Biophysical Journal
Article

Biophysical Society

Escape of a Small Molecule from Inside T4 Lysozyme by Multiple Pathways

Ariane Nunes-Alves,[1] Daniel M. Zuckerman,[2,*] and Guilherme Menegon Arantes[1,*]
[1]Department of Biochemistry, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil and [2]Department of Biomedical Engineering, School of Medicine, Oregon Health & Science University, Portland, Oregon

ABSTRACT The T4 lysozyme L99A mutant is often used as a model system to study small-molecule binding to proteins, but pathways for ligand entry and exit from the buried binding site and the associated protein conformational changes have not been fully resolved. Here, molecular dynamics simulations were employed to model benzene exit from its binding cavity using the weighted ensemble (WE) approach to enhance sampling of low-probability unbinding trajectories. Independent WE simulations revealed four pathways for benzene exit, which correspond to transient tunnels spontaneously formed in previous simulations of *apo* T4 lysozyme. Thus, benzene unbinding occurs through multiple pathways partially created by intrinsic protein structural fluctuations. Motions of several α-helices and side chains were involved in ligand escape from metastable microstates. WE simulations also provided preliminary estimates of rate constants for each exit pathway. These results complement previous works and provide a semiquantitative characterization of pathway heterogeneity for binding of small molecules to proteins.

INTRODUCTION

The binding of small molecules to proteins, a fundamental process in cellular metabolism, is also exploited pharmacologically to treat a large number of diseases. Drug development strategies usually try to improve ligand affinity to a target protein based on the structure and interactions observed in bound ligand-protein complexes. But binding kinetics and ligand residence time may also determine the physiological response and efficacy of a drug (1–4).

Kinetic, affinity, and structural information on protein-ligand association may be obtained from the ensemble of transition pathways (5–8) for the binding process. Each pathway describes the evolution of conformational and orientational degrees of freedom of ligand and protein that leads from an unbound to a bound configuration. As expected from path-ensemble symmetry (9), the reverse bound-to-unbound process should be described by the same pathway ensemble.

Computer simulations have been of great help in revealing possible pathways and kinetic information of protein-ligand binding (10–34). For example, Kubas and co-workers (29) simulated pathways for molecular O_2 binding to a hydrogenase and proposed mutations along the binding protein tunnels that slowed down the access of small molecules, thus reducing the hydrogenase inhibition caused by O_2. Casasnovas and co-workers (30) simulated dissociation of a potent pyrazol inhibitor from p38 MAP kinase and proposed congeneric ligands that could spend longer residence times when bound to the same kinase.

However, sampling of pathway ensembles for ligand-protein (un)binding is difficult because these are rare events on the timescales usually reached in molecular dynamics (MD) simulations. Methods such as weighted ensemble (WE) (35,36), milestoning (37,38), and transition interface sampling (39) are valuable because they enhance the sampling of pathways by increasing computational time spent in conformational regions with low visiting probability without introducing bias into the simulated potential energy (40,41). The WE method has been applied to study pathways and kinetics for protein conformational transitions (42–44), host-guest association (45), protein-peptide association (46), and protein-ligand dissociation (4,24,33).

One of the proteins most often used to study association with small molecules is the T4 lysozyme (T4L) (47–49), the structure of which is shown in Fig. 1. The T4L L99A mutation creates a hydrophobic cavity in the protein C-terminal domain with a volume of ~150 Å^3. This site may accommodate gases such as xenon or O_2 and small molecules such as benzene and its nonpolar derivatives (50–53). The engineered cavity is dehydrated and shows little

Submitted October 24, 2017, and accepted for publication January 8, 2018.
*Correspondence: zuckermd@ohsu.edu or garantes@iq.usp.br
Editor: Markus Buehler.
https://doi.org/10.1016/j.bpj.2018.01.014

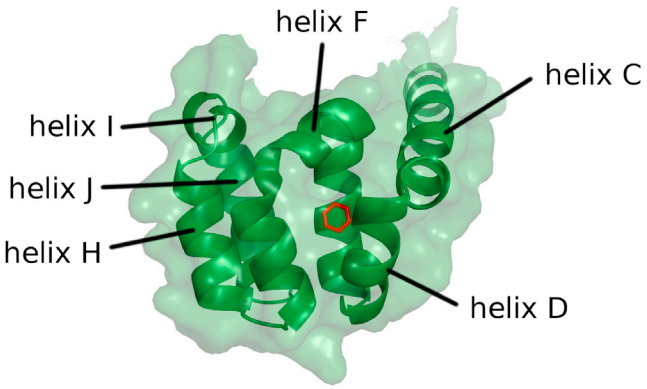


FIGURE 1 Crystal structure of the C-terminal domain of T4 lysozyme L99A mutant bound to benzene (*red*). The protein is represented with its molecular surface (*green transparency*), showing the ligand is fully buried. To see this figure in color, go online.

structural variation upon ligand complexation. Because of this simplicity and ease of experimental manipulation, more than 700 T4L complexes with small molecules have been characterized and used extensively to study structure-affinity relationships (49) and to compare with computational predictions of binding pose and affinity (54–57).

Crystal structures of T4L L99A mutant bound to benzene and a congeneric series of ligands (47,50) show that the binding cavity is buried, without any clear tunnel or channel connecting to the protein surface (Fig. 1). Thus, T4L must undergo structural fluctuations or conformational "breathing" to allow ligand excursion into the binding cavity. The kinetics of benzene entry and exit from T4L L99A mutant has been estimated from NMR spectroscopy ($k_{off} = 950\ s^{-1}$ for benzene at 303 K) (58), implying that the conformational motions involved in benzene escape should be observed in a millisecond timescale.

Several studies have addressed the possible pathways and the structural transitions used by small molecules to escape from inside T4L L99A. A weakly populated (or "excited") state involving motion of the flexible T4L helix F (Fig. 1) was observed from relaxation-dispersion NMR measurements and suggested to be responsible for benzene exit (59). Later structural determination of this weakly populated state showed that the binding cavity remains inaccessible (60), but a recent metadynamics simulation showed that the conformational transition from the crystal structure to the weakly populated state forms a transient protein tunnel along which benzene can exit from T4L (61). An accelerated MD simulation was able to identify a different route for benzene entry (62), and two recent conventional MD simulations identified three other pathways for O_2 unbinding (53) and five putative tunnels formed transiently in *apo* T4L L99A (63) that might accommodate small molecule excursions to the binding cavity.

To resolve the number and nature of ligand exit pathways and the conformational transitions involved in T4L L99A mutant, the WE approach is employed here as a discovery tool to sample trajectories for benzene exit. The ensemble of unbinding pathways obtained shows that benzene escapes via four different protein tunnels, all of which formed spontaneously in *apo* T4L L99A in a previous study that did not sample explicit ligand binding or unbinding (63). We find that motions in helix F, as previously suggested (49,58,59), are involved in ligand escape, but more importantly, so are those in helices C, H, and J, and side chains of several residues. Unbinding rates were estimated for each pathway and allowed a preliminary quantification of their contribution to the overall rate constant for benzene dissociation.

METHODS

Structural model and molecular dynamics simulations

The structure of T4L L99A mutant bound to benzene was obtained from the Protein Data Bank structure 1L83 (47) after removal of water and crystallization molecules. Hydrogen atoms were constructed using the GROMACS Protein Data Bank parser (64).

Interactions were described by the CHARMM36 force field (65–67). All simulations employed implicit solvation in the generalized Born surface area form (68). This should be a reasonable approximation, as the binding cavity in T4L L99A is dehydrated, although any phenomena involving specific water interactions may not be modeled accurately. The OBC method was used to estimate Born radii (69) with a dielectric constant $\epsilon = 80$, and the nonpolar contribution was calculated as in Schaefer et al. (70), with a surface tension of 5.4 cal mol^{-1} $Å^{-2}$ for all atoms.

Molecular dynamics trajectory segments in WE simulations were run using GROMACS 4.5 (64). Dynamics were carried out at 300 or at 400 K, with a 2 fs time-step, using a leapfrog stochastic dynamics integrator with a collision frequency of 10 ps^{-1}. Covalent bonds between hydrogens and heavy atoms were constrained with LINCS (71). Structures were saved every 2 ps for analysis. The total aggregate simulation time was ~29 μs.

