Journal of Computer-Aided Molecular Design manuscript No.
(will be inserted by the editor)

Predicting ligand binding affinity using on- and off-rates for the SAMPL6 SAMPLing challenge

Tom Dixon · Samuel D. Lotz · Alex Dickson

August 6, 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 ligand binding rates (k_{on}) and/or ligand unbinding rates (k_{off}) through explicit simulation of ligand binding pathways, and these methods work by very different mechanisms. 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_{off}/k_{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 maximizing trajectory variance. 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. Our rank ordering of the three host-guest pairs agrees with the reference calculations, however our predicted ΔG values were systematically lower than the reference 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.

Thomas Dixon · Samuel D. Lotz · Alex Dickson
Department of Biochemistry & Molecular Biology, Michigan State University, East Lansing, MI 48824, USA

Tom Dixon · Alex Dickson
Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI 48824

1 Introduction

Binding affinity 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 [1, 2]. The SAMPL challenge is playing an important role to compare tools that predict affinities using blind predictions [3]. 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 [4].

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 [5]. For instance, in many systems it has been observed that drug residence time ($RT = 1/k_{off}$) is the critical factor governing efficacy in living cells [6]. 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 the ligand binding pathway, particularly the ligand binding transition state, which is the highest point in free energy between the bound and unbound states [7]. The binding rate, k_{on}, has an upper bound of 10^9 M^{-1} 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} [8].

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 [9]. The trypsin-benzamidine system has served as a common benchmark application for enhanced sampling methods such as Adaptive Multilevel Splitting [10], SEEKR [11], adaptive [12] and traditional [13, 14] Markov state modeling, funnel metadynamics [15], as well as the WExplore method developed by our group [16]. Recently these efforts have been expanded to more challenging systems such as the unbinding of inhibitors from c-Src kinase [17] and p38 MAP kinase [18] using metadynamics, and the unbinding of the TPPU ligand from the target soluble epoxide hydrolase with WExplore [19]. 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. The SAMPLing challenge required participants to compute free energies as a function of simulation time, to compare the convergence properties and relative computational cost of different free energy calculation methods. Instead of computing free energies through alchemical perturbation, here achieve this by explicitly simulating the binding and release processes, determining the absolute rates k_{on} and k_{off}, and computing the binding affinity as the ratio $k_{\text{off}}/k_{\text{on}}$. We calculate the binding free energy using the binding affinity, or K_D by:

$$\Delta G = kT \ln(\frac{k_{off}}{C_0 k_{on}}) \tag{1}$$

where $kT = 0.597$ kcal/mol corresponding to a temperature of 300 K and C_0 is the reference concentration of 1 mol/L. 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 [20]. 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 [21], 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 [22]). REVO instead directs merging and cloning operations by maximizing trajectory variance 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.

2 Methods

2.1 Host-guest systems

The host-guest systems were selected from the main SAMPL6 challenge. One system is a cucurbit[8]uril (CB8) host [23, 24], 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, referred to as "Octa Acid" (OA), as a host [25]. 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.

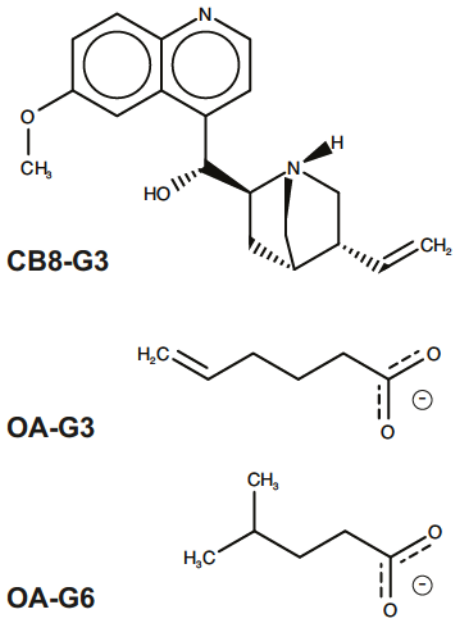

Fig. 1 Structure of the ligands used in this study. (Top) Quinine, referred to herein as CB8-G3. (Middle) 5-hexenoic acid (deprotonated form), referred to herein as OA-G3. (Bottom) 4-methyl pentanoic acid (deprotonated form), referred to herein as OA-G6.

2.2 Dynamics Setup

The fifteen initial configurations (five for each host-guest system) were used as prepared by the organizers of the SAMPLing challenge without modification. The two OA systems had a cubic box with a box length of 45 Å solvated with 2586 water molecules, and contained 12 sodium ions and 3 chloride ions to neutralize the system. The CB8 system had a cubic box with a box length of 42.5 Å solvated with 2149 water molecules, and contained 6 sodium ions and 6 chloride ions to neutralize the system. OpenMM v7.1.1 [26] 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}, a Monte Carlo barostat to keep pressure constant at 1 atm, and a timestep of 2 fs. The non-bonded forces, including electrostatics, had a cutoff of 1 nm, and were calculated using partial mesh ewald. The simulation temperature differs slightly from that used to calculate the reference free energies (298.15 K), although we expect the resulting differences in free energy will be negligible.

