

Predicting ligand binding affinity for the SAMPL6 challenge from on- and off-rates using weighted ensembles of trajectories

Tom Dixon and Alex Dickson*

Department of Biochemistry & Molecular Biology,

Michigan State University, East Lansing, MI 48824 and

Department of Computational Mathematics, Science and Engineering,

Michigan State University, East Lansing, MI 48824

Samuel D. Lotz

Department of Biochemistry & Molecular Biology,

Michigan State University, East Lansing, MI 48824

(Dated: May 28, 2018)

Abstract

Interest in ligand binding kinetics has been growing rapidly, as it is being discovered in more and more systems that ligand residence time is the crucial factor governing drug efficacy. Many enhanced sampling methods have been developed with the goal of predicting k_{on} and/or k_{off} through explicitly simulation of ligand binding pathways, and these methods work by very different mechanisms, and make different trade-offs between accuracy and efficiency. Although there is not yet a blind challenge for ligand binding kinetics, here we take advantage of experimental measurements and rigorously computed benchmarks to compare estimates of K_D calculated as the ratio of two rates: $k_{\text{off}}/k_{\text{on}}$. These rates were determined using a new enhanced sampling method based on the weighted ensemble framework that we call “REVO”: Reweighting of Ensembles by Variance Optimization. This is a further development of the WExplore enhanced sampling method, in which trajectory cloning and merging steps are guided not by the definition of sampling regions, but by the optimization of a variance parameter. Here we obtain estimates of k_{on} and k_{off} that are consistent across multiple simulations, with an average log10-scale standard deviation of 0.28 for on-rates and 0.56 for off-rates, which is well within an order of magnitude and far better than previously observed for previous applications of the WExplore algorithm. However, predicted K_D values were systematically lower than the reference values by an average of 4.2 kcal/mol. Using tree network visualizations of the trajectories in the REVO algorithm, and conformation space networks for each system, we analyze the results of our sampling, and hypothesize sources of discrepancy between our K_D values and the reference. We also motivate the direct inclusion of k_{on} and k_{off} challenges in future iterations of SAMPL, to further develop the field of ligand binding kinetics prediction and modeling.

*Electronic address: alexrd@msu.edu

I. INTRODUCTION

Binding affinity (K_D) has long been seen as the crucial parameter for drug discovery, as it determines the proportion of drug that is bound to a receptor in solution. A wide variety of methods have emerged to predict both absolute and relative binding affinities, each with its own domain of applicability, and tradeoff between efficiency and accuracy. The SAMPL challenge is playing an important role to compare tools that predict affinities using blind predictions [1]. Importantly, errors can arise from both the physical model used to describe the system (e.g. forcefield, thermostat, dynamics engine), and from the sampling methodology used. The SAMPLing challenge, described in this issue, thus serves an important role in comparing the accuracy of computational methods that all employ the same physical model [2].

While the binding affinity is all that is needed to describe the action of a ligand at equilibrium, the on- (k_{on}) and off-rates (k_{off}) are necessary to model drug action in general [3]. For instance, in many systems it has been observed that drug residence time ($RT = 1/k_{\text{off}}$) is the critical factor governing efficacy in living cells [4]. This is due to the number of factors that drive the system out of equilibrium, such as drug metabolism and elimination, the turnover of target protein, and the periodic nature of drug administration. Although $K_D = k_{\text{off}}/k_{\text{on}}$, and lower K_D can be correlated to lower k_{off} , this relationship is governed by the free energy curve of ligand binding, particularly the ligand binding transition state, which is the highest point in free energy between the bound and unbound states [5]. The binding rate, k_{on} , has an upper bound of $10^9 \text{ M}^{-1} \text{ s}^{-1}$, which corresponds to the “diffusion limit”, binding rates of ligands to the same target have been shown to vary over 4 orders of magnitude, which disrupts the correlation between K_D and k_{off} [6].

Prediction of k_{off} and k_{on} is challenging, as they are not state functions: they depend fundamentally on the transition path ensemble between the bound and unbound states. Computational sampling of these transition paths is in general a great challenge for molecular dynamics due to the long timescales of ligand binding and release, although in recent years, a variety of enhanced sampling methods have rose up to meet this challenge [7]. The trypsin-benzamidine system has served as a common benchmark application for enhanced sampling methods such as Adaptive Multilevel Splitting [8], SEEKR [9], adaptive [10] and traditional [11, 12] Markov state modeling, funnel metadynamics [13], as well as the WExplore method

developed by our group [14]. Recently these efforts have been expanded to more challenging systems such as the unbinding of inhibitors from c-Src kinase [15] and p38 MAP kinase [16] using metadynamics, and the unbinding of the TPPU ligand from the target soluble epoxide hydrolase with WExplore [17]. The diversity of computational approaches to handle long timescale ligand binding and release events is a promising sign for the field, but comparison of methodologies is complicated – even for applications to the same system – due to differences in forcefields, boundary conditions, and integrators.

As a step toward the robust comparison of different computational methods for simulation of binding pathways, we participated in the SAMPLing challenge for the prediction of binding affinities. Instead of computing free energies through alchemical perturbation, here we explicitly simulate the binding and release processes, determine the absolute rates k_{on} and k_{off} , and compute the binding affinity as the ratio $k_{\text{off}}/k_{\text{on}}$. As we broadly sample unbinding pathways from multiple starting points, we can also synthesize these results and examine how these poses are connected in the binding network.

We efficiently determine unbinding and binding rates using a further developed variant of the WExplore sampling method [18]. This is the first application of this new method, which we call “Reweighting of Ensembles by Variance Optimization”, or REVO. This new method is also based in the weighted ensemble framework [19], where trajectories are merged and cloned, but it is the first to completely eschew the idea of dividing a space into a set of sampling regions (the possibility has previously been recognized however [20]). REVO instead directs merging and cloning operations by optimizing a variance parameter that describes the instantaneous spread of the ensemble of trajectories, which is described in the Methods section below. We visualize our REVO simulations using a branching tree network diagram, whose layout uses an energy function that takes into account the distances between the trajectories. This allows for the easy visualization of the correlation of exit point ensembles within a weighted ensemble simulation. We compare our binding affinities to computational reference values, and observe that the affinities from REVO are systematically tighter than the reference. We conclude the manuscript with a discussion of possible sources of error.

II. METHODS

A. Host-guest systems

The host-guest systems were selected from the main SAMPL6 challenge. One system is a cucurbit[8]uril host [21, 22], using quinine as a guest ligand (Figure 1). The host is a ring-shaped structure, with 8-fold rotational symmetry about the vertical axis, and two-fold symmetric about the horizontal axis. There are thus 16 symmetry-equivalent atom mappings that we make use of in our distance calculations. The second and third systems both use a Gibb deep cavity cavitand (also referred to as an “Octa Acid”) as a host [23]. Here there is only 4-fold symmetry about the vertical axis. Binding and release of two ligands is examined: 5-hexenoic acid and 4-methyl pentanoic acid, referred to as OA-G3 and OA-G6, respectively. Both of these ligands carry an explicit negative charge.

