Mapping transiently formed and sparsely populated conformations on a complex energy landscape

Yong Wang, Elena Papaleo and Kresten Lindorff-Larsen[1]

1 Structural Biology and NMR Laboratory, Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5 DK-2200 Copenhagen N, Denmark

August 19, 2016

Abstract

Determining the structures, kinetics, thermodynamics and mechanisms that underlie conformational exchange processes in proteins remains extremely difficult. Only in favourable cases is it possible to provide atomic-level descriptions of sparsely populated and transiently formed alternative conformations. Here we benchmark the ability of enhanced-sampling molecular dynamics simulations to determine the free energy landscape of the L99A cavity mutant of T4 lysozyme. We find that the simulations capture key properties previously measured by NMR relaxation dispersion methods including the structure of a minor conformation, the kinetics and thermodynamics of conformational exchange, and the effect of mutations. We discover a new tunnel that involves the transient exposure towards the solvent of an internal cavity, and show it to be relevant for ligand escape. Together, our results provide a comprehensive view of the structural landscape of a protein, and point forward to studies of conformational exchange in systems that are less characterized experimentally.

Keywords

Free Energy Landscape | Conformational Exchange | Kinetics | Nuclear Magnetic Resonance | Metadynamics

Introduction

Proteins are dynamical entities whose ability to change shape often plays essential roles in function. From an experimental point of view, intra-basin dynamics is often described via conformational ensembles whereas larger scale (and often slower) motions are characterized as conformational exchange between distinct conformational states. The latter are often simplified as a two-site exchange process, $G \rightleftharpoons E$, between a highly populated ground (G) state, and a transiently populated minor (or 'excited', E) state. While the structure of the ground state may often be determined by conventional structural biology tools, it is very difficult to obtain atomic-level insight into minor conformations due to their transient nature and low populations. As these minor conformations may, however, be critical to protein functions, including protein folding, ligand binding, enzyme catalysis, and signal transduction [1, 2, 3] it is important to be able to characterize them in detail. While it may in certain cases be possible to capture sparsely populated conformations in crystals under perturbed experimental conditions, or to examine their structures by analysis of electron density maps [4], NMR spectroscopy provides unique opportunities to study the dynamical equilibrium between major and minor conformations [3, 5] via e.g. chemical-exchange saturation transfer [6], Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion [7], or indirectly via paramagnetic relaxation enhancement [2] or residual dipolar coupling [8] experiments. In favorable cases such experiments can provide not only thermodynamic and kinetic information (i.e. the population of G and E states and the rate of exchange between them), but also structural information in the form of chemical shifts (CS), that can be used to determine the structure of the transiently populated state [5].

Despite the important developments in NMR described above, it remains very difficult to obtain structural models of minor conformations, and a substantial amount of experiments are required. Further, it is generally not possible to use such experiments to infer the mechanisms of interconversion, and to provide a more global description of the multi-state free energy landscape [9, 10]. In the language of energy landscape theory [11], free energy basins and their depths control the population and stability of functionally distinct states, while the relative positions of basins and the inter-basin barrier heights determine the kinetics and mechanism of conformational exchange. As a complement to experiments, such functional landscapes can be explored by *in silico* techniques, such as molecular dynamics (MD) simulations, that may both be used to help interpret experimental data and provide new hypotheses for testing [12, 13]. Nevertheless, the general applicability of simulation methods may be limited by

both the accuracy of the physical models (i.e. force fields) used to describe the free energy landscape and our ability to sample these efficiently by computation. We therefore set out to benchmark the ability of simulations to determine conformational free energy landscapes.

The L99A variant of lysozyme from the T4 bacteriophage (T4L) has proven an excellent model system to understand protein structure and dynamics. Originally designed a 'cavity creating' variant to probe protein stability [14] it was also demonstrated that the large (150 $Å^3$) internal cavity can bind hydrophobic ligands such as benzene [15, 16]. It was early established that the cavity is inaccessible to solvent in the ground state, but that ligand binding is rapid [17], suggesting protein dynamics to play a potential role in the binding process. This posts a long-standing question of how the ligands gain access to the buried cavity [18, 19, 20].

NMR relaxation dispersion measurements of L99A T4L demonstrated that this variant, but not the wild type protein, displayed conformational exchange on the millisecond timescale between the ground state and a minor state populated at around 3% (at room temperature) [1]. Such small populations generally lead only to minimal perturbations of ensemble-averaged experimental quantities making structural studies difficult, and hence it was difficult to probe whether the exchange process indeed allowed for ligand access to the cavity. A series of additional relaxation dispersion experiments, however, made it possible to obtain backbone and side chain CSs of the minor E state of L99A [22, 21]. The backbone CS data were subsequently used as input to a CS-based structure refinement protocol (CS-ROSETTA) to produce a structural model of the E state ($E_{ROSETTA}$; Fig. 1) of the L99A mutant [21]. This model was based in part on the crystal structure of the ground state of L99A (referred to in what follows as G_{Xray}), but perturbing the structure in regions that the experiments demonstrated to undergo conformational change in a way so that the final model ($E_{ROSETTA}$) agrees with experiments. The structure was further validated by creating and solving the structure of a triple mutant variant that inverts the populations of the G and E states. The $E_{ROSETTA}$ structure revealed substantial local rearrangements in T4L L99A, in particular near the cavity which gets filled by the side chain of a phenylalanine at position 114 (F_{114}). Because the cavity is filled and solvent inaccessible in the E-state, the structure did, however, not reveal how ligands might access the cavity.

In an attempt to benchmark the ability of simulations to map conformational free energy landscapes, we have here employed a series of *in silico* experiments designed to probe the structure and dynamics of L99A T4L and have compared the results to NMR measurements. We used enhanced-sampling MD simulations in explicit solvent and with state-of-the-art force fields to map the free-energy landscape including the exchange between the major and minor conformations of the protein. We used a series of recently developed metadynamics methods [23] to sample the conformational exchange process and associated structure and thermodynamics, as well as to determine the kinetics and mechanisms of exchange. We obtained additional insight into the structural dynamics of the E state using simulations that employed the experimental CSs as replica-averaged restraints. Our results provide a coherent picture of the conformational dynamics in L99A and extends recent simulations of a triple mutant of T4L[24], by providing new insights into the mechanisms of exchange and the transient exposure of the internal cavity. Together with previous results for Cyclophilin A [25] the results described here reiterate how simulation methods have now reached a stage where they can be used to study slow, conformational exchange processes such as those probed by NMR relaxation dispersion even in cases where less information is available from experiments.

Results and Discussion

Mapping the free-energy landscape

As the average lifetime of the G and E states are on the order of 20–50ms and 1ms, respectively [1, 22, 21], direct and reversible sampling of the G-E transition at equilibrium would be extremely demanding computationally. Indeed, a recent set of simulations of a triple mutant of T4L, which has a substantially faster kinetics, was able only to observe spontaneous transitions in one direction

[24]. We therefore resorted to a set of flexible and efficient enhanced sampling methods, collectively known as 'metadynamics' [23], that have previously been used in a wide range of applications. In metadynamics simulations, a time-dependent bias is continuously added to the energy surface along a small number of user-defined collective variables (CVs). In this way, sampling is enhanced to reach new regions of conformational space and at the same time allows one to reconstruct the (Boltzmann) free-energy surface. The success of the approach hinges on the ability to find a set of CVs that together describe the slowly varying degrees of freedom and map the important regions of the conformational landscape.

We first performed a set of metadynamics simulations in the well-tempered ensemble [26] using so-called path CVs (S_{path} and Z_{path}) [27, 28] with the aid of recently developed adaptive hills to aid in convergence of the sampling [29, 30] (see details in Appendix and Appendix-Table S1). In short, the S_{path} variable describes the progress of the conformational transition between the G_{Xray} and $E_{ROSETTA}$ structures with additional 'interpolation' using an optimal 'reference' path in a simplified model (see details in Appendix and Figure 2-figure supplement 1), while Z_{path} measures the distance to this reference path. In this way, the two-dimensional free energy landscape along S_{path} and Z_{path} provides a useful description on conformational exchange between ground and excited states that does not assume that the initial reference path describes perfectly the actual path(s) taken.

Projecting the sampled free energy landscape along S_{path} (upper panel of Figure 2) reveals a deep, narrow free energy basin around $S_{path} = 0.2$ (labeled by red sphere and corresponding to the G state), and a broader, shallow free energy basin with S_{path} ranging from 0.6 to 0.8 (labeled by blue sphere and corresponding to the E state). Additional information is obtained from the two-dimensional landscape (shown as a negative free energy landscape of -F(S_{path}, Z_{path}) in the lower panel of Figure 2) which reveals a complex and rough landscape with multiple free energy minima (corresponding to mountains in the negative free energy landscape). Subsequently, structural inspection of these minima identified that the conformations in the basins around $S_{path} = 0.2$ and $S_{path} = 0.75$ to correspond to the structures of G_{Xray} and $E_{ROSETTA}$, respectively.

The broad nature of the free energy landscape in the region of the minor state is consistent with the observation that our MD simulations initiated from $E_{ROSETTA}$ display significant conformational fluctuations (RUN20 and RUN22 in Appendix-Table S1). Furthermore, our metadynamics simulations revealed multiple local free energy minima adjacent to the $E_{ROSETTA}$ basin, together composing a wider basin (highlighted by the black curve in Figure 2). Thus, these simulations suggest that the E state displays substantial conformational dynamics, a result corroborated by simulations that have been biased by the experimental data (see section 'Simulations of the minor state using chemical shift restraints').

In addition to free-energy minima corresponding to the G and E states, we also found a free energy minimum around $S_{path} = 0.36$ and $Z_{path} = 0.05nm^2$ (denoted as $I_{0.36}$ and labeled by yellow sphere in Figure 2) that is located between the G and E states on the one-dimensional free-energy surface. We note, however, that it is difficult to infer dominant reaction pathways from such free energy surfaces, and so from this data alone we cannot determine whether $I_{0.36}$ occurs as an intermediate in G-E conformational transitions. Indeed, it appears from the two-dimensional surface that there exist multiple possible pathways between G and E, as illustrated by grey lines along the mountain ridges of the negative free energy landscape in the lower panel of Figure 2. (We also explored the mechanism of exchange by reconnaissance metadynamics simulations [31], the results of which are described and discussed further below.)

Effect of mutations on the free energy landscape

Based on the encouraging results above for L99A T4L, we examined whether simulations could also capture the effect of mutations of the free energy landscape. Using Rosetta energy calculations on the G_{Xray} and $E_{ROSETTA}$ structures it was previously demonstrated that two additional mutations, G113A and R119P, when introduced into the L99A background, cause an inversion in the populations of the two

states [21, 24]. Indeed, NMR data demonstrated that the triple mutant roughly inverts the populations of the two states so that the minor state structure (of L99A) now dominates (with a 96% population) the triple mutant. We repeated the calculations described above for L99A also for the triple mutant. Remarkably, the free energy profile of the triple mutant obtained using metadynamics simulations reveals a free energy landscape with a dominant minimum around S_{path}=0.7 and a higher energy conformation around S_{path}=0.15 (Figure 2-figure supplement 2). Thus, like our previous observations for a 'state-inverting mutation' in Cyclophilin A [25], we find here that the force field and sampling method are sufficiently accurate to capture the effect of point mutations on the free energy landscape. Further, we note that the barrier height for the conformational exchange in the triple mutant is very similar to the value recently estimated using a completely orthogonal approach [24]. Finally, we attempted to determine the free energy landscape of the L99A,G113A double mutant, which has roughly equal populations of the two states[21], but this simulation did not converge on the simulation timescales at which the two other variants converged.

Calculating conformational free energies

With a free-energy surface in hand and a method to distinguish G- and E-state conformations we calculated the free energy difference, ΔG, between the two conformational states, and compared with the experimental values. We divided the global conformational space into two coarse-grained states by defining the separatrix at $S_{path} = 0.46$ which corresponds to a saddle point on the free energy surface, on the basis of the observations above that the E state is relatively broad. Although a stricter definition of how to divide the reaction coordinate certainly helps the precise calculation, here we just used this simple definition to make an approximate estimation of the free energy difference. Further, since the barrier region is sparsely populated, the exact point of division has only a modest effect on the results. By summing the populations on the two sides of the barrier we calculated ΔG as a function of the simulation time (Figure 3). Initially during the simulations the free energy profile varies substantially (Figure 2) and the free energy difference equally fluctuates. As the simulations converge, however, the free energy difference between the two states stabilize to a value at approximately ΔG=3.5 kcal mol^{-1} (Figure 3, black line). This value can be compared to the value of 2.1 kcal mol^{-1} obtained from NMR relaxation dispersion experiments [1], revealing reasonably good, albeit not exact, agreement with the experiments.

Similar calculations using the simulations of the triple mutant also converge, in this case to about -1.6 kcal mol^{-1} (Figure 3, blue line), in excellent agreement with the experimental measurement (-1.9 kcal mol^{-1}) [21]. Combining these two free energy differences we find that the G113A, R119P mutations cause a shift in the G-E free energy of 5.1 kcal mol^{-1} in simulations compared to 4.0 kcal mol^{-1} obtained by experiments. Thus, we find that the simulations with reasonably high accuracy are able to capture the thermodynamics of the conformational exchange between the two states. While the generality of such observations will need to be established by additional studies we note here that comparably good agreement was obtained when estimating the effect of the S99T mutations in Cyclophilin A [25].

In our previous work on Cyclophilin A [25] we sampled the conformational exchange using parallel-tempering metadynamics simulations [32] using four CVs that we chose to describe the structural differences between the G and E states in that protein. We note here that we also tried a similar approach here but unfortunately failed to observe a complete G-to-E transition, even in a relatively long trajectory of about $1\mu s$ per replica (CVs summarized in Appendix-Table S2, parameters shown in Appendix-Table S1). This negative results is likely due to the CVs chosen did not fully capture the relevant, slowly changing degrees of freedom, thus giving rise insufficient sampling even with the use of a parallel tempering scheme.

Calculating the rates of conformational exchange

Enhanced-sampling simulations such as those described above provide an effective means of mapping the free-energy landscape and hence the structural and thermodynamic aspects of conformational exchange. While the same free-energy landscape also determines the kinetics and mechanisms of exchange it may be more difficult to extract this information from e.g. path-CV-based metadynamics (PathMetaD) simulations. To examine how well simulations can also be used to determine the rates of the G-to-E transitions, quantities that can also be measured by NMR, we used the recently developed 'infrequent metadynamics' method (InMetaD, see details in Appendix) [42, 43, 44, 45]. Briefly described, the approach calculates first-passage times for the conformational change in the presence of a slowly-added bias along a few CVs, here chosen as the path CVs also used to map the landscape. By adding the bias slowly (and with lower amplitude) we aim to avoid biasing the transition-state region and hence to increase the rate only by lowering the barrier height; in this way it is possible to correct the first-passage times for the bias introduced.

Using this approach on L99A T4L we collected 42 and 36 independent trajectories with state-to-state transition starting from either the G state or E state, respectively (Appendix-Figure S1 and S2). The (unbiased) rates that we calculated (Table 1 and Appendix-Figure S3) are in good agreement with the experimental rates [1, 21] (within a factor of 10), corresponding to an average error of the barrier height of $\sim$1 kcal mol^{-1}. We also performed similar calculations for the 'population-inverting' triple mutant, where we collected 30 transitions (15 for each direction) using InMetaD simulations. As for L99A, we also here find similarly good agreement with experimental measurements[24] (Table 1 and Appendix-Figure S4). We estimated the reliability of this computational approach using a Kolmogorov-Smirnov test to examine whether the first-passage times conform to the expected Poisson distribution [43], and indeed the results of this analysis suggest good agreement (Table 1-Appendix-Figure S5 and S6).

The ability to calculate forward and backward rates between G and E provided us with an alternative and independent means to estimate the free energy difference between the two states (Table 1), and to test the two-state assumption used in the analysis of the experimental NMR data. We therefore calculated the free energy difference from the ratio of the forward and backward reaction rates. The values obtained (2.9$\pm$0.5 kcal mol^{-1} and -1.2$\pm$1.1 kcal mol^{-1} for L99A and the triple mutant, respectively) are close both to the values obtained above from the equilibrium free energy landscape (3.5 kcal mol^{-1} and -1.6 kcal mol^{-1}) and experiment (2.1 kcal mol^{-1} and -1.9 kcal mol^{-1}). In particular, the relatively close agreement between the two independent computational estimates lends credibility both to the free energy landscape and the approach used to estimate the kinetics. The observation that both values for L99A are slightly larger than the experimental number suggests that this discrepancy (ca. 1 kcal mol^{-1}) can likely be explained by remaining force field deficiencies rather than lack of convergence or the computational approach used.

Simulations of the minor state using chemical shift restraints

While the simulations described above used available structural information of G and E states to guide and enhance conformational sampling, the resulting free energy surfaces represent the Boltzmann distributions of the force field and are not otherwise biased by experimental data. To further refine the structural model of the E state we used the relaxation-dispersion derived CSs that were used to determine of $E_{ROSETTA}$ (BMRB [33] entry 17604) as input to restrained MD simulations. In these simulations, we used the experimental data as a system-specific force field correction to derive an ensemble of conformations that is compatible both with the force field and the CSs. Such replica-averaged simulations use the experimental data in a minimally-biased way that is consistent with the Principle of Maximum Entropy [34, 35, 36, 37].

We performed CS-restrained MD simulations of the E state of L99A averaging the CSs over four replicas. Although the number of replicas is a free parameter, which should in principle be chosen as large as possible, it has been demonstrated that four replicas are sufficient to reproduce the structural

heterogeneity accurately [38] without excessive computational requirements. The agreement between calculated and experimental CSs was quantified by the root-mean-square deviation between the two (Figure 4-figure supplement 1). In particular, it is important not to bias the agreement beyond what can be expected based on the inherent accuracy of the CS prediction methods (we assumed that the error in the experimental CS measurement even for the E state is negligible in comparison). Thus, we compared the experimental CS values of the minor state with the values calculated using the $E_{ROSETTA}$ structure as input to CamShift[39], Sparta+[40] and ShiftX[41] (Figure 4-figure supplement 2). The average RMSDs for five measured nuclei (H_α, H_N, N, C' and C_α) are 0.2, 0.4, 2.0, 0.8 and 1.1ppm, respectively (Appendix-Table S3), which are close to the inherent uncertainty of the CS back-calculation, indicating that the level of agreement enforced is reasonable.

To compare the results of these experimentally-biased simulations with the experimentally-unbiased simulations described above, we projected the CS-restrained MD trajectories onto either one (Figure 4) or both (Figure 4-figure supplement 3) of the S_{path} and Z_{path} variables used in the path-variable-driven simulations (PathMetaD). The distribution of conformations obtained using the E-state CSs as restraints is in good agreement with the broad free energy profile of the E-state obtained in the metadynamics simulations that did not include any experimental restraints. To ensure that this observation is not merely an artifact of both simulations using the same force field (CHARMM22*), we repeated the biased simulations using the Amber ff99SB*-ILDN force field and obtained comparable results. We also verified that the conclusions obtained are reasonably robust to other variables such as the number of replicas and the strength of restraints (Figure 4-figure supplement 4).

As a final and independent test of the structural ensemble of the minor conformation of L99A we used the ground state CSs of the triple mutant (BMRB entry 17603), which corresponds structurally to the E state of L99A, as restraints in replica-averaged CS-biased simulations (Figure 4-figure supplement 5). Although not fully converged, these simulations also cover roughly the same region of conformational space when projected along S_{path} (Figure 4).

Thus, together our different simulations, which employ different force fields, are either unbiased or biased by experimental data, and use either dispersion-derived (L99A) or directly obtained (triple mutant) CS all provide a consistent view of the minor E-state conformation of L99A. We also note that the CS-derived ensembles of the E-state support the way we divided the G- and E-states when calculating conformational free energy differences between the two states.

Mechanisms of conformational exchange

Having validated that our simulations can provide a relatively accurate description of the structure, thermodynamics and kinetics of conformational exchange we proceeded to explore the molecular mechanism of the G-to-E transitions. We used the recently developed reconnaissance metadynamics approach [46], that was specifically designed to enhance sampling of complicated conformational transitions and has been employed to explore the conformational dynamics of complex systems [31, 47].

We performed three independent reconnaissance metadynamics simulations of L99A starting from the G state (summarized in Appendix-Table S1) using the same geometry-based CVs that we also used in the parallel-tempering simulations described above. We observed complete conformational transitions from the G to E state in the reconnaissance simulations in as little as tens of nanoseconds of simulations (Figure 5-figure supplement 1) — at least 1–2 orders of magnitude faster than standard metadynamics. These G-to-E and E-to-G transitions, although biased by the CVs, provide insight into the potential mechanisms of exchange. To ease comparison with the equilibrium sampling of the free energy landscape we projected these transitions onto the free energy surface F(S_{path},Z_{path}) (Figure 5). The results reveal multiple possible routes connecting the G and E states, consistent with the multiple gullies found on the free energy surface (Figure 2). The trajectories also suggested that the G-to-E interconversion can either take place directly without passing the $I_{0.36}$ state or indirectly via it.

In the context of coarse-grained kinetic models the results above would suggest at least two possible

mechanisms operate in parallel: $G \rightleftharpoons E$ or $G \rightleftharpoons I_{0.36} \rightleftharpoons E$. Further inspection of the structures along these different kinetics routes (see the trajectories of other order parameters in Figure 5-figure supplement 2 and Videos 1-4) suggested an interesting distinction between the two. In the $G \rightleftharpoons I_{0.36} \rightleftharpoons E$ route the side chain of F114, which occupies the cavity in the E state, gets transiently exposed to solvent during the transition, whereas in the direct $G \rightleftharpoons E$ transitions F114 can rotate its side chain inside the protein core (see also the solvent accessible surface area calculation of F114 in Figure 5-figure supplement 3).

A potential pathway for ligand binding and escape

As the internal cavity in L99A T4L remains buried in both the G and E states (and indeed occupied by F114 in the E state) it remains unclear how ligands access this internal cavity and how rapid binding and release is achieved. Visual inspection of our trajectories and solvent-accessible surface area analysis revealed structures with transient exposure of the internal cavity towards the solvent. The structures were mostly found in a region of conformational space that mapped onto the $I_{0.36}$ basin (Figure 2), and the events of that basin mostly took place between 430ns and 447ns (see Video 5). Thus, we mapped these structures to the free energy surface (Figure 6-figure supplement 1) and analysed them. Overall, the structure is more similar to the G- than E-state, though is more loosely packed. The similarity to the G-state is compatible with rapid binding and position of F114 in this state.

We used CAVER3[48] (see parameters in Appendix-Table S4) to analyse the structures and found multiple tunnels connecting the cavity with protein surface (Figure 6-figure supplement 1 and 2). The tunnels are relatively narrow with the typical radius of the bottleneck (defined as the narrowest part of a given tunnel) between $\sim$ 1Å – $\sim$ 2Å. We used CAVER Analyst1.0 [49] (see details in Appendix and parameters in Appendix-Table S4) to separate the tunnels into different clusters (Figure 6-figure supplement 3 and Appendix-Table S5) with the dominant cluster (denoted tunnel#1) having a entrance located at the groove between H_F and H_I. A typical representative structure of $I_{0.36}$ is shown in Figure 6A. The radii along the structures in cluster #1 vary, but share an overall shape (Figure 6-figure supplement 1), and we find that the maximal bottleneck radius is $\sim$ 2.5 Å, the average bottleneck radius is $\sim$ 1.3 Å, and the average length $\sim$ 11.2 Å.

Interestingly, a series of structures of L99A were recently described, in which the internal cavity where filled with eight congeneric ligands of increasing size to eventually open the structure size[20]. We performed a comparable tunnel analysis on those eight ligand-bound structures (PDB ID codes: 4W52 – 4W59), revealing the maximal bottleneck radius of 1.8 Å (bound with n-hexylbenzene, 4W59). Although the size of the tunnel in these X-ray structures is slightly smaller than that in $I_{0.36}$ structures, the location of the tunnel exit is consistent with the dominant tunnel#1 in $I_{0.36}$ (Figure 6-figure supplement 3). We note, however, that the tunnels observed in our simulation and in the ligand-induced cavity-open X-ray structure (4W59), are too narrow to allow for unhindered passage of e.g. benzene with its a van der Waals' width of 3.5 Å [15]. Thus, we speculate that the transient exposure in $I_{0.36}$ might serve as a possible starting point for ligand (un)binding, which would induce [50, 19, 10] further the opening of the tunnel.

As an initial step towards characterizing the mechanism of ligand binding and escape we used adiabatic biased molecular dynamics (ABMD) simulations [51, 52] to study the mechanism of how benzene escapes the internal cavity (see Appendix for details). In ABMD the system is perturbed by a 'ratcheting potential', which acts to 'select' spontaneous fluctuations towards the ligand-free state. In particular, the biasing potential is zero when the reaction coordinate (here chosen to be the RMSD of the ligand to the cavity-bound state) increases, but provides a penalty for fluctuations that brings the ligand closer to the cavity. In this way, we were able to observe multiple unbinding events in simulations despite the long lifetime (1.2 ms) of the ligand in the cavity. Most of trajectories (15 of the 20 events observed) reveal that benzene escapes from the cavity by following tunnel #1 (Figure 6-figure supplement 4 and Appendix-Table S6). A typical unbinding path is shown in the right panel of Fig. 6 (see also Video 6). Because the ABMD introduces a bias to speed up ligand escape, we ensured that the observed pathway

was the same at two different values of the biasing force constants (Figure 6-figure supplement 4 and Appendix-Table S6). Future work will be aimed to perform a more quantitative analysis of the ligand binding and unbinding kinetics.

