Article

Markov State Models Reconcile Conformational Plasticity of GTPase with Its Substrate Binding Event

Bhupendra R. Dandekar, Navjeet Ahalawat, Suman Sinha, and Jagannath Mondal*

Cite This: *JACS Au* 2023, 3, 1728−1741

Read Online

| ACCESS | | Metrics & More | | Article Recommendations | | SI Supporting Information |

ABSTRACT: Ras GTPase is an enzyme that catalyzes the hydrolysis of guanosine triphosphate (GTP) and plays an important role in controlling crucial cellular signaling pathways. However, this enzyme has always been believed to be undruggable due to its strong binding affinity with its native substrate GTP. To understand the potential origin of high GTPase/GTP recognition, here we reconstruct the complete process of GTP binding to Ras GTPase via building Markov state models (MSMs) using a 0.1 ms long all-atom molecular dynamics (MD) simulation. The kinetic network model, derived from the MSM, identifies multiple pathways of GTP en route to its binding pocket. While the substrate stalls onto a set of non-native metastable GTPase/GTP

encounter complexes, the MSM accurately discovers the native pose of GTP at its designated catalytic site in crystallographic precision. However, the series of events exhibit signatures of conformational plasticity in which the protein remains trapped in multiple non-native conformations even when GTP has already located itself in its native binding site. The investigation demonstrates mechanistic relays pertaining to simultaneous fluctuations of switch 1 and switch 2 residues which remain most instrumental in maneuvering the GTP-binding process. Scanning of the crystallographic database reveals close resemblance between observed non-native GTP binding poses and precedent crystal structures of substrate-bound GTPase, suggesting potential roles of these binding-competent intermediates in allosteric regulation of the recognition process.

KEYWORDS: GTPase, GTP, Ligand binding, Markov state model, Kinetics, Non-native
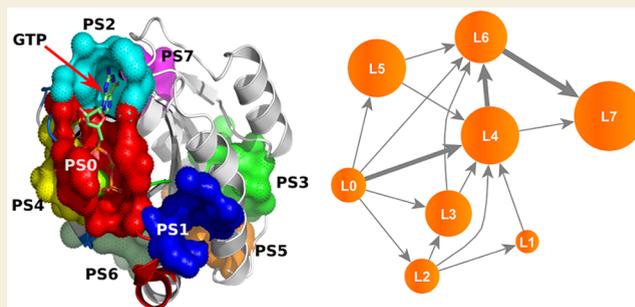
INTRODUCTION

GTPases are a large family of hydrolase enzymes that bind to the nucleotide guanosine triphosphate (GTP) and hydrolyze it to guanosine diphosphate (GDP). The GTP binding and hydrolysis takes place in the highly conserved P-loop "G domain", a protein domain common to many GTPases. Ras superfamily of proteins are class of small G proteins.[1] Ras superfamily GTPases act as molecular switches that generally cycle between a GDP-bound "off" state and a GTP-bound "on" state.[2] The conformational changes associated with these different molecular states are involved in the regulation of multiple cellular processes.[3] Small GTPases function as molecular regulators and control a wide spectrum of cellular activities, such as cell proliferation and cell survival.[4] Ras family proteins are generally observed in a monomeric form with molecular weight of 20−40 kDa,[2] and they contain five α helices (A1−A5), six β strands (B1−B6), and five loop regions (G1−G5)[5] (Figure 1). Contrary to the general trends, where amino-acid sequence is conserved in β-sheets and α-helical regions of the proteins, in case of the GTPases, the most sequence-conserved locations are the loop regions.[1] The P-loop (loop G1) is the structurally most stable location, and it does not undergo any conformational change during GTPase activity.[2] On the
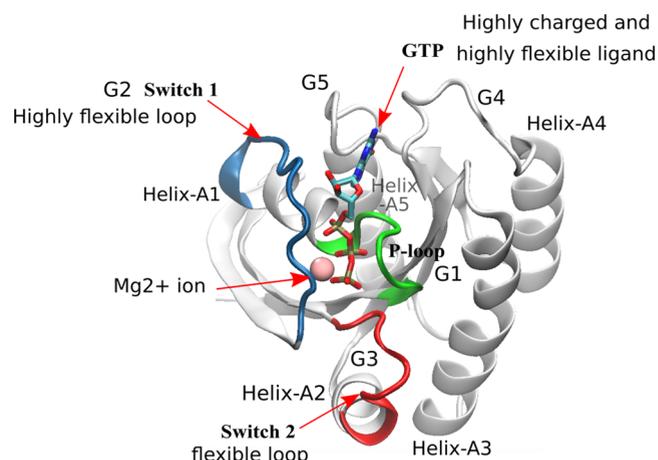
Figure 1. Systems under investigation: GTPase/GTP system, GTP bound crystal structure of P21ras GTPase (PDB: 1QRA).

other hand, switch 1 (loop G2) and switch 2 (loop G3) that surround the γ-phosphate group of the nucleotide show major conformational changes during nucleotide exchange.[6]

The first mutated Ras gene was detected in human cancer in 1982.[7] There are three types of Ras genes which are found to be most frequently mutated genes; these are H-Ras, K-Ras, and N-Ras.[8] Among them K-Ras (85%) is the most frequently mutated gene followed by N-Ras (12%) and H-Ras (3%).[9] The most common oncogenic mutations are observed in the "P-loop" region, residues 12, 13, and 61.[4,10] Despite significant efforts to directly target Ras activity, no anti-Ras drug has been developed and taken into the clinic. Being a nightmare for drug developers, it has been believed for a long time that they are "undruggable" due to the high binding affinity of GTP to Ras and relatively smooth structure of Ras proteins devoid of any deep pocket where a small molecule could bind.[11] More recently, there has been reported the discovery of multiple covalent binders[12–14] such as MRTX849[15,16] which are in clinical trials. Moreover sotorasib[17,18] has become the first FDA approved covalent binder at the switch-2 pocket.

Illustrating the mechanism of GTP binding to GTPase with atomistic detail and information on metastable binding hotspots are in demand. Such an attempt would potentially prompt a revisit to the design principle of new small molecules which can perturb the bound GTP ligand at the active site via allosteric networks or pathways. Moreover the potential metastable hotspots that would be identified in this process can be used for site-directed mutagenesis studies for further modulating the GTP binding mechanism. The present work aims to elucidate the atomistically detailed mechanism of GTP binding to "undruggable" Ras GTPase by combining ∼0.1 ms long unbiased classical molecular dynamics (MD) trajectories with Markov state model (MSM) analysis.[19] MD simulation has remained a method of choice for obtaining an atomistic account of biomolecular recognition processes.[20,21] However, the stochasticity involved in conformational fluctuation of bio-molecular systems needs to be addressed in more quantitative manner. Moreover, the slow event of ligand binding, from solvent to the native pocket of the receptor, is often difficult to capture in an individual time-continuous trajectory. In this regard, adaptively spawned short simulation trajectories, when integrated with MSM, have elucidated statistically robust accounts of state-to-state transition in a recognition event.[22–26] In the present investigation, we implement a more comprehensive scheme to decipher the connection between GTP recognition and conformational transitions. While our simulation captures rare spontaneous GTP-recognition events, the overall process displays the signatures of conformational plasticity. A close investigation reveals the role of flexible switches in modulating the binding event, and the GTP-recognition happens via multiple plausible transition pathways which involve the existence of a set of metastable binding-competent states around the ligand. As a key contribution, the investigation quantitatively dissects the inherent conformational plasticity in the small GTPase and shows how it would influence the GTP recognition process.

■ METHODS AND MATERIALS

Unbiased Binding MD Simulations

The X-ray crystal structure of small GTPase p21ras in complex with GTP (PDB ID 1QRA) is used for conducting unbiased MD simulation. The ligand was removed from the active site, and the protein without the substrate was used as the starting point of all binding simulations.

All unbiased MD simulations were performed using the GROMACS 2018 simulation package[27] using a leapfrog integrator with a time step of 2 fs. The binding simulations were initiated with the substrate-free GTPase placed at the center of a cubic box of dimension 7.1 nm in each direction with explicit solvent. The minimum distance between the protein surface and the box was 10 Å. The system was solvated with 10897 water molecules, and 8 copies of Mg^{2+} were added to the system to keep sufficient magnesium ion concentration. Two GTP molecules were placed at random positions and at random orientations in the bulk solvent media of the simulation system (far away from the protein surface) by using the combination of *insert-molecules* and *editconf* modules of GROMACS software. This corresponds to a GTP concentration of 11.98 mM. There was no direct contact between GTP and the protein at the onset of the simulation. The final system had a total 35406 atoms. The all-atom CHARMM36m[28] force field was used for protein and ions with the backbone CMAP correction. CHARMM parameters was used for GTP and ions while the water molecules were modeled via TIP3P water model.[29] The simulation was performed in the NPT ensemble at an average temperature of 300 K using the Nose−Hoover thermostat[30] with a relaxation time of 1.0 ps and at 1 bar constant pressure with Parrinello−Rahman barostat[31] with a coupling constant of 5.0 ps. The Verlet cutoff scheme[32] was employed with a minimum cutoff of 1.2 nm for the Lennard-Jones interaction and short-range electrostatic interactions throughout the simulation. Long range electrostatic interactions were treated by the Particle Mesh Ewald[33] (PME) summation method. All bonds connected to hydrogen atoms were constrained using the LINCS algorithm.[34] The bonds and the angles of TIP3P[29] water molecules were constrained using the SETTLE algorithm.[35] All the independent simulations were started from different configurations by randomly inserting the GTP molecules in the simulation box. All the particles of each of the configurations are assigned with random velocities. We carried out multiple independent multi-microsecond-long unbiased simulations. The substrate-binding process was verified by computing the Root Mean Squared Deviation (RMSD) of GTP between each simulated conformation of MD-derived trajectories and the crystal structure, after removing the rotational and translational motions of protein. The simulation lengths of the longer binding trajectories were 2.6 μs, 10.2 μs, and 13.6 μs with a total simulation time of 26.4 μs.

