



On Calculating Free Energy Differences Using Ensembles of Transition Paths

Robert Hall¹, Tom Dixon^{1,2} and Alex Dickson^{1,2*}

¹ Department of Biochemistry & Molecular Biology, Michigan State University, East Lansing, MI, United States, ² Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI, United States

OPEN ACCESS

Edited by:

Pratyush Tiwary,
University of Maryland, College Park,
United States

Reviewed by:

Steffen Wolf,
University of Freiburg, Germany
Wei Yang,
State College of Florida,
Manatee-Sarasota, United States

*Correspondence:

Alex Dickson
alexrd@msu.edu

Specialty section:

This article was submitted to
Biological Modeling and Simulation,
a section of the journal
Frontiers in Molecular Biosciences

Received: 26 February 2020

Accepted: 06 May 2020

Published: 05 June 2020

Citation:

Hall R, Dixon T and Dickson A (2020)
On Calculating Free Energy
Differences Using Ensembles of
Transition Paths.
Front. Mol. Biosci. 7:106.
doi: 10.3389/fmolb.2020.00106

The free energy of a process is the fundamental quantity that determines its spontaneity or propensity at a given temperature. In particular, the binding free energy of a drug candidate to its biomolecular target is used as an objective quantity in drug design. Recently, binding kinetics—rates of association (k_{on}) and dissociation (k_{off})—have also demonstrated utility for their ability to predict efficacy and in some cases have been shown to be more predictive than the binding free energy alone. Some methods exist to calculate binding kinetics from molecular simulations, although these are typically more difficult to calculate than the binding affinity as they depend on details of the transition path ensemble. Assessing these rate constants can be difficult, due to uncertainty in the definition of the bound and unbound states, large error bars and the lack of experimental data. As an additional consistency check, rate constants from simulation can be used to calculate free energies (using the log of their ratio) which can then be compared to free energies obtained experimentally or using alchemical free energy perturbation. However, in this calculation it is not straightforward to account for common, practical details such as the finite simulation volume or the particular definition of the “bound” and “unbound” states. Here we derive a set of correction terms that can be applied to calculations of binding free energies using full reactive trajectories. We apply these correction terms to revisit the calculation of binding free energies from rate constants for a host-guest system that was part of a blind prediction challenge, where significant deviations were observed between free energies calculated with rate ratios and those calculated from alchemical perturbation. The correction terms combine to significantly decrease the error with respect to computational benchmarks, from 3.4 to 0.76 kcal/mol. Although these terms were derived with weighted ensemble simulations in mind, some of the correction terms are generally applicable to free energies calculated using physical pathways via methods such as Markov state modeling, metadynamics, milestoning, or umbrella sampling.

Keywords: free energy, molecular dynamics, enhanced sampling, binding kinetics, statistical mechanics, nonequilibrium

1. INTRODUCTION

In recent years there is a growing appreciation for the utility of binding kinetics in the prediction of drug efficacy (Lu and Tonge, 2010; Carroll et al., 2012; Vauquelin et al., 2012; Pei et al., 2014; Ayaz et al., 2016; Copeland, 2016; Costa et al., 2016; Guo et al., 2016; Tonge, 2017; Bruce et al., 2018; Lee et al., 2019; Nunes-Alves et al., 2020). Pharmacokinetic and pharmacodynamic models of drug activity in the body are inherently out of equilibrium: a drug is administered, it is absorbed, distributed to different tissues, metabolized and eliminated from the body. As such, kinetic constants of binding and release—beyond just the equilibrium constants of binding—are required to model drug action when the timescales of binding and release cannot be separated from the other competing processes (Bernetti et al., 2017). The relationship between molecular structure and the kinetics of binding (also called “structure-kinetic relationships” or SKR) is complicated, as small changes to structure can change kinetic constants by orders of magnitude (Ayaz et al., 2016). It is important to note that changes in kinetics are not always tied to changes in affinity (Guo et al., 2014), and that to accurately predict changes in kinetics, models of the ligand-binding transition state are needed to estimate transition-state stabilization or destabilization (Spagnuolo et al., 2017).

Computational methods that reveal structures of transition states and calculate binding (k_{on}) and unbinding (k_{off}) rate constants for real compounds are in their infancy, but are quickly developing (Dickson et al., 2017). It is a tremendous challenge to obtain reliable values for these quantities, as (1) they depend on the entire (un)binding pathway, not just its endpoints, and (2) the timescales of ligand binding and release often exceed the capabilities of molecular dynamics simulations by orders of magnitude. Specialized computing platforms have been applied to generate continuous binding pathways (Dror et al., 2011), although the unbinding process is typically beyond the reach of molecular dynamics simulation for compounds beyond millimolar drug fragments (Guo et al., 2016; Pan et al., 2017). Recent studies have used enhanced sampling methods in molecular dynamics to simulate ligand (un)binding pathways and determine mechanisms and rate constants k_{on} and k_{off} (Casasnovas et al., 2017; Tiwary et al., 2017; Dickson, 2018; Kokh et al., 2018; Lotz and Dickson, 2018; Bruno et al., 2019; Deb and Frank, 2019; Kirberger et al., 2019; Dixon et al., 2020). Some of these rate constants have shown surprisingly good agreement with experiment—given the extraordinarily long timescales involved—however these have the confounding uncertainty of force field accuracy (Yin et al., 2017; Camilloni and Pietrucci, 2018), there is a possibility for fortuitous cancelation of error. Unfortunately, the computational cost required to predict these quantities is typically massive (Camilloni and Pietrucci, 2018), especially for large protein systems and ligands with extremely long residence times, precluding the study of these events under a series of different simulation conditions (e.g., forcefields, water models, polarizability).

In the field of biomolecular modeling, blind challenges—where a series of objectives are released by the organizers, and participants entries are directly judged by their agreement with

experiment—have been useful catalysts for the development of predictive algorithms (Lensink et al., 2017; Synapse, 2018; Croll et al., 2019; Parks et al., 2020). Although no blind challenge currently exists for the prediction of k_{on} and k_{off} , we recently participated in the SAMPL6 SAMPLING challenge, which required participants to compute free energies as a function of simulation time and to compare the computational cost of different free energy calculation methods (Dixon et al., 2018; Rizzi et al., 2018, 2020). This challenge allows sampling methods to be assessed independently of force field accuracy, as all entries used the same initial coordinates, force field parameters and partial charges. Importantly, the challenge makes use of very small model systems (host-guest) that require considerably less computational resources to simulate, which allowed us to efficiently simulate binding and release for a number of systems, determine k_{on} and k_{off} , and predict values for the binding free energy (ΔG) that would then be compared to experimental observables, as well as results from alchemical free energy perturbation methods (Gilson et al., 1997; Shirts and Chodera, 2008).

The standard free energy of binding was determined as a function of rate constants:

$$\Delta G = -k_B T \ln \frac{C^0 k_{\text{on}}}{k_{\text{off}}} \quad (1)$$

where C^0 is a reference concentration of 1 mol/L. In this paper, we revisit this equation in detail and explicitly examine the assumptions made when the rate constants used in Equation (1) are computed through typical simulations with finite box-size and periodic boundary conditions. In section 3.1, we derive three correction terms that can be easily computed and facilitate a better connection with both experiment and alchemical computational free energy calculations. One term accounts for the particular definitions of the bound and unbound states. The second term accounts for residual electrostatic interactions that might still be present between the molecules, which is especially useful if one or both of the molecules carry an explicit charge. The third term accounts for the volume of the unbound state in the simulation box, which is useful to keep the simulated volume as small as possible during rate calculations. These terms were derived particularly with weighted ensemble simulations in mind, where rates are computed using the trajectory flux between two non-equilibrium ensembles. However, the second and third term can be directly applied to other simulation methods which employ physical simulation of the transition path ensemble, such as Markov state modeling (Singhal et al., 2004; Gu et al., 2014), metadynamics (Laio and Parrinello, 2002; Tiwary et al., 2017), milestone (Faradjian and Elber, 2004; Votapka et al., 2017), and umbrella sampling (Torrie and Valleau, 1977; Nishikawa et al., 2018).

To examine questions of convergence, we reproduce our binding and unbinding simulations for a host-guest system with larger numbers of replicas and longer simulation times. We also explore the effects of the Langevin integrator on the prediction of unbinding and binding rates; in particular, how altering the friction coefficient (γ), defined in the Langevin integrator,

impacts the binding and release processes. Although γ does not appear in the internal energy function, and hence cannot affect thermodynamic properties such as the binding free energy, we examine whether lower friction coefficients can accelerate the convergence of unbinding simulations.

2. METHODS

2.1. Host-Guest Systems

The host-guest system utilized in this study is referred to as “OA-G6” (Figure 1), where the host is a Gibb deep cavity cavitand, referred to as an “octa acid” or “OA” (Gan et al., 2011). OA forms a basket-like structure with 4-fold symmetry, functionalized with four benzoic-acid substituents on the top rim of the basket and four more on the bottom. The guest ligand we study here is 4-methyl pentanoic acid (referred to as “G6”). This ligand harbors a negative charge at the carboxyl end of the alkyl chain.

