Absolute FKBP binding affinities obtained via non-equilibrium unbinding simulations

F. Marty Ytreberg*
Department of Physics, University of Idaho, Moscow, ID 83844-0903
(Dated: November 24, 2008)

We compute absolute binding affinities for two ligands bound to the FKBP protein using non-equilibrium unbinding simulations. The methodology is straight-forward, requiring little or no modification to many modern molecular simulation packages. The approach makes use of a physical pathway, eliminating the need for complicated alchemical decoupling schemes. Results of this study are promising. For the ligands studied here the binding affinities are typically estimated within less than 4.0 kJ/mol of the target values; and the target values are within less than 1.0 kJ/mol of experiment. These results suggest that non-equilibrium simulation could provide a simple and robust means to estimate protein-ligand binding affinities.

I. INTRODUCTION

The accurate estimation of binding affinities for protein-ligand systems (ΔG) remains one of the most challenging tasks in computational biophysics and biochemistry.[1] Due to the high computational cost of such free energy computation, it is of interest to understand the advantages and limitations of various ΔG methods.

Many previous studies[2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,18,20,21,22,23] have calculated protein-ligand binding affinities using equilibrium free energy methods such as thermodynamic integration,[22,23,24] and weighted histogram analysis.[22] Due to the introduction of the novel Jarzynski approach[25] it is also possible to estimate ΔG from non-equilibrium simulations. However, the estimation of ΔG for protein-ligand binding using non-equilibrium approaches remains largely untested. Two recent studies used non-equilibrium simulations to test unbinding pathways.[25,26] Studies by other groups found that estimating ΔG via non-equilibrium simulations resulted in a large error compared to experiments.[27] A recent study by Kuyucak and collaborators demonstrated that use of non-equilibrium simulation required longer simulation times than umbrella sampling.[28]

In this report, apparently for the first time, we demonstrate the ability to compute accurate (as compared to experimental data) protein-ligand binding ΔG estimates following a non-equilibrium methodology. The approach relies on performing multiple non-equilibrium unbinding simulations using a physical pathway, i.e., pulling the ligand out of the binding pocket, and then uses the Jarzynski relation[25] to estimate ΔG. The system is an FKBP protein complexed with 4-hydroxy-2-butanone (BUQ) and dimethyl sulfoxide (DMSO). The motivation for using this system is that comparison to experiment is possible[29] and many previous computational studies have been performed.[12,13,15,20,21]

The importance of pursuing non-equilibrium methods such as used in this report is three-fold: (i) The approach is trivially parallelizable since each non-equilibrium unbinding simulation is performed independently. (ii) The method is simple to implement in many existing simulation packages such as GROMACS[30] used here; little or no modification to the code is necessary. (iii) Since a physical pathway is utilized, there is no need to use alchemical decoupling schemes.

This study represents the first stage of a project comparing the efficiencies of various free energy methods for protein-ligand ΔG computation. We note that efficiency studies have been carried out for other non-protein systems.[31,32,33,34,35,36,37,38,39,40,41]

II. THEORY

In general, the absolute binding affinity is defined as the free energy difference between the unbound (apo) and bound (holo) states of the protein-ligand system. We define the apo state as when the protein and ligand are not interacting due to a large separation between them. The holo state is defined to be when the ligand is in the binding pocket of the protein. Experimentally, the binding affinity is measured by determining the equilibrium constant $K_{eq} = [PL]/[P][L]$, where $[PL]$ denotes the concentration of the protein-ligand complex, and $[P]$ and $[L]$ are the concentrations of the apo protein and free ligand, respectively. Then the absolute binding affinity is given by $\Delta G_{bind} = -k_B T \ln(K_{eq}/V_0)$, where $k_B$ is the Boltzmann constant, $T$ is the system temperature, and $V_0$ is the standard volume used for the experiment (typically $V_0 = 1.661 \text{ nm}^3$ corresponding to 1.0 mol/liter concentration).