From our primary analysis of long unbiased GTP binding trajectories, we observed that both GTP and GTPase are modulating conformations and orientations of each other. For exhaustive sampling of the ligand movement and protein conformational space, we adaptively spawned numerous copies of short-length trajectories. Figure S1 provides the adaptive sampling scheme employed in this work. These simulations were initiated after clustering the simulated trajectory frames into a set of cluster centers. We used the regular space clustering[36] algorithm for this purpose. For sampling protein conformations, we used root-mean-squared deviations of switch 1 (residue id 25−36) (denoted here as RMSD1) and of switch 2 (residue id 59−67) (denoted here as RMSD2), after fitting rest of the protein backbone, as metric in the clustering algorithm (see Figure S1A). Similarly, for sampling the ligand native and non-native states, we used a distance cutoff based contact matrix as metric in the clustering algorithm (see Figure S1B). We iteratively performed sampling along both protein and ligand phase space until the explored FES converged to a final landscape. In total, we conducted 401 short simulations each for sampling protein and ligand conformations (i.e., total 802 trajectories) and each of these adaptively sampled simulations was 100 ns in length. These short trajectories, with a cumulative length of 80.2 μs were utilized to improve the statistics of the underlying Markov state model (MSM, discussed later). Overall, an aggregate of about 106 μs of unbiased trajectories was simulated. All binding simulations benefitted from usage of a graphics processing unit (GPU) in the in-house computing facilities.

Development of the Markov State Model

We initially attempted to build a comprehensive MSM for resolving both ligand states and protein conformations simultaneously. But our initial attempts in this regard were not very fruitful as there was mixing

among the states. As would be elucidated later, this is mostly because protein conformational dynamics and ligand binding dynamics in this GTPase/GTP complex involve different time scales, which are difficult to reconcile in a single model. Subsequently, we built two MSMs in this investigation: One of the MSMs was focused on resolving the GTP binding dynamics from its unbound state to its native bound pose. The other MSM involved dissecting the key metastable states that GTPase adopts. We have used PyEMMA software[37,38] for MSM construction. The MSM was performed using the relative location of one of the two GTP molecules. In the case of the long-time-scale binding trajectories, we considered the specific copy of the GTP ligand which had either eventually bound to the native pocket or had approached the binding pocket. For the rest of the short-length adaptively sampled simulations, we randomly selected either of the two GTP ligand copies, to avoid any bias. For the MSM involving ligand dynamics, we have used the residue-wise protein−ligand contact matrix as an input feature, which is based on a 0.6 nm cutoff on the pairwise minimum distances between any protein residue heavy atom with any ligand heavy atom. Thus, each frame of the trajectories was converted into a row vector with 166 elements, each corresponding to a residue−ligand pairwise distance of the 166-residue long protein. To further increase resolution of the underlying binding kinetics and simplify the complexity of the raw dimensions, we used the time lagged independent component analysis (TICA)[39,40] as a dimensional reduction technique. We finally reduced the 166 dimensional input into the top 10 (slowest) TICA dimensions and used them as an input for discretization and MSM construction. Figure S2a shows the MSM reweighted Free Energy Surface (FES) along the first two TIC dimensions. The free energy landscape in Figure S2a shows the projection of dimensionally reduced TICA data along the slowest two TICA-generated dimensions. The FES along this latent TICA space gives us an idea about the number of local minima present and the energetic barrier among them. The FES provided in Figure S2a highlights (and is also annotated with) the locations of 7 distinct free energy basins. We applied a k-means clustering algorithm[41] to discretize the trajectories into 100 microstates. We then built a 100 microstate MSM at varying lag times. Finally, a lag time of around 100 steps (i.e., 10 ns) was chosen as the minimum value of lag time at which the implied time scale (ITS) becomes flatter as a signature of the Markovian nature of the model (see Figure S2b. The chosen lag time suggests the presence of eight time-scale-separated MSM macrostates. In order to coarse-grain the 100-microstates into eight macrostates, we used the Perron Cluster Cluster Analysis (PCCA+)[42,43] method. Figure S1c shows the state map decomposition of the phase space into 8 macrostates along the top two TICA dimensions (TIC1, TIC2). We have also validated the construction of this 8-macrostate MSM by performing the Chapman−Kolmogorov test[44] with Bayesian error estimation (see Figure S3B).

We obtained the stationary state populations for each of the macrostates from the MSM and estimated the binding free energy using following relation:

$$\Delta G = -RT \ln\left(\frac{\pi_{\text{bound}}}{\pi_{\text{unbound}}} \frac{V_{\text{unbound}}}{V_0}\right) \tag{1}$$

where π_{bound} is the stationary state population of the native-bound state, π_{unbound} is the stationary state population of the unbound (ligand in bulk) state. Similarly, V_{unbound} is the simulated solvent volume and $V_0 = 1.663$ nm^3 is the volume at standard temperature and pressure. The ratio of these volumes corresponds to the standard state correction into the ligand binding free energy.

In Table 1, the free energies reported for each of the ligand bound metastable macrostates (S1 to S7) are computed relative to the unbound population (state S0).

The kinetic parameters were calculated from the mean first passage time (mfpt), where the on rate and off rates are defined as

$$k_{\text{on}} = 1/(\text{mfpt}_{\text{on}}C); \; k_{\text{off}} = 1/(\text{mfpt}_{\text{off}}) \tag{2}$$

where C is the concentration of GTP molecules in the simulation box.

Similarly, the dissociation constant (K_d) is defined as

Table 1. Equilibrium Populations of the Key Ligand Binding Competent States Obtained from the MSM

binding competent states	populations (%)	binding free energy (kcal/mol)
State L0 (Unbound)	7.30 ± 0.04	0
State L1	1.52 ± 0.11	−2.25 ± 0.16
State L2	9.41 ± 0.08	−3.35 ± 0.03
State L3	11.24 ± 0.01	−3.45 ± 0.02
State L4	18.17 ± 0.07	−3.74 ± 0.03
State L5	16.13 ± 0.13	−3.67 ± 0.04
State L6	14.02 ± 0.01	−3.59 ± 0.02
State L7 (Bound)	22.18 ± 0.04	−3.86 ± 0.02

$$K_d = \frac{k_{\text{off}}}{k_{\text{on}}} \tag{3}$$

We employed transition path theory (TPT)[22,45,46] to obtain the net transition pathways and their fluxes. Here, we will discuss it briefly and direct readers to other sources[22,45,46] for additional details on how to estimate paths and fluxes. TPT analysis provides a flux network from source state A to sink state B that passes through intermediary states (nodes). Each state (node) has a stationary probability, forward committor, and backward committor (see eq 4). Every edge (path) has an associated flux, which is usually net flux (see eq 5). The pathways and their associated weights (fraction) from A to B can be calculated using the flux network. The TPT analysis was carried out among the ligand macrostates by considering State L0 as the source state (A) and State L7 as the sink state (B). In the TPT calculations, the committor probability functions are used to define the direction of the transitions. The committor probability function q_+ is designated as the forward committor and q_- is designated as the backward committor. If the simulation is started at point z, the intermediate phase space point between path A and B reaches sink state B without visiting source state A, then the forward committor value is assigned to the phase space point (z). Similarly, If the simulation is started at point z and visits source A without visiting sink state B, then the backward committor is assigned to the process. For equilibrium processes, $q_- = 1 - q_+$. The gross flux, i.e., average number of transitions going from microstate $i \rightarrow j$ as a part of the A →B transition is computed as follows:

$$f_{ij} = q_{-,i}\pi_i P_{ij} q_{+,j} \tag{4}$$

where q_+ and q_- are forward and backward committor probabilities as mentioned before and $\pi_i P_{ij}$ is the equilibrium flux. Finally, the net fluxes are computed from the difference between forward gross flux and backward gross flux represented as follows:

$$f_{+,ij} = \max\{0, f_{ij} - f_{ji}\} \tag{5}$$

The dominant pathways and their percentages in the total binding were obtained using the *pyemma.msm.tpt.pathways* module provided in the pyEMMA[37,38] software, which makes use of eq 5 to estimate the maximum net flux occurring among each of these states (nodes) present between states A and B.

