Reconciling conformational heterogeneity and Substrate Recognition in Cytochrome P450

Bhupendra R. Dandekar,[†] Navjeet Ahalawat,[‡] and Jagannath Mondal[*,†]

†*Tata Institute of Fundamental Research, Center for Interdisciplinary sciences, Hyderabad 500107, India*

‡*Department of Molecular Biology, Biotechnology and Bioinformatics, Chaudhary Charan Singh, Haryana Agricultural University, Hisar 125004, India*

E-mail: jmondal@tifrh.res.in,+914020203091

Abstract

The mechanistic basis of substrate-recognition by archtypal metalloenzyme cytochrome P450cam lacks clarity due to reports of contrasting crystallographic poses in its substrate-free form. Here we employ atomic resolution computer simulation to quantitatively dissect the conformational diversities of cytochrome P450cam and reconcile them with its substrate-binding mechanism. By initiating unbiased molecular dynamics simulation trajectories from multiple crystal structures of substrate-free P450cam, we first explore the statistical probabilities of each crystallographic pose of making dynamically transition to the other crystallographic pose and vice versa. Our approach allows us to build a comprehensive conformational ensemble of substrate-free P450cam using an aggregated 54 μs simulation trajectories. A Markov state model constructed using the conformational ensemble identifies multiple key macrostates in a broad and heterogenous free energy landscape and in particular, reveals a dynamic equilibrium between a pair of conformations in which the substrate-recognition sites are closed and open respectively. However, the prediction of a significantly high population

of closed conformation, coupled with higher flux and faster rate from *open* → *closed* transition than its reverse process, dictates that the net conformational equilibrium would be swayed in favour of closed conformation. The investigation infers that, while a potential substrate of P450cam would in principle sample the diverse array of conformations of substrate-free protein, it would mostly encounter a closed conformation and hence would follow an induced-fit based recognition process. Taken together, this work establishes how a quantitative elucidation of conformational heterogeneity in protein would provide crucial insights in the mechanism of potential substrate-recognition process.

Introduction

Cytochrome P450s are key hemoprotein monooxygenase family which catalyzes a diverse range of biochemical processes, including metabolism of almost all drugs, lipid and steroid biosynthesis and decomposition of pollutants.[1] They also play a pivotal role in the detoxification of xenobiotics, cellular metabolism, homeostasis and therefore, are of major clinical significance.[2] These superfamily are especially notable for catalysing the hydroxylation process of unactivated hydrophobic molecules,[3] making them attractive models for investigation of determinants of biological molecular recognition process.[4–6] Among various Cytochrome P450 family, CYP101A1, from the *Pseudomonas Putida* has served as the archtypal model for thermodynamic and kinetic investigation of substrate recognition in its heme centre.[7,8] Popularly known as P450cam, this specific subfamily of cytochrome P450 catalyzes the hydroxylation process of D-camphor to 5-exohydroxycamphor.[9,10] P450cam is known for its high substrate specificity in comparison to the specificities of many members of the P450 superfamily and this has been attributed to its relatively small, rigid active site that showed few binding-induced changes in early crystal structures.[10] However, due to large attention from the community over last several decades, P450cam has been a regular subject of biochemical and crystallographic investigations, which has given rise to multiple reports of crystallo-

graphic structures in both substrate-free[11,12] and substrate-bound form of P450cam,.[9,10,13] A visual inspection of these crystallographic conformations indicates that some of them deviates considerably from early perception of rigid active site of P450cam. As would be discussed in the forthcoming paragraphs, the prevalent heterogeneity in crystallographic structure of substrate-free P450cam has clouded our perception of underlying mechanism of substrate-recognition process in this protein.

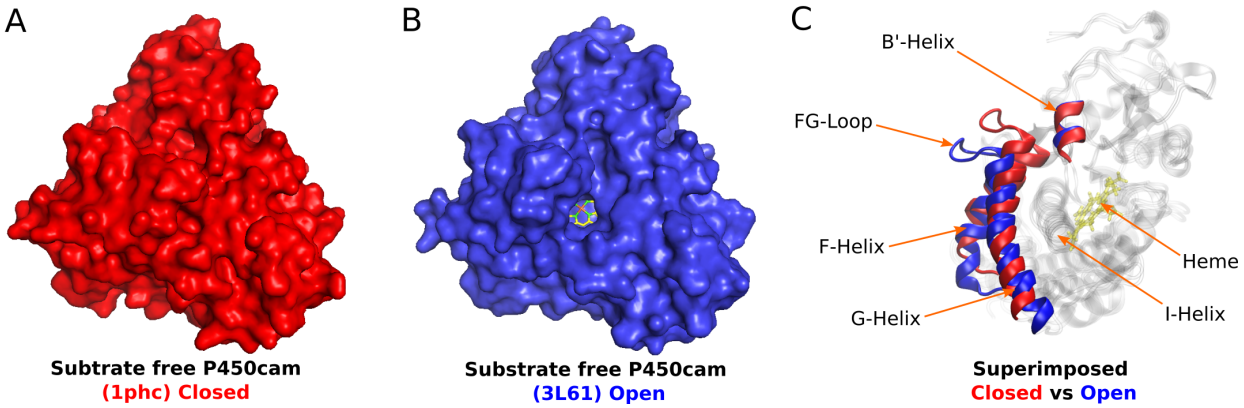


Figure 1: Systems under investigation: (A) Substrate free cytochrome P450cam (PDB:1PHC) with closed pocket (buried cavity). (B) Substrate free cytochrome P450cam (PDB:3L61) with open pocket (solvent accessible cavity). The heme is situated at the catalytic center and is visible in yellow-green color. (C) Comparison of Closed vs Open structures.

The substrate recognition channel of P450cam is formed by F and G helices, the intervening F-G loop, and B' helix, which fold around the heme and helix I to enclose the substrate and position it for binding of substrate to the heme (see figure 1).[9,14] The earliest crystallographic structure of P450cam in substrate-bound pose, resolved by Poulos and coworkers, revealed that the substrate is hidden in a deeply buried cavity near the heme active site, with no open channel accessible from protein exterior.[9] Discovery of crystallographic structure of a substrate-free conformation of P450cam[11] (pdb id: 1PHC) also indicated a substrate binding site that is not readily accessible from the outside. Additionally, the conformations of P450cam describing the intermediate steps involving the catalytic pathways of this metalloenzyme have also been observed in occluded conformation.[15] The prevalence of these closed conformations in various states of activity of P450cam, including the precedences of a closed

substrate-free conformations lead to a perception that during catalytic cycle, P450cam perhaps opens up to allow the camphor entry, closes the active site to perform the catalytic activity and then reopens for product release.[16]

However, the precedences of occluded substrate-free crystallographic pose notwithstanding, report of first structure of an open conformation of P450cam in the presence of large tethered substrate analogs[17] and more significantly a more recent discovery of a crystallographic structure of 'open' conformation of substrate-free P450cam (pdb id: 3L61)[12] have complicated an otherwise simple picture of the conformational landscape of P450cam. Overlay of 'open' (pdb id: 3L61)[12] and 'closed' (pdb id: 1PHC)[11] crystallographic pose of P450cam shows that these two structures mutually differ only in the position of F/G helices and intervening F/G loop near the substrate binding channel (see figure 1): The inward movements of F/G helices and F/G loop towards the heme active site are found to occlude the substrate-binding site in the closed conformation, while the outward movements of F/G helices and F/G loop away from the heme active site are found to open the substrate-binding site in the open conformation. Subsequent investigations of P450cam in solution by double electron-electron resonance spectroscopy suggested that in the physiologically relevant solution state, the substrate-free P450cam exists in the open conformation[18] and it is the binding of substrate which leads to a closed conformation. Recent nuclear magnetic resonance (NMR) studies[19,20] showed that P450cam populates an ensemble of conformations based on the segmental movement of F and G helices. The presence of a mobile active site and speculation of conformational selection based substrate binding mechanism have been supported by Infrared spectroscopic investigations.[21] Conformational fluctuation is known to play key role in substrate recognition across multiple P450s, but only recently have specific details about the conformational space sampled by a given enzyme are emerging.[12,22–24] Together, these findings have introduced the notion of complex 'conformational heterogeneity' in the substrate-free P450cam and its role in modulating substrate recognition by P450cam is slowly being appreciated. In this context, a quantitative analysis of conformational hetero-

geneity substrate-free P450cam at an atomistic resolution is in order and the current work performs this important task.