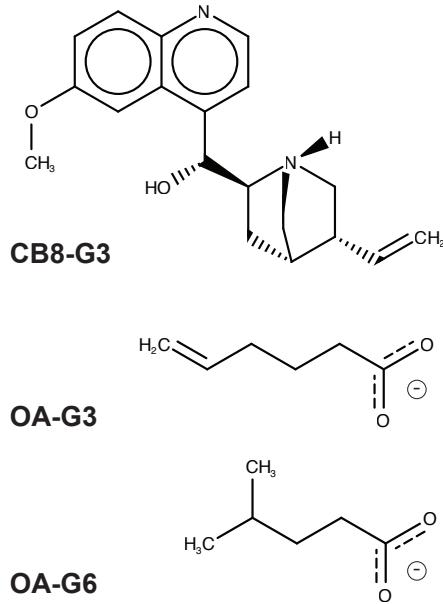


FIG. 1: Structure of the ligands used in this study. (Top) Quinine, referred to herein as CB8-G3. (Middle) 5-hexenoic acid (deprotonated form), referred to herein as OA-G3. (Bottom) 4-methyl pentanoic acid (deprotonated form), referred to herein as OA-G6.

B. Molecular dynamics

The fifteen initial configurations (five for each host-guest system) were used as prepared by the organizers of the SAMPLing challenge. OpenMM v7.1.1 was used to run dynamics on the CUDA v8.0 platform. We use a Langevin integrator, with a thermostat at 300 K, a friction coefficient of 1.0 ps^{-1} , and a timestep of 2 fs. The water boxes were used as prepared by the challenge organizers without modification.

C. Reweighting of Ensembles by Variance Optimization

To encourage the sampling of rare events, we developed a method based on the Weighted Ensemble (WE) framework [19] that we call “Reweighting of Ensembles by Variance Optimization”, or REVO. WE methods use an ensemble of trajectories that are each assigned a statistical weight, and enhance sampling through the introduction of cloning and merging steps. When trajectories are *cloned*, their weight is divided among the progeny. This is typically done in under-sampled regions of space, in order to boost the probability of observing rare events in the simulation. Trajectories are also *merged* together, and their summed weight is given to the resulting merged walker. In practice, merging walkers A and B is accomplished by choosing a survivor (walker A is chosen with probability $\frac{w_A}{w_A+w_B}$), and discarding the other walker. Merging is typically done in over-sampled regions, with walkers that can be seen as “redundant”.

Previous applications of the weighted ensemble methods, proceed by constructing a set of sampling regions, determining their occupancies, and using cloning and merging operations to make the occupancies as even as possible. In general, the free energy landscapes of interest are inherently high-dimensional, which makes it difficult to construct an appropriate set of regions. For this reason we were motivated to discard the notion of “regions” entirely, and direct cloning and merging operations instead by the optimization of a variance measure, V :

$$V = \sum_i \sum_{j>i} (d_{ij})^p \omega_i \omega_j \quad (1)$$

where the double sum is over all pairs of walkers, d_{ij} is a distance metric, p is a parameter set here to 4, and ω_a is a weighting function for walker a :

$$\omega_a = \log(100 * w_a / p_{\min}) \quad (2)$$

where w_a is the weight of trajectory a , and p_{\min} is the lowest probability attainable by a walker, set here to be 10^{-12} . The weighting function ω was designed to be largest for high w_a , and to smoothly decay to a low value as w_a approaches p_{\min} .

The structure of the REVO “resampling” algorithm proceeds as follows. Eq. 1 is used to compute the variance function, and the walker with the highest (“H”) and lowest (“L”) contributions to the variance are identified. The closest walker to “L” is identified, called “C”. A coupled cloning and merging event is proposed, where “C” and “L” are merged and “H” is cloned. Eq. 1 is again used to recompute the variance, and this coupled cloning and merging move is accepted if the variance increases. Further moves are proposed after recomputing “H”, “L” and “C”, and the process continues until a move is rejected. This way, the algorithm automatically determines the optimal number of cloning and merging events. In fact, if the system is already in an optimal configuration, no further cloning and merging operations will take place, and REVO will skip to the next dynamics step.

As in previous WExplore applications, a minimum and maximum walker weight was enforced (p_{\min} and p_{\max} , respectively). This is done by only identifying “H”, “L” and “C” walkers that will not violate these conditions. In these simulations, $p_{\min} = 10^{-12}$ and $p_{\max} = 10^{-1}$, following previous work.

This process is general to any dynamics engine, and to any form of the distance function d_{ij} . Here we use two different distance functions to describe the unbinding and rebinding processes. For unbinding, d_{ij}^U is defined as the RMSD, in Angstroms, of the guest ligand between structures i and j , after aligning to the host. As mentioned in Section II A, there are multiple symmetry-equivalent mappings of the host atoms. We thus compute this distance after alignment of j to each symmetry-equivalent mapping of host i , and use the smallest such value as d_{ij}^U . For rebinding, d_{ij}^R is computed using the RMSD of both i and j to the reference starting structure:

$$d_{ij}^R = |1/d_{iR}^U - 1/d_{jR}^U| \quad (3)$$

where d_{aR}^U is the distance from walker a to the reference structure. The difference between the inverse of these two quantities is used to highlight differences between small values of this quantity (e.g. between RMSD = 1.5 Å and RMSD = 2.0 Å).

Unbinding REVO simulations were run for 2000 cycles, with 48 walkers run for $\Delta t = 20$ ps each cycle. The exit points registered after 1000 cycles were used to initialize the rebinding REVO simulations. In some cases, fewer than 48 exit points were obtained at this point, and

the walkers were randomly cloned in order to create a full set of 48 walkers. The rebinding REVO simulations were run for 200 cycles, with $\Delta t = 200$ ps per cycle. In aggregate, we ran $1.92 \mu\text{s}$ for each of the unbinding and rebinding simulations, $3.84 \mu\text{s}$ for each starting pose, or $57.6 \mu\text{s}$ over the entire set of results presented here.

1. Note about CB8-G3-0 and CB8-G3-4

After the conclusion of the SAMPLing challenge we found an error in the weight normalization procedure that was used to initialize the weights of the rebinding walkers when fewer than 48 exit points were observed. This affected only two simulations: CB8-G3-0 and CB8-G3-4, where only 5 and 7 exit points were observed, respectively, in the first 1000 cycles of the unbinding simulation. Due to an error, the initial weights in these rebinding simulations summed to a value greater than 1, and while this could be accounted for in the rate calculations, it was compounded by the fact that no walker in these simulations had a weight value less than $p_{\max} = 0.1$, and thus no cloning/merging moves could occur.

