RESEARCH ARTICLE

Atomic resolution mechanism of ligand binding to a solvent inaccessible cavity in T4 lysozyme

Jagannath Mondal[1]*, Navjeet Ahalawat[1], Subhendu Pandit[1], Lewis E. Kay[2,3], Pramodh Vallurupalli[1]*

1 Tata Institute of Fundamental Research, Hyderabad, India, 2 Departments of Molecular Genetics, Biochemistry and Chemistry, University of Toronto, Toronto, Ontario, Canada, 3 Hospital for Sick Children Program in Molecular Medicine, Toronto, Ontario, Canada

* jmondal@tifrh.res.in (JM); pramodh@tifrh.res.in (PV)

Abstract

Ligand binding sites in proteins are often localized to deeply buried cavities, inaccessible to bulk solvent. Yet, in many cases binding of cognate ligands occurs rapidly. An intriguing system is presented by the L99A cavity mutant of T4 Lysozyme (T4L L99A) that rapidly binds benzene (~10^6 M^{-1}s^{-1}). Although the protein has long served as a model system for protein thermodynamics and crystal structures of both free and benzene-bound T4L L99A are available, the kinetic pathways by which benzene reaches its solvent-inaccessible binding cavity remain elusive. The current work, using extensive molecular dynamics simulation, achieves this by capturing the complete process of spontaneous recognition of benzene by T4L L99A at atomistic resolution. A series of multi-microsecond unbiased molecular dynamics simulation trajectories unequivocally reveal how benzene, starting in bulk solvent, diffuses to the protein and spontaneously reaches the solvent inaccessible cavity of T4L L99A. The simulated and high-resolution X-ray derived bound structures are in excellent agreement. A robust four-state Markov model, developed using cumulative 60 μs trajectories, identifies and quantifies *multiple* ligand binding pathways with low activation barriers. Interestingly, none of these identified binding pathways required large conformational changes for ligand access to the buried cavity. Rather, these involve transient but crucial opening of a channel to the cavity via subtle displacements in the positions of key helices (helix4/helix6, helix7/helix9) leading to rapid binding. Free energy simulations further elucidate that these channel-opening events would have been unfavorable in wild type T4L. Taken together and via integrating with results from experiments, these simulations provide unprecedented mechanistic insights into the complete ligand recognition process in a buried cavity. By illustrating the power of subtle helix movements in opening up *multiple* pathways for ligand access, this work offers an alternate view of ligand recognition in a solvent-inaccessible cavity, contrary to the common perception of a *single* dominant pathway for ligand binding.

Author summary

Proteins often bind ligands in buried cavities that appear to be inaccessible based on static structures. The mechanisms and pathways by which ligands reach their binding sites in such cases are, thus, often unknown. Yet, ligand recognition by occluded cavities can happen rapidly. A central question remains: How does such a process occur? Experiments that provide insight at atomic resolution are currently lacking. In the current work, we have used a computational approach to capture the process by which a ligand, benzene, binds to a buried cavity in the L99A cavity mutant of T4 Lysozyme. Using multiple long, unbiased atomistic simulations, we have discovered how benzene, starting from bulk solvent, finds and binds the solvent-inaccessible cavity. We find that there is no *single* dominant pathway. Rather, simulated trajectories discover *multiple* binding pathways with low activation barriers, facilitating a rapid recognition process. We highlight the role of subtle movements in helix positions in opening up multiple crucial paths for benzene to reach its binding cavity without the need for large-scale distortions of the protein structure, explaining the small activation energies.

Introduction

Measurements of affinities of small molecules to specific binding sites in proteins have become routine, especially in the context of drug discovery studies where such experiments are often among the first to be performed. These thermodynamic measurements are routinely supplemented by kinetic studies of drug-receptor interactions [1–3]. However, in contrast to these standard kinetic and thermodynamic measurements, experiments that provide atomic level insights into the kinetic pathways by which small cognate molecules find their binding sites and the roles that protein conformational dynamics play in this process are lacking. This is particularly the case for ligand binding sites that are deeply buried in the receptor protein core, precluding access, in the absence of structural rearrangements, even for bulk solvent molecules. Yet, many ligands recognize these deeply buried, solvent-inaccessible cavities very efficiently, and often at rates that approach those which are diffusion limited. Effective methodologies for the study of such binding processes in atomistic detail would have significant implications for the rational design of pharmaceuticals at a practical level and would provide important insights into molecular recognition and the role of protein dynamics in ligand binding, in general.

The current work provides an atomistic view of the complete kinetic processes by which a hydrophobic ligand binds rapidly to the occluded and solvent-inaccessible cavity of a well-known system, namely the L99A mutant of lysozyme from the T4 bacteriophage (T4L L99A). Substitution of Ala for Leu at position 99 results in a 150 Å^3 cavity [5] that can accommodate a range of ligands, including benzene, (Fig 1) [4–7]. Over the past several decades this protein has been used as a model system for understanding how buried cavities affect protein stability and structure, in general, and how ligands might navigate the protein landscape to bind rapidly to their proper sites. To this end, high-resolution X-ray structures of both the apo form of the protein and the benzene-bound complex have been solved by Mathews and coworkers [4,5] and pioneering studies have been undertaken that relate cavity size to protein stability [6]. NMR measurements have shown that the binding of benzene to T4L L99A occurs with an on rate constant of approximately $10^6 \text{ M}^{-1}\text{s}^{-1}$ [8]. While certainly slower than the diffusion-limited rate ($\sim 10^9 \text{ M}^{-1}\text{s}^{-1}$), binding is more rapid than what is generally perceived for ligand-recognition in solvent-inaccessible buried protein cavities. Additional NMR studies, based on
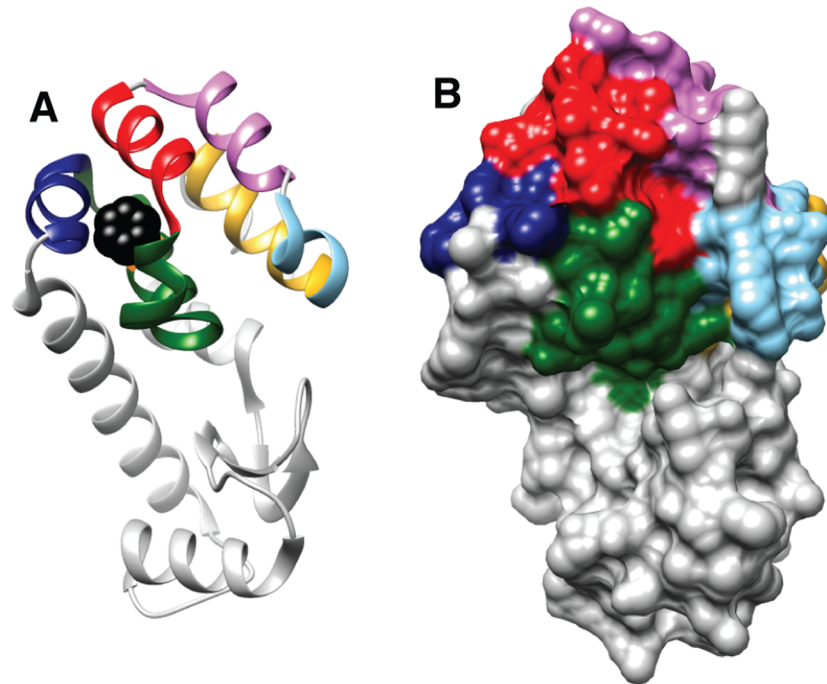
Fig 1. Benzene binds to a buried hydrophobic cavity in T4L L99A that is inaccessible from the surface. A) Ribbon representation of T4L L99A bound to benzene (PDB 3DMX) [4]. Bound benzene is shown in black CPK representation, helix 4 (residues 83 to 91) in navy blue, helix 5 (92–113) in dark green, helix 6 (114–123) in red, helix 7 (125–134) in orchid, helix 8 (136–142) in light blue and helix 9 (142–156) in gold. B) The benzene molecule is not visible in the surface representation of T4L L99A because the binding site is buried in the protein with no pathway to the surface in the crystal structure.

relaxation dispersion approaches [9], established that T4L L99A exchanges in solution between a major conformer similar to the crystal structure of the protein and a minor state that is populated to 3% with a millisecond lifetime at room temperature [10,11]. However, the structure of this rare state shows that the cavity is occluded by the aromatic side-chain of Phe114 that prevents the binding of benzene so that the dynamic process that has been characterized is not relevant to ligand binding [11,12]. Oxygen and xenon binding studies have also been performed that have helped to identify hydrophobic cavities in the protein [13–15]. Further, this system has served as a prototype for many binding free energy calculations [16,17].

Despite the availability of macroscopic kinetic rate constants for aromatic ligand binding, high resolution structures of the bound and free forms of T4L L99A, and NMR spin relaxation data that indicate conformational flexibility in the region surrounding the ligand binding site, a detailed description of how ligands bind to the protein remains elusive. Recent innovations in experimental and computational techniques, notwithstanding, the mechanism by which ligands bind to buried sites in a protein has remained difficult to address, because the process often involves the formation of transient metastable states that are challenging to characterize at high resolution and because the conformational fluctuations leading to binding can be rapid, complicating detailed characterization by most biophysical techniques. To this end, molecular dynamics simulations are emerging as a complementary tool to experiment for the study of molecular recognition processes because of recent innovations in GPU-based technologies, access to distributed computing, and the development of special-purpose computers [18–21], that have facilitated studies of many pertinent biochemical processes at atomic resolution.