The current article quantifies the inherent conformational heterogeneity in P450cam by computer simulation of inter-state transition between 'open' and 'closed' conformation of substrate-free P450cam. The substrate free 'closed' structure of cytochromeP450 (PDB id: 1PHC)[11] and the substrate free 'open' structure of cytochromeP450 (PDB id: 3L61)[12] form the basis of our work and we explore the possibility of transition from one state to the other state, by spawning unbiased Molecular Dynamics (MD) simulation trajectories from the both end. We use an aggregated 54 μs of MD trajectories to reconstruct the underlying conformational landscape of P450cam and decipher both free energetic and kinetic aspects of conformational transition within the framework of Markov state model (MSM).[25–27] In particular, the MSM identifies both 'closed' and 'open' conformation of P450cam as two major macrostates, in addition to mapping three other key metastable macrostates in the free energy landscape. The investigation suggests that, while P450cam can in principle visit both open and closed conformation, the balance of the conformational equilibrium is tilted towards 'closed' conformation. Interestingly, the kinetic flux network of transition paths predicts that an intermediate ensemble facilitates the transition between open and closed conformation. Additionally, the exhaustive nature of the computer simulations also detects an interesting feature of uncoiling of the B' prime helix of P450cam, which is found to be coupled to the open conformation.[12,28] Taken together, the current work renders a quantitative account of conformational heterogeneity in P450cam and provides key insights on its implication in possible mechanism of substrate recognition in P450cam.

Materials and Methods

Unbiased atomistic simulations of substrate free P450cam

The unbiased atomistic simulations were performed using the substrate-free crystallographic poses of the proteins. Simulations were spawned from *both* the substrate free 'closed' (PDB id: 1PHC)[11] and 'open' structure of P450cam (PDB id: 3L61).[12] Both of these structures have 99.9% sequence similarity except at 334th position where 1PHC has Cystine(CYS) residue and the 3L61 has an alanine (ALA) residue, leading to only one atom difference. We therefore computationally mutated cystine to alanine in pdb id 1PHC for comparison of these two crystal structures at equal footing. Further, we modeled the missing coordinates of B' helix in the 3L61 starting structure using the homology modelling program Modeller[29] by using 1PHC as a template.

Charmm27 force fields[30] were employed to model the protein in their all-atom representations. In each cases, a $8.9nm \times 8.9nm \times 8.9nm$ cubic box was solvated with TIP3P[31] water, with the protein centred at the box. Sufficient numbers of sodium and chloride ions were added to keep the sodium chloride concentration at 150 mM and render the system charge neutral. The total number of particles in each system was around 71000.

All simulations were performed using NPT ensemble. The average temperature is maintained at 303 K using Nose Hoover thermostat[32,33] with a relaxation time of 1 ps. The average isotropic pressure of 1 bar was maintained using Parrinello-Rahman barostat.[34] The Verlet cutoff scheme[35] was employed throughout the simulation with the Lenard Jones interaction extending to 1.2 nm and long-range electrostatic interactions being treated by Particle Mesh Ewald summation.[36] All bond lengths involving hydrogen atoms of the proteins and the ligands were constrained using the LINCS algorithm and water molecules were kept rigid using the SETTLE[37] approach.

We initiated a series of long unbiased MD simulations of substrate-free P450cam from both 'closed' crystallographic pose (pdb id: 1PHC) and 'open' crystallographic pose (pdb id:

3L61) by generating random initial velocities according to Maxwell-Boltzmann distribution. From each initial configurations, 14 trajectories of 500 ns length were initiated leading to 7 μs (i.e 14 $\times$ 500 ns) of MD simulations. Hence a total of 14 μs long unbiased MD simulations from both the initial configurations were done. The simulation trajectories were saved every 10 ps for the analysis. Figure 2 shows the schematic of the overall methodology adopted in this work.

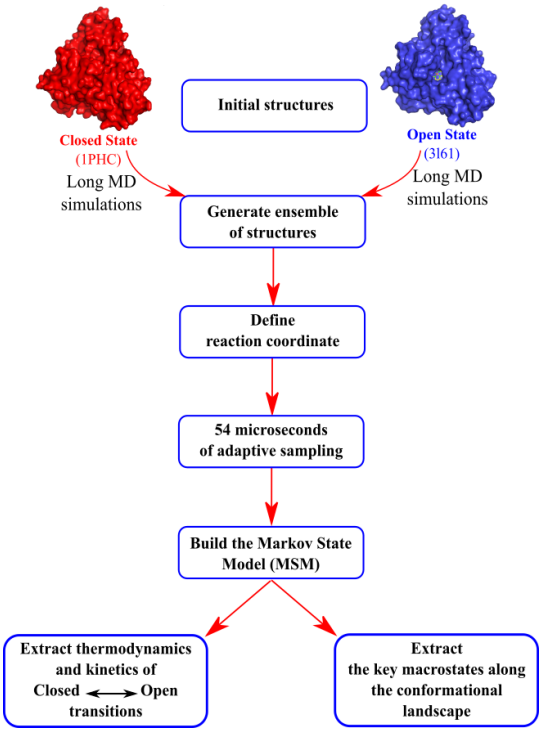


Figure 2: The schematics of the overall methodology adopted to capture conformational heterogeneity in cytochromeP450 system.

We quantified the 'open' and 'closed' conformations based on the collective root-mean squared deviation (RMSD) of F/G-helices plus intervening FG-loop region and B'-helix relative to 'closed' crystallographic pose. Towards this end, we aligned the MD trajectories with the 'closed' crystallographic pose along the backbone residues of the I-helix (resid 234 to 267), "B5"-beta sheet (resid 389 to 401), B'-helix (resid 89 to 96) (see figure 3A) via rotation and translation and subsequently calculated the collective RMSD of F/G-helices plus intervening FG-loop (resid 173 to 214) and B'-helix region (resid 89 to 96) (see figure

3B). To enhance the statistical sampling of the conformational space, we extracted 400 configurations via k-means clustering of 14 μs aggregated trajectories based on this defined RMSD. Subsequently, each of these 400 extracted configurations were subjected to 100 ns of short simulations, leading to a cumulative 40 μs trajectories. All together 54 μs of unbiased all-atoms simulation trajectories were generated. These trajectories were used to quantify the underlying conformational landscape of P450cam and to develop a Markov State Model (MSM) for describing the underlying dynamical process.

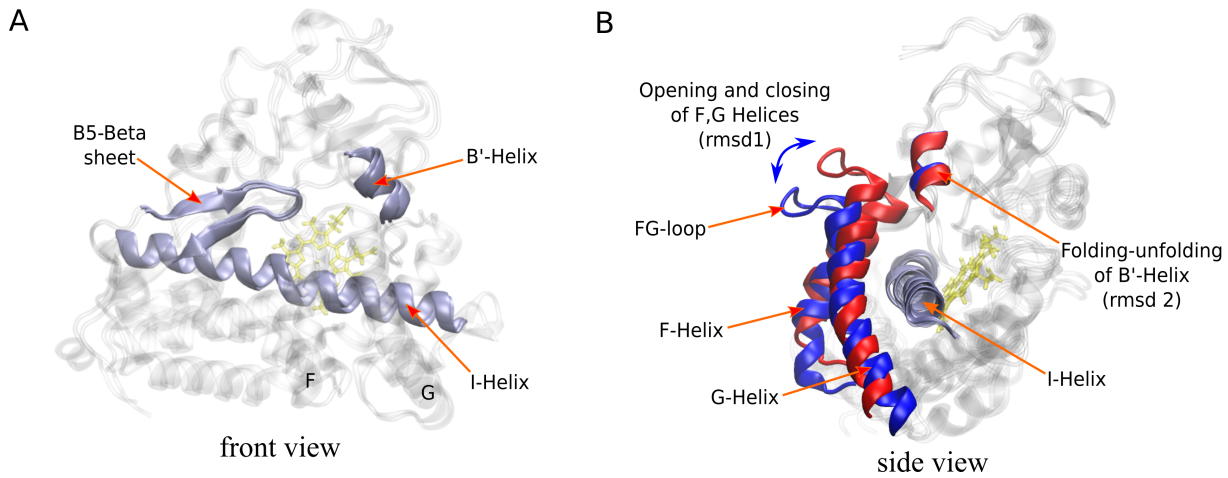


Figure 3: (A) The alignment region used for fitting in rmsd calculations. (B) The reaction coordinates chosen ($RMSD_1$,$RMSD_2$) to capture conformational heterogeneity in cytochromeP450 system. In the figure A the protein is shown from the front side and FG-helices are kept transparent while in the figure B the protein is shown from the side view after rotating figure A by 90° along z-axis and the B5 sheet is kept transparent for clarity

Markov state model analysis

We undertook the task of building a Markov state model (MSM)[26,27] for gaining the detailed understanding of conformational heterogeneity in P450cam using 54 μs data of unbiased all-atoms simulations. We used MSM to identify key conformations and to estimate thermodynamics and kinetics of the underlying dynamics of P450cam. We employed PyEMMA[38](http://pyemma.org) for the purpose of constructing the MSM.