WE simulations and rate estimation

The WE algorithm (35,36) was used to enhance sampling of benzene unbinding events. Briefly, in the WE method, a progress coordinate is defined and divided into bins to describe the process of interest. A group of trajectories of the simulated system is propagated from an initial state with initially equal probabilities or weights. At fixed time intervals of length τ, the occupancy of each coordinate bin is reevaluated. Trajectories may be replicated or merged with a proper weight attribution to keep a given number (M) of trajectories per bin as follows. If there are more than M trajectories in one bin, trajectories to be terminated will be selected according to their weight, and the weight of the terminated trajectories will be shared among the remaining ones in the bin. If a trajectory reaches an unvisited empty bin, it will be replicated in M trajectories, and each daughter trajectory will receive $1/M$ of the weight of the mother trajectory. Based on this unbiased allocation of trajectory weights, a properly weighted trajectory ensemble is generated (72). Maintaining M trajectories per bin evenly distributes computing resources in configuration space in contrast to an ordinary simulation. The number of trajectories per bin and the binning scheme to divide the progress coordinate do not change the results of simulations but affect the efficiency of the WE method in sampling rare events (72).

The WESTPA software package (73) was used to manage trajectory splitting and merging. The initial state was defined as benzene bound to T4L in the crystallographic position, and the target or final state was reached when

benzene had a solvent-accessible surface area (SASA) higher than 60% of its maximal SASA. This definition was enough to see benzene reaching the protein surface. But, as discussed below, this definition is not equivalent to complete unbinding as measured in experiments. Resampling intervals (τ) of 10 ps (for simulations run at 400 K) or 2 ps (for runs at 300 K) were used. The number of trajectories per bin varied from four to five according to the progress coordinate used. WE simulations were run for 4000 iterations, resulting in a maximum trajectory length of 8 ns. More details about the WE method are given in Table S2 (see Supporting Material).

WE simulations were run in a nonequilibrium steady-state scheme with trajectories being recycled back to the initial state once they reached the target state. This scheme allows estimation of transition rates from the initial to the target state (see below) (74). Estimated rate constants were expected to increase transiently in the beginning of a WE simulation after bin occupancy and later remain constant around the steady-state value after convergence was reached.

Time-windowed rate constants ($k_{\Delta t}$) were estimated from the nonequilibrium steady-state set of trajectories obtained from WE simulations (74):

$$k_{\Delta t}(t) = \frac{\Delta p}{\Delta t}, \quad (1)$$

where Δp is the sum of probabilities or weights of the trajectories that reached the target state in the time interval $t - \Delta t$ to t. Transition rates from the bound to the unbound state (dissociation rate constants) were calculated using a Δt of 2 ns. Errors for average rates were obtained from 90% confidence intervals from bootstrapping (75), calculated by resampling the rates obtained from six independent WE simulations.

Definition of progress coordinates for WE simulations

Two sets of WE simulations with different progress coordinates were carried out: an exploratory set to find possible routes for benzene unbinding and a production set to estimate rate constants and determine the protein conformational transitions involved in ligand escape for each unbinding route found.

In the exploratory simulations, two one-dimensional progress coordinates were used: the distance between benzene and binding site center-of-mass and the root mean-squared deviation (RMSD) between the current benzene position and the bound position found in the crystal structure. The binding site was defined by the Cα atoms of L84, V87, R95, A98, A99, V111, L118, N122, A129, and L133 in the artificial binding cavity. Only benzene carbon atoms were considered for RMSD calculations, and structures were previously aligned by superimposing the binding site Cαs just defined. The 12 possible symmetric images of benzene in each configuration were compared to the reference crystal structure, and the lowest RMSD value was adopted.

The exploratory progress coordinates were partitioned into small bins along coordinate regions where benzene movements were more restricted, as observed in initial simulations. Larger bins were employed along regions where benzene could diffuse more freely. Unbound states were defined by a distance or RMSD higher than 2.0 nm, in agreement with the SASA criteria described above. Simulations were run at high temperature (400 K) for 150 iterations, resulting in a maximum trajectory length of 1.5 ns. Bin boundaries and further details on the progress coordinates are given in Table S2.

Based on initial pathways found in the exploratory simulations, production simulations at 300 K were set up by using Voronoi bins (24,33) to study each of four pathways separately. In a Voronoi mapping, a set of centers is specified and all points in configurational space are attributed to bins for which center they are closest to according to a given distance criteria. Here, a Voronoi center was manually selected as the structure of a ligand-protein complex along unbinding trajectories obtained in the exploratory set of simulations. A total of 25, 23, 24, and 26 Voronoi centers were used to sample trajectories for the paths denoted below by the colors blue, orange, pink, and cyan, respectively.

In the production set of simulations, a three-dimensional progress coordinate was employed for a subset of Voronoi bins based on two atomic pair distances. The distances were chosen to delineate small protein conformational transitions observed in the exploratory simulations, as shown in Table S1 (see Supporting Material). To avoid a combinatorial increase in the total number of bins and WE trajectories simulated, the distance dimensions were included in a nested way to only two Voronoi bins for each of the four paths.

Conformational transitions involved in ligand escape from metastable or long-lived microstates along the unbinding pathways were obtained from analysis of the transition structure observed immediately after the Voronoi bin with highest lifetime in a trajectory was unoccupied. Further details for this analysis are described in the Supporting Material.

RESULTS

Identification of pathways for ligand escape

In the exploratory set of simulations run at 400 K, four exit pathways were found for benzene unbinding from T4L, as shown in Fig. 2 and denoted by color-coding. In the blue pathway, the ligand exits from the buried binding site by passing through helices C and F. Benzene mainly transits through helices C and D in the orange pathway, through helices F and I in the pink pathway, and through helices H and J in the cyan pathway.

Table 1 shows that four pathways were found when RMSD was used as the progress coordinate, whereas three of these paths were visited when the distance from the binding site was used as the progress coordinate. The four pathways found

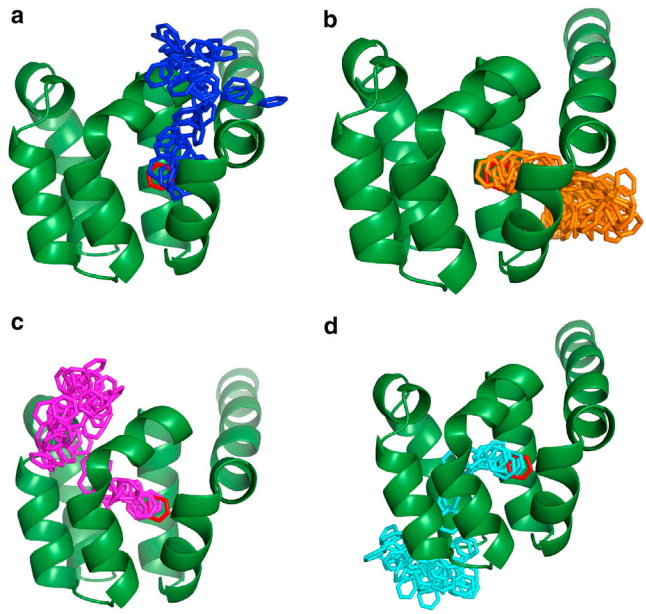


FIGURE 2 Four pathways were found for benzene exit from the buried binding site in T4 lysozyme. Blue (*a*), orange (*b*), pink (*c*), and cyan (*d*) colors show benzene positions sampled during one WE simulation. Each color represents a different exit pathway. The benzene position in the crystal structure is shown in *red sticks*. Only the T4L C-terminal domain is shown, but the complete protein was used in all simulations. To see this figure in color, go online.

TABLE 1 Sampling of Exit Pathways by WE Simulations Performed with Distance to Cavity Center-of-Mass, Ligand RMSD, and Voronoi Bins Progress Coordinates in Different Temperatures

	400 K		300 K					
			V_{blue}		V_{orange}			
Pathway	Dist	RMSD	Before[a]	After	Before	After	V_{pink}	V_{cyan}
Blue	4	2	6	6	6	0	0	0
Orange	3	2	6	0	6	6	0	0
Pink	2	5	0	0	0	0	6	0
Cyan	0	2	0	0	0	0	0	6
Total	4	5	6		6		6	6

The total refers to the total amount of WE simulations. Dist, distance to the center-of-mass; V, Voronoi bin.

[a]Before and after post-production reassignment step, as explained in the main text.

at 400 K could also be sampled at simulations performed at 300 K using a Voronoi mapping of the unbinding process.