In this work, we elucidate the mechanism by which benzene binds to the solvent-inaccessible cavity of T4L L99A by capturing the entire binding process using unbiased molecular dynamics (MD) simulations. Despite the fact that the simulations contained no *a priori* knowledge of the T4L L99A binding site, the resulting bound conformations that were obtained in a series of microsecond-long simulations and independent trajectories matched that of the crystal structure. Underlying Markov state model analyses [22–24] of our simulation trajectories established multiple binding pathways, each with transient opening of a channel involving a distinct pair of helices in the C-terminal domain of T4L L99A. Using an enhanced sampling technique, infrequently biased metadynamics simulations [25,26], the ligand unbinding pathways from the bound state have also been established. The resulting estimates for ligand binding on- and off-rates as well as binding free energies are in reasonable agreement with the experimentally reported values, lending support to the conclusions of this study. Our work establishes that the fast binding of benzene to T4L L99A derives from its access to multiple pathways that are formed from low barrier concerted protein fluctuations that only occur in the mutated protein. It also establishes the utility of molecular dynamics simulations, in general, in providing detailed descriptions of binding processes that are subsequently validated by comparison of the resulting calculated kinetic and thermodynamic parameters with those obtained via experiment.

Results

Long unbiased simulations place benzene at the target-binding site of L99A T4L with atomic precision

We performed six independent all-atom unbiased MD simulations with lengths varying between 2 and 8 μs, resulting in a total simulation time of 29 μs. In each of the simulations a benzene molecule, starting from a random position in solution, was correctly placed in the target-binding site, corresponding to the hydrophobic and solvent-inaccessible cavity created by the L99A mutation in T4L. Ligands were initially positioned at least 4 nm away from the binding pocket in random orientations. As shown in three representative trajectories (Fig 2 and S1–S3 Movies), the ligand diffused extensively in the solvent, occasionally contacting different parts of the protein surface, before entering the binding pocket. Fig 2A quantifies the time-evolution of pocket-ligand distances en route to binding in three representative binding trajectories, where the distance between the binding cavity and benzene gradually decreases as the ligand finds its target. As depicted in Fig 2B, the MD derived structure converges to within 1–2 Å of the X-ray based model for the holo-form of the protein [4], as quantified by the root-mean-squared deviation (RMSD) of heavy atoms from residues defining the cavity and including the ligand (detailed in Materials and methods). The excellent agreement between the simulated and crystallographic binding regions, including the orientation of benzene (Fig 2C), provides confidence in the underlying force field used to model the T4L L99A benzene binding process. Unlike standard docking approaches which search for the best ligand orientation within a predefined binding site, our long and unguided atomistic MD simulations reproducibly identified and maintained the correct ligand-bound pose without user intervention or incorporation of any prior knowledge of the binding site. As shown in Fig A in S1 Supporting Information in each of the trajectories substantial dehydration accompanies benzene entry into the hydrophobic and solvent-inaccessible binding pocket.

Simulation trajectories reveal multiple binding pathways

One of the key findings of the current work is that there are multiple distinct pathways by which benzene binds the solvent-inaccessible cavity in T4L L99A. Contrary to the belief of a
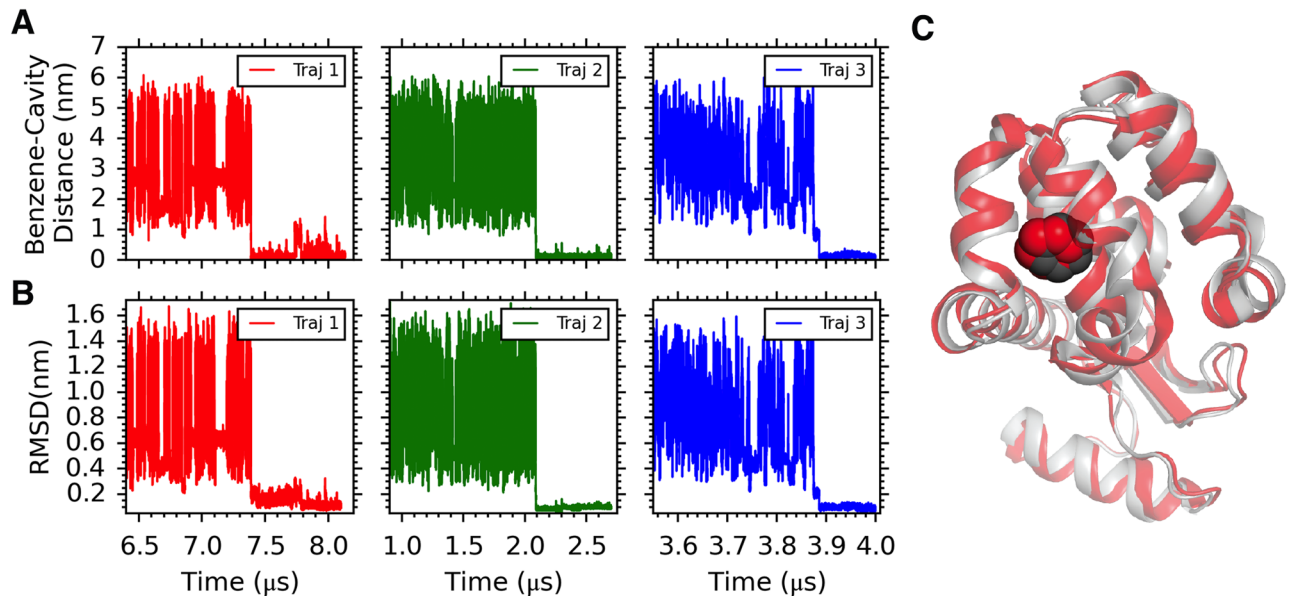
Fig 2. Benzene binds correctly to the L99A cavity in long unbiased MD simulations. A) Distance between the solvent-inaccessible cavity of T4L L99A and benzene shows that benzene diffuses through the solvent before finally binding. Protein heavy atoms within 5 Å of benzene in the holo-protein (PDB id: 3DMX) [4] are defined as the 'binding pocket' or 'cavity'. Radial distances were computed between the respective centers of mass of the pocket and the ligand. B) Time profile of the root-mean-squared deviation (RMSD) between the MD derived structure and the crystal structure of the benzene-bound form of the protein. Heavy atoms from cavity residues plus ligand of the MD derived structure are compared with corresponding positions in the crystal structure of the benzene-bound form of the protein (PDB id: 3DMX [4]) to calculate the RMSD values. Initial RMSDs are high because benzene is not in the binding site and not because the protein is distorted. C) Overlay of a snapshot of the benzene-T4L L99A complex from simulations with the crystal structure (PDB id: 3DMX). The crystal structure is shown in grey with the benzene in black CPK representation while the protein in the MD snapshot is shown in red with the benzene in red CPK representation.

https://doi.org/10.1371/journal.pcbi.1006180.g002

single dominant pathway in protein-ligand binding events, as exemplified by Shaw and coworkers in studies of GPCRs and kinases [18,19], our post simulation analyses of the six successful binding trajectories specifically identified three distinct pathways, each involving opening of channels between pairs of helices near the C-terminus of the protein. Fig 3A and S1–S3 Movies illustrate the three representative binding pathways via the time evolution of a benzene molecule as it reaches the binding pocket. The three pathways involve crucial conformational fluctuations of the protein and specifically, creation of crevices across certain inter-helical interfaces in the C-terminal domain. Notably, these helices were different in each of the three identified pathways as illustrated in Fig 3. For example, in trajectory 1, helices 4 and 6 temporarily move away from one another, creating a pathway to the binding site as seen in the snapshot at 7.388 μs where Ala 99 colored in orange is visible from the surface (Fig 3B). This is in contrast to the snapshot at 7.375 μs where the distance between the two helices is considerably shorter, prohibiting the entry of benzene. In trajectory 2 (Fig 3C) benzene goes through a path created between helices 7 and 9. Benzene first associates with the surface as seen in the snapshot at 2.08875 μs. The subsequent transient displacement between helices 7 and 9 creates a tunnel through which benzene enters, as shown in the snapshot at 2.090 μs. After residing at an alternate site for ~2 ns the ligand reaches the final binding site (Fig 3C). In trajectory 3, benzene enters through an opening created at the junction between helices 5, 6, 7 and 8 (compare snapshots at 3.874 and 3.876 μs, Fig 3D). After a period of approximately 10 ns the ligand localizes to the correct binding site, as illustrated in Fig 3D.

