

Unbiased Molecular Dynamics of 11 min Timescale Drug Unbinding Reveals Transition State Stabilizing Interactions

Samuel D Lotz, Alex Dickson

October 13, 2017

1 Abstract

Ligand (un)binding kinetics is being recognized as a determinant of drug specificity and efficacy in an increasing number of systems. However, the calculation of kinetics and the simulation of drug unbinding is more difficult than computing thermodynamic quantities, such as binding free energies. Here we present the first full simulations of an unbinding process at pharmacologically relevant timescales (11 min), without the use of biasing forces, detailed prior knowledge, or specialized processors using the weighted ensemble based algorithm, WExplore. These simulations show the inhibitor TPPU unbinding from its enzyme target soluble epoxide hydrolase (sEH), which is a clinically relevant target that has attracted interest in kinetics optimization in order to increase efficacy. We make use of conformation space networks (CSNs) that allow us to conceptualize unbinding not just as a linear process, but as a network of interconnected states that connect the bound and unbound states. This allows us to visualize patterns in hydrogen-bonding, solvation, and non-equilibrium free energies, without projection onto progress coordinates. The topology and layout of the network reveal multiple unbinding pathways, and other rare events, such as the reversal of ligand orientation within the binding site. Furthermore, we make a prediction of the transition state ensemble, using transition path theory, and identify protein-ligand interactions which are stabilizing to the transition state. Additionally, we uncover trends in ligand and binding site solvation that corroborate experimental evidence from more classical structure kinetics relationships (SKRs) and generates new questions as to the role of drug modifications in kinetics optimization. Finally, from only 6 μ s of simulation time we observed 75 unbinding events from which we calculate a residence time of 42 s, and a standard error range of 23 s to 280 s. This nearly encompasses the experimental residence time 11 min (660 s). In addition to the insights to sEH inhibitor unbinding, this study shows that simulations of complex processes on timescales as long as minutes are becoming feasible for more researchers to perform.

2 Introduction

The dominant objective in drug design and discovery has historically been to improve drug efficacy through increasing the affinity of drug binding [1]. This has been an effective strategy and

has been extensively addressed computationally through the calculation of binding free energies [2–5]. Thermodynamic drug affinity (K_D) is the "driving force" of ligand binding in living non-equilibrium systems, but a full description of drug action requires kinetic information as well. This need has been recognized and an emerging subdiscipline has begun to focus on optimizing binding and unbinding rates (k_{on} and k_{off} respectively) [1, 6–8]. Of particular interest is the residence time (RT) ($1/k_{\text{off}}$) which has been shown to be particularly good at predicting drug efficacy for some targets [9–12]. Higher RTs, coupled to an appropriate dosing strategy, can be an effective way to protect the drug from being metabolized or eliminated by the body and increase receptor occupancy over longer periods of time [10, 13–15].

However, the measurement and prediction of kinetics presents a much bigger challenge than binding affinity. Rates depend on the free-energy of transition states, which are very short lived, making them nearly impossible to observe through x-ray crystallography or NMR. Technology for running medium-throughput binding kinetics assays (e.g. surface plasmon resonance (SPR) and radioligand binding assays) exist [6, 16], but these are expensive and difficult to execute [17]. Assays for membrane proteins, which are the majority of drug targets (i.e. G-protein coupled receptors (GPCRs)), are even more difficult [6]. Additionally, obtaining kinetics from labelled assays rely on mathematical models [18] that make assumptions on mass action and ligand rebinding, which in practice are hard to satisfy and may affect final rates. Furthermore, experimental kinetics assays can only deliver structural information on the transition state through structure kinetic relationships (SKRs) from a panel of ligands, and requires a statistical model. Molecular dynamics (MD) simulations of ligand (un)binding processes have the potential to both elucidate the structure of transition states and to calculate rates *ab initio*. MD simulations are able to observe these transient states, however they are typically limited to a microsecond timescale [19] – far from the timescales needed to simulate many ligand unbinding processes, which can be as long as minutes to hours [20]. However, with the use of specialized hardware [21, 22] and enhanced sampling algorithms [23–26] these timescales are becoming approachable.

Until recently most simulations of ligand unbinding were limited to model systems, such as small solute-like molecules unbinding from the FK506 binding protein (FKBP) [27–30, 30] and the protease inhibitor benzamidine binding to trypsin [23, 31–36]. Small ligands studied for FKBP have low affinities (250 μM to 20 000 μM) and residence times in the 10^{-9} s timescale, and are useful for testing sampling algorithms or off rate calculations. Benzamidine is a more drug-like ligand with a residence time of 1.7×10^{-3} s on the serine-protease, trypsin [37]. For this reason trypsin-benzamidine is a popular model system for studying ligand unbinding processes and for the development of new sampling algorithms. The use of the MD supercomputer, Anton, has enabled unbiased binding simulations of dasatinib to a kinase receptor (a 200 ms timescale event), but report only 1 to 3 such events using 6 copies of the ligand in 32.5 μs of sampling [21]. Enhanced sampling with metadynamics, however, was able to observe 12 unbinding events of dasatinib, each taking 150 ns to 750 ns of MD [26]. Other enhanced sampling techniques were used to reveal unbinding trajectories for a family of p38 binding drugs with residence times of approximately 7 s including metadynamics [23, 24] and umbrella sampling [25]. These methods (and many other enhanced sampling methods) all require the use of carefully selected biasing potentials and/or progress coordinates to describe the unbinding process. For some methods, correction factors have been developed to recover rates and unbiased free-energy profiles, but make assumptions on the form of the potential energy landscape at the transition state [38]. Suitable progress coordinates for the unbinding process are typically limited to 1-2 degrees of freedom and often

require *a priori* knowledge of the unbinding mechanism [38], or existing sampling to optimize order parameters [39]. There exist some methods that do not require carefully selected progress coordinates, such as temperature accelerated molecular dynamics (TAMD) [40, 41], which have been used, for instance to simulate ligand unbinding from the Adenosine A_{2A} receptor on a residence timescale of 5.04×10^3 s (84 min) on the Anton supercomputer [22]. However, the rates and conformations obtained must be somehow reconstructed from simulations run in an ensemble of different temperatures. We also note that if one is only interested in residence times, there are more efficient methods to predict them in either a relative, or an absolute fashion. Scaled molecular dynamics has shown promise in this regard [42, 43], as well as steered molecular dynamics [44].

Simulations run with the Weighted Ensemble (WE) sampling method [45] do not require biasing potentials or correction factors, instead sampling rare events by cloning trajectories in low probability regions of conformation space. Recently, the WExplore algorithm [46] has been used to successfully simulate ligand unbinding in the FKBP [29] and trypsin [36] test systems. WExplore results for trypsin uncovered a distinct unbinding pathway that was not observed by other methods [36] showcasing the ability of this method to broadly sample many diverse unbinding mechanisms without the use of prior knowledge or intuition.

In this study we present unbiased pathways of a drug development intermediate (1-trifluoromethoxyphenyl-3-(1-propionylpiperidin-4-yl)-urea, or TPPU) unbinding from its receptor, soluble epoxide hydrolase (sEH), using WExplore. TPPU forms several stable interactions with residues that are buried within the sEH protein. The unbinding of TPPU has an experimentally determined residence time of 11 min. With WExplore we generate a series of unbinding trajectories using unbiased dynamics, and construct a portrait of the ligand-protein free-energy landscape through a conformation space network (CSN). We extensively profile ligand-protein hydrogen bonds throughout the network, and use these to develop an understanding of distinct steps in the unbinding mechanism. Furthermore, we identified the likely ensemble of states surrounding the unbinding transition state (TS) and begin to investigate potential TS stabilizing interactions.

3 Methods

3.1 Molecular Dynamics

Initial atomic coordinates were taken from the Protein Data Bank (PDB) entry 4OD0 [15], which has two domains with few inter-domain contacts and are connected by a flexible linker. We removed the domain which does not bind to the ligand of interest, including the associated crystallographic waters, magnesium ion, and phosphate ion, retaining residues Ser231 to Arg546 (PDB serial numbers) and 10 crystallographic waters. To solvate the system, a box of TIP3P water molecules were placed around the protein and ligand with a 12 Å cutoff. The completed system included 316 amino acids (5052 atoms) for the protein, 45 ligand atoms, 16 831 water molecules, and 7 neutralizing sodium ions for a total of 21 935 atoms.

The protein is parametrized by the CHARMM36 forcefield [47] and the ligand forcefield was derived using the homology based CHARMM Generalized Force Field (CGENFF) algorithm [48]. Before dynamics the solvent molecules are minimized using 500 steps of steepest descent followed

by 500 steps of the adopted Newton-Raphson method. The whole system is then minimized in the same way.

Molecular dynamics are run using the CHARMM program [49] with an OpenMM (v6.3) [50] interface to allow use of graphics processing units (GPUs). Timesteps of 2 fs are calculated in the constant pressure and temperature ensemble by coupling to a Langevin heatbath, temperature of 300 K and a friction coefficient of 1 ps^{-1} , and a Monte Carlo barostat with reference pressure at 1 atm where volume moves attempted every 50 timesteps. Nonbonded interactions are calculated using the particle mesh Ewald method with a switching function that scales interactions to zero from 8.5 Å to 10 Å. We start the heatbath at 50 K and run dynamics for 10 ps before increasing the temperature in ten even steps to 300 K. The resulting structure is then equilibrated at 300 K for 500 ps before being used in WExplore simulations.

3.2 WExplore Simulations in the Non-equilibrium Unbinding Ensemble

The WExplore algorithm is explained in Ref. [46] and has been used successfully to generate ligand unbinding pathways [29, 36]. Briefly, WExplore is a variant of the WE algorithm that runs an ensemble of trajectories (walkers) concurrently, and balances sampling over a set of regions by cloning walkers in under-represented regions and merging walkers in over-represented regions. This encourages broad sampling of conformations, including high free-energy transition states that separate stable states of interest, such as the bound and unbound states of a ligand. The key advance of the WExplore algorithm is the hierarchical organization of sampling regions. This allows for the use of a small number of walkers while still dividing a high-dimensional space into a large number of regions.

The distinguishing feature of WExplore is that it does not bias the potential energy function and does not require prior knowledge about the mechanism. All that is required is a means of measuring the distance between two conformations of the system. It improves efficiency if this distance does not include features that are irrelevant to the process of interest. Here the distance is calculated by first aligning the ligand binding site atoms (those within 8 Å of the ligand in the crystal structure) and then calculating the ligand RMSD without further orientation. This distance metric captures the translational and rotational degrees of freedom of the ligand, which does not predispose sampling to any particular unbinding path.