2.3 Reweighting of Ensembles by Variance Optimization

To encourage the sampling of rare events, we developed a method based on the Weighted Ensemble (WE) framework [21] that we call "Reweighting of Ensembles by Variance Optimization", or REVO. WE methods use an ensemble of trajectories (called "walkers") that are each assigned a statistical weight, and enhance sampling through the introduction of cloning and merging steps. Initially the weights of all the walkers are equal, and are defined as $1/N_{\text{walk}}$, where N_{walk} is the total number of walkers. When walkers are *cloned*, their weight is divided among the progeny. The cloned trajectories are identical replicas of the original, with the same atomic positions and velocities. This is typically done in under-sampled regions of space, in order to boost the probability of observing rare events in the simulation. Walkers 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". The trajectory weights are only changed due to merging and cloning operations.

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:

$$V = \sum_i V_i = \sum_i \sum_j (d_{ij})^p \omega_i \omega_j \tag{2}$$

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:

$$\omega_a = \log(100 * w_a / p_{\text{min}}) \tag{3}$$

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 (see also Fig. 2). Eq. 2 is used to

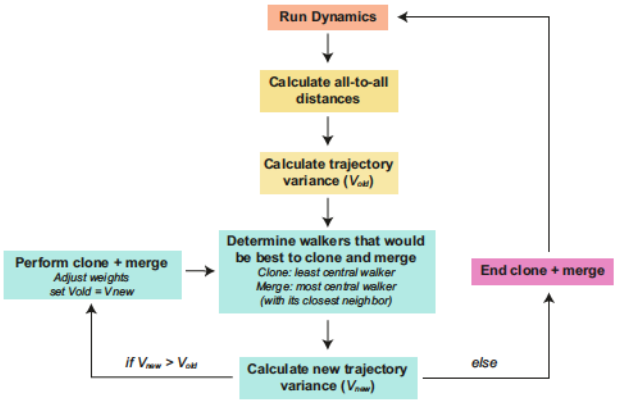

Fig. 2 The REVO algorithm. Each cycle begins by running an ensemble of walkers forward in time using unbiased dynamics. The distances between the walkers are used to calculate a variance (Eq. 2). In the resampling loop (blue), coupled cloning and merging operations are proposed, and they are accepted only if they result in a higher V. If the proposed V is lower, the resampling loop is terminated and dynamics are continued for the next cycle.

compute the variance function, and the walker with the highest ("H") and lowest ("L") contributions to the variance are identified (e.g. with the highest and lowest V_i values). The closest walker to "L" is identified, called "C". A coupled cloning and merging event is proposed, where "C" and "L" would be merged and "H" would be cloned. Eq. 2 is again used to recompute the variance, and this coupled cloning and merging move is only accepted if V increases. Further moves are proposed after recomputing "H", "L" and "C", and the process continues until V decreases, and the 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 [19], a minimum and maximum walker weight was enforced ($p_{\min}$ and $p_{\max}$, respectively). Note that only walkers which will not violate the walker probability boundaries (p_{min} and p_{max}) are eligible to be chosen as walkers "H", "L" and "C". 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^U_{ij} is defined as the root mean square deviation (RMSD), in Å, of the guest ligand between structures i and j, after aligning to the host. As mentioned in Section 2.1, 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^U_{ij}. For rebinding, d^R_{ij} is computed using the RMSD of both i and j to the reference starting structure:

$$d^R_{ij} = \left|1/d^U_{i0} - 1/d^U_{j0}\right| \tag{4}$$

where d^U_{a0} 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).

2.4 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" [27, 28] (also referred to as "tilting" [29], or "coloring" [30, 31]). An equilibrium ensemble is split into two non-equilibrium ensembles by defining two basins, in this case the "bound" basin and the "unbound" basin (Fig. 3). 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. The structure recorded before warping is known as an exit point. This means the atomic coordinates and velocities of the trajectory return to the starting structure. The weight of the trajectory does not change as a result of warping. 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 [32, 30]) that leave the ensemble:

$$k_{\text{off}} = \frac{\sum_i w^U_i}{T} \tag{5}$$

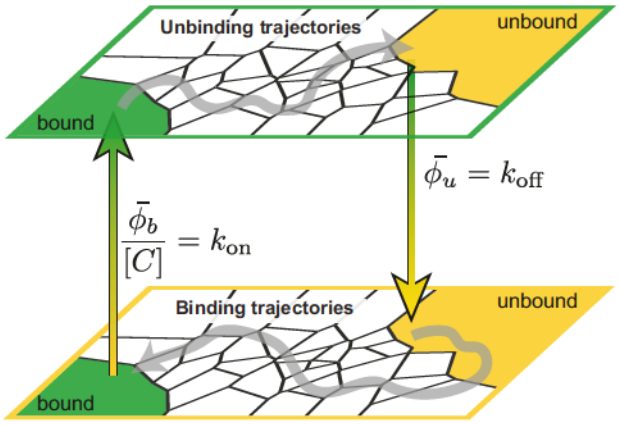

Fig. 3 Ensemble splitting. An equilibrium host-guest binding system is split into two nonequilibrium ensembles for the calculation of on and off rates. This is done by defining "bound" and "unbound" basins (left and right of each ensemble). The "unbinding" ensemble (top) is the set of trajectories that have most recently visited the bound basin. The "binding" ensemble (bottom) is the set of trajectories that most recently visited the unbound basin. The on and off rates are directly computed using the time averaged trajectory flux ($\bar{\phi}_b$ or $\bar{\phi}_u$) between the ensembles.

$$k_{\text{on}} = \frac{\sum_i w_i^R}{CT} \tag{6}$$

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.