In combination, these results show that key conformational fluctuations of T4L L99A lead to the formation of transient conformers where the distance between sets of helices becomes
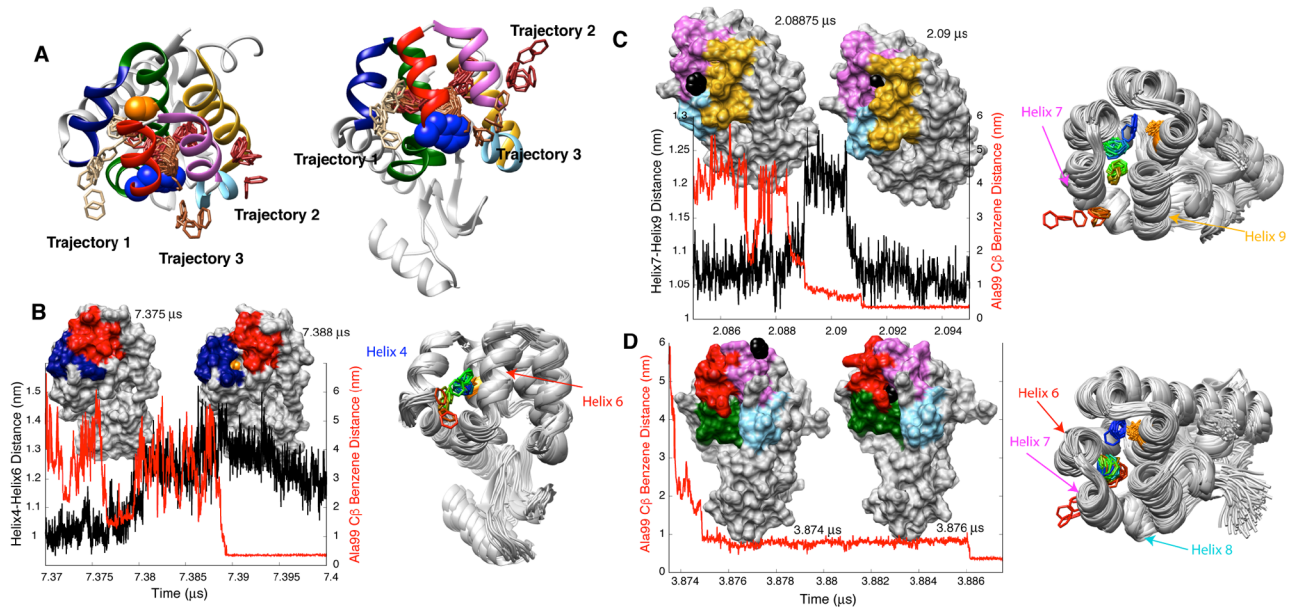
Fig 3. Illustration of the different pathways by which benzene reaches the binding pocket. A) Paths from three highlighted trajectories are illustrated by superimposing the backbone atoms of the protein from snapshots during the trajectories onto the crystal structure of the protein. In trajectory 1 benzene enters between helices 4 and 6, between helices 7 and 9 in trajectory 2 and through the junction region of helices 5, 6, 7 and 8 in trajectory 3. For clarity only the crystal structure of the protein is shown in A, with Ala 99 and Phe 114 in orange and blue CPK representations, respectively. All helices are colored as in Fig 1. B) In trajectory 1 helices 4 and 6 transiently move away from one another, temporarily creating a pathway resulting in Ala 99 (colored in orange) becoming visible from the surface in the snapshot at 7.388 μs. The separation between these helices returns to the starting value of ~1 nm later in the trajectory (see Fig C in S1 Supporting Information). C) In trajectory 2 benzene (shown in black) enters through a pathway created between helices 7 and 9 (2.09 μs snapshot) after associating with the protein surface (2.08875 μs snapshot). D) In trajectory 3 benzene enters through an opening created at the helix 5, 6, 7 and 8 junction region. On the right side of panels B, C, D benzene entry is illustrated as in A, except that the protein is shown as a grey ribbon in each snapshot and the color of benzene changes from red to green to blue as time progresses and benzene reaches the binding site. Ala 99 is shown in orange in the structures.

sufficiently large for benzene to reach the solvent-inaccessible cavity. As depicted in Fig B in S1 Supporting Information, the spatial density profile of benzene around the protein, obtained by combining the long trajectories, provides a cumulative picture of the different pathways that lead to binding. The localization of benzene near the cavity and near the C-terminal domain helices is evident from the time-averaged density profile. However, we also observe certain locations near the N terminus where the ligand resides for significant periods of time (Fig B in S1 Supporting Information). Whether these highly visited locations are ultimately important for ligand binding is not clear.

Markov state model identifies and quantifies transitions between key ligand-binding intermediates and pathways

We have used Markov State Model (MSM)-based analysis[24] of the MD trajectories to obtain mechanistic insights into the ligand binding process and to calculate binding on- and off-rates and thermodynamic parameters. Prior to building a meaningful Markov state model, an additional three hundred, 100 ns trajectories were obtained by initiating independent simulations from different intermediates in the long trajectories. Analysis of the cumulative 59 μs of simulated data (29 μs from long trajectories and another 30 μs from 300 short trajectories, see Materials and methods) using the MSM approach yielded estimates of the thermodynamic and kinetic parameters for the binding process which are in reasonable agreement with experiment (Table 1). The calculated kinetic on and off rate constants (k_{on} and k_{off}) obtained by MSM-

Table 1. Thermodynamic and kinetic parameters estimated from MD simulations are in reasonable agreement with experimentally measured values. The thermodynamic and kinetic parameters were calculated from the four state MSM based on the mean first passage time (MFPT), $k_{on} = 1/(MFPT_{on}C)$, $k_{off} = 1/MFPT_{off}$ where C is the benzene concentration, 9.55 mM. The standard binding free energies were calculated based on the stationary populations (shown in Table 2) of bound and unbound macrostates as obtained from the MSM [20]. Unless specified, the experimental values refer to those from the NMR study by Dahlquist and coworkers [8] and are used for all comparisons.

Methods	k_{on} (M^{-1}s^{-1}) x10^6	k_{off} (s^{-1})	$\Delta G^0_{binding}$ (kcal/mol)
Experimental [7,8]	0.8–1	950 ± 20	-5.2±0.2[7] / -4.2 ± 0.1[8]
Unbiased MD (MSM analysis)	21 ± 9	310 ± 130	−6.9 ± 0.8
Metadynamics		270 ± 100	

derived mean first passage time (*MFPT*) values (see caption of Table 1 and Material and methods for equations), are, respectively, k_{on} = 21 ± 9 x 10^6 M^{-1} s^{-1} (*MFPT$_{on}$* = 5163 ns) and k_{off} = 311 ± 130 s^{-1} (*MFPT$_{off}$* = 3.2 x 10^6 ns). As compared in Table 1, the MSM-derived k_{on} is larger than that measured experimentally, k^{exp}_{on} = 0.8–1 x 10^6 M^{-1} s^{-1}. We also note that the k_{on} value calculated from the average binding times from six long unbiased multi-microsecond MD trajectories yields an on-rate constant of 25 x 10^6 M^{-1} s^{-1}, very close to the MSM-derived k_{on} rate (21 ± 9 x 10^6 M^{-1} s^{-1}). This implies that the underlying deviation of the simulated k_{on} value from the experimental k^{exp}_{on} rate does not result from MSM analysis but is due to limitations in the force field. For example, the TIP3P [27] water model used in our MD simulations leads to water diffusion rates that are two to three times higher than those based on experiment. The ligand unbinding rate constant, i.e. the so-called off-rate constant k_{off}, computed using both MSM (k_{off} = 310 ± 130 s^{-1}) and Metadynamics approaches (k_{off} = 270 ± 100 s^{-1}) (detailed later) is in reasonable agreement with experiment (k^{exp}_{off} = 950 s^{-1}). In this context, we note that there are several reports highlighting systematic deviations of MSM-derived off-rate constants from those determined experimentally [20,21]. However, it should also be emphasized that the MD estimate of k_{off} is less precise than for k_{on} because spontaneous unbinding events were not observed. Finally, we have also computed the standard binding free energy ($\Delta G^0_{binding}$) of benzene to T4L L99A using the MSM-derived stationary population of the unbound and bound conformations (see Table 2). The computed standard binding free energy of -6.9 ± 0.8 kcal/mol predicts slightly higher binding affinity than the experimentally measured $\Delta G^0_{binding}$ = -5.2 ± 0.1 kcal/mol.

A simple four state MSM was constructed from our simulated data, as described in Materials and Methods. The four macrostates are illustrated in Fig 4, where MS0 and MS3 are identified with the unbound solvated benzene state and the final bound conformation, respectively, and MS1 and MS2 are two intermediates, with benzene localized near different entry points. Populations and lifetimes of each of the states, along with the committor values that measure the progress of the binding process, are summarized in Table 2. As shown in Table 2, the stationary population of the bound state, MS3, is highest and as described above, the binding free energy of -6.9 ± 0.8 kcal/mol, derived from the stationary state populations, is in reasonable agreement with the experimental measurement. The location of the benzene ligand is distinct

Table 2. Properties of the four different MSM macrostates, with MS0 and MS3 corresponding to the unbound and bound states, respectively.

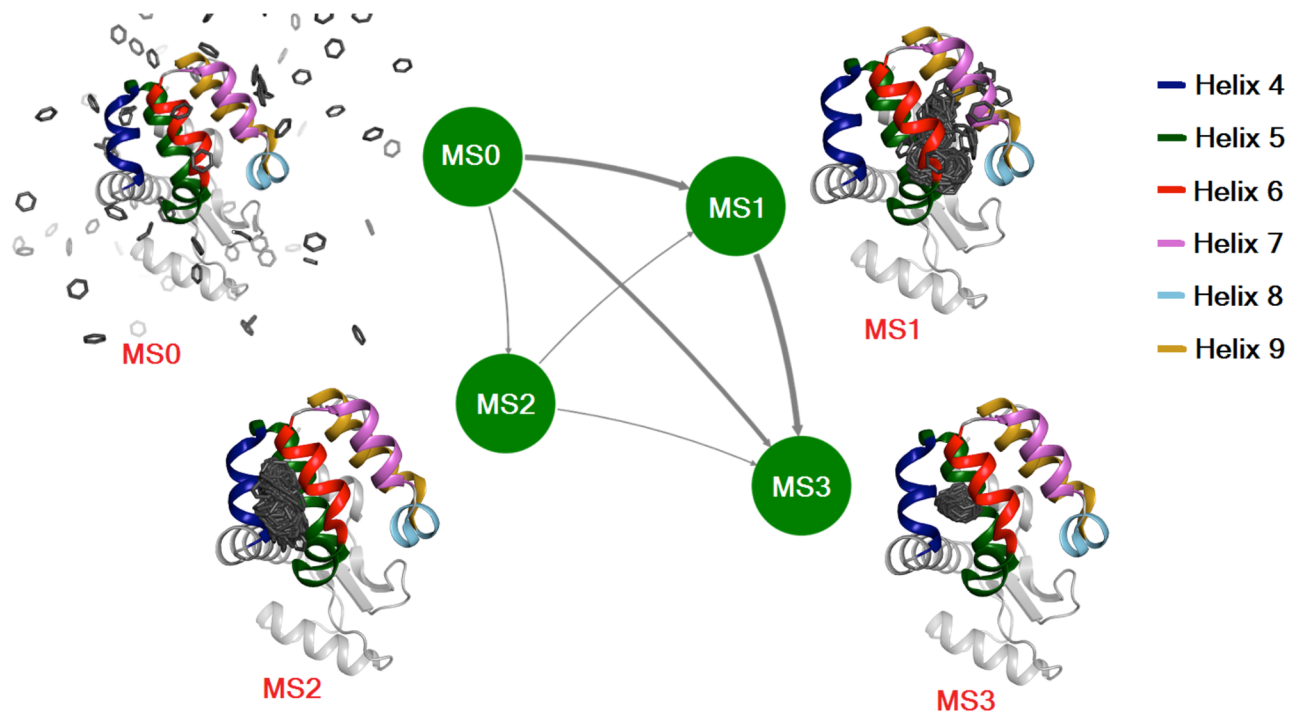| State	Population %	Lifetime (ns)	Committor Probability
MS0	0.11±0.06	2.6±4.4 x10^3	0.0
MS1	6.4±3.0 x 10^{-3}	69±8	0.8±0.1
MS2	4.7±4.5 x 10^{-2}	742±289	0.2±0.1
MS3	99.83±0.05	9±4 x10^5	1