Surprisingly, this did not affect the calculation of the binding rate. Although the number of binding events observed in CB8-G3-0 and CB8-G3-4 (32 and 25, respectively), was much lower than the number observed in CB8-G3-1, CB8-G3-2 and CB8-G3-3 (289, 427 and 190), the total amount of weight that exited was comparable (0.62, 0.43, 0.66, 0.14, 0.50, for starting poses 0 through 4). This goes to show the downhill nature of binding in host-guest systems, as confirmed by the almost diffusion-limited k_{on} (see Table I. The calculated MFPT of binding for the CB8-G3 system was 91 ns, which is well within the aggregate sampling time of each rebinding simulation ($1.92 \mu\text{s}$), again indicating why a group of straight-forward trajectories was able to produce over two dozen binding events each.

D. Calculating rates by ensemble splitting

REVO, like other weighted ensemble methods, can calculate kinetic quantities on the fly, through a technique we call “ensemble splitting” [24, 25] (also referred to as “tilting” [26], or “coloring” [27, 28]). An equilibrium ensemble is split into two non-equilibrium ensembles by defining two basins, in this case the “bound” basin and the “unbound” basin. The unbinding ensemble is defined as the set of trajectories that have most recently visited the

bound basin, and the rebinding ensemble is the set of trajectories that have most recently visited the unbound basin. The unbound basin is the set of structures where the closest host-guest interatomic distance exceeds 10 Å, as in previous work. The bound basin is defined as the set of structures with guest RMSD < 1.0 Å, computed after aligning to the host. Note that a sweep over symmetry-equivalent atom mappings of the host was again conducted, so a binding event can be registered by binding to either the top or bottom of the CB8 host, for example.

In this work, REVO simulations are conducted explicitly either in the unbinding ensemble, or the rebinding ensemble. After each dynamics step, any walker that has exited its ensemble (by entering the opposite basin) is identified. Its weight is recorded, and its structure is “warped” back to the starting structure. In the unbinding ensemble, the starting structure is the initial bound pose. In the rebinding ensemble, the starting structures are exit points that were generated by the unbinding simulations. The rates are simply calculated using the flux of trajectories (sometimes referred to as the Hill relation [27, 29]) that leave the ensemble:

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

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

where T is the elapsed time, and the sums are over the set of exit points observed before time T , and C is the concentration of the ligand, computed as $1/V$ where V is the box volume.

E. Visualization of trajectory trees

We visualize cloning events in a tree graph, where each node represents a walker at a given time point and the edges indicate how walkers are connected through time. Each level (y -position) on the tree represents walkers at the same time step. The initial horizontal placement (x -position) of each node is a direct result of its parent’s position in the previous time step. If no cloning events occurred for that walker, then the node is placed directly above its parent. If the parent was cloned, then the walkers are spread out in a fan pattern. Once the nodes are initially placed, their x -positions are minimized with a steepest descent

algorithm using the following energy function:

$$E = \sum_i \left[b(x_i^t - x_i^{t-1})^2 + cw_i(x_i^t)^2 + \sum_j E_{ij} e^{\frac{(x_i^t - x_j^t)^2}{r_0}} \right] \quad (6)$$

where x_i^t and x_j^t are the positions on the tree of walkers i and j at time t , x_i^{t-1} is the position of the parent at the previous time step, and w_i is the walker weight obtained from the simulation. The variables b, c, r_0 are parameters set here to 0.01, 5 and 1000 respectively. The first term causes the nodes to stay close to their parent's position, allowing trajectories to be visually tracked through the tree more easily. The second term encourages the higher weight trajectories to stay close to $x = 0$. The third term is a pairwise repulsion term, which gives the nodes a “radius” of r_0 , and is scaled by a repulsion energy (E_{ij}) that takes into account the molecular distance between the walkers in the simulation (d_{ij}):

$$E_{ij} = a * \max(0, d_{ij} - d_0) \quad (7)$$

where a and d_0 are parameters set here to 2.5 and 2.0. It is important to note that this energy minimization only affects the x -position of each node. The y -position is determined by the timestep and is not used in the steepest descent algorithm. The graphs were made using NetworkX 2.2 library and visualized using Gephi 0.9.2.

F. Clustering and visualization of conformation space networks

All of the trajectory frames for the five starting poses of each system were clustered together using the MSMBuilder 3.8.0 library. The clustering was done on a featurized space defined by the Canberra distance between all atoms of the host molecule and all atoms of the guest molecule, this amounted to 7056, 3128, and 3496 distances for the CB8-G3, OA-G3 and OA-G6 systems, respectively. A k-centers clustering algorithm was used to generate 1000 clusters using the featurized space and assign each frame of the trajectories to a cluster. A count matrix describing the cluster-cluster transitions was calculated. This corresponds to a Markov state model with a lag-time equal to the cycle length $\Delta t = 20$ ps.

We then construct Conformation Space Networks (CSNs) from the count matrices, which are graphical models of the transition matrix, with a node representing each row, and edges representing non-zero off-diagonal elements. Gephi 0.9.2 was used to visualize the CSN. The size of each node is proportional to the statistical population of the cluster. The

smallest node was 20 times smaller than the largest node. The topology of the network was determined using a force minimization algorithm, Force Atlas, included in Gephi. This algorithm includes repulsive forces for nodes that are not connected and attractive forces proportional to the weight of the edges. The directed edge weights were values between 0.1 and 100 as determined by $w_{ij} = 100p_{ij}$ where p_{ij} is the transition probability of cluster i transitioning to cluster j. Undirected edge weights were then determined as the average between the two directed edge weights. Force Atlas was applied twice, first without adjusting for node sizes which enabled the nodes to overlap, and then a second minimization adjusted for node size which prevented overlap. For visualization, all edge weights were given a uniform value.

III. RESULTS

A. Warped walkers

For each host-guest system we run 5 REVO simulations, each from a different starting pose (Figure S1-S3). As mentioned in Methods, these simulations are run under two different sets of nonequilibrium conditions. In “unbinding simulations”, trajectories are initialized in the bound pose and are terminated when the minimum host-guest interatomic distance exceeded 10 Å. These exit points are used as seeds for the “rebinding simulations”, where trajectories are initialized in the unbound state, and terminated in the bound state, which is marked by the guest RMSD falling below 1.0 Å. Note again that this distance is computed using all symmetry-equivalent orientations of the host. In general we refer to the termination of trajectories as “warping events”, as the trajectory is instantaneously transported in space.