Conclusions

The ability to change shape is an essential part of the function of many proteins, but it remains difficult to characterize alternative conformations that are only transiently and sparsely populated. We have studied the L99A variant of T4L as a model system that displays a complicated set of dynamical processes which have been characterized in substantial detail. Our results show that modern simulation methods are able to provide insight into such processes, paving the way for future studies for systems that are more difficult to study experimentally.

Using a novel method for defining an initial reference path between two conformations, we were able to sample the free energy landscape described by an accurate molecular force field. In accordance with experiments, the simulations revealed two distinct free energy basins that correspond to the major and minor states found by NMR. Quantification of the free energy difference between the two states demonstrated that the force field is able to describe conformational free energies to an accuracy of about 1 kcal mol^{-1}. This high accuracy is corroborated by previous studies of a different protein, Cyclophilin A, where we also calculated conformational free energies and compared to relaxation dispersion experiments and found very good agreement. For both proteins we were also able to capture and quantify the effect that point mutations have on the equilibrium between the two states, and also here found good agreement with experiments. We note, however, that comparable simulations of the L99A/G113A mutant did not reach convergence.

Moving a step further, we here also calculated the kinetics of conformational exchange using a recently developed metadynamics method. For both the L99A variant and a population-inverting triple mutant we find that the calculated reaction rates are in remarkably good agreement with experiments. The ability to calculate both forward and backward rates provided us with the opportunity to obtain an independent estimate the calculated free energy difference. The finding that the free energy differences estimated in this way (for both L99A and the triple mutant) are close to those estimated from the free energy landscape provides an important validation of both approaches, and we suggest that, when possible, such calculations could be used to supplement conventional free energy estimates.

The free-energy landscape suggested that the E state is relatively broad and contains a wider range of conformations. To validate this observation we used the same chemical shift information as was used as input to Rosetta and performed replica-averaged CS-restrained simulations. The resulting ensemble demonstrates that the experiments and force field, when used jointly, indeed are compatible with a broader E state. Thus, we suggest that the $E_{ROSETTA}$ structure and CS-restrained ensemble jointly describe the structure and dynamics of the E state.

While NMR experiments, in favourable cases, can be used to determine the structure, thermodynamics and kinetics of conformational exchange, a detailed description mechanism of interconversion remains very difficult to probe by experiments. We explored potential mechanisms of conformational exchange between the two states, finding at least two distinct routes. One route involved a direct transition with the central F114 entering the cavity within the protein, whereas a different possible mechanism involves transient partial-loosening of the protein. In both cases, the mechanism differ from the reference path that we used as a guide to map the free energy landscape, suggesting that high accuracy of the initial guess for a pathway is not absolutely required in the metadynamics simulations, suggesting also the more general applicability of the approach.

Finally, we observed a set of conformations with a transiently opened tunnel that leads from the exterior of the protein to the internal cavity, that is similar to a recently discovered path that is exposed when the cavity is filled by ligands of increasing size. The fact that such a tunnel can be explored even in

the absence of ligands suggests that intrinsic protein motions may play an important role in ligand binding, and indeed we observed this path to be dominant in simulations of ligand unbinding.

In total, we present a global view of the many, sometimes coupled, dynamical processes present in a protein. Comparison with a range of experimental observations suggests that the simulations provide a relatively accurate description of the protein, demonstrating how NMR experiments can be used to benchmark quantitatively the ability of simulations to study conformational exchange. We envisage that future studies of this kind, also when less is known about the structure of the alternative states, will help pave the way for using simulations to study the structural dynamics of proteins and how this relates to function.

Materials and methods

System preparation

Our simulations were initiated in the experimentally determined structures of the ground state of L99A (G_{Xray}; PDB ID code 3DMV) or minor, E state ($E_{ROSETTA}$; 2LCB). The structure of the ground state of the L99A, G113A, R119P triple mutant, corresponding to the E state of L99A was taken from PDB entry 2LC9 ($G^{Triple}_{ROSETTA}$). Details can be found in Appendix Materials and Methods.

Initial reaction path

Taking G_{Xray} and $E_{ROSETTA}$ as the models of the initial and final structures, we calculated an initial reaction path between them with the MOIL software [53], which has been used to explore the mechanism of conformational change of proteins [54]. Further details can be found in the Appendix and in refs. [55, 54].

Path CV driven metadynamics simulations with adaptive hills

The adaptive-hill version of metadynamics updates the Gaussian width on the fly according to local properties of the underlying free-energy surface on the basis of local diffusivity of the CVs or the local geometrical properties. Here, we used the former strategy. Simulation were performed using Gromacs4.6[56] with the PLUMED2.1 plugin[57]. See parameter details in Appendix-Table S1.

Replica-averaged CS-restrained simulations

We performed replica-averaged CS restrained MD simulations by using GPU version of Gromacs5 with the PLUMED2.1 and ALMOST2.1 [58] plugins. Equilibrated structures of $E_{ROSETTA}$ and $G^{Triple}_{ROSETTA}$ were used as the starting conformations. CS data of $E_{ROSETTA}$ and $G^{Triple}_{ROSETTA}$ were obtained from the BMRB database [33] as entries 17604 and 17603, respectively.

Reconnaissance metadynamics simulations

Reconnaissance metadynamics [46] uses a combination of a machine learning technique to automatically identify the locations of free energy minima by periodically clustering the trajectory and a dimensional reduction technique that can reduce the landscape complexity. We performed several reconnaissance metadynamics simulations with different combinations of CVs starting from G_{Xray} using Gromacs4.5.5 with PLUMED1.3 plugin. See parameter details on Appendix-Table S1.

Calculating kinetics using infrequent metadynamics

The key idea of infrequent metadynamics is to bias the system with a frequency slower than the barrier crossing time but faster than the slow inter-basin relaxation time, so that the transition state region has a low risk of being substantially biased. As the first transition times should obey Poisson statistics, the reliability of the kinetics estimated from InMetaD can be assessed by a statistical analysis based on the Kolmogorov-Smirnov (KS) test [43]. See parameter details on Appendix and Appendix-Table S1.

Acknowledgments

We would like to thank Pratyush Tiwary and Matteo Salvalaglio for the help on the usage and analysis of InMetaD, Gareth Tribello for the help on the usage of reconnaissance metadynamics, Pengfei Tian, Shuguang Yuan, Micha Ben Achim Kunze for their helpful discussions and comments. Ludovico Sutto and Francesco Gervasio are thanked for input on metadynamics simulations.

References

[1] Mulder, F. A, Mittermaier, A, Hon, B, Dahlquist, F. W, & Kay, L. E. (2001) Studying excited states of proteins by NMR spectroscopy. *Nat Struct Mol Biol* **8**, 932–935.

[2] Tang, C, Schwieters, C. D, & Clore, G. M. (2007) Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* **449**, 1078–1082.

[3] Baldwin, A. J & Kay, L. E. (2009) NMR spectroscopy brings invisible protein states into focus. *Nat Chem Biol* **5**, 808–814.

[4] Fraser, J. S, Clarkson, M. W, Degnan, S. C, Erion, R, Kern, D, & Alber, T. (2009) Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462**, 669–673.

[5] Sekhar, A & Kay, L. E. (2013) Nmr paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc Natl Acad Sci USA* **110**, 12867–12874.

[6] Vallurupalli, P, Bouvignies, G, & Kay, L. E. (2012) Studying "invisible" excited protein states in slow exchange with a major state conformation. *J Am Chem Soc* **134**, 8148–8161.

[7] Hansen, D. F, Vallurupalli, P, & Kay, L. E. (2008) Using relaxation dispersion nmr spectroscopy to determine structures of excited, invisible protein states. *J Biomol NMR* **41**, 113–120.

[8] Lukin, J. A, Kontaxis, G, Simplaceanu, V, Yuan, Y, Bax, A, & Ho, C. (2003) Quaternary structure of hemoglobin in solution. *Proc Nat Acad Sci* **100**, 517–520.

[9] Zhuravlev, P. I & Papoian, G. A. (2010) Protein functional landscapes, dynamics, allostery: a tortuous path towards a universal theoretical framework. *Q Rev Biophys* **43**, 295–332.

[10] Wang, Y, Tang, C, Wang, E, & Wang, J. (2012) Exploration of multi-state conformational dynamics and underlying global functional landscape of maltose binding protein. *PLoS Comput Biol* **8**, e1002471.

[11] Onuchic, J. N, Luthey-Schulten, Z, & Wolynes, P. G. (1997) Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* **48**, 545–600.

[12] Karplus, M & Lavery, R. (2014) Significance of molecular dynamics simulations for life sciences. *Israel J Chem* **54**, 1042–1051.

[13] Eaton, W. A & Muñoz, V. (2014) Impact of atomistic molecular dynamics simulations on understanding how proteins fold: an experimentalist's perspective. *Bioinformatics*.

[14] Eriksson, A, Baase, W. A, Zhang, X, Heinz, D, Blaber, M, Baldwin, E, & Matthews, B. (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* **255**, 178–183.

[15] Eriksson, A. E, Baase, W. A, Wozniak, J. A, & Matthews, B. W. (1992) A cavity-containing mutant of T4 lysozyme is stabilized by buried. *Nature* **355**, 371–373.

[16] Liu, L, Baase, W. A, & Matthews, B. W. (2009) Halogenated benzenes bound within a non-polar cavity in T4 lysozyme provide examples of I-S and I-Se halogen-bonding. *J Mol Biol* **385**, 595–605.

[17] Feher, V. A, Baldwin, E. P, & Dahlquist, F. W. (1996) Access of ligands to cavities within the core of a protein is rapid. *Nat Struct Biol* **3**.

[18] Mulder, F. A, Hon, B, Muhandiram, D. R, Dahlquist, F. W, & Kay, L. E. (2000) Flexibility and ligand exchange in a buried cavity mutant of T4 lysozyme studied by multinuclear NMR. *Biochemistry* **39**, 12614–12622.

[19] López, C. J, Yang, Z, Altenbach, C, & Hubbell, W. L. (2013) Conformational selection and adaptation to ligand binding in T4 lysozyme cavity mutants. *Proc Natl Acad Sci USA* **110**, E4306–E4315.

[20] Merski, M, Fischer, M, Balius, T. E, Eidam, O, & Shoichet, B. K. (2015) Homologous ligands accommodated by discrete conformations of a buried cavity. *Proc Natl Acad Sci USA*.

[21] Bouvignies, G, Vallurupalli, P, Hansen, D. F, Correia, B. E, Lange, O, Bah, A, Vernon, R. M, Dahlquist, F. W, Baker, D, & Kay, L. E. (2011) Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* **477**, 111–114.

[22] Mulder, F. A, Hon, B, Mittermaier, A, Dahlquist, F. W, & Kay, L. E. (2002) Slow internal dynamics in proteins: application of nmr relaxation dispersion spectroscopy to methyl groups in a cavity mutant of T4 lysozyme. *J Am Chem Soc* **124**, 1443–1451.

[23] Laio, A & Parrinello, M. (2002) Escaping free-energy minima. *Proc Natl Acad Sci USA* **99**, 12562–12566.

[24] Vallurupalli, P, Chakrabarti, N, Pomes, R, & Kay, L. E. (2016) Atomistic picture of conformational exchange in a T4 lysozyme cavity mutant: an experiment-guided molecular dynamics study. *Chem. Sci.* **7**, 3602–3613.

[25] Papaleo, E, Sutto, L, Gervasio, F. L, & Lindorff-Larsen, K. (2014) Conformational changes and free energies in a proline isomerase. *J Chem Theo Comput* **10**, 4169–4174.

[26] Barducci, A, Bussi, G, & Parrinello, M. (2008) Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys Rev Lett* **100**, 020603.

[27] Branduardi, D, Gervasio, F. L, & Parrinello, M. (2007) From A to B in free energy space. *J Chem Phys* **126**, 054103.

[28] Saladino, G, Gauthier, L, Bianciotto, M, & Gervasio, F. (2012) Assessing the performance of metadynamics and path variables in predicting the binding free energies of p38 inhibitors. *J Chem Theo Comput* **8**, 1165–1170.

[29] Branduardi, D, Bussi, G, & Parrinello, M. (2012) Metadynamics with adaptive gaussians. *J Chem Theo Comput* **8**, 2247–2254.

[30] Dama, J. F, Parrinello, M, & Voth, G. A. (2014) Well-tempered metadynamics converges asymptotically. *Phys Rev Lett* **112**, 240602.

[31] Tribello, G. A, Cuny, J, Eshet, H, & Parrinello, M. (2011) Exploring the free energy surfaces of clusters using reconnaissance metadynamics. *J Chem Phys* **135**, 114109.

[32] Bonomi, M & Parrinello, M. (2010) Enhanced sampling in the well-tempered ensemble. *Phys Rev Lett* **104**, 190601.

[33] Ulrich, E. L, Akutsu, H, Doreleijers, J. F, Harano, Y, Ioannidis, Y. E, Lin, J, Livny, M, Mading, S, Maziuk, D, Miller, Z, et al. (2008) Biomagresbank. *Nucleic Acids Res* **36**, D402–D408.

[34] Pitera, J. W & Chodera, J. D. (2012) On the use of experimental observations to bias simulated ensembles. *Journal of Chemical Theory and Computation* **8**, 3445–3451.

[35] Roux, B & Weare, J. (2013) On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *J Chem Phys* **138**, 084107.

[36] Cavalli, A, Camilloni, C, & Vendruscolo, M. (2013) Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *J Chem Phys* **138**, 094112.

[37] Boomsma, W, Ferkinghoff-Borg, J, & Lindorff-Larsen, K. (2014) Combining experiments and simulations using the maximum entropy principle. *PLoS Comput Biol* **10**, e1003406.

[38] Camilloni, C, Cavalli, A, & Vendruscolo, M. (2013) Replica-averaged metadynamics. *J Chem Theo Comput* **9**, 5610–5617.

[39] Kohlhoff, K. J, Robustelli, P, Cavalli, A, Salvatella, X, & Vendruscolo, M. (2009) Fast and accurate predictions of protein nmr chemical shifts from interatomic distances. *J Am Chem Soc* **131**, 13894–13895.

[40] Shen, Y & Bax, A. (2010) Sparta+: a modest improvement in empirical nmr chemical shift prediction by means of an artificial neural network. *J Biomol NMR* **48**, 13–22.

[41] Neal, S, Nip, A. M, Zhang, H, & Wishart, D. S. (2003) Rapid and accurate calculation of protein 1h, 13c and 15n chemical shifts. *J Biomol NMR* **26**, 215–240.

[42] Tiwary, P & Parrinello, M. (2013) From metadynamics to dynamics. *Phys Rev Lett* **111**, 230602.

[43] Salvalaglio, M, Tiwary, P, & Parrinello, M. (2014) Assessing the reliability of the dynamics reconstructed from metadynamics. *J Chem Theo Comput* **10**, 1420–1425.

[44] Tiwary, P, Limongelli, V, Salvalaglio, M, & Parrinello, M. (2015) Kinetics of protein–ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proc Natl Acad Sci USA* **112**, E386–E391.

[45] Tiwary, P, Mondal, J, Morrone, J. A, & Berne, B. (2015) Role of water and steric constraints in the kinetics of cavity–ligand unbinding. *Proc Natl Acad Sci USA* p. 201516652.

[46] Tribello, G. A, Ceriotti, M, & Parrinello, M. (2010) A self-learning algorithm for biased molecular dynamics. *Proc Natl Acad Sci USA* **107**, 17509–17514.

[47] Söderhjelm, P, Tribello, G. A, & Parrinello, M. (2012) Locating binding poses in protein-ligand systems using reconnaissance metadynamics. *Proc Natl Acad Sci USA* **109**, 5170–5175.

[48] Chovancova, E, Pavelka, A, Benes, P, Strnad, O, Brezovsky, J, Kozlikova, B, Gora, A, Sustr, V, Klvana, M, Medek, P, et al. (2012) Caver 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput Biol* **8**, e1002708.

[49] Kozlikova, B, Sebestova, E, Sustr, V, Brezovsky, J, Strnad, O, Daniel, L, Bednar, D, Pavelka, A, Manak, M, Bezdeka, M, et al. (2014) Caver analyst 1.0: Graphic tool for interactive visualization and analysis of tunnels and channels in protein structures. *Bioinformatics* p. btu364.

[50] Koshland Jr, D. (1958) Application of a theory of enzyme specificity to protein synthesis. *Proc Natl Acad Sci USA* **44**, 98.

[51] Marchi, M & Ballone, P. (1999) Adiabatic bias molecular dynamics: A method to navigate the conformational space of complex molecular systems. *J Chem Phys* **110**, 3697–3702.

[52] Paci, E & Karplus, M. (1999) Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. *J Mol Bio* **288**, 441 – 459.

[53] Elber, R, Roitberg, A, Simmerling, C, Goldstein, R, Li, H. Y, Verkhivker, G, Keasar, C, Zhang, J, & Ulitsky, A. (1995) Moil: A program for simulations of macromolecules. *Comput Phys Commun* **91**, 159–189.

[54] Wang, J, Wang, Y, Chu, X, Hagen, S. J, Han, W, & Wang, E. (2011) Multi-scaled explorations of binding-induced folding of intrinsically disordered protein inhibitor IA3 to its target enzyme. *PLoS Comput Biol* **7**, e1001118.

[55] Majek, P, Elber, R, & Weinstein, H. (2008) Pathways of conformational transitions in proteins in coarse-graining of condensed phase and biomolecular systems. *Voth GA, ed. Boca Raton, FL: CRC Press* pp. 185–203.

[56] 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.

[57] Tribello, G. A, Bonomi, M, Branduardi, D, Camilloni, C, & Bussi, G. (2014) Plumed 2: New feathers for an old bird. *Comput Phys Commun* **185**, 604–613.

[58] Fu, B, Sahakyan, A. B, Camilloni, C, Tartaglia, G. G, Paci, E, Caflisch, A, Vendruscolo, M, & Cavalli, A. (2014) Almost: An all atom molecular simulation toolkit for protein structure determination. *J Comput Chem* **35**, 1101–1105.

Impact statement

Transiently and sparsely populated conformations of proteins are invisible to standard methods of structural biology, yet often play functional roles. We show how molecular simulations can capture key aspects of the conformational exchange in a protein.

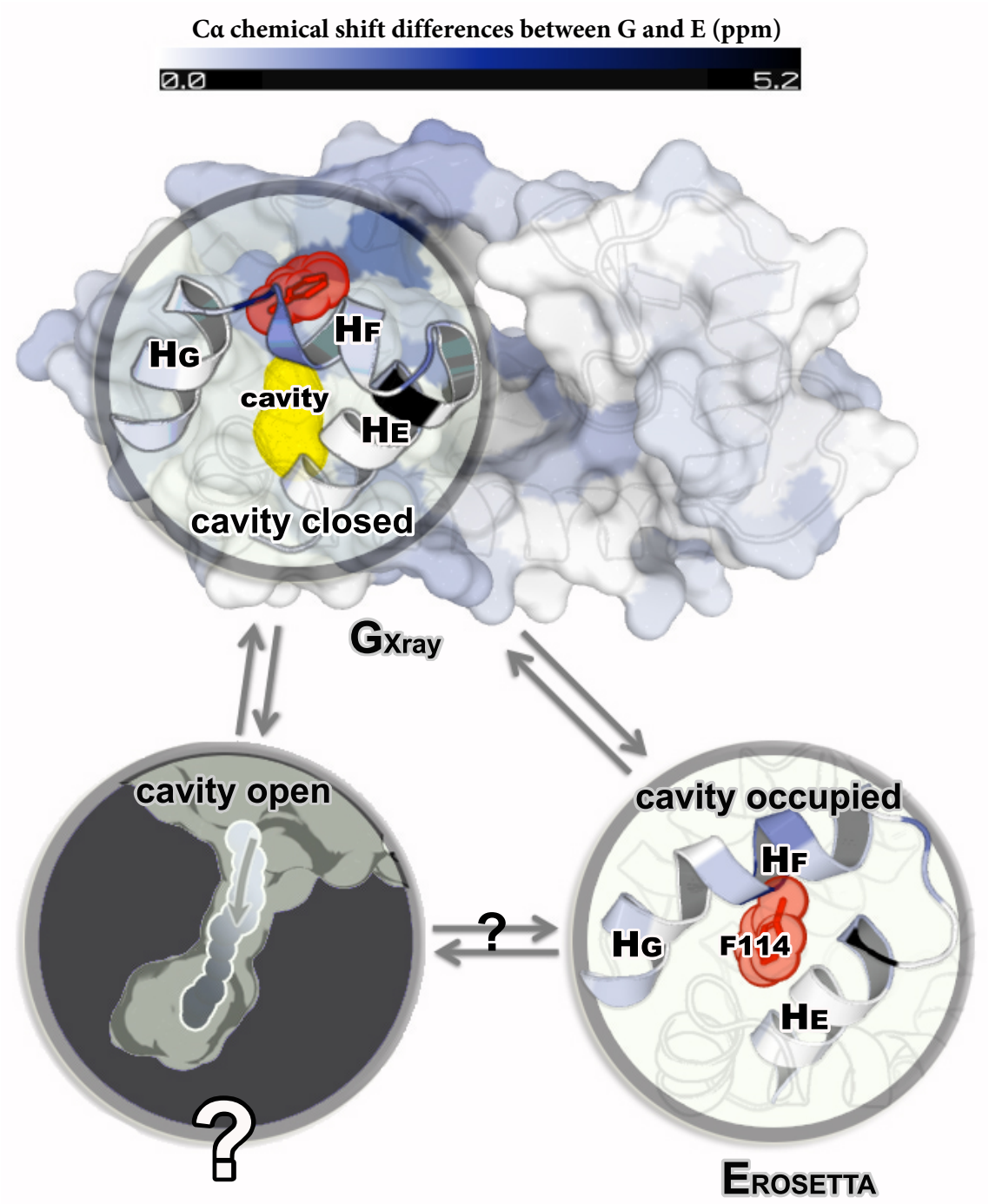


FIGURE 1: **Structures of the major G and minor E states of L99A T4L and the hidden state hypothesis.** The X-ray structure of the G state (G_{Xray}; PDB ID code 3DMV) has a large internal cavity of within the core of the C-terminal domain that is able to bind hydrophobic ligands. The structure of the E state ($E_{ROSETTA}$; PDB ID code 2LC9) was previously determined by CS-ROSETTA using chemical shifts. The G and E states are overall similar, apart from the region surrounding the internal cavity. Comparison of the two structures revealed two remarkable conformational changes from G to E: helix F (denoted as H_F) rotates and fuses with helix G (H_G) into a longer helix, and the side chain of phenylalanine at position 114 (F_{114}) rotates so as to occupy part of the cavity. As the cavity is inaccessible in both the G_{Xray} and $E_{ROSETTA}$ structures it has been hypothesized that ligand entry occurs via a third 'cavity open' state [21, 20].

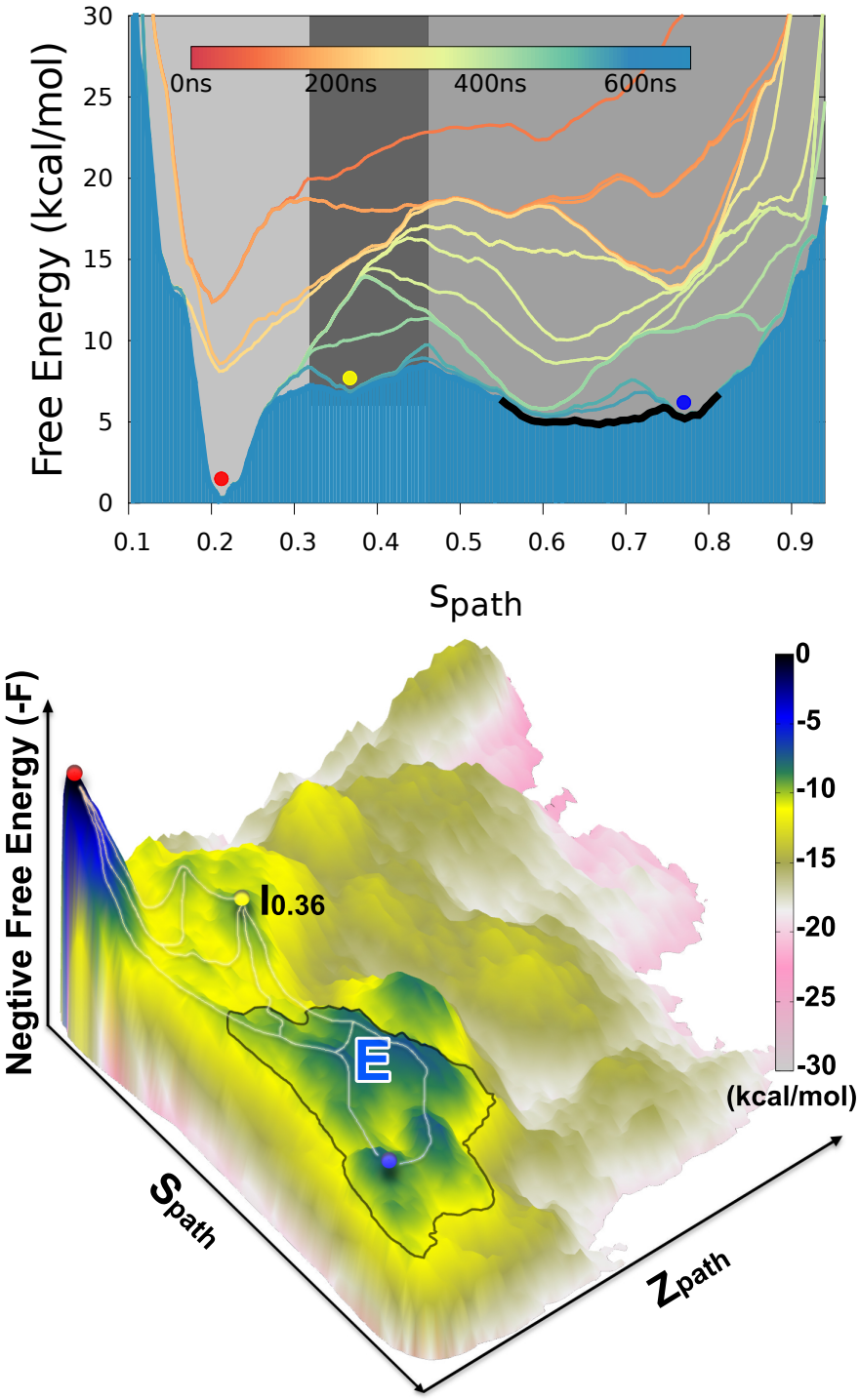


FIGURE 2: **Free energy landscape of the L99A variant of T4L.** In the upper panel, we show the projection of the free energy along S_{path}, representing the Boltzmann distribution of the force field employed along the the progression along of the reference path. Differently colored lines represent the free energy profiles obtained at different stages of the simulation, whose total length was 667ns. As the simulation progressed we rapidly found two distinct free energy basins, and the free energy profile was essentially constant during the last 100 ns of the simulation. Free energy basins around $S_{path} = 0.2$ and $S_{path} = 0.75$ correspond to the previously determined structures of the G- and E-state, respectively (labelled by red and blue dots, respectively). As discussed further below, the E-state is relatively broad and is here indicated by the thick, dark line with S_{path} ranging from 0.55 to 0.83. In the lower panel, we show the three-dimensional negative free energy landscape, -F(S_{path}, Z_{path}), that reveals a more complex and rough landscape with multiple free energy minima, corresponding to mountains in the negative free energy landscape. An intermediate-state basin around $S_{path} = 0.36$ and $Z_{path} = 0.05nm^2$, which we denote $I_{0.36}$, is labeled by a yellow dot.