The Voronoi bin map used as progress coordinate in the production set of simulations allowed sampling of trajectories for a specific pathway, thus facilitating the estimation of rate constants and identification of protein conformational transitions involved in ligand unbinding. Six independent WE simulations were carried out for each set of Voronoi bins defined (Table 1), leading to a total of 11,237, 16,062, 15,777, and 12,642 unbinding events through the blue, orange, pink, and cyan pathways, respectively. These events are correlated within a single WE run because of the WE procedure, and hence uncertainties are estimated based on independent WE runs.

Table 1 shows that WE simulations with a progress coordinate defined from Voronoi bins for the blue pathway also sampled unbinding events for the orange pathway and vice versa. This overlap for two different Voronoi mappings is due to a similar route followed by the ligand in the first half of the orange and blue pathways. Unbinding trajectories were reassigned to the correct pathway (blue or orange) before further analysis by adding two new Voronoi centers to the blue pathway (total of 27 centers) and one new center to the orange pathway (total of 24 centers), as detailed in the Supporting Material. Thus, the correct estimation of rate constants and detection of conformational transitions was possible for each pathway.

Protein structural transitions in long-lived microstates

Considering that small-molecule ligands such as benzene are clearly buried inside a hydrophobic cavity in T4L L99A holo crystal structures (47,52) (Fig. 1), conformational transitions or some "breathing" of the protein structure are expected to allow ligand escape to the protein surface.

Transition structures involved in ligand escape from metastable or long-lived microstates along the unbinding pathways were analyzed. Table 2 shows fractions of unbinding trajectories with transition structures that displayed side chain motions involved in ligand unbinding. Different reference structures were used to show the calculated fractions have a small dependency on the reference. Motions in Y88 and I78 side chains are the most often observed for the blue and orange pathways. These side chains move away from the binding site, allowing the ligand to depart from the buried cavity. Motions of W126, R154, and V111 side chains in the cyan pathway are related to motions in helices H, J, and F, respectively. Motions of the other side chains listed on Table 2 allow the ligand to exit the bound state and reach the protein surface. Fig. 3 shows some of the side chains involved in exit pathways.

Backbone fluctuation is also involved in benzene unbinding. Significant displacements in helix C in transition structures were observed in fractions of 0.35 and 0.65 of unbinding trajectories via blue and orange pathways, respectively. This helix moves away from the binding site and facilitates ligand exit from the buried cavity. The orange pathway also involved displacements of helix D in one-fifth of the transition structures. One-third of trajectories in the cyan pathway presented significant fluctuations in the backbone of helices H and J, increasing the distance between these helices and allowing benzene passage. On the other hand, only 15 and 12% of trajectories for the blue and pink pathways, respectively, presented transition structures with displacements in helix F. Fluctuations along this helix F were previously suggested as essential displacements for small molecule unbinding from T4L (58,60,61).

Preliminary rate constants for each unbinding pathway

Preliminary estimates for dissociation rate constants were calculated as a function of simulation time for each exit

TABLE 2 Fraction of Trajectories in WE Simulations Showing Residue Side Chain Motion Associated with Benzene Transit from the Microstate with Longest Lifetime along an Unbinding Event

Blue			Orange		
Side Chain	100 ps	10 ps	Side Chain	100 ps	10 ps
I78	0.78	0.67	Y88	0.55	0.69
Y88	0.52	0.48	I78	0.57	0.49
Pink			**Cyan**		
Side Chain	100 ps	10 ps	Side Chain	100 ps	10 ps
F114	0.43	0.50	M102	0.36	0.45
M102	0.42	0.25	F153	0.59	0.35
M106	0.35	0.25	R154	0.29	0.31
			W126	0.26	0.23
			V111	0.26	0.22

Reference structures were collected either 100 or 10 ps before the transition structure. Data obtained from six WE simulations for each pathway.

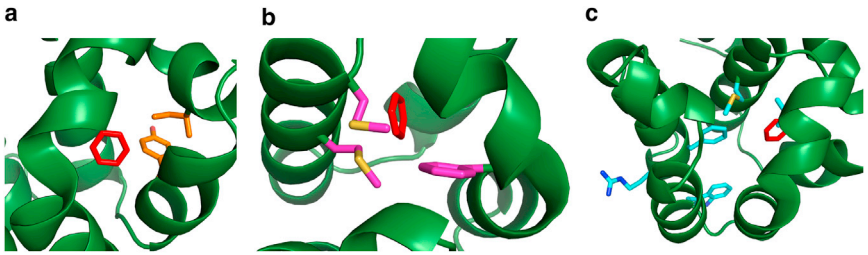


FIGURE 3 Residue side chains involved in benzene unbinding. (*a*) Y88 and I78 (*orange sticks*) contribute to the blue and orange pathways. (*b*) F114, M102, and M106 (*pink*) contribute to the pink pathway. (*c*) M102, F153, R154, W126, and V111 (*cyan*) contribute to the cyan pathway. The benzene position in the crystal structure is shown in *red sticks*. The structure was rotated in each panel to allow a better view. To see this figure in color, go online.

pathway to quantify their relative contributions to the overall ligand escape rate constant, as shown in Fig. 4. However, the relative contributions of each pathway cannot be clearly distinguished, as their rank order changes over simulation time. Note that the probability fluxes, upon which the rates are estimated, are expected to increase with simulation time until they reach steady values. Additionally, these estimates are dominated by only one or two individual WE simulations (Fig. S1). Thus, further sampling appears to be required to enable confident distinctions among the pathways.

The experimental rate constant for benzene dissociation (k_{off}) from T4L L99A mutant has been determined (58) by lineshape analysis of protein NMR as k_{off} =950 s^{-1}. The experimental value is close to the rate calculated for the pink and cyan pathways (at 8 ns in Fig. 4) but lower than the rates estimated for the blue and orange pathways (Fig. 4). Rates calculated here from WE simulations are expected to be higher than the experimental value because a less strict definition of unbound state was used (60% of the ligand maximum SASA; see WE Simulations and Rate Estimation). The calculated rates should be corrected by the committor probability for full ligand unbinding (equivalent to 100% ligand SASA) from the unbound state defined here (6).

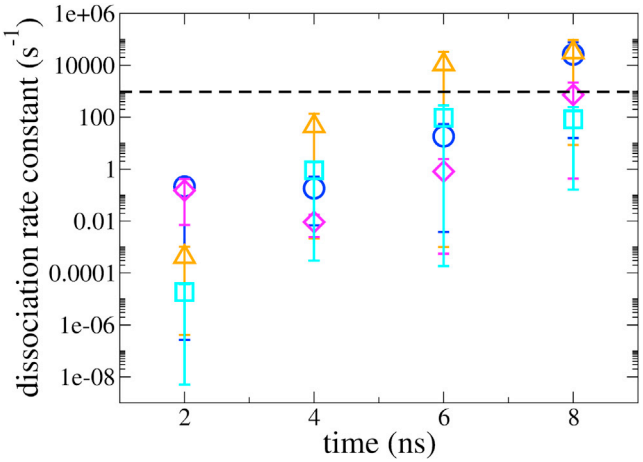


FIGURE 4 Average rate constant estimated for each pathway of benzene unbinding from T4L. Symbols: blue pathway *circle*; orange pathway *triangle*; pink pathway *diamond*; cyan pathway *square*. The bars indicate a 90% confidence interval from bootstrapping. The dotted line indicates the experimental rate constant. Rates were averaged every 2 ns for 6 WE simulations. To see this figure in color, go online.

DISCUSSION

Exit pathways and comparison to previous T4L studies

Four pathways were found here for benzene exit from the binding cavity buried in T4L L99A mutant. Each escape route was characterized by a preliminary rate constant and a different set of protein structural transitions.

Interestingly, a previous long timescale (30 μs) conventional MD simulation of *apo* T4L L99A, without sampling explicit ligand entry or exit, identified five protein tunnels that could be used by small molecules to access the buried cavity (63). Four of these tunnels, named D/F/G (tunnel through helices D, F, and G), C/D, F/G/H, and H/J, correspond respectively to the blue, orange, pink, and cyan pathways found here. An additional D/G tunnel was also identified, but the authors suggested that this D/G and the H/J tunnels would be too narrow to accommodate the passage of benzene (63). The WE simulations shown here were not able to find benzene escape through tunnel D/G but identified benzene transit via tunnel H/J (cyan pathway) with a slower rate and indeed a narrower tunnel than escape through the other pathways.