Following the notation of Roux and collaborators[10] (also see discussion in Refs. [11,12]) the equilibrium constant is given by a ratio of integrals over the apo and holo regions of configurational space

$$
K_{eq} = \frac{\int_{\text{holo}} d\vec{x} \int d\vec{X} \ e^{-\beta U(\vec{x}, \vec{X})}}{\int_{\text{apo}} d\vec{x} \int d\vec{r} \ \delta(\vec{r} - \vec{r}_c) \int d\vec{X} \ e^{-\beta U(\vec{x}, \vec{X})}},
$$

where $\beta = 1/k_B T$, $\vec{x}$ represents the configurational coordinates of the ligand, $\vec{X}$ are the coordinates of the protein.
Eq. (1) can be written as

\[ I_1 = \int_{\text{holo}} \int d\vec{X} e^{-\beta(U + U_a)} \]

where strategies for computing the terms \( I_1 \) and \( I_2 \) will be discussed below.

The first term \( I_1 \equiv e^{-\beta \Delta G_{\text{holo}}} \) in Eq. (2) corresponds to the free energy difference associated with restraining the protein to the binding axis while in the holo state. This free energy can be computed using any standard technique by performing simulations for a range of force constants from 0 to \( k_a \). For the current study we chose to compute this free energy difference via a multi-stage Bennett approach.\(^{12}\)

To determine the second term \( I_2 \) in Eq. (2) we define the potential of mean force (PMF) \( W \), with the restraint potential \( U_a \) present, as a function of the scalar distance \( r \)

\[ e^{-\beta(W(r_2) - W(r_1))} = \int d\vec{x} \delta(r - r_2) \int d\vec{X} e^{-\beta(U + U_a)} \]

Integrating the PMF over both apo and holo regions we can obtain

\[ I_2 = \frac{1}{4\pi r^2} \int_{\text{holo}} d\vec{x} \int d\vec{X} e^{-\beta(U + U_a)} \]

where we have used the fact that \( \delta(r - r_a) = 4\pi r^2 \delta(r - r_a) \) for the apo integral since the PMF is independent of the direction of \( \vec{r} \) when the ligand is not interacting with the protein. Thus \( I_2 \) can be evaluated by estimating the integral of the PMF in Eq. (4),

\[ I_2 = \frac{4\pi r^2}{V_0} \int_{\text{apo}} d\vec{r}_1 \delta(\vec{r}_1 - r_a) e^{-\beta U_a} \int_{\text{holo}} dr_2 e^{-\beta(W(r_2) - W(r_1))} \]

which have been used the fact that \( \delta(r - r_a) = 4\pi r^2 \delta(r - r_a) \) for the apo integral since the PMF is independent of the direction of \( \vec{r} \) when the ligand is not interacting with the protein. Thus \( I_2 \) can be evaluated by estimating the integral of the PMF in Eq. (4),

\[ I_2 = \frac{4\pi r^2}{V_0} \int_{\text{apo}} \int d\theta_1 d\phi_1 \frac{\delta(r_1 - r_a)}{4\pi r^2} e^{-\beta U_a(\theta_1, \phi_1)} \int_{\text{holo}} dr_2 e^{-\beta W(r_2)} \]

the holo integral will be evaluated using quadrature.

Below the apo integral will be evaluated analytically and
With our approximations above the absolute binding free energy can now be estimated via the relation
\[
\Delta G_{\text{bind}} = \Delta G_{a}^{\text{holo}} - \mathcal{W}(r_{a}) - k_{B}T \ln \left[ \frac{V_{0}^{2}}{r_{0}^{2}} \times \right.
\int_{\text{apo}} \cos \theta d\theta d\phi \left. e^{-\beta U_{a}(\theta, \phi)} \int_{\text{holo}} dr \ e^{-\beta \mathcal{W}(r)} \right].
\] (6)

This is our central theoretical result. Thus, estimating \(\Delta G_{\text{bind}}\) in the current framework requires three computations: the PMF must be calculated (detailed below), \(\Delta G_{a}^{\text{holo}}\) must be approximated, and the apo integral must be analytically evaluated.