2.5 REVO simulation details

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. Five simulations were run from each starting pose (see Figure S1-S3 for starting poses). In aggregate, we ran 1.92 μs for each of the unbinding and rebinding simulations, 3.84 μs for each starting pose, or 57.6 μs over the entire set of results presented here.

2.5.1 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 $p_{\text{max}} = 0.1$, 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 wegiht 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 1. The calculated mean first passage time (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 μs), again indicating why a group of straightforward trajectories was able to produce over two dozen binding events each.

2.6 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 as can be seen in Figure 7. 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:

$$E = \sum_i \left[b(x_i^t - x_i^{t-1})^2 + cw_i(x_i^t)^2 + \sum_j E_{ij}^{rep} e^{\frac{(x_i^t - x_j^t)^2}{r_o}} \right] \tag{7}$$

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}^{rep}) that takes into account the molecular distance between the walkers in the simulation (d_{ij}):

$$E_{ij}^{rep} = a * \max(0, d_{ij} - d_0) \tag{8}$$

where a and d_0 are parameters set here to 2.5 and 2.0. d_{ij} can refer to either d_{ij}^U or d_{ij}^R when making trees of unbinding and rebinding simulations respectively. However, only trees generated from unbinding simulations are shown here. The parameters for Eq. 7 were selected to keep the branches generated by cloning events in the same region on the tree, as well as to keep larger weighted walkers towards the center. 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 [33] and visualized using Gephi 0.9.2 [34].

2.7 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 [35]. The clustering was done on a featurized space defined by the Canberra distance defined as:

$$D(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} \frac{|p_i - q_i|}{|p_i| + |q_i|} \tag{9}$$

where $\mathbf{p}$ and $\mathbf{q}$ are vectors at different frames containing the distances between all the guest atoms and all the host atoms and n is the total number of distance pairs. These vectors contain 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 using CSNAnalysis [36]. 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 [37]. 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 $w_{ij} = 100 p_{ij}$ 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. A CSN of each system is shown in Figure 8.

3 Results

3.1 Warped walkers

For each host-guest system we run both unbinding and rebinding REVO simulations originating from five different starting poses (Figure S1-S3), making 30 simulations total. All of these REVO simulations generated a substantial number of warping events. In general these are distributed across a wide range of weight values (Figure 4). 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 4, 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.

The warping points for the unbinding simulation are shown in Figure 5, 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.

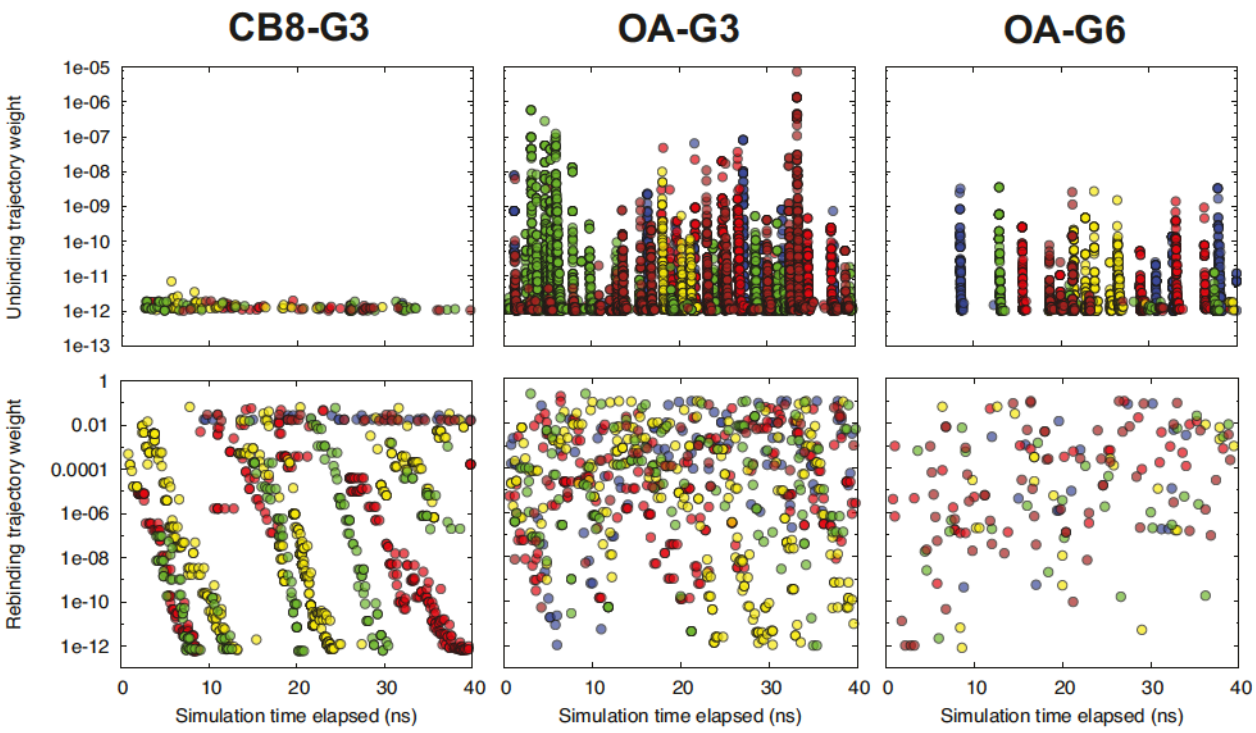

Fig. 4 Weights of warped walkers. Weights of warping events for the unbinding (top row) and rebinding (bottom row) simulations. In both cases the points are colored according to the index of the corresponding starting pose (0, blue; 1, red; 2, yellow; 3, green; 4, brown).