Similarly, for resolving protein conformational dynamics, we used the RMSD of the switch 1 and switch 2 regions of the protein relative to native conformation (denoted previously as RMSD1 and RMSD2) as the features for MSM construction. Figure S2d shows the MSM reweighted Free Energy Surface (FES) along RMSD1 and RMSD2. We then discretized the input phase space into 1000 microstates using the k-means clustering algorithm.[41] Based on the ITS plot, we built the MSM at lag time of 40 ns, i.e., 400 steps (see Figure S2e). Subsequently, based on ITS, a 5 macrostate coarse-grained MSM was built using the PCCA+[42,43] method. Figure S2f shows the state map decomposition of the phase space into 5 macrostates. We further validated the construction of this five-state MSM by performing the Chapman− Kolmogorov test[44] with Bayesian error estimation (see Figure S3B). Finally, in order to elucidate the mechanism of protein conformational changes, TPT[22,45,46] was employed to obtain the net transition
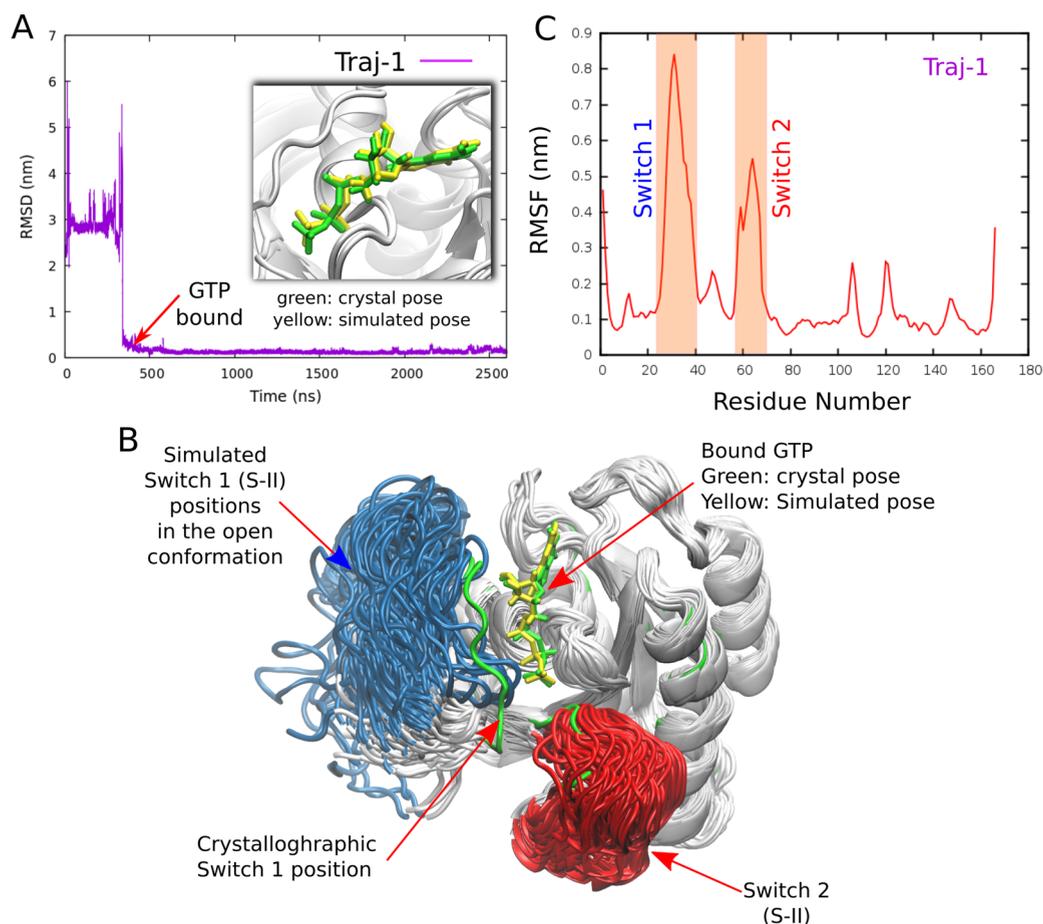
Figure 2. (A) A representative binding trajectory (Traj-1) of GTP binding to GTpase. The rmsd of GTP is calculated with respect to the bound crystal structure as a function of time. (B) Conformational plasticity of GTPase: The highly flexible switch 1 (S-1; Blue) and switch 2 (S-2; Red) in GTP binding trajectory (Traj-1) do not remain stable at crystallographic native pose. (C) The root-mean-square fluctuation (RMSF) analysis of the GTPase protein from the GTP binding trajectory (Traj-1).

pathways and the final rates of the transitions among the protein macrostates.

RESULTS

Spontaneous GTP Binding at the Native Pocket Captured by Unbiased Atomistic MD Simulations

Conformational flexibility of a protein in mediating protein−ligand binding has been demonstrated to be an integral factor in molecular recognition studies. However, the dual role of ligand flexibility in addition to protein flexibility in the overall recognition process has been relatively sparsely explored. GTPase/GTP (Figure 1) represents one such protein/ligand system in which both the protein (GTPase) and the ligand (GTP) are conformationally flexible.

In a bid to capture GTP in the act of finding and binding to its designated cavity on GTPase, we spawned three long unbiased MD trajectories which simulated the kinetic process of diffusive exploration of the ligand in the bulk solvent as well around the protein and ranged between several hundred nanoseconds and multiple microseconds. In one of these three MD trajectories, we discovered that GTP spontaneously identified the native binding site of GTPase at around 400 ns and got bound in a pose which is exactly the same as the crystallographic bound pose of the ligand (PDB ID 1QRA) (see zoomed view (inset) of Figure 2A). A time profile of RMSD of the simulated GTP with respect to its crystallographic bound-pose gradually went well

below 0.2 nm, indicating that the simulated GTP wanders into the aqueous media and around the GTPase and eventually gets bound into a pose that recapitulates the actual crystallographic pose (see Figure 2A).

However, a visual inspection of the bound simulated pose also indicated that while the ligand has already attained the crystallographic bound pose, the corresponding simulated conformation of the protein still deviates considerably from the crystallographic pose of the protein (see Figure 2B). In particular, the simulated trajectory shows that the highly flexible loops in the binding trajectory (see Figure 2C), namely, the switch 1 and switch 2 loops of GTPase, visit their respective native poses (or location) but do not remain stable in the crystallographic native pose. They were found to be relatively open and disordered relative to the crystallographic pose and remained in the open conformation for most of the simulation time as shown in Figure 3A. A residue-wise analysis of the root-mean-square fluctuation (RMSF) of GTPase of this binding trajectory reveals the higher flexibility of the switch 1 and switch 2 loops (see Figure 2C). From this analysis, it can be envisioned that movement of constituent residues of switch 1 (from residue 25 to 36) and switch 2 (from 59 to 67) would be the most crucial factors for guiding the GTP-binding process. Extending this simulation trajectory by one additional microsecond (far beyond the ligand-binding event, i.e., from 1.5 μs onward) did not make the protein conformation converge to its proper native
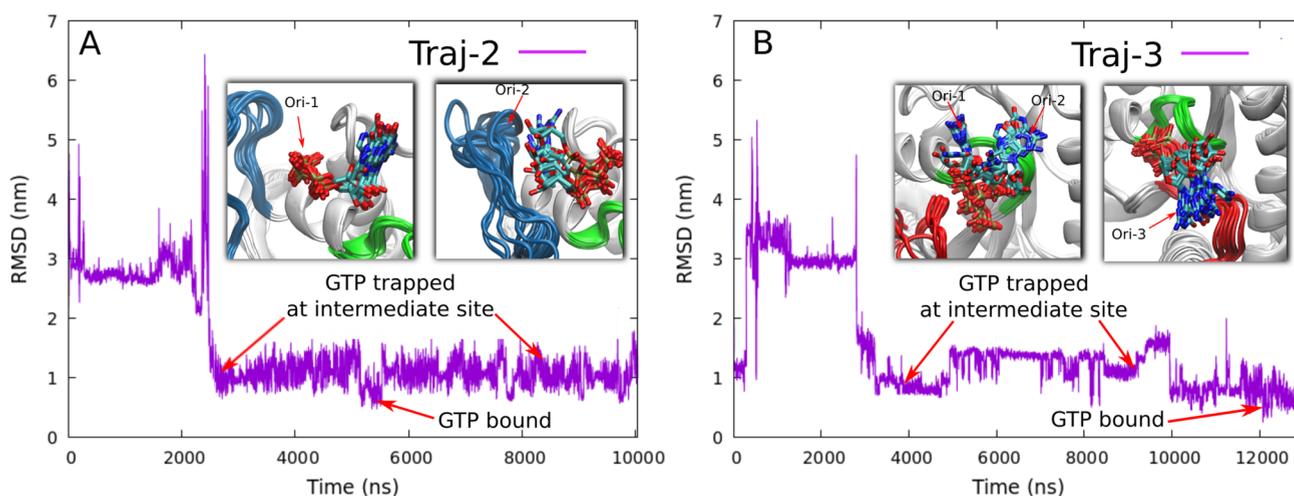
Figure 3. Two other representative long binding trajectories (A) Traj-2 and (B) Traj-3, where the GTP was trapped in metastable states for most of the simulation lengths but also got bound transiently. See GTP orientations-1, and -2 in the inset near switch 1 (S-1; Blue) and switch 2 (S-2; Red)).

crystallographic pose. This occurrence of the protein being trapped in a non-native conformation while the ligand has already attained the native crystallographic pose is the classic instance of protein conformational plasticity.[23,47]

The investigation of the other two long binding trajectories revealed that GTP explores different sites on the dynamically changing protein surface, which are chemically compatible pockets, and in this process of exploration GTP halts on different sites. In fact, in these extensively long trajectories (traj-2, 10.2 μs, and traj-3, 13.6 μs), the ligand became bound transiently during the simulation, before shifting away from the native pose (see Figure 3). Although the GTP came very close to the active site, with its RMSD attaining a value below 0.8 nm (see Figure 3A,B), it remained trapped in metastable states for most of the simulation period. The snapshots extracted from these two trajectories corresponding to these metastable states or locations (see Figure 3A,B insets) show that the nucleotide-based ligand has partially formed stable interactions with the native site residues, either using its semiflexible, polar guanosine head part (shown in Figure 3A as Ori-1, Ori-2) or using its highly flexible, charged triphosphate tail (shown in inset of Figure 3B as Ori-1, Ori-2, Ori-3). The RMSF of these two binding trajectories also shows higher flexibility of switch 1 and switch 2 loop residues (see Figure S5). It is also observed from trajectory visualization that the GTP easily switched back and forth between these metastable states but hardly attains the proper ligand binding pose, and this is due to the higher flexibility of the GTPase loops and the overall protein active site residues. None of the independent multi-microsecond MD simulation trajectories, in aggregate 26.4 μs, produced a crystallographically accurate protein native pose, but they were able to produce ligand-bound pose only.