Other previous simulations also studied small-molecule (un)binding to T4L (53,61,62). Three routes were found for molecular O_2 dissociation in conventional MD simulations (53). One route matches the cyan pathway (H/J tunnel) found here for benzene transit. The other two O_2 routes correspond to the tunnel D/G mentioned above and to another tunnel found between helices D, E, G, H, and J. Both these two last tunnels were not observed here for benzene transit. Two MD simulations combined with enhanced sampling methods separately identified benzene transit from the binding site through the blue (62) and pink (61) pathways, with similar protein structural transitions to the ones observed here. To our knowledge, ligand excursion through the orange pathway was observed for the first time here.

The two previous simulations showing benzene (un)binding to T4L were able to find only one ligand transit pathway each (61,62), probably because of the enhanced sampling procedure employed. The biasing boost potential used by Miao et al. (62) may not efficiently increase sampling of the heterogeneous processes exhibited by the full set of pathways found in our study. Moreover, sampling the binding (rather than unbinding) paths may be intrinsically more

difficult under the conditions used. Wang et al. used metadynamics (76) with a path progress coordinate designed to sample the protein transitions involved in displacements of T4L residue F114 and helix F (60,61). These transitions are associated with benzene unbinding via the pink pathway, which was thus the only ligand exit pathway they observed (61).

Taken together, these results show that small molecules such as benzene unbind from T4L L99A through multiple and kinetically competitive pathways that correspond to ligand passage through transient tunnels spontaneously opened by intrinsic fluctuations of the protein structure observed in the absence of any ligand. This is reminiscent of a conformational selection mechanism (77,78) but for protein tunnels permitting small molecule transit. Our results also show that there is a degree of conformational diversity even within each of the four observed pathways.

It should be noted the results shown here are the first to our knowledge to find together and consistently the four pathways for small molecule exit from T4L by explicitly collecting exit trajectories instead of observing protein structural fluctuations that may lead to putative tunnels. The exploratory set of simulations presented here was carried out in 2016 without any knowledge of the pathways explored in the pink, orange, and cyan pathways. At that time, previous works (53,61,63) were not yet published or known to us. We were only aware of the blue path (62), although no information about it was used during our exploratory simulations. Remarkably, model composition (solvation, for instance) and force field descriptions differ significantly between our and previous studies (53,61–63). Nevertheless, a consistent set of ligand exit tunnels and pathways was found.

Pathways for ligand entry and exit have also been investigated in other proteins containing binding sites with variable surface exposure. Multiple protein tunnels are usually available for small gaseous molecules to diffuse into binding sites. For instance, hydrogenases may react with molecular O_2, and three entry routes for gaseous molecules have been characterized (12,14,29). On the other hand, a dominant route usually emerges for proteins with exposed binding sites and bulky ligands. This is the case for unbinding of a drug-like inhibitor from p38 MAP kinase (30). Binding to T4L lies somewhere in between and is similar to cytochrome P450, which binds organic molecules larger than gases but smaller than bulky drug-like ligands in a buried cavity via three or four different tunnels (10,18,19,23,28,31).

Structural analysis of exit pathways

The analysis of transition structures presented in Protein Structural Transitions in Long-Lived Microstates will only find protein movements related to benzene escape from metastable microstates (or long-lived Voronoi bins). Thus, conformational transitions involved in short-lived microstates have not been analyzed. It is noticeable, however, that no specific conformational gate such as a single side chain or helix displacement is clearly dominant for each pathway. Instead, benzene exit depends on a combination of intrinsic protein motions, and these motions exhibited variability even within each (single color) pathway. Nonetheless, some side chains such as in residues Y88 and I78 (blue and orange pathways) or F114 (pink pathway) were often found to be involved in benzene escape from long-lived microstates. Thus, point mutations that exchange these residues for a less bulky side chain such as alanine could speed up benzene exit by allowing faster transit through the respective protein tunnels and dissociation pathways.

Crystallographic (47,48,52) and NMR (59,60) data has shown that helix F is the most flexible element near the artificial cavity in T4L mutants. Thus, it has been suggested (49,58,79) that displacements along this helix are involved in ligand unbinding. The WE simulations suggest that, although observed, movements on helix F are not determinant or ubiquitous among unbinding pathways. Instead, transitions in helices C, H, and J appear to be more important for ligand transit. These motions might lead to conformational states which are too weakly populated to be detected experimentally.

Methodological limitations

Overall WE sampling led to detailed pathway information of a millisecond process based on a total aggregate simulation time of $\sim$30 μs. However, the large variance of rate estimates for each individual pathway (based on multiple WE runs) precluded discrimination among rates for different paths. This large variance may have been caused by high correlation among trajectories belonging to the same WE simulation, which is a consequence of the splitting and merging scheme of WE simulations to enhance sampling (36,41). Because of correlation among trajectories, slow coordinates orthogonal to the progress coordinate may not be fully sampled, leading to different rate estimates for one single path.

The choice of coordinates and bins is always a concern in WE simulation. Based on our results, it appears that initial use of ligand RMSD as a progress coordinate in combination with high-temperature MD simulation is an efficient and reliable procedure to explore ligand exit pathways from buried cavities in proteins. Ligand distance to the binding site center-of-mass lacks information on ligand orientation and proved a less reliable progress coordinate, as one pathway was not found when this metric was used. It also should be noted that high-temperature MD simulations may generate pathways that differ from those at normal temperatures (80). The pathways found here were ultimately sampled in production runs at 300 K and are in agreement with other MD simulations performed at normal temperature and with different potential energy models (53,61,62).

Although WE simulation is guaranteed to yield unbiased values of the fluxes (and hence rate constants) with sufficient sampling, the relatively short (8 ns) trajectory lengths

may prevent good sampling of certain types of unbinding events. All the WE unbinding trajectories are observed with a maximum length of 8 ns in our WE simulations, which is significantly shorter than the overall time that would be required in direct MD simulations. This overall time generally consists of a waiting period before a transition and the duration of the transition event itself (41). The WE approach enhances sampling in part by reducing the waiting time for the start of transitions, which is corrected by the weights. The event durations seen in WE, although short, may not be unreasonable in light of repeated observations of opening of protein tunnels during the 30 μs MD simulation of *apo* T4L (63).

CONCLUSIONS

Pathways for benzene exit from the buried binding cavity of T4L L99A mutant and the associated protein conformational changes were characterized here with the WE simulation method. For the first time to our knowledge, four separate ligand unbinding processes from T4L were observed in a single study. The four pathways found here are also in agreement with ligand transit routes observed in previous holo T4L simulations (53,61,62). Conformational transitions in several side chains, most notably Y88, I78, and F114, as well as displacements in helices C, D, F, H, and J, are involved in benzene unbinding. Our study also provided preliminary estimates of the dissociation rates for the different pathways, although further sampling is required to narrow statistical uncertainties.

This study used the WE method in two stages: an exploratory or discovery mode using naive progress coordinates to find possible pathways, followed by a production stage with more focused sampling. Combining the WE method with an RMSD progress coordinate and high temperature MD simulations appears to be a plausible approach to find multiple (un)binding pathways for small molecules. Production simulations using Voronoi bins may then be obtained to better characterize the dissociation rates and the conformational transitions involved in ligand transit.

The four protein pathways found for benzene exit may also be used for larger binders such as meta-xylene and N-butylbenzene (49). Similarly, it is also worth investigating if hydrophobic molecules such as 1,3,5-trimethylbenzene, which could still fit inside the binding cavity, experimentally do not bind T4L L99A mutant (49) because of a very slow entry kinetics (57). Such bulkier molecules may require wider protein tunnels to access the binding cavity, which WE simulation could help to rule out if extremely low rates were found.

AUTHOR CONTRIBUTIONS

A.N.-A. performed research and wrote the manuscript. D.M.Z. designed research, analyzed data, and revised the manuscript. G.M.A. designed research, analyzed data, and wrote the manuscript.

SUPPORTING MATERIAL

Supporting Materials and Methods, three figures, two tables, and one data file are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(18)30140-1.

ACKNOWLEDGMENTS

A.N.-A. acknowledges Lillian Chong, Adam Pratt, Rory Donovan, and Ernesto Suarez for discussions and technical help during a recent visit to the University of Pittsburgh, and Lillian Chong and Murilo Teixeira (Universidade De São Paulo) for suggestions on the manuscript.

Grants from São Paulo Research Foundation (FAPESP projects 2014/17008-7, 2014/21900-2, 2015/19912-5, and 2016/24096-5) and from the National Institutes of Health (GM115805) are gratefully acknowledged. Computing resources were in part provided by the University of Pittsburgh Center for Research Computing.