One may note that a different, yet viable, non-equilibrium approach would be to set \(k_{a} = 0\), i.e., remove the axial restraint. This would simplify the \(\Delta G_{\text{bind}}\) calculation since \(\Delta G_{a}^{\text{holo}} = 0\) and the apo integral would be \(4\pi\) in Eq. (6), and thus only the PMF would be needed. Further, the binding axis would not need to be defined by the researcher, which would allow the ligand more flexible unbinding routes. However, there are two important advantages to using the axial restraint. Most important is that since the axial restraint limits the allowable configurational space for the ligand, the PMF converges much more quickly than with no restraint. Also, in cases where the binding pocket is not at the protein surface, or when more than one viable pathway exists, it may be advantageous, or even necessary, to define the binding path to obtain meaningful results.

### A. Estimating the PMF

The PMF will be computed using two different approaches: the Hummer-Szabo method, and the stiff-spring approximation with the second cumulant expansion. Below we summarize these techniques.

In the Hummer-Szabo approach the PMF is estimated by performing multiple non-equilibrium pulling simulations along the reaction coordinate \(r\) by using a time-dependent biasing potential \(U_{r}(r, t) = k_{r}[r - (r_{0} + vt)]^{2}\), where \(k_{r}\) is the force constant, \(r = r(t)\) is the protein-ligand center of mass separation, \(r_{0}\) is an initial reference separation which is constant for all pulling simulations (i.e., \(r_{0} \neq r(0)\)), and \(v\) is the speed at which the biasing center is moved. The PMF is then estimated via \[W_{t}(r) \approx -\int_{t_{0}}^{t} \frac{\partial U_{r}(r, t)}{\partial r} dr - U_{r}(r(0), 0) = k_{r}v \left[ \frac{1}{2}v^{2} t - \int_{0}^{t} [r(t') - r_{0}] dt' \right] - U_{r}(r(0), 0). \] (8)

Note that \(W_{t}\) is the accumulated work minus the initial \(t = 0\) biasing energy.

The stiff-spring approximation utilizes the well-known Jarzynski equality \[e^{-\beta \mathcal{W}(r)} \approx e^{-\beta \Delta G(r)} = \langle e^{-\beta W_{r}} \rangle \times e^{-2 \ln(r)}, \] (9)

where the \(2 \ln(r)\) term is due to the Jacobian correction, and the work is determined by integrating the biasing force over the location of the bias center

\[W_{r} \approx W_{\xi} = \int_{r_{0}}^{r_{0}+vt} \frac{\partial U_{r}(r, t)}{\partial r} dr \approx -\int_{r_{0}}^{r_{0}+vt} \frac{\partial U_{r}(r, t)}{\partial r} d\xi - U_{r}(r(0), 0), \] (10)

where we have used the fact that \(\partial r / \partial \xi \approx 1\). Applying the cumulant expansion to Eq. (9), we obtain the final expression used estimate the PMF for the stiff-spring approach \[W(r) \approx \langle W_{r} \rangle - \frac{\beta}{2} \left[ \langle W_{r}^{2} \rangle - \langle W_{r} \rangle^{2} \right] + 2 \ln(r). \] (11)

For the results given in this report the ligand is pulled out of the binding pocket, and the reverse process of pulling the ligand into the pocket is not considered. Future studies will include the reverse process since the use of bi-directional simulation has been shown to be an effective approach to accurate \(\Delta G\) estimation.

We note two aspects of the relationships embodied in Eqs. (7) and (11): (i) The equality in Eq. (7) holds only in the case of obtaining all possible pulling trajectories. The approximation in Eq. (11) is an equality for the case that the work value distribution is perfectly Gaussian. Thus, it is important to calculate uncertainty estimates for the PMF, and if possible, to compare results to an independent computational measure—below we will compare our results to use of umbrella sampling. (ii) The relation is independent of the speed at which the system is forced, i.e., the unbinding speed. In practice, however, it has been found that the speed chosen can dramatically affect the convergence behavior of the estimate.
B. Use of a physical pathway

It is useful to consider the advantages and disadvantages of using a physical (rather than alchemical) pathway. The regions of configurational space corresponding to apo and holo in Eq. (1) are well-separated with no overlap, thus a pathway connecting them is typically created. For our discussion below, this pathway will be parameterized using the variable $\lambda$.

