

SCIENTIFIC REPORTS

OPEN

Thermodynamics of camphor migration in cytochrome P450cam by atomistic simulations

J. Rydzewski & W. Nowak

Understanding the mechanisms of ligand binding to enzymes is of paramount importance for the design of new drugs. Here, we report on the use of a novel biased molecular dynamics (MD) methodology to study the mechanism of camphor binding to cytochrome P450cam. Microsecond-long MD simulations allowed us to observe reaction coordinates characterizing ligand diffusion from the active site of cytochrome P450cam to solvent via three egress routes. These atomistic simulations were used to estimate thermodynamic quantities along the reaction coordinates and indicate diverse binding configurations. The results suggest that the diffusion of camphor along the pathway near the substrate recognition site (SRS) is thermodynamically preferred. In addition, we show that the diffusion near the SRS is triggered by a transition from a heterogeneous collection of closed ligand-bound conformers to the basin comprising the open conformations of cytochrome P450cam. The conformational change accompanying this switch is characterized by the retraction of the F and G helices and the disorder of the B' helix. These results are corroborated by experimental studies and provide detailed insight into ligand binding and conformational behavior of the cytochrome family. The presented methodology is general and can be applied to other ligand-protein systems.

Cytochrome P450 enzymes are ubiquitous monooxygenases that contain heme¹. They are responsible for a variety of life processes including carcinogenesis and drug metabolism as well as the degradation of xenobiotics and biosynthesis of numerous natural products^{2,3}. Remarkably, 11 cytochrome P450 enzymes contribute to the biotransformation of 70% of clinically used drugs^{4,5}, which makes cytochrome P450 enzymes a potential template to produce biotechnologically relevant chemicals such as an antimalarial drug artemisinin, or statins⁶. The most well understood cytochrome P450cam from soil bacterium *Pseudomonas putida* catalyzes the hydroxylation of camphor (its natural substrate) to 5-hydroxy-camphor. This enzyme has been used as a model for many cytochromes P450 and was the first cytochrome P450 protein structure solved by X-ray crystallography⁷. The active site of cytochromes P450 is situated deep inside the protein matrix on the distal site of the heme prosthetic group, rather than exposed on the protein surface⁸. Thus, the crystal structures of substrate-bound cytochrome P450cam do not show any obvious access or exit channels along which ligands could migrate. Consequently, the enzyme specificity, thermodynamics and conformational states can be influenced by how substrates pass through multiple channels of the protein to access the active site and how products escape toward solvent.

In the proteins with no clear ligand transport pathways, ligand diffusion is challenging to model computationally, because the entrance and egress trajectories of the ligand generally involve migration through binding intermediates linking the dissociated and associated states^{9,10}. Furthermore, ligand diffusion is difficult to assess also experimentally due to the dynamical complexity of protein structures, and in the virtual absence of time-resolved crystallography experiments on ligand intermediates, the actual ligand expulsion pathways remain undetermined¹¹. And yet, detailed knowledge of binding and release mechanisms is fundamental for the comprehension of the biological function of enzymes. From a computational viewpoint, the main complication arises from two reasons. First, the reaction coordinate for a pathway that interpolates the transition between the bound and unbound state is difficult to assess. Second, the process is not at equilibrium. To sample ligand diffusion within timescales accessible in MD simulations computing nonequilibrium trajectories is necessary since ligand binding occurs on the microsecond-millisecond time scale¹². Thus, the ligand diffusion and its escape rates are accelerated, making it difficult to estimate thermodynamic and kinetic properties from the simulations. Because

Institute of Physics, Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University, Grudziadzka 5, 87-100, Torun, Poland. Correspondence and requests for materials should be addressed to J.R. (email: jr@fizyka.umk.pl)