2.2. Molecular Dynamics

The OA-G6 configuration was obtained from the organizers of the SAMPLing challenge (Rizzi et al., 2020). The system was solvated in a (roughly) cubic box with box length 4.28, 4.33, and 4.33 nm in the x , y and z dimensions, respectively. The system provided had a total of 7,976 atoms: 2,586 water molecules to solvate the system, 12 sodium and 3 chloride ions to neutralize the system, and the remaining atoms belonging to either the host or the guest. Forcefield parameters for the system are as provided by the original organizers of the SAMPLing challenge (Rizzi et al., 2018). The system was parameterized using GAFF (Wang et al., 2004) and then converted into Gromacs format. The conversion was done using ParmEd version 2.7.3. OpenMM v7.2.1 (Eastman et al., 2017) was used to run dynamics with the CUDA v9.0.176 platform. A Monte Carlo barostat is used to maintain a constant pressure of 1 atm. A timestep of 2 fs was used across all simulations.

We utilize the Langevin integrator, which uses a drag term and a noise term to account for the friction of solvent molecules and high velocity collisions that perturb the system. Langevin dynamics allows for the temperature to be controlled and can be used as a thermostat; we run all dynamics here at 300 K. Our host-guest system is propagated with the Langevin equation, shown below:

$$\mathbf{F} = m\mathbf{a} = -\nabla U(\mathbf{r}) - m\gamma\mathbf{v} + \sqrt{\frac{2m\gamma k_B T}{\tau}}\mathbf{R}(t) \quad (2)$$

where $U(\mathbf{r})$ is the particle interaction potential, $\mathbf{R}(t)$ is a random Gaussian noise term evaluated every timestep, T is the temperature, k_B is the Boltzmann constant, τ is the timestep and γ is the friction coefficient in units of inverse time. The friction term plays two different roles here, both modulating the second “drag” term, and the Gaussian noise. As γ approaches zero, the noise gets weaker and the dynamics becomes more deterministic. Here we run binding and unbinding simulations with γ values of 1.0, 0.1, and 0.01 ps⁻¹.

2.3. Reweighting of Ensembles by Variance Optimization

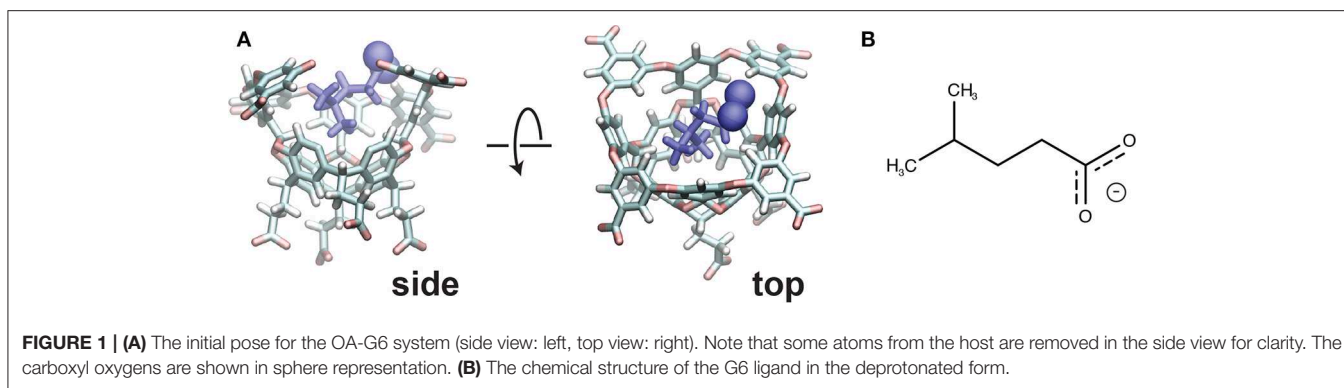
To generate an ensemble of ligand unbinding events, we need to employ enhanced sampling as the timescale of ligand unbinding events in this system is prohibitively long; we found in previous studies a mean first passage time of 2.1 s (Dixon et al., 2018), which is six orders of magnitude longer than the reach of conventional MD sampling. In this work, we implement the REVO resampling method, based on weighted ensemble (WE) framework, to encourage the sampling of rare unbinding/rebinding events. WE accelerates the sampling of rare events using an ensemble of trajectories that are each assigned a statistical weight (Huber and Kim, 1996). The ensemble is integrated forward in time in a parallel fashion, and periodically “resampled” by cloning certain trajectories and merging others. When a trajectory is cloned, its weight is divided amongst the clones, but the multiple copies of the trajectory go on to evolve independently. By repeatedly cloning trajectories that are in undersampled regions of space we can obtain statistics on very long-timescale events using only short-timescale simulations.

The REVO resampling method (Resampling Ensembles by Variance Optimization) was designed to efficiently perform cloning and merging operations on small ensembles of trajectories that are evolving in high-dimensional spaces (Donyapour et al., 2019). This is valuable in situations where it is difficult to define one or two progress variables that capture the long-timescale events of interest. In REVO, coupled cloning and merging operations are proposed (e.g., clone trajectory i , and merge trajectories j and k) and are accepted or rejected based on an objective function called the “trajectory variation”:

$$V = \sum_i V_i = \sum_i \sum_j (d_{ij}/d_0)^\alpha \phi_i \phi_j \quad (3)$$

where d_{ij} is the distance between trajectories i and j , α and d_0 are parameters, and ϕ_x is a function that measures the importance, or “novelty” of a trajectory x , which in our work here is strictly a function of the weight of the trajectory: $\phi_i = \log w_i - C$, where w_i is the weight of trajectory i and C is a constant. Trajectories with the highest V_i values in Equation (3) are chosen for cloning, and those with the lowest V_i are chosen for merging. More information about the algorithm can be found in previous work (Donyapour et al., 2019).

We run separate simulations for the binding and unbinding processes. In our unbinding simulations, the ligands start in the bound state and are terminated as they unbind. In the rebinding simulations, the ligands start in the unbound state and are terminated as they bind. The distance function (d_{ij}) we use in Equation (3) is different for these two simulation types. For the unbinding simulations, we superimpose the hosts from trajectories i and j , and then compute the root mean squared distance (RMSD) between the guest molecules, without any further alignment (Dickson and Lotz, 2016, 2017). As there is 4-fold symmetry in this system, we perform the alignment four times (once for each symmetrically-equivalent mapping) and use the smallest such distance as d_{ij} . For the rebinding simulations, we calculate the distance to the native state for each trajectory



($d_{\text{native}}(\mathbf{X}_i)$), which again takes into account the four symmetry mappings, using the lowest such distance. The distance between trajectories i and j is then calculated as $d_{ij} = |1/d_{\text{native}}(\mathbf{X}_i) - 1/d_{\text{native}}(\mathbf{X}_j)|$, where the inverse is used to prioritize differences between small values of d_{native} .

2.4. Calculating Rates by Ensemble Splitting

A major advantage of the REVO method, much like other weighted ensemble methods, is that it can calculate kinetic parameters in real time as the simulation progresses. This is achieved by running separate simulations for the binding and unbinding processes, and in each case, measuring the trajectory flux into the opposite basin (Dickson et al., 2009, 2011; VandenEijnden and Venturoli, 2009; Costaouec et al., 2013; Suárez et al., 2014). The unbound basin is defined as the set of structures where the closest host-guest interatomic distance is > 1 nm, following previous work (Dickson and Lotz, 2016, 2017; Lotz and Dickson, 2018). The bound basin is defined as the set of structures where the guest RMSD (compared to the native structure) is < 0.1 nm after aligning to the host. Again, this RMSD measurement takes into account the four symmetry-equivalent mappings of OA.

In our studies, the binding and rebinding REVO simulations are conducted separately. However, the methodology of obtaining on and off rates is essentially the same. After each dynamics step, if a walker has entered the opposite basin, as described above, its weight is recorded and its structure is “warped” back to the starting structure at the beginning of the simulation. The atomic coordinates are set to the starting structure and the velocities are reinitialized; however, the weight of the trajectory remains the same. Before the warping event to the starting structure, the structure of the walker is recorded and is referred to as an “exit point.” In our unbinding simulations, the initial starting structure is the initial bound pose provided. In our rebinding simulations, the initial starting structure is chosen from a set of exit points generated from the unbinding simulations. Therefore, the unbinding analyses were performed prior to initialization and the subsequent running of our rebinding simulations.

The off and on rates are calculated by using the flux of trajectories into either the unbound or bound state, respectively.

$$k_{\text{off}}(t) = \frac{\sum_i w_i}{t} \quad (4)$$

$$k_{\text{on}}(t) = \frac{\sum_i w_i}{Ct} \quad (5)$$

where the sum is over the set of “warped” trajectories, t is the elapsed simulation time, and C is the concentration of ligand, computed as $1/V$ where V is the box volume. The box volume was approximately 80.2 nm^3 , corresponding to a concentration of ligand of 0.0207 M .