The sampling regions are Voronoi polyhedra (VP), which are dynamically defined by central "images" as the simulation progresses. Specifically, the distance from the walker to each image is computed, and a new image is defined when the closest such distance is larger than a cutoff distance (d) defined for each level of the hierarchy. In our simulations we use four levels of a hierarchy with cutoffs $d = 10 \text{ \AA}, 5 \text{ \AA}, 3.5 \text{ \AA}$ and 2.5 \AA . Cloning and merging steps occur every $\tau = 20 \text{ ps}$ between the set of 48 walkers. As in previous work [36], we enforce a maximum weight of 0.1 and a minimum weight of 10^{-12} by disallowing cloning and merging steps that would violate these conditions. Each separate run of WExplore was run with 48 replicas and was run for 1 μs of simulation time. The parameters used here, including d , τ , the number of walkers, minimum and maximum weights, and the definition of the unbound state, are identical to those used previously for the trypsin-benzamidine system [36], which emphasizes the generality of the WExplore approach.

Simulations are run such that trajectories start in the bound state (source; A) and end in the unbound state (sink; B). The source state is the initial bound pose and the sink is the set of

conformations in which the shortest protein-ligand distance is greater than 10 Å. Walkers that have crossed the transition barrier normally decay very quickly to this absorbing boundary at the free-energy minima of the unbound state and thus contribute little probability to this portion of the state space, and consequently have very high free-energies. Thus the free energies calculated from the weights in our simulations are not the more familiar equilibrium free-energies. The equilibrium free-energy can be recovered if the non-equilibrium binding ensemble is also calculated and combined with the unbinding ensemble, which has not been done here. For our purposes the unbinding ensemble is sufficient for determining the structure of the transition state, and subsequently calculating the unbinding rate.

3.3 Residence Time Calculation

A trajectory that reaches the sink (B) is called a reactive trajectory γ and has probability $P(\gamma)$ equal to the weight of the reactive walker. In these simulations, $P(\gamma)$ values are very small due to the low probability of unbinding and therefore is functionally equivalent to a nonequilibrium ensemble, even though trajectories are not explicitly reinitialized in the source state. To calculate the actual RT (equivalent to the mean first passage time (MFPT) of unbinding) we assume that our nonequilibrium ensemble is in steady state and use the so-called Hill relation which states that the MFPT is equal to the inverse of the probability flux into the sink state [51–54]. The flux measurement at time t for a single run is calculated by taking the sum of the weighted passage times:

$$\frac{1}{RT(t)} = \text{Flux}_{A \rightarrow B}(t) = \frac{\sum_{\gamma \in \Gamma(t)} P(\gamma)}{t}$$

Where $\Gamma(t)$ is the collection of all reactive trajectories that have occurred before time t . We then compute an average over the rates of all runs to get our final estimate for the rate and invert to get the residence time.

3.4 Clustering and Network Visualization

To reduce the dimensionality of the sampled trajectories and to summarize the contained dynamical information we create a conformation space network (CSN). We first cluster our data according to a feature vector consisting all possible distances between TPPU atoms 2, 5, 10, 13, 15, 22, and 24 (see Fig. 1 for numbering) and all non-hydrogen protein atoms within 8.0 Å of any of those atoms in the bound crystal structure. A total of 2478 atom pairs were included in the MSMBuild [55] featurizer (see Supplemental Information (SI) for details). The KCenters clustering algorithm in MSMBuild [55] was used to create 2000 clusters with a Canberra distance metric to avoid a proliferation of unbound clusters. Edges in the CSN were defined between states that are connected in the sampling trajectories, and given a strength equal to the number of transitions observed between the nodes in either direction.

Network visualization was performed in Gephi version 0.9.1 [56] using the ForceAtlas2 layout algorithm [57], followed by a short minimization to prevent the overlap of nodes. The node sizes are proportional to the weight of each cluster, with a minimum and maximum node size chosen for clear visualization. In CSN plots the nodes and edges can be colored in order to highlight different properties calculated for the clusters. The network layout (node positions and edge

lengths) is not quantitatively reproducible and layouts from separate minimizations may differ slightly due to the stochasticity of the ForceAtlas2 algorithm. However, the network layout is a useful tool for data exploration, allowing us to visualize the entire free energy landscape explored by our simulations in a single plot, and giving insight into the heterogeneity of transition paths connecting the bound and unbound states.

3.5 Interaction Profiling

We profile the frequency of hydrogen bonds across the set of clusters using a single structure from each cluster that was randomly chosen. Possible donor and acceptor atoms in both the protein and ligand are detected by RDKit [58]. A hydrogen bond is counted between a given donor-acceptor pair if the distance between the atoms is less than 4.1 Å and the donor-hydrogen-acceptor angle is within 100° to 180° [59]. To achieve this we use the software Mastic [60], which we have developed, and is parallelized with the SCOOP library [61], which uses the ZeroMQ message passing protocol. See SI for further details and usage.

3.6 Committor Probabilities and Highest Flux Pathways

Using Transition Path Theory (TPT) we estimate the highest flux pathways through the CSN and the clusters near the TS, using the TPT module in MSMBuilder [55]. For this purpose we use a trimmed network containing 1987 nodes, excluding nodes that are not connected to main connected component by both in-flow and out-flow. We first select a set of source and sink nodes, which are shown below in Fig. 7. The set of source nodes have cluster centers that are closer than 0.7 from the center of the cluster containing the initial MD structure according to a Canberra distance metric applied to the clustering feature vectors previously described in Section 3.4. The sink nodes are defined as the set of clusters whose centers' minimum distance between the ligand and protein is greater than 4 Å. Using these we construct a Markov state model (MSM), and calculate the unbinding committor probability ($p_{B \rightarrow U}$) using the `committors` function, then calculate the net forward flux (f^+) using the `net_fluxes` function in MSMBuilder [55, 62, 63]. From the unbinding committor probability we designated the set of clusters with $0.4 \leq p_{B \rightarrow U} \leq 0.6$ as the transition state ensemble (TSE). We also identified the highest flux reactive pathway in the network using Djikstra's algorithm (the `top_path` function in MSMBuilder [55]).

To find the contributions of different pathways to the reactive flux, we identify the edge with the lowest flux (the "bottleneck") along the highest flux reactive pathway. This is then removed, and the highest flux reactive pathway is again computed, identifying the next bottleneck. The reactive flux contribution of a bottleneck is equal to the flux through the bottleneck in the modified network. This procedure is continued until all but $1 - 1 * 10^{-10}$ of the reactive flux is accounted for. This procedure is implemented in the `paths` function in MSMBuilder, except for the identification of the actual edge that is the bottleneck.

3.7 Soluble Epoxide Hydrolase (sEH)

The protein soluble epoxide hydrolase (sEH) is found in most mammalian tissues and catalyzes the conversion of epoxyeicosatrienoic acid (EETs) to dihydroxyeicosatrienoic acids (DHET) [64].

It plays a physiological role in blood pressure, anti-inflammation, neuroprotection, and cardio-protection [64], and as well as being a target for treating chronic obstructive pulmonary diseases (COPD), atrial fibrillation, and diabetic neuropathic pain, it has had many drugs in clinical trials [65]. The binding site of sEH (see Fig. 1(C)) is large and deeply buried [64]. In crystal structures it is partially occluded by a center pinch (CP) formed by two loops coming together. Typical inhibitors of sEH are derived from urea and amide derivatives but other scaffolds include chalcone oxides, carbamates, and acyl hydrazones [64]. In this study we simulate unbinding of the TPPU inhibitor from the piperidyl-urea scaffold family (see Fig. 1(A)). The bound state of TPPU is coordinated by three hydrogen bonds from Asp105 and Tyr236 in the back of the sEH binding pocket and Tyr153 near the front of the pocket (see Fig. 1(B)) [15, 66].

4 Results

4.1 Conformation Space Network (CSN) Features

Six WExplore runs were conducted starting in the equilibrated crystallographic binding pose. All of these runs generated sampling poses that differed significantly from the crystal structure, and five runs sampled complete ligand unbinding pathways. To assess the degree of convergence we visualize the free-energy profile of a relevant observable (solvent accessible surface area (SASA) [67]) at progressive stages across all the runs in Fig. S1. From this we see that the free-energy profile has stabilized and few new structures are being observed. The entire ensemble of conformations generated by all WExplore runs is depicted by a CSN shown in Fig. 2, and the individual contributions from each run are shown in Fig. S2. The sizes of the nodes correspond to their statistical weight determined from sampling and nodes that are close together tend to interconvert more quickly than nodes that are far apart (see section 3.4). This results in a depiction that allows us to visualize the complete free energy landscape sampled by our simulations, as well as identify major collections of states and the paths between them [27, 36, 68–70].

Using a network modularity algorithm [71] we break up the network into communities, where the connections within a community are stronger than the connections between communities. We will refer to these communities using the labels in Fig. 2 (e.g. B_1 , B^* , P_1 , P_2 , and U). The spatial density of each community is compared in Fig. 3 and shown individually in Fig. S3.

The majority of the sampling weight is concentrated in the bound ensemble states (B_1), which includes the crystal structure. The B_1 community forms a tight, compact cluster of states both in the network layout (Fig. 2), indicating that they interconvert quickly, as well as the spatial density plot (Fig. 3 and Fig. S3), which additionally shows that the states are localized deep in the binding pocket. The B_1 ensemble can be characterized by its low free energy and low SASA (Fig. 4). The slowest interconverting states (identified as having the longest edges in the network) are those in R_1 and R_2 in which the ligand is bound, but has a reversed orientation relative to the crystal structure (Fig. S4). The location of these states in the network suggests that they are not directly accessible from B_1 , and that the ligand needs to first visit U to access R_1 and R_2 . These states demonstrate the breadth of sampling that was obtained with WExplore.