The visual inspection and preliminary analysis of the trajectories suggested that, apart

from the opening and closing events of F/G-helices and intervening F/G-loop of P450cam, uncoiling of B'-prime helix is also taking place as an additional dynamical event. We therefore defined two separate collective variables (CVs) for this purpose: RMSD of F/G helices named as $RMSD_1$ (using resid 173 to 185 and 192 to 214 only) and RMSD of B'-helix named as $RMSD_2$ (using resid 89 to 96) (refer figure 3B). We omitted the F/G-loop region in the $RMSD_1$ calculation because of its higher flexibility to decrease the noise in the calculation. Both these RMSDs were computed relative to 'closed' crystallographic pose. Finally, we captured the conformational heterogeneity in cytochromeP450 along these two $RMSD_1$ and $RMSD_2$ dimensions. Before calculating these RMSDs we aligned the protein backbones of the I-helix, B'-helix and "B5"-beta sheet of all frames of the trajectory with the backbone of 'closed' crystallographic pose. Then we used these RMSDs as clustering metric for discretising the simulation trajectories for subsequent MSM analysis. We employed the k-means clustering algorithm[39] to discretise the input data into 2000 clusters. Then a 2000 microstates MSM was constructed at a lag time of 20 ns. The choice of lag time was based upon the onset of levelling of implied time scale, which ensures the Markovianity of the model (see Fig. S1) in SI. The implied time scale plot also suggested the presence of total 5 metastable states at 20 ns lag time. The 2000-microstate MSM was coarse-grained into a 5 macrostate model using a Perron Clusters Cluster Analysis method (PCCA).[40] The MSM is built with 95% of confidence interval and all the values calculated from the model are with the same level of accuracy. We also performed the bootstrapping of the MSM analysis to estimate the error in estimated population values. Stationary populations of all five macrostates were obtained from the MSM.

The mean first passage time (MFPT) for transition between two states was computed from the MSM. The MFPT defines an average timescale for a stochastic event to first occur[41] or the average time taken for a given conformation to fold for the first time.[42] The MFPT, F_{ij} for the transition of state i to state j is given by the following expression[42]

$$F_{ij} = \tau + \sum_{j \neq i} p_{ij} F_{ji} \ ,$$

where (τ) is the lag time used, p_{ij} is the probability of transitioning from state i to state j within the lagtime of (τ) and the F_{ji} is the MFPT for the reverse transition. The rate constants were calculated as

$k_{ij} = \frac{1}{F_{ij}}$,

where F_{ij} is the MFPT for the transition of state i to the state j.

The gross flux, the average number of the transitions from state i to state j were estimated form the Transition Path Theory (TPT)[38,43–45] and are computed as follows

$f_{ij} = q_i^- \pi_i p_{ij} q_j^+$,

where q_i^- and q_i^+ are the backward and forward commitor probabilities, π_i is the equilibrium population of the state and $\pi_i p_{ij}$ is the equilibrium flux. The net fluxes are computed as

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

These net fluxes define a directed path from initial state (A) to final state (B) and hence define the mechanism of the dynamics.

Results

Precedences of crystallographic poses of both closed and open structures of P450cam in its substrate-free form raise a question if these two crystallographic poses are dynamically visited by P450cam in the array of conformational ensemble. A quantitative insight into this question would address mechanistic aspect of recognition process of camphor to P450cam: For example, a prevalence of closed conformation in the conformational ensemble of P450cam would lead to an induced-fit based mechanism for substrate recognition by P450cam. On the other hand, prevalence of open conformation or equal probability of both open and closed conformation in the conformational landscape would imply that the substrate might adopt a conformational selection based mechanism for recognising P450cam. To resolve this question, we first simulate the conformational dynamics of substrate-free P450cam by

initiating a large set of MD trajectories from both open (pdb id: 3L61) and closed (pdb id: 1PHC) crystallographic poses. An overlay of the crystallographic pose of open and closed conformation (see figure 1C) suggests that these two structures mutually differ by an RMSD of 0.45 nm due to difference in positioning of F/G helices region. Accordingly, we monitor the root-mean squared deviation (RMSD) of F/G helices of all trajectories from that of closed conformation (pdb id: 1PHC) (hereby referred as "$RMSD_1$") as a measure of the conformational dynamics of P450cam (see figure 3B).

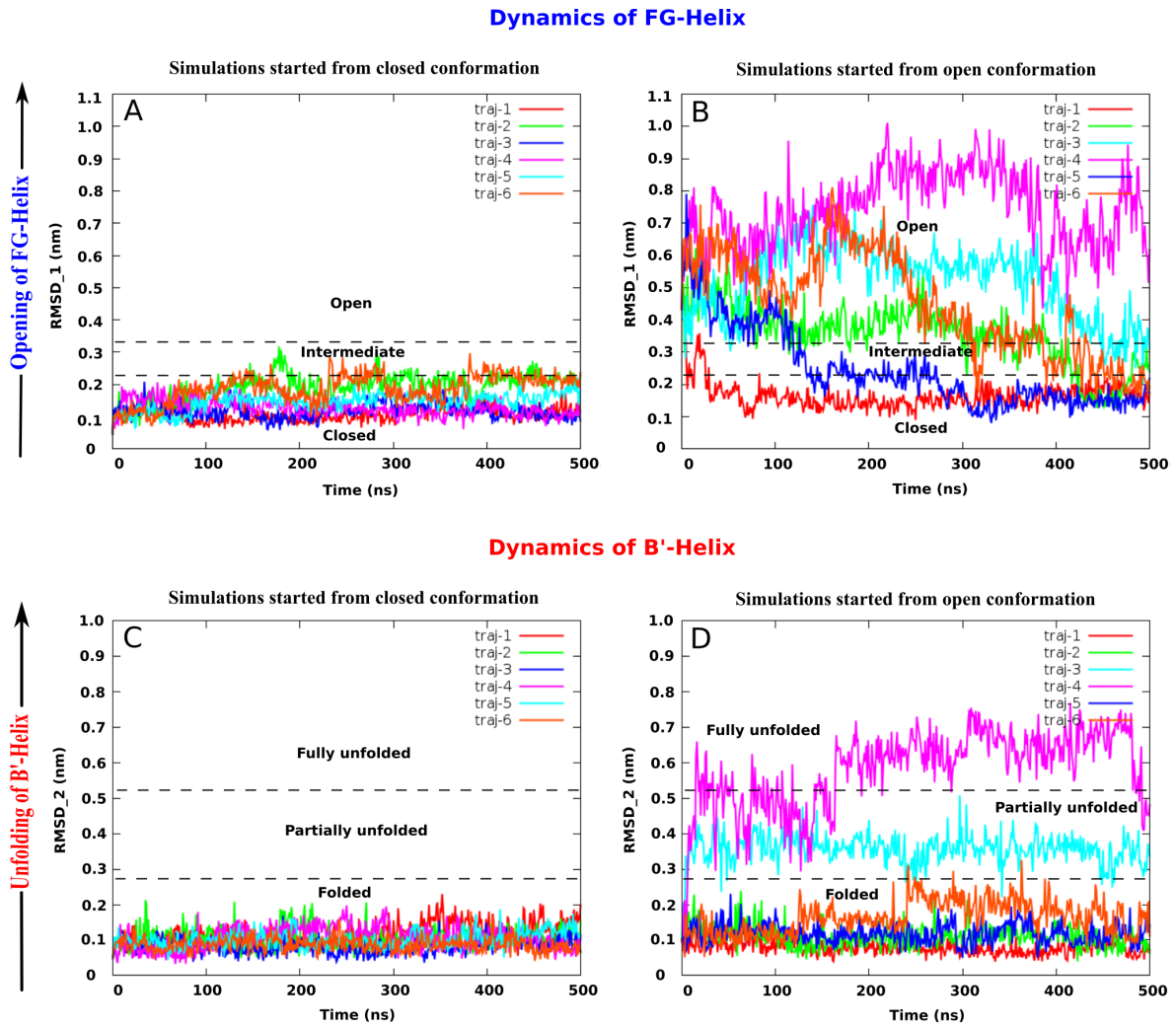


Figure 4: The conformational dynamics of FG-helix and B'-helix in cytochrome P450: (A) The time profile of $RMSD_1$ for simulations starting from closed conformation and (B) for simulations starting from open conformation. (C) The time profile of $RMSD_2$ for simulations starting from closed conformation and (D) for simulations starting from open conformation.