All 30 of our REVO simulations generated a substantial number of warping events. In general these are distributed across a wide range of weight values (Figure 2). For all systems it is observed that rebinding can occur with very high probability walkers ($p > 0.1$), but that unbinding occurs with much lower probability. Indeed it is the probability of the unbinding warped walkers that largely governs differences in K_D and k_{off} between the systems. The minimum weight that is achievable by a walker, p_{\min} , was set to 10^{-12} in all cases. As shown in Figure 2, this could be increased substantially (e.g. to 10^{-3}) in the rebinding case to avoid the integration of low-weight trajectories that will not meaningfully contribute to the

binding flux.

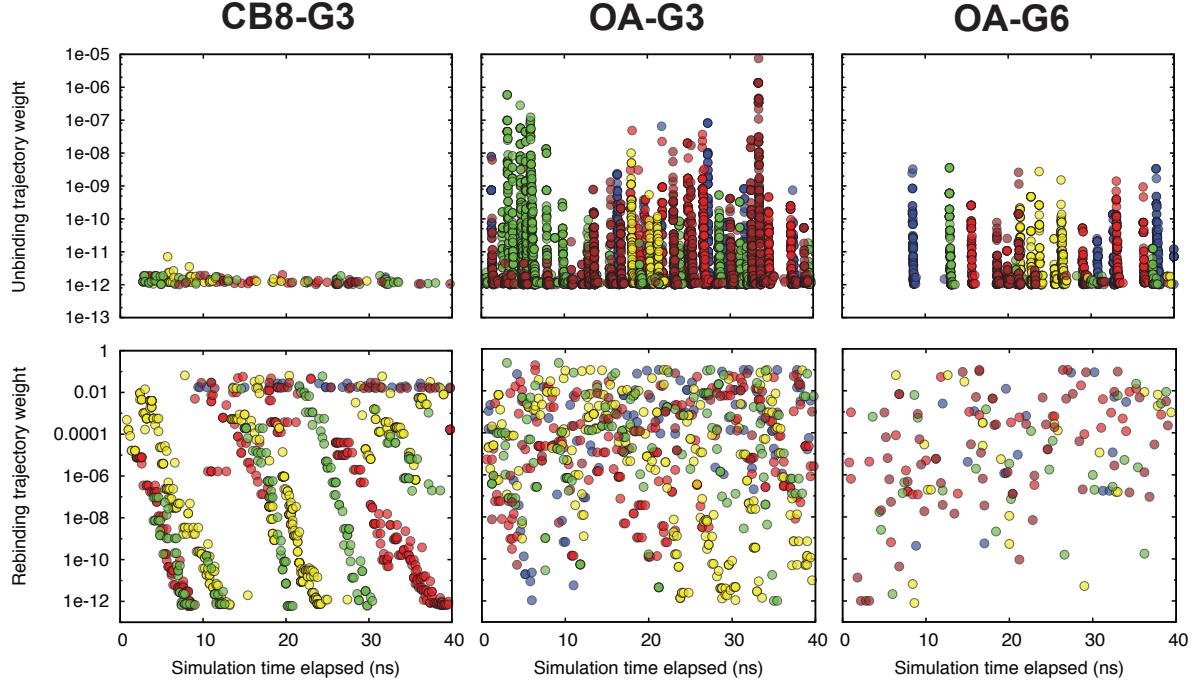


FIG. 2: **Weights of warped walkers.** Weights of warping events for the unbinding (top row) and rebinding (bottom row) simulations. In both cases the points are colored according to the index of the corresponding starting pose (0, blue; 1, red; 2, yellow; 3, green; 4, brown).

The warping points for the unbinding simulation are shown in Figure 3, again using color to indicate the starting pose. Although they exhibit some strong correlation within a REVO run, together they comprise a broad distribution. For CB8-G3, both upward and downward exit pathways are sampled with roughly equal frequency, whereas for the Octa-Acid systems, the exit points are clustered towards the top of the cavitand.

B. Kinetics and free energies

The binding and unbinding rates are calculated using sum of the weights of the warped walkers, divided by the elapsed time (see “Calculating rates by ensemble splitting” in Methods). The binding rate is calculated by dividing the binding trajectory flux by the concentration of the guest in mol/L, calculated as $\frac{1}{N_A V}$, where V is the box volume. The concentration ranged from 0.021 M for OA-G6 to 0.025 M for CB8-G3 and OA-G3, resulting from unit

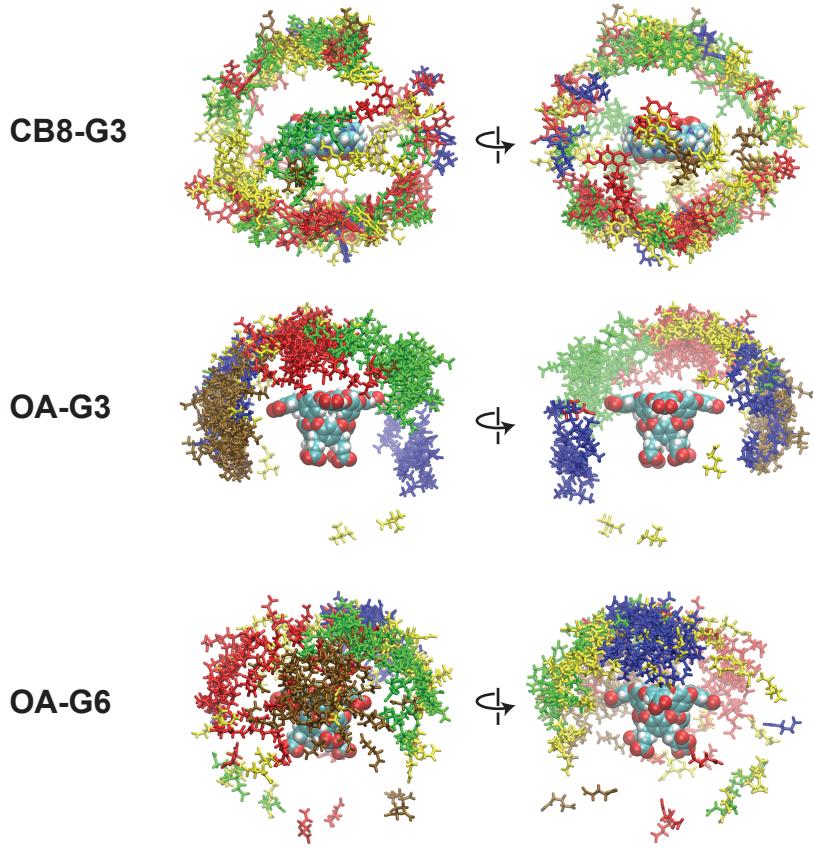


FIG. 3: Spatial distribution of warped walkers. Structures of warping events for the unbinding simulations viewed from the front and back. Guest ligands are colored according to the index of the corresponding starting pose (0, blue; 1, red; 2, yellow; 3, green; 4, brown).

cells with side-length ranging from 4.1 nm to 4.3 nm. Running estimates of k_{on} and k_{off} are shown individually for each REVO simulation in Figure 4, along with their average, which is calculated by averaging the trajectory flux over the set of five simulations. Large, upward jumps are observed in the rate curves whenever an exit point is recorded that has a higher weight than was previously observed.