There are a few key differences between the REVO simulations discussed here and our previous studies (Dixon et al., 2018). For both the unbinding and rebinding simulations in this study, the total simulation time is 2.25 times longer compared to our previous study, as our current unbinding and rebinding simulations were run for 4,500 and 450 cycles, respectively. Additionally, ten independent unbinding simulations were run for each of the four friction coefficients, whereas our previous study only ran five independent simulations for each starting pose. However, only five independent rebinding simulations were run for each of the coefficients, as we observe much less variation in the k_{on} estimates. Finally, 48 walkers were used in both studies and the time per cycle is consistent, where the unbinding simulations are 20 ps/cycle and the rebinding simulations are 200 ps/cycle .

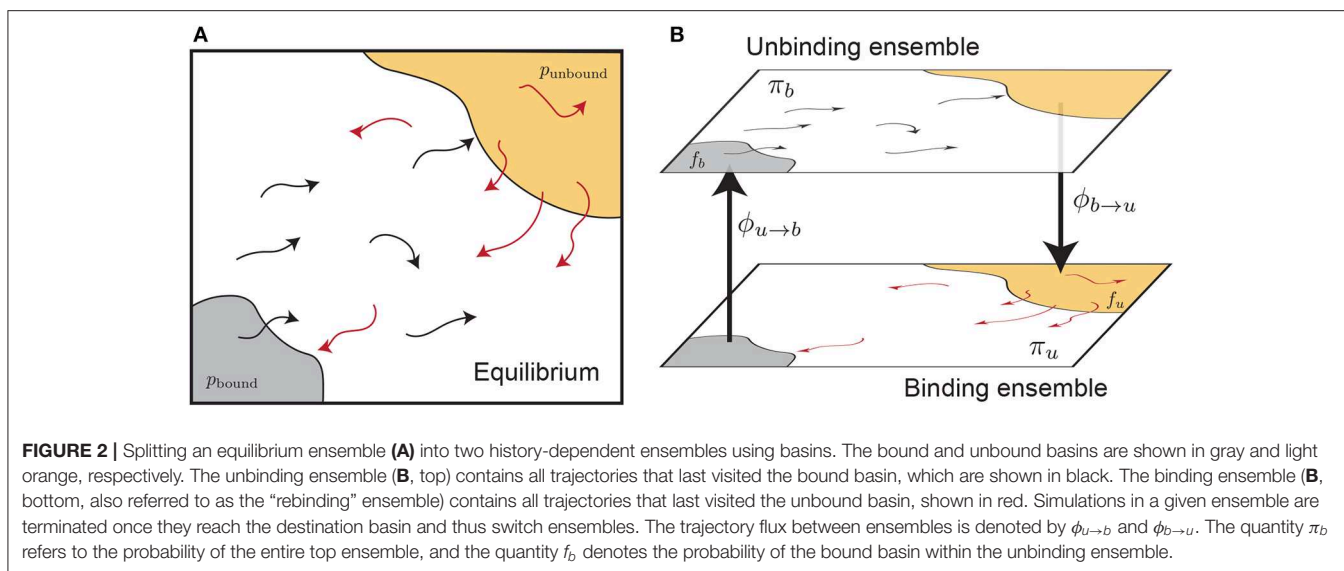
2.5. Calculating Electrostatic Interaction Energies

The electrostatic energy between the host and guest molecules for use in the second correction term was calculated as: $E_{\text{int}} = \frac{1}{4\pi\epsilon_w} \sum_i \frac{Q_i Q_j}{r_{ij}}$ where Q_a is the partial charge of atom a used in the force field during simulation. r_{ij} is the interatomic distance between atoms i and j calculated by using the minimum image convention. $\epsilon_w = 6.88 \times 10^{-10} \text{ F/m}$ is the permittivity of water at 300 K calculated by linear interpolation of the water dielectric constant at 298.15 and 303.15 K (Archer and Wang, 1990).

3. RESULTS

3.1. Derivation of Correction Terms

The binding free energy can be calculated using the rate constants k_{on} and k_{off} as $\Delta G = G_{\text{bound}} - G_{\text{unbound}} = -kT \ln K_{\text{eq}} C_0 = -kT \ln \frac{C_0 k_{\text{on}}}{k_{\text{off}}}$, where K_{eq} is the binding equilibrium constant, C_0 is the reference concentration of 1 mol/L , k is Boltzmann’s constant and T is the temperature in Kelvin. While this relationship is correct in the macroscopic limit, it fails to account for the box



size and the volume of the unbound state in finite simulation environments with periodic boundary conditions. Here we derive a more accurate expression for the binding free energy that accounts for the finite box size in a typical MD simulation.

Our starting point is an expression for K_{eq} , which is valid for a dilute solution in thermodynamic equilibrium. We use the notation of Woo and Roux (see Equation 4 from Woo and Roux, 2005):

$$K_{eq} = \frac{\int_{\text{bound}} d\mathbf{l} \int d\mathbf{X} e^{-\beta U}}{\int_{\text{bulk}} d\mathbf{l} \delta(\mathbf{r}_1 - \mathbf{r}_1^*) \int d\mathbf{X} e^{-\beta U}} \quad (6)$$

where U is the internal energy of the system, $\beta = 1/kT$ is the inverse temperature, \mathbf{r}_1 is the center of mass of the ligand (referred to as a “guest” molecule) and \mathbf{r}_1^* is an arbitrary position of the guest in the bulk. Note that $d\mathbf{l}$ integrates over the guest positions, and $d\mathbf{X}$ integrates over everything else: the host and the solvent degrees of freedom. Note also that K_{eq} has units of volume, as the delta function constraining the center of mass in the denominator removes three spatial degrees of freedom.

Here we examine the calculation of free energies using rates determined from split ensemble calculations (**Figure 2**, see section 2.4 for more details). We denote the probability of these two ensembles as π_b and π_u , where $\pi_b + \pi_u = 1$, and:

$$\frac{\pi_b}{\pi_u} = \frac{\phi_{u \rightarrow b}}{\phi_{b \rightarrow u}} \quad (7)$$

where $\phi_{a \rightarrow b}$ is the time-averaged flux from the a ensemble to the b ensemble (i.e., across the dotted lines in **Figure 2**). The equilibrium probability of a position \mathbf{X} can be obtained by combining estimates from both ensembles:

$$p(\mathbf{X}) = p_u(\mathbf{X})\pi_u + p_b(\mathbf{X})\pi_b \quad (8)$$

where $p_a(\mathbf{X})$ is the probability of conformation \mathbf{X} in ensemble a , which is normalized such that $\int p_a(\mathbf{X}) d\mathbf{X} = 1$.

Let us define the bound state as the domain of the integral in the numerator of Equation (6), and the unbound state as a set of structures considered unbound in simulation (not the same as the bulk state in Equation 6). These states are shown as shaded regions in **Figure 2**. The ratio of the probabilities of these two states, at equilibrium, is given by:

$$\frac{p_{\text{bound}}}{p_{\text{unbound}}} = \frac{\int_{\text{bound}} d\mathbf{l} \int d\mathbf{X} e^{-\beta U}}{\int_{\text{unbound}} d\mathbf{l} \int d\mathbf{X} e^{-\beta U}} \quad (9)$$

which can also be calculated in our ensemble splitting simulations:

$$\frac{p_{\text{bound}}}{p_{\text{unbound}}} = \frac{\pi_b \int_{\text{bound}} p_b(\mathbf{X}) d\mathbf{X}}{\pi_u \int_{\text{unbound}} p_u(\mathbf{X}) d\mathbf{X}} = \frac{\pi_b f_b}{\pi_u f_u} \quad (10)$$

where f_a is the probability of the basin state within ensemble a .

Expanding Equation (6) we have:

$$\begin{aligned} K_{eq} &= \frac{\int_{\text{bound}} d\mathbf{l} \int d\mathbf{X} e^{-\beta U}}{\int_{\text{unbound}} d\mathbf{l} \int d\mathbf{X} e^{-\beta U}} \frac{\int_{\text{unbound}} d\mathbf{l} \int d\mathbf{X} e^{-\beta U}}{\int_{\text{bulk}} d\mathbf{l} \delta(\mathbf{r}_1 - \mathbf{r}_1^*) \int d\mathbf{X} e^{-\beta U}} \\ &= \frac{\pi_b f_b}{\pi_u f_u} \frac{\int_{\text{unbound}} d\mathbf{l} \int d\mathbf{X} e^{-\beta U}}{\int_{\text{bulk}} d\mathbf{l} \delta(\mathbf{r}_1 - \mathbf{r}_1^*) \int d\mathbf{X} e^{-\beta U}}. \end{aligned} \quad (11)$$