3.2 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 $C = \frac{1}{N_A V}$, 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 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 6, 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 1. The uncertainties of k_{on} and k_{off} (δ_{on} and δ_{off}) are determined using the standard error across the five trajectories. The uncertainties in the mean first passage times (MFPT) of binding and unbinding are calculated as $\delta_{\text{on}}/Ck_{\text{on}}^2$ and $\delta_{\text{off}}/k_{\text{off}}^2$, respectively. Finally, the uncertainty in ΔG is as follows:

$$\delta_{\Delta G} = kT\sqrt{\left(\frac{\delta_{\text{off}}}{k_{\text{off}}}\right)^2 + \left(\frac{\delta_{\text{on}}}{k_{\text{on}}}\right)^2}. \tag{10}$$

The 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 860 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, resulting in an acceleration factor of $\approx 9 \times 10^7$.

With k_{off} and k_{on} in hand, the binding affinity is calculated using Eq. 1. This binding affinity is compared to both the experimentally measured binding affinity [38], and a computational reference computed using alchemical free energy calculations with YANK (see [4] for more details). As shown in Figure 6, 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

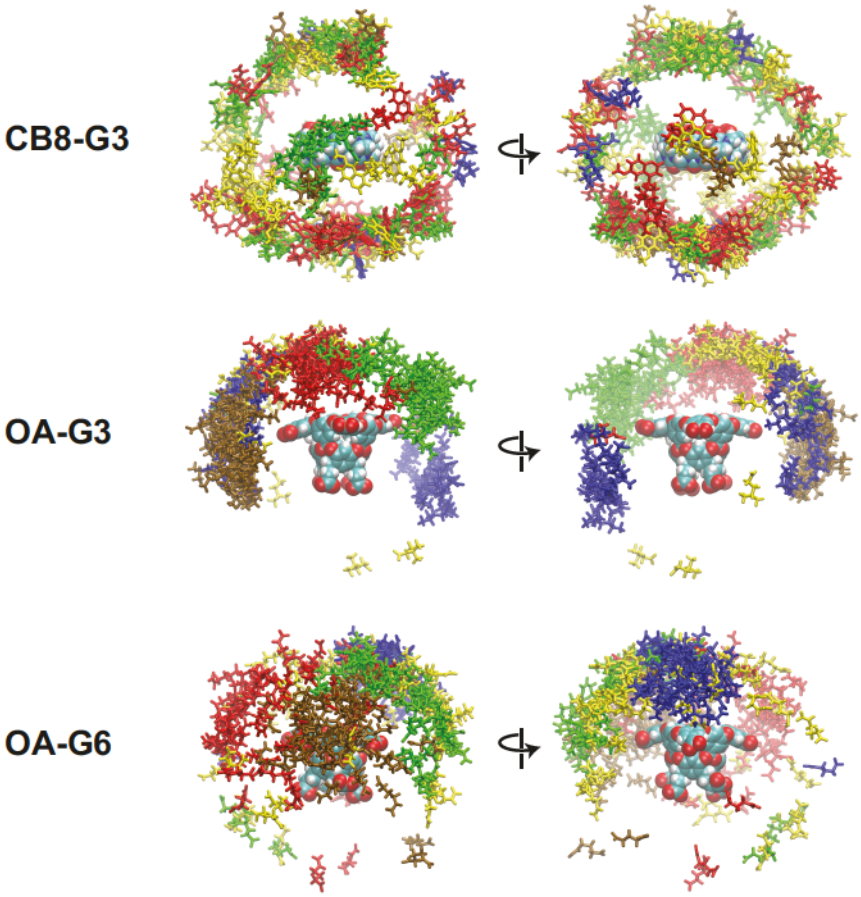

Fig. 5 Spatial distribution of warped walkers. Structures of warping events for the unbinding simulations viewed from the front and back. Guest ligands are colored according to the index of the corresponding starting pose (0, blue; 1, red; 2, yellow; 3, green; 4, brown).

Table 1 Pose-averaged rates and affinities

	k_{off} (s^{-1})	$MFPT_{off}$ (s)	k_{on} (s^{-1} M^{-1})	$MFPT_{on}$ (ns)	ΔG (calc.)	ΔG (ref.) [4]	ΔG (exp.) [38]
CB8-G3	0.0012 ± 0.0003	860 ± 230	$4.7 \pm 0.8 \times 10^8$	92 ± 16	-16.0 ± 0.2	-10.90 ± 0.16	-6.45 ± 0.06
OA-G3	160 ± 110	0.0064 ± 0.0044	$1.2 \pm 0.2 \times 10^9$	36 ± 6	-9.5 ± 0.4	-6.70 ± 0.02	-5.18 ± 0.02
OA-G6	0.48 ± 0.11	2.1 ± 0.5	$2.8 \pm 1.0 \times 10^8$	150 ± 50	-12.1 ± 0.3	-7.18 ± 0.05	-4.97 ± 0.02

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 [9, 19], where rates from individual simulations varied over several orders of magnitude.

3.3 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.