The final average rate values, as well as the corresponding mean first passage times, are given in Table I. The mean first passage times (MFPT) of unbinding demonstrate the power and scope of the REVO method: we estimate that the CB8-G3 system has an average ligand residence time of 830 seconds, and we obtain multiple ligand release events for each of the five starting poses. In total, we used 9.6 μs of sampling in the CB8-G3 unbinding ensemble,

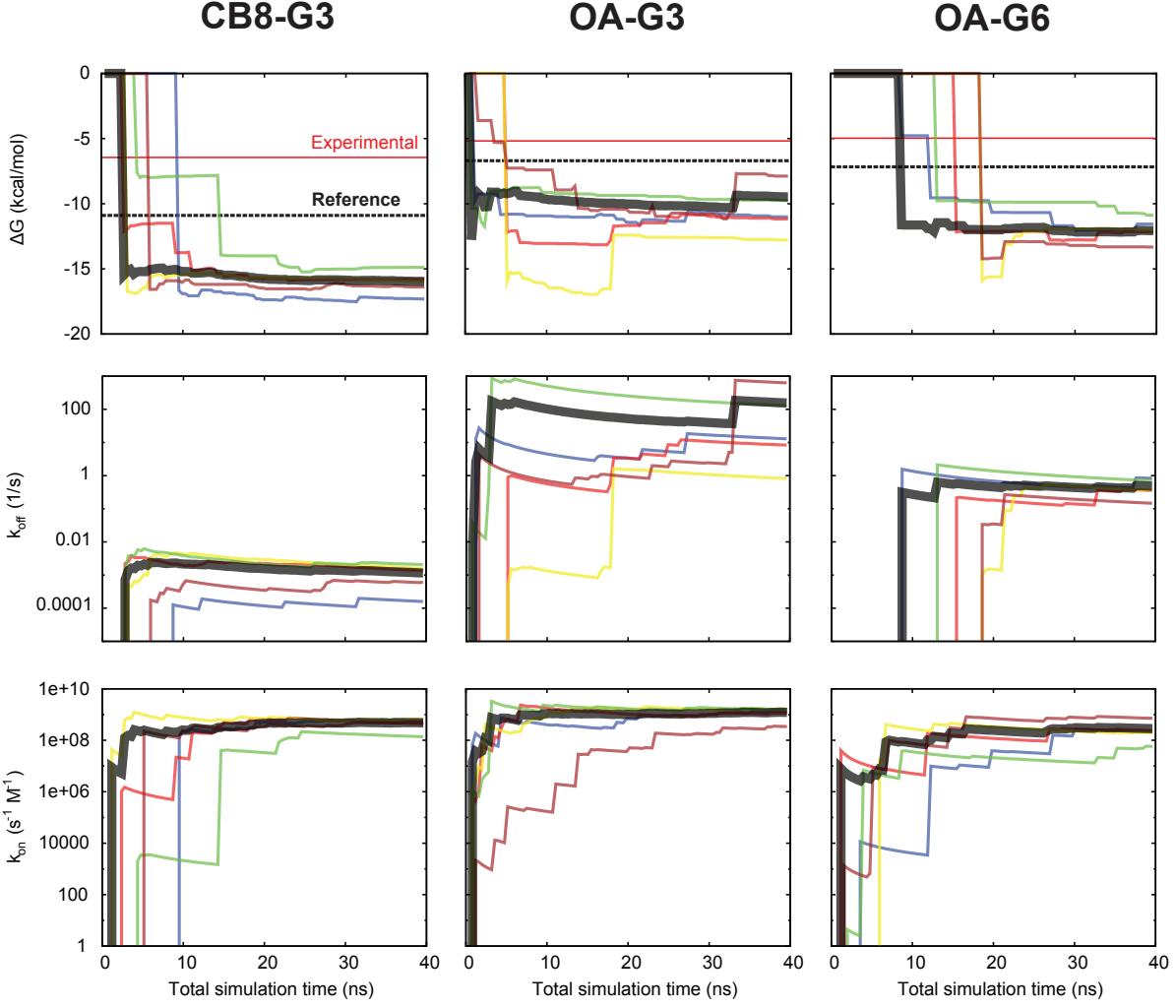


FIG. 4: Predicted kinetics and free energies. The calculated free energies (top), off-rates (middle), and on-rates (bottom) are shown as a function of simulation time for each starting pose in each host-guest system. The curves are colored according to the index of the starting pose as in Figures 3 and 2. The calculated binding free energies are compared with experimental measurements (horizontal red line), and the computational reference (dashed black line) for each system.

resulting in an acceleration factor of $\approx 8.6 \times 10^7$.

With k_{off} and k_{on} in hand, the binding affinity is calculated as $\Delta G = kT \ln \frac{k_{\text{off}}}{Ck_{\text{on}}}$, where C is the reference concentration of 1 mol/L, and $kT = 0.597$ kcal/mol, corresponding to a temperature of 300 K. This binding affinity is compared to both the experimentally mea-

TABLE I: Pose-averaged rates and affinities

	k_{off} (s $^{-1}$)	MFPT $_{\text{off}}$ (s)	k_{on} (s $^{-1}$ M $^{-1}$)	MFPT $_{\text{on}}$ (ns)	ΔG (calc.)	ΔG (ref.)	ΔG (exp.)
CB8-G3	0.0012	830	4.8×10^8	91	-16.0	-10.9	-6.45
OA-G3	160	0.0064	1.2×10^9	36	-9.5	-6.7	-5.18
OA-G6	0.48	2.1	2.8×10^8	160	-12.0	-7.17	-4.97

sured binding affinity, and a computational reference computed using alchemical free energy calculations with YANK (see [2] for more details). As shown in Figure 4, the host-guest affinity calculated by the rate ratio in REVO is systematically too tight when compared both the experimental and reference values. This is possibly due to finite box size effects, which is discussed further in the Discussion and Conclusions section.