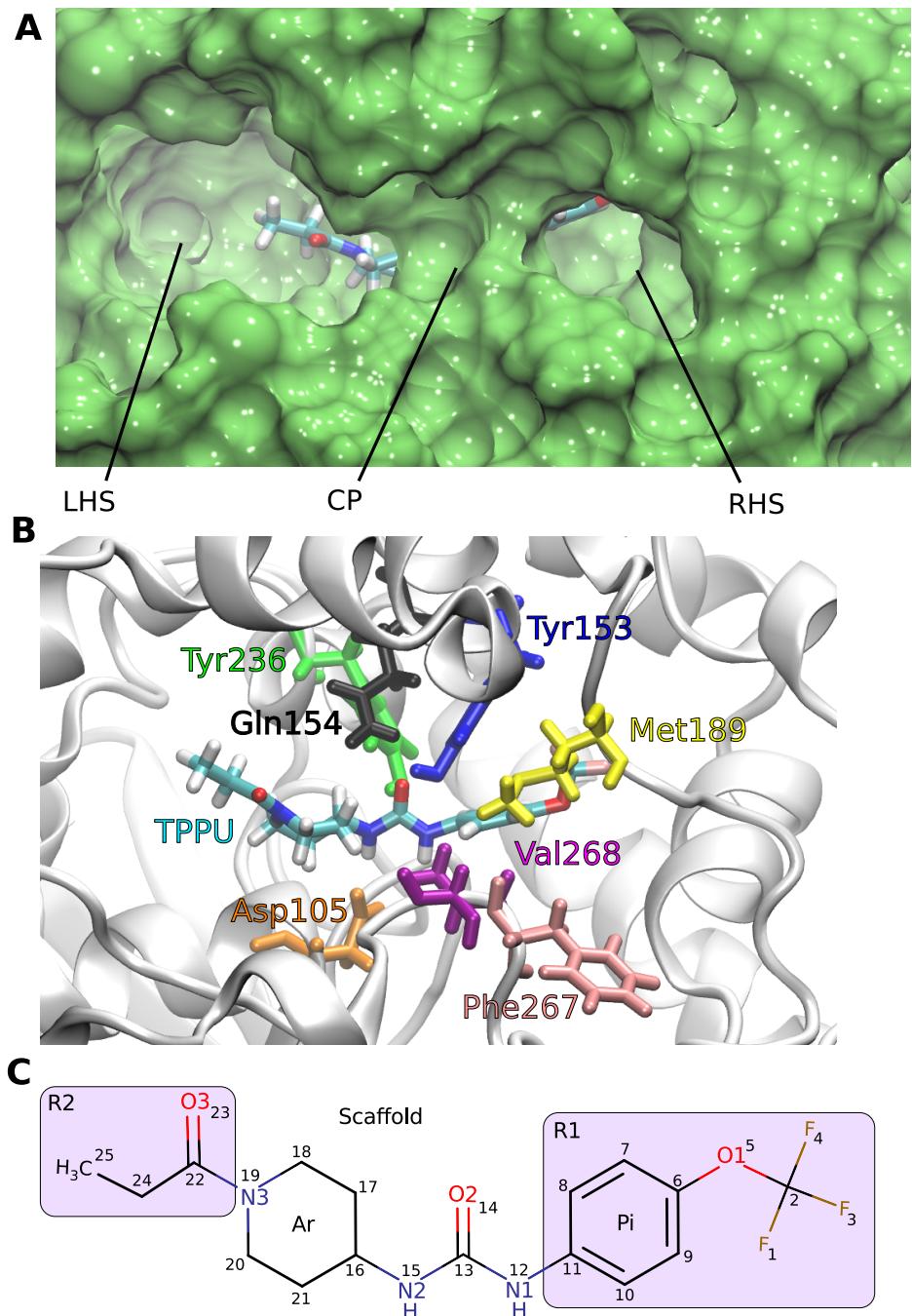


Figure 1: Binding site anatomy of sEH and structure of TPPU. (A) Surface representation (1.4 Å probe radius) of the binding pocket of sEH and the TPPU ligand in licorice representation colored by element, from PDB: 4OD0. The left hand side (LHS), the right hand side (RHS), and the center pinch (CP) of the binding site are labelled. (B) Identification of some important amino acids along the unbinding path of TPPU. The three residues with hydrogen bonds in PDB: 4OD0 include two in the back of the binding site: Tyr236 (green) and Asp268 (orange); and Tyr153 (blue) in the upper front. Other residues that form hydrogen bonds during dynamics are Gln154 (black), Met189 (yellow), Phe267 (pink), and Val268 (purple). (C) TPPU with the oxygens, nitrogens, piperidyl ring (Pi), and aromatic ring (Ar) labelled. The two R groups (R1 and R2) from the piperidyl-urea scaffold are highlighted in the boxes. All heavy atoms are labelled according to the serial numbers in PDB: 4OD0.

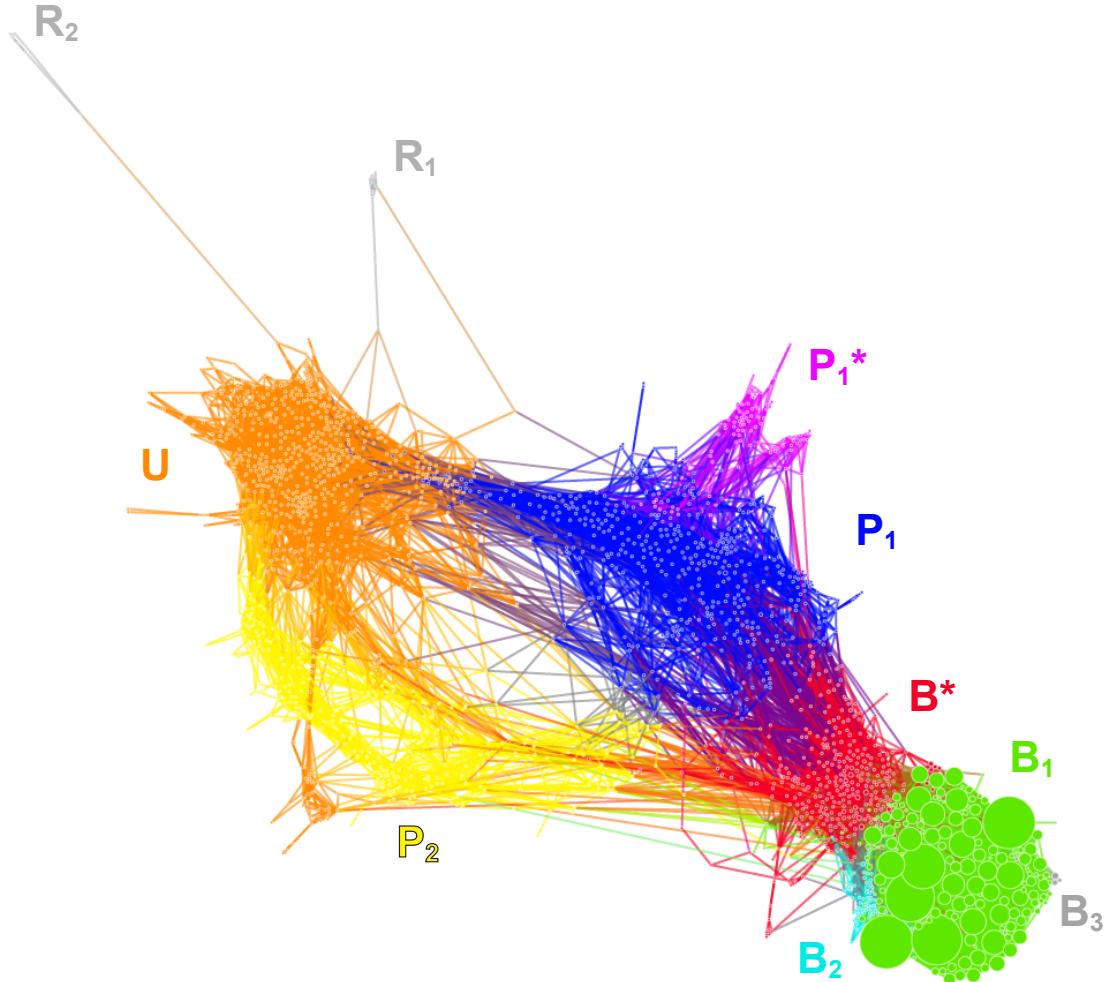


Figure 2: A conformation space network shows all poses of TPPU associated to sEH during sampling, including bound, transitioning, and unbound conformations. Each cluster is represented by a node, whose size is proportional to the weight of the cluster. The layout of the nodes and edges was determined using the ForceAtlas2 algorithm [57] in Gephi [56]. The nodes are colored according to their community: B_1 (Primary bound ensemble, light green), B_2 (secondary bound ensemble, cyan), B^* (bound branchpoint ensemble, red), P_1 (primary pathway 1 ensemble, blue), P_1^* (secondary pathway 1 ensemble, purple), P_2 (pathway 2 ensemble, yellow), and U (near-unbound ensemble, orange). Minor communities detected include B_3 , R_1 , and R_2 (grey). R_1 and R_2 feature ligand poses of reversed orientation relative to the native orientation (see Fig. S4). R_2 is removed from other CSN depictions for convenience of display.

4.2 Release from the Deeply Bound State

The deviation in ligand position is very restricted in B_1 as shown in Fig. 3, but in the near-bound intermediate community B^* , the ligand position density moves outwardly. To understand the driving forces for this transition we performed an interaction profiling across the network, where

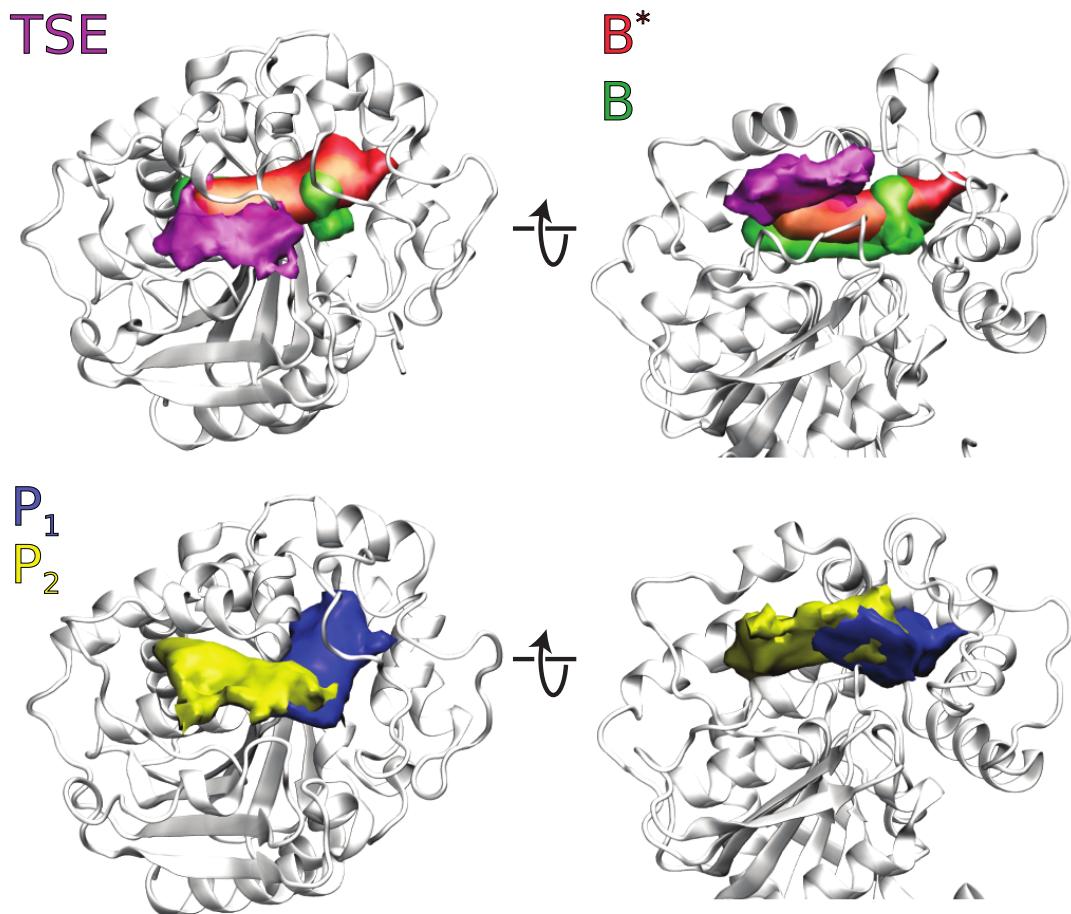


Figure 3: Density isosurfaces for various network subsets. Two views are shown for each: the front in the left column, and the bottom in the right column. The upper panel shows two views containing density isosurfaces for the transitions state ensemble (TSE; purple), the deeply bound community ensemble (B; green), and the near-bound community ensemble (B^* ; red). The lower panel shows density isosurfaces for the two main exit pathway communities. Path 1 (P1) is shown in blue and Path 2 (P2) is shown in yellow. All surfaces shown at a density of $\rho = 0.2$.