Figure 4A, 4B compare the representative time profiles of RMSD of F/G helices as computed from equilibrium simulated trajectories, initiated from closed and open crystallographic pose respectively. We find that for trajectories starting from the open crystallographic pose (refer figure 4B), $RMSD_1$ decreases from 0.45 nm to 0.20 nm (or below that) at a very short time scale, implying that the 'open' conformation of P450cam often makes transition to closed conformation within the simulation time scale. In some of the trajectories initiated with open conformation, we also find that $RMSD_1$ increases from 0.45 nm to 0.9 nm within a very short time scale, implying that the 'open' conformation of P450cam potentially can make transition to even more open conformation also within the simulation time scale. On the other hand, none of the MD trajectories starting from the closed crystallographic pose (refer figure 4A) made transition to open conformation within the simulation period. Rather, in trajectories initiated from closed crystallographic pose, $RMSD_1$ fluctuates between 0.04 and 0.20 nm, within the simulation time scale, at times making occasional visit to an intermediate state($RMSD_1 \approx 0.22$ nm) , implying that trajectories initiated from closed crystallographic pose mostly remained confined around closed conformation within the time-scale of the simulation. Inspection of these simulation trajectories initiated from open conformation also reveals unfolding of B' helix in some cases. As depicted in figure 4D, RMSD of B' helix (referred as $RMSD_2$, see figure 3B) deviates significantly in trajectories initiated from open conformation. However, the $RMSD_2$ does not change substantially in trajectories initiated from closed conformation and the B' helix remains helical. (see figure 4C).

Taking cue from the trends of conformational transitions in these simulation trajectories, we reconstructed the underlying conformational free energy landscape of P450cam to get detailed insights of complex conformational dynamics. Towards this end, we combined long simulation trajectories initiated from both ends (closed and open crystallographic pose) and adaptively sampled conformational space by spawning numerous short simulations from various conformations clustered from the long trajectories (see method and figure 2 for

schematics). These aggregated trajectories were then discretised and a Markov state Model (MSM) was built (see methods for details) and a free energy landscape along $RMSD_1$ and $RMSD_2$ is derived. This free energy landscape is then finally reweighed according to MSM derived relative weights for each microstate.

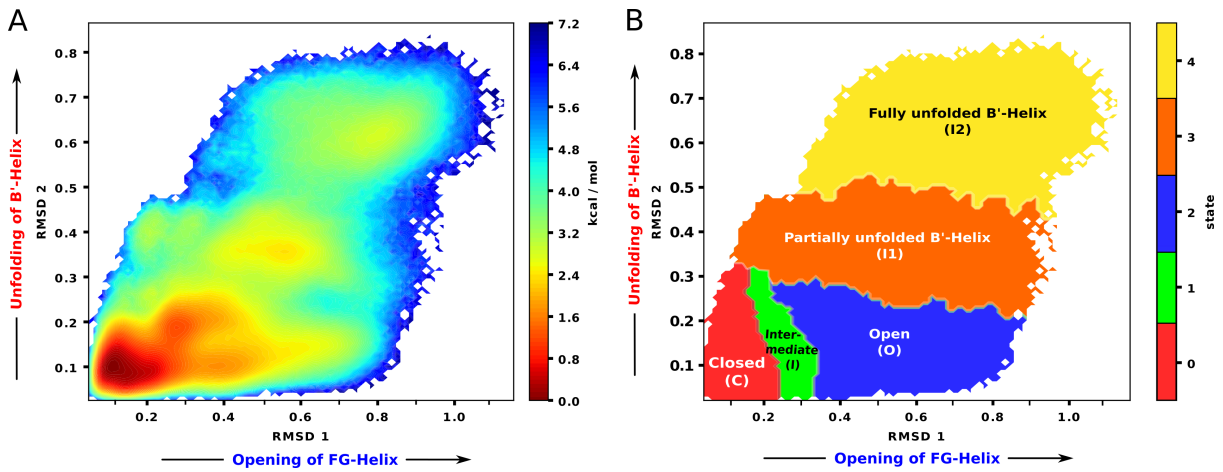


Figure 5: The thermodynamics of the conformational heterogeneity in P450cam: (A) MSM reweighed free energy landscape. It shows a presence of five distinct energy basins along $RMSD1, RMSD2$. (B) The state decomposition obtained from MSM calculation. Each macrostate shown in different colour and labelled with the description of conformations it represents.

Figure 5A depicts the two-dimensional map of the free energy landscape along $RMSD_1$ and $RMSD_2$ and provides a quantitative thermodynamic account of underlying conformational heterogeneity in P450cam. We find that the location of global minimum of the free energy landscape corresponds to both $RMSD_1$ and $RMSD_2$ close to 0.1 nm, which is very close to the RMSD of the 'closed' crystallographic pose (pdb id: 1phc). Apart from this, an important basin at a relatively higher free energy value, centred around $RMSD_1 = 0.45$ nm and $RMSD_2$= 0.10 nm, corresponds to the crystallographic pose of pdb id 3L61. The estimated free energy barrier height separating the two minima corresponding to closed and open conformation is 2.5 kcal/mol. Additionally, the free energy surface also showcases three other local minima at other values of $RMSD_1$ and $RMSD_2$. One of the minima lies in between open and closed state ($RMSD_1$ =0.3 nm and $RMSD_2$=0.20 nm), which we would eventually identify as an intermediate macrostate. The remaining two minima belongs to

relatively larger value of $RMSD_1$ and $RMSD_2$: one with partial unfolded B'-Helix corresponding to $RMSD_1$ and $RMSD_2$ values of 0.55 nm and 0.38 nm respectively and other with fully unfolded B' helix corresponding to $RMSD_1$ and $RMSD_2$ values 0.80 and 0.62 nm respectively. Together, the presence of multiple basins at different free energy values reflects a conformationally heterogenous landscape of substrate-free P450cam.

To quantify the state-space decomposition of the conformational landscape in more comprehensive manner, the microstates constituting the underlying MSM are coarse grained into finite number of macrostates. The implied time scale plot as a function of lag time (see figure S1) suggested that a 5-state decomposition would be optimal for aptly describing the conformational landscape of P450cam. Figure 5B depicts how these five macrostates (namely C, O, I, I1 and I2) map the entire conformational landscape along $RMSD_1$ and $RMSD_2$. We find that each of these macrostates corresponds to a distinct basin of the free energy landscape. Table 1 enumerates relative stationary populations of these macrostates. We find that the closed conformation (C) is the most populated (65.20 %) macrostate and is spatially well-separated from the open conformation (O) (14.14 %) via an intermediate conformation (I) (12.89 %). Interestingly, apart from identifying the 'closed (C)', 'open(O)' and intermediate (I) states, we also discover two additional macrostates (I1 and I2) of relatively lower populations at larger values of $RMSD_1$ and $RMSD_2$.

In figure 6, we depict distinct five representative snapshots corresponding to most probable states from each macrostates. These conformations are also annotated on the free energy landscape in figure S2. Each of the snapshot consists of ten conformations randomly pruned from ensemble of each macrostates and overlaid with the closed crystallographic pose (1PHC, red) for comparison. Visual inspection of these conformations (refer figure S2) shows that there is gradual opening of the F/G helices from close conformation (C) to the intermediate confromation (I) and to the open conformation (O). In all these three states the helicity of B'-helix is preserved. The overlay of conformations of macrostate C, macrostate O and macrostate I in figure 7 (left) clearly shows that the state-space decomposition can distin-

guish the conformation based on the extent of outward movement of F/G helices (away from heme group) from the closed crystallographic pose: the F/G helices move the most, away from heme group, in macrostate O, followed by that in macrostate I and least in macrostate C. Visual inspection of the conformations corresponding to I1 and I2 indicates that these two macrostates appear due to occasional uncoiling or unfolding of B' prime helix as F/G helix opens. The overlay of conformations of macrostate I1 and macrostate I2 with the macrostate C (refer figure S2) indicates partial to complete uncoiling of B' helix, with outward movement of F/G helices in these conformations resembling that of macrostate O.

Table 1: The stationary populations for MSM macrostates.

MSM macrostate	Populations in %
State C	65.20 ±0.04
State I	12.89 ±0.11
State O	14.14 ±0.08
State I1	5.28 ±0.01
State I2	2.49 ±0.01

Table 2: The kinetic rate of mutual transitions between state C and state O.

State transitions	Kinetic constants in s^{-1}
State C →State O	$k^{co} = 0.22 \times 10^{-4}$
State O →State C	$k^{oc} = 8.69 \times 10^{-4}$

To unravel the underlying kinetic mechanism of conformational heterogeneity in cytochrome P450 we constructed the kinetic network of transition among the macrostates, using TPT. We considered closed state(C) and open state(O) as two extreme ends of the conformation that P450cam chooses to explore and extracted the net fluxes of the transitions among the macrostates. Figure 6 depicts the kinetic mechanism of conformational transitions in substrate-free P450cam based on the direction of the net flux and the amount (arrow thickness) of net flux among the macrostates. These macrostates are represented as a circular discs with their sizes proportional to their stationary populations and are placed according to their committor probability value.