Moderate variation in k_{on} and k_{off} is observed across the sets of simulations for each host-guest system, which contributes to some uncertainty in the predicted rates and affinities. However, the average standard deviation in the log10 final rates ($\log 10(k)$) is 0.28 for on-rates and 0.56 for off-rates, both well under an order of magnitude. This compares very favorably with recent studies using WExplore [7, 17], where rates from individual simulations varied over several orders of magnitude.

C. Trajectory trees reveal correlation between exit points

Rates are derived from exit points, and while points from different starting poses are guaranteed to be independent, it is unclear how correlated the observations are within a given REVO simulation. We can use a tree network to observe the entire set of merging and cloning events that occur during a simulation, and to determine how closely related walkers are to one another. Additionally, one can visualize the state of the walkers through coloring the tree based on physical properties observed during the simulation, such as the solvent accessible surface area (SASA) of the guest molecule, which can help evaluate how close the guest is to unbinding from or rebinding to the host. Using this coloring, and how closely related walkers are to one another, we can visualize the correlation between a set of unbinding or rebinding events.

Figure 5 shows a trajectory tree for the OA-G3-0 unbinding simulation. From the tree it is clear that the majority of sampling time is spent sampling the bound state (dark green

structures). However, the top inset shows that this sampling is still very active, with outliers being detected and cloned nearly every cycle, although the vast majority of these clones are merged one or two cycles later, which implies that the outlying property corresponded to a fast degree of freedom. The middle inset shows a breakout event that led to a series of exit points. The vertical “branches” show individual trajectories. Termination of a branch with high SASA (orange) correspond to exit points.

The OA-G3-0 simulation generated 966 exit points, 534 of which can be seen in Figure 5, which captures only the first 1329 cycles. From the tree it can be seen that many of these exit points are correlated, as they were recently cloned from common ancestors. Using the tree analysis one can observe that there are likely at least seven distinct groups of exit points that can be treated as independent observations of unbinding pathways.

In the bottom inset we see a trajectory that demonstrates transient rebinding behavior. That is, the SASA goes high (≈ 320 , orange), to medium (≈ 160 , light green), back to high again. This behavior results from a transient, loose association with the exterior of the host molecule. This trajectory is shown as Supporting Movie 1.

D. Conformation space networks reveal connection between starting poses

Here we obtain combined estimates of k_{on} , k_{off} and K_D by averaging the transition flux from simulations with different starting poses, and in the case of the rebinding simulations, different boundary conditions. This is only appropriate if the five starting poses are all part of the same basin of attraction, and can interconvert on timescales much faster than the unbinding process. If two poses form distinct basins of attraction, then we cannot expect that the poses will have the same k_{off} , k_{on} , or K_D . To examine the connectivity of starting poses, we use the REVO trajectory segments to construct a Markov state model. We then visualize conformation space networks (CSNs) to examine how the starting poses are connected, whether they are in the same basin of attraction, and whether they share the same (un)binding pathways.

Figure 6 shows CSNs for the unbinding simulations of all three host-guest systems. For both OA systems a large, densely-connected ensemble of bound states is observed. As the entire set of host-guest distances was used to featurize our dataset, this heterogeneity arises from motions of the flexible chemical groups on the bottom and around the rim of the Octa

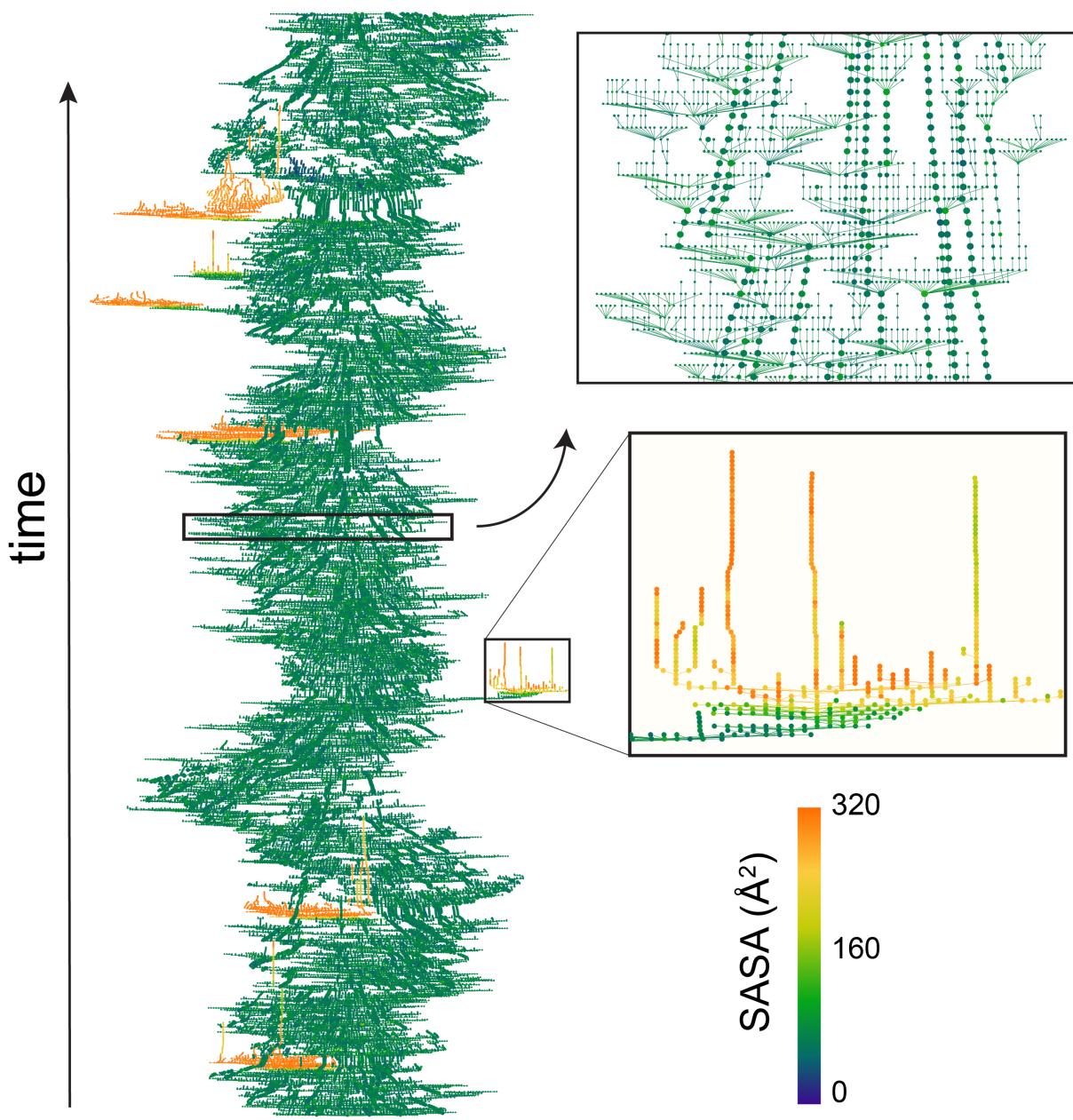


FIG. 5: Trajectory trees show all cloning and merging events in a simulation. The trajectory tree for the first 1329 cycles of the OA-G3-0 unbinding simulation is shown. Each horizontal row in this tree represents a cycle, and the placement of all 48 nodes in the row is determined by minimizing an energy function (see “Visualization of trajectory trees” in Methods). SASA is used to color the nodes, with blue and dark green indicating bound structures, and yellow to orange indicating unbound.