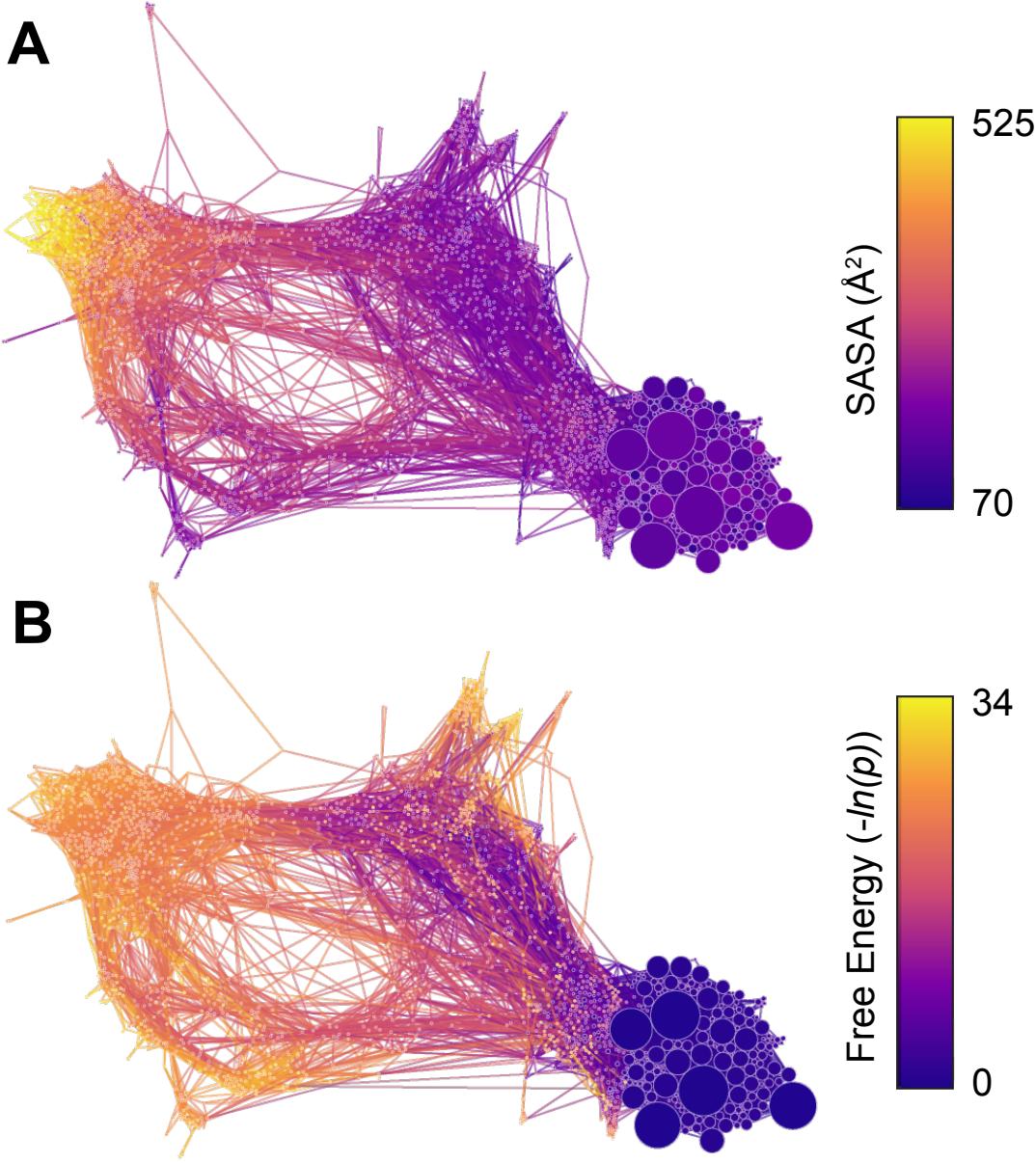


Figure 4: Conformation space networks colored according to A) solvent accessible surface area (SASA) of the TPPU ligand (\AA^2) calculated with the Shrake-Rupley method [67], and B) non-equilibrium unbinding ensemble free energy ($-\ln(p)$).

every protein-ligand interaction is identified from a representative frame from each node in the CSN. Fig. 5 shows ligand position density isosurfaces for some of the CSN's highest frequency ligand-protein hydrogen bonds (Tyr153-OH:Lig-O2, Asp105-OD2:Lig-N2, Val268-O:Lig-N1, and Met189-N:Lig-O2; see Fig. 5 caption for explanation of interaction nomenclature), as well as the distribution in the CSN. The frequencies of these interactions are shown separately for each network community in Fig. 6. Interactions which occur deep in the binding site (e.g. Asp105-OD2:Lig-N2) are shown to be much more compact than interactions with Tyr153. The occurrence

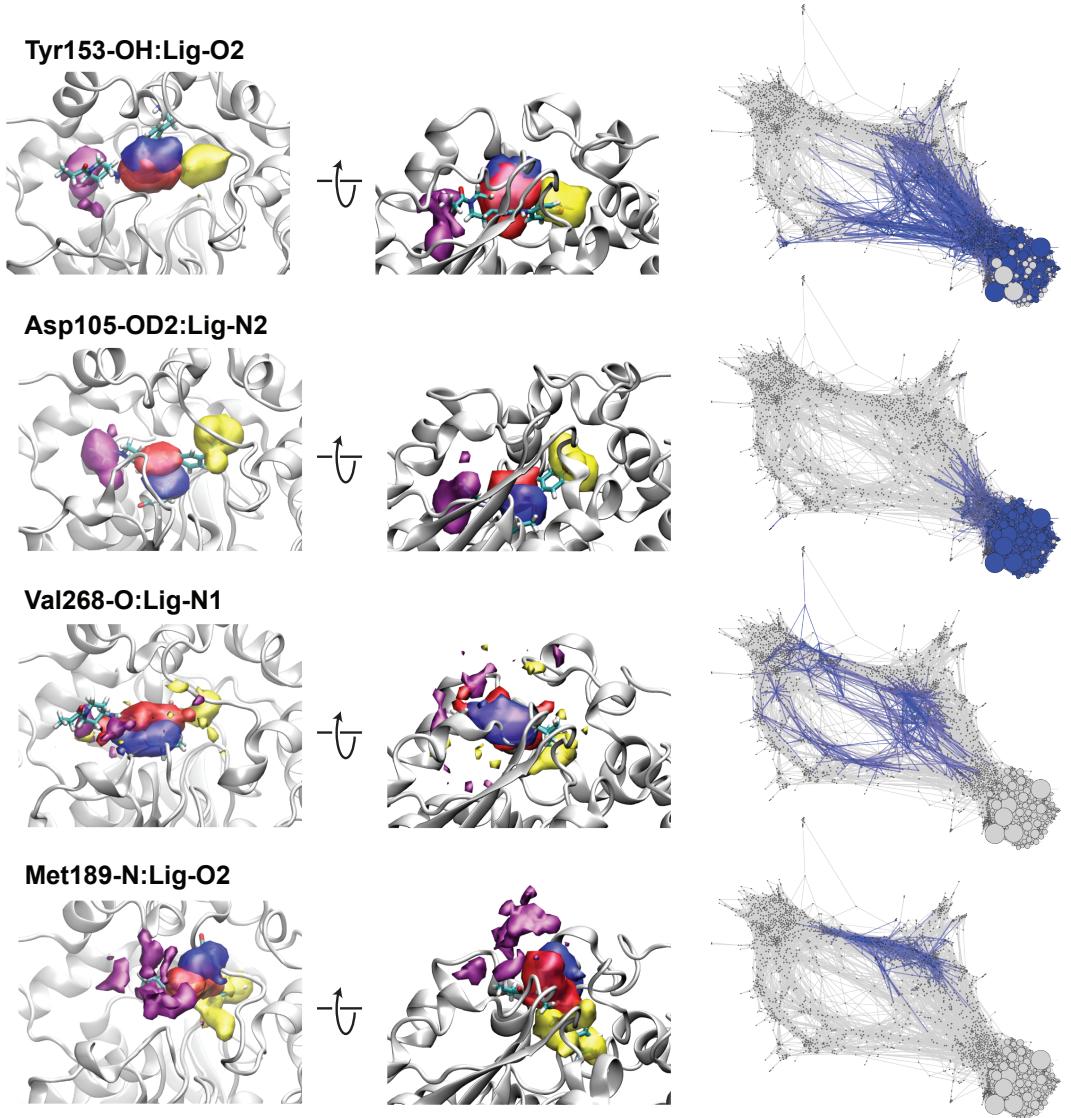


Figure 5: Density isosurfaces of the ligand (TPPU) (left and middle) and conformation space network distributions (right) for specific hydrogen bond acceptor-donor pairs. The interaction labels for each row specify the amino acid type and PDB residue index followed by a hyphen (-) followed by the PDB atom type of the acceptor/donor. The ligand (Lig) PDB atom type acceptor/donor follows the colon (:) (see Fig. 1(C)). The left column shows the front view of the binding site as shown in Fig. 1(A-B). The middle column shows a rotated view of the bottom side of the binding site. The acceptor/donor atom density isosurfaces for the protein are shown in blue and the ligand in red. Two density isosurfaces for distal ligand atoms C16 and C1 are shown in purple and yellow, respectively. The TPPU ligand is shown in licorice representation colored by atom type and the protein backbone is shown as a cartoon in white. Density isosurfaces are calculated from the positions of cluster representatives in which the interaction was detected, and surfaces of density $\rho = 0.025$ are shown. In the right column nodes in the conformation space network (CSN) are colored blue (as well as adjacent edges) by the presence of the interaction in the cluster representative. Nodes in which the interaction was not observed in the cluster representative are in grey.

of Tyr153-OH:Lig-O2 across the bound ensembles (B_1), near-bound ensemble (B^*), and parts of the path 1 (P_1) and path 2 (P_2) ensembles correspond with the positioning of Tyr153 at the front of the binding site. Similarly, Asp105-OD2:Lig-N2 (Fig. 5) and Tyr236 interactions (not shown), are focused in the bound ensemble and part of B^* due to the deep location of those residues. Thus it is likely that the breaking of hydrogen bonds to Asp105 and Tyr236 – which are found in the crystal structure – are necessary first steps in unbinding, while the positioning and flexibility of Tyr153 may stabilize this outward motion.

