How robust is the ligand binding transition state?

Samik Bose,[†] Samuel D. Lotz,[†,‡] Indrajit Deb,[†,¶] Megan Shuck,[§] Kin Sing Stephen Lee,[§,‖,⊥] and Alex Dickson[*,†,#]

†*Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, United States*

‡*Current address: Examol Corporation, Cranberry Township, Pennsylvania 16066, United States*

¶*Current address: Drug Discovery Unit, School of Life Sciences, University of Dundee, Dundee, United Kingdom*

§*Department of Pharmacology and Toxicology, Michigan State University, East Lansing, Michigan 48824, United States*

‖*Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States*

⊥*Institute of Integrative Toxicology, Michigan State University, East Lansing, MI 48824, United States*

#*Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, Michigan 48824, United States*

E-mail: alexrd@msu.edu

Abstract

For many drug targets, it has been shown that the kinetics of drug binding (e.g. on rate and off rate) is more predictive of drug efficacy than thermodynamic quantities alone. This motivates the development of predictive computational models that can be

used to optimize compounds on the basis of their kinetics. The structural details underpinning these computational models are not only found in the bound state, but also in the short-lived ligand binding transition state: the highest free energy point along the (un)binding pathway. Although this transition state cannot be directly observed experimentally, due to its extremely short lifetime, recent successes have demonstrated that modeling of the ligand binding transition state is possible with the help of enhanced sampling methods for molecular dynamics. In our previous work we analyzed the transition state ensemble for an inhibitor of soluble epoxide hydrolase (sEH) with a residence time of 11 minutes. Here we computationally modeled unbinding events for five additional inhibitors of sEH with residence times ranging from 14.25 to 31.75 minutes, with our results recapitulating these experimental residence times to within an order of magnitude. The unbinding ensembles are analyzed in detail, focusing on features of the ligand binding transition state. We find that ligands with similar structures and similar bound poses can show significant differences in their ligand binding transition states, in terms of their spatial location and their interactions with specific protein residues. However we also find similarities across the transition state ensembles when examining more general features such as ligand degrees of freedom. Together these findings show significant challenges for rational, kinetics-based drug design.

Introduction

Structure based drug design (SBDD) has matured over the past few decades from a handful of success stories[1–5] into a near-ubiquitous tool to guide the discovery and optimization of potential drug molecules.[6] SBDD approaches – including docking- and AI-assisted virtual screens,[7,8] MM-PBSA methods[9] and free energy perturbation[10,11] – all utilize the knowledge of the 3D structure of a target protein and its probable binding site to design potential drug molecules by optimizing the binding free energy, or approximations thereof. In principle, this provides an incomplete picture of the drug-target interaction; since the *in vivo* environment

is far from thermodynamic equilibrium, the binding kinetics are also necessary to thoroughly model drug efficacy.[12–14] In practice, it has been shown that residence time – the average duration of a given drug-target binding event – can be the central feature related to drug efficacy in some systems,[15] including soluble epoxide hydrolase (sEH),[16,17] studied here. In contrast to the binding free energy, which is a path-independent state function relying only on the bound and unbound states, the unbinding rate constant depends on the details of the transition path ensemble of (un)binding events. Specifically, it is related by the Arrhenius equation to the free energy of activation of the unbinding event, which is the difference in free energy of bound state and the ligand binding transition state. Hence, to fully engage the tools of SBDD for kinetics-based rational design, we need to consider the molecular structures of both the bound state and the transition state.

This poses a monumental challenge, because unlike ligand bound states, for which there are hundreds of thousands of available experimentally-determined structures, there are no experimental observations of ligand binding transition state structures, due to their extremely short lifetimes. Ligand binding transition states are also challenging to model *in silico*. Although the transition state lifetimes are short, they are often at the top of extremely large energy barriers, with mean first passage times (MFPTs) of unbinding events that range up to minutes or even hours in duration. This is 6-8 orders of magnitude beyond the current capabilities of even specialized supercomputers for molecular dynamics, which are still restricted to the μs to ms regime.[18] Also, transition state ensembles for ligand binding are likely to be much more diverse than bound ensembles and are unable to be captured in a single structure. There is substantial possibility of a ligand unbinding from a protein using multiple pathways and within each pathway there are conformational fluctuations that change particular ligand-protein interactions.[19–23] This requires not just generation of a single unbinding event, but of a representative ensemble of unbinding events. Finally, even with an ensemble of (un)binding transition paths, identification of the transition state ensemble requires additional analysis techniques that calculate the unbinding committor probability for each

conformation. The introduction of biasing forces in methods such as metadynamics[24] and τ-random acceleration molecular dynamics[25] introduces further uncertainty to the definition of the transition state, as biasing forces can change the underlying energy landscapes. Recent studies have sought to bypass explicit identification of the TSE by combining biased dynamics methods with machine learning approaches that can identify relevant structural features of the transition in a data-driven manner.[26–29] While these could be promising approaches to predict rates for specific systems, it is difficult to extract direct structural insight into the TSE of the unbiased transition paths.

Here we apply an alternative enhanced sampling method that can generate long-timescale ligand unbinding events without applying biasing forces. The weighted ensemble method[30] is a general framework for path sampling where an ensemble of trajectories, each with an associated statistical weight, is evolved forward together in time.[31] Periodically, the ensemble of trajectories is "resampled": the number copies of each trajectory are changed in order to better direct the computational effort towards a pre-defined objective,[32] here, the generation of ligand unbinding events. We have previously shown that variants of the weighted ensemble algorithm[33,34] can efficiently generate a thorough sampling of possible ligand unbinding transition paths.[19,34–36] As these paths are all generated using the unbiased energy function, this provides us with the clearest window through which to examine properties of the TSE. We analyze the resulting trajectory sets with Markov state models (MSM)[37–39] that are constructed using the trajectory weights from the WE sampling method, and are history-augmented (haMSM,[40]) in that they only include trajectories originating from the bound state. This allows for quantitative predictions of ligand unbinding rate constants (k_{off}) both through the MSMs and directly from the WE simulations.[41–43]

The target protein studied here is the enzyme soluble epoxide hydrolase (sEH), which is present in mammalian tissues and metabolizes epoxy fatty acids (EpFAs) to their corresponding dihydroxy fatty acids.[44] EpFAs are a novel class of lipid mediators that play critical roles in blood pressure regulation, inflammation, pain perception and ER stress.[45]

Epoxyeicosatrienoic acid (EET), one of the most studied EpFAs, is antihypertensive, anti-inflammatory, analgesic and neuroprotective.[46] Inhibitors of sEH can raise EET levels and thus have been developed as potential treatments for these medical conditions.[47,48] As it has been previously shown that the residence time is a key quantity for determining the efficacy of sEH inhibitors in these contexts,[16,17] our long-term goal is the development of new molecules with longer sEH residence times.

The binding site of sEH is large and deeply buried inside the protein. The reference crystal structure (PDB ID: 4OD0), which is used in our study, reveals that the large binding site of sEH is separated into two compartments by a center pinch resulting from two flexible loops of the protein (Fig. 1). Previously we also used weighted ensemble simulation techniques to simulate the unbinding mechanism and estimate unbinding rates for 1-trifluoromethoxyphenyl-3-(1-propionylpiperidin-4-yl)-urea (TPPU), achieving a rough agreement with the experimental rate (42 seconds predicted versus 660 seconds experimental), and offering the first structural hints for the TSE.[35] We found that while the TSE was structurally diverse, there were a small number of specific protein ligand interactions that could potentially be targeted for kinetics-driven SBDD. However, a key question remained: could we extend this information to give us insight into the transition states of other structurally-related inhibitors? Put another way, how robust is the ligand binding transition state?

Here we tackle this question using simulations of five sEH inhibitors that are structurally related to TPPU. Their inhibition constants and residence times were measured experimentally and directly compared with residence times from simulations. We develop broad models of the ligand binding energy landscape for each ligand by maximizing the average agreement with experimental residence times. The transition path ensembles are then identified, analyzed and compared across the set of ligands, focusing on 1) spatial location, 2) ligand-protein interactions, 3) internal ligand degrees of freedom, and 4) solvent accessibility. We then discuss the implications of these results for kinetics-driven SBDD.

Methods

Soluble epoxide hydrolase inhibitors

The inhibitors studied here share a common piperidyl phenyl-urea scaffold with a few different functional groups at the open-ends of the scaffold. The N-atom of the piperidine group connects to a secondary keto-alkyl group. On the other side of the scaffold the phenyl group has substitution in the meta and/or para position. Fig. 1 describes the chemical structure of the inhibitor candidates. The inhibitor candidates (Ligands 10, 28, 29, 50 and 51) are referred to using an internal numbering scheme, which does not correspond to any ligand properties, or have any special significance.