SUPPORTING CITATIONS

Reference (81) appears in the Supporting Material.

REFERENCES

1. Copeland, R. A., D. L. Pompliano, and T. D. Meek. 2006. Drug-target residence time and its implications for lead optimization. *Nat. Rev. Drug Discov.* 5:730–739.
2. Copeland, R. A. 2016. The drug-target residence time model: a 10-year retrospective. *Nat. Rev. Drug Discov.* 15:87–95.
3. Schuetz, D. A., W. E. A. de Witte, …, G. F. Ecker. 2017. Kinetics for drug discovery: an industry-driven effort to target drug residence time. *Drug Discov. Today.* 22:896–911.
4. Tang, Z., C. C. Roberts, and C. A. Chang. 2017. Understanding ligand-receptor non-covalent binding kinetics using molecular modeling. *Front. Biosci.* 22:960–981.
5. Dellago, C., P. G. Bolhuis, and D. Chandler. 1998. Efficient transition path sampling: application to Lennard-Jones cluster rearrangements. *J. Chem. Phys.* 108:9236–9245.
6. Bolhuis, P. G., D. Chandler, …, P. L. Geissler. 2002. Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* 53:291–318.
7. Singhal, N., C. D. Snow, and V. S. Pande. 2004. Using path sampling to build better Markovian state models: predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. *J. Chem. Phys.* 121:415–425.
8. Zhang, B. W., D. Jasnow, and D. M. Zuckerman. 2007. Efficient and verified simulation of a path ensemble for conformational change in a united-residue model of calmodulin. *Proc. Natl. Acad. Sci. USA.* 104:18043–18048.
9. Bhatt, D., and D. M. Zuckerman. 2011. Beyond microscopic reversibility: are observable non-equilibrium processes precisely reversible? *J. Chem. Theory Comput.* 7:2520–2527.
10. Winn, P. J., S. K. Lüdemann, …, R. C. Wade. 2002. Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine. *Proc. Natl. Acad. Sci. USA.* 99:5361–5366.
11. Martínez, L., M. T. Sonoda, …, I. Polikarpov. 2005. Molecular dynamics simulations reveal multiple pathways of ligand dissociation from thyroid hormone receptors. *Biophys. J.* 89:2011–2023.

12. Cohen, J., K. Kim, …, K. Schulten. 2005. Finding gas diffusion pathways in proteins: application to O2 and H2 transport in CpI [FeFe]-hydrogenase and the role of packing defects. *Structure.* 13:1321–1329.
13. Buch, I., T. Giorgino, and G. De Fabritiis. 2011. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA.* 108:10184–10189.
14. Wang, P. H., R. B. Best, and J. Blumberger. 2011. Multiscale simulation reveals multiple pathways for H2 and O2 transport in a [NiFe]-hydrogenase. *J. Am. Chem. Soc.* 133:3548–3556.
15. Shan, Y., E. T. Kim, …, D. E. Shaw. 2011. How does a drug molecule find its target binding site? *J. Am. Chem. Soc.* 133:9181–9183.
16. Huang, D., and A. Caflisch. 2011. The free energy landscape of small molecule unbinding. *PLoS Comput. Biol.* 7:e1002002.
17. Kang, M., C. Roberts, …, C. E. Chang. 2011. Gating and intermolecular interactions in ligand-protein association: coarse-grained modeling of HIV-1 protease. *J. Chem. Theory Comput.* 7:3438–3446.
18. Cojocaru, V., P. J. Winn, and R. C. Wade. 2012. Multiple, ligand-dependent routes from the active site of cytochrome P450 2C9. *Curr. Drug Metab.* 13:143–154.
19. Yu, X., V. Cojocaru, and R. C. Wade. 2013. Conformational diversity and ligand tunnels of mammalian cytochrome P450s. *Biotechnol. Appl. Biochem.* 60:134–145.
20. Bisha, I., A. Rodriguez, …, A. Magistrato. 2014. Metadynamics simulations reveal a Na+ independent exiting path of galactose for the inward-facing conformation of vSGLT. *PLoS Comput. Biol.* 10:e1004017.
21. Huang, Y. M., M. Kang, and C. E. Chang. 2014. Switches of hydrogen bonds during ligand-protein association processes determine binding kinetics. *J. Mol. Recognit.* 27:537–548.
22. Tiwary, P., V. Limongelli, …, M. Parrinello. 2015. Kinetics of protein-ligand unbinding: predicting pathways, rates, and rate-limiting steps. *Proc. Natl. Acad. Sci. USA.* 112:E386–E391.
23. Yu, X., P. Nandekar, …, R. C. Wade. 2016. Ligand tunnels in *T. brucei* and human CYP51: insights for parasite-specific drug design. *Biochim. Biophys. Acta.* 1860:67–78.
24. Dickson, A., and S. D. Lotz. 2016. Ligand release pathways obtained with WExplore: residence times and mechanisms. *J. Phys. Chem. B.* 120:5377–5385.
25. Teo, I., C. G. Mayne, …, T. Lelièvre. 2016. Adaptive multilevel splitting method for molecular dynamics calculation of benzamidine-trypsin dissociation time. *J. Chem. Theory Comput.* 12:2983–2989.
26. Paloncýová, M., V. Navrátilová, …, M. Otyepka. 2016. Role of enzyme flexibility in ligand access and egress to active site: bias-exchange metadynamics study of 1,3,7-trimethyluric acid in cytochrome P450 3A4. *J. Chem. Theory Comput.* 12:2101–2109.
27. Rydzewski, J., and W. Nowak. 2017. Ligand diffusion in proteins via enhanced sampling in molecular dynamics. *Phys. Life Rev.* 22–23:58–74.
28. Rydzewski, J., and W. Nowak. 2017. Thermodynamics of camphor migration in cytochrome P450cam by atomistic simulations. *Sci. Rep.* 7:7736.
29. Kubas, A., C. Orain, …, C. Léger. 2017. Mechanism of O2 diffusion and reduction in FeFe hydrogenases. *Nat. Chem.* 9:88–95.
30. Casasnovas, R., V. Limongelli, …, M. Parrinello. 2017. Unbinding kinetics of a p38 MAP kinase type II inhibitor from metadynamics simulations. *J. Am. Chem. Soc.* 139:4780–4788.
31. Magistrato, A., J. Sgrignani, …, A. Cavalli. 2017. Single or multiple access channels to the CYP450s active site? An answer from free energy simulations of the human aromatase enzyme. *J. Phys. Chem. Lett.* 8:2036–2042.
32. Tiwary, P., J. Mondal, and B. J. Berne. 2017. How and when does an anticancer drug leave its binding site? *Sci. Adv.* 3:e1700014.
33. Dickson, A., and S. D. Lotz. 2017. Multiple ligand unbinding pathways and ligand-induced destabilization revealed by WExplore. *Biophys. J.* 112:620–629.
34. Huang, Y. M., M. A. V. Raymundo, …, C. A. Chang. 2017. Mechanism of the association pathways for a pair of fast and slow binding ligands of HIV-1 protease. *Biochemistry.* 56:1311–1323.
35. Huber, G. A., and S. Kim. 1996. Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophys. J.* 70:97–110.
36. Zuckerman, D. M., and L. T. Chong. 2017. Weighted ensemble simulation: review of methodology, applications, and software. *Annu. Rev. Biophys.* 46:43–57.
37. Faradjian, A. K., and R. Elber. 2004. Computing time scales from reaction coordinates by milestoning. *J. Chem. Phys.* 120:10880–10889.
38. Votapka, L. W., and R. E. Amaro. 2015. Multiscale estimation of binding kinetics using Brownian dynamics, molecular dynamics and milestoning. *PLoS Comput. Biol.* 11:e1004381.
39. van Erp, T. S., D. Moroni, and P. G. Bolhuis. 2003. A novel path sampling method for the calculation of rate constants. *J. Chem. Phys.* 118:7762–7774.
40. Zuckerman, D. M. 2011. Equilibrium sampling in biomolecular simulations. *Annu. Rev. Biophys.* 40:41–62.
41. Chong, L. T., A. S. Saglam, and D. M. Zuckerman. 2017. Path-sampling strategies for simulating rare events in biomolecular systems. *Curr. Opin. Struct. Biol.* 43:88–94.
42. Bhatt, D., and D. M. Zuckerman. 2010. Heterogeneous path ensembles for conformational transitions in semi-atomistic models of adenylate kinase. *J. Chem. Theory Comput.* 6:3527–3539.
43. Adelman, J. L., A. L. Dale, …, M. Grabe. 2011. Simulations of the alternating access mechanism of the sodium symporter Mhp1. *Biophys. J.* 101:2399–2407.
44. Suárez, E., S. Lettieri, …, D. M. Zuckerman. 2014. Simultaneous computation of dynamical and equilibrium information using a weighted ensemble of trajectories. *J. Chem. Theory Comput.* 10:2658–2667.
45. Zwier, M. C., J. W. Kaus, and L. T. Chong. 2011. Efficient explicit–solvent molecular dynamics simulations of molecular association kinetics: methane/methane, Na(+)/Cl(-), methane/benzene, and K(+)/18–crown–6 ether. *J. Chem. Theory Comput.* 7:1189–1197.
46. Zwier, M. C., A. J. Pratt, …, L. T. Chong. 2016. Efficient atomistic simulation of pathways and calculation of rate constants for a protein-peptide binding process: application to the MDM2 protein and an intrinsically disordered p53 peptide. *J. Phys. Chem. Lett.* 7:3440–3445.
47. Eriksson, A. E., W. A. Baase, …, B. W. Matthews. 1992. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature.* 355:371–373.
48. Morton, A., and B. W. Matthews. 1995. Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: linkage of dynamics and structural plasticity. *Biochemistry.* 34:8576–8588.
49. Baase, W. A., L. Liu, …, B. W. Matthews. 2010. Lessons from the lysozyme of phage T4. *Protein Sci.* 19:631–641.
50. Morton, A., W. A. Baase, and B. W. Matthews. 1995. Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme. *Biochemistry.* 34:8564–8575.
51. Mulder, F. A. A., B. Hon, …, L. E. Kay. 2000. Flexibility and ligand exchange in a buried cavity mutant of T4 lysozyme studied by multinuclear NMR. *Biochemistry.* 39:12614–12622.
52. Merski, M., M. Fischer, …, B. K. Shoichet. 2015. Homologous ligands accommodated by discrete conformations of a buried cavity. *Proc. Natl. Acad. Sci. USA.* 112:5039–5044.
53. Kitahara, R., Y. Yoshimura, …, F. A. Mulder. 2016. Detecting O2 binding sites in protein cavities. *Sci. Rep.* 6:20534.
54. Wei, B. Q., W. A. Baase, …, B. K. Shoichet. 2002. A model binding site for testing scoring functions in molecular docking. *J. Mol. Biol.* 322:339–355.
55. Mobley, D. L., A. P. Graves, …, K. A. Dill. 2007. Predicting absolute ligand binding free energies to a simple model site. *J. Mol. Biol.* 371:1118–1134.