The aforementioned efforts in capturing the GTP in the act of recognizing the protein via long time-scale all-atom simulations implied that the conformational flexibilities of both the protein (GTPase) and the ligand (GTP) create bottlenecks trapping the protein and the ligand in diverse non-native poses. The inherent conformational plasticity of the protein, as displayed in Figure 2, suggested that the protein-conformational as well as ligand-diffusional landscapes need to be exhaustively sampled, well beyond the long simulation trajectories. Accordingly, as illustrated in Figure S1 we undertook an endeavor of adaptively sampling the unexplored phase-space via initiating short-time-scale MD trajectories from various frames of the existing long-time-scale trajectories. In particular, many short independent trajectories were initiated from different intermediate states obtained after clustering of long binding trajectories by considering both protein conformation and ligand location (see Methods and Materials). A set of simulations were also initiated from the crystallographic native ligand-bound pose of the protein to ensure the sampling of this crucial location as well. The process of spawning new trajectories was iterated in multiple cycles so that the protein conformational and ligand binding landscapes were sufficiently explored. These short trajectories, with a cumulative length of 80.2 μs were utilized to improve the statistics of the underlying Markov state model (discussed later). Overall, an aggregate of about 106 μs of unbiased trajectories was performed. For a comprehensive understanding of the ligand binding in this flexible system, these cumulative data were used to build a kinetic network of transition among key binding-competent states within the framework of the MSM.

A Microscopic Network Model Identifies Native Bound State and Non-Native Binding Hotspots

We initially had aimed to develop a comprehensive statistical model which would connect the key ligand locations around GTPase with the possible native and non-native conformations of the protein. However, our initial attempts to resolve the kinetics by combining both protein conformation and ligand location around the protein were futile and resulted in intermediates which were spatially and temporally mixed. In fact there are precedents of such scenarios[48] and generally these are pathological signs of protein conformational plasticity. Accordingly, we decided to build two individual MSM-based kinetic networks by separately considering protein conformation and ligand-location. (See Methods and Materials)

The statistical model obtained by discretizing the trajectories based on the ligand location suggested a kinetic network involving eight key binding-competent states in which ligand locations were predicted (see Methods and Materials for details). Figure 4 annotates the representative snapshots covering a wide spectrum of each of the ligand bound states interpolating between two limiting cases: the GTP-unbound state (denoted here as State L0) in bulk solution and the perfectly bound native state (denoted here as State L7). Apart
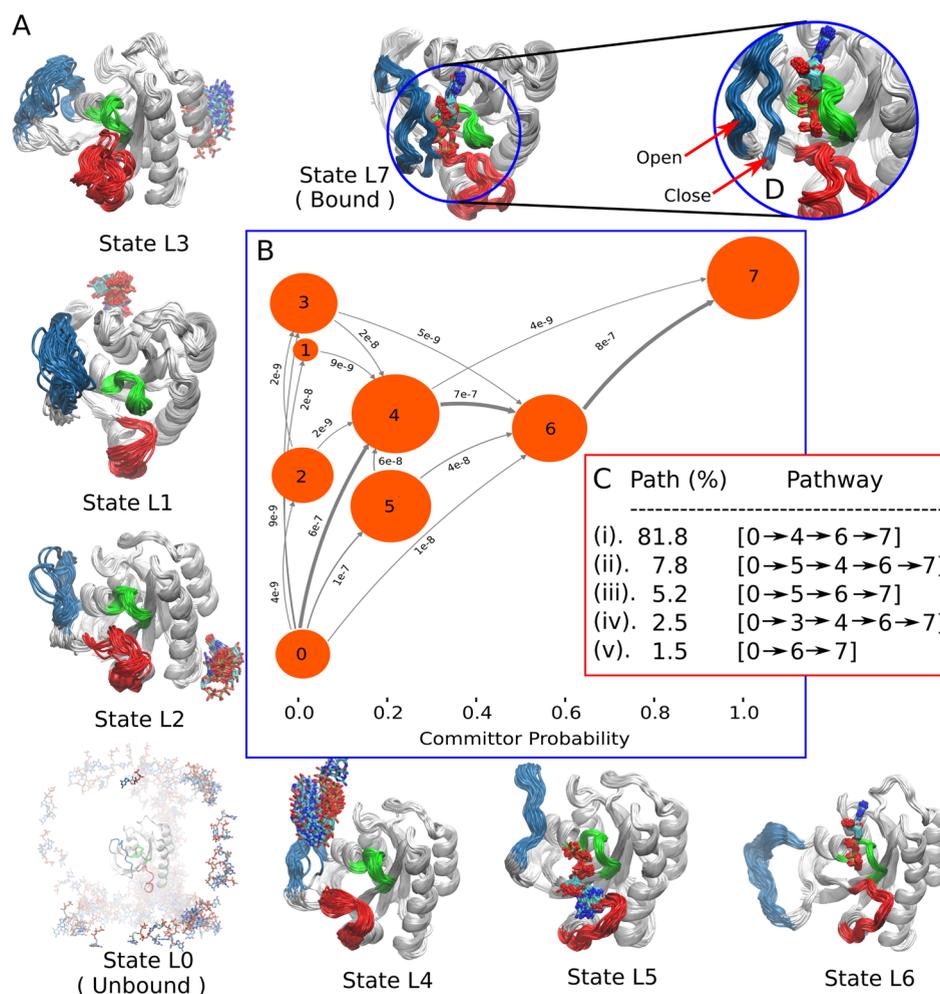
Figure 4. Kinetic mechanism of GTP binding to GTPase protein: The net flux network is obtained from TPT (Transition Path Theory) analysis. These TPT calculations were done by considering State L0 as the source state and State L7 as the sink state. (A) Eight distinct representative snapshots of the ligand macrostates obtained from MSM along with their corresponding protein conformations. Only 50 snapshots of the ligands are shown for simplicity. (B) The network connects eight macrostates that are represented as a circular discs. The sizes of the discs are proportional to their stationary populations (see Table 1) and are placed according to their respective committor value. The committor scale is drawn at the bottom of the panel. The committor value of 0 corresponds to the unbound State L0 and value of 1 corresponds to the bound State L7, while the committor values between 0 and 1 correspond to the intermediate states. The states are connected with the arrows and whose thickness is proportional to the amount of net flux of the transition within the macrostates. The flux values of corresponding transitions are written over the arrow. (C) The resulting major pathways along with their path percentage. (D) The enlarged view of State L7 showing intrinsic flexibility of switches.

from the solvated and native-bound pose, the kinetic model also predicts the presence of a set of metastable intermediates (denoted here as State L1 to State L6) each localized on different parts of the dynamically changing protein surface.

Figure 5 shows the key interactions of each of the GTP bound metastable intermediate states and Table S1 reports the type of interactions and interacting residues from these non-native sites. In general, we observe that in all of the macrostates GTP is mainly stabilized by strong polar interactions like hydrogen bonds, salt bridges, and cation−π interactions between polar and charged residues of the protein sites (see Figure 5 and Table S1). In the native bound pose State L7 (Figure 4), GTPase attains its perfect native conformation, where switch 1 and switch 2 are positioned in their stable native locations. However, we observe that there exists an intrinsic flexibility in the proper native conformation. In the native ligand-bound pose of State L7, GTPase is found to be present in dynamic equilibrium of two conformations similar to the GDP-bound "off-open" state and GTP-bound "on-closed" state (see Figure 4D; enlarged view of