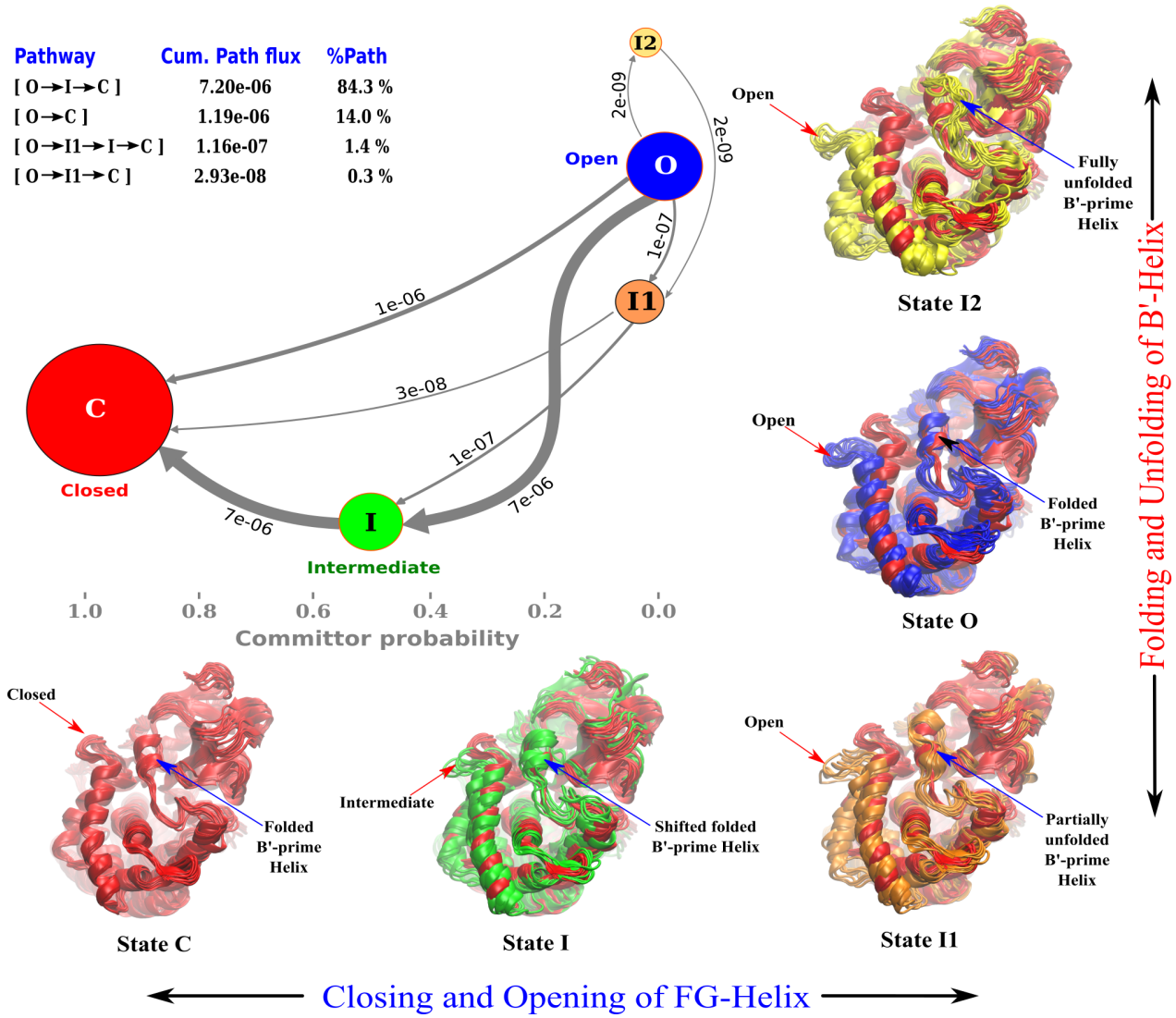


Figure 6: The kinetic mechanism of conformational heterogeneity in cytochrome P450cam: The net flux network is obtained from TPT analysis. The network connects five macrostates and are represented as a circular discs. The sizes of the discs are proportional to their stationary populations and are placed according to their respective committor value. The committor scale is drawn at the bottom of the figure. The committor value of 1 corresponds to the closed state(C) and 0 corresponds to the open state(O), while committor value of 0.5 corresponds to the intermediate (I) state. The states are connected with the arrows and 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. The resulting pathways are shown at the top left of the figure along with their cumulative path fluxes and percentage path contribution. The figure also shows the distinct five representative snapshots corresponding to most probable states from each macrostates. The snapshots are shown in comparison with the first state, the closed state C (red colord) to indicate the relative differences in the conformations. In each states 10 snapshots are shown for simplicity.

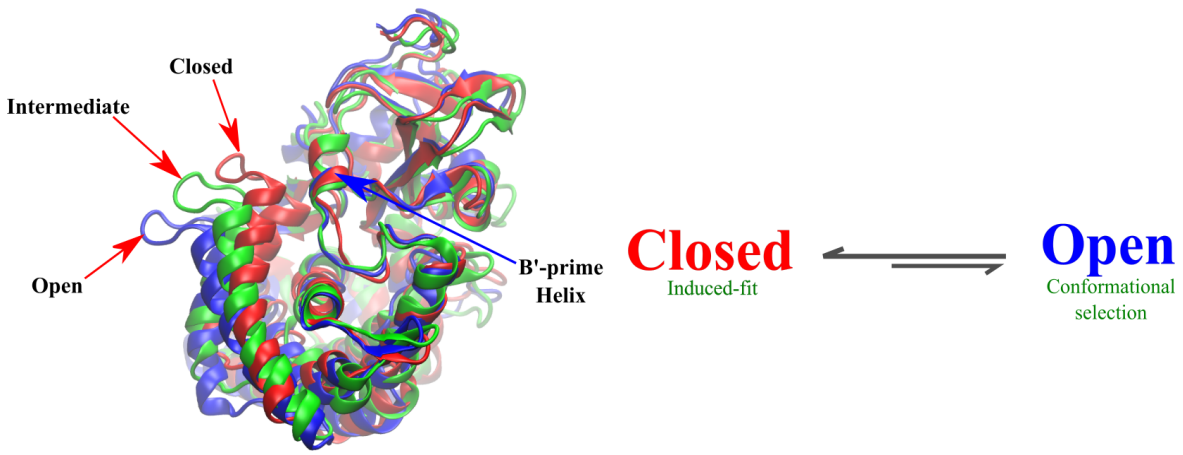


Figure 7: Left: Comparison of major MSM macrostates, macrostate C (closed), macrostate I (intermediate) and macrostate O (open). In all these three states the helicity of B'-helix is preserved. Right: Proposed substrate recognition mechanism

While there are finite probabilities of transitions between open and closed state on both directions, the direction of arrows dictates that the overall net flux of transition will be from O $\rightarrow$ C, implying that the net equilibrium of ($C \rightleftharpoons O$) transition would be tilted in favour of closed (C) conformation. The MFPT values for state-to-state transitions depicted in figure S2 also support that rate of $C \rightarrow O$ transition is slower than the reverse process $O \rightarrow C$ transition. Figure 6 also depicts that there are *multiple* pathways for transition between the closed (C) and open states (O). The arrow thickness in figure 6 indicates the net flux among the macrostates. Especially, comparison of net flux values among different pathways dictates that the *direct* transition O $\rightarrow$ C is less likely, with only a 14 % contribution to the total path. On the other hand, the net flux values predict that the transition from open (O) state to closed (C) state would be predominantly mediated by intermediate macrostate (I). In particular, O $\rightarrow$ I $\rightarrow$ C would emerge as the most dominant pathway with a 84.3 % of contribution in the total path. The kinetic network also suggests additional pathways of transition mediated by I1 ((O $\rightarrow$ I1 $\rightarrow$ I $\rightarrow$ C) and (O $\rightarrow$ I1 $\rightarrow$ C)), but their contributions in the overall pathways are very small. Taken together, higher stationary populations of closed macrostate (table 1), coupled with faster rate (figure S2, table 2) and net flux (figure 6) of open (O) to closed (C) transition (with or without the intermediate(I)) unambiguously

indicate that the dynamic equilibrium of conformational transition is tilted in favour of closed conformation of P450cam. This is also schematically shown in figure7 (right).