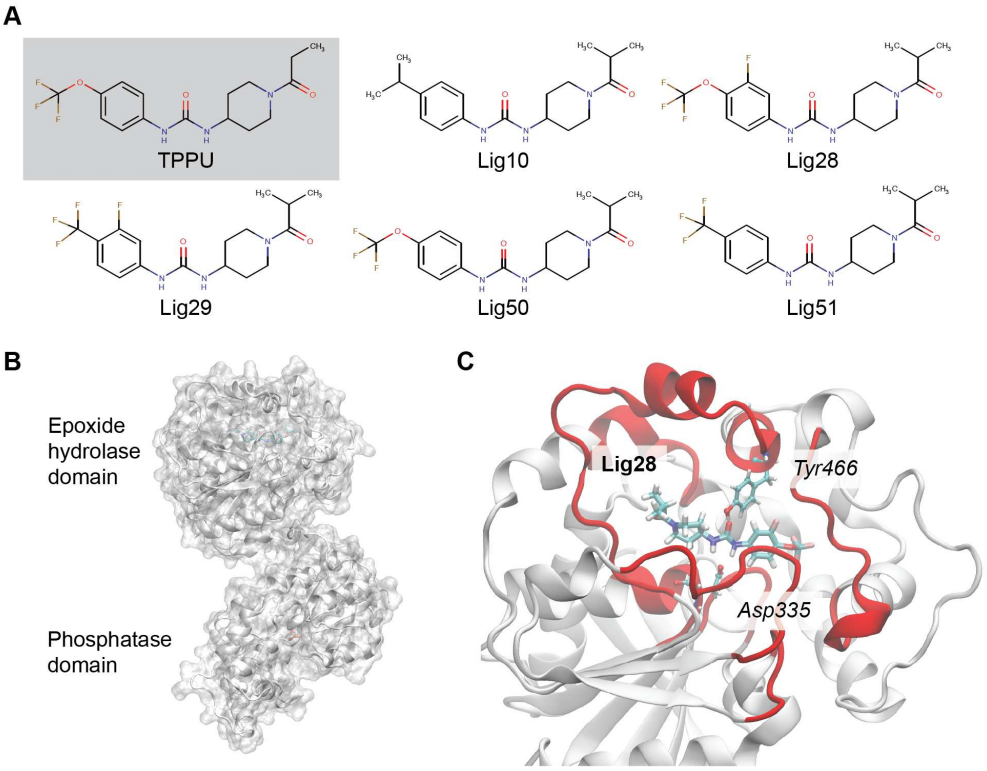

Figure 1: sEH protein and its inhibitors: (A) 2-D representation of sEH inhibitors, (B) The combined epoxide-phosphatase domain, (C) The Ligand28 bound sEH protein with the two most interacting residues in the bound state.

Experimental determination of rates

General information about the sourcing, synthesis and characterization of the ligand compounds is given in Supplemental Methods. The measurement of unbinding rates (k_{off}) for ligands 28, 29, 50 and 51 were previously reported.[16,49] The unbinding rate for ligand 10 is reported here and was determined using the same FRET-displacement assay. The sEH enzyme (8 μM) was pre-incubated with the selected inhibitor (8.8 μM, 100 mM PB buffer, pH 7.4, 0.01% gelatin) for 1.5 h at room temperature. The sEH-inhibitor complex was diluted 40 times with ACPU (20 μM, 100 mM sodium phosphate buffer, pH 7.4, 0.01% gelatin). The fluorescence ($\lambda_{\text{excitation}}$ at 280 nm, $\lambda_{\text{emission}}$ at 450 nm) intensity was monitored immediately for every 30 s up to 5100 s. The resulting $\lambda_{\text{emission}}$ versus time curve was fitted to a single exponential growth equation to calculate the relative k_{off}.

To prevent the leaching of fluorescence impurities from the plastic tube and loss of sEH inhibitor due to non-specific binding, the inhibitor stock solution (10 mM, DMSO) was stored in glass vials. All buffer used in this assay was filtered with a sterilized filtration unit (Millipore®Durapore PVDF Membrane, pore size: 0.22 um). All the measurements for the FRET-based displacement assays in a 96-well plate format was done in a Biotek Synergy Neo Plate Reader. To prevent non-specific binding of sEH or inhibitor on the 96-well plate, the 96-well plates were pre-incubated with PB buffer with 0.1% gelatin overnight at room temperature. The gelatin coats the plate and prevents the non-specific binding of sEH and sEH inhibitors to the plate. The buffer was discarded and the plate was dried before use.

System preparation for molecular dynamics

The bound pose of the inhibitors inside sEH were obtained by aligning the central scaffold of each inhibitor with the bound pose of TPPU in sEH from Ref. Lotz and Dickson[35] and then minimizing with CHARMM36 force field. This conserves the previously mentioned important bound pose interactions between the ligand and protein amino acids (Asp334, Tyr383 and Tyr466). The structure from PDB ID 4od0 was used to generate initial conformations for

all ligands examined here.[16] The catalytic domain was isolated by selecting residues 231 - 547 and preserving the positions of associated water molecules in the crystal structure. The ligands were parameterized using CGenFF.[50,51] We ensured that the key previously reported protein-ligand non-bonded interactions in the bound pose are conserved after alignment across all the ligands. A representation of the bound pose of one of the ligands (28) is provided in Fig. 1C. The system was solvated in TIP3 water up to a cutoff of 10 Å from the protein to the edge of the periodic box. The system was charge neutralized in slightly different ways across the ligand sets. Ligand 10 used 7 sodium atoms, ligands 28 and 29 used 7 potassium atoms, and ligands 50 and 51 added 17 potassium and 10 chlorine atoms to achieve an ionic strength of 150 mM. We do not expect that these differences had a significant impact on the kinetics or the transition paths.

The OpenMM simulation engine[52] was used for all of the minimization and dynamics steps in this work. CHARMM-GUI was used to generate the systems, as well as the scripts for minimization and heating.[53] The system was energy minimized using the L-BFGS algorithm with a maximum number of 5000 steps and an energy tolerance of 100 kcal/mol. The system was run at 303.15 K for 1 ns using a 0.001 ps timestep with harmonic positional restraints on the protein backbone (force constant 400 kJ/mol/nm^2) and protein sidechains (force constant 40 kJ/mol/nm^2). A force switch method was used to handle the non-bonded interactions with a switch-on distance of 10 Å and a switch-off distance of 12 Å. The Particle mesh Ewald method was used to handle the electrostatic cutoff, with an error tolerance of 0.0005. All covalent bonds with hydrogen were constrained. The protein restraints were then removed and the system was equilibrated for 10 ns using a 0.002 ps timestep. An isotropic Monte Carlo barostat was used to maintain a constant pressure of 1.0 bar, with a pressure coupling frequency of 100 steps. The final structure was used to initialize subsequent weighted ensemble simulations, which use the same simulation conditions as the second equilibration step.

Generation of ligand unbinding paths with the REVO weighted ensemble method

As mentioned above, weighted ensemble methods seek to shift the focus of the ensemble towards under-sampled regions. It achieves this by "cloning" certain members of the ensemble, dividing the weight of the parent walker to be distributed evenly across the clones. Typically, trajectories are run with a stochastic integrator, such as a Langevin integrator, so that the clones diverge to explore independent paths as the simulation continues. To save computational expense, pairs of trajectories can also be "merged". This typically occurs in over-sampled regions of space near local or global free energy minima. When two trajectories A and B are merged, the resulting walker takes on the sum of the weights $(w_A + w_B)$, and adopts either the conformation of walker A (with probability $w_A/(w_A + w_B)$) or walker B (with probability $w_B/(w_A + w_B)$). The exact nature of this random choice is important to ensure that the expectation value of the flow of probability is zero between any two regions of space.[54]

The REVO algorithm is a particular implementation of weighted ensemble that was designed to efficiently sample rare events while using an ensemble size that is as small as possible.[34] It achieves this by using a fixed ensemble size (here, 48) and proposing coupled merging and cloning events that are either accepted or rejected. To decide whether to accept these events it computes the value of an objective variable termed the "trajectory variation", V:

$$V = \sum_i \sum_j \frac{d_{ij}}{d_0}^{\alpha} \phi_i \phi_j \tag{1}$$

both before and after the proposed events. This quantifies the variation between members of the trajectory set, using a measurement of distance d_{ij}, which is discussed in the next paragraph. The constant d_0 is a characteristic distance to make the variation unit-less, but does not affect resampling outcomes. The function ϕ_i determines the importance of individual trajectories and is defined as a function of the walker weight: $\phi(w_i) = \log(w_i) -$

$\log(p_{min}/C)$, where p_{min} is a predefined minimum walker weight allowed in the simulation, and C is a constant, set here to 100 following previous work.[55–57] If the proposed cloning and merging event increases the value of V, then it is executed and another coupled merging and cloning event is proposed. This continues until V reaches a local maximum, at which point the ensemble is propagated forward in time by the molecular dynamics integrator. Here, a round of 20 ps of dynamics for each trajectory followed by a round of resampling is called a "cycle". For each ligand, we run between 5 and 6 independent runs, each containing at least 2000 cycles with an ensemble size of 48. The ensemble size was chosen to be large enough to capture a diversity of snapshots along the ligand unbinding pathway, while being as small as possible to enable extension of the runs as far as possible in time. For efficient implementation, an ensemble size that is divisible by the number of GPU cards on a node (in our case, 8) is also ideal. A summary of the number of cycles in each run is given in Table 1. In total, the results presented here combined 82.9 μs of total sampling time.

The distance between trajectories (d_{ij}) is calculated by aligning the binding site residues of the two trajectories and computing the root mean squared distance (RMSD) between the ligand atoms. The set of binding site residues are defined as those within a cutoff of 5.0 Å from the ligand in the equilibrated bound pose. This distance metric captures both 1) movements of the ligand with respect to the binding site and 2) movement of ligand internal degrees of freedom. By maximizing the variation with respect to this distance, we can enhance observations of not only one unbinding path, but a broad ensemble of ligand unbinding paths.