The unbound state in simulation is far enough that the host and guest do not interact directly through van der Waals interactions, although if both molecules carry an explicit charge—as in the example considered here—there could still be significant host-guest electrostatic interactions. To account for these, we introduce another intermediate state with an altered energy function (U^*) which is the same as U except that it does not

include electrostatic interactions between the host and the guest:

$$K_{eq} = \frac{\pi_b f_b \int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U}}{\pi_u f_u \int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U^*}} \frac{\int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U^*}}{\int_{\text{bulk}} d\mathbf{1} \delta(\mathbf{r}_1 - \mathbf{r}_1^*) \int d\mathbf{X} e^{-\beta U}} \quad (12)$$

$$= \frac{\pi_b f_b}{\pi_u f_u} \langle e^{\beta E_{\text{int}}} \rangle_{\text{unb}}^{-1} \frac{\int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U^*}}{\int_{\text{bulk}} d\mathbf{1} \delta(\mathbf{r}_1 - \mathbf{r}_1^*) \int d\mathbf{X} e^{-\beta U}} \quad (13)$$

where $E_{\text{int}} = U - U^*$ and the subscript “unb” indicates an ensemble average over structures in the unbound state obtained with the normal energy function U . Note the final step used the relation:

$$\frac{\int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U^*}}{\int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U}} = \frac{\int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{\beta E_{\text{int}}} e^{-\beta U}}{\int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U}} = \langle e^{\beta E_{\text{int}}} \rangle_{\text{unb}}. \quad (14)$$

We can now reasonably assume that the guest in the unbound state is non-interacting with the host. This allows us to write $e^{-\beta U}$ as $e^{-\beta U_G} e^{-\beta U_{HS}}$, where U_G are the terms in the energy function that depend only on the coordinates of the guest, and U_{HS} are terms that only depend on the host and the solvent. We can then pull the integral $\int d\mathbf{X} e^{-\beta U_{HS}}$ out of the numerator and denominator of the last term of Equation (11):

$$\frac{\int_{\text{unbound}} d\mathbf{1} \int d\mathbf{X} e^{-\beta U^*}}{\int_{\text{bulk}} d\mathbf{1} \delta(\mathbf{r}_1 - \mathbf{r}_1^*) \int d\mathbf{X} e^{-\beta U}} = \frac{\int_{\text{unbound}} d\mathbf{1} e^{-\beta U_G}}{\int_{\text{bulk}} d\mathbf{1} \delta(\mathbf{r}_1 - \mathbf{r}_1^*) e^{-\beta U_G}}. \quad (15)$$

The bottom integral has the center of mass of the ligand fixed and is only over internal and rotational degrees of freedom of the ligand. This can also be separated and removed from the numerator, which simplifies the ratio to be the volume of the unbound state, defined as:

$$V_{\text{unbound}} = \frac{\int_{\text{unbound}} d\mathbf{1} e^{-\beta U_G}}{\int_{\text{guest}} d\mathbf{G}_1 e^{-\beta U_G}} = \int_{\text{box}} d\mathbf{R} \phi_u(\mathbf{R}) \quad (16)$$

where we use \mathbf{G}_1 to denote the internal and rotational degrees of freedom of the guest that remain after specification of \mathbf{r}_1 . The quantity $\phi_u(\mathbf{R})$ is the fraction of conformers with center of mass \mathbf{R} that satisfy the unbound boundary conditions: here, that the guest atoms are all farther than a cutoff distance of 1 nm away from the host. This integral can be calculated by Monte Carlo, where a center of mass position and orientation of the ligand is randomly generated, and the number of successful unbound conformers is recorded:

$$V_{\text{unbound}} = V_{\text{box}} \frac{N_{\text{unbound}}}{N_{\text{trials}}}. \quad (17)$$

Note that for large boxes $V_{\text{unbound}} \approx V_{\text{box}}$.

Putting this all together we have:

$$K_{eq} = \frac{\pi_b f_b}{\pi_u f_u} \langle e^{\beta E_{\text{int}}} \rangle_{\text{unb}}^{-1} V_{\text{unbound}}, \quad (18)$$

which differs from the straightforward interpretation used in our previous work (Dixon et al., 2018):

$$K_{eq}^0 = \frac{\pi_b}{\pi_u [L]} = \frac{\pi_b}{\pi_u} V_{\text{box}} \quad (19)$$

Using $\Delta G = -kT \ln(K_{eq} C_0)$, we have:

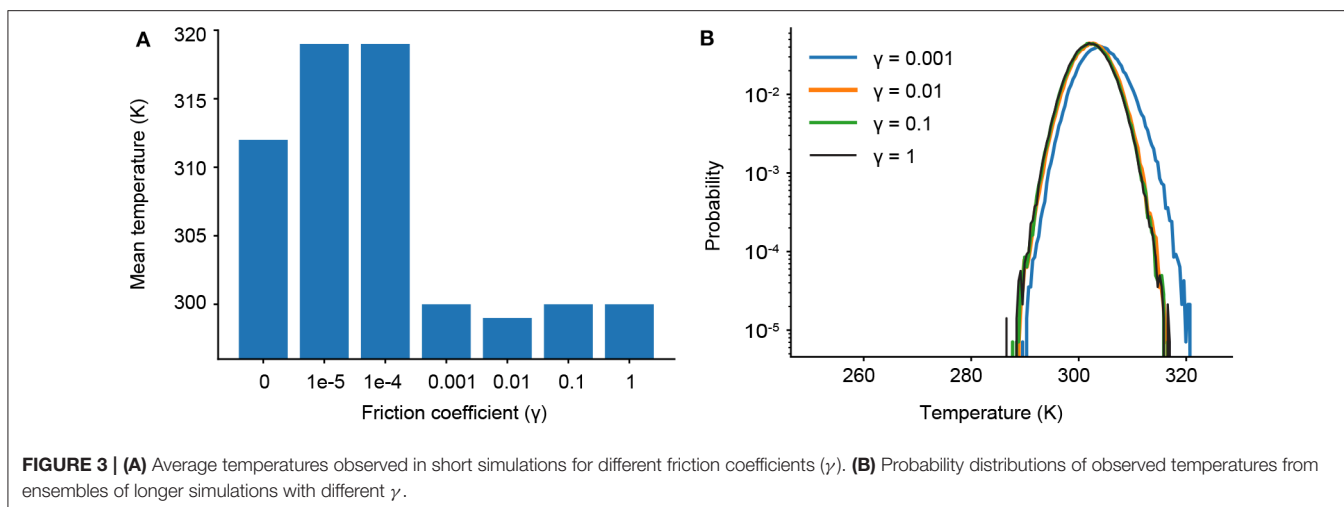
$$\Delta G = \Delta G^0 - kT \ln\left(\frac{f_b}{f_u}\right) + kT \ln \langle e^{\beta E_{\text{int}}} \rangle_{\text{unb}} - kT \ln\left(\frac{V_{\text{unbound}}}{V_{\text{box}}}\right) \quad (20)$$

which explicitly shows ΔG as the sum of $\Delta G^0 = -kT \ln(K_{eq}^0 C_0)$ and the three newly derived correction terms. The first term will go to zero in the limit that the basin states are chosen to represent the vast majority of the probability in both the binding and unbinding ensembles. In other words, this term goes to zero when both f_b and f_u approach one. The second term is likely to only be non-negligible in the case of explicitly charged host and guest molecules and regardless would go to zero as the definition of the unbound state is moved to farther and farther distances. The third term would also go to zero for large simulation boxes, but in practice this is often not feasible due to computational constraints. Consequently, $V_{\text{unbound}}/V_{\text{box}}$ could be much less than one, introducing a correction in the positive direction. Below we calculate these three correction terms and apply them to free energy calculations.

3.2. Extended Trajectory Ensembles With Lower Friction Coefficients

In previous work, we used a Langevin integrator with a value of $\gamma = 1 \text{ ps}^{-1}$ for the friction coefficient. As the simulations already have explicit solvent, this adds extra friction into the system that is not physical. Here we investigate whether reducing γ to values less than one will significantly affect our rate calculations. We thus run a set of trajectory ensembles at multiple values of γ and extend each ensemble to be larger and longer than those published in our prior study (Dixon et al., 2018) to more fully examine questions of convergence.

As γ governs the coupling to the Langevin thermostat, we determine the minimum value of γ where our target temperature (300 K) is maintained. We first ran a series of short simulations (one 10 ns trajectory for each γ) and find that temperature control is completely lost for friction coefficients less than $\gamma = 0.001$ (Figure 3A). We then ran longer simulations for $\gamma = 1, 0.1, 0.01$, and 0.001 , examining not only the mean temperature, but the probability of significant temperature fluctuations, which could spur anomalous results in our ligand dissociation simulations. Figure 3B shows the probability distribution of observed temperatures over an ensemble of 240 trajectories run for 90 ns each. For $\gamma = 0.01, 0.1$ and 1 ps^{-1} , the temperature distribution is normally distributed around the mean (300 K) as seen by the parabolic curves on a log scale. Temperature control is not fully maintained for $\gamma = 0.001 \text{ ps}^{-1}$, as shown by a rightward shift and slight widening of the parabolic distribution. We thus



restrict our analysis to three values of the friction coefficient: $\gamma = 0.01, 0.1, \text{ and } 1 \text{ ps}^{-1}$.