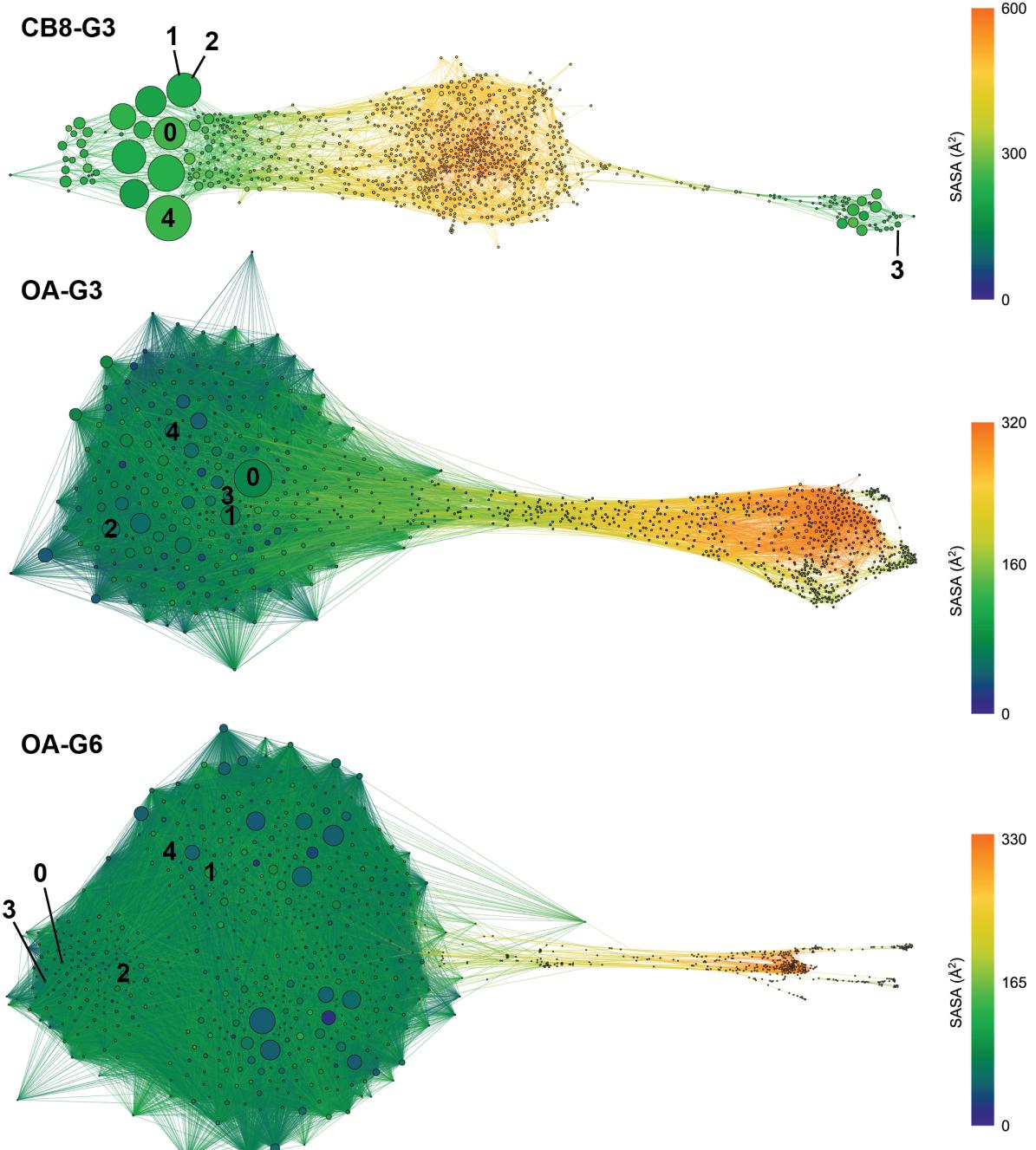


FIG. 6: Conformation space networks for the unbinding simulations. Each node in a CSN represents a cluster of host-guest structures. Edges in the networks connect clusters that are seen to interconvert in the REVO simulations. The size of each node is proportional to the number of times it was observed in the unbinding simulations. Nodes are colored according to the solvent accessible surface area of the guest molecule, as shown in the color-bars on the right. The clusters corresponding to the starting poses are labeled in each network.

acid host molecule. Starting structure 3 in CB8-G3 is bound in the opposite orientation from the others (see Figure S1), although the host-molecule is symmetric to inversion about the horizontal plane. While this did not affect the kinetic measurements (which took into account symmetry-equivalent atom mappings of the host), in the CSN it forms a distinct basin from the other starting poses. This allows us to observe that the ligand cannot flip between these two structures inside the host, and instead converts between the two poses only through the quasi-bound and unbound states (yellow and orange). Although here we conclude that all structures are part of the same (or symmetry-equivalent) fast-interconverting bound ensemble, this type of analysis is useful to reveal the interconversion of binding poses, and whether we should expect them to have the same calculated residence times.

IV. DISCUSSION

Although we obtained much information about the binding and release processes of these host-guest systems, our predicted K_D values were systematically lower than those of a reference calculation employing the same forcefield (average 4.2 kcal/mol). These reference K_D values were themselves systematically lower than the experimentally calculated dissociation constants (average 2.7 kcal/mol), likely arising from inaccuracies in the forcefield. The nature of the SAMPLing challenge gives us a unique opportunity to isolate these different sources of error. Below we discuss different possible sources of error in light of the analyses presented above.

In weighted ensemble simulations that calculate kinetic quantities, convergence is often the first question. Here we devoted the same amount of sampling time to the binding and unbinding processes (1.92 μ s per system per starting pose). This is more than sufficient to capture the binding process, which has a mean first passage time ranging from 36 to 160 ns. The unbinding process was much more challenging, and it is possible that longer simulations would have captured higher weight walkers exiting from the bound state. This would increase our k_{off} estimates, and K_D as well. Significantly extending the unbinding simulations and monitoring their exit rates could provide additional insight.