The simulations were run in the "unbinding ensemble", in which trajectories originate in the bound state and are terminated when they cross into the unbound state, which at run time was defined as having at least 10 Å of clearance between the ligand and the protein. Trajectories that unbind are re-initialized in the bound state, but keeping the same weight. In practice, these are quickly merged into other high weight trajectories in the bound state

Table 1: Details of the REVO simulations run for each ligand. The aggregated sampling is summed across all trajectories in a given run.

Lig. ID	Num. Walkers	Run Index	Num. Cycles	Aggregated sampling (μs)
Lig10	48	0	2986	2.97
		1	2751	2.64
		2	2453	2.35
		3	2224	2.13
		4	2450	2.35
		5	2470	2.37
Lig28	48	0	3257	3.13
		1	3295	3.16
		2	3274	3.14
		3	3294	3.16
		4	3275	3.14
Lig29	48	0	4000	3.84
		1	4000	3.84
		2	3885	3.73
		3	4000	3.84
		4	4000	3.84
Lig50	48	0	4000	3.84
		1	4000	3.84
		2	3414	3.28
		3	2799	2.69
		4	3518	3.38
Lig51	48	0	3400	3.26
		1	3400	3.26
		2	3400	3.26
		3	3400	3.26
		4	3400	3.26

by the resampling algorithm. The simulations were conducted using the Wepy software,[58] which is a Python implementation of the REVO resampler.

Markov state modeling and transition state definition

An overview of the Markov state modeling workflow is given in Fig. 2. Each frame of each trajectory is projected onto a set of features, which form the basis of the Markov state models used to define the transition state ensembles and calculate the unbinding rates. The features are a set of 336 distances calculated between a set of 56 backbone atoms in the binding site from residues Phe267, Asp335, Trp336, Ile363, Phe381, Tyr383, Gln384, Phe387, Met419, Leu328, Tyr466, Val498, Leu499, and Met503 and six ligand atoms (as shown in Fig. S1) for all the frames in each simulation trajectories. The ligand atoms were chosen along the common central scaffold to facilitate comparison of features between ligands.

For analyses where time-independent component analysis (tICA[59,60]) is used, we process the features into a Deeptime[61] time-lagged dataset object using the `sliding_windows` function from Wepy. Separately for each ligand, we randomly choose a subset of 500,000 time-lagged dataset members to train a tICA model, which is then used to transform the entire dataset for that ligand. The reduced tICA variables are clustered into a number of states using the KMeans algorithm and each frame of the trajectories is labeled with a cluster index. For analyses where tICA is not used, the complete set of feature vectors are used as the basis for clustering with KMeans.

In both cases, transition count matrices are build by counting the inter-state transitions between two states across a certain lag time, again using the `sliding_windows` function from Wepy. Unless specified otherwise below, transitions contribute to the count matrix according to the weight of the trajectory at the end of the time interval. These are used to generate conformation space network ("CSN") objects from the CSNAnalysis package.[62] For all clusters we compute the average ligand RMSD to the initial reference structure. If this RMSD value is less than 2.5 Å, the cluster is labeled as "bound". We also compute the

minimum distance $d_{\text{unb-min}}$ between the ligand and the sEH binding site for each frame in the simulation. If any members of a cluster have $d_{\text{unb-min}} > 5.0$ Å then the cluster is labeled as "unbound". Note that this is a more relaxed definition than the minimum distance of 10 Å used during run time. We consider the $d_{\text{unb-min}} > 5.0$ Å to be more appropriate for describing unbinding rates that are measured by competitive binding assays, as it is more sensitive as to whether the binding site has been vacated. Using these definitions of the bound and unbound basins, unbinding committor probabilities[63] and the unbinding mean first passage times are computed using CSNAnalysis. Transition state ensembles were chosen by selecting all conformations belonging to clusters within a given committor range across the unbinding committor of 0.5.

Direct rate calculation

In addition to the rates from the Markov state models we also directly calculate unbinding rates using the flux into the unbound state. The flux is simply calculated as the sum of the weights of the unbinding trajectories divided by the elapsed time.

$$k_{\text{off}} = \frac{\sum_{(i,t)\in\mathcal{U}} w_{i,t}}{T} \tag{2}$$

where $\mathcal{U}$ is the set of tuples denoting the trajectory indices i and the time points t where the unbinding events occurred, $w_{i,t}$ is the weight of the trajectory at that time point, and T is the total elapsed time of the simulation. Eq. 2 is also known as the "Hill's equation".

For comparison, we also compute the fluxes corresponding to the more relaxed unbound state definition: $d_{\text{unb-min}} > 5.0$ Å, defined in the previous subsection. This is done by identifying the set of crossing points into the relaxed unbound state ($\mathcal{U}^*$). To mimic an absorbing boundary condition, we add crossing points starting from cycle 0 and only add a new crossing point if none of its predecessors have been added to $\mathcal{U}^*$.

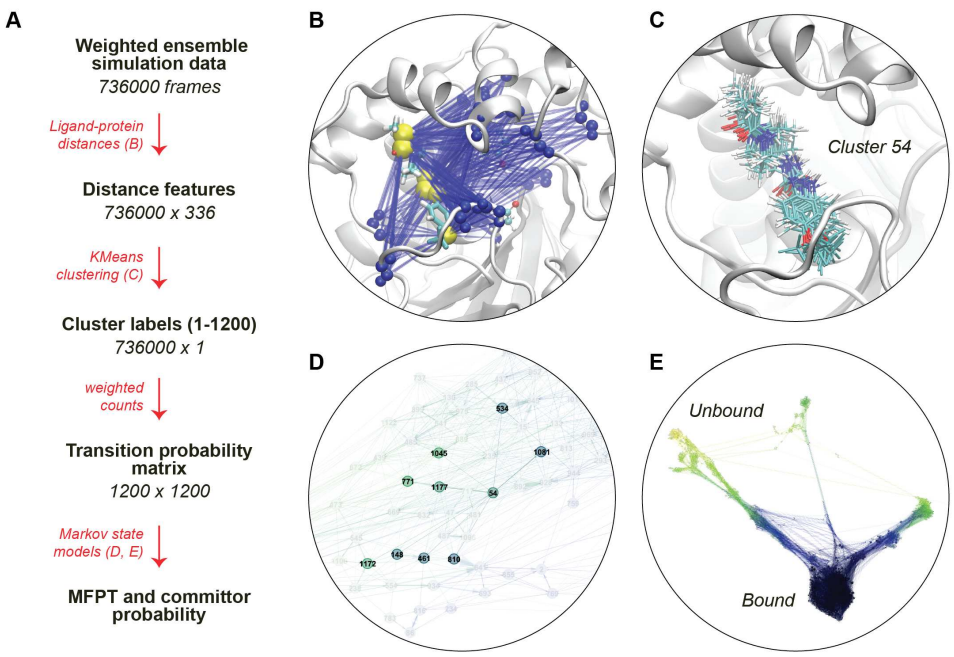

Figure 2: (A) The scheme of extracting TSE from weighted ensemble MD data with Markov state models. (B) Visual representation of the interatomic ligand-binding site distance features for a ligand. (C) All ligand conformations from a particular microstate after clustering all the frames based on distance features. (D) Microstates and their connectivity from a transition probability matrix, (E) Conformational space network (CSN) of ligand 50, with ligand RMSD being the scale of color. The densely populated bound state is shown in dark blue and the sparsely connected unbound states are shown in yellow/red.

Results

Unbinding Pathways of five sEH ligands

For each of the five ligands, we performed a set of weighted ensemble simulations with the REVO algorithm that starts from an equilibrated bound pose. We observed unbinding events for all five ligands that we simulated. In our study, a ligand is defined to be unbound from the protein when it has a minimum distance of 5 Å or more from the residues that make up the binding site (Fig. S2). The number of unbinding events for each ligand along with the cumulative unbinding weights and simulation time are provided in the Table 2. The number of cycles in each run varies, ranging from from 2224 to 4000 (see Table 1), with all the ligands having achieved at least 14.7 μs of combined sampling.

Table 2: Details of the REVO simulations for each ligand. The cumulative simulation time for each ligand is combined over all 48 trajectories in each independent run. The total unbinding weight is the sum of the weights of all trajectories at the point of unbinding. Specific unbinding events were manually labeled as utilizing either the "left" or "right" pathway to determine the weight fractions.

Lig. ID	Cumulative sim. time (μs)	No of unb. events	Runs_{unb} / $\text{Runs}_{\text{total}}$	Total unb. weight	Weight fraction (left pathway)	Weight fraction (right pathway)
Lig10	14.72	34	2/6	3.39e-10	0.000	1.000
Lig28	15.74	7	2/5	1.19e-11	0.499	0.501
Lig29	19.09	4	1/5	6.92e-12	1.000	0.000
Lig50	17.02	8	3/5	3.28e-10	0.993	0.007
Lig51	16.32	4	2/5	1.45e-07	0.000	1.000

Upon inspection of the unbinding trajectories for each ligand, we noticed that the ligand unbinding in sEH can be broadly divided into two categories: (i) Unbinding through the right side of the cavity, (ii) unbinding through left side of the cavity of sEH. For Ligand 10 and 51, all of their unbinding events occur through the right side of the cavity, whereas ligand 29 accesses the left side only during unbinding. Ligands 28 and 50 have at least one unbinding pathway through either side of the cavity. The right and left sides of the sEH cavity are highlighted in Fig. S3 and exemplary unbinding trajectories through each cavity are shown in Fig. S4. In a weighted ensemble simulation, the cavity-specificity of a ligand should not be determined based on the mere existence of binding paths, but should take into account the relative weights of the trajectories that exit through those pathways. For example, the weight fractions from Table 2 indicate that despite having at least one unbinding pathway through the right, ligand 50 will almost always preferentially unbind through the left cavity. In contrast, we find that ligand 28 has an almost equal probability of unbinding through either of the cavities.