We run both unbinding and rebinding REVO simulations for the OAG6 system. For unbinding, we ran 10 simulations for each of the three friction coefficients; for rebinding, we ran five simulations for each coefficient, yielding a total of 30 simulations for unbinding and 15 simulations for rebinding. A set of binding and unbinding simulations were also run for $\gamma = 0.001$ —despite the impaired temperature control—which are reported in the **Supplemental Information**. The estimates for the unbinding and binding fluxes are depicted in **Figure 4**, where each curve represents an individual REVO simulation. The averages, illustrated with a bolded line, are calculated by averaging the trajectory flux over the entire set of simulations for that value of γ . The upward jumps on these plots indicate that an exit point was recorded that has a higher weight than was previously observed.

By reducing γ to values < 1 , we observed no change in the binding rates, and small changes to the unbinding rates which are on the border of significance. With regard to unbinding rates, the two largest friction coefficients yielded the smallest error and similar k_{off} values, where $\gamma = 1$ yielded an average off rate of 16.4 s^{-1} and $\gamma = 0.1$ yielded an off rate of 11.5 s^{-1} . The off-rate increased by 10-fold for $\gamma = 0.01$, although this is mostly driven by exit points observed in a single simulation. In our previous OA-G6 results using $\gamma = 1$, we calculated an unbinding rate of 0.48 s^{-1} which slightly differs from the value calculated in this study using $\gamma = 1$ (**Table 1**). Unbinding rates for $\gamma = 0.001 \text{ ps}^{-1}$ were approximately 1000-fold higher, although these are known to be affected by a higher average temperature (**Supplemental Information**). Taking a closer look at the binding rates, we saw no discernible difference across the friction coefficients. The binding rate was approximately $10^9 \text{ s}^{-1} \text{ M}^{-1}$, for all friction coefficients, which was about 5-fold larger when compared to our previous study using $\gamma = 1$. For both binding and unbinding rates we have more confidence in the results obtained here, as they are based on more extensive simulation data.

TABLE 1 | Binding and unbinding rates as a function of friction coefficient (γ).

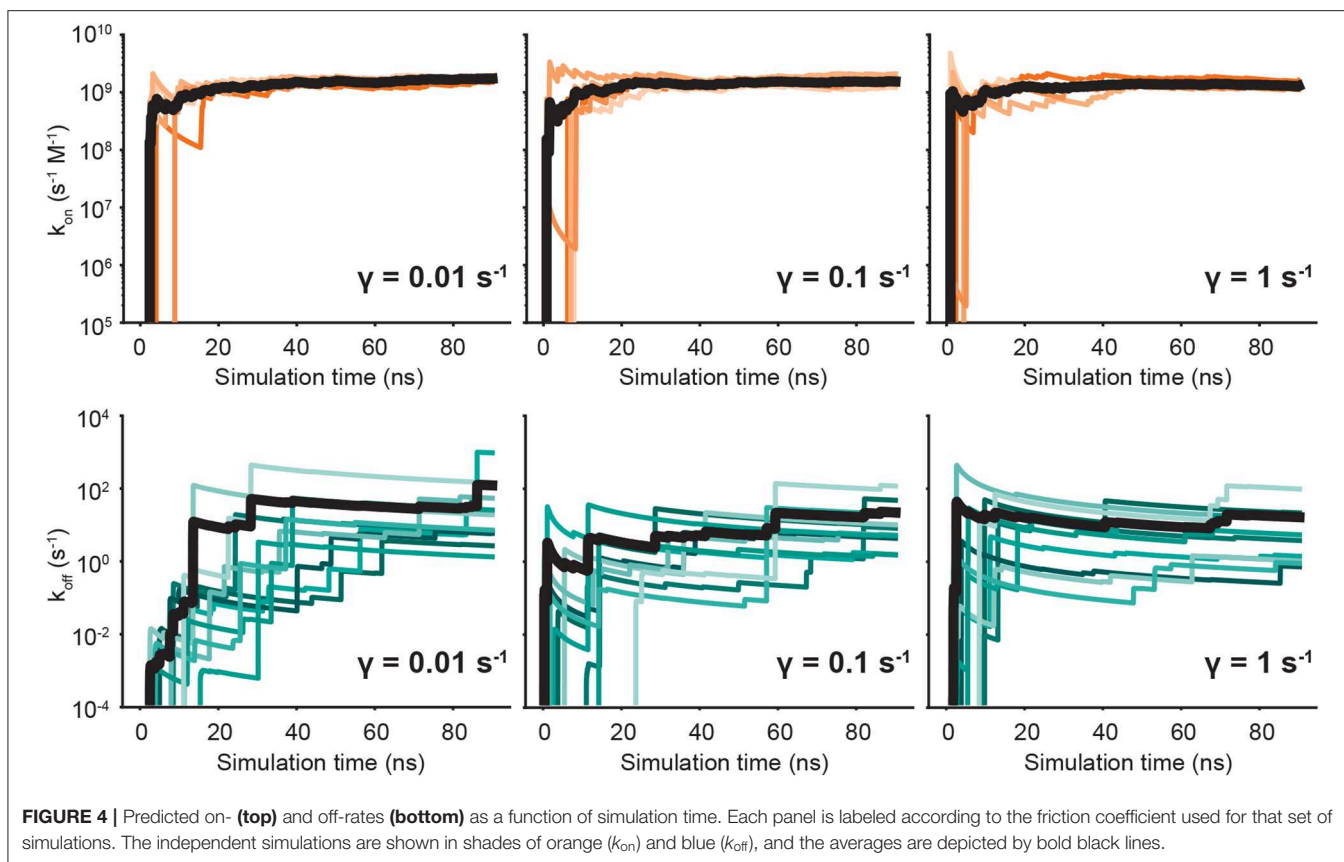
	$k_{\text{on}} (10^8 \text{ M}^{-1} \text{ s}^{-1})$	$k_{\text{off}} (\text{s}^{-1})$
$\gamma = 0.01$	17 ± 1	122 ± 94
$\gamma = 0.1$	16 ± 2	22 ± 12
$\gamma = 1$	13 ± 1	16.4 ± 9.4
Dixon et al. (2018) ($\gamma = 1$)	2.8 ± 1.0	0.48 ± 0.11

The uncertainties shown use the standard error of the mean calculated from 5 and 10 independent REVO runs for binding and unbinding, respectively. The quantities from Dixon et al. (2018) were obtained with 5 REVO runs that used different initial conformations, each of which were 2,000 cycles in length.

For both the unbinding and rebinding simulations, across all friction coefficients, we observed at least 1,000 warping events (**Figure S4**). As expected, we observe that rebinding occurs with a much higher probability when compared to unbinding, by several orders of magnitude. The unbinding walker weights are limited at the low end by the minimum walker probability (p_{min}), which is set to 10^{-12} . The rebinding walker weights are limited at the high end by the maximum walker probability (p_{max}), which is set to 10^{-1} , respectively. **Figure S4** shows that the 10-fold larger unbinding rate for $\gamma = 0.01$ was largely due to a single unbinding point in a single simulation, which underscores the sensitivity and uncertainty of rate calculations using trajectory fluxes. **Figure S2** shows unbinding fluxes for $\gamma = 0.001$, which is known to have elevated temperatures. There we see a large number of high-weight unbinding events in two different simulations, leading to the 1,000-fold increase in k_{off} .

3.3. Free Energy Estimates, Correction Terms, and Comparison With Previous Benchmarks

As the friction coefficient unevenly affected the rates of binding and unbinding, there was a net effect on the binding free energies. As shown in **Figure 5** and **Table 2**, the binding free energy increases as the friction coefficient is lowered, independent of the



free energy correction terms derived in section 3.1. **Table 2** shows the free energies computed using the averaged fluxes across all simulations at each γ value. For all friction coefficients, the calculated free energy was always higher than that from our previous study (-12.1 kcal/mol; red line), even for $\gamma = 1$, signifying that extending the simulation time aided in predicting experimentally determined binding free energies.

The correction terms are calculated using data obtained from the simulations, but they are mostly functions of geometric properties of the simulation box and boundary conditions, and are not expected to change as a function of γ . The first term, $-kT \ln f_b/f_u$, was calculated to be 0.74 ± 0.10 kcal/mol, with f_b and f_u taking on values of 0.157 and 0.54, respectively. As described in section 3.1, f_b is the probability of the being in the bound basin given that you are in the unbinding ensemble, which is calculated using the sum of the weights of trajectories in the bound basin, divided by the total sum of the weights of the trajectories considered. The f_b value in particular was lower than expected, indicating that our definition of the bound state might be too restrictive, even though we did account for all symmetry-equivalent conformations in our calculation of f_b .