Fig 4. MSM analysis shows that the binding process proceeds via distinct intermediates. The MD binding data can be described by a four state (MS0-MS3) MSM network representation showing how binding proceeds via different paths through distinct intermediate states, from the free form (MS0) to the bound form (MS3). Each node represents a macrostate, with the thickness of the arrows denoting the net flux between states. 500 snapshots for each of the states are superimposed on the crystal structure of the holo enzyme (PDB 3DMX) [4], where only benzene molecules from the 500 snapshots are rendered for clarity. The population, lifetime and committor value of each of the states is shown in Table 2. Coloring is as in Fig 1 and the color-code of the helices is provided in the panel.

in the intermediate structures, MS1 and MS2, highlighting again the different pathways of entry (see above). Notably, benzene is bound to different positions of the protein in the MS1 state, potentially interconverting rapidly between the different sites. These include locations near a pair of entry points that are comprised of helices 7 and 9 or the junction between helices 5, 6, 7 and 8 that place the ligand close to the cavity. The positions of benzene in MS1 are, thus, as found in MD trajectories 2 and 3 (Fig 3C and 3D), so that MS1 is a crucial intermediate. In contrast, benzene is further from the cavity in MS2, positioned instead on the surface of helices 4 and 6. MS2 is thus an intermediate of trajectory 1 that is elucidated from MD simulations (Fig 3B). The computed committor probabilities from transition-path-theory based analysis suggest a higher (0.8) commitment of MS1 towards the bound state than that of MS2 (0.2) (Table 2). The conversion from unbound (MS0) to bound (MS3) macrostates proceeds via multiple pathways involving states MS1 and MS2. Transition path theory was used to identify paths connecting unbound and bound states and to calculate the flux through the different pathways. The kinetic pathways are illustrated in Fig 4. The contributions of the four binding pathways [MS0 → MS1 → MS3], [MS0 → MS3], [MS0 → MS2 → MS1 → MS3] and [MS0 → MS2 → MS3] to the total flux from MS0 to MS3 are 52%, 39%, 5% and 4% respectively, showing that there is no single dominant pathway. The direct conversion MS0 → MS3 reflects the very rapid transformation of ligand from unbound to bound macrostates with no detectable intermediates. For the MS0 → MS3 path there have been occurrences where the ligand traverses rapidly through the helix7/helix9 or helix4/helix6 gateways or via the helix5/helix6/helix7/helix8 interface. The direct transitions from unbound to bound conformations thus

often involve ligand passage via helix gateways. Out of four pathways derived by the MSM analysis two are dominant, involving MS0, MS1 and MS3, and jointly account for 90% of the flux. However, long simulation trajectories also showed a distinct ligand-binding pathway via MS2, even though its contribution to the overall flux is less. The observation of multiple pathways for this system argues that binding of ligands to proteins in general is likely to be more complex than via a single pathway which is often the model used in the analysis of binding data. As a final note, the [MS0 → MS2 → MS3] path can be identified with trajectory 1 (S1 Movie), while the [MS0 → MS1 → MS3] pathway is a composite of trajectories 2 and 3 (S2 and S3 Movies).

Insight into the low activation barrier for ligand binding to T4L L99A

The fact that benzene is able to rapidly bind to T4L L99A is consistent with a low activation barrier. The activation free energy ΔG^* is not readily available from experimental data, as the temperature dependencies of measured rate constants provide only an estimate of the activation enthalpy ΔH^* [28]. By assuming a diffusion controlled ligand-protein binding rate of $\sim 10^9$ s^{-1}M^{-1}, and comparing this value with the experimentally measured $k_{on} \sim 10^6$ s^{-1}M^{-1} rate, Feher et. al. estimated the free energy barrier for binding to be relatively small, 4–5 kcal/mol (6.5 to 8.5 RT) [8]. Consequently, the barrier for the unbinding process, $\Delta G^*_{unbinding}$, can be calculated to be $\sim$8–9 kcal/mol (13–15 RT), based on $\Delta G^0_{binding}$ = -4.2 kcal/mol (Table 1). The MD simulations are also consistent with a small barrier as the observed binding rates are faster than the experimentally measured ones, even after accounting for the faster diffusion constant of the TIP3P-modeled water used here. The unbinding activation barrier can be estimated from the MD simulations using the relation $k_{off} = (1/\tau_{TPT})e^{-\Delta G^*_{unbinding}/RT}$ [29], where the transition path time τ_{TPT} is the time required for benzene to transition from the cavity to the unbound form. Note that τ_{TPT} is obtained from simulation while k_{off} can be measured experimentally or estimated by simulation. Values of τ_{TPT} have been quantified directly from the six binding trajectories. In this approach, a benzene molecule was considered to be bound to T4L L99A if any of its carbon atoms was less than 0.38 nm from the Ala 99 Cβ position and, conversely, unbound when every carbon atom was greater than 0.6 nm from the Ala 99 Cβ. Based on this criterion, τ_{TPT} values for the six trajectories vary from 0.2 to 24 ns with an average of 7 ns. Using this average value along with an experimental value for k_{off} of 0.95x10^3 s^{-1} [8], $\Delta G^*_{unbinding}$ = 11.9 RT was estimated, in agreement with the barrier obtained from the analysis of Feher et al. The small value of $\Delta G^*_{binding}$ $\sim$5 RT, that is estimated from $\Delta G_{binding} = \Delta G^*_{binding} - \Delta G^*_{unbinding}$, further emphasizes the small activation barrier to binding. In the above analysis barriers were estimated assuming a single binding pathway but the conclusion regarding small barriers will hold for the multiple paths as each of them occurred spontaneously during the simulations.

The MD trajectories described above provide insights into the origin of the small activation barrier for ligand binding. In all three binding routes identified in this study (Fig 3) the ligand reaches the cavity through the transient formation of pathways that are linked to the displacement of specific helices by 2–3 Å, without large scale changes in secondary structure or unfolding. We hypothesized that the L99A mutation not only creates the binding cavity but might also increase the flexibility of the cavity-containing C terminal domain in general, leading to an increased probability for the adoption of conformations that allow ligand binding. To test this hypothesis we calculated backbone order parameters squared, S^2, from chemical shifts for both T4L L99A and T4L WT (referred to as T4L WT* in what follows, see Material and methods) [30]. S^2 values report on the amplitudes of ps-ns motions of amide bond vectors
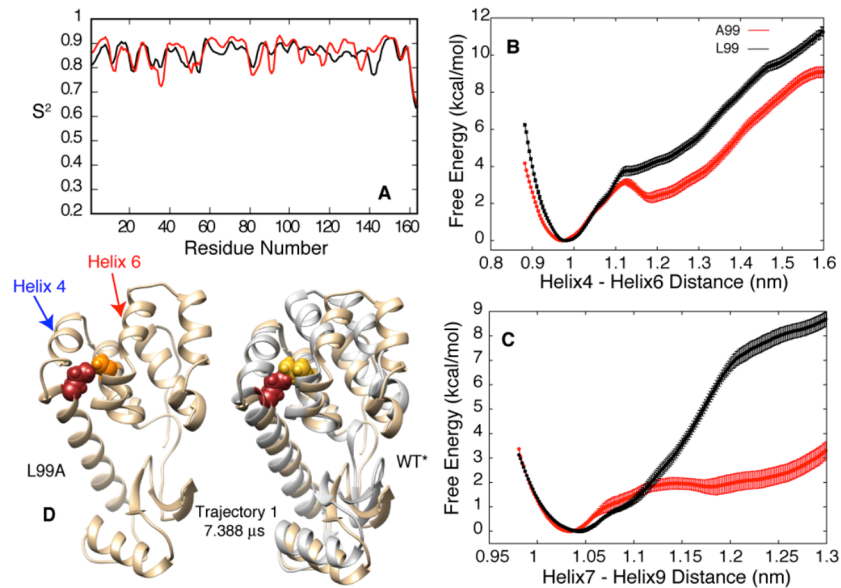
Fig 5. The L99A mutation stabilizes rare conformers that allow benzene to reach the cavity without substantial changes to ps-ns timescale backbone dynamics of the C terminal domain. A) NMR chemical shift derived S^2 values [30] are very similar for T4L WT* and T4L L99A showing that the L99A mutation does not change the overall flexibility of the protein backbone on the ps-ns timescale. B,C) Free energy profiles as a function of distance between helices 4 and 6 and helices 7 and 9 showing that a larger separation between helices is more likely in the T4L L99A mutant. D) Steric clashes between Leu99 and Ile78 destabilize conformers of the WT* protein when the inter-helical distance has increased to allow benzene access to the cavity. The snapshot from trajectory 1 at 7.388 µs (L99A) has a helix 4—helix 6 distance of 1.32 nm, with no steric clash between Ile78 (dark red) and Ala99 (orange). In contrast, in the WT* protein the larger Leu residue at the 99 position (yellow) leads to a steric clash between Ile78 and Leu99, destabilizing the conformation. The A99 to L99 mutation was performed by superimposing the backbone coordinates of residues 95 to 103 from the WT* crystal structure in grey on the snapshot from the simulation in brown.