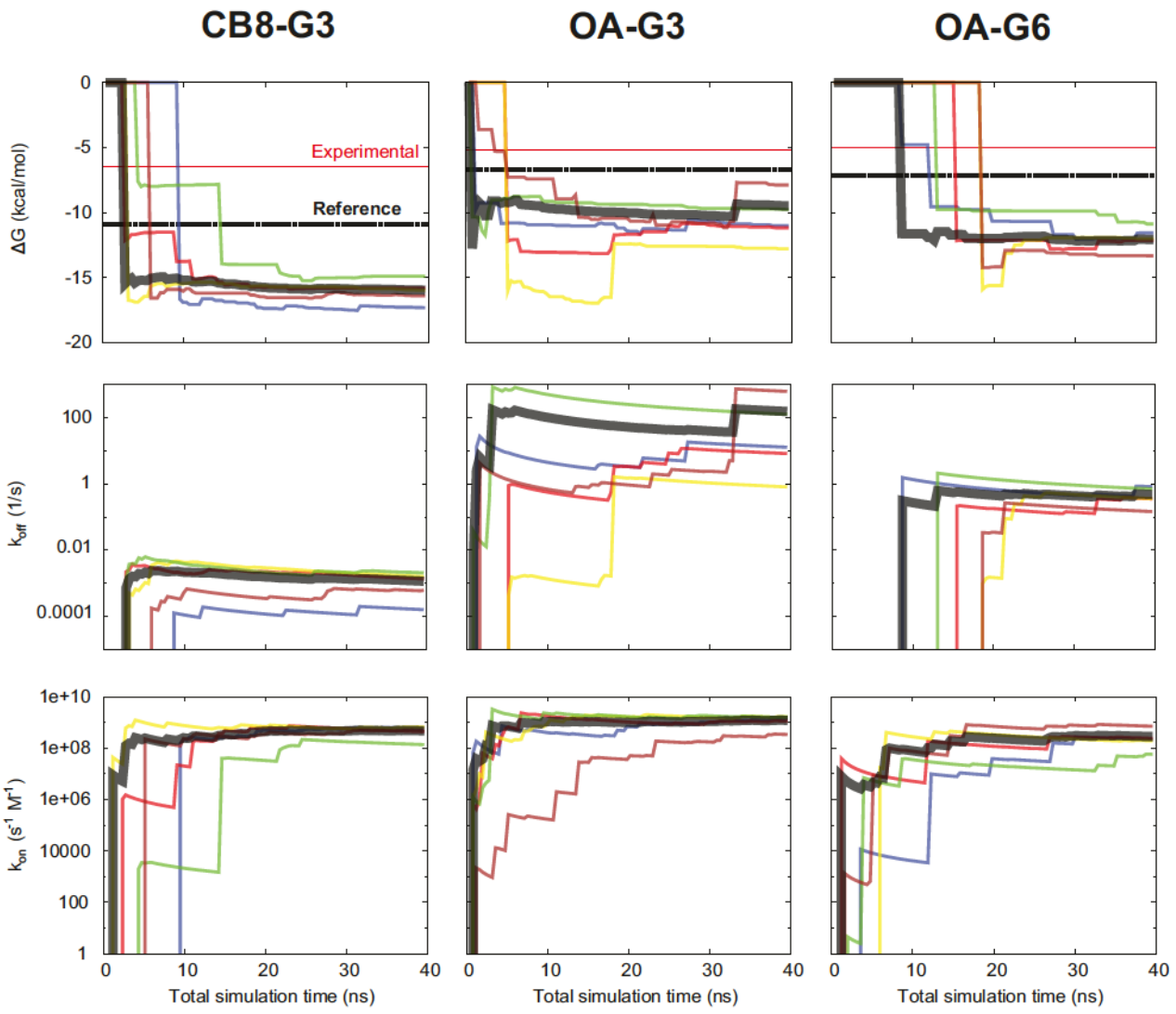

Fig. 6 Predicted kinetics and free energies. The calculated free energies (top), off-rates (middle), and on-rates (bottom) are shown as a function of simulation time for each starting pose in each host-guest system. The curves are colored according to the index of the starting pose as in Figures 5 and 4. The calculated binding free energies are compared with experimental measurements (horizontal red line) [38], and the computational reference (dashed black line) for each system.

Figure 8 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 7, 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 ($\approx$ 320, orange), to medium ($\approx$ 160, light green), back to high again. This behavior results from