The role of Tyr153 as a stabilizer of dynamics is supported by the frequency patterns of all hydrogen bond interactions in each community shown in Fig. 6. Comparing B_1 and B^* we see that the highest frequency interaction, Tyr153-OH:Lig-O2, is represented in B^* at a higher frequency than in B_1 across network nodes. The other high frequency interactions in B_1 correspond to deep contacts with Asp105 and Tyr236, and are found in B^* at a lower frequency.

The sharp increase in binding site SASA seen at committor probabilities between 0.0 to 0.2 in Fig. S5 corresponds to B^* (Fig. 7). This suggests that concomitant with the breaking of deep native contacts to Asp105 and Tyr236 the binding site becomes more exposed to solvent. This transition is potentially a high free-energy barrier to unbinding, especially given the hydrophobicity of the binding site [15], and Tyr153 is likely crucial in stabilizing the outward motion of the ligand. Furthermore, B^* also acts as a branching point before the ligand commits to an exit path (i.e. P_1 , P_2).

4.3 Ligand Exit Pathways

Due to the shape and width of the sEH binding pocket [15, 64, 72] there is a range of possible exit trajectories. Here exit trajectories are centered along two dominant pathways, P_1 and P_2 , although we also find a fair number of states between the two pathways. P_1 trajectories are closer to the right hand side (RHS) of the pocket and P_2 trajectories are closer to the left hand side (LHS) of the pocket (see Fig. 3).

These two exit pathways are not topologically distinct relative to the protein backbone, as we found previously for benzamidine exit pathways from trypsin [36]. The two major modes of exit are likely due to the pinched center (CP) of the "lips" of the binding pocket which likely steers the ligand through the more open RHS and LHS corners of the "mouth" (Fig. 1). P_1 is more favorable in terms of free energy (Fig. 4), and is the shortest path through 3D space from the crystal structure state. P_1 is also traversed along the highest flux pathway through the CSN according to the transition matrix (see Methods section 3.6), as shown in Fig. S7 and S8. Exit through P_2 first requires a sliding motion in the bound state where a hydrogen bond with Asp105 is shifted between urea moiety nitrogens (N2 to N1) on TPPU. This exchange is not required in P_1 , and seems to be associated early on with interactions with Gln154 (see network plot in Fig. S6). We note however that analyzing individual pathways through the network or from continuous trajectories is likely misleading as can be seen by the low individual contributions of individual paths shown in Fig. S8. The bottleneck of the highest flux pathway explains only 0.8624 % of the flux through the network, the next highest being 0.5934 % and a sum of 367 such bottlenecks is needed to explain 80 % of the total flux through the network.

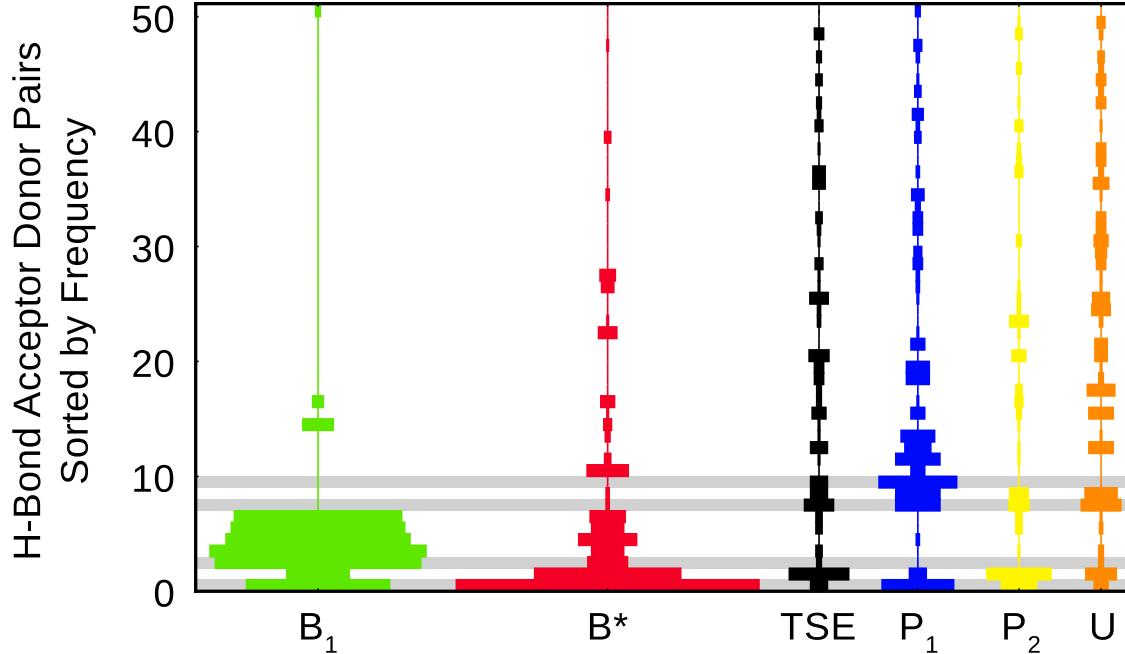


Figure 6: Interaction frequency violin bar graphs for the top 50 highest frequency interactions. Horizontal bars are for individual hydrogen bond acceptor-donor pairs, which are indexed by the vertical axis and sorted by total frequency from the bottom to the top. Violins are shown for each major community as labeled on the horizontal axis, and the color corresponds to those in Fig 2, except for the Transition State Ensemble (TSE) violin which is taken from Fig. 7. The total width of the bar corresponds to the total frequency of the interaction in the community. The gray highlighted rows are, from bottom to top are Tyr153-OH:Lig-O2, Asp105-OD2:Lig-N2, Val268-O:Lig-N1, and Met189-N:Lig-O2, which are detailed in Fig. 5 and some others not highlighted are detailed in Fig. S6. Complete results of interaction profiling are given in a file in the SI.

4.4 Transition State Stabilizing Interactions

Focusing on P_1 we find that concomitant with the breakage of Tyr153 hydrogen bonds, hydrogen bond interactions are formed between the ligand and the backbone nitrogens of Met189 (Fig. 5) and Phe267 (Fig. S6). Both of these residues are located on the "lips" of the binding site, whereas Met189 is on the far RHS, Phe267 is more towards the center, and adjacent to Val268 which is located directly at the center pinch (Fig. 1). The formation of a P_1 -specific interaction pattern that is distinct from the P_2 interactions can be seen in Fig. 6. For both P_1 and P_2 , the total number of states with specific interactions is much less than in B_1 or B^* , likely due to the higher solvent exposure (Fig. 4). Interestingly, while the ligand clouds of the residues located near the center pinch, like Val268, are more spread out and heterogeneous, the Met189 ligand cloud is fairly dense and tightly focused on the RHS corner with the ligand turned perpendicular relative to the bound pose. This ligand-protein interaction seems to be a key step in the ultimate release of the ligand as this interaction is found in states directly adjacent to the U ensemble. We summarily observe that for the P_1 pathway, interactions with Tyr153 and Met189 (e.g. Tyr153-OH:Lig-O2 and Met189-N:Lig-O2) play major roles in stabilizing the step-wise unbinding of TPPU from sEH.

The implication of Tyr153 or Met189 as the key interactions in stabilizing the overall unbind-

ing transition state is an important one, as destabilization of these interactions could result in longer ligand residence times. Although the positioning and flexibility of Tyr153 seems to be vital for unbinding, it is also a native contact and contributes significantly to ligand affinity, and thus not a favorable target for disruption. Met189 on the other hand is a much more desirable target, because it is not a native contact and could likely be disrupted with little impact on affinity. To ascertain the role these residues in kinetics, we predicted the transition state by first calculating the committor probabilities in a Markov state model framework [73], shown in Fig. 7 using the bound source and unbound sink basin nodes shown in Fig. 7(B). We then identified states with committor probabilities between 0.4 and 0.6 and use this as our prediction of the transition state ensemble (TSE) shown in Fig. 7(B). From this we can immediately tell that the TSE lies closer to U than to B_1 and is colocalized closely with the distribution of Met189 interactions (from Fig. 5). This suggests that Met189 is a transition state stabilizer, potentially along with Val268 and Phe267. The identification of transition state stabilizing interactions is important for kinetics-oriented drug design [6] and the disruption of such interactions has already been used to produce slow-onset inhibitors with very long residence times [20, 74, 75].

4.5 Unbound States and Unbinding Rates

The near-unbound ensemble (U) has the most varied ligand positions, as expected, and includes fully unbound poses as well poses around the outside of the "lips" of the binding site, potentially highlighting residue interactions (Fig. 6) that might be important for the initiation of binding events. While the other communities have mapped reasonably well to distinct phases of unbinding (e.g. B_1 , P_1 , etc.), the U ensemble is not merely the collection of unbound clusters, as can be seen by the protrusion below P_2 in Fig. 2. This is a result of the complexity of the CSN topology in the U region and the fact that the communities are determined using only properties of the network as a graph and not properties of the molecular structure. Additionally, as trajectories leave the surface of the protein, they are killed and their weight is recorded along with the time elapsed. Thus states near such an absorbing boundary are not sampled as thoroughly as the other communities and consequently fewer edges connect those states. Accordingly states in the U ensemble are likely short lived and weight is not able to accumulate and lower the free energy, even though, thermodynamically, an unbound solvated ligand in U should have a lower free energy than a transition state pose. This explains the high free energies shown in Fig. 4.