[31], range between 0 (isotropic motion)– 1 (rigid), and are routinely used to identify flexible regions of proteins [32,33]. The similar S^2 vs residue profiles for both T4L L99A and WT* (Fig 5A) indicate that fast timescale backbone dynamics are not affected by the L99A mutation, while the high S^2 values (0.8 to 0.9) show that both forms of T4L are rigid and that rapid back-bone motions are not linked directly to ligand binding. In order to ascertain whether the L99A mutation increases the population of transiently formed high energy conformers that would allow benzene to move past helices and enter the cavity we have computed free energy profiles for both T4L L99A and WT* as a function of the distances between helices 4 and 6 (Fig 5B) and between helices 7 and 9 (Fig 5C). Interestingly, we observe that the free energy penalty for larger inter-helical distances in T4L WT* is greater than for T4L L99A, and thus the conforma-tions that facilitate binding are less probable for T4L WT*. For example, the probability of a Helix 4/6 separation of 1.30 nm, that is required for benzene entry (Fig 2B and 2C) increases by ∼ 40 fold in the L99A mutant, as it is 2.21 kcal/mol more unfavorable to separate this helix pair by 1.30 nm for the WT* sequence ($e^{\Delta G/RT} = 39$ for $\Delta G = 2.21$ kcal/mol). It is thus unlikely that a molecule of benzene could 'squeeze' between the helices in T4L WT* before reaching the occluded binding site due to the steric effect of residue L99. As the helices forming the C terminal domain of T4L are packed against one another, creating openings between helices that are required for binding (Fig 3) results in subtle positional changes to the other helices as well. A closer inspection of the trajectories showed that movement of helices 4 and 6 or 7 and 9 that is required to accommodate binding reduces the distance between side-chains of Ile78 and Ala99. A similar change in these inter-helical distances in the WT* protein would lead to a

steric clash between Ile78 and the larger Leu99, as illustrated in Fig 5D in the case of Trajectory 1 where the helix 4-helix 6 distance transiently increases for only a few nanoseconds. We would like to emphasize that the population of the conformer where helices 4 and 6 are separated by 1.30 nm is extremely low, even for T4L L99A, ($e^{-\Delta G/RT} \sim 0.0046\%$ with $\Delta G = 3.23$ kcal/mol, Fig 5B) relative to the equilibrium conformer corresponding to a helix4-helix6 distance of ~1 nm. Thus, this ligand-accessible state cannot be quantified through experiment, further emphasizing the importance of large scale MD simulations like those performed here. A similar scenario is also found for the separation of helices 7 and 9 (Trajectory 2, Fig 3) where increasing the inter-helical distance from 1.05 nm to 1.25 nm, that is required for ligand entry, is more unfavorable by 5.9 kcal/mol for WT* than for T4L L99A.

Unbinding trajectories validate and add to the ensemble of pathways generated from unbiased MD sampling

The computational expense of unbiased binding simulations prohibits the exhaustive survey of all of the possible kinetic pathways. Hence we explored a complementary approach of validating the identified pathways and of enriching the ensemble of binding trajectories using a recent implementation of metadynamics simulation [25] to enable the extraction of the kinetics of benzene unbinding from the cavity. Using the radial distance between the binding pocket and the ligand as the collective variable, this simulation technique infrequently deposits repulsive, history-dependent Gaussian bias along the pocket-ligand distance so as to efficiently accelerate the unbinding process. The corresponding acceleration factor, multiplied by the simulation time, provides an estimate of the time for unbinding. As depicted in Fig D in S1 Supporting Information, the p-value analysis [26], as suggested by Tiwary and Parrinello, provided a good Poisson fit with a high p-value of 0.73, when averaged over numerous independent trajectories. The computed off rate, 369 s^{-1}, is in reasonable agreement with the experimentally measured off-rate, Table 1, showing that the ligand-pocket distance is a relevant collective variable for exploring benzene-T4L L99A unbinding pathways. Significantly, as depicted in Fig D in S1 Supporting Information, the unbinding trajectories revealed two unbinding pathways of the ligand from the pocket. These involved benzene exiting through small openings between helices 7 and 9 and the juncture of helices 5, 6 7 and 8, pathways that are the reverse of the subset of those previously identified for binding.

Discussion

Although ligand binding to receptors is critically important for many biological processes an atomistic understanding of how this might occur in the context of occluded binding sites is lacking. This reflects the fact that while the main biophysical tools that are used to study molecular interactions are powerful for characterizing and providing atomic resolution structures of the end points of the binding process (free and bound states) they are often much less robust in generating a description of the binding mechanism. Improvements in both computer technology and MD simulation methodology have made it possible to obtain μs-ms MD trajectories of proteins in explicit solvent so that it is possible to study processes such as ligand binding, conformational exchange, and protein folding using MD simulations [12,18,19,34]. Here, using unbiased MD simulations, we have addressed the longstanding question of how hydrophobic molecules rapidly reach buried cavities in proteins by working with a model system in which benzene binds to a 150 Å^3 cavity mutant in T4L L99A. Central to every MD-based approach is that the force field used should be able to model the underlying free energy surface accurately. We chose to use the CHARMM force field [35] as previous MD studies had shown that the Phe114 buried to exposed conformational transition in T4L L99A and in the

related protein, T4L L99A/G113A/R119P, is well modeled with it [12,36]. The structures of the benzene bound conformers obtained from the unbiased simulations are in excellent agreement with those determined by crystallography (backbone RMSD < 2 Å, Fig 2C), with the position of the benzene within 2 Å of the binding site residues established experimentally [4]. Further, kinetic and thermodynamic parameters obtained from the simulations are in reasonable agreement with the experimental values (Table 1), providing confidence in the results of the MD study.

Our unbiased μs timescale MD simulations establish that benzene, initially fully hydrated, reaches the internal cavity created by the L99A mutation via at least three different trajectories. Notably, the process does not involve a large-scale deformation to the T4L L99A structure, nor significant changes to secondary structure, but rather small perturbations to the positions of two or more helices in the C terminal domain of the protein that create pathways for binding (Fig 3). Similar tunnels from the protein surface to the cavity were also observed in a 30 μs MD simulation of T4L L99A using the AMBER force field [37]. The existence of multiple pathways coupled with a small net activation barrier explains how benzene can bind the occluded cavity rapidly.

Previously we had used combined MD simulations and NMR relaxation experiments to study how the side-chain of residue Phe114 fills the cavity created by the L99A mutation, that in some sense serves as a surrogate to hydrophobic ligands such as the benzene molecule considered here. It might be expected, therefore, that the binding pathways of Phe114 and benzene would share some similar features. Notably, in one of the trajectories (Trajectory 3) benzene enters the cavity in a manner similar to Phe114 [12] from its solvent exposed conformation.

The binding of ligands to occluded sites in other proteins, such as oxygen binding to the buried heme group in myoglobin, has been studied using a variety of experimental and computational techniques [38,39] In the case of myoglobin, the mechanism of binding also involves small barriers, multiple pathways and secondary binding sites in the protein [38,39]. Multiple pathways have also been observed by Dickson and Lotz in recent studies of an epoxide hydrolase inhibitor [40] and a trypsin-benzamidine system [41] that exploited a weighted ensemble based algorithm. The plasticity of T4L L99A allows the larger benzene molecule to reach the cavity in a manner similar to how oxygen connects with the heme group in myoglobin. The origin of the conformational plasticity in T4L L99A that facilitates this process is, in fact, the L99A mutation itself (Fig 5). Replacing the Leu99 residue with the smaller Ala not only creates the cavity but also allows helices to move with respect to one another to generate the pathways that are required for binding. In contrast, the presence of Leu99 destabilizes the binding competent conformations by clashing with Ile78 (Fig 5). It remains of interest to investigate if the benzene binding rate can be modulated by mutating Ile78 to smaller hydrophobic residues. Similar to the binding of benzene, protein plasticity and cavities have also been implicated in oxygen binding to T4L L99A [14].

Preliminary observations into the mechanism of ligand unbinding from T4L L99A have recently been published and are in agreement with the results of this study. Kitahara et. al. [14] developed a simple and elegant ^{15}N NMR based method to detect O_2 binding sites in proteins using T4L L99A as a model system. In order to quantify the ^{15}N chemical shift changes that were observed upon oxygen binding MD simulations were performed to understand the rotational and translational diffusion properties of oxygen molecules in the T4L L99A cavities. Notably, an oxygen molecule escaped from the cavity via a pathway that is the reverse of Trajectory 1 (Fig 3). In MD simulations performed to understand how water interacts with the cavity at high pressures, water molecules were found to escape from the cavity due to transient openings formed at the junction of helices 5, 6 and 7 [42], as observed in the creation of a pathway to the cavity in Trajectory 3. In adiabatic-biased MD simulations of benzene unbinding

from T4L L99A a pathway that is the reverse of Trajectory 3 was observed [36]. In a separate 27 μs MD trajectory of T4L L99A performed to understand the Phe114 exposed to buried conformational transition, tunnels leading to the cavity from solvent were transiently formed [37], as observed in the present work. A recent implementation of a biased simulation approach by Lindorf-Larson and coworkers also led to the extraction of the kinetics of the benzene recognition process [43]. McCammon and coworkers [44] have employed an accelerated MD-based biased simulation approach to explore the thermodynamics of benzene binding to T4L L99A cavity. Very recently, Arantes and coworkers [45] have reported weighted ensemble based biased simulations of unbinding of benzene from the T4L cavity and revealed multiple benzene exit pathways, in agreement with the current study. However, none of these simulation-based studies address the mechanism by which ligands *spontaneously* bind to buried protein cavities and consequently insights into multiple pathways and low free energy barriers were not obtained.

Recent studies suggest that diverse processes such as protein folding and protein interconversion between different compact conformers proceed via multiple pathways and/or small barriers [12,46,47]. More studies are required to determine if small activation barriers and multiple paths are common features of biomolecular recognition processes and to characterize at atomic resolution what these pathways are.