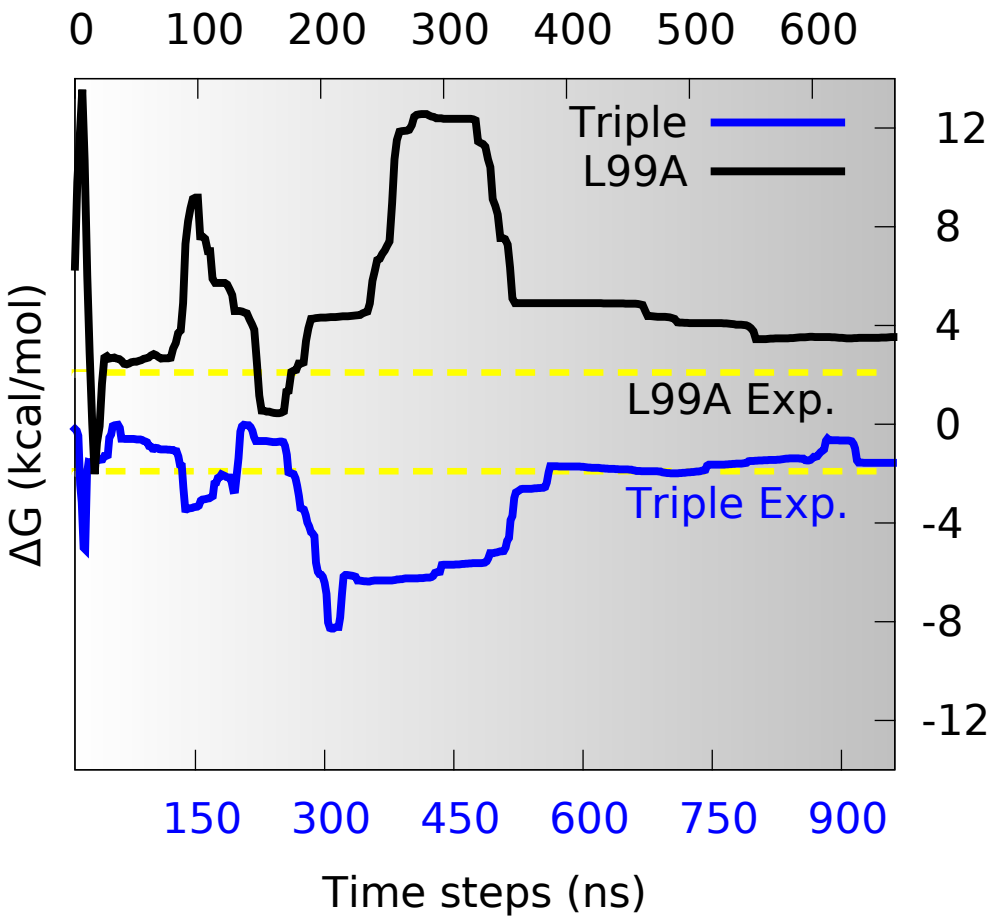


FIGURE 3: **Estimation of free energy differences and comparison with experimental measurements.** We divided the global conformational space into two coarse-grained states by defining the separatrix at $S_{path} = 0.46$ (0.48 for the triple mutant) in the free energy profile (Figure 2-figure supplement 2) which corresponds to a saddle point of the free energy surface, and then estimated the free energy differences between the two states (ΔG) from their populations. The time evolution of ΔG of L99A (upper time axis) and the triple mutant (lower axis) are shown as black and blue curves, respectively. The experimentally determined values (2.1 kcal mol^{-1} for L99A and -1.9 kcal mol^{-1} for the triple mutant) are shown as yellow dashed lines.

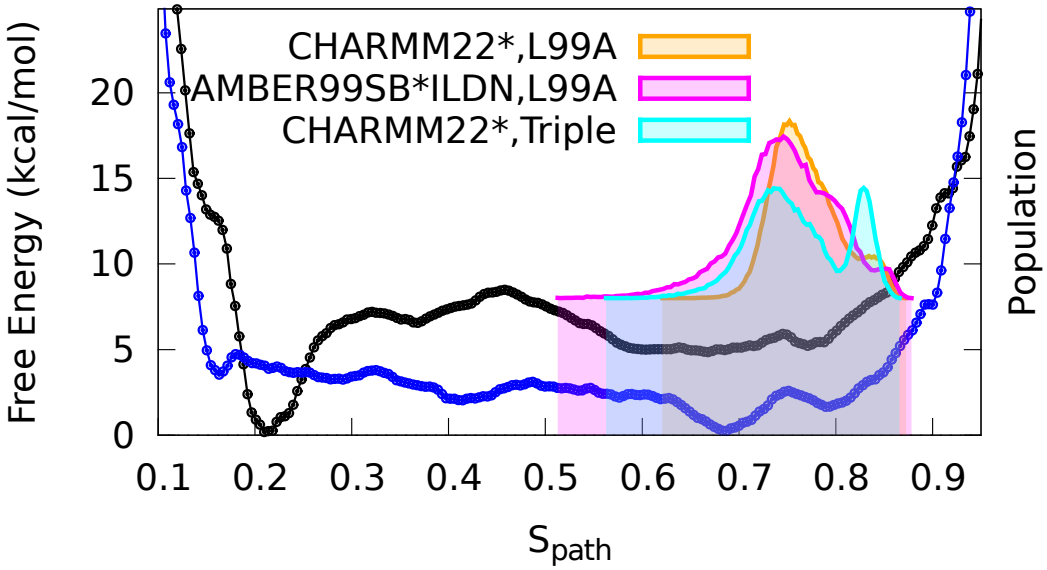


FIGURE 4: **Conformational ensemble of the minor state as determined by CS biased, replica-averaged simulations.** We determined an ensemble of conformations corresponding to the E-state of L99A T4L using replica-averaged CSs as a bias term in our simulations. The distribution of conformations was projected onto the S_{path} variable (orange) and is compared to the free energy profile obtained above from the metadynamics simulations without experimental biases (black line). To ensure that the similar distribution of conformations is not an artifact of using the same force field (CHARMM22*) in both simulations, we repeated the CS-biased simulations using also the Amber ff99SB*-ILDN force field (magenta) and obtained similar results. Finally, we used the ground state CSs of a triple mutant of T4L, which was designed to sample the minor conformation (of L99A) as its major conformation, and also obtained a similar distribution along the S_{path} variable (cyan).

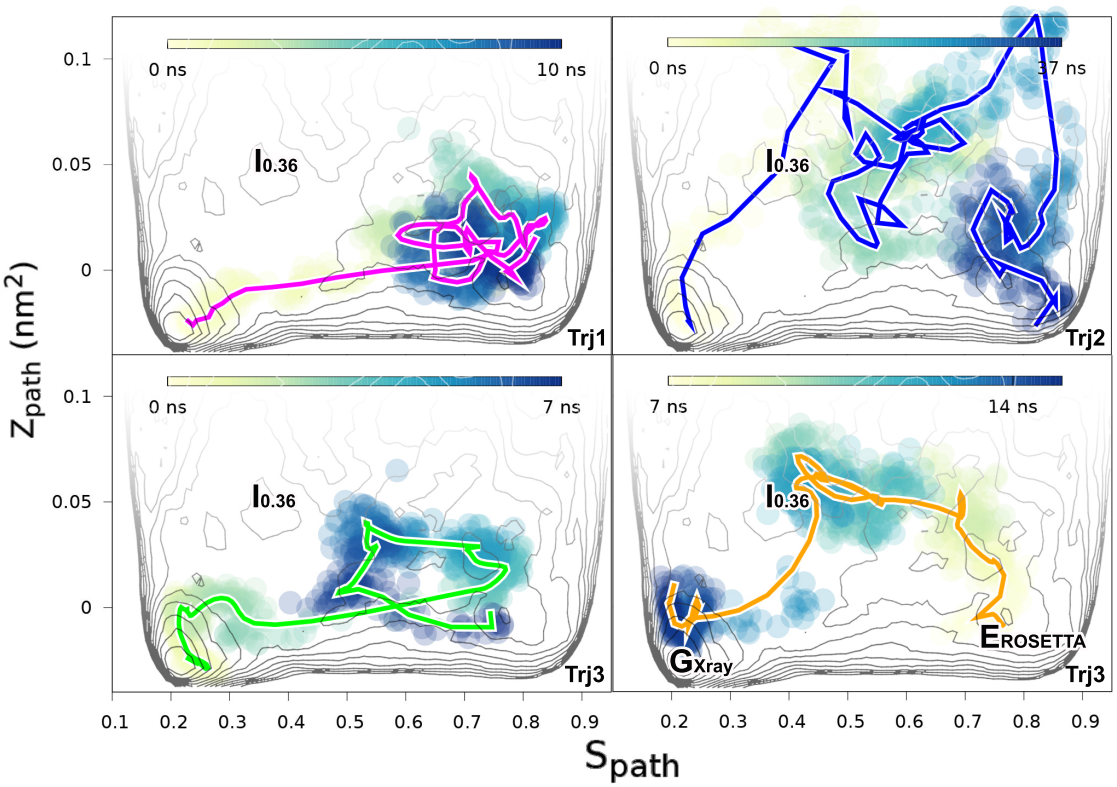


FIGURE 5: **Mechanisms of the G-E conformational exchanges explored by reconnaissance metadynamics.** Trajectories labeled as Trj1 (magenta), Trj2 (blue) and Trj3 (green and orange) are from the simulations RUN10, RUN11 and RUN12 (Appendix-Table S1), respectively. There are multiple routes connecting the G and E states, whose interconversions can take place directly without passing the $I_{0.36}$ state or indirectly via it.

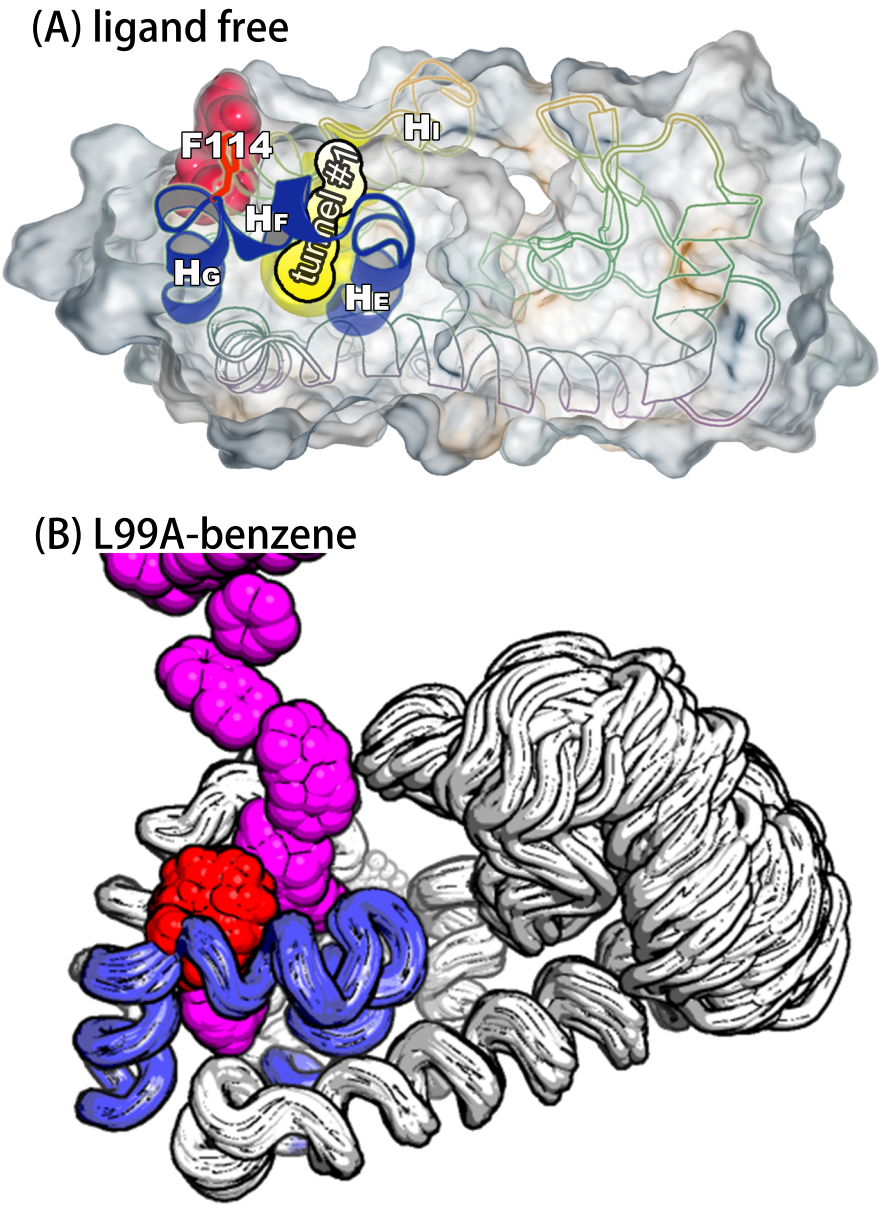


FIGURE 6: **A transiently formed tunnel from the solvent to the cavity is a potential ligand binding pathway.** (A) We here highlight the most populated tunnel structure (tunnel#1), that has an entrance located at the groove between helix F (H_F) and helix I (H_I). Helices E, F and G (blue) and F114 (red) are highlighted. (B) The panel shows a typical path of benzene (magenta) escaping from the cavity of L99A, as seen in ABMD simulations, via a tunnel formed in the same region as tunnel #1 (see also Video 6).

TABLE 1: **Free energy differences and rates of conformational exchange**

	$\tau_{G\to E}$ (ms)	$\tau_{E\to G}$ (ms)	ΔG (kcal mol^{-1})
	L99A		
NMR	20	0.7	2.1
InMetaD	175±56	1.4±0.6	2.9±0.5
PathMetaD			3.5
	L99A/G113A/R119P		
NMR	0.2	4	-1.9
InMetaD	2.0±1.7	14.3±8.3	-1.2±1.1
PathMetaD			-1.6

Captions of figure supplements

Figure 2-figure supplement 1: Approximately equidistant frames along the reference path. The plot reveals a 'gullwing' shape of the matrix of pairwise RMSDs of the frames of the reference path, indicating that frames along the reference path are approximately equidistant. We used 31 structures to discretize the path.

Figure 2-figure supplement 2: One and two dimensional free energy landscape of L99A and the triple mutant. (A) The two-dimensional free energy surface F(S_{path},Z_{path}) of L99A sampled by a 667 ns PathMetaD simulation. (B) The two-dimensional free energy surface F(S_{path},Z_{path}) of the triple mutant sampled by a 961 ns PathMetaD simulation. (C) The free energy profiles as a function of S_{path} of both L99A (black) and the triple mutant (blue).

Figure 4-figure supplement 1: Equilibrium sampling of conformational regions of the E state of L99A by CS-restrained replica-averaged simulation. We calculated the RMSD between the experimental CSs and the values back-calculated by CamShift [39] as implemented in ALMOST [58]. We projected a 250ns MD trajectory sampled using the CHARMM22* force field (RUN3 in Appendix-Table S1) was projected onto the RMSDs. The average RMSDs for the five measured nuclei (H_α, H_N, N, C' and C_α) are 0.23ppm, 0.38ppm, 1.97ppm, 0.83ppm and 1.06ppm, respectively (Appendix-Table S2), which are close to the inherent uncertainty of the chemical shift calculation (Figure 3-figure supplement 2). This indicates the simulation yielded an ensemble in good agreement with experiments.

Figure 4-figure supplement 2: Estimation of the inherent uncertainty of the chemical shift calculation by different softwares: CamShift[39], ShiftX[41] and Sparta+[40]. Using $E_{ROSETTA}$ as the reference structure, we calculated the chemical shifts using different algorithms and compared the correlation coefficients and RMSD between them.

Figure 4-figure supplement 3: Dependence of replica averaged MD simulations of L99A with chemical shift restraints on force fields. The chemical shifts of the E state of L99A (BMRB 17604) were used. (A) The simulation with CHARMM22* force field. (B) The simulation with Amber ff99SB*-ILDN force field.

Figure 4-figure supplement 4: Effect of changing the force constant and number of replicas in CS-restrained simulation of L99A. (A) N=4, $\epsilon_{CS} = 24kJ{\cdot}mol^{-1}$. (B) N=2, $\epsilon_{CS} = 24kJ{\cdot}mol^{-1}$. (C) N=2, $\epsilon_{CS} = 12kJ{\cdot}mol^{-1}$. N refers to the number of replicas that the CS values are averaged over. The CHARMM22* force field was used in these simulations. The results also support the conclusion that the conformational space of the minor (E) state covers a relatively wide set of structures including the $E_{ROSETTA}$ structure.

Figure 4-figure supplement 5: Replica-averaged CS-restrained MD simulation of a T4L triple mutant (L99A/G113A/R119P). Chemical shift restraints were from BMRB 17603 and CHARMM22* force field was used.

Figure 5-figure supplement 1: Complete G-to-E transitions of L99A obtained by reconnaissance metadynamics simulations. The state-specific fraction of contacts [10], Q_G and Q_E, were employed to monitor the conformational transitions to G and E state, respectively. Trajectories Trj1, Trj2 and Trj3 are from the simulations RUN10, RUN11 and RUN12 (Table S1), respectively.

Figure 5-figure supplement 2: Conformational transitions between the G and E states monitored by other order parameters. Trajectories Trj1 (magenta), Trj2 (blue) and Trj3 (green and orange) are from the simulations RUN10, RUN11 and RUN12 (Table S1), respectively. The steepest descent path (SDP, black) used to define the initial path in PathMetaD is also shown as a reference. To measure the distance between helix F and helix I, and between F144 and helix D, we employed order parameters R_{HF-HI} and $R_{F114-HD}$. R_{HF-HI} is defined as the C_α distance between E108 and R137, while $R_{F114-HD}$ is defined as the distance between the $C_{\delta4}$ atom of F114 and the C_α atom of Y88.

Figure 5-figure supplement 3: Solvent accessible surface area (SASA) calculation of the side chain of F114. The figure suggests in the direct $G \leftrightharpoons E$ transitions (Trj1 and first half of Trj3) F114 can rotate its side chain inside the protein core. In contrast, in the $G \leftrightharpoons I_{0.36} \leftrightharpoons E$ route (Trj2 and second half of Trj3) the side chain of F114, which occupies the cavity in the E state, gets transiently exposed to solvent during the transition.

Figure 6-figure supplement 1: A transiently formed tunnel from the solvent to the cavity forms in the $I_{0.36}$ state. (A) Typical structures from the $I_{0.36}$ state sampled in the simulation (between 430ns and 447ns) are mapped onto the free energy surface, and represented by yellow points. (B) The cavity-related regions (helix E, F and G) are coloured in blue, while F114 is coloured in red. F114 tends to be partially solvent exposed in $I_{0.36}$, resulting in the internal cavity to be open. The tunnel#1 connecting the internal cavity and protein surface is coloured in yellow, and has a peanut-shell like shape. (C) shows the radius along the tunnel of structures belong to the cluster of tunnel#1. Lines in different colours represent different structures. Grey dotted line represents the average tunnel radius.

Figure 6-figure supplement 2: Representative structures of the cavity region in the $I_{0.36}$ state. The figure shows six representative structures of the cavity region revealing multiple tunnels that connect the cavity with the protein surface. The different colours correspond to different tunnels observed, and a structure can have different tunnels with different widths present at the same time. The colours represent the relative size with yellow, purple and green corresponding to tunnels of decreasing width.

Figure 6-figure supplement 3: Tunnel clustering analysis on $I_{0.36}$ state. The clustering of tunnels was performed using the CAVER Analyst software [49] and the average-link hierarchical algorithm based on the calculated matrix of pairwise tunnel distances. We found that the most weighted tunnel (denoted as tunnel#1) populates 27% of the $I_{0.36}$ basin. The second and third tunnels populate 20% and 15%, respectively, but their maximal bottleneck radii are 1.4 and 1.3 Å, in contrast to the maximal bottleneck radius of tunnel#1 of 2.5 Å. (A) Heat map visualization of the tunnel profile of tunnel#1. The colour map represents the radius of tunnel#1 along the tunnel length. (B) Average tunnel radius and minimal tunnel radius of individual structures belonging to tunnel#1 cluster. Note that the gaps indicate these snapshots do not have tunnels. (C) The tunnel radius as a function of the tunnel index which is ranked by the average radius (R). The second widest tunnel (tunnel#1) has the highest population and is highlighted in yellow. (D) A typical structure of $I_{0.36}$ with an open tunnel#1. H_E, H_F and H_G are coloured in blue, F114 is coloured in red, and tunnel#1 is coloured in yellow. (E) The figure shows the location of an alkylbenzene (magenta) in a crystal structure of L99A T4L (PDB ID: 4W59). The figure further shows (in yellow) the tunnel induced in the structure by the alkyl chain, as revealed by CAVER3 when applied to the structure after removing the ligand. Because the tunnel overlaps with the alkyl chain of the ligand, only the phenyl moiety of the ligand is visible.

Figure 6-figure supplement 4: Ligand unbinding pathways revealed by ABMD simulations. The figure shows how ABMD simulations allow us to observe the ligand benzene escape from the internal binding site. We performed two sets of 20 simulations using two different force constants for the ABMD (upper: 50 $kJ \cdot mol^{-1} \cdot nm^{-2}$; lower: 20 $kJ \cdot mol^{-1} \cdot nm^{-2}$); note also the different time scales on the two plots. The simulations used the RMSD of the ligand to the bound state as reaction coordinate, but are here shown projected onto the distance between the benzene molecule and the side chain of F114. The three structures in the bottom panel provide representative structures.

Video 1: Trajectory of the G-to-E conformational transition observed in Trj1, corresponding to the red trajectory in Figure 5. The backbone of L99A is represented by white ribbons, Helices E, F and G are highlighted in blue, while F114 is represented by red spheres.

Video 2: Trajectory of the G-to-E conformational transition observed in Trj2, corresponding to the blue trajectory in Figure 5. The backbone of L99A is represented by white ribbons, Helices E, F and G are highlighted in blue, while F114 is represented by red spheres.

Video 3: Trajectory of the G-to-E conformational transition observed in Trj3, corresponding to the green trajectory in Figure 5. The backbone of L99A is represented by white ribbons, Helices E, F and G are highlighted in blue, while F114 is represented by red spheres.

Video 4: Trajectory of the E-to-G conformational transition observed in Trj3, corresponding to the yellow trajectory in Figure 5. The backbone of L99A is represented by white ribbons, Helices E, F and G are highlighted in blue, while F114 is represented by red spheres.

Video 5: Movie of the calculated two-dimensional free energy landscape of L99A as a function of simulation time. The figure shows the time evolution of the free energy surface as a function of S_{path} and Z_{path} sampled in a 667 ns PathMetaD simulation of L99A.

Video 6: A typical trajectory of the benzene escaping from the buried cavity of L99A via tunnel #1 revealed by ABMD simulations. The backbone of L99A is represented by white ribbons, Helices E, F and G are highlighted in blue, while F114 and benzene are represented by spheres in red and magenta, respectively.