Fig. 3 shows the conformation space networks (CSN) of all the ligands, where each node represents a particular cluster of ligand-protein conformations. These are obtained from the same transition probability matrices used to build Markov state models, as described

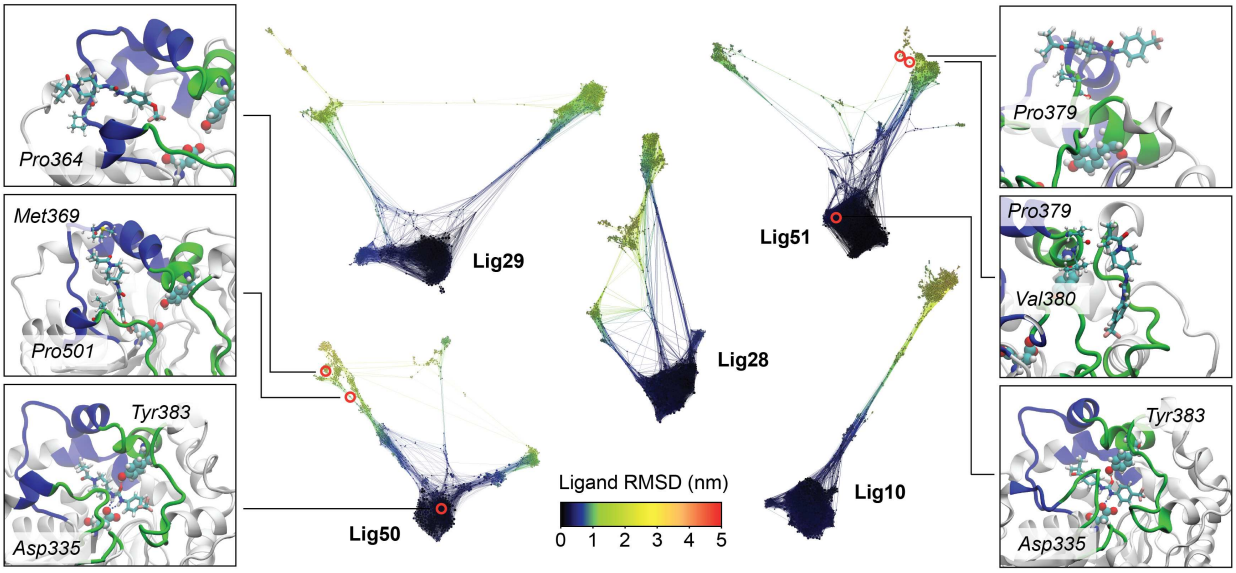

Figure 3: Conformational space networks of ligand unbinding from sEH in the scale of ligand RMSD. Three frames from the most probable unbinding pathways are highlighted for ligand 50 (cavity specificity: left) and ligand 51 (cavity specificity: right). The states corresponding to those frames are highlighted in the CSNs. In each panel, the ligands are shown in licorice while the amino acid residues within 2.5 Å of the ligands are depicted in CPK representation, with the binding site Asp335 and Tyr383 highlighted in VdW representation.

in Methods. Each network shows 1200 nodes that are colored according to ligand RMSD, with the dark blue clusters corresponding to the bound states. The networks are oriented such that the left and right branches extending from the bound region in each network correspond to structures with the ligand occupying the left and right cavities. This reveals that although ligands 29 and 51 did not register full unbinding trajectories through the right and left cavities, respectively, we were able to sample trajectories that progressed along both directions. In contrast, ligand 10 only accesses the right side of the cavity and hence has no left branch.

The unbinding pathways with the highest weights contribute most strongly to the transition path ensemble. We present three snapshots from the highest-weighted unbinding pathways obtained for ligand 51 (right panels) and ligand 50 (left panels) in Fig. 3. The snapshots at the bottom show the bound pose for both the ligands. The middle snapshots on each side show vertical ligand poses inside the cavity which is a characteristic feature for

all unbinding pathways of all the ligands. The top snapshots are further along the pathway and show only sparse interactions with the outer surface of the protein. Additional sets of snapshots for ligands 10 and 28 are shown in Fig. S5. Interestingly, ligand 29 has equal unbinding probability in all four of its pathways. These unbinding events all originated from the same starting trajectory, sharing a common vertical pose as an unbinding intermediate (Fig. S6).

We observe a slight correlation between the total weight of the unbinding trajectories and the number of unbinding events generated. Opposing this trend, ligand 51 showed the highest total unbinding weight, but registered only 4 unbinding events. As trajectories within a run are inter-related through cloning and merging events in the weighted ensemble algorithm, not all of the unbinding events are independent observations. However, unbinding events between runs can be considered to be completely independent. For all ligands except for ligand 29, we obtained unbinding events from at least two runs.

Kinetics of Ligand Unbinding

Rates of unbinding can be calculated either directly from the sum of the transition rates (Eq. 2), or indirectly through construction of a Markov state model. These can be compared with unbinding rates determined experimentally using a FRET-displacement assay, which correspond to mean first passage times ranging from 14 to 32 min. We note that these MFPTs are at least tens of millions of times longer than the cumulative simulation times from our MD simulations. Despite this extreme difference of timescales, we are able to achieve root mean squared log-errors (RMSLE) of 2.3 for our direct rate calculations and 0.9 for our Markov state model results. The latter indicates an average agreement with experimental quantities that is within an order of magnitude. Values for experimental and computational MFPTs are summarized in Table 3 and log-scale RMSEs are shown in Fig. 4.

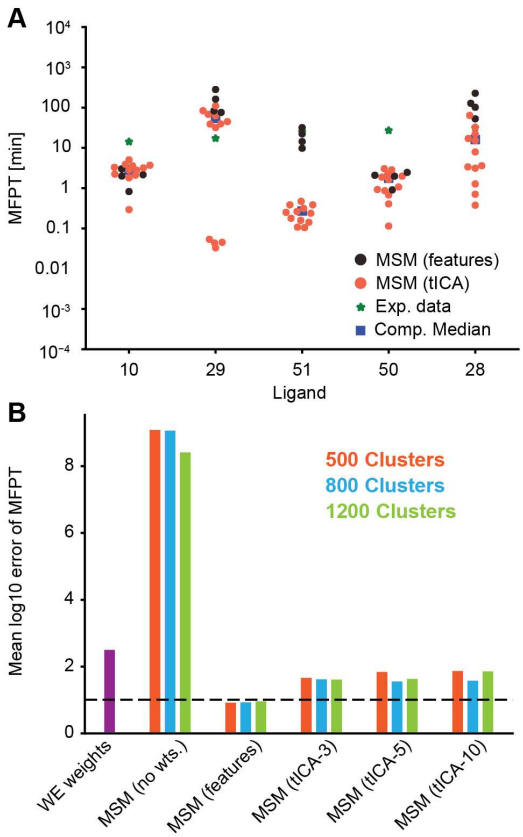

Figure 4: (A) Computational prediction of MFPTs for each ligand with all Markov state models are plotted together as a swarmplot. The full feature space MSMs (black circles) perform better compared to tICA-based MSMs (red circles). The experimental data (green asterices) are plotted along with the median of all computational estimates (blue boxes). (B) Comparison of the root mean squared log-10 error plotted for various MSMs (red, blue and green) and direct estimates from WE weights (violet). The MSM using the full feature set has the lowest RMS log-10 error, while the unweighted MSM has the highest RMS log-10 error. The horizontal line marks an average error of one order of magnitude.

The larger error for our direct rate calculations is primarily due to the inadequate sampling of exit points. In the direct rate calculation method, the rates are entirely dependent on the weights of the unbinding trajectories. Although we observe unbinding events for all the ligands, the number of these events is limited to only a handful (a range of 4 to 34) per ligand. On the other hand, MSMs are not as susceptible to the sampling of the unbinding events, as they are built with all the simulation data and consider the non-reactive trajectories as well. Hence, trajectories far away from the bound pose that may not go on to unbind because of the finite length of the simulation will still contribute statistically in rates

Table 3: MFPT (in minutes) of ligand unbinding by experimental assay, Hill's equation and Markov state modeling. The "best model" for the WE-MSM uses the full feature set (without tICA) and 1200 clusters.

Lig. ID	Expt. MFPT (unbinding kinetics assay)	Comp. MFPT (Hill's eq., 5 Å cutoff)	Comp. MFPT (Hill's eq., 10 Å cutoff)	Comp. MFPT (WE-MSM, best model)	Comp. MFPT (WE-MSM, median)
Lig10	14.25	723.44	1346.9	3.1	2.1
Lig28	31.75	21984.8	40971.1	227.1	115.1
Lig29	17.31	45952.9	45952.9	282.1	122.7
Lig50	27.14	864.1	1083.3	1.2	1.1
Lig51	25.37	1.87	12512.8	11.6	6.7

determined from the MSM.