56. Wang, K., J. D. Chodera, …, M. R. Shirts. 2013. Identifying ligand binding sites and poses using GPU-accelerated Hamiltonian replica exchange molecular dynamics. *J. Comput. Aided Mol. Des.* 27:989–1007.

57. Nunes-Alves, A., and G. M. Arantes. 2014. Ligand-receptor affinities computed by an adapted linear interaction model for continuum electrostatics and by protein conformational averaging. *J. Chem. Inf. Model.* 54:2309–2319.

58. Feher, V. A., E. P. Baldwin, and F. W. Dahlquist. 1996. Access of ligands to cavities within the core of a protein is rapid. *Nat. Struct. Biol.* 3:516–521.

59. Mulder, F. A. A., A. Mittermaier, …, L. E. Kay. 2001. Studying excited states of proteins by NMR spectroscopy. *Nat. Struct. Biol.* 8:932–935.

60. Bouvignies, G., P. Vallurupalli, …, L. E. Kay. 2011. Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature*. 477:111–114.

61. Wang, Y., E. Papaleo, and K. Lindorff-Larsen. 2016. Mapping transiently formed and sparsely populated conformations on a complex energy landscape. *eLife*. 5:e17505.

62. Miao, Y., V. A. Feher, and J. A. McCammon. 2015. Gaussian accelerated molecular dynamics: unconstrained enhanced sampling and free energy calculation. *J. Chem. Theory Comput.* 11:3584–3595.

63. Schiffer, J. M., V. A. Feher, …, R. E. Amaro. 2016. Capturing invisible motions in the transition from ground to rare excited states of T4 lysozyme L99A. *Biophys. J.* 111:1631–1640.

64. Pronk, S., S. Páll, …, E. Lindahl. 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics.* 29:845–854.

65. MacKerell, A. D., D. Bashford, …, M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.

66. MacKerell, A. D., Jr., M. Feig, and C. L. Brooks, III. 2004. Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* 126:698–699.

67. Best, R. B., X. Zhu, …, A. D. Mackerell, Jr. 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles. *J. Chem. Theory Comput.* 8:3257–3273.

68. Still, W. C., A. Tempczyk, …, T. Hendrickson. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 112:6127–6129.

69. Onufriev, A., D. Bashford, and D. A. Case. 2004. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins.* 55:383–394.

70. Schaefer, M., C. Bartels, and M. Karplus. 1998. Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. *J. Mol. Biol.* 284:835–848.

71. Hess, B., H. Bekker, …, J. G. E. M. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.

72. Zhang, B. W., D. Jasnow, and D. M. Zuckerman. 2010. The "weighted ensemble" path sampling method is statistically exact for a broad class of stochastic processes and binning procedures. *J. Chem. Phys.* 132:054107.

73. Zwier, M. C., J. L. Adelman, …, L. T. Chong. 2015. WESTPA: an interoperable, highly scalable software package for weighted ensemble simulation and analysis. *J. Chem. Theory Comput.* 11:800–809.

74. Bhatt, D., B. W. Zhang, and D. M. Zuckerman. 2010. Steady-state simulations using weighted ensemble path sampling. *J. Chem. Phys.* 133:014110.

75. Efron, B., and R. J. Tibshirani. 1993. An Introduction to the Bootstrap, First Edition. Chapman & Hall/CRC, Boca Raton, Florida.

76. Laio, A., and M. Parrinello. 2002. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA.* 99:12562–12566.

77. Weikl, T. R., and F. Paul. 2014. Conformational selection in protein binding and function. *Protein Sci.* 23:1508–1518.

78. Gianni, S., J. Dogan, and P. Jemth. 2014. Distinguishing induced fit from conformational selection. *Biophys. Chem.* 189:33–39.

79. Vallurupalli, P., N. Chakrabarti, …, L. E. Kay. 2016. Atomistic picture of conformational exchange in a T4 lysozyme cavity mutant: an experiment-guided molecular dynamics study. *Chem. Sci.* 7:3602–3613.

80. Zhang, B. W.; Jasnow, D.; Zuckerman, D. M. 2009, Weighted ensemble path sampling for multiple reaction channels. arXiv, arXiv:0902.2772, https://arxiv.org/abs/0902.2772.

81. Roux, B. 1995. The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* 91:275–282.

Biophysical Journal, Volume *114*

Supplemental Information

Escape of a Small Molecule from Inside T4 Lysozyme by Multiple Pathways

Ariane Nunes-Alves, Daniel M. Zuckerman, and Guilherme Menegon Arantes

Escape of a small molecule from inside T4 lysozyme by multiple pathways

Ariane Nunes-Alves[1], Daniel M. Zuckerman[2]* and Guilherme Menegon Arantes[1]*

[1]*Department of Biochemistry, Instituto de Química, Universidade de São Paulo, Av. Prof. Lineu Prestes 748, 05508-900, São Paulo, SP, Brazil*

[2]*Department of Biomedical Engineering, School of Medicine, Oregon Health & Science University, 2730 SW Moody Avenue, 97239, Portland, OR, US*

**Corresponding authors: zuckermd@ohsu.edu, garantes@iq.usp.br*

Supporting Information

S1.1 Choice of exploratory temperature and WE parameters

High temperature exploratory simulations were performed at 400 K. This temperature was chosen intuitively to increase the rate of transitions over barriers and lead to a more even sampling of pathways with different probabilities for ligand escape. It is expected that the ideal temperature to explore unbinding pathways is system-dependent, and a systematic test of the dependence of the pathways found on different temperatures was not performed here.

A low resampling interval (τ) may increase the chances of detecting trajectories occupying unvisited bins, but it will also increase the total computational cost due to the more resources spent in trajectory analysis and bin attribution. Initial WE simulations with τ = 2, 5 and 10 ps showed that bin occupancy is reached in shorter simulation time for lower τ values. Reaching full bin occupancy in short simulation times leads to faster sampling of unbinding events. WE simulations with different τ values had similar computational cost (defined as the time to propagate the trajectories plus the time for trajectory analysis and bin attribution) to perform 10 ps of MD simulation time. Thus, τ = 2 ps was adopted for the production WE simulations here.