Figure 6 further quantifies the respective committor probabilities of each of the five macrostates. We find the closed state (C) and the open state(O) are located at a committor probability of 1 and 0 respectively. Most interestingly, the intermediate (I) state corresponds to a committor probability of 0.5, suggesting that it is the saddle point in the transition path (O $\rightarrow$ I $\rightarrow$ C) with 50% (or equal) transition probability for both the direction. This is also consistent with the relative location of intermediate (I) macrostate in the middle of closed (C) and open (O) state in the free energy surface (see figure S2). The committor probabilities of intermediate states (I1) and (I2) are 0.02 and 0.01 (very close to 0), which also indicates that they are closely associated with the open state (O). This is consistent with the longer MFPT values for C $\rightarrow$ I1 transitions than O $\rightarrow$ I1 or O $\rightarrow$ I2 transitions (see figure S2). Taken together, the observation of macrostates I1 and I2 in the conformational ensemble correlates very well with the partial to complete uncoiling of B' helix with the outward movement of F/G helices.

The observation of macrostates with partially or completely unfolded B' helix in open conformation (i.e. I1 and I2 macrostates) is consistent with prior crystallographic investigations.[12,13] In prior investigations by Lee et. al.,which involved crystallisation of the proteins in presence of camphor with low and high potassium ion concentration[12] or comparison of the crystallised structures of cytochrome P450cam bound to library of active substrates[13], the B' helix has been found to provide key contacts with the substrate. On the contrary, in absence of substrate/ligand, B' helix was found in disordered (unfolded) state along with loss of its site specific potassium ion interaction due to loss of its interaction with the substrate.[12,13] On a related note, it is also found that the B' helix regains its folded state in the presence of substrate and finally converts to the closed state (PDB id: 3L63).[12,13] These experimental observations also justify why the closed substrate-free conformation (pdb id: 1PHC), which is structurally identical with substrate-bound closed state, is always associated with a folded

state of the B'-helix and that its interactions with the substrate were necessary to access closed state from the open state.[12]

The functional implication of the current results can be assessed by exploring the most relevant question: How does the conformational landscape of P450cam, as portrayed in the current work, impact the substrate recognition mechanism by P450cam? While the current work predicts a conformationally heterogeneous landscape of substrate-free P450cam, a considerably larger population of 'closed' conformation than 'open' conformation, net fluxes from *open* $\rightarrow$ *closed* via intermediate coupled with the higher rate of *open* $\rightarrow$ *closed* transition, would tilt the conformational equilibrium towards the 'closed' conformation. As a result, we infer that a potential substrate like camphor would mostly visit the more populated 'closed' conformation of P450cam, which would require an induced-fit based mechanism for substrate-recognition (see figure 7, right for a schematic). This inference of 'induced-fit' based recognition mechanism is in line with Ahalawat et al's recent investigation[46] of substrate-recognition process by P450cam. In Ahalawat et al's multi-microsecond long simulation starting with 'closed' conformation of P450cam, camphor first diffused from bulk to the F/G helices of the protein and then spent a significant time to induce an opening, before landing on the designated heme active site. Similar to current work, in their investigation also, the F/G loop did not open up completely during multi-microsecond long trajectories. However, the current work's observation of a dynamical equilibrium between closed and open conformation (with the equilibrium being tilted towards closed conformation) also does not completely rule out the possibility of 'conformational selection' based substrate-binding mechanism, in which the substrate can potentially visit an open conformation and then closes the pocket after the binding is complete.[13,18]

Discussion

Conformational heterogeneity plays important role in defining the structural and functional dynamics of the enzymes. In Cytochrome P450s, the identification and characterisation of the potential channels connecting its active site to the protein exterior is very important and the conformational heterogeneity has a critical role in their functional mechanism. It has also become apparent that the binding mechanism of substrate is linked to the enzyme's ability to access its conformational ensemble in the presence and absence of substrate. The static three-dimensional crystallographic structures of enzymes solved in different conditions and/or environments are crucial to provide the conformational substates of enzymes but are not sufficient to understand the kinetics and thermodynamics of these substates and their role in complexation.

In this regard, the current article investigates the conformational heterogeneity in P450cam by considering both 'closed' and 'open' crystallographic poses of substrate-free P450cam at equal footing. Via spawning unbiased equilibrium MD simulation trajectories from both crystal structures, we explore the possibilities of each crystallographic conformation of making dynamical interconversion to the other crystallographic conformation and build an exhaustive conformational ensemble of substrate-free conformation. A Markovian model based upon cumulative 54 μs MD data of P450cam quantitatively maps the underlying conformational landscape and identifies multiple key macrostates in the landscape. In particular, the model recovers both 'closed' conformation and 'open' conformation as major macrostates in the free energy landscape, with the 'closed' conformation estimated to be the most populated. The model also predicts a dynamical equilibrium of transition between 'open' and 'closed' macrostate. However, the rate of *closed* $\rightarrow$ *open* transition is found to be considerably slower than the reverse (*open* $\rightarrow$ *closed*) transition. An important finding of current work is that the *direct* transition between open (O) and closed (C) conformation is not the preferred route. Rather, our work predicts that the transition is facilitated via an intermediate (I) macrostate, which is also spatially positioned between these two macrostates. This can also

be seen in figure 6, where major contribution to the net flux is going from open (O) to close (C) conformation via intermediate (I). The observation of an intermediate macrostate and a relatively broader distribution of conformational space in open macrostate, are also in line with the previous hypothesis of protein conformations accessible during substrate recognition.[13] Additionally, the current work elucidates the possible unfolding of B' helix, predominantly from open conformations, which is in line with earlier experimental reports of B' helix exhibiting higher thermal motion[11] and being disordered in open form in absence of substrate.[12] More over, comparison of multiple existing cytochrome P450 structures showed the well preserved protein fold but differing in the orientation of B' helix and its secondary structure state (folded or unfolded).[28] Besides, observation of a disordered B' helix during substrate-egress in previous computer simulation[47] also provides key support to the current result (using a different protein force field than the current work). Therefore, as quantified in the current work, the dynamics of B' helix also plays an important role in defining the conformational heterogeneity in cytochrome p450. Together, the present work provides a quantitative perspective to the notion of conformational heterogeneity in substrate-free P450cam, previously proposed by multiple spectroscopic investigations.[19–21]

The current work's prediction of a dynamic equilibrium between closed and open conformation of substrate-free P450cam, with the equilibrium tilted towards closed conformation, suggest an interesting functional implication related to substrate-recognition process in cytochrome P450cam: while substrate like camphor would visit both open and closed conformations of P450cam, it would mostly encounter the closed conformation due to its significantly higher chances of occurrences than other populated macrostates and hence substrate needs to find an opening to the heme via an induced fit based mechanism. However, the presence of dynamic equilibrium between closed and open macrostate also suggests a finite, albeit low, possibility of conformational selection based mechanism for substrate recognition. We note that previous computational investigations,[47,48] the majority of which are biased MD simulations, have suggested diverse insights on conformational plasticity in P450cam and

its related implication in substrate recognition processes. Rydzewski and coworkers[47] have explored the egress pathways of substrate from the closed cavity using metadynamics simulation and showed that substrate exit from the heme cavity leads to a transition from a heterogeneous ensemble of closed substrate-bound conformers to the basin comprising of the open conformations of P450cam. Based on the observation of shift of conformation from closed to open state during substrate egress, the work suggested an induced-fit based substrate recognition mechanism. On the other hand, Markwick and coworkers,[48] using adaptive accelerated molecular dynamics simulation approach, have indicated that while closed conformation are free energetically more stable, substrate-free closed protein can potentially access a fully open and partially open conformation and suggested induced-fit and conformational selection based mechanism for binding small and large substrate respectively. Our results contradict their prediction that substrate-free open protein can not access a closed structure at all. However, none of previous works has simultaneously considered spawning unbiased trajectories from both closed and open crystallographic poses for exploring their interconversion, as has been done in the current work.

Finally, the proposed role of active site flexibility in substrate recognition in P450 family goes beyond P450cam. The substrate specificities across family of P450s show strong diversity and are believed to largely result from differences in structure, flexibility and dynamics of substrate access channel that connects the protein surface to the deeply buried heme centre. Especially, significant diversities have been noted in the sequence of the substrate-binding channel across various forms of P450,[49] and as a result, the structure and flexibility of the substrate binding channel varies considerably from one form to another. For example, the small closed substrate access channels seen in many prokaryotic P450s[9,50–52] are in marked contrast with the bigger and more wide access channels for mammalian microsomal enzymes involved in drug metabolism.[53–55] Apart from P450cam, investigations have revealed several examples of bacterial P450s that exist both in an open conformation in the absence of substrate and in a closed conformation in the presence of substrate or ligand.[23,54,56–61] Hence

it is expected that the extent of dynamical heterogeneity that current work has discovered in the conformational ensemble of P450cam (which was long perceived to have narrow substrate channel), is relevant and can be more pronounced in P450s with relatively large access channels. Evidences both supporting and refuting the conformational change associated with substrate recognition processes exist among drug-metabolising microcosmal P450s. For example, while P450 CYP2B4 has been reported to undergo significant conformational change upon substrate-recognition,[24,54,55] CYP3A4 P450[62] has been observed in both the substrate-bound and -free forms with relatively small changes in structures. Together, these observations point to a greater functional role of conformational heterogeneity in dictating the mechanism of substrate recognition in cytochrome P450.