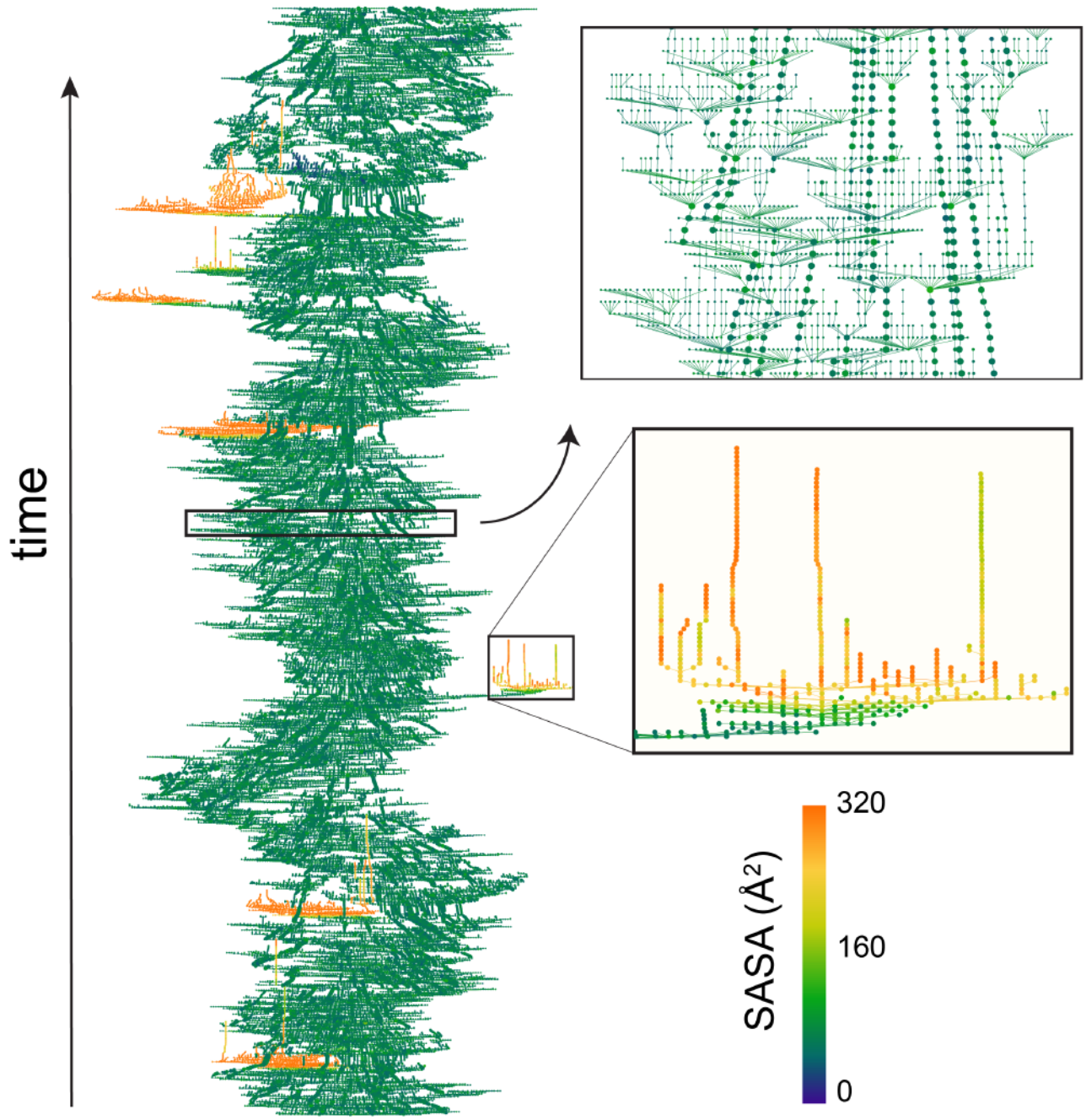

Fig. 7 Trajectory trees show all cloning and merging events in a simulation. The trajectory tree for the first 1329 cycles of the OA-G3-0 unbinding simulation is shown. Each horizontal row in this tree represents a cycle, and the placement of all 48 nodes in the row is determined by minimizing an energy 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.

a transient, loose association with the exterior of the host molecule. This trajectory is shown as Supporting Movie 1.

3.4 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 8 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 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.

4 Discussion

Although we obtained much information about the binding and release processes of these host-guest systems, our predicted ΔG values were systematically lower than those of a reference calculation employing the same forcefield (average 4.2 kcal/mol). These reference ΔG 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 μs 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 [39]. 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 [40, 41], and the SEEKR method (Simulation Enabled Estimation of Kinetic Rates) [11].

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 OA-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 difficul-