Materials and methods

Here, as in most biophysical studies involving T4 lysozyme, the cysteine free version of the protein [48] was used, T4L WT*, where Cys54 and Cys97 are replaced by Thr and Ala, respectively. In T4L L99A the L99A mutation is introduced into the T4L WT* background.

MD simulations and analysis

Unbiased binding simulations. The X-ray crystallographic structures of T4L L99A free and benzene bound forms (PDB code: 3DMX and 3DMV) were used as the starting points for all of the simulations [4]. All binding simulations were initiated by placing the apo form of L99A T4L at the center of a cubic box of dimension 7.2 nm with the empty space then filled with water and ion molecules. The system was solvated with 11613 triatomic water molecules and sufficient numbers of sodium and chloride ions were added to keep the sodium chloride concentration at 150 mM and render the system charge neutral. Two benzene molecules, corresponding to a concentration of 9.55 mM, that is less than the benzene solubility limit in water (20 mM) [49], were placed in random positions in the solvent. The two molecules were allowed to diffuse freely and no artificial bias was introduced throughout the simulation. The system included a total of 37541 atoms. The protein, benzene and ions were parameterized with the charmm36 force field [35] and the TIP3P water model [27] was used.

All MD simulations were performed with the Gromacs 5.0.6 simulation package [50], in most cases benefiting from usage of Graphics processing units [51]. During the simulation, the average temperature was maintained at 303K using the Nose-Hoover thermostat [52,53] with a relaxation time of 1.0 ps and an average isotropic pressure of 1 bar was maintained with the Parrinello-Rahman barostat [54]. The Verlet cutoff scheme [55] was employed throughout the simulation with the Lenard Jones interaction extending to 1.2 nm and long-range electrostatic interactions treated by Particle Mesh Ewald (PME) summation [56]. All bond lengths involving hydrogen atoms of the protein and the ligand benzene were constrained using the LINCS algorithm [57] and water hydrogen bonds were fixed using the SETTLE approach [58]. Simulations were performed using the leapfrog integrator with a time step of 2 fs and initiated by randomly assigning the velocities of all particles.

Initially six long independent unbiased trajectories were performed, differing in the assignment of initial velocity seeds. These ranged between 2.5 and 8.1 μs with a total simulation length of 29 μs. The simulations were only terminated after one of the two copies of the ligand settled into the target binding site. The binding process was ascertained via inspection and by checking (i) the radial distance between respective centers of mass of benzene and the cavity and (ii) the root-mean-squared deviation (RMSD) of the simulated conformation of the protein-ligand system from that of X-ray crystal structure of holo T4L L99A (PDB id: 3DMX) [4] that included benzene atoms and those of the protein cavity. The 'cavity' or the 'binding pocket' was defined by the protein heavy atoms within 5Å of benzene in the X-ray structure of holo T4L L99A (PDB id: 3DMX). For the calculation of RMSD, the cavity plus ligand of each frame of the trajectory was first translationally and rotationally aligned with that of the crystal structure of the benzene-bound form of protein (PDB id: 3DMX) and then the deviation was computed. The ligand was ascertained to be bound to the cavity when the RMSD remained below 0.5 nm and the cavity-ligand distance was below 0.2 nm for a simulation duration of at least 100 ns.

Apart from the six μs-long trajectories discussed above, we have also performed 300 short independent trajectories, each 100 ns in duration. These short trajectories were initiated from different intermediates observed from the long binding simulations and improved the statistics of the MSM that was derived from analysis of the MD trajectories (see below). Overall, an aggregate of about 59 μs of unbiased trajectory was recorded.

MSM analysis of the MD data. The generation of a MSM is a powerful way to automatically identify the kinetically relevant states and their interconversion rates from the simulated trajectories [59–62]. We employed PyEMMA [63] (http://pyemma.org) to construct and analyze the MSM obtained from the combination of all the recorded MD trajectories. The nearest-neighbor (cut-off 0.5 nm) heavy-atom contacts between benzene and protein residues were used as input coordinates. The time-lagged or time-structure-based independent component analysis (tICA) [64–66] with a correlation lag time of 20 ns was used for dimensionality reduction, which linearly transforms the high dimensional input data into a slow linear subspace (i.e. collective coordinates sorted by 'slowness'). The high dimensional input coordinates were subsequently projected onto 12 tICA components, covering more than 95% of the kinetic variance. We employed the k-mean clustering algorithm [67] to discretize the tICA data into 100 clusters. Then 100 microstates MSMs were constructed at variable lag times to identify the appropriate lag time. A lag time of 10 ns was used to construct the final MSM as the implied time scale leveled off at about 10 ns which ensures the Markovianity of the model (see Fig E in S1 Supporting Information). To better understand the dynamics of ligand binding a coarse-grained kinetic model with four metastable states was constructed using a hidden Markov model (HMM) [68].

The kinetic parameters reported in Table 1 were calculated using the mean first passage time (MFPT) obtained from the MSM analysis based on four derived macrostates. The MFPT was computed as the average time taken for the transition from the initial to the final macrostate [69]. The calculation included both the direct transition from the initial state to the final state and transitions through other intermediate states. The on-rate and off-rate constants are respectively calculated as $k_{on} = 1/(MFPT_{on}C)$ and $k_{off} = 1/MFPT_{off}$ where C is the benzene concentration, 9.55 mM. The binding free energies (ΔG) were calculated based on the stationary populations (shown in Table 2) of bound and unbound macrostates as obtained from the MSM [20],

$$\Delta G = -k_B T \, log\left(\frac{\pi_{bound}}{\pi_{unbound}}\right)$$

where π_{bound} and $\pi_{unbound}$ are stationary populations of bound and unbound macrostates, k_B is Boltzmann's constant, and T is the absolute temperature. We subsequently converted the result to a standard free energy ΔG^0 (one molar concentration) for comparison with experiment. Finally, the transition path theory as proposed by Vanden-Eijnden and coworkers [70] was employed to calculate the transition path fluxes.

Free energies of helix opening. The umbrella sampling simulation technique [71] was used to map the free energy surface underlying the opening of a pair of helices to facilitate ligand binding to the cavity in T4L L99A. To this end, we performed two independent umbrella sampling simulations where the distance between centers of mass of Cα-atoms of helices 4 and 6 or of helices 7 and 9 was used as the collective variable (CV). The value of the CV ranged from 0.9 nm—1.6 nm for the distance between helices 4 and 6 and from 1 to 1.3 nm for helices 7 and 9. The starting configuration for each window was picked from the prior unbiased simulation trajectories, with spacing between consecutive windows of 0.05 nm. Each window was restrained to the desired value of the CV using a harmonic potential of force constant 8000 kJ/mol/nm² to ensure that the distribution of the CV for each window remains Gaussian with sufficient overlap in the distribution of two consecutive windows so that the entire space of the CV is sampled. Each of the windows was sampled by the umbrella sampling technique for 20 ns. Finally, the weighted histogram analysis method (WHAM) [72,73] was employed to reweight the umbrella-sampled windows and to obtain the unbiased potential of mean force as a function of the CV. In order to investigate the effect of the L99A mutation on the free energetics of helix opening, the entire series of umbrella samplings was repeated by in-silico mutation of Ala99 in each of the starting configurations to Leu99.

Unbinding simulations using an enhanced sampling technique. The recent implementation of metadynamics by Tiwary and Parrinello [25] was used to compute unbinding rates of benzene from the hydrophobic cavity of L99A T4L and to extract unbinding pathways. In this novel method, by making the bias deposition slower than the time in bottlenecks, and by using a suitable CV capable of distinguishing between bound and unbound states, it is possible to keep the transition state relatively bias-free through the course of the metadynamics simulation.

This so-called 'infrequent metadynamics' approach allows one to accelerate the transition from ligand-bound to unbound states, without affecting the transition state. More importantly, this technique provides an estimate of the acceleration factor, which when multiplied with metadynamics simulation times, provides the true unbinding time. As elaborated in the original work of Tiwary and Parrinello [25], the acceleration factor is given by the time $\alpha = \langle e^{\beta V(s,t)} \rangle_t$ where s is the CV being biased, β is the inverse temperature, $V(s,t)$ is the bias experienced at time t and the subscript t indicates that averaging is performed under the time-dependent potential. The above expression is valid even if there are multiple intermediate states and numerous alternative reaction pathways. In our current work, we employed the distance between the centers of mass of the cavity and benzene as the CV using a Gaussian width of 0.025 Å. One of the reasons for choosing the distance between the binding pocket and the ligand as the CV was that it can distinguish between bound and unbound states very well, while providing an estimate of binding free energies that agree with those obtained from experiment. During the 'infrequent metadynamics' simulation approach, the Gaussians were deposited every 10 ps with a starting height of 1.2 kJ/mol and gradually decreased on the basis of a well-tempered metadynamics biasing factor, $\Gamma = 6$. A total of 15 such 'infrequent metadynamics' simulations having different initial velocity distributions were spawned with the ligand fully bound to the cavity and simulations were stopped only when the ligand became fully unbound, reaching bulk-solvent i.e. the distance between the binding pocket and the ligand exceeded 3 nm. Acceleration factors, as described earlier and elaborated on in reference [25],

were computed during the simulations and multiplied with the Metadynamics run time to get the true unbinding time.

The transition time statistics so obtained were then subject to a Poisson analysis [26] to ascertain their reliability (based on the Kolmogorov-Smirnov statistical analysis) and to obtain the average unbinding time, from which the unbinding rate is easily calculated. Our approach of performing metadynamics simulations using the cavity-ligand separation as a single CV was sufficient to pass the so-called p-value test which suggested that the frequency of bias was slow enough and the CV was appropriate to accelerate unbinding. All simulations were carried out using Gromacs 5.0.6 patched with the PLUMED plugin [74].

NMR experiments and analysis

NMR samples. ^{15}N,^{13}C labeled T4L WT* and T4L L99A were overexpressed and purified as described previously [75]. The protein concentration in the NMR samples (50 mM sodium phosphate, 25 mM NaCl, 2mM EDTA, 2mM NaN$_3$, pH 5.5, 10% D$_2$O buffer) was ∼1.5 mM.