While predicting the MFPTs, we have considered both (i) tICA-based MSMs and (ii) feature distance based MSMs. Fig. 4A is a swarmplot showing MFPTs estimated from all MSMs constructed for each ligand. The experimental data and the median of the computational estimates are also plotted. The RMSLE between the experimental data and the computational medians is 0.97. We observe that the tICA-based models perform worse than the full feature distance clustered MSMs, particularly for ligand 51. We discussed probable reasons for this below in the Discussion section. We also calculated MFPTs from MSMs where the transitions in the counts matrix were not scaled by the weighted ensemble probability, rather, all transitions between microstates were given an equal weight of 1. As shown in Fig. 4B, unweighted MSMs deviate from the MFPTs by over 8 orders of magnitude. This is due to a systematic underestimation of the MFPT, resulting from improper weighting of state-to-state transitions in the MSM.

Both types of weighted MSMs perform significantly better than the MFPT calculated from only the unbinding flux. The feature distance based MSMs ("no-TICA") have RMSLE 0.93 for all three different cluster numbers, while tICA based MSMs have a higher RMSLE of 1.8. In examining different numbers of clusters for the MSM, we find that 500, 800 and 1200 all have similarly low RMSLEs. We find MSMs with 1200 clusters have the highest Spearman's rank coefficient and Kendall's rank coefficient (Table S1). MFPTs from these

MSMs are shown in Table 3 as "WE-MSM best model".

Robustness of Transition State Ensembles

The structural determinants that underlie the unbinding rates are found in the transition state ensembles (TSEs) for each ligand. We developed a workflow to isolate and characterize the TSEs from the weighted ensemble simulation data (Fig. 2), which was guided by the kinetics results in the previous section. We use a Markov state model constructed using the complete set of distance features (e.g. no tICA) and 1200 states, as it resulted in the best agreement with experimental unbinding rates. However, the number of states was not found to significantly impact the location of the TSE (Fig. S7). We also estimated the variance among the TSE ligand poses and their average RMSD to the bound state and observed that those structural parameters are similar across MSMs built with different numbers of microstates (Tables S2 and S3).

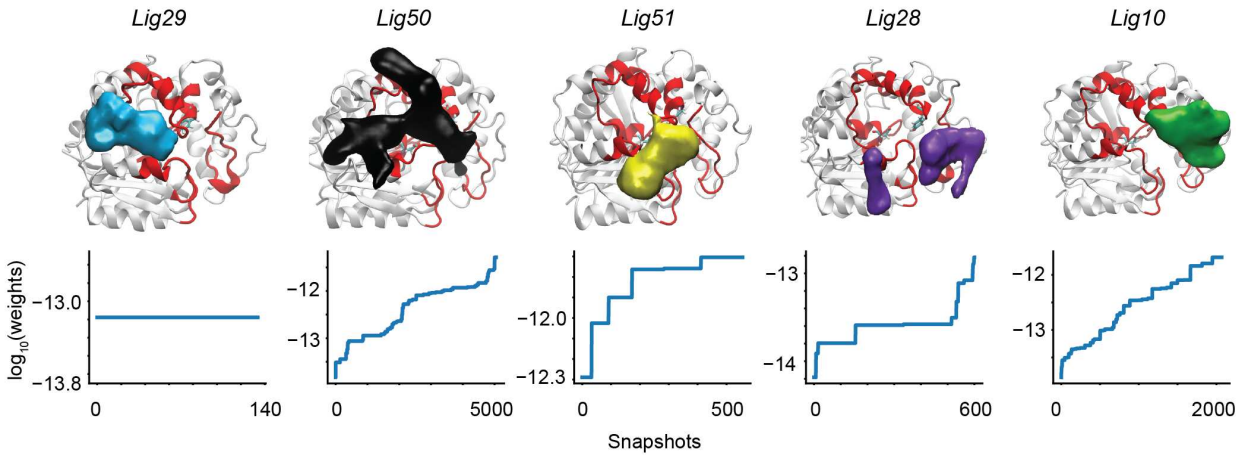

Figure 5: (Top row) Density plots of the ligand unbinding TSEs: different ligands are plotted in different colors. The two binding site residues Asp335 and Tyr383 are shown in licorice representation, while the overall binding region is highlighted by red color. Each surface is plotted in VMD using the same density cutoff ("isovalue") of 0.05. (Bottom row) Weights of conformations used to build the TS ensembles are plotted in log-scale for each ligand. The horizontal axis shows the number of independent snapshots in the TSE. The vertical axis shows the $\log_{10}$ of the weight of that snapshot.

Density volume maps for all TSEs in sEH are illustrated in Fig. 5. These TSEs were

constructed using all trajectory frames that were assigned to a cluster with a committor probability in the range $(0.3, 0.7)$. Fig. 5B shows the distribution of probabilities of these trajectory frames, which were computed using the equilibrium probabilities of the clusters computed from the MSM. Although the ligands share a common scaffold, the spatial density of the TSEs show significant variation. These densities are in accordance with the corresponding most probable unbinding pathways mentioned earlier. Ligands 10 and 51 have spatial density primarily on the right side of the cavity, ligand 29 is mostly to the left side, while ligands 28 and 50 are distributed across both sides. The numbers of snapshots in the TSEs range from 140 to 5000. Higher numbers of snapshots indicate better sampling of transition paths, though not necessarily higher unbinding rates. We do not observe a strong relationship between the number of shapshots and the volume of the TSE density plots. This is expected as the TSEs are generated probabilistically using the MSM weights and the density plots show only the regions of space with a probability density above a cutoff of 0.05 for all ligands. Generally, we find that the TSEs of all the ligands are structurally closer to the unbound ensemble compared to the bound ensemble. This can be observed in the probability distributions of ligand RMSDs within the bound, TS and unbound ensembles (Fig. S8). The heterogeneity in the TSEs can lead to a wide variety of specific sEH-ligand interactions, which we study next.

Fig. 6A shows the protein residues with the most stable contacts with the ligand for the bound ensemble and the TSE. We define a contact to be present when the minimum distance for a residue-ligand pair is below 2.5 Å; the fraction of frames in which the contact is formed is shown as a heatmap. The cutoff is carefully chosen based on the maximum range of H-bond (2.2 Å) and VdW ($\sim$ 3.0 Å) interatomic distances. For both the bound state and transition state, we show all residues that have a probability of interaction greater than 5% for at least one of the ligands. The axis denotes the ligand IDs with the right cavity specific ligands (10, 28 and 51) positioned first followed by the left cavity specific ligands (29 and 50). For the TSE heatmap, the residues are arranged so as to move from residues

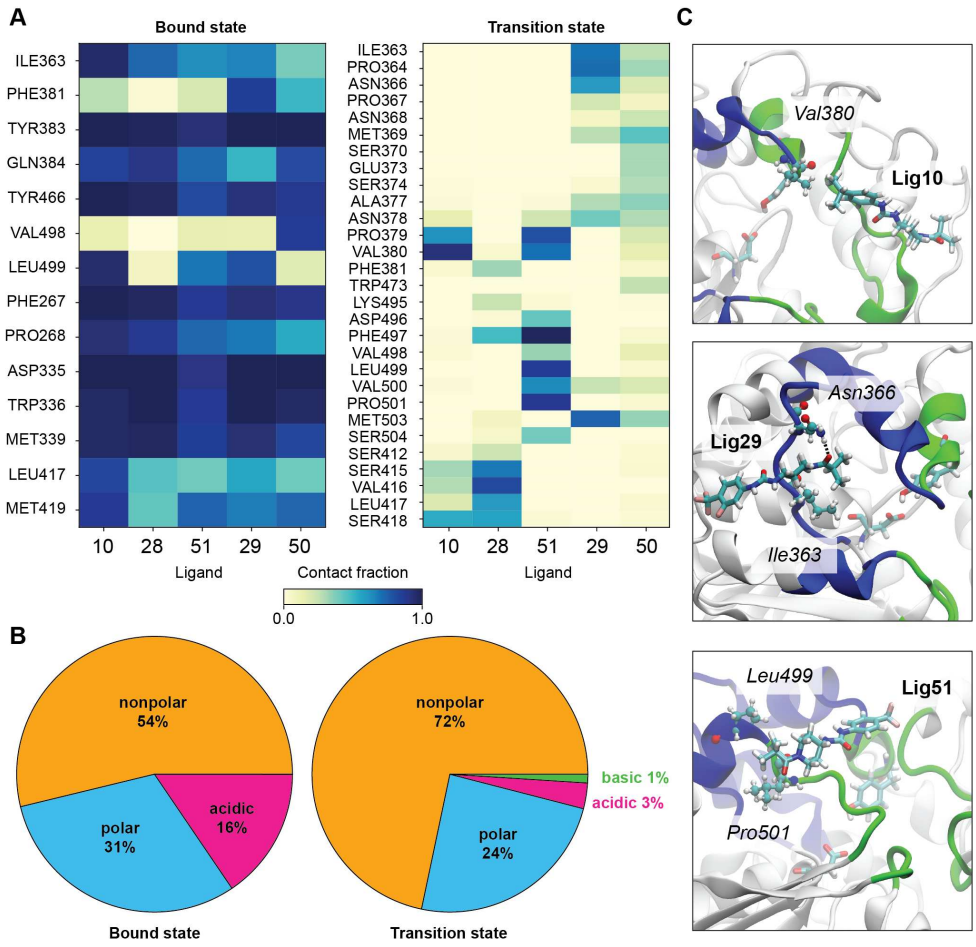

Figure 6: Protein-ligand interactions in the ligand unbinding TS ensembles of sEH: (A) Heatmap of interaction probabilities in bound and TSE for the ligands, colorbar denotes the measure of probability, (B) Pie-chart describing the category of protein-ligand interactions based on the type of the amino acids, averaged over all the ligands, (C) Representations of a few of the most probable interactions in TS ensembles for ligands 10, 29 and 51.