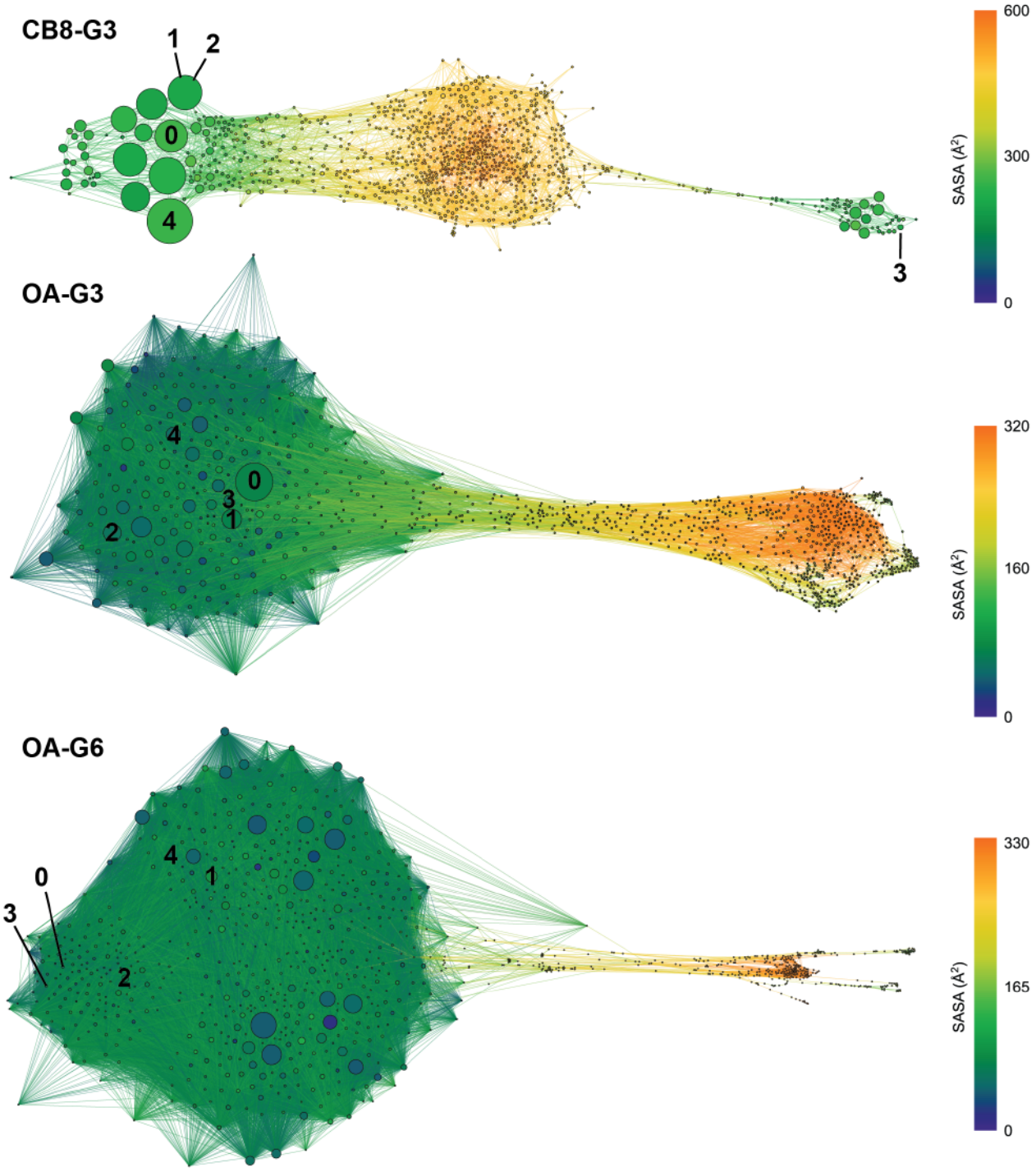

Fig. 8 Conformation space networks for the unbinding simulations. Each node in a CSN represents a cluster of host-guest structures. Edges in the networks connect clusters that are seen to interconvert in the REVO simulations. The size of each node is proportional to the number of times it was observed 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.

ties 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.

References

1. De Ruiter A, Oostenbrink C (2011) Free energy calculations of protein-ligand interactions. Curr Opion Chem Biol 15(4):547–552
2. Gapsys V, Michielssens S, Peters JH, de Groot BL, Leonov H (2015) Calculation of binding free energies. Methods in Mol Bio (Clifton, NJ) 1215:173–209
3. Geballe MT, Skillman AG, Nicholls A, Guthrie JP, Taylor PJ (2010) The SAMPL2 Blind Prediction Challenge: Introduction and Overview. J Comput Aided Mol Des 24(4):259–279
4. Rizzi A, Chodera J, Mobley D, Jensen T, Shirts M, Dixon T, Lotz S, Dickson A, Michel J, Gilson M, Walker R (2018) The SAMPL6 SAMPLing Challenge: Assessing the Reliability, Reproducibility, and Efficiency of Free Energy Calculations. J Comput Aided Mol Des p To appear
5. Copeland RA (2016) The Drug-Target Residence Time Model: A 10-Year Retrospective. Nat Rev Drug Discovery 15(February):87–95
6. Costa B, Da Pozzo E, Giacomelli C, Barresi E, Taliani S, Da Settimo F, Martini C (2016) TSPO Ligand Residence Time: A New Parameter to Predict Compound Neurosteroidogenic Efficacy. Sci Rep 6(August 2015):18164
7. Pan AC, Borhani DW, Dror RO, Shaw DE (2013) Molecular Determinants of Drug-Receptor Binding Kinetics. Drug Discov Today 18(13-14):667–673
8. Kokh DB, Amaral M, Bomke J, Grädler U, Musil D, Buchstaller HP, Dreyer MK, Frech M, Lowinski M, Vallée F, Bianciotto M, Rak A, Wade RC (2018) Estimation of Drug-Target Residence Times by T-Random Acceleration Molecular Dynamics Simulations. J Chem Theory Comput 14(7):3859–3869
9. Dickson A, Tiwary P, Vashisth H (2017) Kinetics of Ligand Binding Through Advanced Computational Approaches: A Review. Curr Top Med Chem 17:2626–2641
10. Teo I, Mayne CG, Schulten K, Lelièvre T (2016) Adaptive Multilevel Splitting Method for Molecular Dynamics Calculation of Benzamidine-Trypsin Dissociation Time. J Chem Theory Comput 12(6):2983–2989
11. Votapka LW, Jagger BR, Heyneman AL, Amaro RE (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(15):3597–3606
12. Doerr S, De Fabritiis G (2014) On-the-Fly Learning and Sampling of Ligand Binding by High-Throughput Molecular Simulations. J Chem Theory Comput 10(5):2064–2069
13. Buch I, Giorgino T, De Fabritiis G (2011) Complete Reconstruction of an Enzyme-Inhibitor Binding Process by Molecular Dynamics Simulations. Proc Natl Acad Sci USA 108(25):10184–10189
14. Plattner N, Noé F (2015) Protein Conformational Plasticity and Complex Ligand-Binding Kinetics Explored by Atomistic Simulations and Markov Models. Nat Commun 6(May):7653
15. Limongelli V, Bonomi M, Parrinello M (2013) Funnel Metadynamics as Accurate Binding Free-Energy Method. PNAS 110(16):6358–6363
16. Dickson A, Lotz S (2017) Multiple Ligand Unbinding Pathways and Ligand-Induced Destabilization Revealed by WExplore. Biophys J 112(4)
17. Tiwary P, Mondal J, Berne BJ (2017) How and When Does an Anticancer Drug Leave Its Binding Site? Sci Adv 3:e1700014
18. Casasnovas R, Limongelli V, Tiwary P, Carloni P, Parrinello M (2017) Unbinding Kinetics of a P38 MAP Kinase Type II Inhibitor from Metadynamics Simulations. J Am Chem Soc 139(13):4780–4788
19. Lotz SD, Dickson A (2018) Unbiased Molecular Dynamics of 11 Min Timescale Drug Unbinding Reveals Transition State Stabilizing Interactions. J Am Chem Soc 140(2):618–628.
20. Dickson A, Brooks III CL (2014) WExplore: Hierarchical Exploration of High-Dimensional Spaces Using the Weighted Ensemble Algorithm. J Phys Chem B 118(13):3532–3542
21. Huber GGA, Kim S (1996) Weighted-Ensemble Brownian Dynamics Simulations for Protein Association Reactions. Biophys J 70(1):97–110
22. Zuckerman DM, Chong LT (2017) Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. Annu Rev Biophys 46(1):43–57, DOI 10.1146/annurev-biophys-070816-033834, URL https://doi.org/10.1146/annurev-biophys-070816-033834
23. Muddana HS, Varnado CD, Bielawski CW, Urbach AR, Isaacs L, Geballe MT, Gilson MK (2012) Blind Prediction of Host-Guest Binding Affinities: A New SAMPL3 Challenge. In: J Comput Aided Mol Des, vol 26, pp 475–487