Appendix
for
Mapping transiently formed and sparsely populated conformations on a complex energy landscape

Yong Wang, Elena Papaleo and Kresten Lindorff-Larsen[1]

19th August 2016

1 Molecular modeling and system preparation

We used the crystal structure of T4L L99A (PDB ID code 3DMV) as starting point for simulations of the G state of L99A. For the E state of L99A and the G state of the L99A, G113A, R119P triple-mutant we used the CS-ROSETTA structures with PDB ID code 2LCB and 2LC9, respectively.

Each protein was solvated in a dodecahedral box of TIP3P water molecules with periodic boundary conditions. The protein-solvent box had a distance of 10 $\mathring{A}$ from the solute to the box boundary in each dimension, which results in approximately 10,000 water molecules and more than 32,000 atoms. Chloride counter-ions were included to neutralize the overall electric charge of the system. We used the CHARMM22* force field[1] for most of our simulations, but also used the Amber ff99SB*-ILDN[2, 3, 4] for some simulations to examine the dependency of the results on the choice of force field.

The van der Waals interactions were smoothly shifted to zero between 0.8 and 1.0 nm, and the long-range electrostatic interactions were calculated by the means of the particle mesh Ewald (PME) algorithm with a 0.12 nm mesh spacing combined with a switch function for the direct space between 0.8 and 1.0 nm. The bonds involving hydrogen atoms were constrained using the LINCS algorithm. We employed the V-rescale thermostat [5] as the temperature control and simulated the system in the canonical ensemble.

2 Reference transition path and path collective variables

We used path collective variables both to enhance sampling in path driven metadynamics (see below) as well as to represent the conformational landscape sampled by other means. Path collective variables have previously been shown to be very useful in finding free energy channels connecting two metastable states, and also able to construct the global free energy surfaces even far away from the initial path [6]. A reference path is defined by a set of conformations along the path, and the progress along this path can be described mathematically as:

$$S_{path}(X) = \frac{\sum_{i=1}^{N-1} i e^{-\lambda M_i(X)}}{\sum_{i=1}^{N-1} e^{-\lambda M_i(X)}}$$

Here X are the coordinates of the instantaneous protein conformer sampled by MD simulations, N is the number of frames used to describe the reference path (often dependent on the length scale of the conformational transition process), $M_i(X)$ is the mean-square deviation (after optimal alignment) of a subset of the atoms from the reference structure of i'th frame, and λ is a smoothing parameter whose value should roughly be proportional to the inverse of the average mean square displacement between two successive frames along the reference path. With this definition, S_{path} quantifies how far the instantaneous conformer, X, is from the reactant state and the product state, thus monitoring the progress of the system along the conformational transition channel.

Using S_{path} as the sole CV would assume that the initial reference path contains a sufficient description of the important degrees of freedom between the two states. It is, however, rarely possible to guess such a path because determining the actual pathway taken is a goal of the simulation. Thus, S_{path} is supplemented by a second CV, Z_{path}, which measures the deviation away from the structures on the reference path. I.e. if S_{path} quantifies the progress along the path, Z_{path} measures the distance away from the reference path:

$$Z_{path}(X) = -\frac{1}{\lambda} ln \sum_{i=1}^{N} e^{-\lambda M_i(X)}$$

The combination of S_{path} and Z_{path} thus maps the entire conformational landscape to a two-dimensional projection, which can also be thought of as a tube connecting the two end states, and where S measures the progress along the tube and Z the width of the tube. The usefulness of the path CVs is, however, dependent on the quality of reference path, which is determined amongst other things by two factors: (1) the relative accuracy of the reference path and (2) how uniform reference structures are distributed along the path. Because of the explosion in number of possible conformations as one progresses along Z_{path}, simulations are mostly enhanced when S_{path} provides

a relatively good description of the pathway taken. Further, if reference structures are only placed sparsely along the path, one looses resolution of the free energy surface and also decreases the ability to enhance sampling.

To obtain a good reference path, without knowing beforehand the mechanism of conversion, we here developed a new method to construct snapshots along a possible initial path. Taking G_{Xray} and $E_{ROSETTA}$ as initial and final structures, we calculated the optimal reaction paths between them with the MOIL software [7], which has previously been used to explore the mechanism of conformational change of proteins [8]. After minimizing endpoint structures, we employed the minimum-energy-path self-penalty walk (SPW) [9] functional embedded in the CHMIN module to obtain an initial guess for the conformational transition path. This path was subsequently optimized in the SDP (steepest descent path) module by minimizing the target function T consisting of two terms S and C. S is an action function that provides approximate most probable Brownian trajectories connecting the reactant and product states, while C is a restraint function aimed to distribute framesapproximately uniformly along the path. They can be expressed by:

$$S = \sum_{i=1}^{N-1} \sqrt{H_s + (\frac{\partial U}{\partial x_i})^2} |x_{i+1} - x_i|$$
$$C = \lambda \sum_i (\Delta l_{i,i+1} - < \Delta l >)^2$$

where N is the number of frames along the reference path, x_i is the entire vector of conformational coordinates of frame i, U is a potential energy as a function of the mass-weighted coordinate vector, H_s is a constant with an arbitrary positive value, which can be tuned to generate the optimal paths with different thermal energies. $\Delta l_{i,i+1} = M^{1/2}|x_i - x_{i+1}|$ is the arc-length between consecutive frames. λ controls the strength of the restraint function C.

The SDP or minimum energy path is the limiting path in which the action S is optimized with $H_s \to 0$. An important advantage of an SDP is that it is capable of giving a good guess of the minimal energy path which can reflect the major mechanism, only with inexpensive computation. Further details can be found in Refs. [8, 10].

The SDP was approximated as 31 discrete conformations. Most regions of the G_{Xray} and $E_{ROSETTA}$ structures are very similar, with the exception of the cavity-related atoms whose movement determines the conformational transitions between the G and E states. To minimize computations, we used only a subset of heavy atoms around the cavity from amino acid residues 99 to 126 to define S_{path} and Z_{path}, resulting a 'light version' of the reference path which included only 212 atoms. We used the C_α atoms of the whole peptide chain to align the molecule. It is important to note that by focusing only on atoms surrounding the cavity in the calculation of the path-variables we only enhance sampling of the conformational changes relating to the cavity. We used $\lambda = 56.0$ based on the consideration that smooth change of the function of S_{path} can be achieved when $e^{-\lambda < M_i(X) >} \geq 0.1$ [11].

We defined S_{path} and Z_{path} using the SDP as described above. The equidistant requirement of the path is satisfied by the penalty function as is evident from the RMSD matrix for path frames which has a gullwing shape, indicating that each frame is closest to its neighbor and more different from all other reference frames.

3 Metadynamics simulations

Metadynamics discourages the system from sampling already visited conformational regions by continuously adding an external history-dependent repulsive potential at the present value of the reaction coordinates or CVs, which are assumed to include the slowly varying degrees of freedom and thus describe the main features of the dynamics [12]. The biasing potential in metadynamics results in an artificial (enhanced) dynamics but makes it possible to reconstruct the free energy surfaces by removing the bias introduced. The bias is typically added as Gaussians at regular time intervals, τ_G,

and is given by:

$$V_G(S,t) = \sum_{k=1}^{|t/\tau_G|} \omega_k \delta(S(t), S(k \cdot \tau_G))$$

Here S denotes the CVs, k is the index of the individual Gaussians , ω_k is the height of the k'th Gaussian, and $\delta(S(t), S(k \cdot \tau_G))$ is a short-ranged kernel function of the CVs:

$$\delta(S(t), S(k \cdot \tau_G)) = e^{\sum_{j=1}^{n} -\frac{|s_j(t) - s_j(k \cdot \tau_G)|^2}{2\sigma_{kj}^2}}$$

where n is the number of CV, j is the index of a CV, and σ_{kj} and $s_j(k \cdot \tau_G)$ are the width and the position of the Gaussian hills, respectively.

We here used a range of different metadynamics approaches to determine the free energy landscape, and the mechanism and kinetics of conformational exchange (see below).

3.1 Well-tempered metadynamics

In the well-tempered version of metadynamics, the height of the individual Gaussians, ω_k, is decreased as the total bias accumulates over time, in order to improve the convergence of the free energy:

$$\omega_k = \omega_0 e^{-\frac{1}{\gamma - 1} \frac{V_G(S, k \cdot \tau_G)}{k_B T}}$$

Here ω_0 is the initial height, $\gamma = (T + \Delta T)/T$ is referred as the bias factor, which can be tuned to control the speed of convergence and diminish the time spent in lesser-relevant, high-energy states. Thus, the quantity $T + \Delta T$ is often referred as the fictitious CV temperature.

3.2 Adaptive-width metadynamics

In contrast to standard metadynamics in which the width of the Gaussians is constant, the adaptive-width version of metadynamics updates the Gaussian width σ_{kj} on the fly according to local properties of the underlying free-energy surface on the basis of local diffusivity of the CVs or the local geometrical properties[13].

In the region of conformational space near the endpoints of the path CVs many conformations are compressed on similar CV values, leading to high-density but low-fluctuation boundaries. It is apparent that the use of a fixed width might give an inaccurate estimation of the free energy profile in the boundaries where the free energy basins of reactant and product states are located, also makes it more difficult for the simulations to converge. Therefore, the feature of shape adaptive of Gaussian potential is particularly helpful for the case of using path CVs that have significant boundary effects.

3.3 Metadynamics with path variables (PathMetaD)

We sampled the free energy landscape along S_{path} and Z_{path}, as defined above, using adaptive-width metadynamics, which resulted in a finer resolution and faster convergence of the free energy landscape, in particular near the path boundaries, than standard metadynamics. The production simulations were performed at 298K in well-tempered ensemble (Table S1: RUN1 and RUN2).

3.4 Reconnaissance metadynamics

To explore the mechanism of conformational exchange we used reconnaissance metadynamics[14]. This is a 'self-learning' approach which combines of a machine learning technique to automatically identify the locations of free energy minima by periodically clustering the trajectory and a dimensional reduction technique that can reduce the complex locations to a locally-definfed one-dimensional CV by using information collected during the clustering. It has previously been shown that reconnaissance metadynamics makes it possible to determine a path from a large set of input collective variables [15, 16].

3.5 Infrequent Metadynamics (InMetaD)

We used the recently described 'infrequent metadynamics' (InMetaD) to obtain the rates of the conformational exchange process[17]. In standard metadynamics simulations it is very difficult to obtain kinetic properties because the biasing potential is added both to the free energy basins as well as the barriers that separate them. While it is potentially possible to determine the rates from the height of the free energy barriers, this requires both that the CVs used represent the entire set of slowly varying degrees of freedom, and also a good estimate of the pre-exponential factor to convert barrier height to a rate.

The key idea in InMetaD to circumvent these problems is to attempt to add the bias to the system more slowly than the barrier crossing time but faster than the slow inter-basin relaxation time, so that the transition state region has a lower risk of being biased, and therefore the transitions are less affected. By filling up a free energy basin by a known amount it is possible to determine how much the barrier has been decreased, and hence remove this bias from the rates determined. Thus, as described in more detail below, the approach works by performing a large number of individual simulations to obtain first passage times between the individual basins, which are then corrected by the known enhancement factors to obtain estimates of the unbiased rates. This method has been successfully used to reproduce the kinetics of conformational change of alanine dipeptide[17], unbinding of the inhibitor benzamidine from trypsin [18], and slow unbinding of a simple hydrophobic cavity-ligand model system[19].

In these simulations we used a deposition frequency of 80 or 100 ps (see parameters in Table S1), a value much lower than the deposition frequency of 1 ps used in the PathMetaD simulations described above. In this way we lower the risk of substantially corrupting the transition state region. In addition, a tight upper wall potential on Z_{path}=0.10 nm^2 is used to confine the sampling based on our converged free energy surface which shows the conformational change majorly occurs within this region.

With these parameters we collected dozens of trajectories that have a state-to-state transition in the G-to-E and E-to-G directions. The passage times observed in each of these were then corrected for the metadynamics bias as follows.

First, we calculate the acceleration factor α from:

$$\alpha = \tau/\tau_M = < e^{V(s,t)/kT} >_M$$

where the anguluar brackets denote an average over a metadynamics run before the first transition, and $V(s,t)$ is the metadynamics time-depedent bias. The evolution of the acceleration factor $\alpha(t)$ can be expressed by:

$$\alpha(t) = (1/t) \int_0^t dt' e^{V(s,t')/kT}$$

Then the observed passage time, t, is reweighted by:

$$\tau_{true} = \alpha(t) * t = \int_0^t dt' e^{V(s,t')/kT}$$

In principle, the transition time should be a Poisson-distributed random variable, and its mean, μ, standard deviation σ and median $t_m/ln2$ all should be equal to each other. In practice, however, they are somewhat sensitive to insufficient sampling[20]. So rather than simply calculating averages of the individual times, we estimated the average rate and transition time τ from a fit of the empirical cumulative distribution function (ECDF) with the theoretical cumulative distribution function (TCDF):

$$TCDF = 1 - e^{-\frac{t}{\tau}}$$

It has previously been shown that τ estimated in this way converges more quickly than the simple average, μ. This is also consistent with our observation, and we find that 10–15 samples appear sufficient to get a reasonably accurate estimation of the transition time. We used a bootstrap approach to estimate the errors.

To examine whether the observed times indeed follow the expected Poission distribution we used a Kolmogorov-Smirnov (KS) test to obtain a p-value that quantifies the similarity between the empirical and theoretical distributions. Traditionally, a threshold value typically of 0.05 or 0.01 (the significance level of the test) is used to judge if the theoretical (TCDF) and empirical (ECDF) distributions are in agreement. If the p-value is equal to or larger than the threshold value, it suggests that the estimated transition time is quite reliable. If (a) the transition regions were perturbed significantly with infrequent biasing or (b) there are hidden unidentified timescales at play (e.g. the CVs do not capture the slow degrees of freedom) the KS test for time-homogeneous Poisson statistics would fail.

3.6 CS-restrained replica-averaged simulation

The simulation methods described above constitute different ways of exploring the thermodynamics (PathMetaD), kinetics (InMetaD) and mechanism (Reconnaissance metadynamics) of conformational exchange. In all of these simulations, sampling is determined by the molecular energy function (force field), and the experimental information on T4L is used only in the construction of the path variables. When additional experimental information is available, one may introduce an additional energy term so as to bias the simulations to be in agreement with this information[21]. As the experimental values are ensemble averages we apply these restraints only to averages calculated over a number of 'replicas' that are simulated in parallel. In this way, the information from the experimental data is incorporated into the simulation as a perturbation following the maximum entropy principle[22, 23, 24, 25].

We used this approach to obtain conformational ensembles that include not only information from the molecular force field, but also experimental NMR chemical shifts (CS). In particular, we used either the chemical shifts of the E state of L99A obtained from the analysis of the CPMG experiments, or the native state chemical shifts of a triple mutant that populates the same state as its ground state[26]. The CS restraints were imposed by adding a pseudo-energy term (E_{CS}) to a standard molecular-mechanics force field (E_{FF}).

$$E_{CS} = \epsilon_{CS} \sum_{i=1}^{N} \sum_{j=1}^{6} (\delta_{ij}^{Exp} - \frac{1}{M} \sum_{k=1}^{M} \delta_{ijk}^{Sim})^2$$

Here ϵ_{CS} is strength of the CS restraints, i indicates the residue number (total of N), j indicates each of the the six backbone atoms whose chemical shifts were used (C_α, C_β, C', H_α, HN and N), k is an index for the total of M replicas, and δ^{Exp} and δ^{Sim} are the experimental and simulated CSs, respectively. The latter quantity, δ^{Sim}, was calculated by CamShift[27] as a plugin of PLUMED. The CS values of Pro, Gly, Asp, Glu, His and terminal residues were not included because the accuracy in their predictions are too low to contain sufficient information in this approach[27]. We set ϵ_{CS}=24 kJ $mol^{-1}ppm^{-2}$) and used M=4 replicas. In principle, the number of replicas is a free parameter that should be set as large as possible when the experimental data and the method for calculating it is noise free [25]. In practice, one uses a finite set of replicas and it has been shown that M=4 replicas is sufficient to capture the dynamics accurately [28].

We performed replica-averaged CS restrained MD simulations using GROMACS4.6 and the PLUMED2.1 plugin at 298K. The equilibrated structures of $E_{ROSETTA}$ and $G^{Triple}_{ROSETTA}$ were used as the starting configuration (of each of the four replicas) in the CS-restrained simulations of L99A and the L99A,G113A,R119P triple mutant, respectively. The CS data for $E_{ROSETTA}$ and $G^{Triple}_{ROSETTA}$ were obtained from the Biological Magnetic Resonance Bank (BMRB) database [29] with entries 17604 and 17603 , respectively.

3.7 PT-WT-MetaD failed to get the converged free energy landscape

In practice, the choice of CVs plays a fundamental role in determining the accuracy, convergence and efficiency of metadynamics simulations. If an important CV is missing, the exploration of the free energy landscape will be difficult due to hysteresis. Finding a minimal set of CVs that include all important degrees of freedom is a highly nontrivial task and one often has to proceed by several rounds of trial simulations.

At the beginning, we followed a strategy which had previously been successfully used in the exploration of protein conformational transitions [30, 31], to design a set of CVs on the basis of static structural comparison between G_{Xray} and $E_{ROSETTA}$. In particular, by comparing these two structures we defined several CVs that described structural differences by individual dihedral angles and hydrogen bonds, as well as dihedral correlation and coordination number (state-specific contact map) [32] (summarized in Table S1). We combined these in a multiple-replica, parallel tempering approach in the well-tempered ensemble (PT-WT-metaD) [33], to further enhance the sampling. In PT-WT-metaD, the energy fluctuations are enlarged by using energy as a biased CV but the average energy is the same as the canonical ensemble, allowing the use of a larger spacing between temperatures and a much fewer number of replicas than normal PT simulations[34]. Coordinate exchange with high temperature replicas can enhance the sampling of all the degrees of freedom, even those not included in the biased CVs, and one may include a 'neutral' replica (without energy bias, at 298K). We performed a series of simulations with different combinations of CVs starting from the G state of L99A (Table S2). However, unfortunately, we only observed partial G-to-E transitions, even in a relatively long trajectory of about $1\mu s$ for each replica. This negative results suggested that these manually chosen CVs did not contain all the necessary slow degrees of freedom.

3.8 Tunnel analysis

We used CAVER3 [35] to analyse the structures and CAVER Analyst1.0 [36] (http://www.caver.cz/, see also parameters in Table S4) to separate the tunnels into different clusters. CAVER Analyst is a standalone program based on CAVER3.0 algorithm [35]. The settings for the tunnel calculations can be set through the Tunnel Computation window, while the advanced parameters can be set in the Tunnel Advanced Settings window. We used the center of mass of the cavity-related region (residue 93-124) as the position of the starting point. Average-link hierarchical clustering algorithm is performed to build a tree hierarchy of tunnel axes based on their pairwise distances. The size of the resulting clusters is dependent on the Clustering threshold parameter which specifies the level of detail at which the tree hierarchy of tunnel clusters will be cut. We used the default value so that the tree hierarchy of tunnel clusters is cut at the value of 3.5.

3.9 Adiabatic bias molecular dynamics

Adiabatic biased molecular dynamics (ABMD) [37, 38] is an algorithm developed to accelerate the transition from the reactant state to the productive state, here corresponding to the ligand bound state and ligand-free state, respectively. In ABMD the system is perturbed by a 'ratcheting potential', which acts to 'select' spontaneous fluctuations towards the ligand-free state. The ratcheting potential is implemented in PLUMED2.2 as

$$V(\rho(t)) = \begin{cases} 0.5K(\rho(t) - \rho_m(t))^2, & \rho(t) > \rho_m(t) \\ 0, & \rho(t) \leq \rho_m(t) \end{cases}$$

where

$$\rho(t) = (S(t) - S_{target})^2$$

and

$$\rho_m(t) = min_{0 \leq \tau \leq t}\rho(\tau) + \eta(t)$$

K is the force constant, $S(t)$ is the instantaneous CV value, S_{target} is the target value of the CV and $\eta(t)$ is an additional white noise acting on the minimum position of $\rho(t)$. Here, we used the RMSD of the ligand to the cavity-bound state as the CV, and set $S_{target} = 4.0nm$ and K=20 $kJ \cdot mol^{-1} \cdot nm^{-2}$ or 50 $kJ \cdot mol^{-1} \cdot nm^{-2}$ to check the dependency of the force constant chosen. The biasing potential is zero when the CV increases but provides a penalty when the CV decreases. In this way, we were able to observe multiple unbinding events in simulations despite the long lifetime (1.2 ms) of the ligand in the cavity.

4 References

References

[1] Piana, S, Lindorff-Larsen, K, & Shaw, D. E. (2011) How robust are protein folding simulations with respect to force field parameterization? *Biophys J* **100**, L47–L49.

[2] Hornak, V, Abel, R, Okur, A, Strockbine, B, Roitberg, A, & Simmerling, C. (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* **65**, 712–725.

[3] Best, R. B & Hummer, G. (2009) Optimized molecular dynamics force fields applied to the helixcoil transition of polypeptides. *J Phys Chem B* **113**, 9004–9015.

[4] Lindorff-Larsen, K, Piana, S, Palmo, K, Maragakis, P, Klepeis, J. L, Dror, R. O, & Shaw, D. E. (2010) Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins: Struct Funct Bioinf* **78**, 1950–1958.

[5] Bussi, G, Donadio, D, & Parrinello, M. (2007) Canonical sampling through velocity rescaling. *J Chem Phys* **126**, 014101.

[6] Branduardi, D, Gervasio, F. L, & Parrinello, M. (2007) From a to b in free energy space. *J Chem Phys* **126**, 054103.

[7] Elber, R, Roitberg, A, Simmerling, C, Goldstein, R, Li, H. Y, Verkhivker, G, Keasar, C, Zhang, J, & Ulitsky, A. (1995) Moil: A program for simulations of macromolecules. *Comput Phys Commun* **91**, 159–189.

[8] Wang, J, Wang, Y, Chu, X, Hagen, S. J, Han, W, & Wang, E. (2011) Multi-scaled explorations of binding-induced folding of intrinsically disordered protein inhibitor ia3 to its target enzyme. *PLoS Comput Biol* **7**, e1001118.

[9] Czerminski, R & Elber, R. (1990) Self-avoiding walk between 2 fixed-points as a tool to calculate reaction paths in large molecular-systems. *Int J Quantum Chem* **38**, 167–186.

[10] Majek, P, Elber, R, & Weinstein, H. (2008) Pathways of conformational transitions in proteins in coarse-graining of condensed phase and biomolecular systems. *Voth GA, ed. Boca Raton, FL: CRC Press* pp. 185–203.

[11] Tribello, G. A, Bonomi, M, Branduardi, D, Camilloni, C, & Bussi, G. (2014) Plumed 2: New feathers for an old bird. *Comput Phys Commun* **185**, 604–613.

[12] Laio, A & Parrinello, M. (2002) Escaping free-energy minima. *Proc Natl Acad Sci USA* **99**, 12562–12566.

[13] Branduardi, D, Bussi, G, & Parrinello, M. (2012) Metadynamics with adaptive gaussians. *J Chem Theo Comput* **8**, 2247–2254.

[14] Tribello, G. A, Ceriotti, M, & Parrinello, M. (2010) A self-learning algorithm for biased molecular dynamics. *Proc Natl Acad Sci USA* **107**, 17509–17514.

[15] Söderhjelm, P, Tribello, G. A, & Parrinello, M. (2012) Locating binding poses in protein-ligand systems using reconnaissance metadynamics. *Proc Natl Acad Sci USA* **109**, 5170–5175.

[16] Tribello, G. A, Cuny, J, Eshet, H, & Parrinello, M. (2011) Exploring the free energy surfaces of clusters using reconnaissance metadynamics. *J Chem Phys* **135**, 114109.

[17] Tiwary, P & Parrinello, M. (2013) From metadynamics to dynamics *Phys Rev Lett* **111**, 230602.

[18] Tiwary, P, Limongelli, V, Salvalaglio, M, & Parrinello, M. (2015) Kinetics of protein–ligand unbinding: Predicting pathways, rates, and rate-limiting steps *Proc Natl Acad Sci USA* **112**, E386–E391.

[19] Tiwary, P, Mondal, J, Morrone, J. A, & Berne, B. (2015) Role of water and steric constraints in the kinetics of cavity–ligand unbinding *Proc Natl Acad Sci USA* p. 201516652.

[20] Salvalaglio, M, Tiwary, P, & Parrinello, M. (2014) Assessing the reliability of the dynamics reconstructed from metadynamics *J Chem Theo Comput* **10**, 1420–1425.

[21] Robustelli, P, Kohlhoff, K, Cavalli, A, & Vendruscolo, M. (2010) Using nmr chemical shifts as structural restraints in molecular dynamics simulations of proteins *Structure* **18**, 923–933.

[22] Pitera, J. W & Chodera, J. D. (2012) On the use of experimental observations to bias simulated ensembles *Journal of Chemical Theory and Computation* **8**, 3445–3451.

[23] Roux, B & Weare, J. (2013) On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *J Chem Phys* **138**, 084107.

[24] Cavalli, A, Camilloni, C, & Vendruscolo, M. (2013) Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *J Chem Phys* **138**, 094112.

[25] Boomsma, W, Ferkinghoff-Borg, J, & Lindorff-Larsen, K. (2014) Combining experiments and simulations using the maximum entropy principle. *PLoS Comput Biol* **10**, e1003406.

[26] Bouvignies, G, Vallurupalli, P, Hansen, D. F, Correia, B. E, Lange, O, Bah, A, Vernon, R. M, Dahlquist, F. W, Baker, D, & Kay, L. E. (2011) Solution structure of a minor and transiently formed state of a t4 lysozyme mutant. *Nature* **477**, 111–114.

