

(36) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. Molecular dynamics study of unbinding of the avidin–biotin complex. *Biophys. J.* **1997**, *72*, 1568–1581.

(37) Hyre, D. E.; Amon, L. M.; E.Penzotti, J.; Trong, I. L.; Stenkamp, R. E.; Lybrand, T. P.; Stayton, P. S. Early mechanistic events in biotin dissociation from streptavidin. *Nat. Struct. Biol.* **2002**, *9*, 582−585.
(38) Allen, T.; Andersen, O.; Roux, B. Energetics of ion conduction through the gramicidin channel. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 117−122.
(39) Lee, M. S.; Olson, M. A. Calculation of absolute protein−ligand binding affinity using path and endpoint approaches. *Biophys. J.* **2006**, *90*, 864−877.
(40) Shoup, D.; Szabo, A. Role of diffusion in ligand binding to macromolecules and cell-bound receptors. *Biophys. J.* **1982**, *40*, 33−39.
(41) Dadarlat, V. M.; Skeel, R. D. Dual role of protein phosphorylation in DNA activator/coactivator binding. *Biophys. J.* **2011**, *100*, 469−477.
(42) Maffeo, C.; Luan, B.; Aksimentiev, A. End-to-end attraction of duplex DNA. *Nucleic Acids Res.* **2012**, *40*, 3812−3821.
(43) Pisabarro, M. T.; Serrano, L. Rational design of specific high-affinity peptide ligands for the Abl-SH3 domain. *Biochemistry* **1996**, *35*, 10634−10640.
(44) Pisabarro, M. T.; Serrano, L.; Wilmanns, M. Crystal structure of the Abl-SH3 domain complexed with a designed high-affinity peptide ligand: Implications for SH3-ligand interactions. *J. Mol. Biol.* **1998**, *281*, 513−521.
(45) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(46) Roux, B. The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* **1995**, *91*, 275−282.
(47) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(48) Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **2000**, *113*, 6042−6051.
(49) Darve, E.; Pohorille, A. Calculating free energies using average force. *J. Chem. Phys.* **2001**, *115*, 9169−9183.
(50) Hénin, J.; Chipot, C. Overcoming free energy barriers using unconstrained molecular dynamics simulations. *J. Chem. Phys.* **2004**, *121*, 2904−2914.
(51) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(52) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(53) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(54) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, L.; Kalé, R. D.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(55) Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulations—The Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613−4621.
(56) Darden, T. A.; York, D. M.; Pedersen, L. G. Particle mesh Ewald: An *N* log *N* method for ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(57) Tuckerman, M. E.; Berne, B. J.; Martyna, G. J. Reversible multiple time scale molecular dynamics. *J. Phys. Chem. B* **1992**, *97*, 1990−2001.
(58) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994**, *222*, 529−539.
(59) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J. Chem. Phys.* **1994**, *100*, 9025−9031.
(60) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Chem. Phys.* **1976**, *22*, 245−268.
(61) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.
(62) Hénin, J.; Forin, G.; Chipot, C.; Klein, M. L. Exploring multidimensional free energy landscapes using time-dependent biases on collective variables. *J. Chem. Theory Comput.* **2010**, *6*, 35−47.
(63) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(64) Grossfield, A. WHAM: the Weighted Histogram Analysis Method, version 2.0.6; Grossfield Lab: Rochester, NY, 2012. Available online: http://membrane.urmc.rochester.edu/content/wham.
(65) Zhu, F.; Hummer, G. Convergence and error estimation in free energy calculations using the weighted histogram analysis method. *J. Comput. Chem.* **2012**, *33*, 453−465.
(66) Yu, Y. B.; Privalov, P. L.; Hodges, R. S. Contribution of translational and rotational motions to molecular association in aqueous solution. *Biophys. J.* **2001**, *81*, 1632−1642.
(67) Zheng, L.; Chen, M.; Yang, W. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *51*, 20227−20232.
(68) Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. Constrained reaction coordinate dynamics for the simulation of rare events. *Chem. Phys. Lett.* **1989**, *156*, 472−477.
(69) Khavrutskii, I. V.; Dzubiella, J.; McCammon, J. A. Computing accurate potentials of mean force in electrolyte solutions with the generalized gradient-augmented harmonic Fourier beads method. *J. Chem. Phys.* **2008**, *128*, 044106.
(70) Pohorille, A.; Jarzynski, C.; Chipot, C. Good practices in free-energy calculations. *J. Phys. Chem. B* **2010**, *114*, 10235−10253.
(71) Hou, T.; Chen, K.; McLaughlin, W. A.; Lu, B.; Wang, W. Computational analysis and prediction of the binding motif and protein interacting partners of the Abl SH3 domain. *PLoS Comput. Biol.* **2006**, *2*, 0046−0055.
(72) Jorgensen, W.; Ravimohan, C. Monte Carlo simulation of differences in free energies of hydration. *J. Chem. Phys.* **1985**, *83*, 3050−3054.
(73) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. String method in collective variables: Minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.* **2006**, *125*, 024106.
(74) Gan, W.; Roux, B. Binding specificity of SH2 domains: insight from free energy simulations. *Proteins: Struct., Func., Bioinf.* **2009**, *74*, 996−1007.
(75) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ, and side-chain χ_1 and χ_2 dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257−3273.

NOTE ADDED AFTER ASAP PUBLICATION

This article was published ASAP on November 16, 2012. Additional changes have been made throughout the manuscript. The correct version was published on November 19, 2012.