Supplemental Information

Supporting information with figures of Implied time scale plot, rate network among macrostates.

Acknowledgements

This work was supported by computing resources obtained from shared facility of TIFR Centre for Interdisciplinary Sciences, India. JM would like to acknowledge research intramural research grants obtained from TIFR, DAE, India, Ramanujan Fellowship and Core Research grants provided by the Department of Science and Technology (DST) of India (CRG/2019/001219).

References

(1) Guengerich, F. P.; Waterman, M. R.; Egli, M. Recent Structural Insights into Cytochrome P450 Function. *Trends in Pharmacological Sciences* **2016**, *37*, 625–640.

(2) Manikandan, P.; Nagini, S. Cytochrome P450 Structure, Function and Clinical Significance: A Review. *Current Drug Targets* **2018**, *19*, 38–54.

(3) Coon, M. J. Cytochrome P450: Nature's Most Versatile Biological Catalyst. *Annu Rev. Pharmaco.* **2005**, *45*, 1–25.

(4) Wade, R. C.; Motiejunas, D.; Schleinkofer, K.; Sudarko,; Winn, P. J.; Banerjee, A.; Kariakin, A.; Jung, C. Multiple molecular recognition mechanisms. Cytochrome P450 - A case study. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2005**, *1754*, 239–244.

(5) Miao, Y.; Yi, Z.; Cantrell, C.; Glass, D.; Baudry, J.; Jain, N.; Smith, J. Coupled Flexibility Change in Cytochrome P450cam Substrate Binding Determined by Neutron Scattering, NMR, and Molecular Dynamics Simulation. *Biophys. J* **2012**, *103*, 2167 – 2176.

(6) Miao, Y.; Baudry, J. Active-Site Hydration and Water Diffusion in Cytochrome P450cam: A Highly Dynamic Process. *Biophys. J* **2011**, *101*, 1493 – 1503.

(7) Griffin, B. W.; Peterson, J. A. Camphor binding of Pseudomonas putida cytochrome P-450. Kinetics and thermodynamics of the reaction. *Biochemistry* **1972**, *11*, 4740–4746, PMID: 4655251.

(8) Helms, V.; Wade, R. C. Thermodynamics of water mediating protein-ligand interactions in cytochrome P450cam: a molecular dynamics study. *Biophys. J* **1995**, *69*, 810–824.

(9) Poulos, T. L.; Finzel, B. C.; Howard, A. J. High-resolution crystal structure of cytochrome P450cam. *J. Mol. Biol* **1987**, *195*, 687 – 700.

(10) Raag, R.; Poulos, T. L. Crystal structures of cytochrome p-450cam complexed with camphane,thiocamphor,and adamantane,factors controlling p-450 substrate hydroxylation. *Biochemistry* **1991**, *30*, 2674–2684.

(11) Poulos, T. L.; Finzel, B. C.; Howard, A. J. Crystal structure of substrate-free Pseudomonas putida cytochrome P-450. *Biochemistry* **1986**, *25*, 5314–5322, PMID: 3768350.

(12) Lee, Y.-T.; Wilson, R. F.; Rupniewski, I.; Goodin, D. B. P450cam Visits an Open Conformation in the Absence of Substrate,. *Biochemistry* **2010**, *49*, 3412–3419, PMID: 20297780.

(13) Lee, Y.-T.; Glazer, E. C.; Wilson, R. F.; Stout, C. D.; Goodin, D. B. Three Clusters of Conformational States in P450cam Reveal a Multistep Pathway for Closing of the Substrate Access Channel,. *Biochemistry* **2011**, *50*, 693–703, PMID: 21171581.

(14) Groves, J. T. Key elements of the chemistry of cytochrome- P-450: The oxygen rebound mechanism. *J. Chem. Educ* **1985**, *62*, 928–931.

(15) Schlichting, I.; Berendzen, J.; Chu, K.; Stock, A. M.; Maves, S. A.; Benson, D. E.; Sweet, R. M.; Ringe, D.; Petsko, G. A.; Sligar, S. G. The Catalytic Pathway of Cytochrome P450cam at Atomic Resolution. *Science* **2000**, *287*, 1615–1622.

(16) Skinner, S. P.; Liu, W.-M.; Hiruma, Y.; Timmer, M.; Blok, A.; Hass, M. A. S.; Ubbink, M. Delicate conformational balance of the redox enzyme cytochrome P450cam. *Proc. Natl. Acad. Sci. U.S.A* **2015**, *112*, 9022–9027.

(17) Dunn, A. R.; Dmochowski, I. J.; Bilwes, A. M.; Gray, H. B.; Crane, B. R. Probing the open state of cytochrome P450cam with ruthenium-linker substrates. *Proc. Natl. Acad. Sci. U.S.A* **2001**, *98*, 12420–12425.

(18) Stoll, S.; Lee, Y. T.; Zhang, M.; Wilson, R. F.; Britt, R. D.; Goodin, D. B. Double electron-electron resonance shows cytochrome P450cam undergoes a conformational change in solution upon binding substrate. *Proc. Natl. Acad. Sci. U.S.A* **2012**, *109*, 12888–12893.

(19) Asciutto, E. K.; Young, M. J.; Madura, J.; Pochapsky, S. S.; Pochapsky, T. C. Solution structural ensembles of substrate-free cytochrome P450 cam. *Biochemistry* **2012**, *51*, 3383–3393.

(20) Colthart, A. M.; Tietz, D. R.; Ni, Y.; Friedman, J. L.; Dang, M.; Pochapsky, T. C. Detection of substrate-dependent conformational changes in the P450 fold by nuclear magnetic resonance. *Sci. Rep.* **2016**, *6*, 1–11.

(21) Basom, E. J.; Manifold, B. A.; Thielges, M. C. Conformational Heterogeneity and the Affinity of Substrate Molecular Recognition by Cytochrome P450cam. *Biochemistry* **2017**, *56*, 3248–3256.

(22) Savino, C.; Montemiglio, L. C.; Sciara, G.; Miele, A. E.; Kendrew, S. G.; Jemth, P.; Gianni, S.; Vallone, B. Investigating the Structural Plasticity of a Cytochrome P450: Three-Dimentional Structures of P450 EryK and Binding to its Physiological Substrates. *J. Biol. Chem.* **2009**, *284*, 29170–29179.

(23) Sherman, D. H.; Li, S.; Yermalitskaya, L. V.; Kim, Y.; Smith, J. A.; Waterman, M. R.; Podust, L. M. The Structural Basis for Substrate Anchoring, Active Site Selectivity, and Product Formation by P450 PikC from Streptomyces venezuelae. *J. Biol. Chem.* **2006**, *281*, 26289–26297.

(24) Zhao, Y.; White, M. A.; Muralidhara, B. K.; Sun, L.; Halpert, J. R.; Stout, C. D. Structure of Microsomal Cytochrome P450 2B4 Complexed with the Antifungal Drug Bifonazole: Insights into P450 Conformational Plasticity and Membrane Interaction. *J. Biol. Chem.* **2006**, *281*, 5973–5981.

(25) Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* **2018**, *140*, 2386–2396, PMID: 29323881.

(26) Bowman, G. R.; Pande, V. S.; Noé, F. An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation. Adv. Exp. Med. Biol. 2014.

(27) Chodera, J. D.; Noé, F. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struc. Biol.* **2014**, *25*, 135–144.

(28) de Montellano, P. R. O., Ed. *Cytochrome P450: Structure, Mechanism and Biochemistry*; Springer International Publishing, 2015.

(29) Sali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol* **1993**, *234*, 779 – 815.

(30) Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem* **2004**, *25*, 1400–1415.

(31) 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.

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

(33) Hoover, W. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695.

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

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

(36) Darden, T.; York, D.; Pederson, L. G. *J. Chem. Phys.* **1993**, *98*, 952.

(37) 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.

(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) Lloyd, S. Least Squares Quantization in PCM. *IEEE Trans. Inf. Theor.* **2006**, *28*, 129–137.

(40) 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.

(41) Polizzi, N. F.; Therien, M. J.; Beratan, D. N. Mean First-Passage Times in Biology. *Israel Journal of Chemistry* **2016**, *56*, 816–824.