[27] Fu, B, Sahakyan, A. B, Camilloni, C, Tartaglia, G. G, Paci, E, Caflisch, A, Vendruscolo, M, & Cavalli, A. (2014) Almost: An all atom molecular simulation toolkit for protein structure determination. *J Comput Chem* **35**, 1101–1105.

[28] Camilloni, C, Cavalli, A, & Vendruscolo, M. (2013) Replica-averaged metadynamics. *J Chem Theo Comput* **9**, 5610–5617.

[29] Ulrich, E. L, Akutsu, H, Doreleijers, J. F, Harano, Y, Ioannidis, Y. E, Lin, J, Livny, M, Mading, S, Maziuk, D, Miller, Z, et al. (2008) Biomagresbank. *Nucleic Acids Res* **36**, D402–D408.

[30] Sutto, L & Gervasio, F. L. (2013) Effects of oncogenic mutations on the conformational free-energy landscape of egfr kinase. *Proc Natl Acad Sci USA* **110**, 10616–10621.

[31] Papaleo, E, Sutto, L, Gervasio, F. L, & Lindorff-Larsen, K. (2014) Conformational changes and free energies in a proline isomerase. *J Chem Theo Comput* **10**, 4169–4174.

[32] Bonomi, M, Branduardi, D, Bussi, G, Camilloni, C, Provasi, D, Raiteri, P, Donadio, D, Marinelli, F, Pietrucci, F, Broglia, R. A, et al. (2009) Plumed: A portable plugin for free-energy calculations with molecular dynamics. *Comput Phys Commun* **180**, 1961–1972.

[33] Bonomi, M & Parrinello, M. (2010) Enhanced sampling in the well-tempered ensemble. *Phys Rev Lett* **104**, 190601.

[34] Barducci, A, Pfaendtner, J, & Bonomi, M. (2015) in *Molecular modeling of proteins.* (Springer), pp. 151–171.

[35] Chovancova, E, Pavelka, A, Benes, P, Strnad, O, Brezovsky, J, Kozlikova, B, Gora, A, Sustr, V, Klvana, M, Medek, P, et al. (2012) Caver 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput Biol* **8**, e1002708.

[36] Kozlikova, B, Sebestova, E, Sustr, V, Brezovsky, J, Strnad, O, Daniel, L, Bednar, D, Pavelka, A, Manak, M, Bezdeka, M, et al. (2014) Caver analyst 1.0: Graphic tool for interactive visualization and analysis of tunnels and channels in protein structures. *Bioinformatics* p. btu364.

[37] Marchi, M & Ballone, P. (1999) Adiabatic bias molecular dynamics: A method to navigate the conformational space of complex molecular systems *J Chem Phys* **110**, 3697–3702.

[38] Paci, E & Karplus, M. (1999) Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations1 *J Mol Bio* **288**, 441 – 459.

Table S1: **Simulation Details**

Method	label	system	init	length	force field	CVs	parameters	T (K)	software version
PathMetaD	RUN1	L99A	G	667ns	CHARMM22*	S_{path},Z_{path}	γ=20[a], τ_D=1 ps[b], τ_G=1 ps[c], ω_0=1.5[d]	298	PLUMED2.1, GMX4.6
PathMetaD	RUN2	L99A, G113A, R119P	G	961ns	CHARMM22*	S_{path},Z_{path}	γ=20, τ_D=1 ps, τ_G=1ps, ω_0=1.5	298	PLUMED2.1, GMX5.0
CS-restrained MD	RUN3	L99A	E	252ns	CHARMM22*		N[e]=4, $\epsilon_{CS} = 24$[f]	298	PLUMED2.1, ALMOST2.1, GMX5
CS-restrained MD	RUN4	L99A	E	233ns	CHARMM22*		N=2, $\epsilon_{CS} = 24$	298	PLUMED2.1, ALMOST2.1, GMX5
CS-restrained MD	RUN5	L99A	E	221ns	CHARMM22*		N=2, $\epsilon_{CS} = 12$	298	PLUMED2.1, ALMOST2.1, GMX5
CS-restrained MD	RUN6	L99A	E	125ns	AMBER99sb*ILDN		N=4, $\epsilon_{CS} = 24$	298	PLUMED2.1, ALMOST2.1, GMX5
CS-restrained MD	RUN7	L99A	E	190ns	AMBER99sb*ILDN		N=2, $\epsilon_{CS} = 12$	298	PLUMED2.1, ALMOST2.1, GMX5
CS-restrained MD	RUN8	L99A, G113A, R119P	G	204ns	CHARMM22*		N=4, $\epsilon_{CS} = 24$	298	PLUMED2.1, ALMOST2.1, GMX5
CS-restrained MD	RUN9	L99A, G113A, R119P	G	200ns	CHARMM22*		N=2, $\epsilon_{CS} = 24$	298	PLUMED2.1, ALMOST2.1, GMX5
Reconnaissance MetaD	RUN10	L99A	G	120ns	CHARMM22*	2,3,4,5	ΔT=10000[g], Δt=250	298	PLUMED1.3, GMX4.5
Reconnaissance MetaD	RUN11	L99A	G	85ns	CHARMM22*	2,3,4,5	10000, 250	298	PLUMED1.3, GMX4.5
Reconnaissance MetaD	RUN12	L99A	G	41ns	CHARMM22*	2,3,4,5,7,8	100000, 500	298	PLUMED1.3, GMX4.5
PT-WT-MetaD	RUN13	L99A	G	961ns	CHARMM22*	1,2,3,7,8	γ=20, ω_0=0.5	297 298[h]303 308 316 325 341 350 362	PLUMED 1.3, GMX4.5
PT-WT-MetaD	RUN14	L99A	G	404ns	CHARMM22*	1,2,3,6,7	γ=20, ω_0=0.5	297 298 303 308 316 325 333 341 350	PLUMED 1.3, GMX4.5
PT-WT-MetaD	RUN15	L99A	G	201ns	CHARMM22*	1,2,3,6	γ=20, ω_0=0.5	297 298 303 308 316 325 333 341 350	PLUMED 1.3, GMX4.5
PT-WT-MetaD	RUN16	L99A	G	511ns	CHARMM22*	1,2,3,4,5	γ=5, ω_0=0.5	295 298 310 325 341 358 376	PLUMED 1.3, GMX4.5
PT-WT-MetaD	RUN17	L99A	G	383ns	CHARMM22*	1,2,3,4,5	γ=20, ω_0=0.5	298 305 313 322 332 337 343	PLUMED 1.3, GMX4.5
PT-WT-MetaD	RUN18	L99A	E	365ns	CHARMM22*	1	γ=20, ω_0=0.5	298 302 306 311 317 323 330 338	PLUMED 1.3, GMX4.5
plain MD	RUN19	L99A	G	400ns	CHARMM22*			298	GMX5.0
plain MD	RUN20	L99A	E	400ns	CHARMM22*			298	GMX5.0
plain MD	RUN21	L99A	G	400ns	AMBER99sb*ILDN			298	GMX5.0
plain MD	RUN22	L99A	E	400ns	AMBER99sb*ILDN			298	GMX5.0
plain MD	RUN23	L99A,G113A,R119P	G	400ns	CHARMM22*			298	GMX5.0
plain MD	RUN24	L99A,G113A,R119P	E	400ns	CHARMM22*			298	GMX5.0
plain MD	RUN25	L99A,G113A,R119P	G	400ns	AMBER99sb*ILDN			298	GMX5.0
plain MD	RUN26	L99A,G113A,R119P	E	400ns	AMBER99sb*ILDN			298	GMX5.0
InMetaD	RUN27-68[i]	L99A	G		CHARMM22*	S_{path},Z_{path}	γ=20, τ_G=80ps, ω_0=1.0, σ_S=0.016[j], σ_Z=0.002[k]	298	PLUMED2.1, GMX5
InMetaD	RUN69-104[l]	L99A	E		CHARMM22*	S_{path},Z_{path}	γ=20, τ_G=80ps, ω_0=1.0, σ_S=0.016, σ_Z=0.002	298	PLUMED2.1, GMX5
InMetaD	RUN105-119	L99A,G113A,R119P	G		CHARMM22*	S_{path},Z_{path}	γ=15, τ_G=100ps, ω_0=0.8, σ_S=0.016[j], σ_Z=0.002[k]	298	PLUMED2.2, GMX5.1.2
InMetaD	RUN120-134	L99A,G113A,R119P	E		CHARMM22*	S_{path},Z_{path}	γ=15, τ_G=100ps, ω_0=0.8, σ_S=0.016[j], σ_Z=0.002[k]	298	PLUMED2.2, GMX5.1.2
ABMD	RUN135-154	L99A-Benzene	bound		CHARMM22*	$RMSD_{benzene}$	K=50 $kJ \cdot mol^{-1} \cdot nm^{-2}$, $S_{target} = 4.0nm$	298	PLUMED2.2, GMX5.1.2
ABMD	RUN155-174	L99A-Benzene	bound		CHARMM22*	$RMSD_{benzene}$	K=20 $kJ \cdot mol^{-1} \cdot nm^{-2}$, $S_{target} = 4.0nm$	298	PLUMED2.2, GMX5.1.2

[a] γ is the bias factor. [b] τ_D is the characteristic decay time used for the dynamically-adapted Gaussian potential. [c] τ_G is the deposition frequency of Gaussian potential. [d] ω_0 is the height of the deposited Gaussian potential (in $kJ \cdot mol - 1$). [e] N is the number of replicas.
[f] ϵ_{CS} is the strength of chemical shift restraints (in $kJ \cdot mol^{-1} \cdot ppm^{-2}$). [g] ΔT means RUN_FREQ and Δt means STORE_FREQ. Other parameters: BASIN_TOL=0.2, EXPAND_PARAM=0.3, INITIAL_SIZE=3.0. [h] Replica at 298 K is the neutral replica without energy bias.
[i] 42 trajectories collected for G-to-E transitions. [j] σ_S is the Gaussian width of S_{path}. [k] σ_Z is the Gaussian width of Z_{path} (in nm^2). [l] 36 trajectories collected for E-to-G transitions.

Table S2: **Definition of collective variables**

CV	definitions	parameters	purpose
1	total energy	bin=500	enhance energy fluctuations
2	dihedral angle of C_α atoms of consecutive residues F104-Q105-M106-G107	σ=0.1	
3	dihedral angle of C_α atoms of consecutive residues G113-F114-T115-N116	σ=0.1	
4	Q_G, distance in contact map space to the G_{Xray} structure	σ=0.5	
5	Q_E, distance in contact map space to the $E_{ROSETTA}$ structure	σ=0.5	
6	distance between Q_G and Q_E	σ=0.5	
7	number of backbone hydrogen bonds formed between M102 and G107	σ=0.1	
8	dihedral correlation between the C_α dihedral angles of consecutive residues in segment N101-G107	σ=0.1	
9	global RMSD to the whole protein	wall potential	avoid sampling unfolding space

Table S3: **Average root-mean-square deviation ($< RMSD >$ in units of ppm) between experimental CSs and those from the CS-restrained replica-averaged simulations**

Nucleus	RUN3	RUN4	RUN5	RUN6	RUN7	RUN8	RUN9
C'	0.833	0.655	0.776	0.854	0.793	0.907	0.727
C_α	1.055	0.879	0.929	1.065	0.940	1.103	0.894
N	1.966	1.707	1.771	1.967	1.780	2.011	1.828
H_N	0.379	0.275	0.291	0.368	0.284	0.414	0.286
H_A	0.232	0.183	0.186	0.242	0.182	0.246	0.183

Table S4: **Parameter set used in tunnel analysis using CAVER3.0[35] and CAVER Analyst1.0[36]**

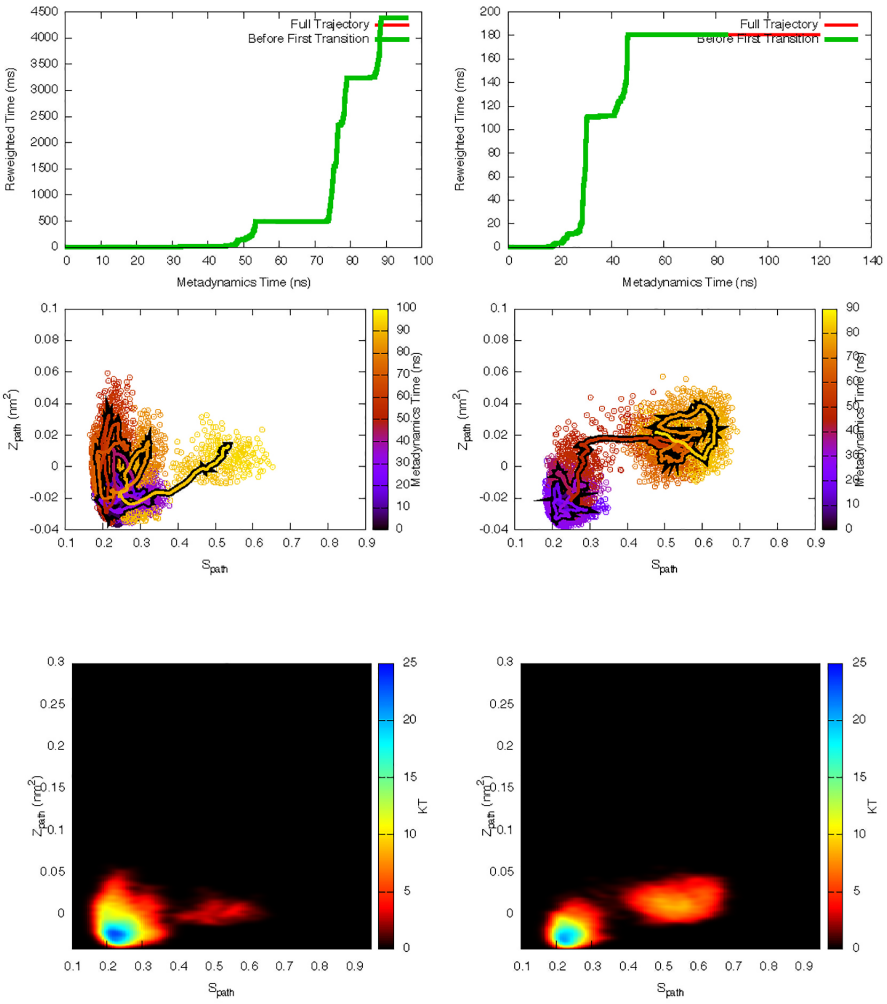
Minimum probe radius	0.9 $\mathring{A}$
Shell depth	4
Shell radius	3
Clustering threshold	3.5
Starting point optimization	
Maximum distance	3 $\mathring{A}$
Desired radius	5 $\mathring{A}$

Table S5: **Clustering Analysis of Tunnels (Top Three Listed)**

Index	Population	Maximal bottleneck radius ($\AA$)	Average bottleneck radius ($\AA$)
#1	27%	2.5	1.3
#2	20%	1.4	1.0
#3	15%	1.3	1.0

Table S6: **Unbinding Pathways Explored by ABMD ($RMSD_{BNZ}$ as CV)**

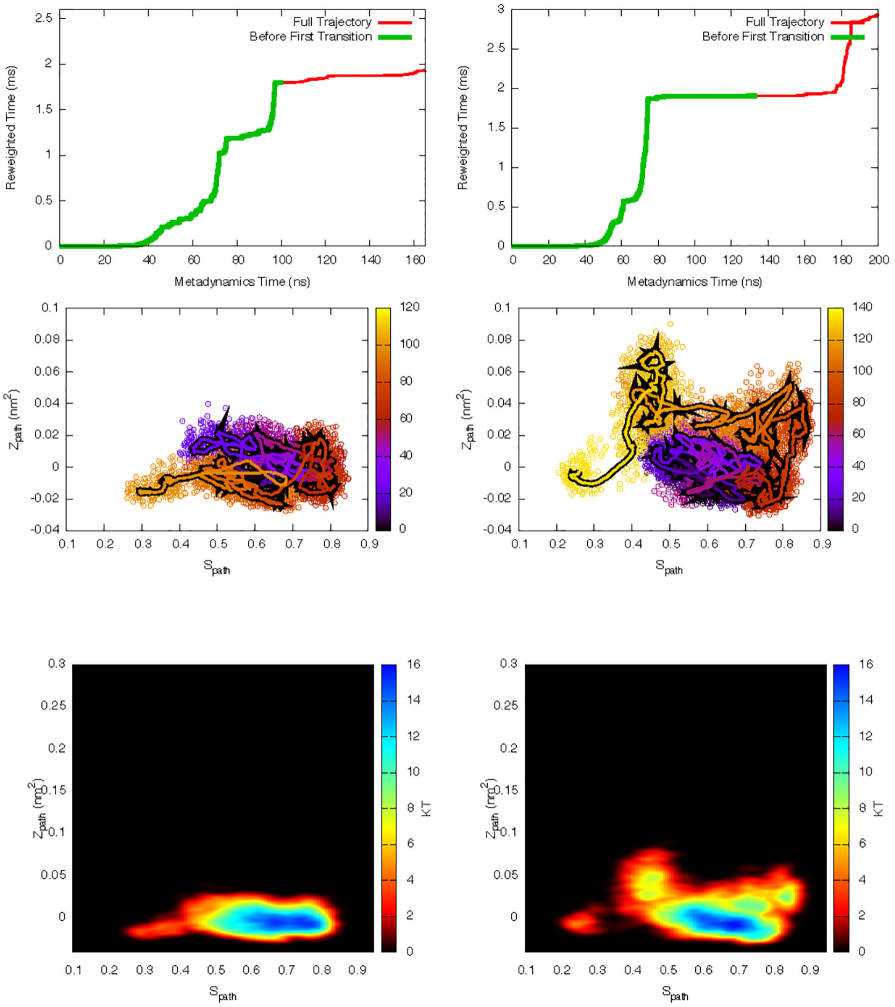
	k=20 $kJ/(mol \cdot nm^2)$		k=50 $kJ/(mol \cdot nm^2)$	
Index	Length	Path	Length	Path
RUN1	56 ns	P1	27 ns	P2
RUN2	36 ns	P2	78 ns	P1
RUN3	43 ns	P1	6 ns	P1
RUN4	43 ns	P1	35 ns	P1
RUN5	77 ns	P2	10 ns	P1
RUN6	176 ns	P1	44 ns	P1
RUN7	41 ns	P1	18 ns	P1
RUN8	106 ns	P1	15 ns	P1
RUN9	72 ns	P1	7 ns	P1
RUN10	107 ns	P1	2 ns	P1
RUN11	61 ns	P1	20 ns	P2
RUN12	58 ns	P2	26 ns	P1
RUN13	64 ns	P1	31 ns	P2
RUN14	173 ns	P2	20 ns	P1
RUN15	172 ns	P1	34 ns	P1
RUN16	74 ns	P2	22 ns	P1
RUN17	20 ns	P1	17 ns	P1
RUN18	34 ns	P1	35 ns	P2
RUN19	91 ns	P1	21 ns	P2
RUN20	61 ns	P1	18 ns	P1
Cost	1.6 μs		0.5 μs	
	Summary			
P1	75% (15/20)		75% (15/20)	
P2	25% (5/20)		25% (5/20)	



(A) example 1

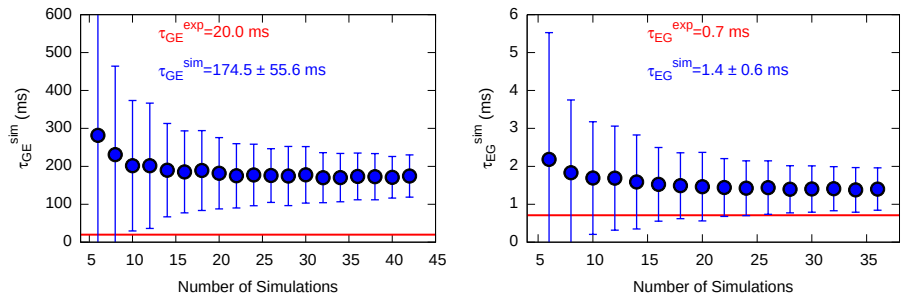
(B) example 2

Figure S1: **Two representative InMetaD trajectories of L99A with G to E transitions.** The time point, t', for the first transition from G to E is identified when the system evolves into conformational region of $S_{path} > 0.55$ and $Z_{path} < 0.01$. We then calculate the unbiased passage time by multiplying t' by the corresponding accelerate factor $\alpha(t')$. Upper panels show the evolution of reweighted time as a function of metadynamics time. The kinks usually indicate a possible barrier-crossing event. Middle panels show the trajectories starting from the G state and crossing the barrier towards the E state. Lower panels show the biasing landscape reconstructed from deposited Gaussian potential, which can be used to check the extent to which the transition state regions are affected by deposited bias potential.



(A) example 1 (B) example 2

Figure S2: **Two representive InMetaD trajectories of L99A with E to G transitions of L99A.** First transition time for G to E transition is identified when the system evolves into conformational region of $S_{path} < 0.28$ and $Z_{path} < -0.01$.



(A) G → E (B) E → G

Figure S3: **Characteristic transition times between G and E states of L99A.** The error bars represent the standard deviation of τ obtained from a bootstrap analysis, and suggest that ten simulations are sufficient to give a reliable estimation of the transition time.

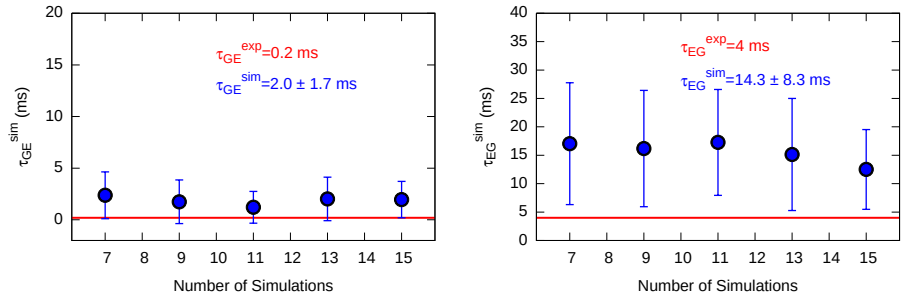


Figure S4: **Characteristic transition times between G and E states of the triple mutant.** The figure shows the characteristic transition time $\tau_{G\to E}$ (right panel) and $\tau_{E\to G}$ (right panel) of the triple mutant as a function of the size of a subsample of transition times randomly extracted from the main complete sample. Errorbars represent the standard deviation of characteristic transition times obtained by a bootstrap analysis. The calculated and experimental values of the transition times are shown in blue and red texts, respectively.