24. Biedermann F, Scherman OA (2012) Cucurbit[8]uril Mediated Donor-Acceptor Ternary Complexes: A Model System for Studying Charge-Transfer Interactions. J Phys Chem B 116(9):2842–2849
25. Gan H, Benjamin CJ, Gibb BC (2011) Nonmonotonic Assembly of a Deep-Cavity Cavitand. J Am Chem Soc 133(13):4770–4773
26. Eastman P, Swails J, Chodera JD, McGibbon RT, Zhao Y, Beauchamp KA, Wang LP, Simmonett AC, Harrigan MP, Stern CD, Wiewiora RP, Brooks BR, Pande VS (2017) OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. PLoS Comput Biol DOI 10.1371/journal.pcbi.1005659
27. Dickson A, Warmflash A, Dinner A (2009) Separating Forward and Backward Pathways in Nonequilibrium Umbrella Sampling. J Chem Phys 131(15)
28. Dickson A, Maienschein-Cline M, Tovo-Dwyer A, Hammond JR, Dinner AR (2011) Flow-Dependent Unfolding and Refolding of an RNA by Nonequilibrium Umbrella Sampling. J Chem Theory Comput 7:2710–2720
29. Vanden-Eijnden E, Venturoli M (2009) Exact Rate Calculations by Trajectory Parallelization and Tilting. J Chem Phys 131(4):044120
30. Suárez E, Lettieri S, Zwier MC, Stringer CA, Subramanian SR, Chong LT, Zuckerman DM (2014) Simultaneous Computation of Dynamical and Equilibrium Information Using a Weighted Ensemble of Trajectories. J Chem Theory Comput 10(7):2658–2667
31. Costaouec R, Feng H, Izaguirre J, Darve E (2013) Analysis of the accelerated weighted ensemble methodology. DCDS pp 171–181
32. Hill T (1989) Free Energy Transduction and Biochemical Cycle Kinetics. Academic Press
33. Hagberg AA, Schult DA, Swart PJ (2008) Exploring network structure, dynamics, and function using NetworkX. Proceedings of the 7th Python in Science Conference (SciPy 2008)
34. Bastian M, Heymann S, Jacomy M (2009) Gephi: An Open Source Software for Exploring and Manipulating Networks. 3rd ICWSM DOI 10.1136/qshc.2004.010033
35. Harrigan MP, Sultan MM, Hernández CX, Husic BE, Eastman P, Schwantes CR, Beauchamp KA, McGibbon RT, Pande VS (2017) MSMBuilder: Statistical Models for Biomolecular Dynamics. Biophys J 112(1):10–15, DOI 10.1016/j.bpj.2016.10.042
36. Dickson AR (2018) CSNAnalysis. URL https://github.com/ADicksonLab/CSNAnalysis
37. Cherven K (2013) Network Graph Analysis and Visualization with Gephi. Packt Publishing
38. Murkli S, McNeill JN, Isaacs L (2018) Cucurbit[8]uril-Guest Complexes: Blinded Dataset for the SAMPL6 Challenge. Supramol Chem p submitted
39. Northrup SH, Allison SA, McCammon JA (1984) Brownian Dynamics Simulation of Diffusion-Influenced Bimolecular Reactions. J Chem Phys 80(4):1517–1524
40. Rojnuckarin A, Livesay DR, Subramaniam S (2000) Bimolecular Reaction Simulation Using Weighted Ensemble Brownian Dynamics and the University of Houston Brownian Dynamics Program. Biophys J 79(2):686–693
41. Saglam AS, Chong LT (2015) Highly Efficient Computation of the Basal kon Using Direct Simulation of Protein-Protein Association with Flexible Molecular Models. J Phys Chem B 120:117–122