In the case of a physical pathway, such as in the current study, $\lambda = r$ represents the protein-ligand separation. By contrast, for an alchemical pathway $\lambda$ is generally a parameter that scales the strength of the interactions between the ligand and rest of the system. Our use of a physical pathway is motivated by several factors. Alchemical pathways are typically much more difficult to implement than physical pathways since interactions must be scaled carefully. In addition, restraints must often be employed such that the non-interacting parts do not drift away from the region of interest.

We note that there are disadvantages to using physical pathways. Physical pathways may require the researcher to determine the pulling direction such that the ligand exits the binding pocket, i.e., determined by choice of $U_a$ in this report. Alchemical pathways do not require such a choice. Perhaps most important, physical pathways require larger system sizes when explicit solvent is used, as in the current report. The size of the system must be large enough that the ligand can be pulled to a distance such that interactions between the ligand and protein are negligible.

In cases where the binding site is buried deep within the protein, alchemical methods should be much more efficient than physical approaches. However, when the binding pocket is close to the protein surface, as for the current study, it is not clear where alchemical or physical approaches are more efficient and/or accurate.

Another important consideration is that the use of a physical pathway allows the researcher to determine the PMF along the pathway. This PMF can give insights into binding that are simply not possible when using alchemical methods, e.g., determining the preferred binding pathway when multiple pathways are present.

C. Use of a non-equilibrium approach

Non-equilibrium approaches, such as used in the current study, rely on computing the work required to force the system from one state to the other rapidly enough that equilibrium is not attained at any value of $\lambda$. This process is repeated many times and the resulting distribution of work values is used to estimate $\Delta G$. By contrast, equilibrium free energy methodologies such as thermodynamic integration, free energy perturbation, and weighted histogram analysis, share the common strategy of generating equilibrium ensembles of configurations for multiple values of the scaling parameter $\lambda$. It is important when performing such $\Delta G$ computation that enough simulation time is spent to equilibrate at each value of $\lambda$ such that the resulting ensemble is valid for the current $\lambda$.

It is not currently known whether equilibrium or non-equilibrium methodologies offer an efficiency advantage for typical protein-ligand binding affinity computation. Equilibrium methods have been widely used to generate accurate $\Delta G$ estimates for protein-ligand binding. However, if equilibrium is not attained the resulting $\Delta G$ estimate can be heavily biased. With few very recent exceptions, non-equilibrium methods are largely untested on protein-ligand systems. In previous calculations of relative solvation free energies non-equilibrium methods were proven to be equal or superior in efficiency to commonly used equilibrium methods.

A key advantage of non-equilibrium methodologies is the ease that one can parallelize the $\Delta G$ calculation. Since each work value must necessarily be generated independently, the corresponding simulations can be run in parallel with no loss of accuracy to the final $\Delta G$ estimate. Equilibrium $\Delta G$ computations, by contrast, are not trivially parallelizable. One can imagine performing each $\lambda$ simulation in parallel, however one must be very careful about the configurations used to start each $\lambda$ simulation. In typical cases it is necessary to start the current $\lambda$ simulation using the final snapshot from the previous $\lambda$ simulation; thus, the $\lambda$ simulations are performed in a serial fashion. If this is not done, the amount of time needed to equilibrate at each value of $\lambda$ could be heavily dependent on the chosen starting structures. The $\Delta G$ estimate could be heavily biased if the time spent for equilibration at each $\lambda$ value is inadequate.

III. METHODS

A. Computational details

The initial coordinates for the FKBP-ligand complexes were obtained from the Protein Data Bank, 1D7H for FKBP-DMSO, and 1D7J for FKBP-BUQ. The topologies for DMSO and BUQ were then generated by the PRODRG server with partial charges modified by the author.