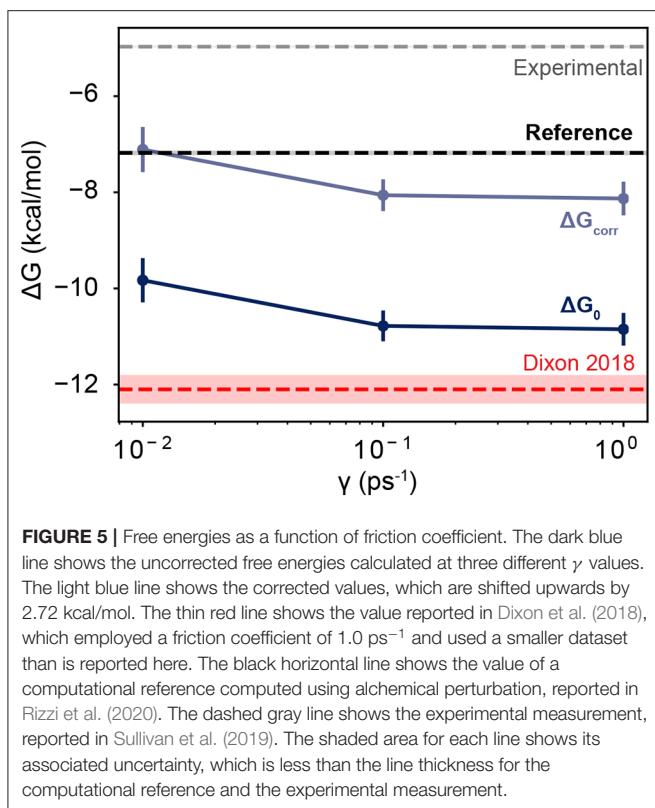
The second term, $+kT \ln \langle e^{\beta E_{int}} \rangle_{unb}$, was calculated to be 1.64 ± 0.002 kcal/mol. This was calculated by determining the electrostatic interaction energies (see section 2.5) for the set of unbound states observed in the rebinding simulations. The expectation value in the correction term again accounted for trajectory weights and was computed using 71,428 interaction

energy measurements that were selected from the unbound ensemble. The uncertainty was computed as the standard error of the mean of this set of energies. To calculate the third correction term, $-kT \ln \left(\frac{V_{unbound}}{V_{box}} \right)$, we directly estimated $V_{unbound}/V_{box}$ using the Monte Carlo procedure described in section 3.1. The ratio was computed as 0.56 ± 0.0037 using five batches of 10,000 trials each, where the uncertainty is the standard error of the mean across the sets of trials.

Together these three terms sum to 2.72 kcal/mol, which is a significant correction to the binding free energies computed here. Over half of this comes from the residual electrostatic interaction energy between the host and the guest. Note that both the host and the guest have negative charges, and the residual interaction between the two molecules is repulsive. Turning this interaction off releases 1.64 kcal/mol of energy, which lowers the free energy gap between the bound and unbound states. The corrected and uncorrected free energies are shown as a function of γ in **Figure 5**. For $\gamma \geq 0.01$ the calculated free energies are almost equal to within standard error and the correction terms significantly reduce the error with respect to the computational reference value (Rizzi et al., 2018, 2020).

4. DISCUSSION AND CONCLUSION

In this study, we sought to better connect the calculation of binding and unbinding rates with the calculation of binding



free energies. The rate calculations measured the microscopic fluxes of trajectories from one basin to another. These fluxes can be visualized in an extended history-dependent conformation space, where trajectories change their “color” based on which basin (“bound” or “unbound”) they have most recently visited (Dickson et al., 2009, 2011; Vanden-Eijnden and Venturoli, 2009; Costaouec et al., 2013; Suárez et al., 2014). The ratio of these rates gives a ratio of two populations: the trajectories that have most recently visited the “bound” basin and the trajectories that have most recently visited the “unbound” basin. The first correction term adjusts this ratio to instead only account for the probability contained within the basins themselves and is particular to rates that are calculated using this history-dependent formalism. The third term can be seen as a volume correction term, which is used to accurately account for the volume in the unbound state. This is done in other approaches where restraints are used, such as methods based on calculation of the potential of mean force (Deng and Roux, 2009). In our case the unbound state cannot be easily approximated by a geometric object, such as the volume of a spherical shell.

The second term accounts for residual interactions in the unbound ensemble. This could be used by other approaches that directly determine free energy differences between bound and unbound conformations, such as Markov state modeling, metadynamics, milestoning, and umbrella sampling. The conventional approach is to define a simulation box that is large enough such that the interactions between the host and guest are negligible in the unbound state. However, this can significantly increase the cost of the simulation. It is worth

TABLE 2 | Raw (ΔG^0) and corrected (ΔG_{corr}) free energy values using simulation data from three different friction coefficients.

	ΔG^0 (kcal/mol)	ΔG_{corr} (kcal/mol)
$\gamma = 0.01$	-9.83 ± 0.46	-7.11 ± 0.47
$\gamma = 0.1$	-10.78 ± 0.32	-8.06 ± 0.33
$\gamma = 1$	-10.85 ± 0.34	-8.13 ± 0.36
Dixon et al. (2018) ($\gamma = 1$)	-12.1 ± 1.0	-9.38 ± 1.0
Comp. (Rizzi et al., 2020)	–	-7.0 ± 0.1
Exp. (Sullivan et al., 2019)	–	-4.97 ± 0.02

Values are in kcal/mol and uncertainties are calculated using propagation of the standard error of the mean.

noting that umbrella sampling results for this system (OA-G6) obtained by Song et al. (2018), -8.50 kcal/mol, were also below both the computational benchmark and the experimental value. Their unbound state was defined as a 20 Å distance between an atom in the guest and a dummy atom in the center of the host, which is roughly comparable to our unbound basin of 10 Å of clearance between the host and the guest. Assuming a similar value for the electrostatic correction term, it would have brought their prediction to -6.86 kcal/mol, which is in line with the computational benchmarks (Rizzi et al., 2020).

The electrostatic term can also be viewed as a sort of “decoupling” between the host and the guest, and it is warranted to discuss similarities and differences with similar procedures in alchemical free energy methods. They are similar in that we are computing a free energy between two Hamiltonians, one in which an interaction is turned off. We could thus use similar techniques for computing these free energy differences, such as thermodynamic integration (Kirkwood, 1935; Bhati et al., 2019), BAR (Gutiérrez et al., 2019), MBAR (Shirts and Chodera, 2008; Bhati et al., 2019), or MM/PBSA (Rifai et al., 2019), although here we effectively use a simple free energy perturbation (FEP) expression (Zwanzig, 1954; Jorgensen and Thomas, 2008). The approaches are different in that we are only considering ensembles of structures where the interactions being turned off are relatively weak. We are assuming here—as is always the case with FEP—that the conformational ensembles of both the host and the guest are highly overlapping between the two Hamiltonians, which considerably simplifies the problem. We also note that although we employ electrostatic decoupling to compute free energies, our simulations still reveal important information about the (un)binding kinetics and mechanism.

Given these correction terms for the binding affinity, it is reasonable to ask if and how the rate constants should be modified. The correction terms each have the effect of “loosening” the interaction, indicating that either the corrected off-rate should increase, or the corrected on-rate should decrease. It is reasonable to assume that lowering the on-rate should account for the vast majority of this correction, as the on-rate measured starts from an unbound conformation that is much closer (a clearance of 1 nm between the host and guest) than is likely in experimental conditions. More accurate calculations of the binding rate can be achieved with better sampling of the unbound state, for instance using the Northrup-Allison-McCammon method (Northrup et al., 1984). It would be

interesting to see whether such calculations can recapitulate part of the free energy differences observed here.

We also examined the role that the Langevin integrator plays in the prediction of kinetic and thermodynamic quantities. In particular, we adjusted the friction coefficient (γ), defined in the Langevin integrator, while maintaining the stability of temperature at 300 K. We did not expect that altering the friction coefficient would have an impact on the calculation of equilibrium quantities. As γ does not appear in the Hamiltonian of the system, it should not affect the probability of a given microstate $P(\mathbf{X})$, which is given by the Canonical probability density $\exp(-\beta U(\mathbf{X}))$. While we did expect it to affect rates, we expected that these effects would offset: that if unbinding was accelerated 10-fold, we would observe the binding process to be sped up by the same factor. However, we observe that the on-rate was very stable as a function of γ , while the off-rate changed slightly. One explanation is that unbinding is a much more rare event when compared to rebinding, and estimates of k_{off} were not converged. Lower friction coefficients could be accelerating sampling of these events and making it easier to observe higher probability walkers unbind in our simulations.

Convergence is of utmost priority in weighted ensemble simulations that calculate kinetic quantities. In our previous study, we hypothesized that it was possible that extending the time of the unbinding simulations could capture more high weight walkers exiting from the bound state. Indeed, we observe a higher unbinding flux in this study across all friction coefficients. In **Figure 4**, we observe large upward jumps, for all γ values, even after 40 ns of simulation time per walker, which was sampling limit in our previous study. These upward jumps, as previously described, signify that an exit point was recorded that has a higher weight than previously observed. This highlights the challenges involved in accurate determination of rate fluxes for rare events. It is worth noting that by using our correction terms to account for small unbound volumes and persistent but small electrostatic interactions in the unbound state, we can keep box sizes small, allowing for better convergence of rate fluxes at fixed computational cost.