in the left cavity (on top) to the middle region to the right cavity (on the bottom). In the bound ensemble, the protein-ligand interactions are largely consistent from one ligand to another, with binding site residues Tyr383 and Asp335 having the maximum probabilities of contact formation. The protein-ligand interactions are substantially more varied in the TS ensembles. The types of amino acids and atomic interactions also change considerably: we notice that non-polar amino acids such as VAL(380, 500, 416), PRO(364, 379, 501), ILE(363), LEU(499), PHE(497) are more highly represented in the TSE, in contrast to interactions with polar amino acids. This is quantified in Fig. 6B, which shows protein-

ligand interactions based on the type of the amino acids, averaged over all the ligands. It is evident that – moving from the bound ensemble to the TSE – there is a significant increase (18%) in the interactions with non-polar residues, which are predominantly hydrophobic interactions. Visual representations of some of the most probable contacts in ligand 10, 29 and 51 are provided in Fig. 6C. These snapshots show the high ligand-to-ligand variation in the specific interactions formed in the TSE as well as the predominance of hydrophobic interactions. Interestingly, although the specific interactions vary from ligand-to-ligand, the shift towards hydrophobic protein-ligand interactions in the TSE is consistent across the set of ligands examined here (Fig. S9).

Irrespective of the cavity preference of the ligand unbinding TSEs, all ligands have substantial hydrophobic interaction with sEH. The non-polar isopropyl group attached to the aryl end of ligand 10 has a higher hydrophobicity compared to mildly lipophilic -OCF3 or -CF3 counterparts connected to other ligands. Consistently, we find that this isopropyl group has stable hydrophobic interactions with sEH non-polar residues such as Pro379, Val380 (Fig. 6A, C) in the TSE, while in other ligand TSEs, the piperidyl end primarily accounts for the protein-ligand hydrophobic interactions. Fig. 6C illustrates the interactions between the piperidyl end of ligand 29 and 51 with non-polar residues such as Ile363, Leu499 and Pro501. Interestingly, we notice that protein-ligand interactions mediated by -OCF3 or -CF3 groups are not significant in the TSE. From the perspective of rational kinetics driven drug discovery, to increase the residence time one can consider chemical changes to the ligand that destabilize favorable interactions in the TSE. The shift from isopropyl to -OCF3 groups could be seen in this context, where -OCF3 disrupts some favorable transition state interactions without destabilizing the bound ensemble, thus contributing to a longer residence time.

In the context of structure based drug design, it is also important to understand the changes in the ligand degrees of freedom along the transition pathway. For instance, differences in the orientation of a rotatable bond between the bound state and TSE could be

exploited to destabilize transition states, leading to longer residence times. Here we examine a set of eight common rotatable bonds and we measure their corresponding angular probability distributions in both the bound and transition state ensembles. All the rotatable bonds are illustrated in Fig. S10, with the four constituent atoms highlighted. Wasserstein distances are computed between the bound and TS probability distributions for each rotatable bond across all the ligands (Fig. S11). This is a metric of dissimilarity between two histograms, with a higher value indicating higher dissimilarity between the distributions. We find that the C1-N2-C2-C3 rotatable bond has the most significant dissimilarity in the bound and TS ensemble (Fig. 7). This dihedral is more restricted in the bound ensemble due to steric effects for all of the ligands examined here. In the TSE, these distributions are substantially broadened, showing a more heterogeneous ensemble of conformations. It is interesting to note that the variation in the rotatable bond in TSE is unidirectional in nature, i.e., the lower bound of the angle decreases in the TSE compared to bound ensemble. The differences between these distributions implies that making chemical changes that hinder or freeze this rotational degree of freedom could entropically destabilize the transition state (as well as the unbound state) with respect to the bound state.

Discussion

Mean first passage times of pharmacologically relevant ligands often occur on time-scales ranging up to hours in length. Although molecular dynamics can be a powerful tool to characterize biomolecular processes and predict mean first passage times, it is still a challenge even for state-of-art enhanced sampling methods to model long-timescale unbinding events with statistical significance. We have previously found that the enhanced sampling method used here ("REVO") performs excellently for standard millisecond-timescale protein-ligand unbinding events such as trypsin-benzamidine. Unbinding transition rates for this system quickly and reproducably converged to within an order of magnitude of the experimental

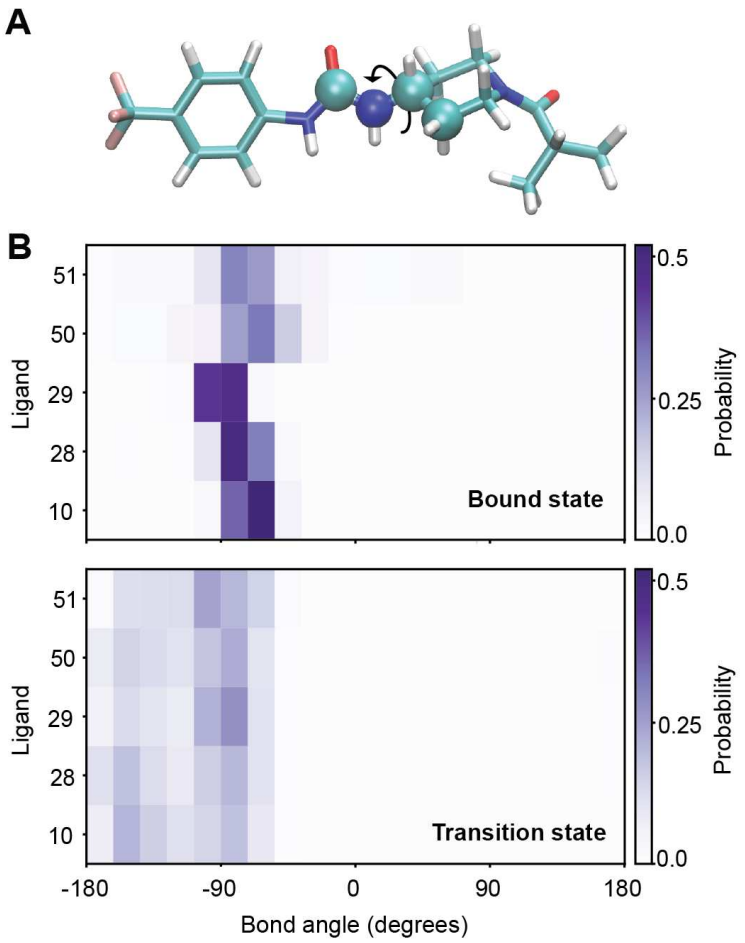

Figure 7: (A) The atoms corresponding to the rotatable bond with the largest difference between the bound ensemble and the TSE are shown in van der Waals representation, with other atoms in licorice. (B) Probability distributions of this angle are shown for each ligand in both the bound ensemble (top) and TSE (bottom).

value.[19,34] The sEH-inhibitor systems studied here pose additional challenges, as the ligands are larger, with more rotatable bonds, and the sEH binding pocket is deeply buried, which requires a multi-step unbinding process. These systems thus display a significantly increased complexity and variety in the unbinding mechanism. In addition, the experimental values for these mean first passage times are six orders of magnitude larger than the trypsin-benzamidine system. Here, using a combination of REVO and Markov state modeling, our root mean squared log error (RMSLE) in the unbinding rate averaged 0.94, indicating that our average agreement with experiment was also within an order of magnitude. This was

much better than we originally anticipated, given that our individual trajectories are only tens of nanoseconds in length.

With a total of 82.9 μs of ligand unbinding REVO simulations, we obtain a handful of unbinding events for five pharmacologically relevant ligands from sEH protein. Although the weights of these trajectories can be used to directly compute MFPTs using the Hill relation, the low number of trajectories results in a high uncertainty and increased RMSLE > 2 compared to experimental values. This is consistent with previous results on the sEH-TPPU system studied with the WExplore algorithm,[33,35] where the MFPT (42 s) underestimated the experimental value (660 s) by more than 15 times. As all of the underlying trajectories are generated without biasing forces, we can use them to build a history-augmented Markov state model ("haMSM"[40]) where all trajectories originate in the bound state. The rates calculated by these haMSMs are potentially more accurate and robust, as they take into account not only the small set of fully reactive trajectories, but all of the transitions from the non-reactive trajectories as well. Here we find that the RMSLE of MFPTs in the MSM-REVO scheme are significantly lower by more than an order of magnitude than the directly estimated MFPTs from the weights of reactive trajectories. An important note is that the trajectory weights from the REVO simulations were used to build the transition count matrix of the haMSM. This led to a dramatic reduction in the RMSLE and demonstrates the powerful synergy of weighted ensemble methods with Markov state modeling.