The exit times and weights of the exited walkers are used to calculate the rate constant predictions shown in Fig. 8 (see Methods section 3.3). Details of the number of exit points and descriptive statistics of exit point weights in each run are shown in Table S2, the 3D positions of the exit points are shown according to run in Fig. S9, and the frequency of exit points according to free energy is shown in Fig. S10. The spatial distribution of exit points is fairly broad even within runs suggesting not all exit points are highly correlated. The variation in cumulative unbound probability between runs is shown in Fig. 8(A) as a function of simulation time. The average cumulative unbound probability is shown as a thick black line, with its standard error in blue. The final total aggregated probability was $3.8 \times 10^{-10} \pm 3.1 \times 10^{-10}$. Aggregate probabilities and passage times are used to calculate the residence time prediction, as well as its standard error, shown in Fig. 8(B). From the observed 75 exit points we calculate a residence time of 42 s, compared to the experimentally observed residence time of 660 s (11 min) [15], which agrees to within 1 to 2 orders of magnitude and is just outside the standard error (which ranges from 23 s to

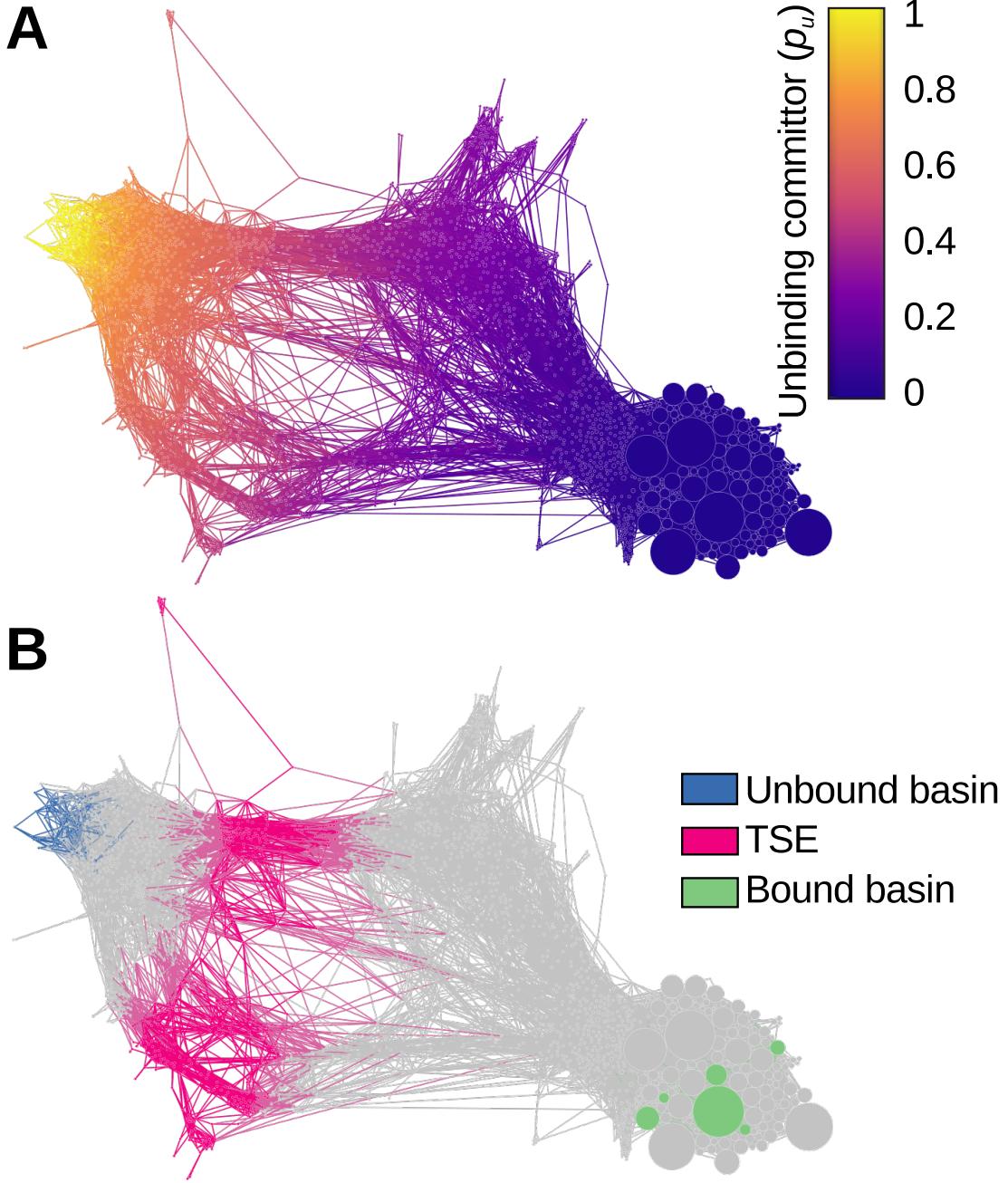


Figure 7: CSN showing A) forward committor probabilities (probability of unbinding, p_u) for each node, and B) the bound and unbound basins used to compute the committor probabilities and the transition state ensemble (TSE) where $0.4 \leq p_u \leq 0.6$.

280 s) of our calculation. To assess the convergence of this residence time we performed subsampling on all possible combinations of the different runs and recalculated the resulting residence time (Fig. S11). We find that while individually most runs are inaccurate, the average over multiple runs tends to converge towards the experimental value. This is comparable to results from trypsin-benzamidine residence time calculations, which is a 10^3 s faster process [36]. The vari-

ation between runs in the number of collected unbinding events, however, varies significantly (Fig. 8(A)) and further sampling would be necessary to obtain more precise predictions.

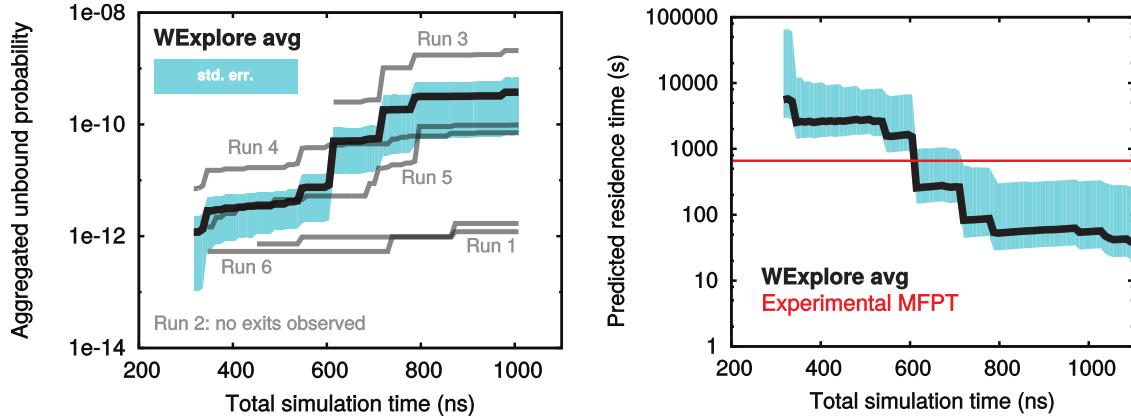


Figure 8: Left) The aggregated probability of unbound states across all WExplore runs. The average aggregated unbound probability of all runs is shown in the black line, with the standard error in light blue as a function of simulation time. The total aggregated weight was $3.8 \times 10^{-10} \pm 3.1 \times 10^{-10}$, and the final residence time was 42 s with standard error ranging from 23 s to 280 s. The aggregated probability in individual runs are shown in grey. No exits were observed for run 2. Right) The predicted residence time of the ligand (black with standard error in light blue) vs. the simulation time passed. This is compared to the experimentally measured residence time shown by the red line.

5 Discussion

Lee *et al.* [15] examined a series of inhibitors with different substituents for the R2 group of TPPU (see Fig. 1). It was observed that larger, more hydrophobic moieties had longer residence times and higher affinities as summarized in Fig. S12 (data from [15]). Thus, the increase in hydrophobic bulk certainly lowers the free energy of the bound state relative to the unbound and to the TS. This is intuitive as the R2 group sits in a pocket that is both solvent exposed and hydrophobic in the crystallographic bound state. This does not address the potential (de)stabilizing effects of the hydrophobic modifications to the TS. From our simulation data we may begin to probe whether or not this is plausible by observing correlations between ligand and binding site solvation with committor probabilities around the TS of unbinding.

When we plot the SASA of the entire TPPU ligand as a function of committor probability (Fig. S13) we observe a sharp increase in SASA at intermediate committor probabilities ($0.4 \leq p_u \leq 0.6$) around the predicted transition state. This transition is colocalized with the TS more than the binding site SASA (Fig. S5), which has a sharp increase only at low committor probabilities ($0.0 \leq p_u \leq 0.2$). In contrast, other potentially useful observables such as the distance between the centers of the two lips of the binding site (Fig. S14), show no such correlation. However, when we plot the SASA for only the TPPU-R2 group this trend disappears as shown in Fig. S15. This shows that TPPU-R2 solvation is not correlated with the transition state, and from Fig. S16 we see that the SASA trend for all non-R2 TPPU atoms recapitulates the trend for the whole

ligand shown in Fig. S13. Interestingly, when TPPU-R2 SASA is plotted on the CSN (Fig. S17) we see a difference between P_1 and P_2 suggesting the role of R2 is heterogeneous across the global transition state, and perhaps plays dual roles in different unbinding pathways.

When compared to other receptors of importance to drug design, the binding site of sEH is similar in some ways to both GPCRs (e.g. [22], PDB: 4EIY) and kinases (e.g. [25], PDB: 3G5D and 1KV2). GPCRs have a well-defined long, deep binding site with a small orifice and binds the ligand very tightly leaving very low SASA. In contrast, kinases have a more open binding region between two domains with high SASA. Comparably sEH has a well defined binding pocket like a GPCR, but is shallower and wider and is not occupied completely by the ligand. Binding pockets and channels which are somewhat restricted, like in GPCRs and sEH, are potentially more amenable to kinetics oriented drug design because the unbinding TS is likely more well defined than that of the open kinases.

In this study we identified two distinct unbinding paths that are not topologically separated, as were those found for trypsin [33, 35, 36] and for type-II kinases [25]. Destabilizations to the TS through ligand modification become more complicated in systems with multiple unbinding pathways as each transition state likely involves different interactions with the ligand. Indeed, here we find an intersection of 13 specific hydrogen-bonding interactions between P_1 and P_2 , from the total union of 76, in which there is no correlation between pathway ensembles as shown in Fig. S18. In systems with multiple major unbinding mechanisms one will be the globally lowest free-energy TS and will contribute the most to the unbinding flux. If the transition states of multiple unbinding pathways are comparable in free energy, then perturbations to drug-TS interactions could potentially shift which pathway is the dominant unbinding path, a phenomenon we name **pathway hopping**. Notably, this phenomenon has been observed previously in the context of protein folding [76]. The complications pathway hopping adds to drug design is an important consideration to studies of drug binding kinetics, and emphasizes the importance of methods which identify the full spectrum of unbinding pathways. For instance, as suggested above, larger, more hydrophobic R2 groups could stabilize the P_2 pathway while destabilizing the P_1 pathway, potentially changing the set of TS-stabilizing interactions to consider during drug design. WExplore is particularly useful in this regard because it emphasizes the discovery of these alternative pathways, and may preemptively address drug-design concerns. A discovery-oriented sampling regime will likely improve the robustness of drug-design for kinetics by both anticipating pathway hopping, and by providing unbiased information for the construction of better order parameters as suggested in [39].