The GROMACS simulation package version 3.3.3 was used with the default GROMOS-96 43A1 forcefield. The software was slightly modified to provide the biasing potential $U_r$ which depends only on the center of mass separation between the ligand and the protein. Protonation states for the histidine residues were selected by the GROMACS program pdb2gmx: HIS25 was protonated at Nδ1, and HIS87 and HIS94 were protonated at Ne2. The protein-ligand complexes were then solvated in a cubic box of SPC water of approximate initial size 6.8 nm a side. A single chloride

\[ \text{G} \]

\[ \text{b} \]

\[ \text{a} \]

\[ \text{c} \]

\[ \text{d} \]

\[ \text{e} \]

\[ \text{f} \]

\[ \text{g} \]

\[ \text{h} \]

\[ \text{i} \]

\[ \text{j} \]

\[ \text{k} \]

\[ \text{l} \]

\[ \text{m} \]

\[ \text{n} \]

\[ \text{o} \]

\[ \text{p} \]

\[ \text{q} \]

\[ \text{r} \]

\[ \text{s} \]

\[ \text{t} \]

\[ \text{u} \]

\[ \text{v} \]

\[ \text{w} \]

\[ \text{x} \]

\[ \text{y} \]

\[ \text{z} \]
ion was randomly placed in each water box to give a net neutral charge, and then each system was minimized using steepest decent for 500 steps. To allow for equilibration of the water, each system was then simulated for 1.0 ns with the positions of all heavy atoms in the ligand and protein harmonically restrained with a force constant of 1000 kJ/mol/nm$^2$. The temperature was maintained at 300 K using Langevin dynamics$^{25}$ with a friction coefficient of 1.0 amu/ps. The pressure was maintained at 1.0 atm using the Berendsen algorithm$^{26}$. We note that the Berendsen algorithm does not produce canonically distributed structures, however, none of the resulting simulation frames were used for generating $\Delta G$ estimates, as will be seen below. The LINCS algorithm$^{62}$ was used to constrain hydrogens to their ideal lengths and heavy hydrogens were used—the hydrogen mass was increased by a factor of four and this increase was subtracted from the bonded heavy atom so that the mass of the system remained unchanged—allowing the use of a 4.0 fs timestep. Particle mesh Ewald$^{63}$ was used for electrostatics with a real-space cutoff of 1.0 nm and a Fourier spacing of 0.1 nm. Van der Waals interactions used a cutoff with a smoothing function such that the interactions smoothly decayed to zero between 0.75 nm and 0.9 nm. Dispersion corrections for the energy and interactions for use in the PMF calculations. Each FKBP-protein complex was simulated with parameters chosen identical to the position restrained simulation above except that the volume was fixed at the value of the final configuration from the position restrained simulations. Importantly, for Eq. (7) and (11) to be used these equilibrium simulations must include the restraints, i.e., both $U_a$ and $U_v$ were present. For both the DMSO and BUQ systems the axial restraint used a force constant $k_a = 1000$ kJ/mol/nm$^2$, and $\theta_0, \phi_0$ were chosen to be equal to the values from the final snapshot of the position restrained simulation. For both systems the biasing potential $U_v$ used a force constant $k_v = 3000$ kJ/mol/nm$^2$ and $r_0 = 0.5$ nm and a speed $v = 0$.

Starting structures for the unbinding simulations were chosen to be equally spaced within the 4.0 ns equilibrium simulation. So, if 40 starting structures were desired, then the spacing between snapshots was 100 ps. The pulling simulations were performed using identical parameters to the 4.0 ns equilibrium simulation, except that the bias center was moved at a constant speed $v$ ranging from $1.0 \times 10^{-4}$ nm/ps to $8.0 \times 10^{-4}$ nm/ps. The pulling simulations were discontinued when the bias center was at a position of 2.5 nm.

The non-equilibrium unbinding simulations provided us with the protein ligand separation $r$ at every time step, which we used to compute the work and thus the resulting PMF. An example of this is shown in Fig. 2 where we have used Eq. (8) to compute the work as a function of simulation time $W_t$. 