In the context of Markov state modeling, dimensionality reduction is particularly important to account for redundancy and noise in the features before they are clustered. For studying kinetic properties associated with slow dynamical motions in biological systems, tICA has been an excellent tool to identify important collective variables.[64,65] When dimension reduction is carried out with 5 or 10 time-lagged independent components, we notice that tIC-1 separates the bound and unbound frames and tIC-2 distinguishes the cavity specificity consistently for all the ligands (Fig. S12). However, the RMSLE in MFPTs obtained from haMSMs built with clustered tICA data is higher compared to its full feature space

counterpart. Moreover, the variation of MFPTs obtained from tIC-clustered MSMs is much larger compared to variation of MFPTs obtained from the full feature space MSMs (Fig. 4A). Although we are unable to identify a particular reason behind the lower accuracy of MSMs with tICA components, it should be noted that our feature space itself (the set of ligand-protein distances) is carefully chosen to describe the ligand unbinding process. Hence, although our feature set was highly redundant, we found that the transformation into a smaller set of linearly independent components decreased the quality of the clustering, likely grouping together trajectories that were less similar in their unbinding committor values. Dimension reduction schemes, including tICA, but also machine learning approaches[66] such as VAMPnets[67] or RAVE[68] could be more useful for more heterogeneous sets of input features, such as those that describe solvation, ion densities, distances and ligand degrees of freedom. These approaches, and their combination with weighted ensemble sampling algorithms, are the subject of ongoing work.

The mean first passage time depends upon the bound to transition state activation energy barrier. Hence to engineer ligands with higher residence time, one needs to understand how changes to the ligand will differentially affect the bound and transition state ensembles. This can include both protein-ligand molecular interactions as well as conformational changes of the ligand. The five ligands have an identical aryl piperidyl-urea scaffold, which may lead to the assumption that the transition states along the ligand unbinding pathways could be similar for these ligands. However, in our molecular simulations we find many differences between the transition paths. We did not observe a connection between the path specificity and the MFPT: ligands with both shorter (ligand 10) and longer (ligand 28) MFPTs were found to unbind through the right-side transition path. The location of the transition state as well as the specific ligand-protein interactions formed, varied considerably from ligand-to-ligand. However, we find a number of similarities in the transition state ensembles that could be exploited for kinetics-driven drug design. TSEs for all of the ligands show an increased extent of hydrophobic interactions. Additionally, the probability distributions along the

C1-N2-C2-C3 dihedral angle show a similar shift between the ligands. Finally, although not completely conserved, the most probable protein-ligand contacts in TSEs show some common elements. Ligands 10, 28 and 51 unbind through the right side of the sEH cavity and have substantial contacts with Pro739, Val380 and Phe497 , while the ligands preferring the left side of the cavity (29 and 50) have contacts with Pro364 and Met503.

These findings present a mixed outlook for kinetics-oriented drug design. On the one hand, the diversity of specific protein-ligand interactions formed in the TSE from ligand-to-ligand makes attempts to rationally modulate the strength of TSE interactions unfeasible. More rigorous attempts such as free energy perturbation calculations for entire ensembles of bound and transition state structures will likely also suffer from poor overlap of the transition state ensembles between ligands. On the other hand, we have identified some structural properties of the TSEs that are consistent across all ligands examined here. The rotatable bond C1-N2-C2-C3 is rigid in the bound ensemble but shows considerable fluctuation in the transition state ensemble. This suggests that restricting rotation of this bond could stabilize the bound state with respect to the transition state, increasing the free energy barrier to dissociation.

A characteristic feature of the sEH binding pocket is a deeply buried binding cavity. As a result, ligand unbinding is hindered by multiple stable interactions along the pathway. The depth of the binding site in the cavity could increase the number of probable unbinding pathways, making it more difficult to thoroughly sample the transition state ensemble. Moreover, contrary to intuition, we observe that even when the bound state ligand-protein contacts are completely broken, the ligands can still be far from committing to the unbound states. This results in TS ensembles for all the ligands that are closer to the unbound state, on the surface of the protein. It remains to be seen whether TSEs from shallower protein-ligand interactions will have similar characteristics in terms of solvent accessibility and conformational heterogeneity. It is possible that the TSE of a moiety unbinding from a protein with a shallower binding site can be more robust, with a more focused set of unbinding pathways.

References

(1) Greer, J.; Erickson, J. W.; Baldwin, J. J.; Varney, M. D. Application of the Three-Dimensional Structures of Protein Target Molecules in Structure-Based Drug Design. *Journal of Medicinal Chemistry* **1994**, *37*.

(2) Rutenber, E. E.; Stroud, R. M. Binding of the anticancer drug ZD1694 to E. coli thymidylate synthase: Assessing specificity and affinity. *Structure* **1996**, *4*.

(3) Wlodawer, A.; Vondrasek, J. Inhibitors of HIV-1 protease: A major success of structure-assisted drug design. *Annual Review of Biophysics and Biomolecular Structure* **1998**, *27*.

(4) Clark, D. E. What has computer-aided molecular design ever done for drug discovery? *Expert Opinion on Drug Discovery* **2006**, *1*.

(5) Veldkamp, C. T.; Ziarek, J. J.; Peterson, F. C.; Chen, Y.; Volkman, B. F. Targeting SDF-1/CXCL12 with a Ligand That Prevents Activation of CXCR4 through Structure-Based Drug Design. *Journal of the American Chemical Society* **2010**, *132*, 7242–7243.

(6) Sadybekov, A. V.; Katritch, V. Computational approaches streamlining drug discovery. *Nature* **2023**, *616*, 673–685.

(7) Gentile, F.; Yaacoub, J. C.; Gleave, J.; Fernandez, M.; Ton, A. T.; Ban, F.; Stern, A.; Cherkasov, A. Artificial intelligence–enabled virtual screening of ultra-large chemical libraries with deep docking. *Nature Protocols* **2022**, *17*.

(8) Chisholm, T. S.; Mackey, M.; Hunter, C. A. Discovery of High-Affinity Amyloid Ligands Using a Ligand-Based Virtual Screening Pipeline. *Journal of the American Chemical Society* **2023**, *145*, 15936–15950.

(9) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery* **2015**, *0441*, 1–13, NULL.

(10) Chen, W.; Cui, D.; Jerome, S. V.; Michino, M.; Lenselink, E. B.; Huggins, D. J.; Beautrait, A.; Vendome, J.; Abel, R.; Friesner, R. A.; Wang, L. Enhancing Hit Discovery in Virtual Screening through Absolute Protein–Ligand Binding Free-Energy Calculations. *Journal of Chemical Information and Modeling* **2023**, *63*, 3171–3185.

(11) de Vaca, I. C.; Zarzuela, R.; Tirado-Rives, J.; Jorgensen, W. L. Robust Free Energy Perturbation Protocols for Creating Molecules in Solution. *Journal of Chemical Theory and Computation* **2019**, *15*, 3941–3948.

(12) Pan, A. C.; Borhani, D. W.; Dror, R. O.; Shaw, D. E. Molecular determinants of drug-receptor binding kinetics. *Drug Discovery Today* **2013**, *18*, 667–673.

(13) Tonge, P. J. Drug–Target Kinetics in Drug Discovery. *ACS Chemical Neuroscience* **2017**, acschemneuro.7b00185.

(14) Bai, F.; Jiang, H. Computationally Elucidating the Binding Kinetics for Different AChE Inhibitors to Access the Rationale for Improving the Drug Efficacy. *The Journal of Physical Chemistry B* **2022**, *126*, 7797–7805.

(15) Copeland, R. A. Conformational adaptation in drug–target interactions and residence time. *Future Medicinal Chemistry* **2011**, *3*, 1491–1501.

(16) Lee, K. S. S. et al. Optimized inhibitors of soluble epoxide hydrolase improve in vitro target residence time and in vivo efficacy. *Journal of Medicinal Chemistry* **2014**, *57*, 7016–7030.

(17) Lee, K. S. S.; Yang, J.; Niu, J.; Ng, C. J.; Wagner, K. M.; Dong, H.; Kodani, S. D.; Wan, D.; Morisseau, C.; Hammock, B. D. Drug-Target Residence Time Affects in Vivo Target Occupancy through Multiple Pathways. *ACS Central Science* **2019**, *5*, 1614–1624.

(18) Shaw, D. E. et al. Anton 3: Twenty Microseconds of Molecular Dynamics Simulation Before Lunch. **2021**,

(19) Dickson, A.; Lotz, S. D. Multiple Unbinding Pathways and Ligand-Induced Destabilization Revealed by WExplore. *Biophysical Journal* **2017**, *112*, 620–629.

(20) Nunes-Alves, A.; Zuckerman, D. M.; Arantes, G. M. Escape of a Small Molecule from Inside T4 Lysozyme by Multiple Pathways. *Biophysical Journal* **2018**, *114*, 1058–1066.

(21) Rydzewski, J.; Jakubowski, R.; Nowak, W.; Grubmüller, H. Kinetics of Huperzine A Dissociation from Acetylcholinesterase via Multiple Unbinding Pathways. *Journal of Chemical Theory and Computation* **2018**, *14*, 2843–2851.

(22) Rydzewski, J.; Valsson, O. Finding multiple reaction pathways of ligand unbinding. *Journal of Chemical Physics* **2019**, *150*.

(23) Hu, G.; Zhou, H. X. Binding free energy decomposition and multiple unbinding paths of buried ligands in a PreQ1riboswitch. *PLoS Computational Biology* **2021**, *17*.

(24) Tiwary, P.; Parrinello, M. From Metadynamics to Dynamics. *Physical Review Letters* **2013**, *111*, 230602.

(25) Kokh, D. B.; Doser, B.; Richter, S.; Ormersbach, F.; Cheng, X.; Wade, R. C. A workflow for exploring ligand dissociation from a macromolecule: Efficient random acceleration molecular dynamics simulation and interaction fingerprint analysis of ligand trajectories. *Journal of Chemical Physics* **2020**, *153*, 125102.