We consider order parameter driven methods, such as metadynamics, to fundamentally be in a refinement-oriented regime in which a particular unbinding mechanism is sampled extensively for kinetics estimations. We thus emphasize that discovery- and refinement-oriented enhanced sampling methods can be used in tandem to both elucidate the structural elements of complex unbinding mechanisms as well as their rates. Particularly, refinement-oriented methods control for the sensitivity of discovery-oriented methods like WExplore to initial conditions and subsequent divergence between different runs. The combination of both regimes to solve increasingly difficult processes will require significant collaboration in order to unify enhanced sampling implementations and share data.

6 Conclusion

This work describes advances in general methods applicable to the simulation of complex macromolecular processes, such as ligand unbinding, protein folding, and conformational changes. To our knowledge this is the first simulation of drug unbinding that occurs at pharmacologically relevant timescales (11 min) with no specific knowledge of the mechanism and no biasing forces, using only commodity hardware. Furthermore, our results were achieved in only 6 μ s of simulation time, which compares favorably to similar simulations using either brute force or Markov state modeling that typically use one to two orders of magnitude more simulation time [21, 33].

We report the simulation of the drug-like inhibitor TPPU unbinding (and rebinding) from the soluble epoxide hydrolase (sEH) enzyme using the WExplore algorithm. We visualize the conformational ensemble as a network which we use to identify specific interactions involved in the unbinding process. Some of these interactions are implicated in the stabilization of the unbinding transition state, which potentially could be used in kinetics-oriented drug design. With these results as a guide, a given ligand-protein system can be examined using about 6 μ s of sampling (or 6000 cycles). Using one node with 4 NVIDIA K80 graphics processing units we are able to obtain 120 cycles per day for this system. With a modest cluster of 5 such nodes, additional ligands can be examined at 10 days each. A larger cluster with 36 such nodes could simulate a new ligand every 33 hours. Thus, we have also proposed a general way forward for obtaining and improving the accuracy of kinetic models from MD using both discovery- and refinement-oriented enhanced sampling methods.

Supporting Information. Supplemental figures and tables can be found in the supplemental pdf. Simulation systems, data, and analysis code will be made available on zenodo.org

7 Acknowledgments

We acknowledge support from the High Performance Computing Center at Michigan State University.

References

- [1] Robert A. Copeland, David L. Pompliano, and Thomas D. Meek. Drug-target residence time and its implications for lead optimization. *Nat Rev Drug Discov*, 5(9):730–739, sep 2006.
- [2] Hideaki Fujitani, Yoshiaki Tanida, Masakatsu Ito, Guha Jayachandran, Christopher D. Snow, Michael R. Shirts, Eric J. Sorin, and Vijay S. Pande. Direct calculation of the binding free energies of fkbp ligands. *The Journal of Chemical Physics*, 123(084108), 2005.
- [3] James C. Gumbart, Benoît Roux, and Christophe Chipot. Standard binding free energies from computer simulations: What is the best strategy? *Journal of Chemical Theory and Computation*, 9(1):794–802, 2013. PMID: 23794960.
- [4] Yuqing Deng and Benoît Roux. Computations of standard binding free energies with molecular dynamics simulations. *The Journal of Physical Chemistry B*, 113(8):2234–2246, 2009. PMID: 19146384.

- [5] Ryoji Takahashi, Víctor A. Gil, and Victor Guallar. Monte carlo free ligand diffusion with markov state model analysis and absolute binding free energy calculations. *Journal of Chemical Theory and Computation*, 10(1):282–288, 2014. PMID: 26579911.
- [6] Doris A. Schuetz, Wilhelmus Egbertus Arnout de Witte, Yin Cheong Wong, Bernhard Knasmueller, Lars Richter, Daria B. Kokh, S. Kashif Sadiq, Reggie Bosma, Indira Nederpelt, Laura H. Heitman, Elena Segala, Marta Amaral, Dong Guo, Dorothee Andres, Victoria Georgi, Leigh A. Stoddart, Steve Hill, Robert M. Cooke, Chris De Graaf, Rob Leurs, Matthias Frech, Rebecca C. Wade, Elizabeth Cunera Maria de Lange, Adriaan P. IJzerman, Anke Müller-Fahrnow, and Gerhard F. Ecker. Kinetics for drug discovery: an industry-driven effort to target drug residence time. *Drug Discovery Today*, 22(6):896 – 911, 2017.
- [7] Robert A Copeland. The dynamics of drug-target interactions: drug-target residence time and its impact on efficacy and safety. *Expert Opinion on Drug Discovery*, 5(4):305–310, 2010. PMID: 22823083.
- [8] Robert A. Copeland. The drug-target residence time model: a 10-year retrospective. *Nature Reviews: Drug Discovery*, 15(2):87–95, feb 2016.
- [9] Dong Guo, Thea Mulder-Krieger, Adriaan P IJzerman, and Laura H Heitman. Functional efficacy of adenosine a_{2a} receptor agonists is positively correlated to their receptor residence time. *British Journal of Pharmacology*, 166(6):1846–1859, 2012.
- [10] Hao Lu and Peter J Tonge. Drug-target residence time: critical information for lead optimization. *Current Opinion in Chemical Biology*, 14(4):467–474, 2010.
- [11] Hao Lu, Kathleen England, Christopher am Ende, James J. Truglio, Sylvia Luckner, B. Gopal Reddy, Nicole L. Marlenee, Susan E. Knudson, Dennis L. Knudson, Richard A. Bowen, Caroline Kisker, Richard A. Slayden, and Peter J. Tonge. Slow-onset inhibition of the fabi enoyl reductase from francisella tularensis: Residence time and in vivo activity. *ACS Chemical Biology*, 4(3):221–231, 2009.
- [12] B Costa, E Da Pozzo, C Giacomelli, E Barresi, S Taliani, F Da Settimo, and C Martini. Tspo ligand residence time: a new parameter to predict compound neurosteroidogenic efficacy. *Sci Rep*, 6:18164–18164, 2016.
- [13] Georges Vauquelin and Isabelle Van Liefde. Slow antagonist dissociation and long-lasting in vivo receptor protection. *Trends in Pharmacological Sciences*, 27(7):355 – 359, 2006.
- [14] Johan Gabrielsson, Hugues Dolgos, Per-Göran Gillberg, Ulf Bredberg, Bert Bentheim, and Göran Duker. Early integration of pharmacokinetic and dynamic reasoning is essential for optimal development of lead compounds: strategic considerations. *Drug Discovery Today*, 14(7–8):358 – 372, 2009.
- [15] K S Lee, J Y Liu, K M Wagner, S Pakhomova, H Dong, C Morisseau, S H Fu, J Yang, P Wang, A Ulu, C A Mate, L V Nguyen, S H Hwang, M L Edin, A A Mara, H Wulff, M E Newcomer, D C Zeldin, and B D Hammock. Optimized inhibitors of soluble epoxide hydrolase improve in vitro target residence time and in vivo efficacy. *J Med Chem*, 57(16):7016–7030, aug 2014.

- [16] John Comley. Progress in the implementation of label-free detection. Technical report, 2008.
- [17] Wilma Keighley. The need for high throughput kinetics early in the drug discovery process. Technical report, Drug Discovery World, 2011.
- [18] H J Motulsky and L C Mahan. The kinetics of competitive radioligand binding predicted by the law of mass action. *Molecular Pharmacology*, 25(1):1–9, 1984.
- [19] David E. Shaw, Martin M. Deneroff, Ron O. Dror, Jeffrey S. Kuskin, Richard H. Larson, John K. Salmon, Cliff Young, Brannon Batson, Kevin J. Bowers, Jack C. Chao, Michael P. Eastwood, Joseph Gagliardo, J. P. Grossman, C. Richard Ho, Douglas J. Ierardi, István Kolossváry, John L. Klepeis, Timothy Layman, Christine McLeavey, Mark A. Moraes, Rolf Mueller, Edward C. Priest, Yibing Shan, Jochen Spengler, Michael Theobald, Brian Towles, and Stanley C. Wang. Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM*, 51(7):91–97, jul 2008.
- [20] Huei-Jiun Li, Cheng-Tsung Lai, Pan Pan, Weixuan Yu, Nina Liu, Gopal R. Bommineni, Miguel Garcia-Diaz, Carlos Simmerling, and Peter J. Tonge. A structural and energetic model for the slow-onset inhibition of the mycobacterium tuberculosis enoyl-acp reductase inha. *ACS Chemical Biology*, 9(4):986–993, 2014. PMID: 24527857.
- [21] Yibing Shan, Eric T. Kim, Michael P. Eastwood, Ron O. Dror, Markus A. Seeliger, and David E. Shaw. How does a drug molecule find its target binding site? *Journal of the American Chemical Society*, 133(24):9181–9183, 2011. PMID: 21545110.
- [22] Dong Guo, Albert C. Pan, Ron O. Dror, Tamara Mocking, Rongfang Liu, Laura H. Heitman, David E. Shaw, and Adriaan P. IJzerman. Molecular basis of ligand dissociation from the adenosine a2a receptor. *Molecular Pharmacology*, 89(5):485–491, 2016.
- [23] Vittorio Limongelli, Massimiliano Bonomi, and Michele Parrinello. Funnel metadynamics as accurate binding free-energy method. *Proceedings of the National Academy of Sciences*, 110(16):6358–6363, 2013.
- [24] Rodrigo Casasnovas, Vittorio Limongelli, Pratyush Tiwary, Paolo Carloni, and Michele Parrinello. Unbinding kinetics of a p38 map kinase type ii inhibitor from metadynamics simulations. *Journal of the American Chemical Society*, 139:4780–4788, 2017. PMID: 28290199.
- [25] Huiyong Sun, Sheng Tian, Shunye Zhou, Youyong Li, Dan Li, Lei Xu, Mingyun Shen, Peichan Pan, and Tingjun Hou. Revealing the favorable dissociation pathway of type ii kinase inhibitors via enhanced sampling simulations and two-end-state calculations. *Scientific Reports*, 5(8457), 2015.
- [26] Pratyush Tiwary, Jagannath Mondal, and B. J. Berne. How and when does an anticancer drug leave its binding site? *Science Advances*, 3(5), 2017.
- [27] Danzhi Huang and Amedeo Caflisch. The free energy landscape of small molecule unbinding. *PLOS Computational Biology*, 7:1–12, February 2011.