A high number of trajectories per bin (traj/bin, in Table S2) may result in a better occupancy of bins and also increase the chances of reaching unvisited bins, but it also increases the total number of trajectories and the computational cost of a WE simulation. Given that our simulations ran on computer nodes with 48 threads and we tried to keep one trajectory run-

ning per thread, a total of 144 trajectories (N_{tot} =3×48) was used for each resampling interval. Thus, the number of trajectories per bin was chosen as 4 (= 144/36 ≈ N_{tot}/bins) which limited the amount of correlated trajectories and the computational cost, and also allowed a reasonable bin occupancy.

The binning scheme to partition the progress coordinate was designed with the help of a rough potential of mean force (PMF) obtained from umbrella sampling simulations for benzene unbinding along the blue pathway. The binning scheme was chosen so that each bin had to transpose ∼ 0.6 kcal/mol steps of the PMF. Umbrella sampling was done with the GROMACS pull code (1), using the distance between ligand and binding site COM as a progress coordinate. Eleven (11) windows spaced at 0.2 nm bins were used, with reference distances ranging from 0.4 to 2.4 nm and a force constant of 1000 kJ/(mol nm^2). Each window run for 30 ns and the last 28.4 ns were used to generate the PMF with the weighted histogram analysis method (2).

S1.2 Description of Voronoi centers

Voronoi centers that describe each unbinding pathway were manually chosen from structures of ligand-protein complexes obtained from unbinding trajectories in the exploratory set of simulations. Each structure was aligned to the T4L L99A mutant crystal structure (PDB 1L83) using the Cα atoms of the binding site as a reference (section 2.3) and later employed in the WE simulations as Voronoi centers, to compute distances and define bins along the progress coordinate. One file for each pathway (blue, pink, cyan and orange) containing the coordinates of protein-ligand (T4L-benzene) complexes used for the definition of the Voronoi centers is included as Supporting Information.

S1.3 Reassignment of the unbinding trajectories to the correct pathways

Visual inspection of the trajectories showed that WE simulations with a progress coordinate defined from Voronoi bins for the blue pathway also sampled unbinding events for the orange pathway, and vice-versa. A post-production step was performed to properly separate unbinding events and reassign to the correct exit pathway. The unbinding event was reassigned to the pathway which contained the Voronoi center closest to the unbound state configuration. This

was possible by using 27 Voronoi centers (instead of 25 used in the WE production runs) to the blue pathway and 24 (instead of 23) centers to the orange pathway. Reassignment allowed proper separation of pathways and correct estimation of rate constants and conformational changes for each pathway.

S1.4 Identification of protein conformational transitions involved in ligand unbinding

Conformational changes allowing ligand exit from T4L were analyzed. Metastable microstates were associated to Voronoi bins with high lifetimes and the protein transitions involved in ligand progression along the unbinding pathway were found by:

- identification of the metastable microstate with the Voronoi bin with highest lifetime for each successful unbinding trajectory;
- collection of the structure in the frame immediately after the metastable state was unoccupied (transition structure);
- construction of a list of the side chains contacting benzene in the transition structure, using a distance cutoff of 0.5 nm;
- computation of the (heavy atom) RMSD between transition structure and reference structure for each side chain, after structure alignment using the binding site C_{alpha} defined in section 2.2;
- selection of 3 side chains with highest RMSD.

Reference structures were collected either 100 ps or 10 ps before the transition structure in the same trajectory.

Analysis revealed that movement of helices C, D, F, H and J could also be involved in ligand unbinding. This was quantified by:

- identification of the metastable microstate and collection of transition structure for each successful unbinding trajectory;

- check if one of the side chains of helices C (residues 69-81), D (residues 83-90) F (residues 107-114), H (residues 126-135) or J (residues 143-155) was contacting benzene in the transition structure, using a cutoff of 0.5 nm;
- check the helix displacement.

The criteria for helix displacement was based on average atom-pair distance distributions shown in figure S2. Helix C was considered displaced when the average distance was higher than 1.01 nm, helix D when higher than 1.02 nm, helix F when higher than 1.25 nm and helices H and J when higher than 0.67 nm.

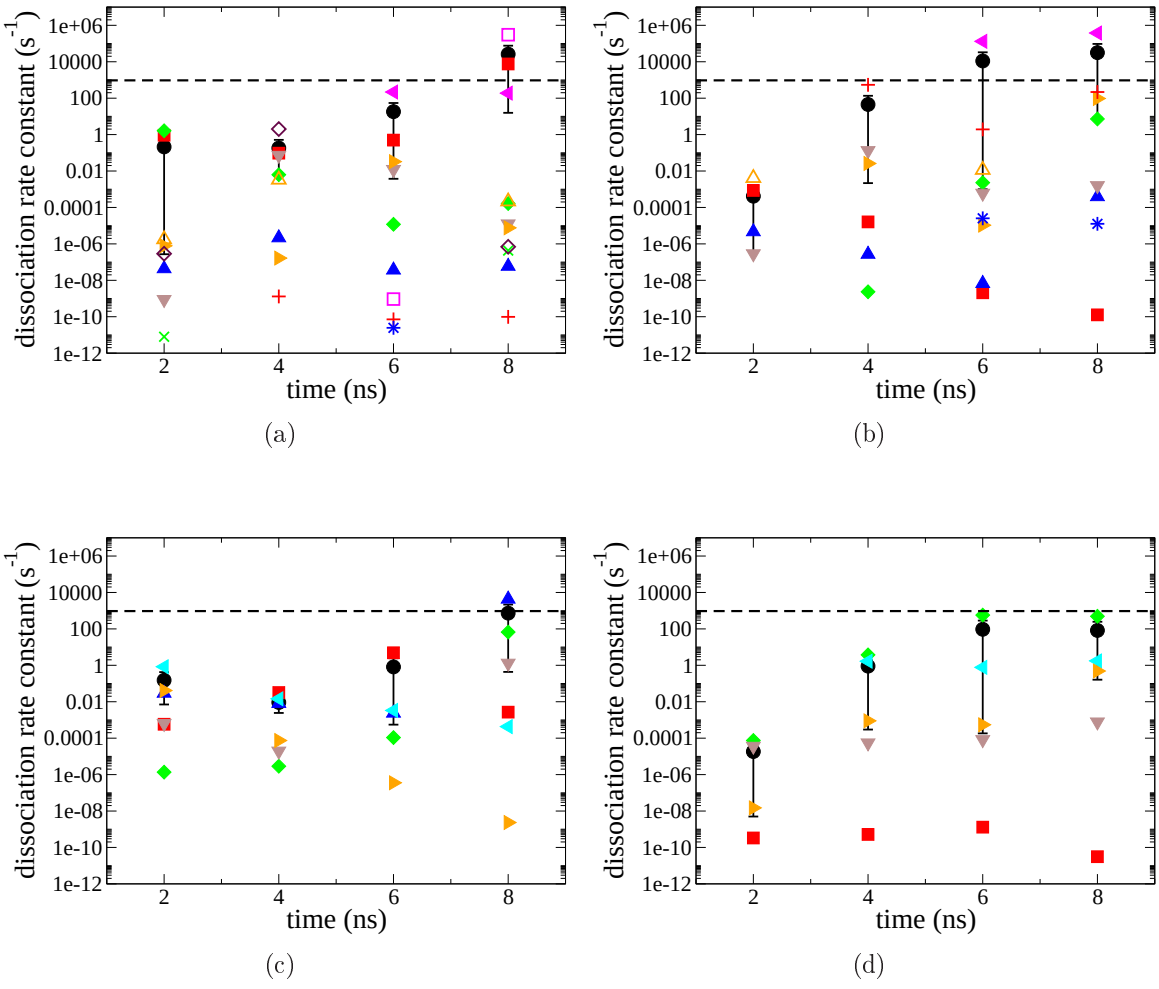


Figure S1: Average and individual rate constant estimated for each pathway of benzene unbinding from T4L. (a) Blue pathway, (b) orange pathway, (c) pink pathway and (d) cyan pathway. Black circles indicate averages, remaining symbols indicate individual estimates from one WE simulation. Bars indicate a 90% confidence interval from bootstrapping for the averages, which are dominated by larger values. Dotted lines indicate the experimental rate constant.