NMR experiments. Assignments for T4L WT* were obtained using standard triple resonance experiments [76,77], with assignments for T4L L99A published earlier [11]. All experiments were performed at 22°C on a 700 MHz Bruker Avance III HD spectrometer equipped with a triple resonance cold probe.

NMR data processing and analysis. NMR data were processed using the NMRPipe software package [78] and the resulting spectra were visualized and analyzed using SPARKY [79]. To overcome complications due to chemical exchange, order parameters were estimated from chemical shifts according to the method of Berjanskii and Wishart [30], as implemented in TALOS+ [80].

Supporting information

S1 Supporting Information. PDF file containing five supporting figures including Fig A, B, C, D and E.
(PDF)

S1 Movie. Movie highlighting trajectory 1. Benzene binds to the cavity via the gateway between helices 4 and 6, created by transient distance fluctuations between these elements of structure. The protein is shown in cartoon representation, the cavity by a solid surface representation and the benzene molecule via a licorice representation.
(MP4)

S2 Movie. Movie of trajectory 2 showing benzene binding via a pathway between helices 7 and 9 created by transient helix 7–9 distance fluctuations. The protein is shown in cartoon representation, the designated cavity by a solid surface representation and the benzene molecule via a licorice representation.
(MP4)

S3 Movie. Movie of trajectory 3 whereby benzene enters T4L L99A through the junction region of helices 5–8. The protein is shown in a cartoon representation, the designated cavity by a solid surface representation and the benzene molecule via a licorice representation.
(MP4)

Author Contributions

Conceptualization: Jagannath Mondal, Lewis E. Kay, Pramodh Vallurupalli.

Data curation: Jagannath Mondal, Navjeet Ahalawat, Subhendu Pandit, Pramodh Vallurupalli.

Formal analysis: Jagannath Mondal, Navjeet Ahalawat, Subhendu Pandit, Lewis E. Kay, Pramodh Vallurupalli.

Investigation: Jagannath Mondal, Navjeet Ahalawat, Subhendu Pandit, Lewis E. Kay, Pramodh Vallurupalli.

Methodology: Jagannath Mondal, Navjeet Ahalawat, Subhendu Pandit, Pramodh Vallurupalli.

Supervision: Jagannath Mondal, Pramodh Vallurupalli.

Validation: Jagannath Mondal, Pramodh Vallurupalli.

Visualization: Jagannath Mondal, Navjeet Ahalawat, Pramodh Vallurupalli.

Writing – original draft: Jagannath Mondal, Navjeet Ahalawat, Lewis E. Kay, Pramodh Vallurupalli.

References

1. Lu H, Tonge PJ. Drug-target residence time: Critical information for lead optimization. Current Opinion in Chemical Biology. 2010; 14: 467–474. https://doi.org/10.1016/j.cbpa.2010.06.176 PMID: 20663707

2. Swinney DC. Applications of Binding Kinetics to Drug Discovery. Pharmaceut Med. 2008; 22: 23–34. https://doi.org/10.1007/BF03256679

3. Pan AC, Borhani DW, Dror RO, Shaw DE. Molecular determinants of drug-receptor binding kinetics. Drug Discov Today. 2013; 18: 667–73. https://doi.org/10.1016/j.drudis.2013.02.007 PMID: 23454741

4. Liu L, Baase WA, Matthews BW. Halogenated Benzenes Bound within a Non-polar Cavity in T4 Lysozyme Provide Examples of I ∙∙∙ S and I ∙∙∙ Se halogen-bonding. J Mol Biol. 2009; 385: 595–605. https://doi.org/10.1016/j.jmb.2008.10.086 PMID: 19014950

5. Eriksson AE, baase WA, Wozniac JA, Matthews BW. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. Nature. 1992; 355: 371–373. https://doi.org/10.1038/355371a0 PMID: 1731252

6. Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Blaber M, Baldwin EP, et al. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. Science. 1992; 255: 178–183. https://doi.org/10.1126/science.1553543

7. Morton A, Baase WA, Matthews BW. Energetic origins of specificity of ligand binding in an interior non-polar cavity of T4 lysozyme. Biochemistry. 1995; 34: 8564–8575. https://doi.org/10.1021/bi00027a006 PMID: 7612598

8. Feher V, EP B, Dahlquist FW. Access of Ligands to cavities within the core of a protein is rapid. Nat Struct Biol. 1996; 3: 516–521. PMID: 8646537

9. Sekhar A, Kay LE. NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. Proc Natl Acad Sci. 2013; 110: 12867–12874. https://doi.org/10.1073/pnas.1305688110 PMID: 23868852

10. Mulder FA, Mittermaier A, Hon B, Dahlquist FW, Kay LE Studying excited states of proteins by NMR spectroscopy. Nat Struct Biol. 2001; 8: 932–5. https://doi.org/10.1038/nsb1101-932 PMID: 11685237

11. Bouvignies G, Vallurupalli P, Hansen DF, Correia BE, Lange O, Bah A, et al. Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. Nature. 2011; 477: 111–114. https://doi.org/10.1038/nature10349 PMID: 21857680

12. Vallurupalli P, Chakrabarti N, Pomes R, Kay LE. Atomistic picture of conformational exchange in a T4 lysozyme cavity mutant: an experiment-guided molecular dynamics study. Chem Sci. 2016; 7: 3602–3613. https://doi.org/10.1039/C5SC03886C

13. Mulder FAA, Hon B, Muhandiram DR, Dahlquist FW, Kay LE. Flexibility and Ligand Exchange in a Buried Cavity Mutant of T4 Lysozyme Studied by Multinuclear NMR. Biochemistry. 2000; 39: 12614–12622. https://doi.org/10.1021/bi001351t PMID: 11027141

14. Kitahara R, Yoshimura Y, Xue M, Kameda T, Mulder FAA. Detecting O2 binding sites in protein cavities. Sci Rep; 2016; 6: 20534. Available: http://dx.doi.org/10.1038/srep20534

15. Quillin ML, Breyer WA, Griswold IJ, W MB. Size versus polarizability in protein-ligand interactions: binding of noble gases within engineered cavities in phage T4 lysozyme. J Mol Biol. 2000; 302: 955–977. https://doi.org/10.1006/jmbi.2000.4063 PMID: 10993735

16. Deng Y, Roux B. Computations of standard binding free energies with molecular dynamics simulations. J Phys Chem B. 2009; 113: 2234–46. https://doi.org/10.1021/jp807701h PMID: 19146384

17. Mobley DL, Graves AP, Chodera JD, McReynolds AC, Shoichet BK, Dill K a. Predicting absolute ligand binding free energies to a simple model site. J Mol Biol. 2007; 371: 1118–34. https://doi.org/10.1016/j.jmb.2007.06.002 PMID: 17599350

18. Shan Y, Kim ET, Eastwood MP, Dror RO, Seeliger MA, Shaw DE. How does a drug molecule find its target binding site? J Am Chem Soc. 2011; 133: 9181–9183. https://doi.org/10.1021/ja202726y PMID: 21545110

19. Dror RO, Pan AC, Arlow DH, Borhani DW, Maragakis P, Shan Y, et al. Pathway and mechanism of drug binding to G-protein-coupled receptors. Proc Natl Acad Sci USA. 2011; 108: 13118–13123. https://doi.org/10.1073/pnas.1104614108 PMID: 21778406

20. Buch I, Giorgino T, De Fabritiis G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. Proc Natl Acad Sci USA. 2011; 108: 10184–10189. https://doi.org/10.1073/pnas.1103547108 PMID: 21646537

21. Plattner N, Noé F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. Nat Comm; 2015; 6: 7653. Available: http://dx.doi.org/10.1038/ncomms8653

22. Bowman GR. An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation. Advances in Experimental Medicine and Biology. 2014; 797: 7–23. https://doi.org/10.1007/978-94-007-7606-7

23. Chodera JD, Noé F. Markov state models of biomolecular conformational dynamics. Curr Opin Struct Biol. 2014; 25: 135–144. https://doi.org/10.1016/j.sbi.2014.04.002 PMID: 24836551

24. Pande VS, Beauchamp K, Bowman GR. Everything you wanted to know about Markov State Models but were afraid to ask. Methods. 2010; 52: 99–105. https://doi.org/10.1016/j.ymeth.2010.06.002 PMID: 20570730

25. Tiwary P, Parrinello M. From Metadynamics to Dynamics. Phys Rev Lett. 2013; 111: 230602–230606. https://doi.org/10.1103/PhysRevLett.111.230602 PMID: 24476246

26. Salvalaglio M, Tiwary P, Parrinello M. Assessing the Reliability of the Dynamics Reconstructed from Metadynamics. J Chem Theory Comp. 2014; 10: 1420–1425. https://doi.org/10.1021/ct500040r PMID: 26580360

27. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein MW. Comparison of simple potential functions for simulating liquid water. JChemPhys. 1983; 79: 926–935.

28. Moore PB. How should we think about the ribosome? Annu Rev Biophys. 2012; 41: 1–19. https://doi.org/10.1146/annurev-biophys-050511-102314 PMID: 22577819

29. Chandler D. Introduction to Modern Statistical Mechanics. Oxford: Oxford University Press; 1987.

30. Berjanskii M V, Wishart DS. A simple method to predict protein flexibility using secondary chemical shifts. J Am Chem Soc. 2005; 127: 14970–14971. https://doi.org/10.1021/ja054842f PMID: 16248604

31. Lipari G, Szabo A. Model-Free Approach To The Interpretation Of Nuclear Magnetic-Resonance Relaxation In Macromolecules .1. Theory And Range Of Validity. J Am Chem Soc. 1982; 104: 4546–4559.