- [28] Min Xu, Amedeo Caflisch, and Peter Hamm. Protein structural memory influences ligand binding mode(s) and unbinding rates. *Journal of Chemical Theory and Computation*, 12(3):1393–1399, 2016. PMID: 26799675.
- [29] Alex Dickson and Samuel D. Lotz. Ligand release pathways obtained with wexplore: Residence times and mechanisms. *The Journal of Physical Chemistry B*, 120(24):5377–5385, 2016. PMID: 27231969.
- [30] Albert C. Pan, Huafeng Xu, Timothy Palpant, and David E. Shaw. Quantitative characterization of the binding and unbinding of millimolar drug fragments with molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 13(7):3372–3377, 2017. PMID: 28582625.
- [31] Ivan Teo, Christopher G. Mayne, Klaus Schulten, and Tony Lelièvre. Adaptive multilevel splitting method for molecular dynamics calculation of benzamidine-trypsin dissociation time. *Journal of Chemical Theory and Computation*, 12(6):2983–2989, 2016. PMID: 27159059.
- [32] Ignasi Buch, Toni Giorgino, and Gianni De Fabritiis. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proceedings of the National Academy of Sciences*, 108(25):10184–10189, 2011.
- [33] Nuria Plattner and Frank Noe. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and markov models. *Nat Commun*, 6, jul 2015.
- [34] S. Doerr and G. De Fabritiis. On-the-fly learning and sampling of ligand binding by high-throughput molecular simulations. *Journal of Chemical Theory and Computation*, 10(5):2064–2069, 2014. PMID: 26580533.
- [35] Pratyush Tiwary, Vittorio Limongelli, Matteo Salvalaglio, and Michele Parrinello. Kinetics of protein-ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proceedings of the National Academy of Sciences*, 112(5):E386–E391, 2015.
- [36] Alex Dickson and Samuel D. Lotz. Multiple ligand unbinding pathways and ligand-induced destabilization revealed by wexplore. *Biophysical Journal*, 112(4):620–629, February 2017.
- [37] Florent Guillain and Darwin Thusius. Use of proflavine as an indicator in temperature-jump studies of the binding of a competitive inhibitor to trypsin. *Journal of the American Chemical Society*, 92(18):5534–5536, 1970.
- [38] Pratyush Tiwary and Michele Parrinello. From metadynamics to dynamics. *Phys. Rev. Lett.*, 111(23):230602, dec 2013.
- [39] Pratyush Tiwary and B. J. Berne. Spectral gap optimization of order parameters for sampling complex molecular systems. *Proceedings of the National Academy of Sciences*, 113(11):2839–2844, 2016.
- [40] Cameron F. Abrams and Eric Vanden-Eijnden. Large-scale conformational sampling of proteins using temperature-accelerated molecular dynamics. *Proceedings of the National Academy of Sciences*, 107(11):4961–4966, 2010.

- [41] Luca Maragliano and Eric Vanden-Eijnden. A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chemical Physics Letters*, 426(1):168 – 175, 2006.
- [42] Luca Mollica, Isabelle Theret, Mathias Antoine, Francoise Perron-Sierra, Yves Charton, Jean-Marie Fourquez, Michel Wierzbicki, Jean A. Boutin, Gilles Ferry, Sergio Decherchi, Giovanni Bottegoni, Pierre Ducrot, and Andrea Cavalli. Molecular dynamics simulations and kinetic measurements to estimate and predict protein? ligand residence times luca mollica? *Journal of Medicinal Chemistry*, 59(15):7167–7176, 2016.
- [43] Luca Mollica, Sergio Decherchi, Syeda Rehana Zia, Roberto Gaspari, Andrea Cavalli, and Walter Rocchia. Kinetics of protein-ligand unbinding via smoothed potential molecular dynamics simulations. *Nature Scientific Reports*, 5(11539), June 2015.
- [44] Anna Maria Capelli and Gabriele Costantino. Unbinding pathways of vegfr2 inhibitors revealed by steered molecular dynamics. *Journal of Chemical Information and Modeling*, 54(11):3124–3136, 2014. PMID: 25299731.
- [45] G A Huber and S Kim. Weighted-ensemble brownian dynamics simulations for protein association reactions. *Biophys J*, 70(1):97–110, jan 1996.
- [46] A Dickson and C L Brooks. Wexplore: hierarchical exploration of high-dimensional spaces using the weighted ensemble algorithm. *J Phys Chem B*, 118(13):3532–3542, apr 2014.
- [47] Robert B. Best, Xiao Zhu, Jihyun Shim, Pedro E. M. Lopes, Jeetain Mittal, Michael Feig, and Alexander D. MacKerell. Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles. *Journal of Chemical Theory and Computation*, 8(9):3257–3273, 2012. PMID: 23341755.
- [48] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell. Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *Journal of Computational Chemistry*, 31(4):671–690, 2010.
- [49] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. Charmm: The biomolecular simulation program. *Journal of Computational Chemistry*, 30(10):1545–1614, 2009.
- [50] P Eastman, M S Friedrichs, JD Chodera, R J Radmer, C M Bruns, J P Ku, K A Beauchamp, T J Lane, L P Wang, D Shukla, T Tye, M Houston, T Stich, C Klein, M R Shirts, and V S Pande. Openmm 4: A reusable, extensible, hardware independent library for high performance molecular simulation. *J Chem Theory Comput*, 9(1):461–469, jan 2013.
- [51] Alex Dickson, Aryeh Warmflash, and Aaron R. Dinner. Separating forward and backward pathways in nonequilibrium umbrella sampling. *The Journal of Chemical Physics*, 131(154104), 2009.

- [52] Eric Vanden-Eijnden and Maddalena Venturoli. Exact rate calculations by trajectory parallelization and tilting. *The Journal of Chemical Physics*, 131(044120), 2009.
- [53] Daniel M. Zuckerman and Lillian T. Chong. Weighted ensemble simulation: Review of methodology, applications, and software. *Annual Review of Biophysics*, 46(1):43–57, 2017. PMID: 28301772.
- [54] Terrell L Hill. *Free Energy Transduction and Biochemical Cycle Kinetics*. Springer, 1989.
- [55] Kyle A. Beauchamp, Gregory R. Bowman, Thomas J. Lane, Lutz Maibaum, Imran S. Haque, and Vijay S. Pande. Msmbuilder2: Modeling conformational dynamics on the picosecond to millisecond scale. *Journal of Chemical Theory and Computation*, 7(10):3412–3419, 2011. PMID: 22125474.
- [56] Mathieu Bastian, Sébastien Heymann, and Mathieu Jacomy. Gephi: an open source software for exploring and manipulating networks. *ICWSM*, 8:361–362, 2009.
- [57] Mathieu Jacomy, Tommaso Venturini, Sébastien Heymann, and Mathieu Bastian. Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PLOS ONE*, 9(6):1–12, 06 2014.
- [58] Greg Landrum. Rdkit: Open-source cheminformatics software. 2016.
- [59] Roderick E Hubbard and Muhammad Kamran Haider. Hydrogen bonds in proteins: Role and strength. In *Encyclopedia of Life Science*. John Wiley & Sons, Ltd., Chichester, 2010.
- [60] Samuel Lotz. Multi-atom structure/selection toolkit with interaction capabilities (mastic) v0.2.1-beta release. 2017.
- [61] Yannick Hold-Geoffroy, Olivier Gagnon, and Marc Parizeau. Once you scoop, no need to fork. In *Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment*, page 60. ACM, 2014.
- [62] Weinan E. and Eric Vanden-Eijnden. Towards a theory of transition paths. *Journal of Statistical Physics*, 123(3):503–523, 2006.
- [63] Philipp Metzner, Christof Schütte, and Eric Vanden-Eijnden. Transition path theory for markov jump processes. *Multiscale Modeling & Simulation*, 7(3):1192–1219, 2009.
- [64] Hang Chen, Ying Zhang, Liang Li, and Ju-Guang Han. Probing ligand-binding modes and binding mechanisms of benzoxazole-based amide inhibitors with soluble epoxide hydrolase by molecular docking and molecular dynamics simulation. *The Journal of Physical Chemistry B*, 116(34):10219–10233, 2012. PMID: 22857012.
- [65] Kin Sing Stephen Lee, Niel M. Henriksen, Connie J. Ng, Jun Yang, Weitao Jia, Christophe Morisseau, Armann Andaya, Michael K. Gilson, and Bruce D. Hammock. Probing the orientation of inhibitor and epoxy-eicosatrienoic acid binding in the active site of soluble epoxide hydrolase. *Archives of Biochemistry and Biophysics*, 613:1 – 11, 2017.

- [66] Li Xing, Joseph J. McDonald, Steve A. Kolodziej, Ravi G. Kurumbail, Jennifer M. Williams, Chad J. Warren, Janet M. O’Neal, Jill E. Skepner, and Steven L. Roberds. Discovery of potent inhibitors of soluble epoxide hydrolase by combinatorial library design and structure-based virtual screening. *Journal of Medicinal Chemistry*, 54(5):1211–1222, 2011. PMID: 21302953.
- [67] A. Shrake and J.A. Rupley. Environment and exposure to solvent of protein atoms. lysozyme and insulin. *Journal of Molecular Biology*, 79(2):351 – 371, 1973.
- [68] Francesco Rao and Amedeo Caflisch. The protein folding network. *Journal of Molecular Biology*, 342(1):299 – 306, 2004.
- [69] Alex Dickson, Anthony M. Mustoe, Loic Salmon, and Charles L. Brooks III. Efficient in silico exploration of rna interhelical conformations using euler angles and wexplore. *Nucleic Acids Research*, 42(19), 2014.
- [70] Loïc Salmon, Logan S. Ahlstrom, Scott Horowitz, Alex Dickson, Charles L. Brooks, and James C. A. Bardwell. Capturing a dynamic chaperone–substrate interaction using nmr-informed molecular modeling. *Journal of the American Chemical Society*, 138(31):9826–9839, 2016. PMID: 27415450.
- [71] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.
- [72] Martin Kotev, Robert Soliva, and Modesto Orozco. Challenges of docking in large, flexible and promiscuous binding sites. *Bioorganic & Medicinal Chemistry*, 24(20):4961 – 4969, 2016.
- [73] Nina Singhal, Christopher D. Snow, and Vijay S. Pande. Using path sampling to build better markovian state models: Predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. *The Journal of Chemical Physics*, 121(1):415–425, 2004.
- [74] Lauren A. Spagnuolo, Sandra Eltschkner, Weixuan Yu, Fereidoon Daryae, Shabnam Davoodi, Susan E. Knudson, Eleanor K. H. Allen, Jonathan Merino, Annica Pschibul, Ben Moree, Neil Thivalapill, James J. Truglio, Joshua Salafsky, Richard A. Slayden, Caroline Kisker, and Peter J. Tonge. Evaluating the contribution of transition-state destabilization to changes in the residence time of triazole-based inha inhibitors. *Journal of the American Chemical Society*, 139(9):3417–3429, 2017. PMID: 28151657.
- [75] Andrew Chang, Johannes Schiebel, Weixuan Yu, Gopal R. Bommineni, Pan Pan, Michael V. Baxter, Avinash Khanna, Christoph A. Sotriffer, Caroline Kisker, and Peter J. Tonge. Rational optimization of drug-target residence time: Insights from inhibitor binding to the staphylococcus aureus fabi enzyme–product complex. *Biochemistry*, 52(24):4217–4228, 2013. PMID: 23697754.
- [76] James S. Butler and Stewart N. Loh. Kinetic partitioning during folding of the p53 dna binding domain. *Journal of Molecular Biology*, 350(5):906 – 918, 2005.

8 For Table of Contents Only