(26) Kokh, D. B.; Kaufmann, T.; Kister, B.; Wade, R. C. Machine learning analysis of tauRAMD trajectories to decipher molecular determinants of drug-target residence times. *Frontiers in Molecular Biosciences* **2019**, *6*.

(27) Ribeiro, J. M. L.; Provasi, D.; Filizola, M. A combination of machine learning and infrequent metadynamics to efficiently predict kinetic rates, transition states, and molecular

determinants of drug dissociation from G protein-coupled receptors. *Journal of Chemical Physics* **2020**, *153*.

(28) Badaoui, M.; Buigues, P. J.; Berta, D.; Mandana, G. M.; Gu, H.; Földes, T.; Dickson, C. J.; Hornak, V.; Kato, M.; Molteni, C.; Parsons, S.; Rosta, E. Combined Free-Energy Calculation and Machine Learning Methods for Understanding Ligand Unbinding Kinetics. *Journal of Chemical Theory and Computation* **2022**, *18*.

(29) Wolf, S. Predicting Protein–Ligand Binding and Unbinding Kinetics with Biased MD Simulations and Coarse-Graining of Dynamics: Current State and Challenges. *Journal of Chemical Information and Modeling* **2023**, *63*, 2902–2910.

(30) Huber, G.; Kim, S. Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophysical Journal* **1996**, *70*, 97–110.

(31) Chong, L. T.; Saglam, A. S.; Zuckerman, D. M. Path-sampling strategies for simulating rare events in biomolecular systems. *Current Opinion in Structural Biology* **2017**, *43*, 88–94.

(32) Adhikari, U.; Mostofian, B.; Copperman, J.; Subramanian, S. R.; Petersen, A. A.; Zuckerman, D. M. Computational Estimation of Microsecond to Second Atomistic Folding Times. *Journal of the American Chemical Society* **2019**, *141*, 6519–6526.

(33) Dickson, A.; III, C. L. B. WExplore: Hierarchical exploration of high-dimensional spaces using the weighted ensemble algorithm. *The Journal of Physical Chemistry B* **2014**, *118*, 3532–42.

(34) Donyapour, N.; Roussey, N. M.; Dickson, A. REVO: Resampling of ensembles by variation optimization. *Journal of Chemical Physics* **2019**, *150*, 244112.

(35) Lotz, S. D.; Dickson, A. Unbiased Molecular Dynamics of 11 min Timescale Drug

Unbinding Reveals Transition State Stabilizing Interactions. *Journal of the American Chemical Society* **2018**, *140*, 618–628.

(36) Dickson, A. Mapping the Ligand Binding Landscape. *Biophysical Journal* **2018**, *115*, 1707–1719.

(37) Prinz, J. H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schtte, C.; Noé, F. Markov models of molecular kinetics: Generation and validation. *Journal of Chemical Physics* **2011**, *134*.

(38) Chodera, J. D.; Noé, F. Markov state models of biomolecular conformational dynamics. *Current Opinion in Structural Biology* **2014**, *25*, 135–144.

(39) Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science. *Journal of the American Chemical Society* **2018**, *140*, 2386–2396.

(40) Copperman, J.; Zuckerman, D. M. Accelerated Estimation of Long-Timescale Kinetics from Weighted Ensemble Simulation via Non-Markovian "Microbin" Analysis. *Journal of Chemical Theory and Computation* **2020**, *16*, 6763–6775.

(41) Dickson, A.; Warmflash, A.; Dinner, A. R. Separating forward and backward pathways in nonequilibrium umbrella sampling. *Journal of Chemical Physics* **2009**, *131*, 154104.

(42) Zuckerman, D. M.; Chong, L. T. Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. *https://doi.org/10.1146/annurev-biophys-070816-033834* **2017**, *46*, 43–57.

(43) Sohraby, F.; Nunes-Alves, A. Advances in computational methods for ligand binding kinetics. *Trends in Biochemical Sciences* **2023**, *48*, 437–449.

(44) Newman, J. W.; Morisseau, C.; Harris, T. R.; Hammock, B. D. The soluble epoxide hydrolase encoded by EPXH2 is a bifunctional enzyme with novel lipid phosphate

phosphatase activity. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 1558–1563.

(45) Yu, Z.; Xu, F.; Huse, L. M.; Morisseau, C.; Draper, A. J.; Newman, J. W.; Parker, C.; Graham, L.; Engler, M. M.; Hammock, B. D.; Zeldin, D. C.; Kroetz, D. L. Soluble Epoxide Hydrolase Regulates Hydrolysis of Vasoactive Epoxyeicosatrienoic Acids. **2000**,

(46) Wagner, K.; Lee, K. S.; Yang, J.; Hammock, B. D. Epoxy fatty acids mediate analgesia in murine diabetic neuropathy. *European Journal of Pain* **2017**, *21*, 456–465.

(47) Mahapatra, A. D.; Choubey, R.; Datta, B. Small Molecule Soluble Epoxide Hydrolase Inhibitors in Multitarget and Combination Therapies for Inflammation and Cancer. *Molecules 2020, Vol. 25, Page 5488* **2020**, *25*, 5488.

(48) Hammock, B. D.; McReynolds, C. B.; Wagner, K.; Buckpitt, A.; Cortes-Puch, I.; Croston, G.; Lee, K. S. S.; Yang, J.; Schmidt, W. K.; Hwang, S. H. Movement to the Clinic of Soluble Epoxide Hydrolase Inhibitor EC5026 as an Analgesic for Neuropathic Pain and for Use as a Nonaddictive Opioid Alternative. *Journal of Medicinal Chemistry* **2021**, *64*, 1856–1872.

(49) Lee, K. S. S.; Ng, J. C.; Yang, J.; Hwang, S. H.; Morisseau, C.; Wagner, K.; Hammock, B. D. Preparation and evaluation of soluble epoxide hydrolase inhibitors with improved physical properties and potencies for treating diabetic neuropathic pain. *Bioorganic and Medicinal Chemistry* **2020**, *28*, 115735.

(50) Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM general force field (CGenFF) I: Bond perception and atom typing. *J. Chem. Info. Model.* **2012**, *52*, 3144–3154.

(51) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM

General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Info. Model.* **2012**, *52*, 3155–3168.

(52) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Computational Biology* **2017**, *13*, 1–17.

(53) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **2008**, *29*, 1859–1865.

(54) Zhang, B. W.; Jasnow, D.; Zuckerman, D. M. The "weighted ensemble" path sampling method is statistically exact for a broad class of stochastic processes and binning procedures. *Journal of Chemical Physics* **2010**, *132*, 1–7.

(55) Dixon, T.; Lotz, S. D.; Dickson, A. Predicting ligand binding affinity using on- and off-rates for the SAMPL6 SAMPLing challenge. *Journal of Computer-Aided Molecular Design* **2018**, *32*, 1001–1012.

(56) Hall, R.; Dixon, T.; Dickson, A. On Calculating Free Energy Differences Using Ensembles of Transition Paths. *Frontiers in Molecular Biosciences* **2020**, *7*, 106.

(57) Dixon, T. et al. Predicting the structural basis of targeted protein degradation by integrating molecular dynamics simulations with structural mass spectrometry. *Nature Communications* **2022**, *13*.

(58) Lotz, S. D.; Dickson, A. Wepy: A Flexible Software Framework for Simulating Rare Events with Weighted Ensemble Resampling. *ACS Omega* **2020**, *5*, 31608–31623.

(59) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; Fabritiis, G. D.; Noé, F. Identification of slow molecular order parameters for Markov model construction. *Journal of Chemical Physics* **2013**, *139*.

(60) Schwantes, C. R.; Pande, V. S. Improvements in Markov State Model construction reveal many non-native interactions in the folding of NTL9. *Journal of Chemical Theory and Computation* **2013**, *9*, 2000–2009.

(61) Hoffmann, M.; Scherer, M.; Hempel, T.; Mardt, A.; de Silva, B.; Husic, B. E.; Klus, S.; Wu, H.; Kutz, N.; Brunton, S. L.; Noé, F. Deeptime: a Python library for machine learning dynamical models from time series data. *Machine Learning: Science and Technology* **2022**, *3*, 015009.

(62) Dickson, A.; Lotz, S. CSNAnalysis. 2020; `https://github.com/ADicksonLab/CSNAnalysis`.

(63) Metzner, P.; Schütte, C.; Vanden-Eijnden, E. Transition Path Theory for Markov Jump Processes. *Multiscale Modeling & Simulation* **2009**, *7*, 1192–1219.

(64) Naritomi, Y.; Fuchigami, S. Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: The case of domain motions. *Journal of Chemical Physics* **2011**, *134*, 65101.

(65) Schwantes, C. R.; Pande, V. S. Modeling Molecular Kinetics with tICA and the Kernel Trick. *Journal of Chemical Theory and Computation* **2015**, *11*, 600–608.

(66) Bhakat, S. Collective variable discovery in the age of machine learning: reality, hype and everything in between. *RSC Advances* **2022**, *12*, 25010–25024.

(67) Mardt, A.; Pasquali, L.; Wu, H.; Noé, F. VAMPnets for deep learning of molecular kinetics. *Nature Communications* **2018**, *9*, 1–11.

(68) Ribeiro, J. M. L.; Bravo, P.; Wang, Y.; Tiwary, P. Reweighted autoencoded variational Bayes for enhanced sampling (RAVE). *Journal of Chemical Physics* **2018**, *149*.