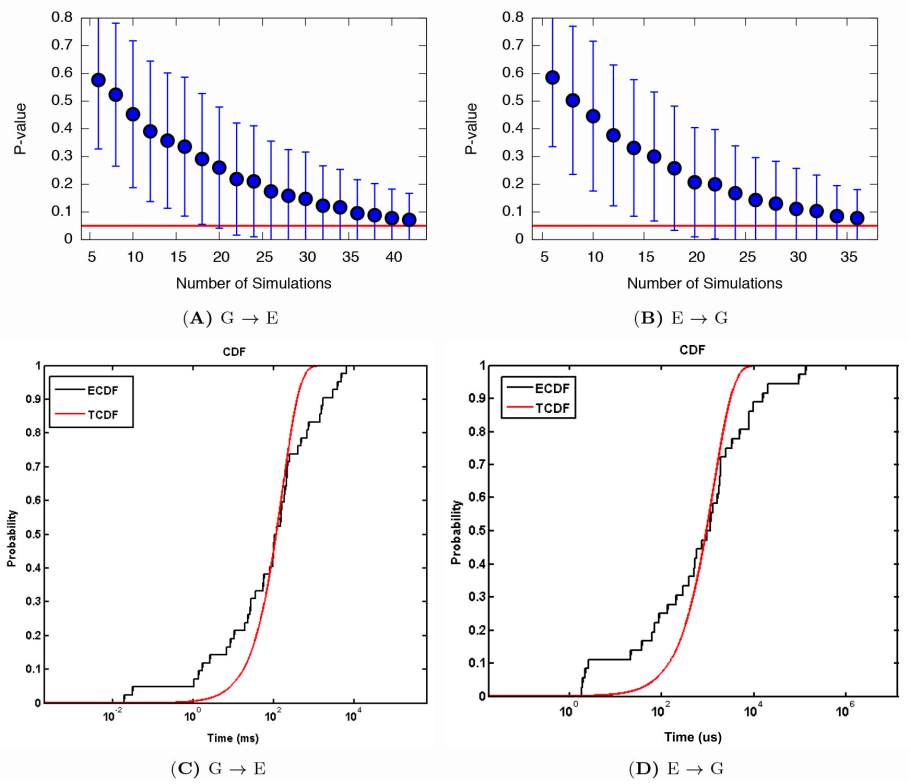
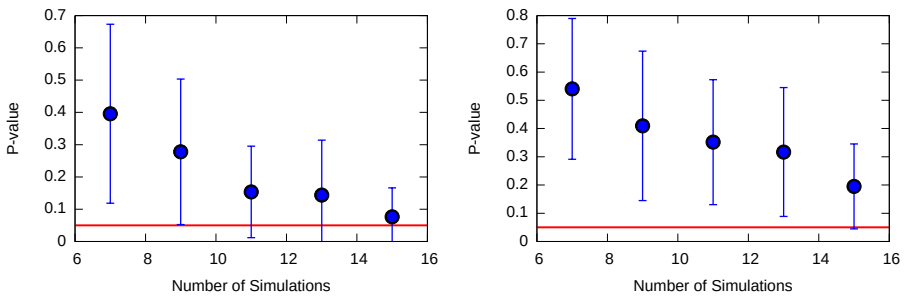
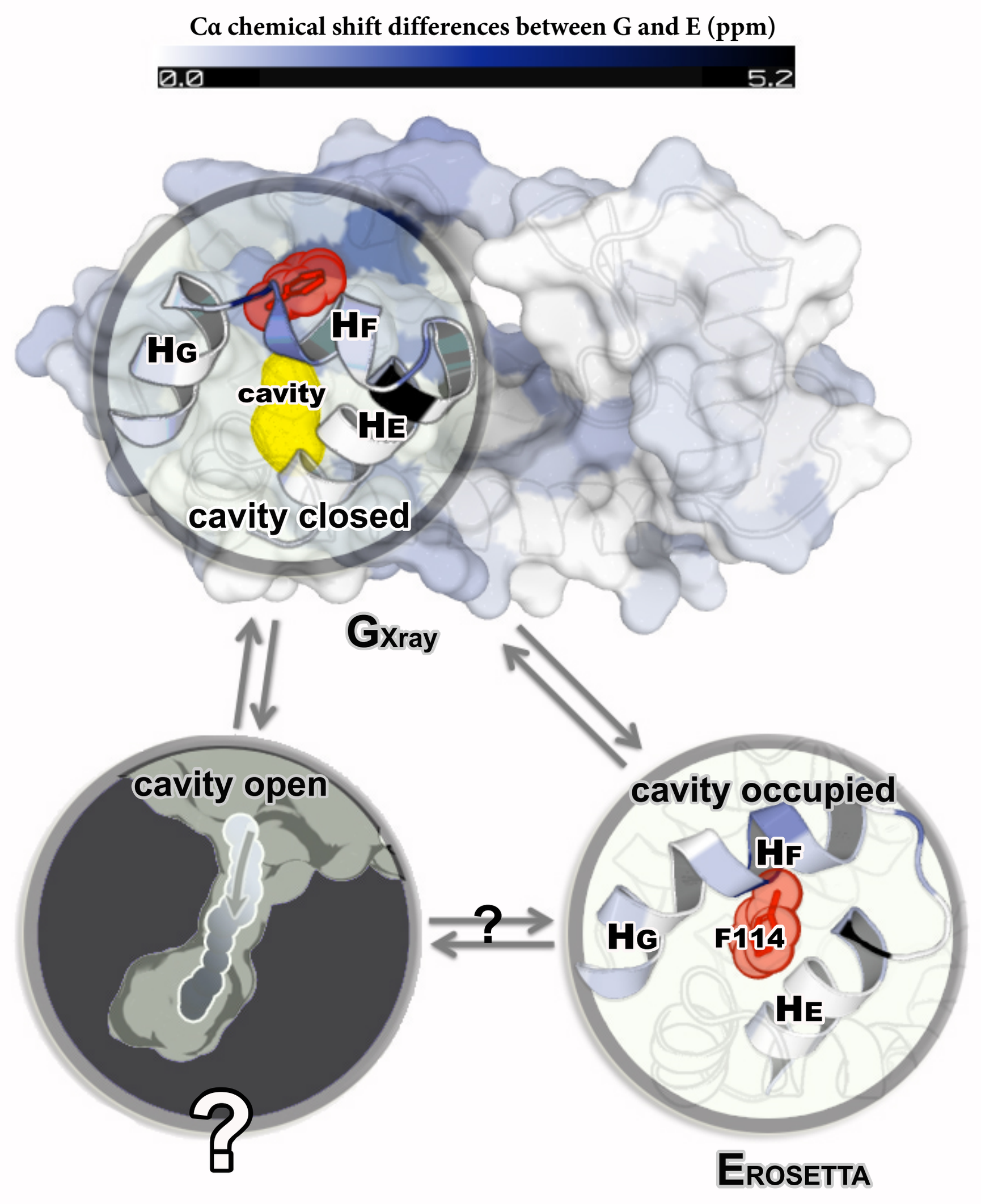


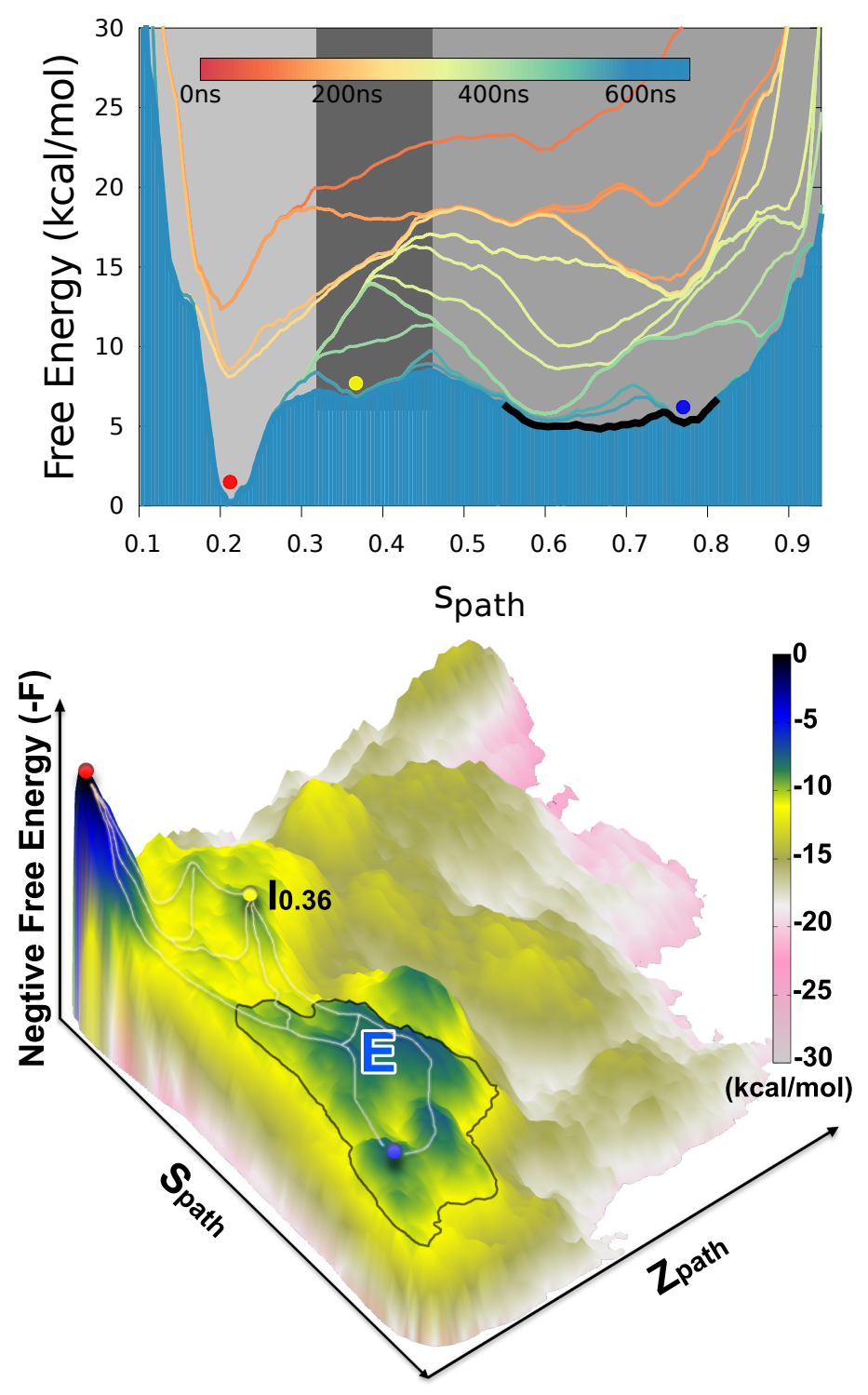
Figure S5: **Poisson fit analysis for G to E transitions and E to G transitions of L99A.** We show the ECDF (the empirical cumulative distribution function) and TCDF (the theoretical cumulative distribution function) in black and blue lines, respectively. The respective p-values are reasonably, albeit not perfectly, well above the statistical threshold of 0.05 or 0.01, indicating the kinetics is not substantially modified by the deposited bias potential in InMetaD. Error bars are the standard deviation obtained by a bootstrap analysis.

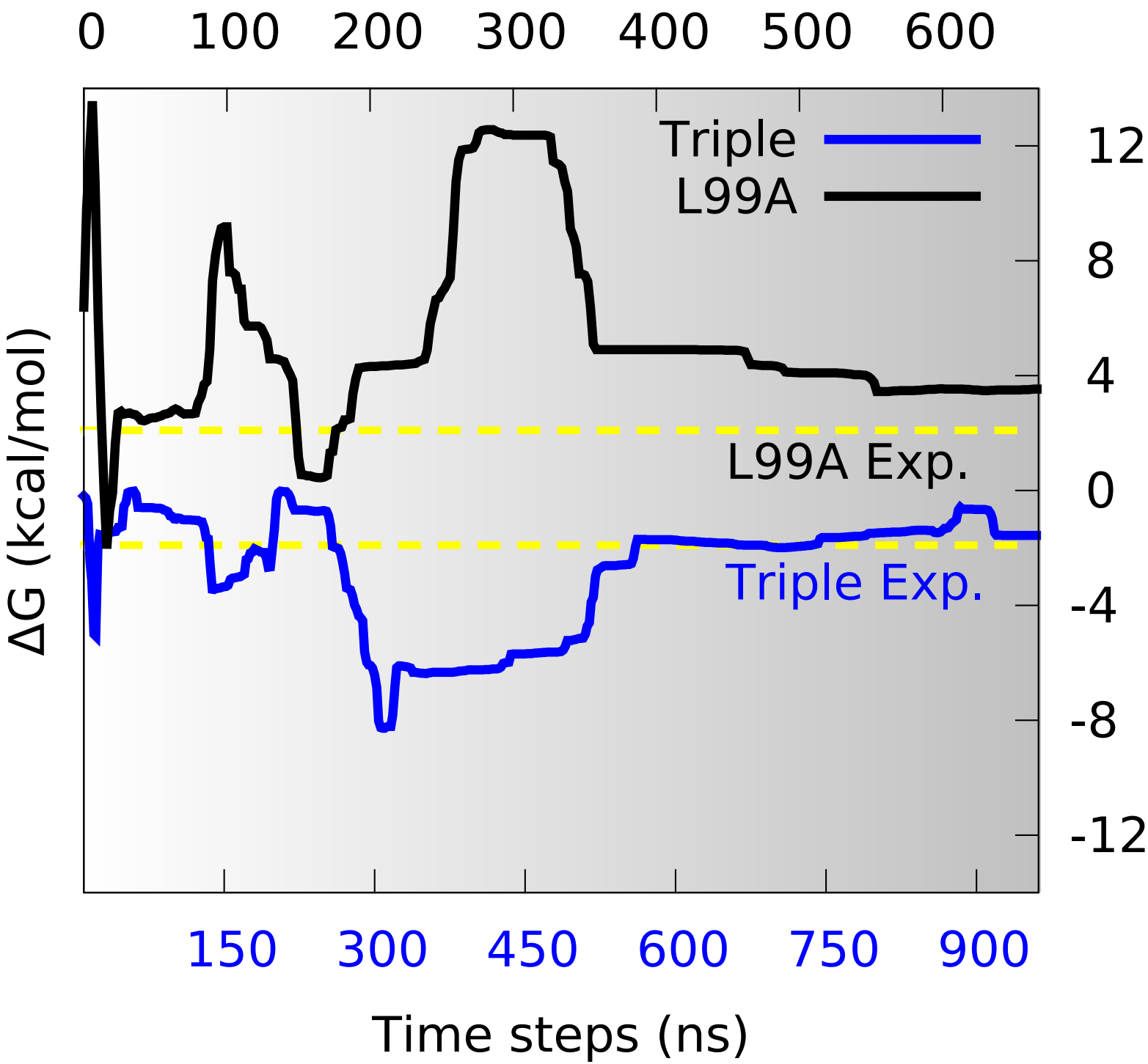


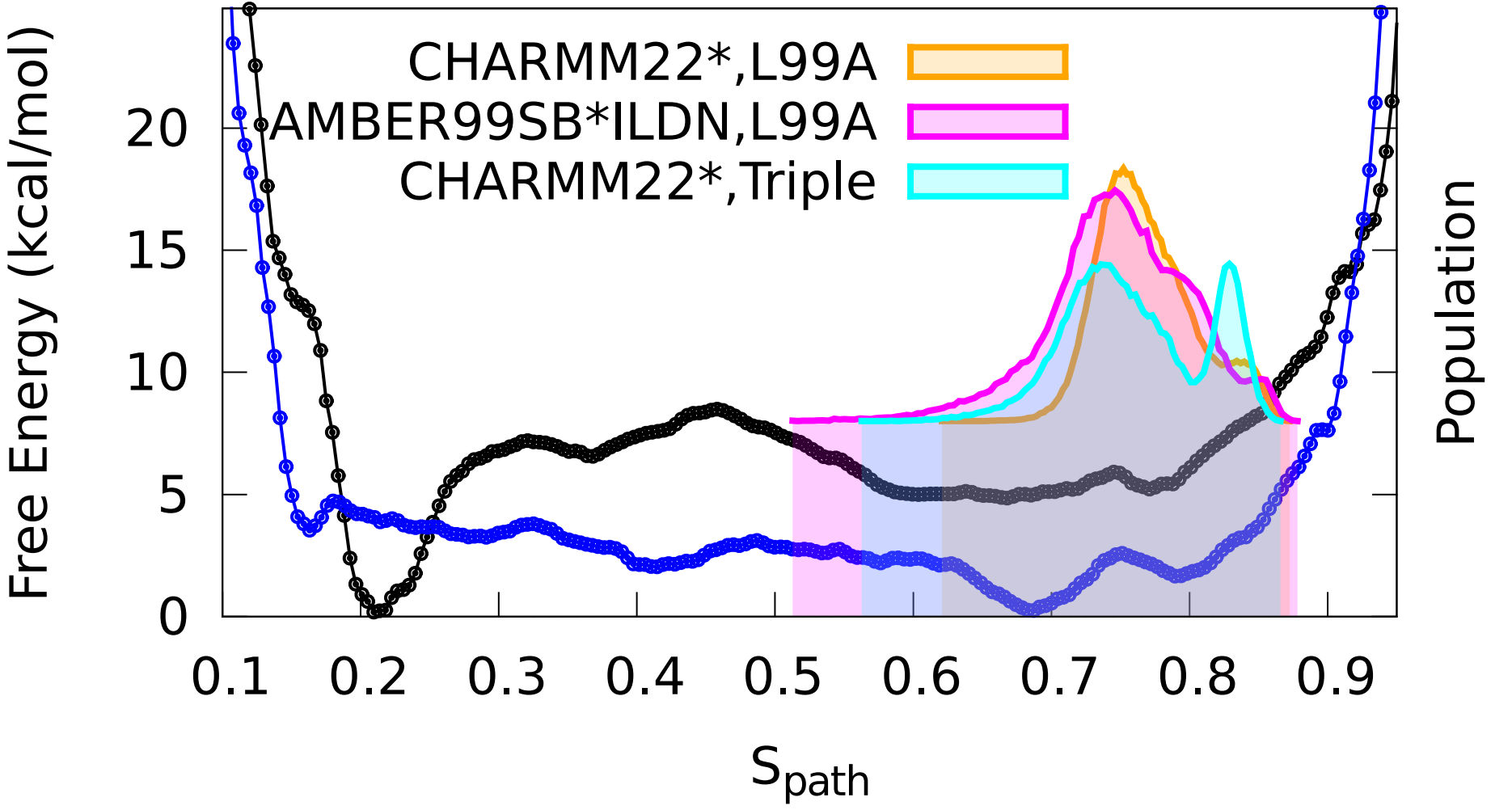
(**A**) G→E (**B**) E→G

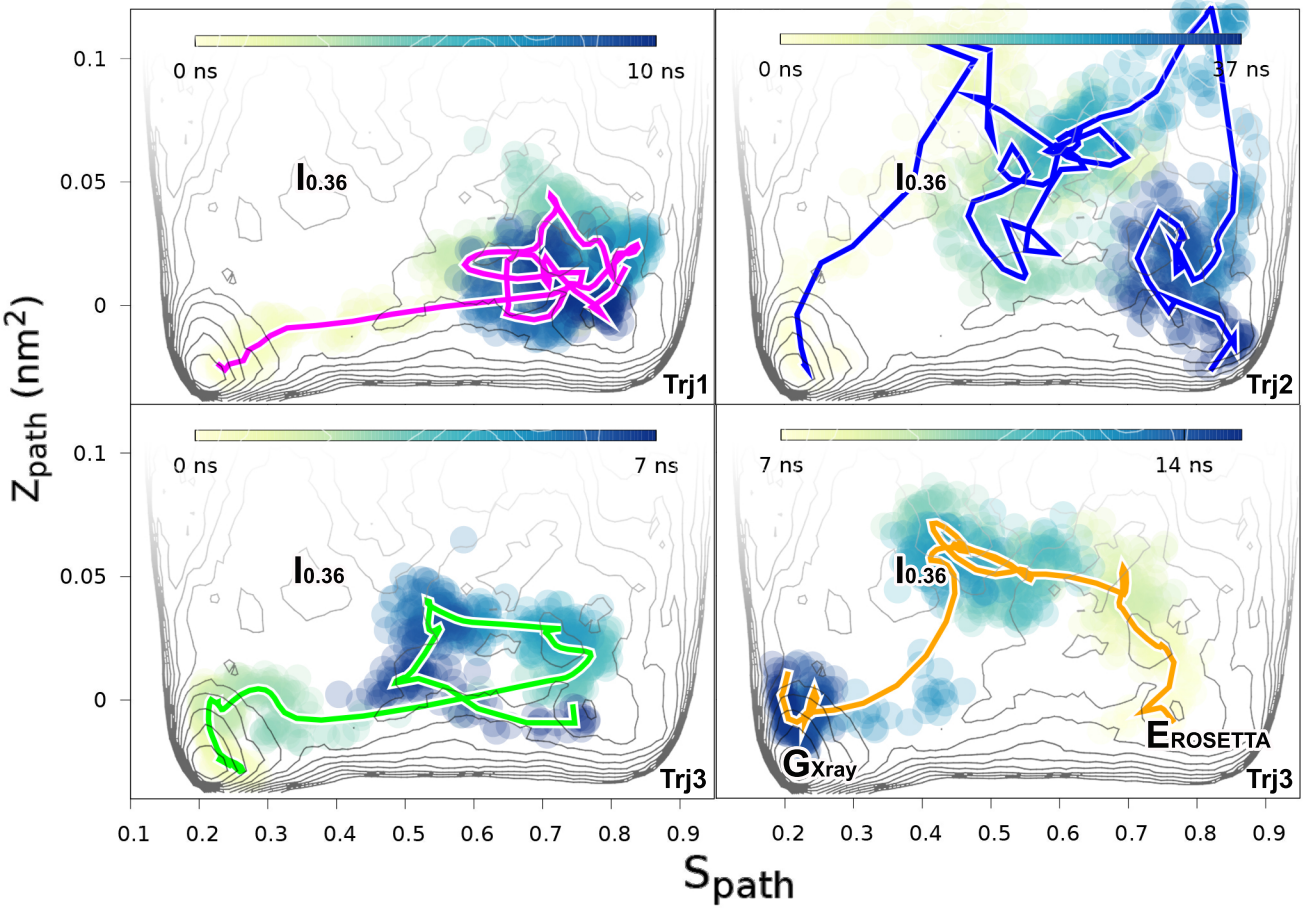
Figure S6: **Poisson fit analysis for G to E transitions and E to G transitions of the triple mutant.** The figure shows the p-values of the Poisson fit analysis of $G \rightarrow E$ (A) and $E \rightarrow G$ (B) transition times as a function of the size of a subsample of transition times randomly extracted from the main complete sample.


Cα chemical shift differences between G and E (ppm)
0.0
5.2
HF
HG
cavity
HE
cavity closed
GXray
cavity open
?
cavity occupied
HF
HG
F114
HE
?
EROSETTA

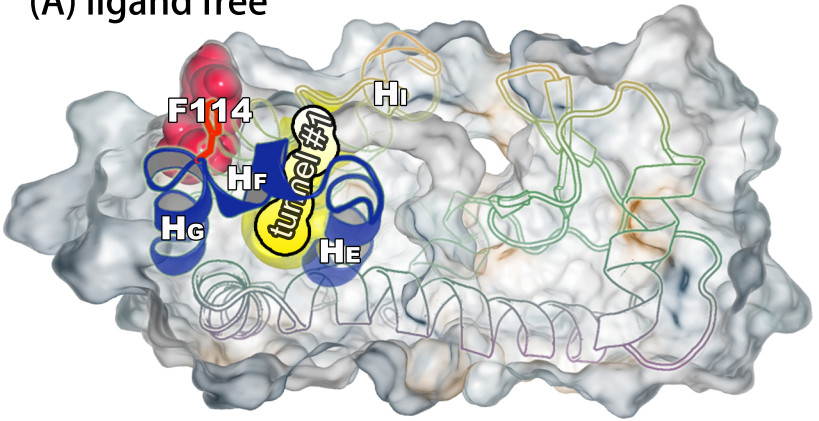

Free Energy (kcal/mol)
0ns
200ns
400ns
600ns
Spath
Negtive Free Energy (-F)
I0.36
E
(kcal/mol)
Spath
Zpath


0
100
200
300
400
500
600
Triple
L99A
L99A Exp.
Triple Exp.
12
8
4
0
-4
-8
-12
ΔG (kcal/mol)
150
300
450
600
750
900
Time steps (ns)

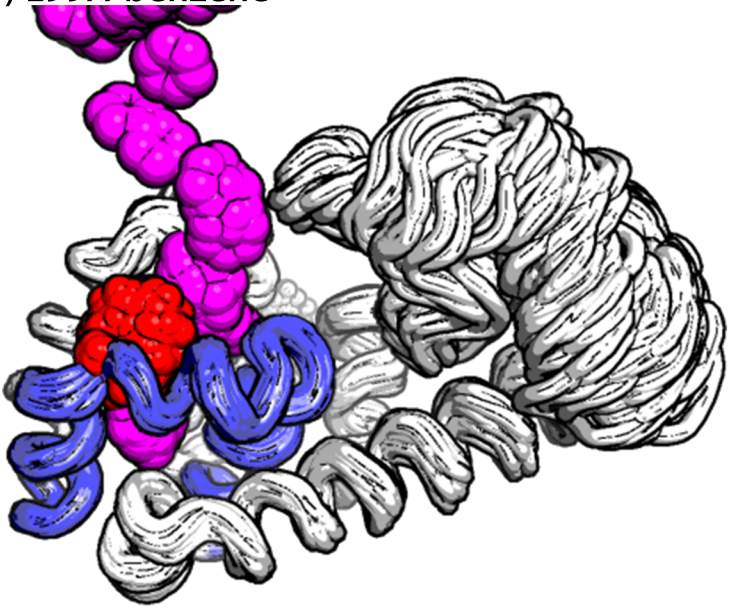

CHARMM22*,L99A
AMBER99SB*ILDN,L99A
CHARMM22*,Triple
Free Energy (kcal/mol)
Population
0
5
10
15
20
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
S path

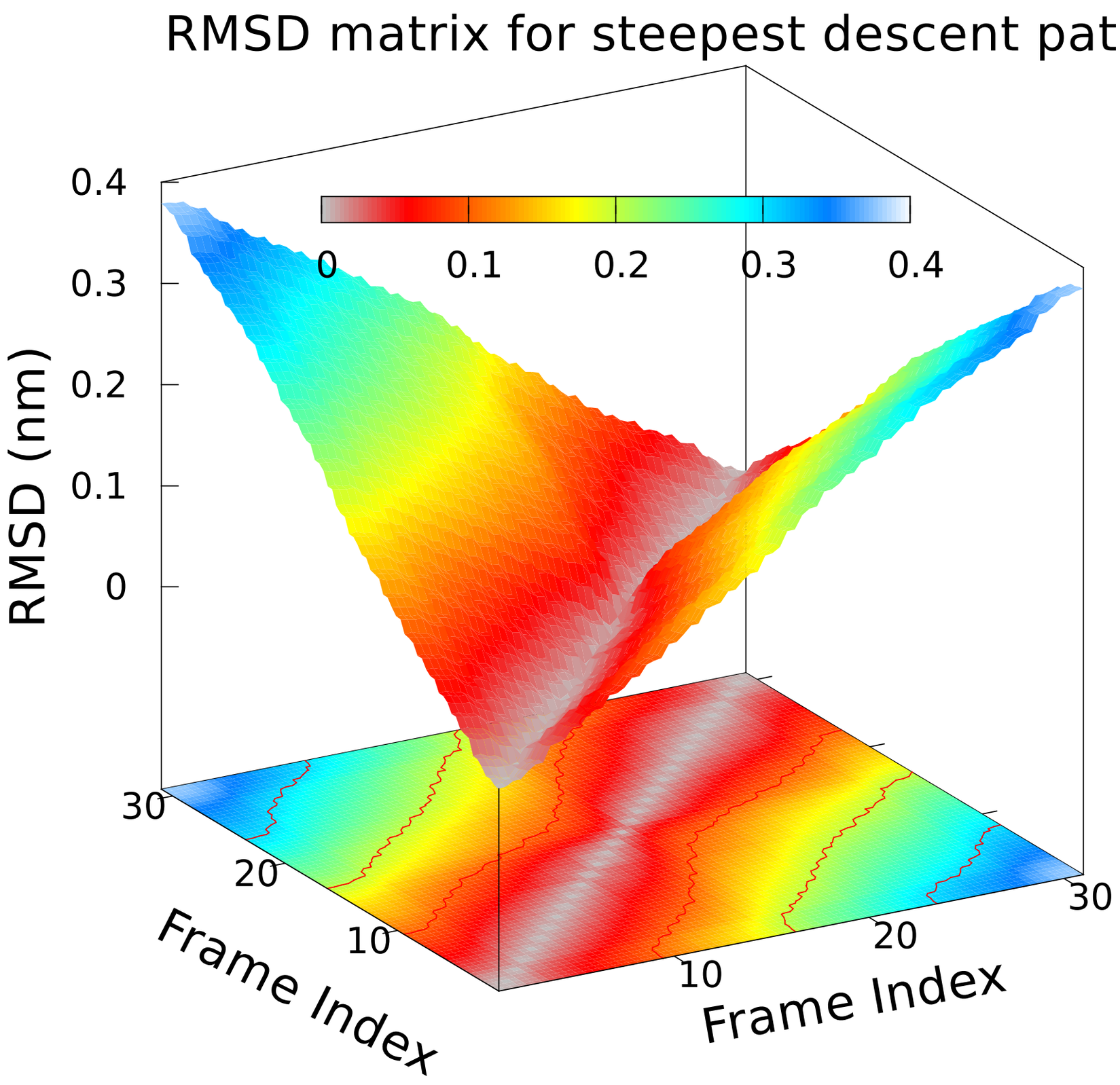

0 ns
10 ns
I0.36
Trj1
0 ns
37 ns
I0.36
Trj2
0 ns
7 ns
I0.36
Trj3
7 ns
14 ns
I0.36
EROSETTA
GXray
Trj3
0.1
0.05
0
Zpath (nm2)
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
Spath