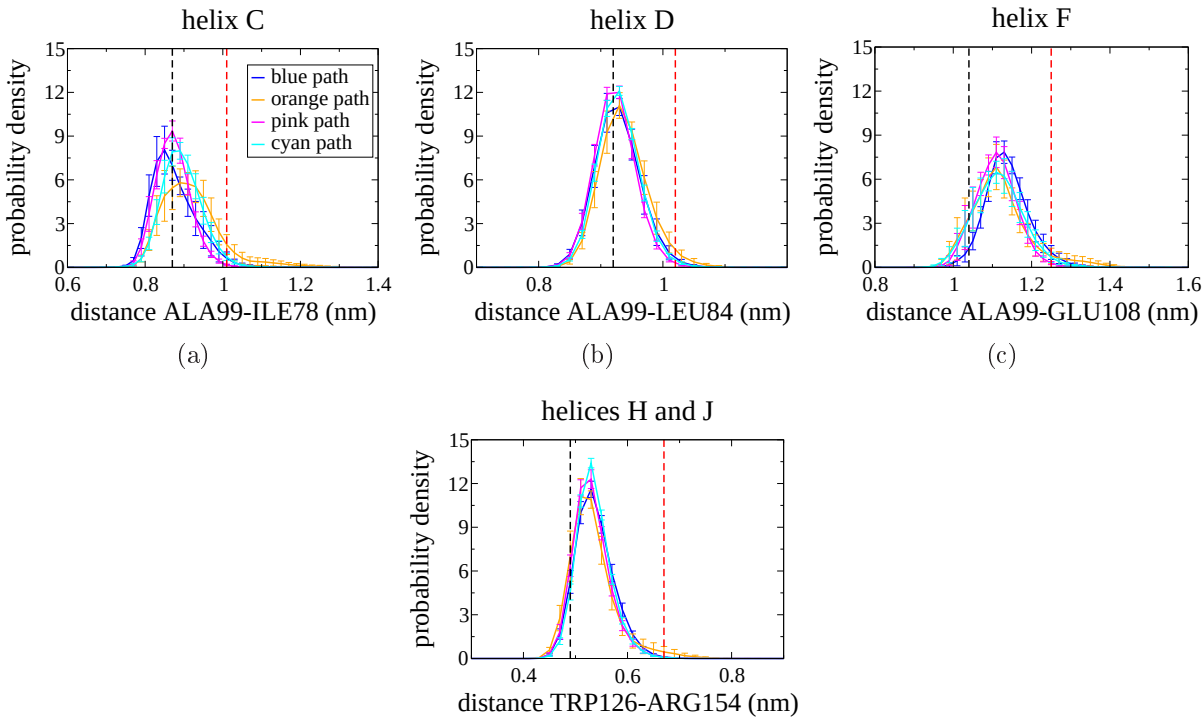
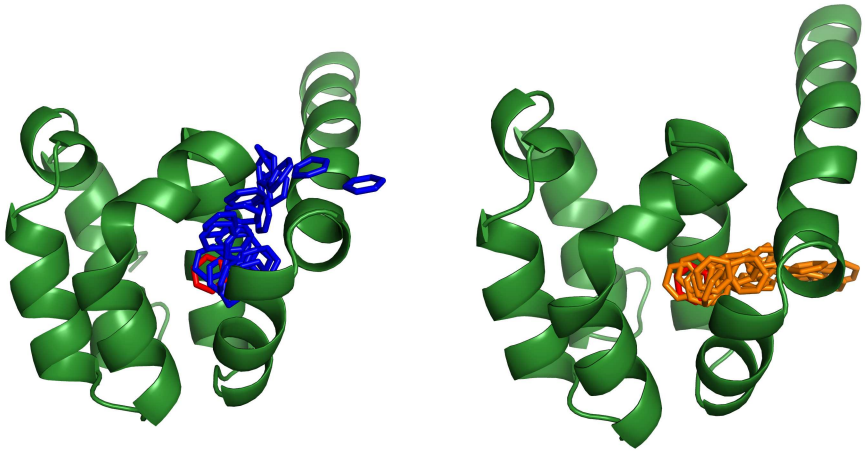
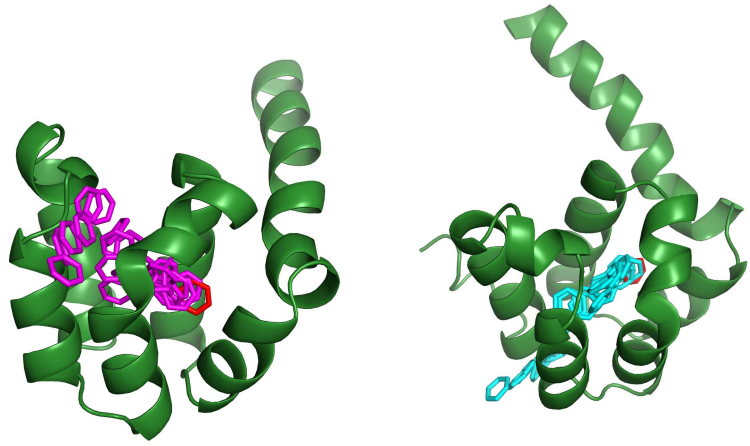


Figure S2: Average distance distributions from 4 to 8 ns of six WE simulations for each pathway (color coded as shown in the insert). Bars represent the standard error in the averages. Black and red lines represent the distance in the crystal structure and the distance criteria for helix displacement, respectively.



(a) (b)



(c) (d)

Figure S3: Benzene positions in the structures used as Voronoi centers to sample the blue (a), orange (b), pink (c) and cyan (d) pathways. Benzene position in the crystal structure is shown in red sticks and was also used as a Voronoi center in each path. Only the T4L C-terminal domain is shown.

Table S1: Definition of additional progress coordinates for benzene unbinding. Two atom-pair distances were selected for each pathway. The bin boundaries (in nm) and the Voronoi bins where the progress coordinate was added are also shown. Voronoi bins are numbered in order of increasing distance from the binding site.

pathway	**distance**	**bin boundaries**	**Voronoi bins**
blue	CZ:Y88 - CA:A99	0.70, 0.83, 0.95	1, 2
	CA:L84 - CA:A99	0.90, 0.97, 1.05	5, 6
orange	CZ:Y88 - CA:A99	0.70, 0.83, 0.95	1, 2
	CB:Y88 - CB:I78	0.50, 0.58, 0.65	7, 8
pink	CB:V111 - CA:A99	0.80, 0.85, 0.90	2, 3
	CE:M102 - CA:A146	0.70, 0.77, 0.85	7, 11
cyan	CG:L133 - CA:S117	0.62, 0.72, 0.80	8, 9
	CD:R154 - CA:W126	0.60, 0.70, 0.80	19, 20

Table S2: Parameters used to run WE simulations for different progress coordinates (PC). WE sim.: number of WE simulations run, traj/bin: number of trajectories per bin, τ: resampling frequency (in ps), iter: number of resampling steps of one WE simulation, max length: maximum length of one WE simulation (in ns, calculated as τ*iter), bins: amount of bins of the progress coordinate calculated as (N-M)+M*J when Voronoi bins are used as progress coordinate, where N is the total number of Voronoi bins and M=4 is the number of Voronoi bins which were partitioned by an additional progress coordinate composed of J=4 bins, as shown in table S1, aggreg. time: aggregate simulation time of the progress coordinate in ns, calculated as (WE sim.)*(traj/bin)*(max length)*bins.

PC	**WE sim.**	**traj/bin**	τ	**iter**	**max length**	**bins**	**aggreg. time**
dist	4	5	10	150	1.5	12	360
RMSD	5	5	10	150	1.5	17	637.5
V_{blue}	6	4	2	4000	8	37	7104
V_{orange}	6	4	2	4000	8	35	6720
V_{pink}	6	4	2	4000	8	36	6912
V_{cyan}	6	4	2	4000	8	38	7296

Bin boundaries for distance PC (in nm): 0.30, 0.35, 0.38, 0.41, 0.50, 0.55, 0.60, 0.65, 0.80, 1.60, 1.68. Boundaries for RMSD PC (nm): 0.15, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.70, 0.80, 0.90, 1.00, 1.10, 1.50, 1.90.

Supporting References

1. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. 2013, GROMACS 4.5: a high–throughput and highly parallel open source molecular simulation toolkit. Bioinformatics, 29, 845–854.

2. Roux, B. 1995, The calculation of the potential of mean force using computer simulations. Comp. Phys. Comm., 91, 275–282.