32. Kay LE, Torchia DA, Bax A. Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry. 1989; 28: 8972–8979. https://doi.org/10.1021/bi00449a003

33. Palmer AG. NMR Characterization of the Dynamics of Biomacromolecules. Chem Rev. 2004; 104: 3623–3640. https://doi.org/10.1021/cr030413t PMID: 15303831

34. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, et al. Atomic-level characterization of the structural dynamics of proteins. Science. 2010; 330: 341–6. https://doi.org/10.1126/science.1187409 PMID: 20947758

35. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, et al. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain $\chi1$ and $\chi2$ Dihedral Angles. JChemTheo Comput. 2012; 8: 3257–3273. https://doi.org/10.1021/ct300400x PMID: 23341755

36. Wang Y, Papaleo E, Lindorff-Larsen K. Mapping transiently formed and sparsely populated conformations on a complex energy landscape. Shan Y, editor. Elife. 2016; 5: e17505. https://doi.org/10.7554/eLife.17505 PMID: 27552057

37. Schiffer JM, Feher VA, Malmstrom RD, Sida R, Amaro RE. Capturing Invisible Motions in the Transition from Ground to Rare Excited States of T4 Lysozyme L99A. Biophys J. 2016; 111: 1631–1640. https://doi.org/10.1016/j.bpj.2016.08.041 PMID: 27760351

38. Elber R. Ligand diffusion in globins: simulations versus experiment. Curr Opin Struct Biol. 2010; 20: 162–167. https://doi.org/10.1016/j.sbi.2010.01.002 PMID: 20116995

39. Yu TQ, Lapelosa M, Vanden-Eijnden E, Abrams CF. Full kinetics of CO entry, internal diffusion, and exit in myoglobin from transition-path theory simulations. J Am Chem Soc. 2015; 137: 3041–3050. https://doi.org/10.1021/ja512484q PMID: 25664858

40. Lotz SD, Dickson A. Unbiased Molecular Dynamics of 11 min Timescale Drug Unbinding Reveals Transition State Stabilizing Interactions. J Am Chem Soc. 2018; 140: 618–628. https://doi.org/10.1021/jacs.7b08572 PMID: 29303257

41. Dickson A, Lotz SD. Multiple Ligand Unbinding Pathways and Ligand-Induced Destabilization Revealed by WExplore. Biophys J. 2017; 112: 620–629. https://doi.org/10.1016/j.bpj.2017.01.006 PMID: 28256222

42. Collins MD, Hummer G, Quillin ML, Matthews BW, Gruner SM. Cooperative water filling of a nonpolar protein cavity observed by high-pressure crystallography and simulation. Proc Natl Acad Sci. 2005; 102: 16668–16671. https://doi.org/10.1073/pnas.0508224102 PMID: 16269539

43. Wang Y, Martins JM, Lindorff-Larsen K. Biomolecular conformational changes and ligand binding: from kinetics to thermodynamics. Chem Sci. 2017; 8: 6466–6473. https://doi.org/10.1039/C7SC01627A PMID: 29619200

44. Miao Y, Feher VA, McCammon JA. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. J Chem Theory Comput. 2015; 11: 3584–3595. https://doi.org/10.1021/acs.jctc.5b00436 PMID: 26300708

45. Nunes-alves A, Zuckerman DM, Arantes GM. Escape of a Small Molecule from Inside T4 Lysozyme by Multiple Pathways. Biophysj. 2018; 114: 1058–1066. https://doi.org/10.1016/j.bpj.2018.01.014 PMID: 29539393

46. Wolynes PG, Eaton WA, Fersht AR. Chemical physics of protein folding. Proc Natl Acad Sci U S A. 2012; 109: 17770–17771. https://doi.org/10.1073/pnas.1215733109 PMID: 23112193

47. Udgaonkar JB. Multiple routes and structural heterogeneity in protein folding. Annu Rev Biophys. 2008; 37: 489–510. https://doi.org/10.1146/annurev.biophys.37.032807.125920 PMID: 18573092

48. Matsumura M, Matthews BW. Control of enzyme activity by an engineered disulfide bond. Science. 1989; 243: 792 LP–794. Available: http://science.sciencemag.org/content/243/4892/792.abstract

49. Arnold D, Plank C, Erickson E, Pike F. Solubility of Benzene in Water. Ind Eng Chem Chem Eng Data Ser. 1958; 3: 253–256. https://doi.org/10.1021/i460004a016

50. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput. 2008; 4: 435–447. https://doi.org/10.1021/ct700301q PMID: 26620784

51. Kutzner C, Pall S, Fechner M, Esztermann A, de Groot BL, Grubmuller H. Best bang for your buck: GPU nodes for GROMACS biomolecular simulations. JComputChem. 2015; 36: 1990–2008. https://doi.org/10.1002/jcc.24030 PMID: 26238484

52. Nosé S. A molecular dynamics method for simulations in the canonical ensemble. Mol Phys. 1984; 52: 255.

53. Hoover WG. Canonical dynamics: equilibrium phase-space distributions. Phys Rev A. 1985; 31: 1695.

54. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys. 1981; 52: 7182–7190.

55. Pail S, Hess B. A flexible algorithm for calculating pair interactions on {SIMD} architectures. Comput Phys Comm.s. 2013; 184: 2641–2650. http://dx.doi.org/10.1016/j.cpc.2013.06.003

56. Darden T, York D, Pedersen L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. J Chem Phys. 1993; 98: 10089.

57. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. JComputChem. 1997; 18: 1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

58. Miyamoto S, K P A. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. J Comput Chem. 1992; 13: 952–962.

59. Noe F, Horenko I, Schitte C, Smith JC. Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states. JChemPhys. 2007; 126: 155102. https://doi.org/10.1063/1.2714539 PMID: 17461666

60. Chodera JD, Singhal N, Pande VS, Dill KA, Swope WC. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. JChemPhys. 2007; 126: 155101. https://doi.org/10.1063/1.2714538 PMID: 17461665

61. Sriraman S, Kevrekidis IG, Hummer G. Coarse Master Equation from Bayesian Analysis of Replica Molecular Dynamics Simulations. J Phys Chem B. 2005; 109: 6479–6484. https://doi.org/10.1021/jp046448u PMID: 16851726

62. Beauchamp KA, McGibbon R, Lin Y-S, Pande VS. Simple few-state models reveal hidden complexity in protein folding. Proc Natl Acad Sci USA. 2012; 109: 17807–17813. https://doi.org/10.1073/pnas.1201810109 PMID: 22778442

63. Scherer MK, Trendelkamp-Schroer B, Paul F, Pérez-Hernández G, Hoffmann M, Plattner N, et al. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. JChemTheo Comput. 2015; 11: 5525–5542. https://doi.org/10.1021/acs.jctc.5b00743 PMID: 26574340

64. Molgedey L, Schuster HG. Separation of a mixture of independent signals using time delayed correlations. Phys Rev Lett. 1994; 72: 3634–3637. https://doi.org/10.1103/PhysRevLett.72.3634 PMID: 10056251

65. Pérez-Hernández G, Paul F, Giorgino T, De Fabritiis G, Noé F. Identification of slow molecular order parameters for Markov model construction. JChemPhys. American Institute of PhysicsAIP; 2013; 139: 15102. https://doi.org/10.1063/1.4811489 PMID: 23822324

66. Schwantes CR, Pande VS. Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. JChemTheo Comput. 2013; 9: 2000–2009. https://doi.org/10.1021/ct300878a PMID: 23750122

67. Lloyd S. Least squares quantization in PCM. IEEE Trans Inf Theory. 1982; 28: 129–137. https://doi.org/10.1109/TIT.1982.1056489

68. Noé F, Wu H, Prinz J-H, Plattner N. Projected and hidden Markov models for calculating kinetics and metastable states of complex molecules. JChemPhys. 2013; 139: 184114. https://doi.org/10.1063/1.4828816 PMID: 24320261

69. Da L-T, Sheong FK, Silva D-A, Huang X. Application of Markov State Models to Simulate Long Time-scale Dynamics of Biological Macromolecules. Han K, Zhang X, Yang M, editors. Adv Exp Med Biol. 805: 29–66. https://doi.org/10.1007/978-3-319-02970-2_2 PMID: 24446356

70. Metzner P, Schütte C, Vanden-Eijnden E. Transition Path Theory for Markov Jump Processes. Multiscale Model Simul. 2009; 7: 1192–1219. https://doi.org/10.1137/070699500

71. Torrie GM, Valleau JP. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. J Comput Phys. 1977; 23: 187–199. http://dx.doi.org/10.1016/0021-9991(77)90121-8

72. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. THE weighted histogram analysis method for free-energy calculations on biomolecules. {I.} The method. J Comput Chem. 1992; 13: 1011–1021.

73. A.Grossfield. WHAM: the weighted histogram analysis method, version 2.0.

74. Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G. {PLUMED} 2: New feathers for an old bird. Comput Phys Comm.s. 2014; 185: 604–613. http://dx.doi.org/10.1016/j.cpc.2013.09.018

75. Vallurupalli P, Hansen DF, Lundstrom P, Kay LE. CPMG relaxation dispersion NMR experiments measuring glycine 1H alpha and 13C alpha chemical shifts in the "invisible" excited states of proteins. J Biomol NMR. 2009; 45: 45–55. https://doi.org/10.1007/s10858-009-9310-6 PMID: 19319480

76. Ikura M, Kay LE, Bax A. A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. Biochemistry. 1990; 29: 4659–4667. PMID: 2372549

77. Cavanagh J, Fairbrother WJ, Palmer AG, Rance M, Skelton NJ. Protein NMR Spectroscopy, Principles and Practice. 2nd ed. Academic Press; 2006.

78. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe—a Multidimensional Spectral Processing System Based on Unix Pipes. JBiomol NMR. 1995; 6: 277–293. Available: %3CGo

79. Goddard TD, Kneller DG. SPARKY 3 University of California, San Francisco. 2014.

80. Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR. 2009; 44: 213–223. https://doi.org/10.1007/s10858-009-9333-z PMID: 19548092