State L7). These are reminiscent of the GTPase conformations denoted as "STATE-1" and "STATE-2" by Geyer and co-workers:[49,50] in STATE-1 the Tyr32 residue is oriented outside or away from the GTP phosphate tail (see Figure 5H), while in STATE-2 the Tyr32 residue is oriented inside (see Figure 5G), toward the GTP phosphate tail also called "Tyr32-out" and "Tyr32-in" state.[49,50] These two orientations of Tyr32 are predominantly observed in many GDP bound open, and GTP or GTP-derivative bound closed GTPase structures are manifested after the Thr35 from switch 1 and Gly60 from switch 2 lose the coordinated interaction with the manganese ion and the third phosphate of the GTP ligand after its hydrolysis. More interestingly, in simulations also it is found that the "Tyr32-in" conformation is present in higher population than the "Tyr32-out" conformation, as was observed in the polarization transfer ^{31}P NMR experiments by Geyer and co-workers.[50] More interestingly, in State L4, we observe that the GTP is simultaneously interacting with switch 1 via hydrogen bonding with Glu31 and also with the G5 loop via a salt-bridge
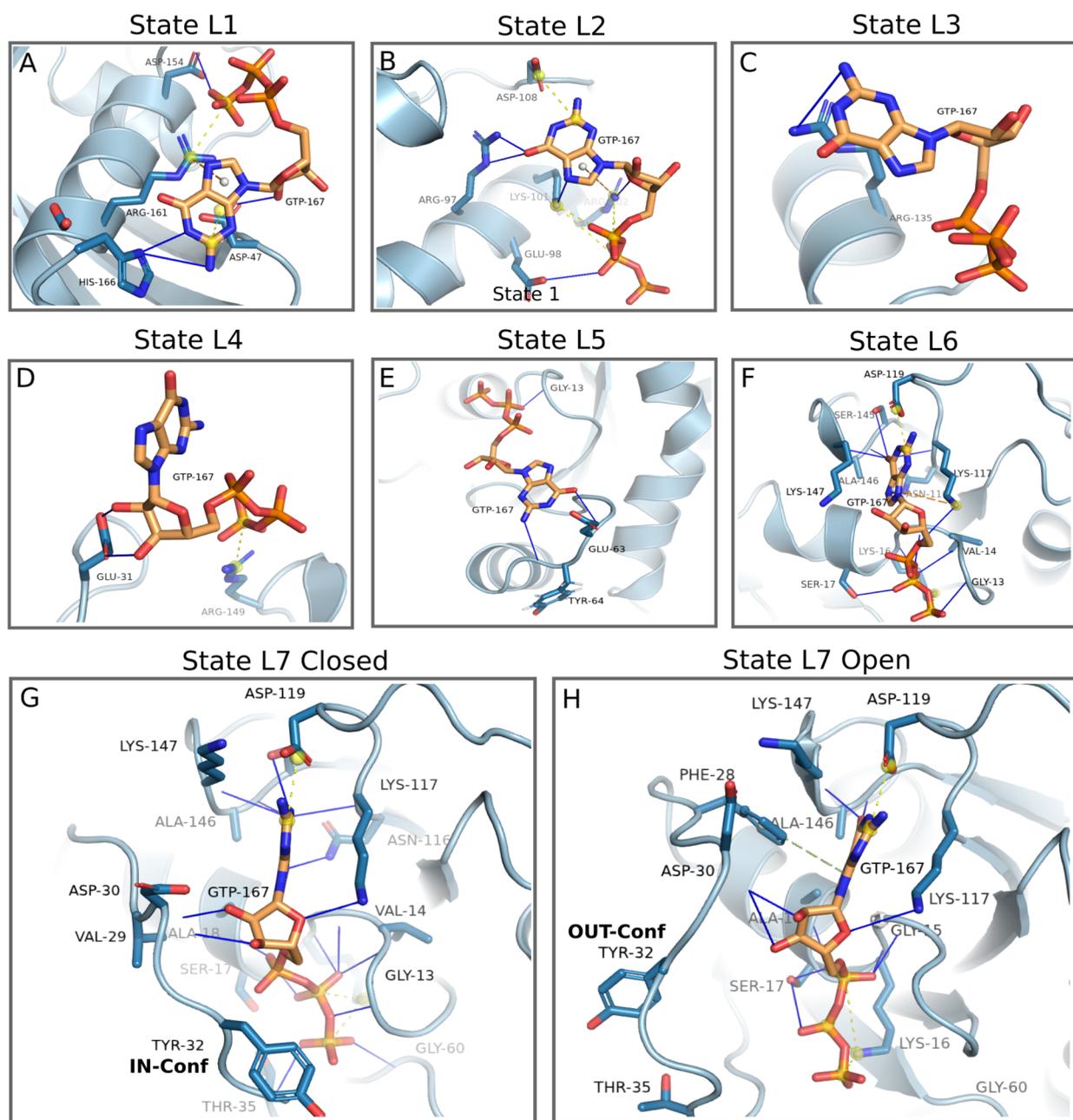
Figure 5. (A−H) Representative atomistic interactions of each ligand-bound metastable state. In all of the states, the GTP is mainly stabilized by strong polar interactions like hydrogen bonds, salt bridges, and cation−π interactions between polar and charged residues of switch 1 and switch 2 and the other protein sites. Hydrogen bonds are shown with blue lines, salt bridges are shown with yellow dashed lines, and cation−π interactions are shown with orange dashed lines. The interactions are evaluated using PLIP-server.[53]

interaction with Arg149 of the protein (see Figure 4D; enlarged view of State L7). Table 1 provides the equilibrium populations and ligand-binding free energies (relative to unbound state) of these 8 macrostates. Additionally, while the model correctly predicts the highest population for the native ligand-bound pose ascertaining a free-energy favorable GTP binding process, we find that (see Table 1) the population of this bound state (equivalent to 22.18%) is just relatively higher than other non-native intermediates. In particular, we observe that many of these non-native intermediates are considerably populated (10−18%), relatively comparably to the native bound state population, suggesting that these non-native macrostates would potentially mediate the GTP binding process. Overall,

this further shows inherent complexity of GTP binding dynamics due to its larger size and highly flexible, highly charged nature.

GTP Recognition Takes Place via Multiple Binding Pathways

We simulated the GTPase−GTP binding process in the absence of other GTPase activity regulators such as guanine nucleotide exchange factors (GEFs) and GTPase-activating proteins (GAPs). Therefore, we need to compare the kinetic constants of intrinsic GTPase activity only (i.e., studies conducted of the apoprotein or in absence of GTPase regulators). The binding on-rate constant, as predicted by the MSM, ($k_{on} = 0.1088 \times 10^6$ M^{-1} s^{-1}) is in good agreement (within one order) with the
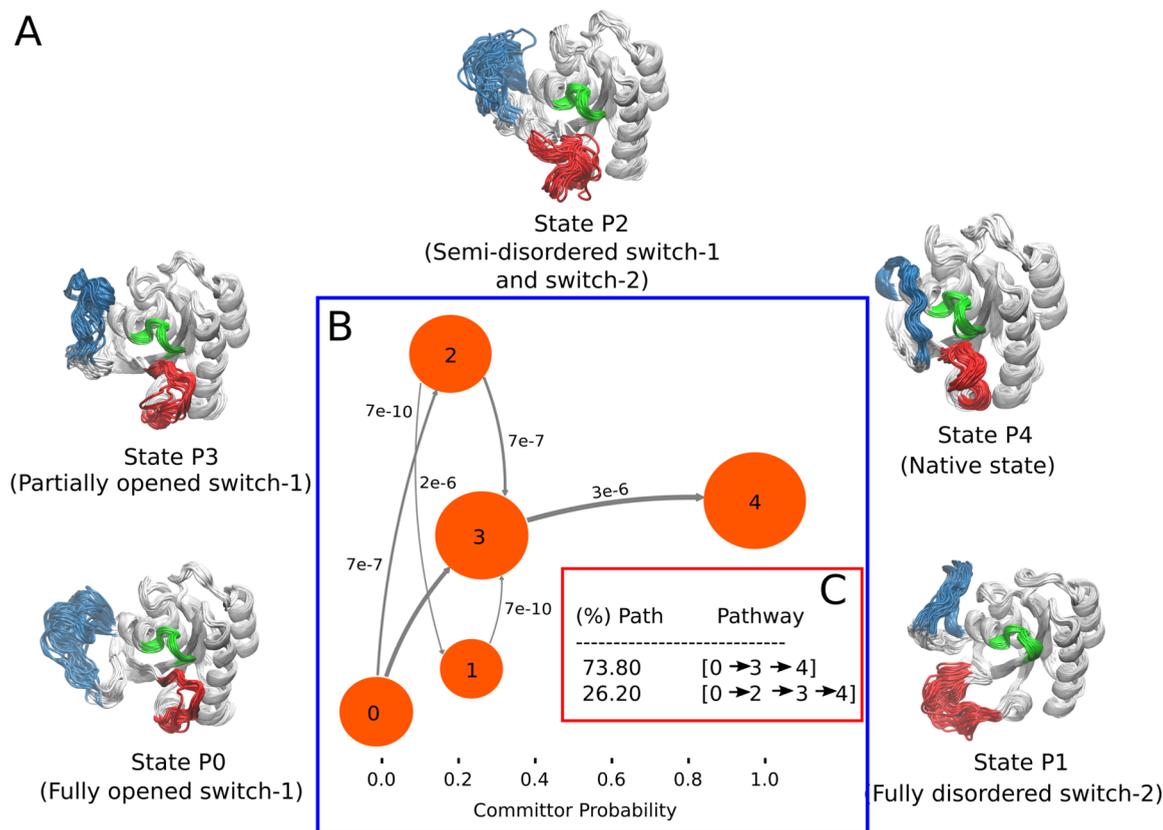
Figure 6. Kinetic mechanism of conformational plasticity of GTPase protein: The net flux network is obtained from TPT (transition path theory) analysis. These TPT calculations were done by considering State P0 as the source state and State P4 as the sink state. (A) Five distinct representative snapshots corresponding to most probable states for each of the protein macrostates. In each state, 50 snapshots are shown for simplicity. (B) The network connects five protein macrostates, States P0, P1, P2, P3, and P4. State P4 corresponds to the native state of the protein, while other states corresponds to non-native states. These states are represented as circular discs, and the sizes of the discs are proportional to their stationary population. The states are connected with the arrows, and the thickness of the arrows is proportional to the amount of net flux of the transition within the macrostates. The flux values of corresponding transitions are written over the arrow. (C) The resulting major pathways along with their path percentage.

experimental value $(k_{on} = 2.9 \times 10^6 \text{ M}^{-1} \text{ s}^{-1})$,[51] indicating adequate sampling of GTP movement around the protein. However, as displayed in eq 1 and eq 2, the MSM significantly overestimates the ligand dissociation constant (k_d) of the GTP, when compared to the experimentally reported k_d value of 6.0×10^{-10} M. This is because, the off-rate constant, as predicted by the MSM, $(k_{off} = 101.18 \text{ s}^{-1})$ is very far from the experimental value $(k_{off} = 8.0 \times 10^{-5} \text{ s}^{-1})$.[51] Since in our unbiased MD simulation set up, we were unable to observe (or sample) sufficient unbinding events of a highly charged, flexible tight binder like GTP, the MSM was unable to characterize this reverse unbinding process very well, and this could be the source of the discrepancy in the K_d and k_{off} values compared to the experimental values.[24,52]

The kinetic ligand binding network shows multiple recognition pathways of GTP to the GTPase protein (see Figure 4B,C). Each of these pathways shows connection from the unbound State L0 to the bound State L7 via diverse sets of metastable intermediates (see Figure 4A). We found that there are multiple on-pathway metastable intermediates of GTP−protein complex on the GTPase surface before GTP lands into the native site of the protein. However, there is only one major ligand binding pathway. It involves GTP transitioning through unbound State L0 → State L4 → State L6 → State L7 (see Figure 4B). This pathway accounts for 81% of the total binding

pathways (see Figure 4C). In this major pathway, the ligand moves from bulk (State L0) to a location close to switch 1 of GTPase and resides there creating a metastable State L4. It then transitions from State L4 to another metastable State L6. Here, although the ligand has already been bound at the crystallographic binding site and has formed many stable native interactions with active site residues (see Figure 5), the native interaction of the ligand with switch 1 of the GTPase is still missing in this state. As discussed before, switch 1 does not remain stable in the native conformation because of its higher flexibility and mostly prefers being in its fully open conformation, but it plays an important role in the GTP recognition. This becomes more evident after plotting the kinetic rate matrix (or transition matrix) of the underlying MSM (see Figure S4). We can see that the rate of transition is more on the path which involves interaction with the switch 1, also mfpt's are slightly faster on this path relative to all other paths.