**B. Computing $\Delta G_a^{\text{holo}}$**

We used the Bennett acceptance ratio approach to compute the free energy differences $\Delta G_a^{\text{holo}}$ associated with the axial restraints$^{12}$ With the ligand bound to the protein we performed 1.0 ns equilibrium simulations for each of the values $k_a = 0, 25, 40, 60, 90, 150, 200, 300, 450, 700, 1000$. The first 0.5 ns of each simulation were discarded for equilibration, and the remaining 0.5 ns were used to compute $\Delta G_a^{\text{holo}}$. We did not attempt to optimize efficiency of the $\Delta G_a^{\text{holo}}$ computations, our only concern was accurate values, so it may be possible to reduce the total computational time from that described above.

**C. Uncertainty estimation**

We estimated the uncertainty in our $\Delta G_{\text{bind}}$ estimates using the bootstrap approach applied to the PMF: (i) The reference value of the PMF given by $W(r_*)$ was computed via Eq. (11) using $N$ work values chosen at random (with replacement) from a dataset containing $N$ values; (ii) The above step was repeated until the mean and standard deviation of the free energy estimates converged; around 100,000 trials in our study. (iii) The uncertainty is given by the converged standard deviation of the free energy estimates.

For comparison, we also used the uncertainty analysis obtained by Zuckerman and Woolf$^{52}$ and the Bustamante group$^{55}$. These uncertainty estimates are reported to be accurate when the variance in the estimate dominates over the bias (as in the case of large $N$).
D. Generating a target PMF

Since the purpose of the current study was to test the effectiveness of non-equilibrium strategies it is important to have an independent estimate of the PMF. Thus, we computed the PMF using umbrella sampling and weighted histogram analysis (WHAM). Simulations were performed with the restraints \( U_a \) and \( U_r \). For the umbrella sampling simulations the speed was set to \( v = 0 \), all other parameters were identical to the non-equilibrium simulations, and 41 windows were used \( r_0 = 0.50, 0.55, 0.60, ..., 2.45, 2.50 \). Each window was simulated for a total time of 12 ns; 6 ns were discarded for equilibration and 6 ns were used for the WHAM analysis. Thus, the total simulation time was nearly three times greater than the non-equilibrium simulations detailed above. No attempt was made to test the efficiency since the goal was to generate the most accurate PMF. Note that the \( 2 \ln(r) \) Jacobian correction from Eqs. (7) and (11) were also used for the target PMF.

IV. RESULTS AND DISCUSSION

The results of this study are very encouraging. Using the non-equilibrium methodology outlined above we estimated the binding affinity for the FKBP-DMSO and FKBP-BUQ complexes typically within less than 4.0 kJ/mol of the target values; and the target values are within less than 1.0 kJ/mol of experiment.

Figure 3 shows the PMF as a function of protein-ligand separation for all systems studied here. Data is shown for both DMSO and BUQ systems, with pulling speeds indicated on each plot. Note that the same amount of total simulation time was spent on each non-equilibrium PMF estimate, but that the target PMF utilized three times as much simulation as the non-equilibrium estimates, thus we are not attempting to compare non-equilibrium and equilibrium approaches in this study. For both systems the non-equilibrium estimates tend to overestimate the target PMF, and underestimate the broadness of the PMF minimum. This suggests that the pulling speeds were not slow enough to properly sample the PMF. Also, use of stiff-spring approximation with the second cumulant expansion does tend to improve the non-equilibrium PMF curves.