REFERENCES

- Archer, D., and Wang, P. (1990). The dielectric constant of water and debye-hückel limiting law slopes. *J. Phys. Chem. Refer. Data* 19, 371–411. doi: 10.1063/1.555853
- Ayaz, P., Andres, D., Kwiatkowski, D. A., Kolbe, C. C., Lienau, P., Siemeister, G., et al. (2016). Conformational adaption may explain the slow dissociation kinetics of rooniclib (BAY 1000394), a type I CDK inhibitor with kinetic selectivity for CDK2 and CDK9. *ACS Chem. Biol.* 11, 1710–1719. doi: 10.1021/acschembio.6b00074
- Bernetti, M., Cavalli, A., and Mollica, L. (2017). Protein-Ligand (un)binding kinetics as a new paradigm for drug discovery at the crossroad between experiments and modelling. *Med. Chem. Commun.* 8, 534–550. doi: 10.1039/C6MD00581K
- Bhati, A. P., Wan, S., and Coveney, P. V. (2019). Ensemble-based replica exchange alchemical free energy methods: the effect of protein mutations on inhibitor binding. *J. Chem. Theory Comput.* 15, 1265–1277. doi: 10.1021/acs.jctc.8b01118

Of course the binding free energy alone is still an important quantity for drug design (Homeyer et al., 2014). If one is only interested in the absolute binding free energy, calculating it through the ratio of rates is needlessly complicated; free energy is a state function and thus only depends on the endpoints of the binding pathway. The prediction of k_{off} and k_{on} themselves is challenging, since they are not state functions: they depend on the transition path ensemble between the bound and unbound state. Sampling of these physical pathways is a large challenge for molecular dynamics, largely due to the long timescales of the binding and release processes. Ensuring that the ratio of rates is consistent with binding free energy calculations—as done here—provides an additional, powerful consistency check. In particular, comparing to well-converged computational benchmarks is more useful than experimental quantities, as we avoid an additional layer of uncertainty associated with the force field used to describe the system.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

AD designed the project. RH, TD, and AD conducted the research, analysis, and wrote the manuscript.

FUNDING

This work was funded by the National Institutes of Health (R01GM130794) and the National Science Foundation (DMS 1761320).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2020.00106/full#supplementary-material>

- Bruce, N. J., Ganotra, G. K., Kokh, D. B., Sadiq, S. K., and Wade, R. C. (2018). New approaches for computing ligand-receptor binding kinetics. *Curr. Opin. Struct. Biol.* 49, 1–10. doi: 10.1016/j.sbi.2017.10.001
- Bruno, A., Barresi, E., Simola, N., Da Pozzo, E., Costa, B., Novellino, E., et al. (2019). Unbinding of translocator protein 18 KDa (TSPO) ligands: from *in vitro* residence time to *in vivo* efficacy via *in silico* simulations. *ACS Chem. Neurosci.* 10, 3805–3814. doi: 10.1021/acschemneuro.9b00300
- Camilloni, C., and Pietrucci, F. (2018). Advanced simulation techniques for the thermodynamic and kinetic characterization of biological systems. *Adv. Phys.* 2018:1477531. doi: 10.1080/23746149.2018.1477531
- Carroll, M. J., Mauldin, R. V., Gromova, A. V., Singleton, S. F., Edward, J., and Lee, A. L. (2012). Evidence for dynamics in proteins as a mechanism for ligand dissociation. *Nat. Chem. Biol.* 8, 246–252. doi: 10.1038/nchembio.769
- Casasnovas, R., Limongelli, V., Tiwary, P., Carloni, P., and Parrinello, M. (2017). Unbinding kinetics of a P38 MAP kinase type II inhibitor from metadynamics simulations. *J. Am. Chem. Soc.* 139, 4780–4788. doi: 10.1021/jacs.6b12950
- Copeland, R. A. (2016). The drug-target residence time model: a 10-year retrospective. *Nat. Rev. Drug Discov.* 15, 87–95. doi: 10.1038/nrd.2015.18

- Costa, B., Da Pozzo, E., Giacomelli, C., Barresi, E., Taliani, S., Da Settimo, F., et al. (2016). TSPO ligand residence time: a new parameter to predict compound neurosteroidogenic efficacy. *Sci. Rep.* 6:18164. doi: 10.1038/srep18164
- Costaouec, R., Feng, H., Izaguirre, J., and Darve, E. (2013). Analysis of the accelerated weighted ensemble methodology. *Discrete Cont. Dyn. Syst.* 2013, 171–181. doi: 10.3934/proc.2013.2013.171
- Croll, T. I., Sammito, M. D., Kryshchuk, A., and Read, R. J. (2019). Evaluation of template-based modeling in CASP13. *Proteins* 87, 1113–1127. doi: 10.1002/prot.25800
- Deb, I., and Frank, A. T. (2019). Accelerating rare dissociative processes in biomolecules using selectively scaled MD simulations. *J. Chem. Theory Comput.* 15, 5817–5828. doi: 10.1021/acs.jctc.9b00262
- Deng, Y., and Roux, B. (2009). Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem. B* 113, 2234–2246. doi: 10.1021/jp807701h
- Dickson, A. (2018). Mapping the ligand binding landscape. *Biophys. J.* 115, 1707–1719. doi: 10.1016/j.bpj.2018.09.021
- Dickson, A., and Lotz, S. (2017). Multiple ligand unbinding pathways and ligand-induced destabilization revealed by WExplore. *Biophys. J.* 112, 620–629. doi: 10.1016/j.bpj.2017.01.006
- Dickson, A., and Lotz, S. D. (2016). Ligand release pathways obtained with WExplore: residence times and mechanisms. *J. Phys. Chem. B* 120, 5377–5385. doi: 10.1021/acs.jpcc.6b04012
- Dickson, A., Maienschein-Cline, M., Tovo-Dwyer, A., Hammond, J. R., and Dinner, A. R. (2011). Flow-dependent unfolding and refolding of an RNA by nonequilibrium umbrella sampling. *J. Chem. Theory Comput.* 7, 2710–2720. doi: 10.1021/ct200371n
- Dickson, A., Tiwary, P., and Vashisth, H. (2017). Kinetics of ligand binding through advanced computational approaches: a review. *Curr. Top. Med. Chem.* 17, 2626–2641. doi: 10.2174/1568026617666170414142908
- Dickson, A., Warmflash, A., and Dinner, A. R. (2009). Separating forward and backward pathways in nonequilibrium umbrella sampling. *J. Chem. Phys.* 136:154104. doi: 10.1063/1.3244561
- Dixon, T., Lotz, S. D., and Dickson, A. (2018). Predicting ligand binding affinity using on and off-rates for the SAMPL6 SAMPLing challenge. *J. Comput. Aided Mol. Design* 32, 1001–1012. doi: 10.1007/s10822-018-0149-3
- Dixon, T., Uyar, A., Ferguson-miller, S., and Dickson, A. (2020). Membrane-mediated ligand unbinding of the PK-11195 ligand from the translocator protein (TSPO). *bioRxiv*. doi: 10.1101/2020.01.21.914127
- Donyapour, N., Roussey, N. M., and Dickson, A. (2019). REVO: resampling of ensembles by variation optimization. *J. Chem. Phys.* 150:244112. doi: 10.1063/1.5100521
- Dror, R. O., Pan, A. C., Arlow, D. H., Borhani, D. W., Maragakis, P., Shan, Y., et al. (2011). Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. U.S.A.* 108, 13118–13123. doi: 10.1073/pnas.1104614108
- Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., et al. (2017). OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* 13:e1005659. doi: 10.1371/journal.pcbi.1005659
- Faradjian, A. K., and Elber, R. (2004). Computing time scales from reaction coordinates by milestoneing. *J. Chem. Phys.* 120:10880. doi: 10.1063/1.1738640
- Gan, H., Benjamin, C. J., and Gibb, B. C. (2011). Nonmonotonic assembly of a deep-cavity cavitand. *J. Am. Chem. Soc.* 133, 4770–4773. doi: 10.1021/ja200633d
- Gilson, M. K., Given, J. A., Bush, B. L., and McCammon, J. A. (1997). The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* 72, 1047–1069. doi: 10.1016/S0006-3495(97)78756-3
- Gu, S., Silva, D. A., Meng, L., Yue, A., and Huang, X. (2014). Quantitatively characterizing the ligand binding mechanisms of choline binding protein using markov state model analysis. *PLoS Comput. Biol.* 10:e1003767. doi: 10.1371/journal.pcbi.1003767
- Guo, D., Heitman, L. H., and Ijzerman, A. P. (2016). The added value of assessing ligand-receptor binding kinetics in drug discovery. *ACS Med. Chem. Lett.* 7, 819–821. doi: 10.1021/acsmchemlett.6b00273
- Guo, D., Xia, L., van Veldhoven, J. P. D., Hazeu, M., Mocking, T., Brussee, J., et al. (2014). Binding kinetics of ZM241385 derivatives at the human adenosine A2A receptor. *ChemMedChem* 9, 752–761. doi: 10.1002/cmdc.2013.00474
- Gutiérrez, M., Vallejos, G. A., Cortés, M. P., and Bustos, C. (2019). Bennett acceptance ratio method to calculate the binding free energy of bace1 inhibitors: theoretical model and design of new ligands of the enzyme. *Chem. Biol. Drug Des.* 93, 1117–1128. doi: 10.1111/cbdd.13456
- Homeyer, N., Stoll, F., Hillisch, A., and Gohlke, H. (2014). Binding free energy calculations for lead optimization: assessment of their accuracy in an industrial drug design context. *J. Chem. Theory Comput.* 10, 3331–3344. doi: 10.1021/ct5000296
- Huber, G. G. A., and Kim, S. (1996). Weighted-ensemble brownian dynamics simulations for protein association reactions. *Biophys. J.* 70, 97–110. doi: 10.1016/S0006-3495(96)79552-8
- Jorgensen, W. L., and Thomas, L. L. (2008). Perspective on free-energy perturbation calculations for chemical equilibria. *J. Chem. Theory Comput.* 6, 869–876. doi: 10.1021/ct800011m
- Kirberger, S. E., Ycas, P. D., Johnson, J. A., Chen, C., Ciccone, M. F., Woo, R. W., et al. (2019). Selectivity, ligand deconstruction, and cellular activity analysis of a BPTF bromodomain inhibitor. *Organ. Biomol. Chem.* 17, 2020–2027. doi: 10.1039/C8OB02599A
- Kirkwood, J. G. (1935). Statistical mechanics of fluid mixtures. *J. Chem. Phys.* 3, 300–313. doi: 10.1063/1.1749657
- Kokh, D. B., Amaral, M., Bomke, J., Grädler, U., Musil, D., Buchstaller, H. P., et al. (2018). Estimation of drug-target residence times by τ -random acceleration molecular dynamics simulations. *J. Chem. Theory Comput.* 14, 3859–3869. doi: 10.1021/acs.jctc.8b00230
- Laio, A., and Parrinello, M. (2002). Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* 99, 12562–12566. doi: 10.1073/pnas.202427399
- Lee, K. S. S., Yang, J., Niu, J., Ng, C. J., Wagner, K. M., Dong, H., et al. (2019). Drug-target residence time affects *in vivo* target occupancy through multiple pathways. *ACS Central Sci.* 5, 1614–1624. doi: 10.1021/acscentsci.9b00770
- Lensink, M. F., Velankar, S., and Wodak, S. J. (2017). Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition. *Proteins* 85, 359–377. doi: 10.1002/prot.25215
- Lotz, S. D., and Dickson, A. (2018). Unbiased molecular dynamics of 11 min timescale drug unbinding reveals transition state stabilizing interactions. *J. Am. Chem. Soc.* 140, 618–628. doi: 10.1021/jacs.7b08572
- Lu, H., and Tonge, P. J. (2010). Drug-target residence time: critical information for lead optimization. *Curr. Opin. Chem. Biol.* 14, 467–474. doi: 10.1016/j.cbpa.2010.06.176
- Nishikawa, N., Han, K., Wu, X., Tofoleanu, F., and Brooks, B. R. (2018). Comparison of the umbrella sampling and the double decoupling method in binding free energy predictions for SAMPL6 octa-acid host-guest challenges. *J. Comput. Aided Mol. Design* 32, 1075–1086. doi: 10.1007/s10822-018-0166-2
- Northrup, S. H., Allison, S. A., and McCammon, J. A. (1984). Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *J. Chem. Phys.* 80:1517. doi: 10.1063/1.446900
- Nunes-Alves, A., Kokh, D. B., and Wade, R. C. (2020). Recent progress in molecular simulation methods for drug binding kinetics. *arXiv*. arXiv: 2002.08983.
- Pan, A. C., Xu, H., Palpant, T., and Shaw, D. E. (2017). Quantitative characterization of the binding and unbinding of millimolar drug fragments with molecular dynamics simulations. *J. Chem. Theory Comput.* 13, 3372–3377. doi: 10.1021/acs.jctc.7b00172
- Parks, C. D., Gaieb, Z., Chiu, M., Yang, H., Shao, C., Walters, W. P., et al. (2020). D3R grand challenge 4: blind prediction of protein-ligand poses, affinity rankings, and relative binding free energies. *J. Comput. Aided Mol. Des.* 34, 99–119. doi: 10.1007/s10822-020-00289-y
- Pei, J., Yin, N., Ma, X., and Lai, L. (2014). Systems biology brings new dimensions for structure-based drug design. *J. Am. Chem. Soc.* 136, 11556–11565. doi: 10.1021/ja504810z
- Rifai, E. A., Dijk, M., Vermeulen, N. P. E., Yanuar, A., and Geerke, D. P. (2019). A comparative linear interaction energy and MM/PBSA study on SIRT1-ligand binding free energy calculation. *J. Chem. Inform. Model.* 59, 4018–4033. doi: 10.1021/acs.jcim.9b00609
- Rizzi, A., Jensen, T., Slochow, D. R., Aldeghi, M., Gapsys, V., Ntekoumes, D., et al. (2020). The SAMPL6 SAMPLing challenge: assessing the reliability and efficiency of binding free energy calculations. *J. Comput. Aided Mol. Des.* 34, 601–633. doi: 10.1007/s10822-020-00290-5