Apart from the major pathway involving metastable intermediates of ligand location near switch 1 and switch 2, there exist multiple secondary competing pathways (see pathways 2 and 3 from Figure 4) and minor pathways (see pathways 4 and 5 from Figure 4) where the GTP is transitioning from the unbound State L0 to multiple combinations of protein non-native sites. In these secondary pathways, the GTP is transitioning from the unbound State L0 to different protein

non-native sites such as near helix-5 in state L1, near helix-3 in State L2, and near helix-4 in State L3, similarly, near switch 2 in State L5 (see state snapshots from Figure 4C). These pathways show nearly equal contributions in the overall binding. The pathways 2 and 3 contribute nearly 5−8% and pathways 4 and 5 nearly 2%, which indicates a complex and competitive GTP binding mechanism among these paths. It is noteworthy to highlight in this context that, the common presence of State L4 in most of these binding pathways (1, 2, 4, and 5) suggests that the highly flexible switch 1 (loop 1) serves as the crucial intermediate location for GTP to latch onto before making the move to the native site. Similarly, the common presence of State L5 in the subsequent pathways (2, 3) suggests that switch 2 (loop 2) also helps GTP to anchor onto the GTPase surface before jumping into the stable, native bound pose. This involvement of both switch 1 and switch 2 in the stabilization of metastable states clearly implies their role in the GTP binding mechanism. Although the comparable stationary state populations of State L4 and State L5 (see Table 1) also suggest definite role of switches (switch 1 and switch 2) in the GTP binding mechanism, according to major pathway we conclude that State L4 and hence switch 1 plays a greater role in the GTP binding than the switch 2. Similarly, comparable stationary state populations of the other macrostates (State L2 and State L3) also suggest the important role of these non-native binding sites in the GTP binding (see Table 1).

GTPase Displays Conformational Plasticity in the Presence of GTP

A key finding of our initial investigation has been that the GTPase protein is highly flexible. More importantly, as shown in Figure 2, in all of the binding trajectories of the GTPase protein, switch 1 and switch 2 loops were found to be highly flexible and disordered. These two loops remained deviated from the proper crystallographic native conformation for most of the simulation time, while the ligand GTP became bound in its native pose. Accordingly, for a comprehensive understanding of the protein conformation in the presence of GTP, we built a separate kinetic network model by discretizing all simulation trajectories along the RMSD of switch 1 and switch 2 loops (see Methods and Materials). We finally constructed a five-state model to resolve the complex protein conformational dynamics (see Figure S2d,e,f).

The resulting network shows that ensembles of conformations of GTPase involve multiple disordered and open conformations of both switch 1 and switch 2 loops, apart from recovering the native tertiary fold of the GTPase (see Figure 6). In particular, State P0 represents a GTPase conformational ensemble in which switch 1 is opened at its full extent, while State P4 represents the ensemble of the properly folded native state (relative to both switch 1 and switch 2). Similarly, State P3 represents a conformational ensemble where switch 1 is partially (or half) in the open state, while State P2 represents a protein conformation where both switches are open and in a disordered state. Again, State P1 represents the conformation where switch 2 is in a highly disordered state with respect to the native state. The statistical model recovers the native fold of GTPase (State P4) as the highest populated conformation. However, the model also indicates the presence of conformational heterogeneity involving competitive protein ensembles. This can be anticipated from the reweighed FES along RMSD1 and RMSD2 (see Figure S2d), which shows five distinct minima with similar free energies separated by an accessible thermody-

namic barrier. Similarly, having significant equilibrium state populations for each macrostate (see Table 2) further indicates

Table 2. Equilibrium Populations of the Key Protein Conformations Obtained from Protein MSM

protein states	population (%)	
State P0	15.29 ± 0.02	(fully open)
State P1	11.16 ± 0.01	
State P2	19.20 ± 0.07	
State P3	24.63 ± 0.01	
State P4	29.69 ± 0.03	(native)

the importance and definite role of each of these macrostates in the GTPase conformational dynamics. Figure 6B,C shows the net transition (net flux) pathways among protein macrostates. We find that the conformational transition of the GTPase mainly occurs via two paths and the net conformational flow is in favor of the native state conformation State P4 (see Figure 6C). In the key major path (pathway 1), GTPase protein transitions from the fully open state (State P0) to the perfect native state (State P4) via an intermediate State P3. This path accounts for almost 74% of the total path. However, there also exists another major pathway, in which the GTPase conformation shuttles to the native fold via the intermediate State P2, where both switch 1 and switch 2 are in a semidisordered state, and then to the State P3 before transitioning to the native pose. This path accounts for 26.20% of the total flux. On the committor scale plotted in the x-axis (bottom of the Figure 6B), we find that State P1, State P2, and State P3 have a similar committor value of 0.25, which is close to zero. This suggests that these intermediates are conformationally closer to State P0 and are therefore more easily susceptible to undergoing further opening of switch 1 (or to visiting State P0) than the closing of switch 1 (or to visiting native State P4) of the GTPase. While the transition rate matrix (or rate network from Figure S6) shows mutual transitions between all of the protein macrostates, the overall protein fluctuations are in slight favor of the native state (State P4). Note that there are more inward transitions coming from State P1, State P2, or State P3 toward the final native State P4 than outward transitions from State P4 to these states. This is also reflected in the equilibrium state populations shown is Table 2, where all of the protein macrostates hold significant populations due to these mutual transitions, with the native state (State P4) showing slight dominance over other conformations due to its preferred thermodynamic stability.

■ DISCUSSION

"Undruggability" of GTPase is a puzzle for the drug discovery community. Although these oncoproteins may be cracked using innovative systems biology and high-throughput techniques, understanding the sequence of events toward molecular recognition and binding of the Ras GTPase protein could lead to the design of GTP-competitive inhibitors. This article examines this process at an atomistic level. While the simulation could capture a rare event of GTP binding to its crystallographic pose, a major bottleneck was found as the protein conformation could not be reconciled with the crystallographic conformation. To that end, the data of around 0.1 ms simulation trajectories were processed using a statistical model. In one of the models, we explored the transition of the ligand from the solvated state to the crystallographic bound pose, where the key finding revealed the important roles of switch 1 and switch 2 in the recognition
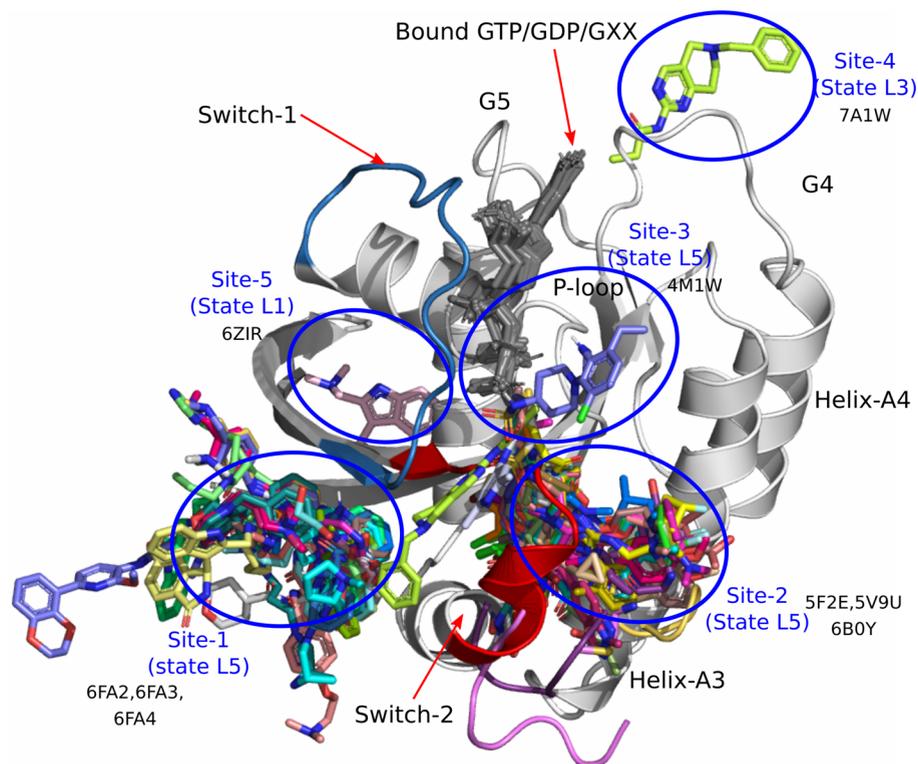
Figure 7. Comparison of ligand MSM macrostates with existing PDB references. For simplicity, only ligands are highlighted from these reference PDBs (a few of them are labeled here). A more comprehensive list of PDB IDs is provided in Table S2.
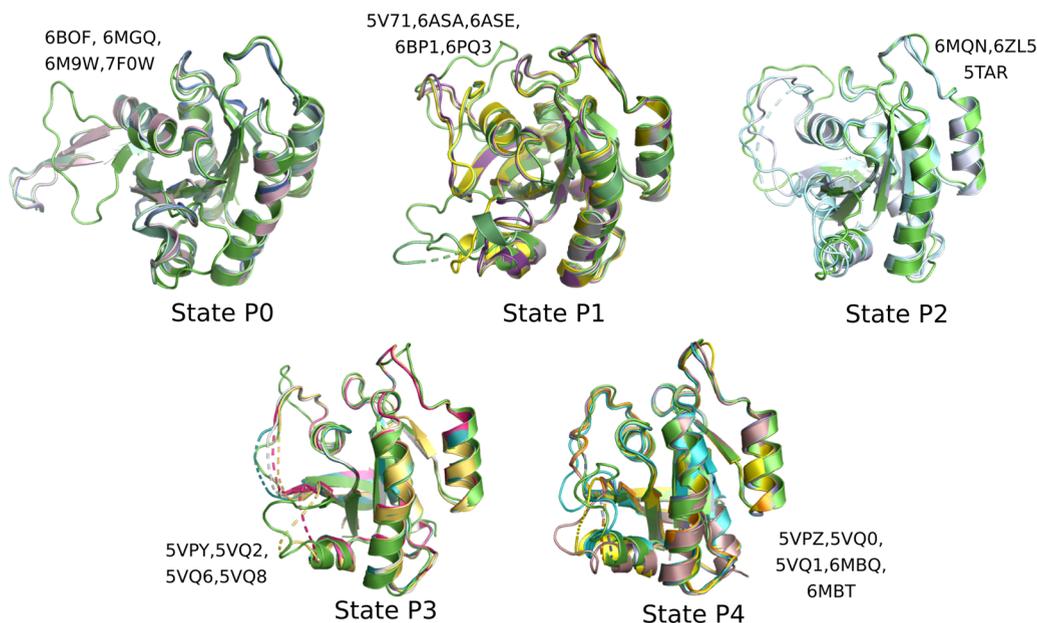
Figure 8. Comparison of protein MSM macrostates with existing PDB references. The reference PDB structures are aligned on top of only a single representative macrostate snapshot using pymol and are shown in green color in all of the state images. Also, each of the states is further labeled with respective PDB IDs.