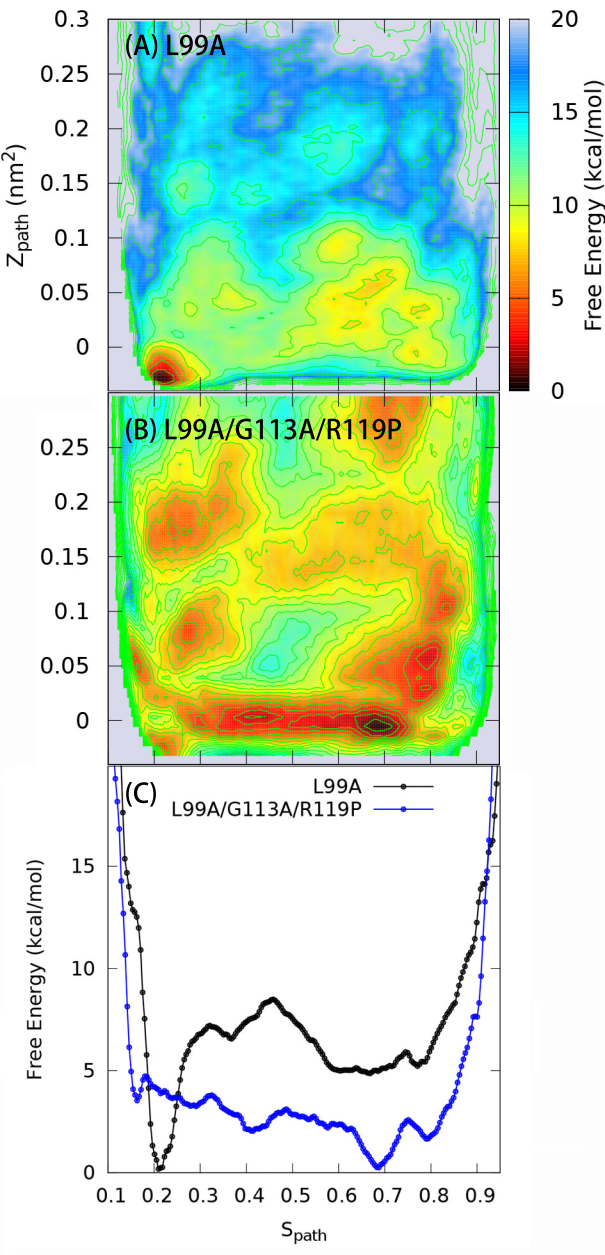
(A) ligand free

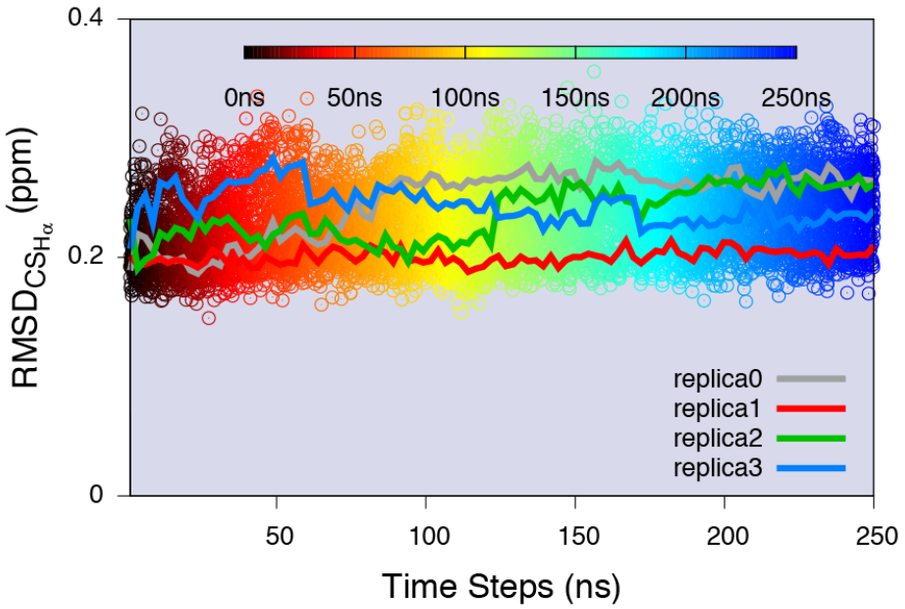


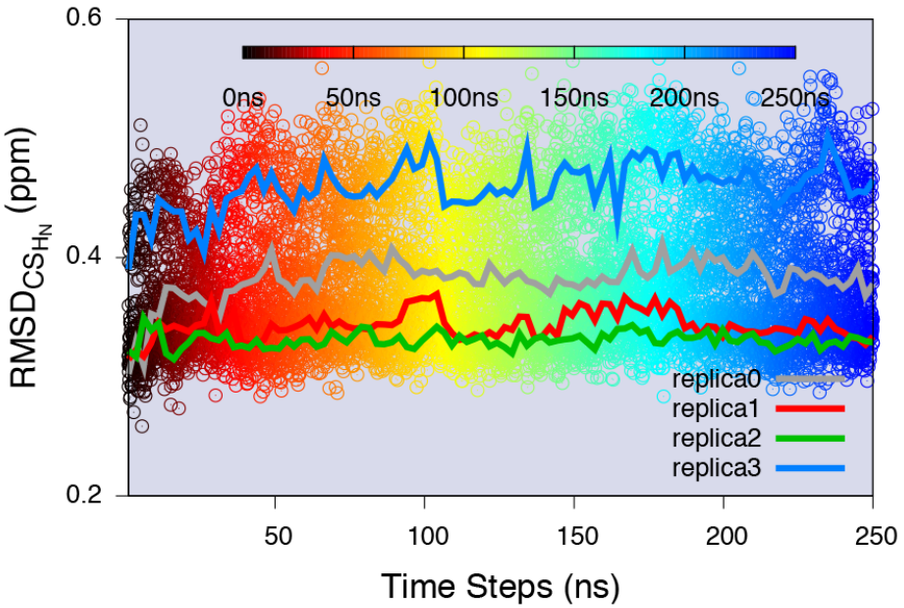
(B) L99A-benzene

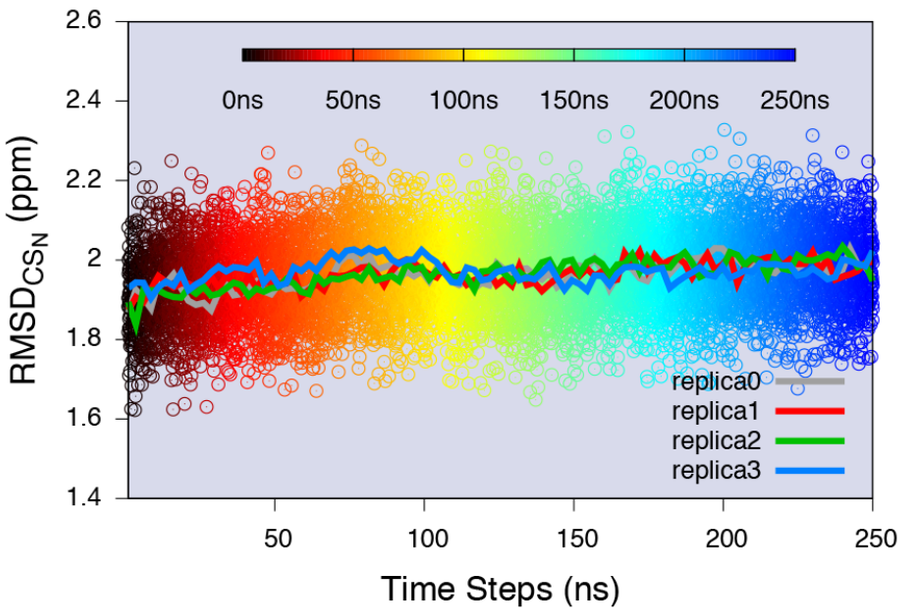


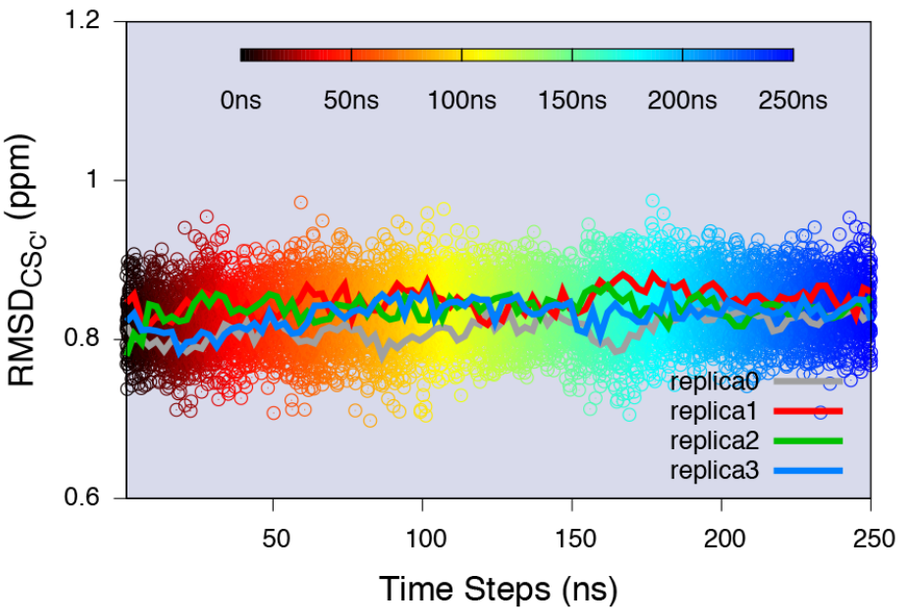

RMSD matrix for steepest descent path
RMSD (nm)
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
30
20
10
Frame Index
10
20
30
Frame Index

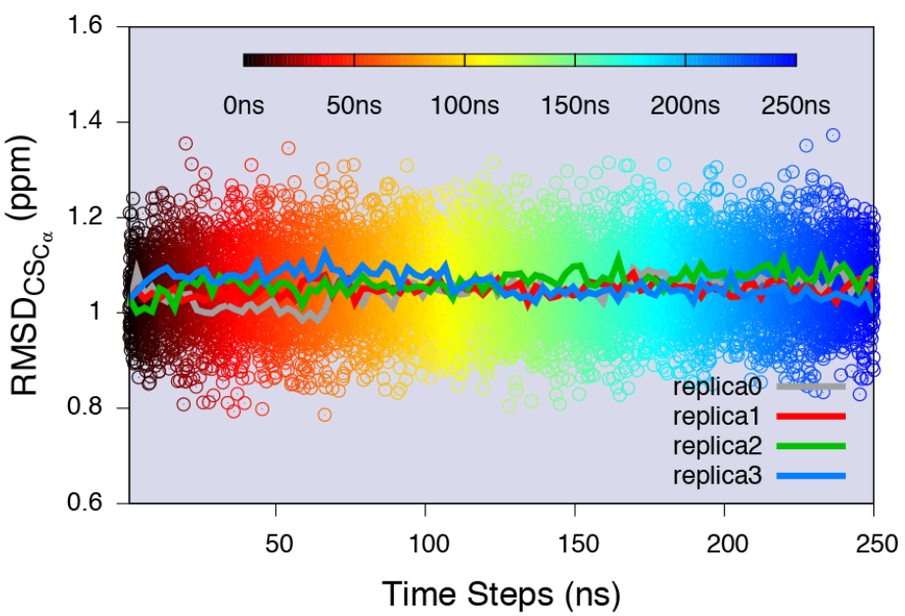

(A) L99A
(B) L99A/G113A/R119P
(C)
L99A
L99A/G113A/R119P
Z_{path} (nm^2)
Free Energy (kcal/mol)
S_{path}
0.3
0.25
0.2
0.15
0.1
0.05
0
20
15
10
5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

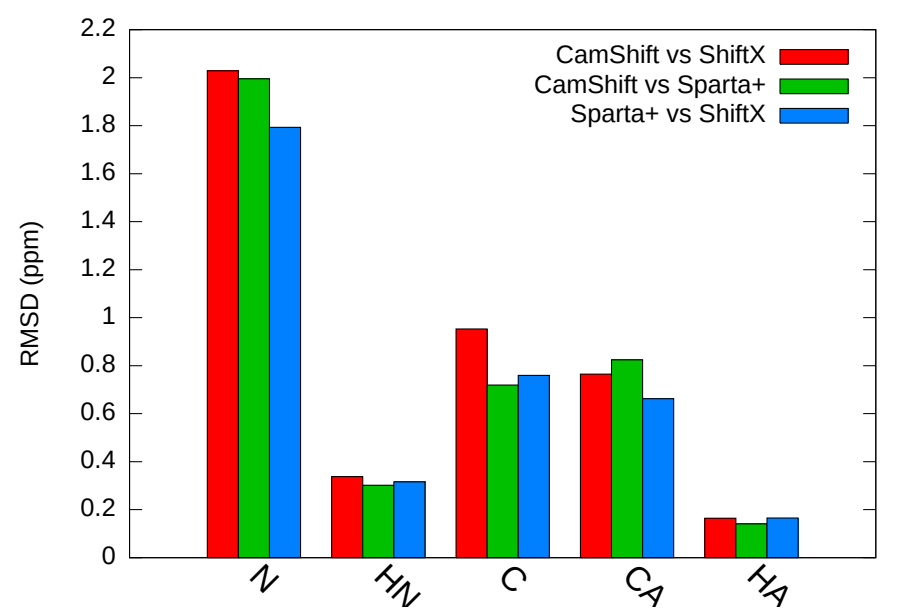

0ns
50ns
100ns
150ns
200ns
250ns
replica0
replica1
replica2
replica3
RMSD$_{CS_{H_\alpha}}$ (ppm)
Time Steps (ns)

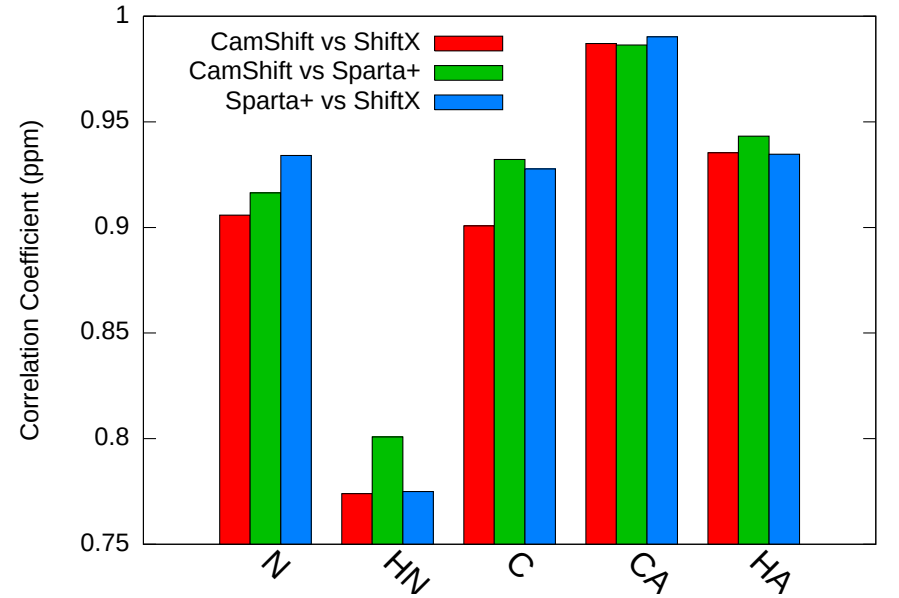

0ns
50ns
100ns
150ns
200ns
250ns
replica0
replica1
replica2
replica3
RMSD$_{CS_{H_N}}$ (ppm)
Time Steps (ns)

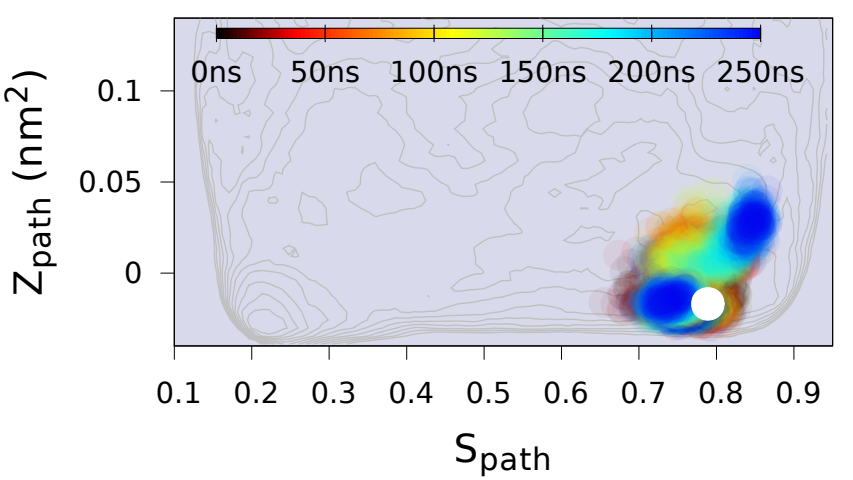

0ns
50ns
100ns
150ns
200ns
250ns
replica0
replica1
replica2
replica3
RMSD$_{CS_N}$ (ppm)
Time Steps (ns)


0ns
50ns
100ns
150ns
200ns
250ns
replica0
replica1
replica2
replica3
RMSD$_{CS_{C'}}$ (ppm)
Time Steps (ns)

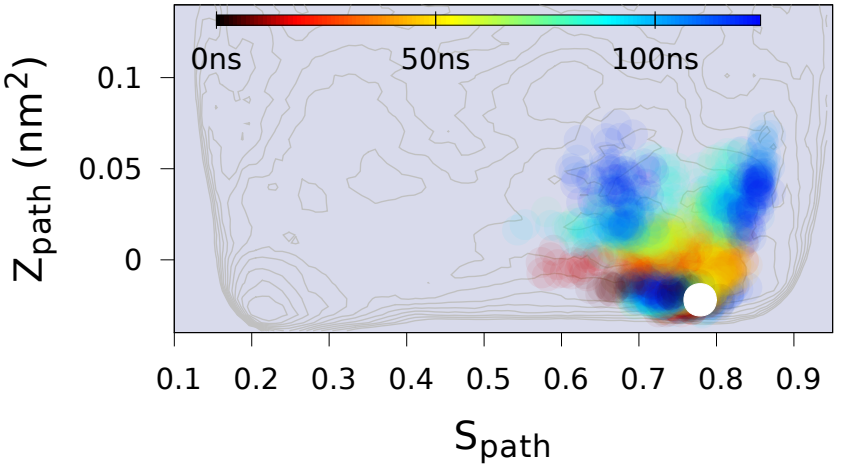

0ns
50ns
100ns
150ns
200ns
250ns
replica0
replica1
replica2
replica3
RMSD$_{CS_{C_\alpha}}$ (ppm)
Time Steps (ns)


RMSD (ppm)
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
CamShift vs ShiftX
CamShift vs Sparta+
Sparta+ vs ShiftX
N
HN
C
CA
HA

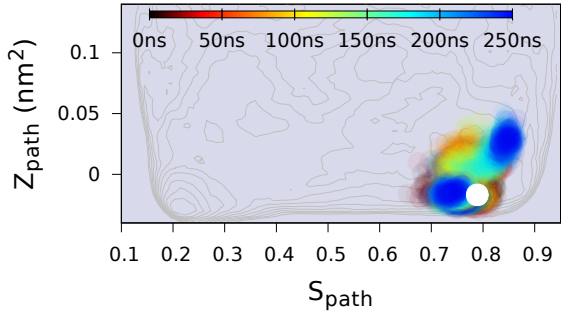

Correlation Coefficient (ppm)
1
0.95
0.9
0.85
0.8
0.75
CamShift vs ShiftX
CamShift vs Sparta+
Sparta+ vs ShiftX
N
HN
C
CA
HA



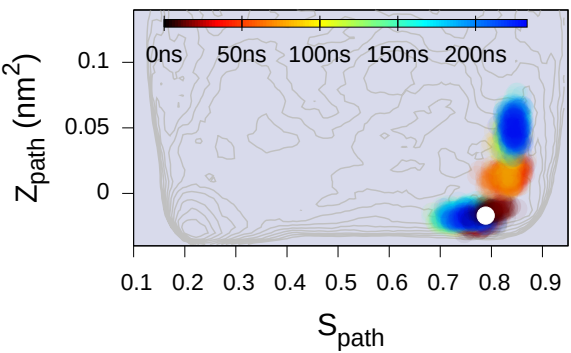
(**A**) CHARMM22*



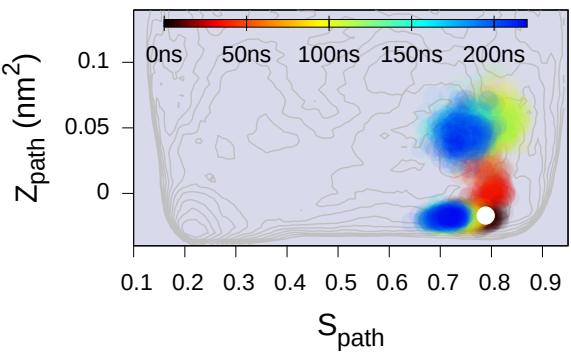
(**B**) AMBER99sb*-ILDN



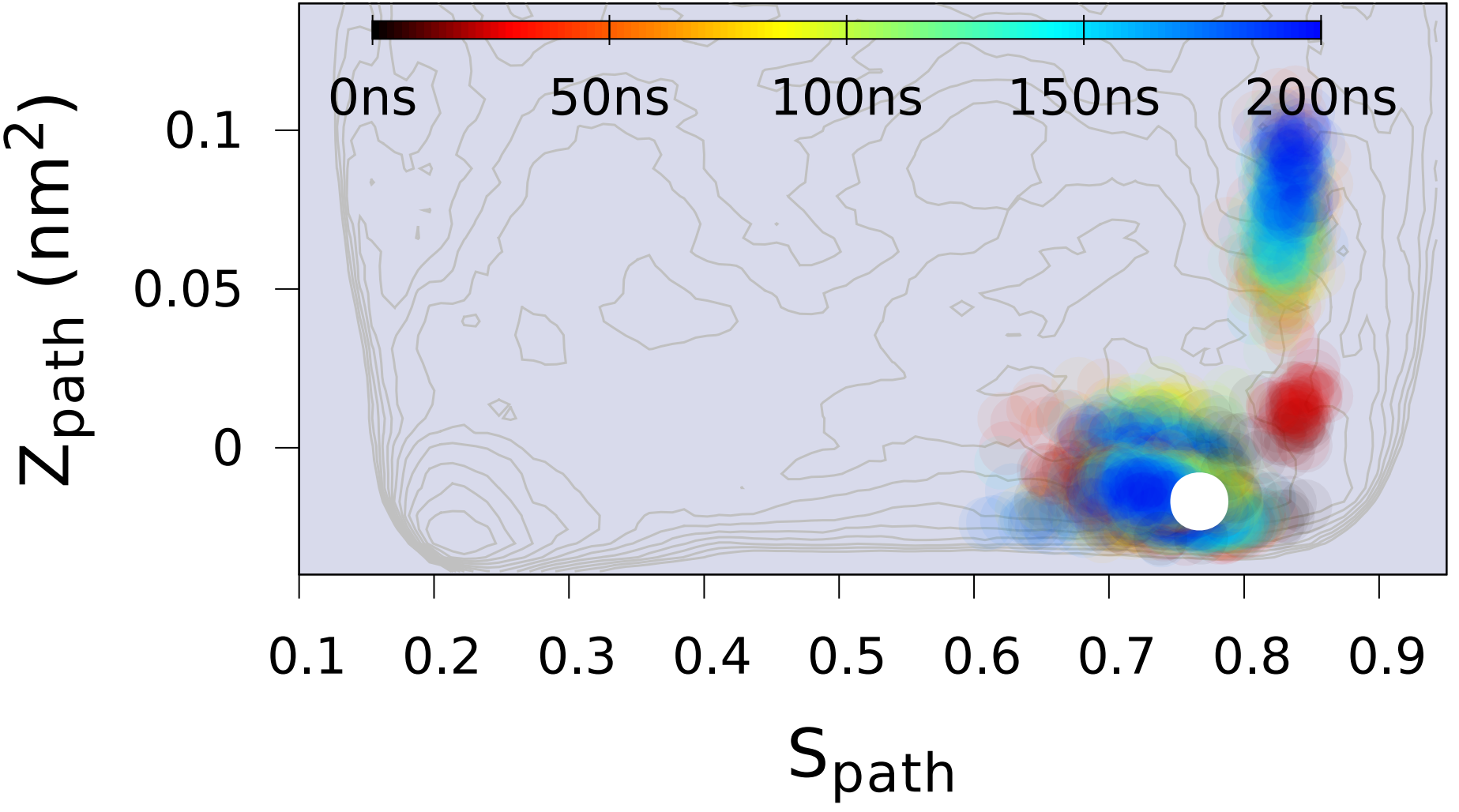
(**A**) CHARMM22*,4 replica,$\epsilon_{CS} = 24$