- Rizzi, A., Murkli, S., McNeill, J. N., Yao, W., Sullivan, M., Gilson, M. K., et al. (2018). Overview of the SAMPL6 host-guest binding affinity prediction challenge. *J. Comput. Aided Mol. Des.* 32, 937–963. doi: 10.1007/s10822-018-0170-6
- Shirts, M. R., and Chodera, J. D. (2008). Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* 129:124105. doi: 10.1063/1.2978177
- Singhal, N., Snow, C. D., and Pande, V. S. (2004). Using path sampling to build better Markovian state models: predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. *J. Chem. Phys.* 121, 415–425. doi: 10.1063/1.1738647
- Song, L. F., Bansal, N., Zheng, Z., and Merz, K. M. (2018). Detailed potential of mean force studies on host-guest systems from the SAMPL6 challenge. *J. Comput. Aided Mol. Des.* 32, 1013–1026. doi: 10.1007/s10822-018-0153-7
- Spagnuolo, L. A., Eltschkner, S., Yu, W., Daryaei, F., Davoodi, S., Knudson, S. E., et al. (2017). Evaluating the contribution of transition-state destabilization to changes in the residence time of triazole-based InHa inhibitors. *J. Am. Chem. Soc.* 139, 3417–3429. doi: 10.1021/jacs.6b11148
- Suárez, E., Lettieri, S., Zwier, M. C., Stringer, C. A., Subramanian, S. R., Chong, L. T., et al. (2014). Simultaneous computation of dynamical and equilibrium information using a weighted ensemble of trajectories. *J. Chem. Theory Comput.* 10, 2658–2667. doi: 10.1021/ct401065r
- Sullivan, M. R., Yao, W., and Gibb, B. C. (2019). The thermodynamics of guest complexation to octa-acid and tetra-endo-methyl octa-acid: reference data for the sixth statistical assessment of modeling of proteins and ligands (SAMPL6). *Supramol. Chem.* 31, 184–189. doi: 10.1080/10610278.2018.1549327
- Synapse (2018). *IDG-DREAM Drug-Kinase Binding Prediction Challenge*.
- Tiwary, P., Mondal, J., and Berne, B. J. (2017). How and when does an anticancer drug leave its binding site? *Sci. Adv.* 3:e1700014. doi: 10.1126/sciadv.1700014
- Tonge, P. J. (2017). Drug-target kinetics in drug discovery. *ACS Chem. Neurosci.* 9, 29–39. doi: 10.1021/acscchemneuro.7b00185
- Torrie, J. M., and Valleau, J. P. (1977). Non-physical sampling distributions in Monte-Carlo free-energy estimation umbrella sampling. *J. Comput. Phys.* 23, 187–199. doi: 10.1016/0021-9991(77)90121-8
- Vanden-Eijnden, E., and Venturoli, M. (2009). Exact rate calculations by trajectory parallelization and tilting. *J. Chem. Phys.* 131:044120. doi: 10.1063/1.3180821
- Vauquelin, G., Bostoen, S., Vanderheyden, P., and Seeman, P. (2012). *Clozapine, Atypical Antipsychotics, and the Benefits of Fast-Off D2 Dopamine Receptor Antagonism, Vol. 385*. Brussels: Springer. doi: 10.1007/s00210-012-0734-2
- Votapka, L. W., Jagger, B. R., Heyneman, A. L., and Amaro, R. E. (2017). SEEKR: Simulation Enabled Estimation of Kinetic Rates, a computational tool to estimate molecular kinetics and its application to trypsin-benzamidine binding. *J. Phys. Chem. B* 121, 3597–3606. doi: 10.1021/acs.jpcc.6b09388
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comput. Chem.* 25, 1157–1174. doi: 10.1002/jcc.20035
- Woo, H. J., and Roux, B. (2005). Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U.S.A.* 102, 6825–6830. doi: 10.1073/pnas.0409005102
- Yin, J., Henriksen, N. M., Slochow, D. R., Shirts, M. R., Chiu, M. W., Mobley, D. L., et al. (2017). Overview of the SAMPL5 host-guest challenge: are we doing better? *J. Comput. Aided Mol. Des.* 31, 1–19. doi: 10.1007/s10822-016-9974-4
- Zwanzig, R. W. (1954). High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.* 22, 1420–1426. doi: 10.1063/1.1740409

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Hall, Dixon and Dickson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.