Figure 9. Potential allosteric binding pocket prediction algorithm: (A) The druggable pocket sites (PS) predicted for native input structure PDB 1QRA. The enclosed table lists the pockets according to the descending order of the druggability probability and also provides the pocket volume and corresponding label of the GTP bound state. (B) Recurrences of druggable pocket sites capturing GTP molecule in various binding competent states obtained from MSM. The displayed pockets are the consistent pockets estimated using two independent methods, fpocket[56] and DoGSite[57] estimation method integrated in the PockDrug online web server.[55]

process. The other network model showed the conformational transitions of the protein before it reaches its crystallographic conformation.

Our investigation indicated the presence of a set of non-native binding-competent states of GTPase/GTP complex (States L1−L5), which act as metastable hotspots (or halting points) in the GTP binding mechanism. The relevant questions are Are there any precedents of these identified states in previous experimental or crystallographic reports? What could be the realistic implications of these states? To address these questions, we surveyed the RCSB database to identify K-Ras entries with additional allosteric ligands (PDBs with ligands other than GTP/GDP/GXX). We found 85 entries with allosteric ligands, which we have reported in Table S2. Figure 7 also shows the structure of these 85 protein−ligand complexes. For simplicity, only allosteric ligands along with parent ligand GTP/GDP/GXX are shown after aligning the respective proteins onto the

single GTPase protein structure (from PDB 1QRA). After obtaining this basic alignment of all the PDB structures, we observed that multiple allosteric ligands clustered at specific sites on the GTPase protein (see Figure 7, sites 1−5.) We then visually compared these GTP binding site locations from the PDB structures with the GTP bound intermediate sites obtained from the ligand MSM after aligning the protein using Pymol software.[54] Interestingly, these allosteric ligand binding sites (locations on the protein) coincided with many of the MSM-derived metastable hotspots. For example, site-3 from Figure 7 accurately matches with the location of State L5 (see State L5 snapshot from Figure 4). Similarly, site-4 and site-5 match with the non-native State L3 and State L1 identified in the present work, respectively. Furthermore, site-1 and site-2 are surprisingly found to match with the subensembles of macrostate State L5. One functional implication of our work could be the use of all of these binding hotspot locations and GTP conformations at

these sites to identify new allosteric ligands for this undruggable protein.

Sites 1, 2, and 3 are biologically crucial because these are the sites where the switch 1 and 2 pockets manifest due to the conformational changes happening during the GTP hydrolysis and GDP−GTP exchange.[1,2] Similarly, sites 1, 2, 3, and 5 correspond to the sites where the GTPase regulating proteins or the effector proteins such as GAPs and GEFs bind to the GTPase, which is essential for its functional activity.[3] Any prior binding of allosteric GTP or GTP-like allosteric ligand at these protein sites can significantly hinder the proper recognition of these GTPase regulating proteins by GTPase and hence can alter or define the fate of the GTPase "on−off" cycle. On similar note, site 4 is biologically significant because it includes the G4-loop, which is important for GTP recognition. Site 4 is the site where the G4-loop residues participate in the hydrogen bonding network for stabilizing the guanine ring of the native bound GTP (in its native pose).[1,2] Having an alternative, allosteric (GTP or GTP-like) ligand at site 4 which interacts with G4-loop residues will make G4-loop residues unavailable for stabilization of the native GTP pose, thereby perturbing the stability of the native bound GTP. Therefore, all of the sites 1−5 are biologically important considering the GTPase activity. Hence, obtaining detailed atomistic insights into GTP interactions at these sites is of utmost importance for improving our understanding of GTP recognition and can pave the way for developing new allosteric binders for this clinically challenging system.

Similar to ligand location, for comparing MSM-derived GTPase protein conformations with the experimental RCSB database, we curated K-Ras PDB entries with keywords such as "open conformation", "closed conformation", "disordered", and "fully opened or extended conformations". We then compared each of the PDB structures with the MSM obtained protein conformation by aligning these conformations. Finally we attributed individual PDB structures to each of the protein conformations based on the best fitting of both structures. Figure 8 shows comparisons of protein MSM macrostates with existing PDB references, and Table S3 provides lists of the PDB IDs.

In order to test our hypothesis that the binding competent sites identified in the present investigation corresponds to GTP bound at potential allosteric binding sites around GTPase, we employed the pocket prediction algorithm using the "PockDrug-Server",[55] an online pocket Web server (http://pockdrug.rpbs. univ-paris-diderot.fr.). The Web server provides options for pocket estimation based on either ligand proximity or the input protein structure. We made use of the second approach for testing our previously stated hypothesis. The server uses two pocket predictions algorithms, "fpocket"[56] and "DoGSite"[57] which are not biased by the ligand position in the input structure. We first performed pocket estimation on the native GTP bound input structure (PDB 1QRA) to use its result as a reference. We then repeated the same approach on the respective protein conformations corresponding to each of the GTP bound non-native macrostates (denoted as States L1−L7 in Figure 4). The server predicts the presence of multiple pockets both in the crystallographic bound pose (Figure 9A) and in the ligand-bound non-native macrostates (Figure 9B). Very gratifyingly, we find that the location of GTP in each of the ligand-bound non-native poses coincides with one of the pockets predicted by the pocket-finding tools (see table present in Figure 9). In addition, the server also provides an empirical "druggability score"[55] for each of the pockets identified by these tools. Almost all of the

pockets predicted for the native PDB structures hold significant druggability probability values. Together the analysis highlights the effective role of MSM to systematically integrate information from diverse reported structures and emphasize the potential druggability of discovered non-native sites in GTPase.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacsau.3c00151.

Supplemental figures describing adaptive sampling scheme, ligand binding MSM, validations of MSMs using CK-test, transition rate network of GTP binding process, conformational plasticity of GTPase, and transition rate network of conformational dynamics of GTPase conformational dynamics in the presence of GTP. Supplemental table describing the experimental reference PDB IDs corresponding to the ligand competent binding states obtained from MSM (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Jagannath Mondal − *Tata Institute of Fundamental Research, Hyderabad, Telangana 500046, India;* ⓞ orcid.org/0000-0003-1090-5199; Phone: +914020203091; Email: jmondal@tifrh.res.in

Authors

Bhupendra R. Dandekar − *Tata Institute of Fundamental Research, Hyderabad, Telangana 500046, India;* ⓞ orcid.org/0000-0002-2734-6090

Navjeet Ahalawat − *Department of Bioinformatics and Computational Biology, College of Biotechnology, CCS Haryana Agricultural University, Hisar 125004 Haryana, India;* ⓞ orcid.org/0000-0002-9665-8709

Suman Sinha − *Institute of Pharmaceutical Research, GLA University, Mathura 281406 Uttar Pradesh, India*

Complete contact information is available at:
https://pubs.acs.org/10.1021/jacsau.3c00151

Author Contributions

CRediT: **Bhupendra R. Dandekar** data curation, formal analysis, investigation, methodology, writing-original draft; **Navjeet Ahalawat** conceptualization, data curation; **Suman Sinha** formal analysis; **Jagannath Mondal** conceptualization, methodology, project administration, resources, supervision, writing-original draft, writing-review & editing.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Paduch, M.; Jeleń, F.; Otlewski, J. Structure of small G proteins and their regulators. *Acta Biochimica Polonica* **2001**, *48*, 829−850.