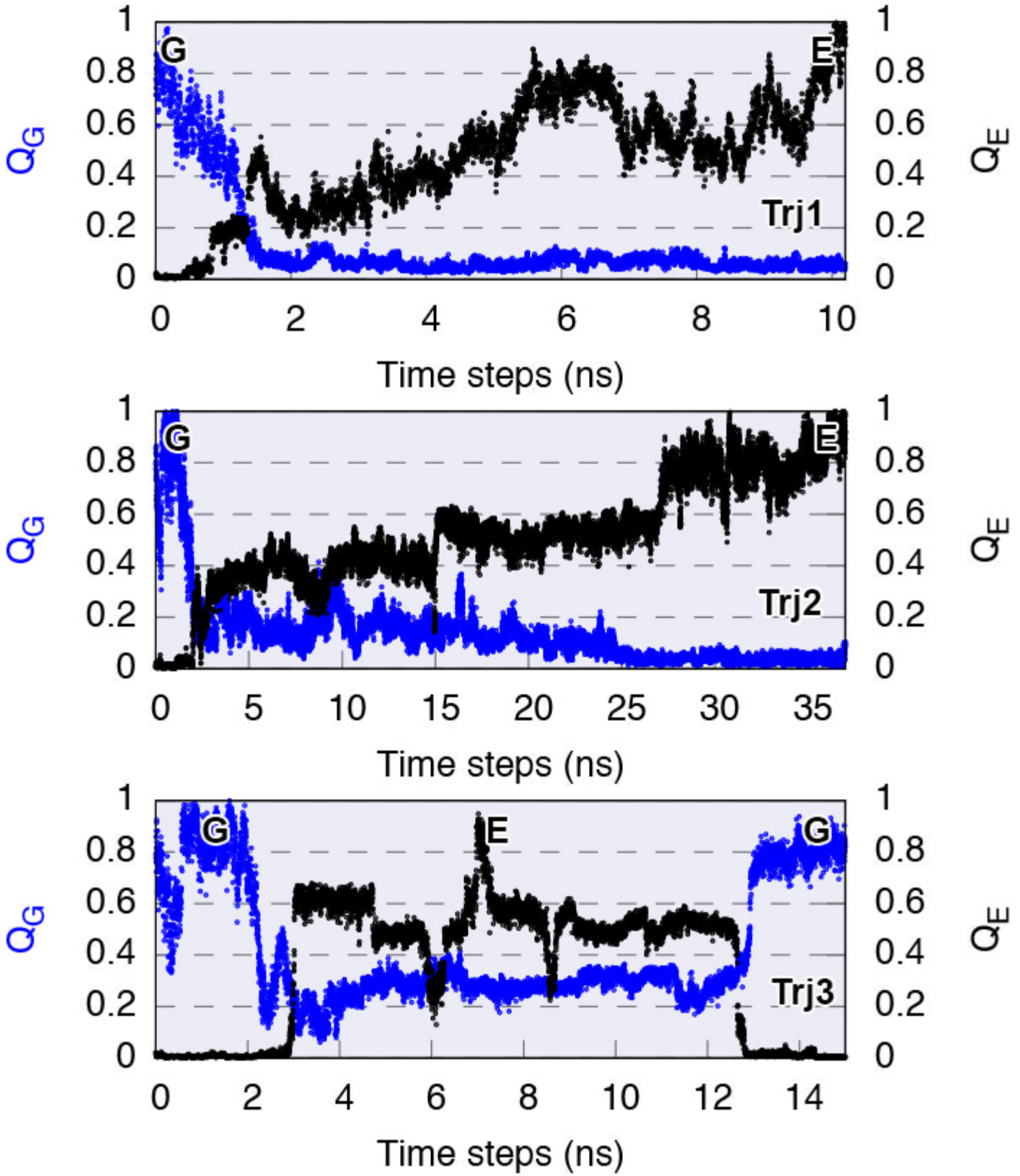


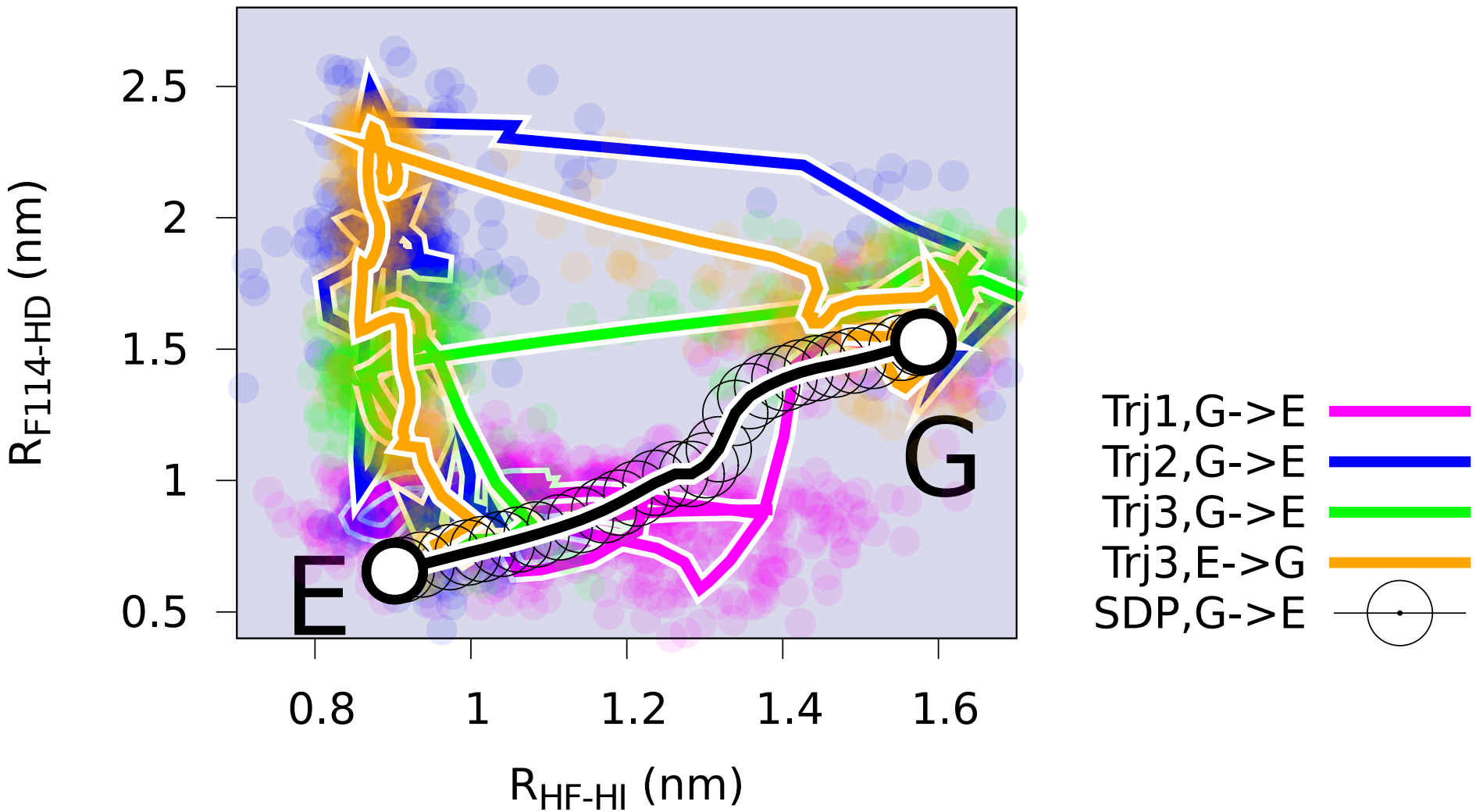
(**B**) CHARMM22*,2 replica,$\epsilon_{CS} = 24$

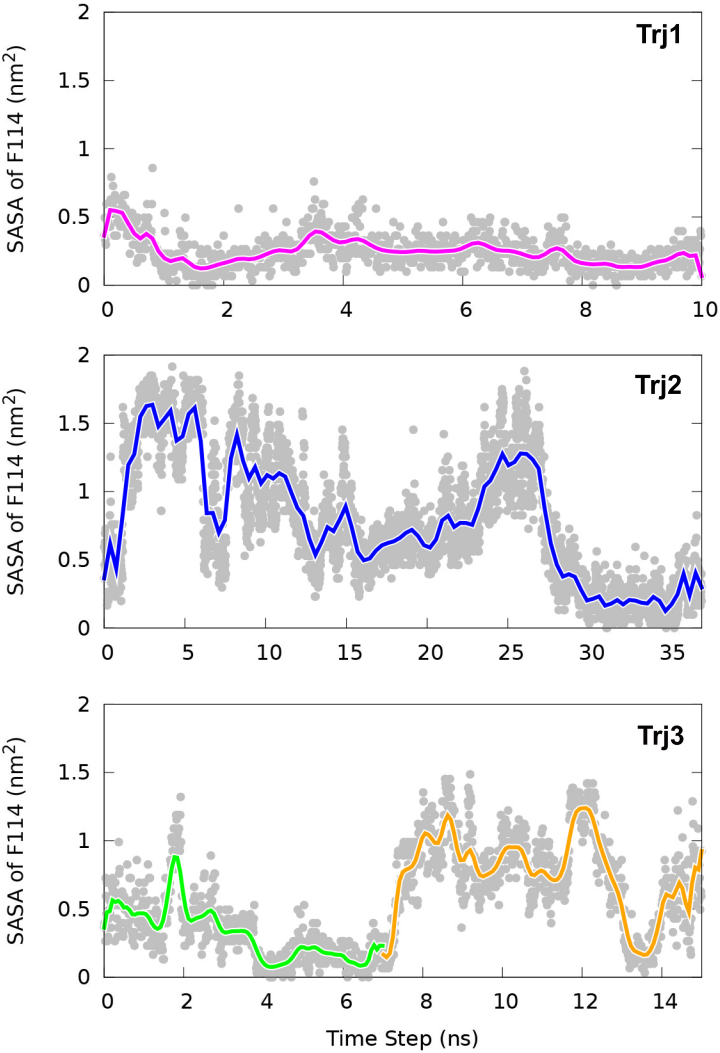


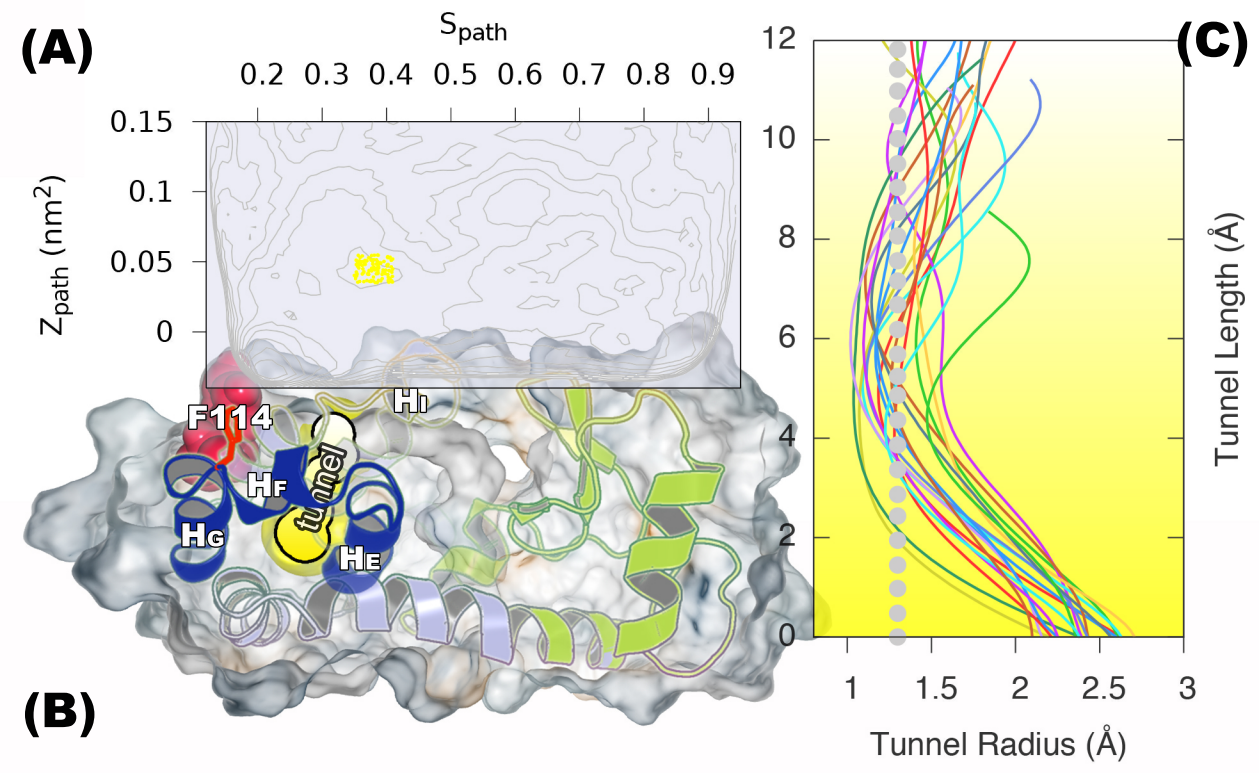
(**C**) CHARMM22*,2 replica,$\epsilon_{CS} = 12$

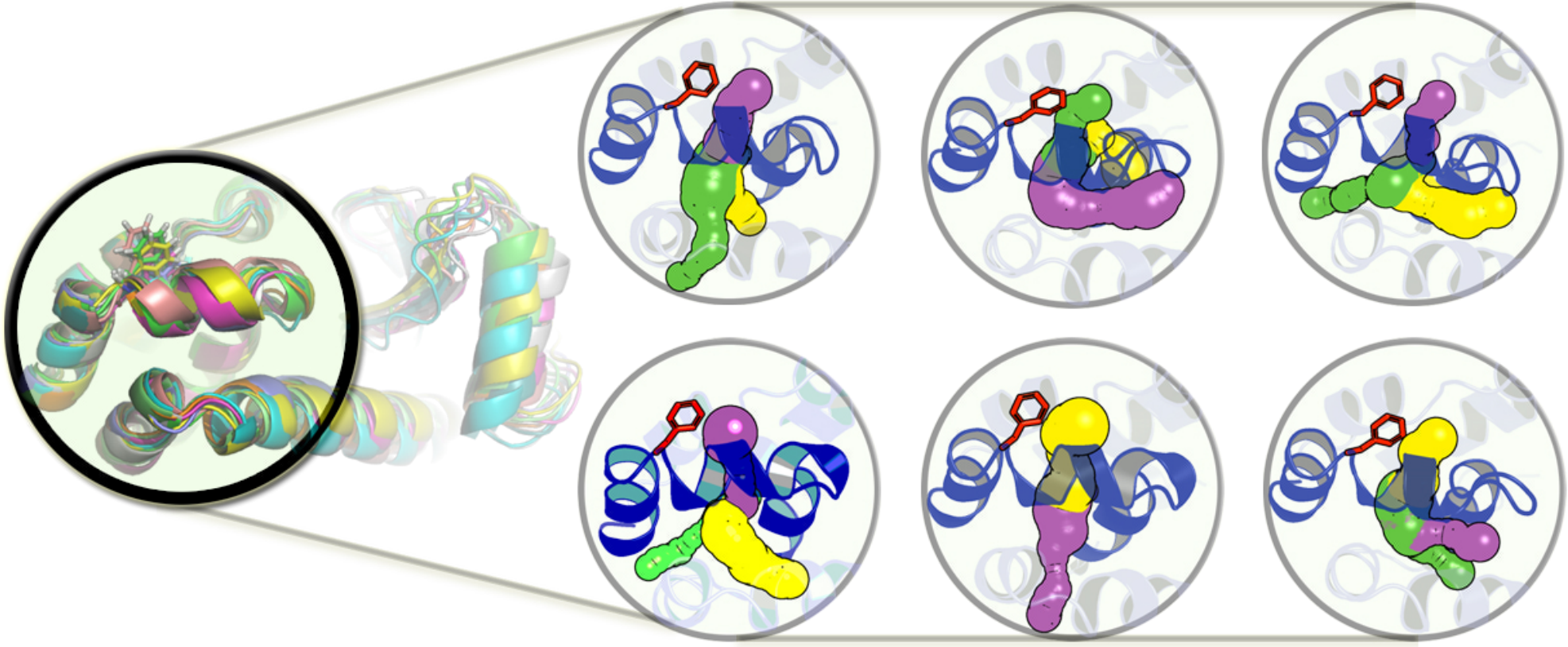

0ns
50ns
100ns
150ns
200ns
0.1
0.05
0
Z_{path} (nm^2)
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
S_{path}

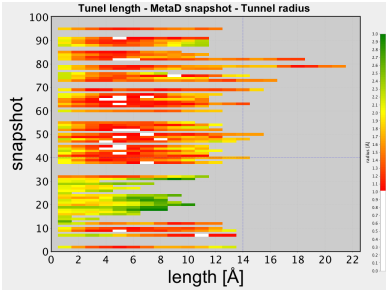

G
E
Trj1
Q_G
Q_E
Time steps (ns)
G
E
Trj2
Q_G
Q_E
Time steps (ns)
G
E
G
Trj3
Q_G
Q_E
Time steps (ns)


2.5
2
1.5
1
0.5
$R_{F114\text{-}HD}$ (nm)
0.8
1
1.2
1.4
1.6
$R_{HF\text{-}HI}$ (nm)
E
G
Trj1,G->E
Trj2,G->E
Trj3,G->E
Trj3,E->G
SDP,G->E

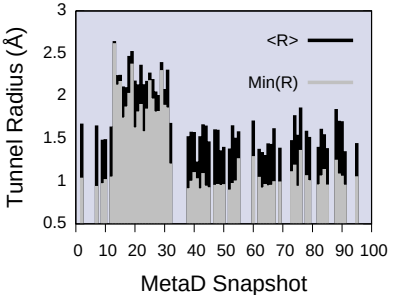

Trj1
Trj2
Trj3
SASA of F114 (nm^2)
Time Step (ns)
0
0.5
1
1.5
2
0
2
4
6
8
10
0
5
10
15
20
25
30
35
0
2
4
6
8
10
12
14


(A)
S_{path}
0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
0.15
0.1
0.05
0
Z_{path} (nm^2)
F114
H_I
H_F
tunnel
H_G
H_E
(B)
(C)
12
10
8
6
4
2
0
1 1.5 2 2.5 3
Tunnel Radius (Å)
Tunnel Length (Å)

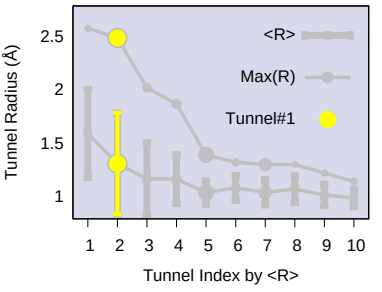




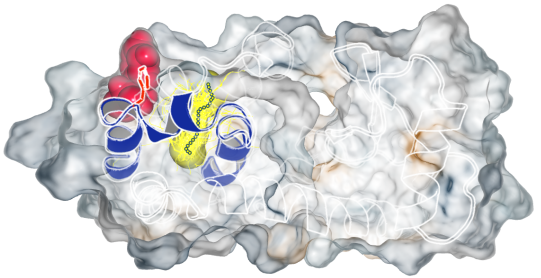
(**A**) heat map



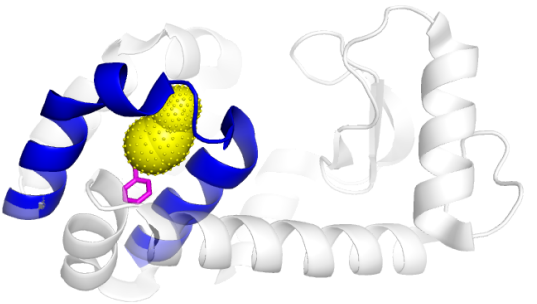
(**B**) Radius



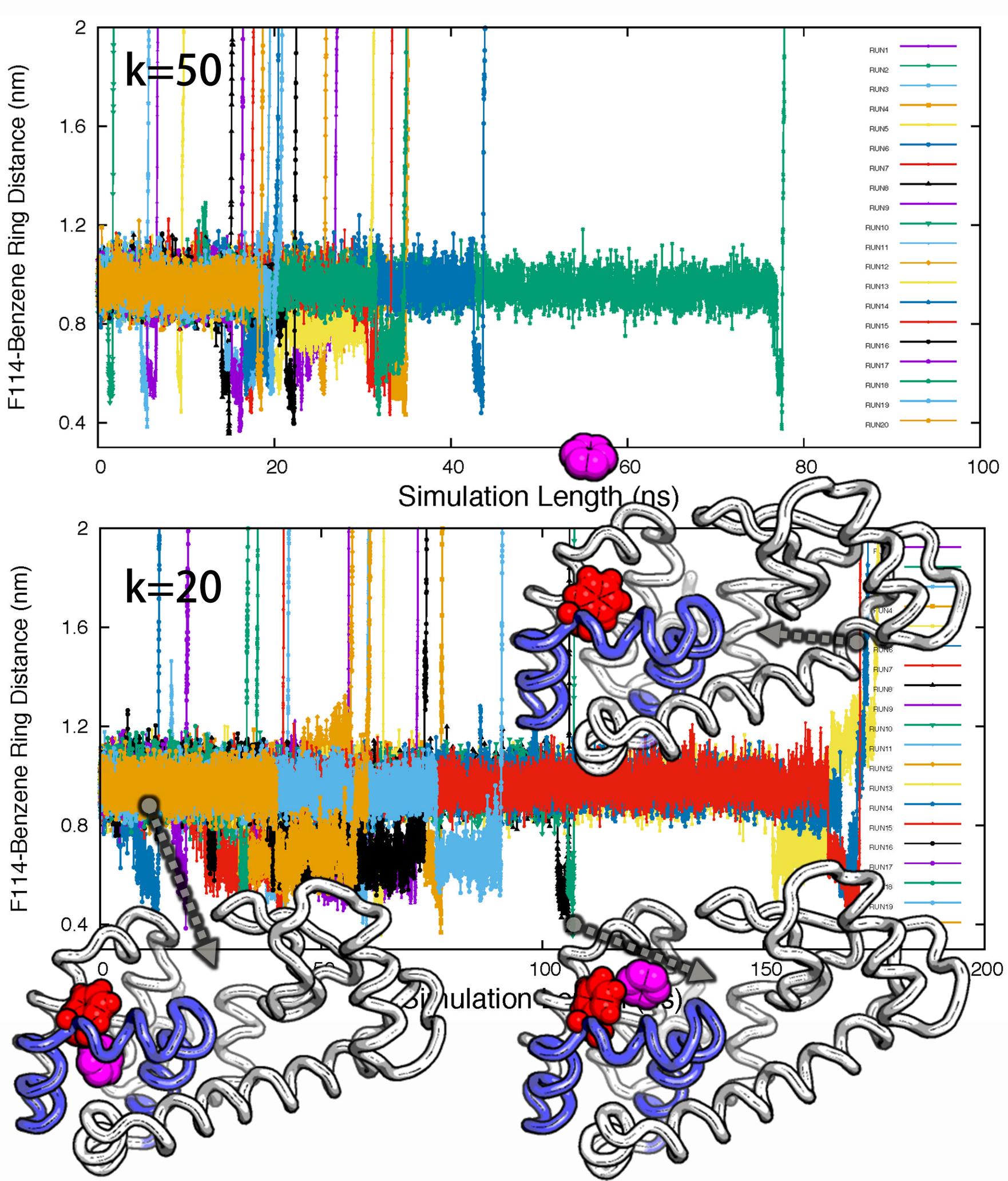
(**C**)



(**D**) InMetaD



(**E**) PDB:4W59


k=50
F114-Benzene Ring Distance (nm)
Simulation Length (ns)
RUN1
RUN2
RUN3
RUN4
RUN5
RUN6
RUN7
RUN8
RUN9
RUN10
RUN11
RUN12
RUN13
RUN14
RUN15
RUN16
RUN17
RUN18
RUN19
RUN20
k=20
F114-Benzene Ring Distance (nm)