Table I shows the binding affinity results obtained via Eq. (6), with the PMF computed using both the Hummer-Szabo approach of Eq. (7) and the stiff-spring second cumulant approximation of Eq. (11). The computational estimates of \( \Delta G_{\text{bind}} \) for the target PMF are in excellent agreement with experimental data. The non-equilibrium estimates are typically within less than 4.0 kJ/mol of the target values. Reference distances were chosen as \( r_s = 2.4 \) nm for both DMSO and BUQ, and the value of the restraint free energy was found to be \( \Delta G_a^{\text{apo}} = -9.0 \) kJ/mol for FKBP-DMSO and \( \Delta G_a^{\text{apo}} = -7.8 \) kJ/mol for FKBP-BUQ. Finally, the value of the apo integral in Eq. (6) was computed analytically to be 10.6 kJ/mol for FKBP-DMSO and 10.7 kJ/mol for FKBP-BUQ. The results show that non-equilibrium estimates for the FKBP-DMSO system are more accurate than the FKBP-BUQ estimates suggesting that the FKBP-BUQ system requires more simulation time to converge. Uncertainty estimates were obtained using both a bootstrap method and the approach described in Refs. [55,65].

Table I includes the standard deviation of the work values \( \sigma_W \) measured at the reference distance \( r_s = 2.4 \) nm. Previous studies have suggested that the optimal efficiency for use of the Jarzynski relation is when the speed is slow enough that \( \sigma_W \approx 1.0 \ k_B T \approx 2.5 \) kJ/mol. Apparently the speeds attempted for the current study were not slow enough to generate work values with such a small \( \sigma_W \). We note however, for the current study, the uncertainty does not appear to correlate with \( \sigma_W \). Future studies will be carried out to determine if there is an optimal pulling speed for these systems.

The results from Tab. I suggest that use of the Hummer-Szabo approach, while exact in the limit of infinite sampling, is not feasible for the current study. This is likely due to the fact that the pulling speeds used were too fast to generate work values with \( \sigma_W \approx 1.0 \ k_B T \).
TABLE I: Comparison between non-equilibrium binding affinity estimates and target and experimental binding affinities. All energy values are shown in units of kJ/mol. The first column describes the ligand used. The second column contains the number of work values \( N \) used in the estimate, and the third and fourth columns are respectively, the corresponding speed of the restraint attached to the ligand, and the standard deviation of the work values. The rightmost column gives the experimental results reported in Ref. \([54]\).

\( a \) Binding affinity estimate obtained via the Hummer-Szabo approach using Eqs. \([6]\) and \([7]\).

\( b \) Binding affinity estimate obtained using the stiff-spring second cumulant expansion approximation using Eqs. \([6]\) and \([11]\).

\( c \) Uncertainty estimate computed via the bootstrap method.

\( d \) Uncertainty estimate computed from the approach described in Refs. \([55, 56]\).

A. Note on simulation time

Each non-equilibrium estimate in Table I was generated using a total simulation time of 216.0 ns (1.0 ns position restrained + 4.0 ns equilibrium to generate starting configurations + 200.0 ns for unbinding simulations + 11.0 ns for \( \Delta G_a^{\text{holo}} \) estimation). Note however, that the unbinding simulations were performed in parallel. So, for example, at a speed of \( 2.0 \times 10^{-4} \) nm/ps, twenty independent 10.0 ns simulations were performed in parallel. Therefore, all the simulation data needed to compute \( \Delta G_{\text{bind}} \) can be obtained in around a day of wall-clock time with the use of a computer cluster.
Acknowledgments

Funding for this research was provided by the University of Idaho, Idaho NSF-EPSCoR and BANTech. Computing resources were provided by IBEST at University of Idaho, and by the TeraGrid Advanced Support Program. FMY would like to thank Ronald White, David Mobley, and Daniel Zuckerman for helpful discussion.
I. G. Tironi, *Biomolecular Simulation: The GROMOS96 manual and user guide* (Hochschulverlag, Zürich, 1996).

59 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, *Intermolecular Forces* (Reidel, Dordrecht, 1981).

60 W. F. van Gunsteren, H. J. C. Berendsen, and J. A. C. Rullmann, Mol. Phys. **44**, 69 (1981).

61 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).

62 B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, J. Comput. Chem. **18**, 1463 (1997).

63 T. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).

64 M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1989).

65 D. M. Zuckerman and T. B. Woolf, Phys. Rev. Lett. **89**, 180602 (2002).

66 D. D. L. Minh and J. A. McCammon, J. Phys. Chem. B **112**, 5892 (2008).