(42) Singhal, N.; Pande, V. S. Error analysis and efficient sampling in Markovian state models for molecular dynamics. *J. Chem. Phys.* **2005**, *123*, 204909.

(43) E., W.; Vanden-Eijnden, E. Towards a Theory of Transition Paths. *Journal of Statistical Physics* **2006**, *123*, 503.

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

(45) Noé, F.; Schütte, 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.

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

(47) Rydzewski, J.; Nowak, W. Thermodynamics of camphor migration in cytochrome P450cam by atomistic simulations. *Sci. Rep.* **2017**, *7*, 7736.

(48) Markwick, P. R. L.; Pierce, L. C. T.; Goodin, D. B.; McCammon, J. A. Adaptive Accelerated Molecular Dynamics (Ad-AMD) Revealing the Molecular Plasticity of P450cam. *J. Phys. Chem. Lett* **2011**, *2*, 158–164, PMID: 21307966.

(49) Gotoh, O. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J. Biol. Chem.* **1992**, *267*, 83–90.

(50) Meharenna, Y. T.; Li, H.; Hawkes, D. B.; Pearson, A. G.; De Voss, J.; Poulos, T. L. Crystal Structure of P450cin in a Complex with Its Substrate, 1,8-Cineole, a Close Structural Homologue to d-Camphor, the Substrate for P450cam,. *Biochemistry* **2004**, *43*, 9487–9494, PMID: 15260491.

(51) Pylypenko, O.; Schlichting, I. Structural Aspects of Ligand Binding to and Electron Transfer in Bacterial and Fungal P450s. *Annu. Rev. Biochem* **2004**, *73*, 991–1018, PMID: 15189165.

(52) Poulos, T. L.; Cupp-Vickery, J.; Li, H. In *Cytochrome P450: Structure, Mechanism, and Biochemistry*; de Montellano, P. R. O., Ed.; Springer US: Boston, MA, 1995; pp 125–150.

(53) Johnson, E. F.; Stout, C. D. Structural diversity of human xenobiotic-metabolizing cytochrome P450 monooxygenases. *Biochem. Biophys. Res. Commun* **2005**, *338*, 331 – 336, Celebrating 50 Years of Oxygenases.

(54) Scott, E. E.; He, Y. A.; Wester, M. R.; White, M. A.; Chin, C. C.; Halpert, J. R.; Johnson, E. F.; Stout, C. D. An open conformation of mammalian cytochrome P450 2B4 at 1.6-Å resolution. *Proc. Natl. Acad. Sci. U.S.A* **2003**, *100*, 13196–13201.

(55) Scott, E. E.; White, M. A.; He, Y. A.; Johnson, E. F.; Stout, C. D.; Halpert, J. R. Structure of Mammalian Cytochrome P450 2B4 Complexed with 4-(4-Chlorophenyl)imidazole at 1.9 angstrom Resolution: Insight into the range of P450

conformations and the coordination of redox partner binding. *J. Biol. Chem.* **2004**, *279*, 27294–27301.

(56) Li, H.; Poulos, T. The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. *Nat. Struct. Mol. Biol.* **1997**, *4*, 140–146.

(57) Zhao, B. et al. Binding of Two Flaviolin Substrate Molecules, Oxidative Coupling, and Crystal Structure of Streptomyces coelicolor A3(2) Cytochrome P450 158A2. *J. Biol. Chem.* **2005**, *280*, 11599–11607.

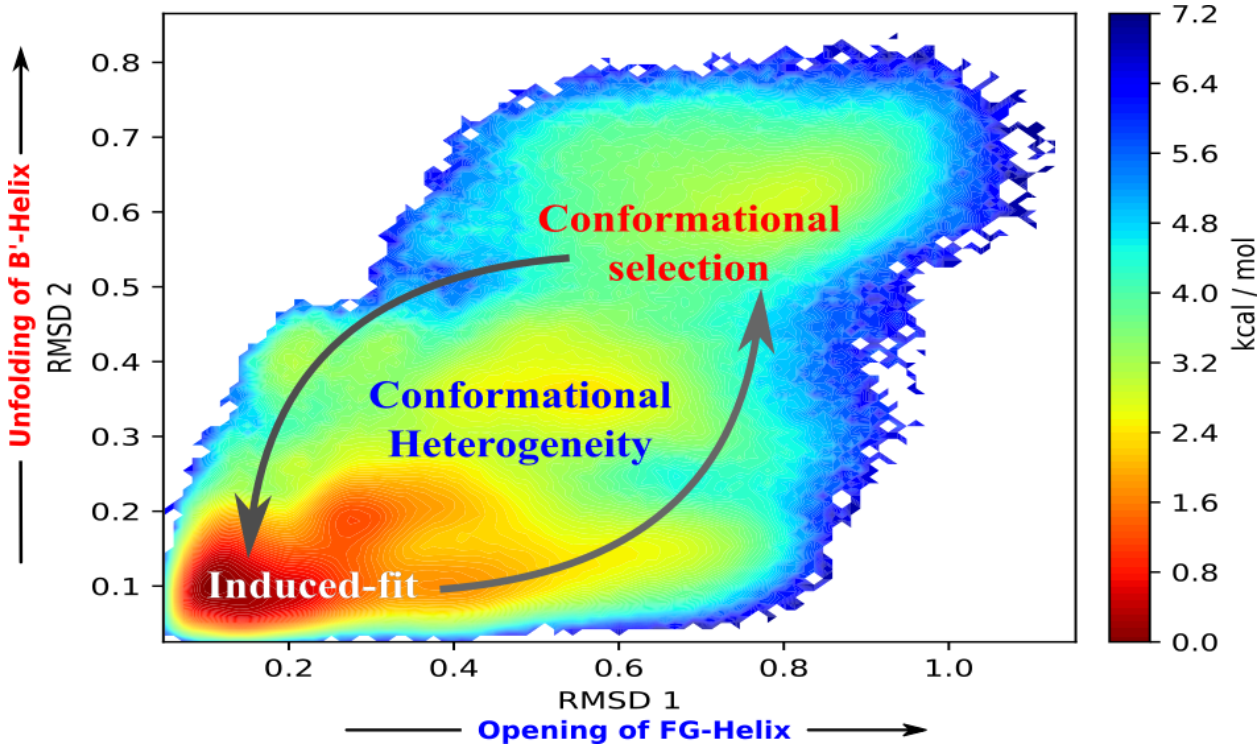
(58) Park, S.-Y.; Yamane, K.; ichi Adachi, S.; Shiro, Y.; Weiss, K. E.; Maves, S. A.; Sligar, S. G. Thermophilic cytochrome P450 (CYP119) from Sulfolobus solfataricus: high resolution structure and functional properties. *J. Inorg. Biochem* **2002**, *91*, 491 – 501.

(59) Xu, L.-H.; Fushinobu, S.; Ikeda, H.; Wakagi, T.; Shoun, H. Crystal Structures of Cytochrome P450 105P1 from Streptomyces avermitilis: Conformational Flexibility and Histidine Ligation State. *Journal of Bacteriology* **2009**, *191*, 1211–1219.

(60) Ouellet, H.; Podust, L. M.; de Montellano, P. R. O. Mycobacterium tuberculosis CYP130: Crystal Structure, Biophysical Characterisation, and Interactions With Antifungal AZOLE Drugs. *J. Biol. Chem.* **2008**, *283*, 5069–5080.

(61) Zerbe, K.; Pylypenko, O.; Vitali, F.; Zhang, W.; Rouset, S.; Heck, M.; Vrijbloed, J. W.; Bischoff, D.; Bister, B.; Sussmuth, R. D.; Pelzer, S.; Wohlleben, W.; Robinson, J. A.; Schlichting, I. Crystal Structure of OxyB, a Cytochrome P450 Implicated in an Oxidative Phenol Coupling Reaction during Vancomycin Biosynthesis. *J. Biol. Chem.* **2002**, *277*, 47476–47485.

(62) Williams, P. A.; Cosme, J.; Vinković, D. M.; Ward, A.; Angove, H. C.; Day, P. J.; Vonrhein, C.; Tickle, I. J.; Jhoti, H. Crystal Structures of Human Cytochrome P450 3A4 Bound to Metyrapone and Progesterone. *Science* **2004**, *305*, 683–686.


Unfolding of B'-Helix
RMSD 2
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Conformational
selection
Conformational
Heterogeneity
Induced-fit
0.2
0.4
0.6
0.8
1.0
RMSD 1
Opening of FG-Helix
7.2
6.4
5.6
4.8
4.0
3.2
2.4
1.6
0.8
0.0
kcal / mol

TOC Graphic