We also have concerns related to the size of the simulation box. This was chosen to be appropriate for standard alchemical free energy perturbations, and not for simulations of

full unbinding and binding pathways. A more accurate determination of the binding rate could be obtained with the Northrup-Allison-McCammon (NAM) method, which combines the rate of first hitting points with a committor probability to determine the binding rate [30]. Diffusion at long distances is typically efficiently simulated using Brownian dynamics. This approach has been used successfully to determine binding rates with both the weighted ensemble method [31, 32], and the SEEKR method (Simulation Enabled Estimation of Kinetic Rates) [9].

An important point is that although the reference calculations were performed with the same forcefield, the rates can sensitively depend on aspects of the forcefield that are not relevant to alchemical measurements of the affinity. As an example, in CB8-G3 unbinding trajectory trees we observe long “tendrils” of unbound trajectories that are stuck in intermediate SASA values, where the guest ligand is bound to the outer surface of the host. The strength of these interactions can significantly affect our calculations of k_{off} , although they will not affect the alchemical K_D calculations.

In general, to successfully predict k_{on} and k_{off} will require optimizing the ligand forcefield terms that govern interactions that occur along binding pathways. By analogy, it is known that protein forcefields that are only trained on *folded* protein structures have difficulties representing unfolded and intrinsically disordered structures. As a community we must take care not to over-emphasize the ligand bound state in forcefield development. An extension of the SAMPL challenge to include the prediction of kinetic quantities would thus be tremendously valuable to the development of both sampling methodologies and forcefields.

- [1] M. T. Geballe, A. G. Skillman, A. Nicholls, J. P. Guthrie, and P. J. Taylor, Journal of Computer-Aided Molecular Design **24**, 259 (2010).
- [2] A. Rizzi, J. Chodera, D. Mobley, T. Jensen, M. Shirts, T. Dixon, S. Lotz, A. Dickson, J. Michel, M. Gilson, et al., Journal of Computer-Aided Molecular Design p. To appear (2018).
- [3] R. A. Copeland, Nature Reviews Drug Discovery **15**, 87 (2016).
- [4] B. Costa, E. Da Pozzo, C. Giacomelli, E. Barresi, S. Taliani, F. Da Settim, and C. Martini, Scientific Reports **6**, 18164 (2016).
- [5] A. C. Pan, D. W. Borhani, R. O. Dror, and D. E. Shaw, Drug Discovery Today **18**, 667 (2013).

- [6] D. B. Kokh, M. Amaral, J. Bomke, U. Grädler, D. Musil, H.-P. Buchstaller, M. K. Dreyer, M. Frech, M. Lowinski, F. Vallée, et al., *Journal of Chemical Theory and Computation* (2018).
- [7] A. Dickson, P. Tiwary, and H. Vashisth, *Current Topics in Medicinal Chemistry* **17**, 2626 (2017).
- [8] I. Teo, C. G. Mayne, K. Schulten, and T. Lelièvre, *Journal of Chemical Theory and Computation* p. acs.jctc.6b00277 (2016).
- [9] L. W. Votapka, B. R. Jagger, A. L. Heyneman, and R. E. Amaro, *Journal of Physical Chemistry B* **121**, 3597 (2017).
- [10] S. Doerr and G. De Fabritiis, *Journal of Chemical Theory and Computation* **10**, 2064 (2014).
- [11] I. Buch, T. Giorgino, and G. De Fabritiis, *Proceedings of the National Academy of Sciences of the United States of America* **108**, 10184 (2011).
- [12] N. Plattner and F. Noé, *Nature Communications* **6**, 7653 (2015).
- [13] V. Limongelli, M. Bonomi, and M. Parrinello, *Proceedings of the National Academy of Sciences* **110**, 6358 (2013).
- [14] A. Dickson and S. Lotz, *Biophysical Journal* **112** (2017).
- [15] P. Tiwary, J. Mondal, and B. J. Berne, *Science Advances* **3**, e1700014 (2017).
- [16] R. Casasnovas, V. Limongelli, P. Tiwary, P. Carloni, and M. Parrinello, *Journal of the American Chemical Society* **139**, 4780 (2017).
- [17] S. D. Lotz and A. Dickson, *Journal of the American Chemical Society* p. jacs.7b08572 (2018).
- [18] A. Dickson and C. L. Brooks III, *The Journal of Physical Chemistry B* **118**, 3532 (2014).
- [19] G. G. A. Huber and S. Kim, *Biophysical Journal* **70**, 97 (1996).
- [20] D. M. Zuckerman and L. T. Chong, *Annual Review of Biophysics* **46**, 43 (2017), pMID: 28301772, URL <http://www.annualreviews.org/doi/abs/10.1146/annurev-biophys-070816-033834>.
- [21] H. S. Muddana, C. D. Varnado, C. W. Bielawski, A. R. Urbach, L. Isaacs, M. T. Geballe, and M. K. Gilson, in *Journal of Computer-Aided Molecular Design* (2012), vol. 26, pp. 475–487, ISBN 1573-4951 (Electronic)\r0920-654X (Linking).
- [22] F. Biedermann and O. A. Scherman, *Journal of Physical Chemistry B* **116**, 2842 (2012).
- [23] H. Gan, C. J. Benjamin, and B. C. Gibb, *Journal of the American Chemical Society* **133**, 4770 (2011).
- [24] A. Dickson, A. Warmflash, and A. Dinner, *Journal of Chemical Physics* **131** (2009).

- [25] A. Dickson, M. Maienschein-Cline, A. Tovo-Dwyer, J. R. Hammond, and A. R. Dinner, Journal of Chemical Theory and Computation **7**, 2710 (2011).
- [26] E. Vanden-Eijnden and M. Venturoli, The Journal of Chemical Physics **131**, 044120 (2009).
- [27] E. Suárez, S. Lettieri, M. C. Zwier, C. A. Stringer, S. R. Subramanian, L. T. Chong, and D. M. Zuckerman, Journal of Chemical Theory and Computation **10**, 2658 (2014).
- [28] R. Costaouec, H. Feng, J. Izquierre, and E. Darve, Discrete and Continuous Dynamical Systems pp. 171–181 (2013).
- [29] T. L. Hill, *Free Energy Transduction and Biochemical Cycle Kinetics* (Springer, 1989), ISBN 978-1-4612-3558-3.
- [30] S. H. Northrup, S. A. Allison, and J. A. McCammon, Journal of Chemical Physics **80**, 1517 (1984).
- [31] A. Rojnuckarin, D. R. Livesay, and S. Subramaniam, Biophysical Journal **79**, 686 (2000).
- [32] A. S. Saglam and L. T. Chong, The Journal of Physical Chemistry B **120**, 117 (2015).