(2) Sprang, S. R. G. protein mechanisms: Insights from structural analysis. *Annu. Rev. Biochem.* **1997**, *66*, 639−678.

(3) Cromm, P. M.; Spiegel, J.; Grossmann, T. N.; Waldmann, H. Direct Modulation of Small GTPase Activity and Function. *Angewandte Chemie - International Edition* **2015**, *54*, 13516−13537.

(4) Munoz-Maldonado, C.; Zimmer, Y.; Medova, M. A Comparative Analysis of Individual RAS Mutations in Cancer Biology. *Frontiers in Oncology* **2019**, *9*, 1088.

(5) Kjeldgaard, M.; Nyborg, J.; Clark, B. F. C. The GTP binding motif: The existence. *FASEB J.* **1996**, *10*, 1347−1368.

(6) Fernando Díaz, J.; Wroblowski, B.; Schlitter, J.; Engelborghs, Y. Calculation of pathways for the conformational transition between the GTP- and GDP-bound states of the Ha-ras-p21 protein: Calculations with explicit solvent simulations and comparison with calculations in vacuum. *Proteins: Struct., Funct., Genet.* **1997**, *28*, 434−451.

(7) Cox, A. D.; Der, C. J. Ras history. *Small GTPases* **2010**, *1*, 2−27.

(8) Cox, A. D.; Fesik, S. W.; Kimmelman, A. C.; Luo, J.; Der, C. J. Drugging the undruggable RAS: Mission Possible? *Nat. Rev. Drug Discovery* **2014**, *13*, 828−851.

(9) Papke, B.; Der, C. J. Drugging RAS: Know the enemy. *Science* **2017**, *355*, 1158−1163.

(10) Prior, I. A.; Lewis, P. D.; Mattos, C. A Comprehensive Survey of Ras Mutations in Cancer. *Cancer Res.* **2012**, *72*, 2457−2467.

(11) Ledford, H. The ras renaissance. *Nature* **2015**, *520*, 278−280.

(12) Mathieu, M.; et al. KRAS G12C fragment screening renders new binding pockets. *Small GTPases* **2022**, *13*, 225−238. PMID: 34558391.

(13) Molina-Arcas, M.; Samani, A.; Downward, J. Drugging the Undruggable: Advances on RAS Targeting in Cancer. *Genes* **2021**, *12*, 899.

(14) Kessler, D.; Bergner, A.; Böttcher, J.; Fischer, G.; Döbel, S.; Hinkel, M.; Müllauer, B.; Weiss-Puxbaum, A.; McConnell, D. B. Drugging all RAS isoforms with one pocket. *Future Medicinal Chemistry* **2020**, *12*, 1911−1923.

(15) Fell, J. B.; et al. Identification of the Clinical Development Candidate MRTX849, a Covalent KRASG12C Inhibitor for the Treatment of Cancer. *J. Med. Chem.* **2020**, *63*, 6679−6693.

(16) Hallin, J.; et al. The KRASG12C Inhibitor MRTX849 Provides Insight toward Therapeutic Susceptibility of KRAS-Mutant Cancers in Mouse Models and Patients. *Cancer Discovery* **2020**, *10*, 54−71.

(17) Skoulidis, F.; et al. Sotorasib for Lung Cancers with KRAS p.G12C Mutation. *New England Journal of Medicine* **2021**, *384*, 2371−2381.

(18) Chan, C.-H.; Chiou, L.-W.; Lee, T.-Y.; Liu, Y.-R.; Hsieh, T.-H.; Yang, C.-Y.; Jeng, Y.-M. PAK and PI3K pathway activation confers resistance to KRASG12C inhibitor sotorasib. *Br. J. Cancer* **2023**, *128*, 148−159.

(19) Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* **2018**, *140*, 2386−2396.

(20) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. How Does a Drug Molecule Find Its Target Binding Site? *J. Am. Chem. Soc.* **2011**, *133*, 9181−9183.

(21) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 13118−13123.

(22) Noe, F.; Schutte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci. U.S.A* **2009**, *106*, 19011−19016.

(23) Plattner, N.; Noé, F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat. Commun.* **2015**, *6*, 7653.

(24) Mondal, J.; Ahalawat, N.; Pandit, S.; Kay, L. E.; Vallurupalli, P. Atomic resolution mechanism of ligand binding to a solvent inaccessible cavity in T4 lysozyme. *PLOS Comput. Biol.* **2018**, *14*, e1006180.

(25) Ahalawat, N.; Mondal, J. Mapping the Substrate Recognition Pathway in Cytochrome P450. *J. Am. Chem. Soc.* **2018**, *140*, 17743−17752.

(26) Ahalawat, N.; Mondal, J. An Appraisal of Computer Simulation Approaches in Elucidating Biomolecular Recognition Pathways. *J. Phys. Chem. Lett.* **2021**, *12*, 633−641.

(27) Abraham, M. J.; Murtola, T.; Schulz, R.; Pall, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to super-computers. *SoftwareX* **2015**, *1−2*, 19−25.

(28) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71−73.

(29) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(30) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255.

(31) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.

(32) Pall, S.; Hess, B. A flexible algorithm for calculating pair interactions on {SIMD} architectures. *Comput. Phys. Comm.s* **2013**, *184*, 2641−2650.

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

(34) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(35) Miyamoto, S.; Kollman, P. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952−962.

(36) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schutte, C.; Noe, F. Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **2011**, *134*, 174105.

(37) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Perez-Hernandez, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noe, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput* **2015**, *11*, 5525−5542.

(38) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Perez-Hernandez, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noe, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput* **2015**, *11*, 5525−5542.

(39) Molgedey, L.; Schuster, H. G. Separation of a mixture of independent signals using time delayed correlations. *Phys. Rev. Lett.* **1994**, *72*, 3634−3637.

(40) Naritomi, Y.; Fuchigami, S. Slow dynamics of a protein backbone in molecular dynamics simulation revealed by time-structure based independent component analysis. *J. Chem. Phys.* **2013**, *139*, 215102.

(41) Lloyd, S. Least Squares Quantization in PCM. *IEEE Trans. Inf. Theor.* **1982**, *28*, 129−137.

(42) Deuflhard, P.; Weber, M. Robust Perron cluster analysis in conformation dynamics. *Linear Algebra and its Applications* **2005**, *398*, 161−184.

(43) Röblitz, S.; Weber, M. Fuzzy spectral clustering by PCCA+: application to Markov state models and data classification. *Adv. Data Anal. Classi.* **2013**, *7*, 147−179.

(44) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schutte, C.; Noe, F. Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **2011**, *134*, 174105.

(45) E, W.; Vanden-Eijnden, E. Towards a Theory of Transition Paths. *J. Stat. Phys.* **2006**, *123*, 503.

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

(47) Huse, M.; Kuriyan, J. The Conformational Plasticity of Protein Kinases. *Cell* **2002**, *109*, 275−282.

(48) Plattner, N.; Noe, F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat. Commun.* **2015**, *6*, 7653.

(49) Geyer, M.; Schweins, T.; Herrmann, C.; Prisner, T.; Wittinghofer, A.; Kalbitzer, H. R. Conformational Transitions in p21ras and in Its Complexes with the Effector Protein Raf-RBD and the GTPase Activating Protein GAP. *Biochemistry* **1996**, *35*, 10308−10320.

(50) Spoerner, M.; Hozsa, C.; Poetzl, J. A.; Reiss, K.; Ganser, P.; Geyer, M.; Kalbitzer, H. R. Conformational States of Human Rat Sarcoma (Ras) Protein Complexed with Its Natural Ligand GTP and Their Role for Effector Interaction and GTP Hydrolysis. *J. Biol. Chem.* **2010**, *285*, 39768−39778.

(51) Feuerstein, J.; Goody, R. S.; Wittinghofer, A. Preparation and characterization of nucleotide-free and metal ion-free p21 "apoprotein. *J. Biol. Chem.* **1987**, *262*, 8455−8458.

(52) Dandekar, B. R.; Mondal, J. Capturing Protein-Ligand Recognition Pathways in Coarse-Grained Simulation. *J. Phys. Chem. Lett.* **2020**, *11*, 5302−5311.

(53) Salentin, S.; Schreiber, S.; Haupt, V. J.; Adasme, M. F.; Schroeder, M. PLIP: fully automated protein-ligand interaction profiler. *Nucleic Acids Res.* **2015**, *43*, W443−W447.

(54) Schrodinger *PyMOL Molecular Graphics System*,Version 2.5.0. BioGrids Consortium.

(55) Hussein, H. A.; Borrel, A.; Geneix, C.; Petitjean, M.; Regad, L.; Camproux, A.-C. PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins. *Nucleic Acids Res.* **2015**, *43*, W436−W442.

(56) Schmidtke, P.; Barril, X. Understanding and Predicting Druggability. A High-Throughput Method for Detection of Drug Binding Sites. *J. Med. Chem.* **2010**, *53*, 5858−5867.

(57) Volkamer, A.; Griewel, A.; Grombacher, T.; Rarey, M. Analyzing the Topology of Active Sites: On the Prediction of Pockets and Subpockets. *J. Chem. Inf. Model.* **2010**, *50*, 2041−2052.