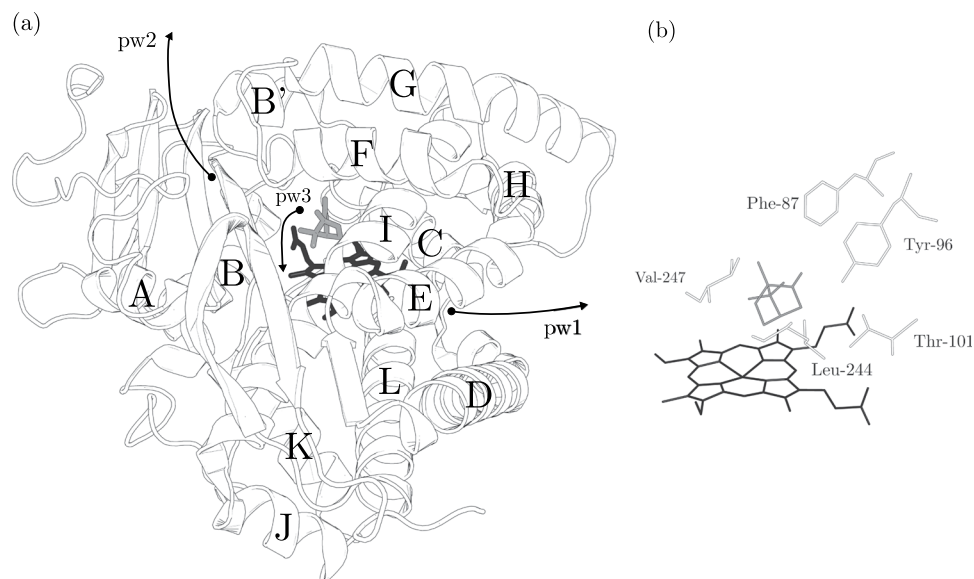


Figure 1. (a) Structure of cytochrome P450cam with camphor (gray) and heme (black). Helices are labeled A, B, B' and C–L. The diffusion exits of pw1–3 are indicated by arrows. (b) The active site with camphor at the crystallographic pose and the aligning residues in the distal heme moiety. For a detailed picture of the camphor diffusion from cytochrome P450cam, see Fig. S1 in the Supplementary Information.

of that, several methods have been proposed to tackle the ligand diffusion problem, ranging from steered MD¹³ and its variants^{14–16} through locally enhanced sampling^{17–19}, metadynamics^{20–22} and, recently, the memetic sampling (MS)^{9,23,24}.

Due to their significance, the ligand diffusion pathways have been identified previously and mechanisms of enzymatic trafficking have been investigated in various cytochromes. These studies have provided valuable insights into the ligand diffusion processes in cytochrome P450 enzymes^{15,16,25–31}. Recently, we have introduced the MS method capable of sampling the ligand diffusion pathways, in which the process of ligand diffusion from a protein is reconstructed by efficient conformational space sampling of the protein matrix in such a way that the obtained egress trajectories minimize effective interaction energy between the ligand and protein^{23,24}. In MS a biasing force with a direction determined by this minimization process is used to guide a ligand along a curved egress route. With MS one can calculate reaction coordinates of ligand diffusion which atomistically describe transitions of a ligand between bound and unbound states. Our simulations revealed three essential classes of the camphor diffusion trajectories in cytochrome P450cam. They were named pathways pw1, pw2 and pw3. The migration of camphor along pw1 involves passing between the I helix, known for its small thermal motions, and the distal side of the heme. The most frequent pathway (pw2) engages the region involving migration near the SRS site that is variable in sequence and structure between different cytochromes in the P450 family³². Its high flexibility in all P450s suggests a role in selectivity of a substrate. There are theoretical results suggesting that pw2 is preferred by substrates^{15,16,26,32}. Along pw3 camphor diffuses from the active site to the proximal side of heme, near the C and L helices. Despite this, transport along the pathways remains poorly understood, perhaps due to the absence of exhaustive sampling of the cytochrome configuration space that is crucial to compute thermodynamic quantities.

In what follows, we continued the research by aiming at understanding how the egress pathways and the fluctuations of the residues of cytochrome P450cam active in the transport contribute to the egress pathway preference and its thermodynamics. We addressed the camphor preference for each ligand diffusion pathway. In order to achieve this goal, we used an unprecedented combination of enhanced sampling methods, including calculating the reaction coordinates of ligand diffusion on-the-fly during MD simulations and metadynamics. To the best of our knowledge, we calculated for the first time the free energy along the previously identified camphor diffusion pathways. Combined with our previous studies of the camphor migration within the matrix of cytochrome P450cam and searching for escape via diffusion pathways, we obtained a full quantitative description of the camphor migration from the active site including structural and thermodynamic information.

Results

The present investigation builds on our prior studies of ligand diffusion in cytochrome P450cam and focuses on recovery of free-energy barriers along the identified camphor pathways and structural changes in the enzyme accompanying diffusion (cf. Methods).

Mechanism of the camphor diffusion in cytochrome P450cam. The crystal structure of cytochrome P450cam has 13 α -helices (A, B, B' and C–L) and five β -sheets (β 1– β 5) and it is shown in Fig. 1. We used terminology introduced by Poulos *et al.* in the article describing the crystallographic camphor-bound structure of cytochrome P450cam (PDB ID: 2CPP)³³. The active site of cytochrome P450cam is located above the proximal

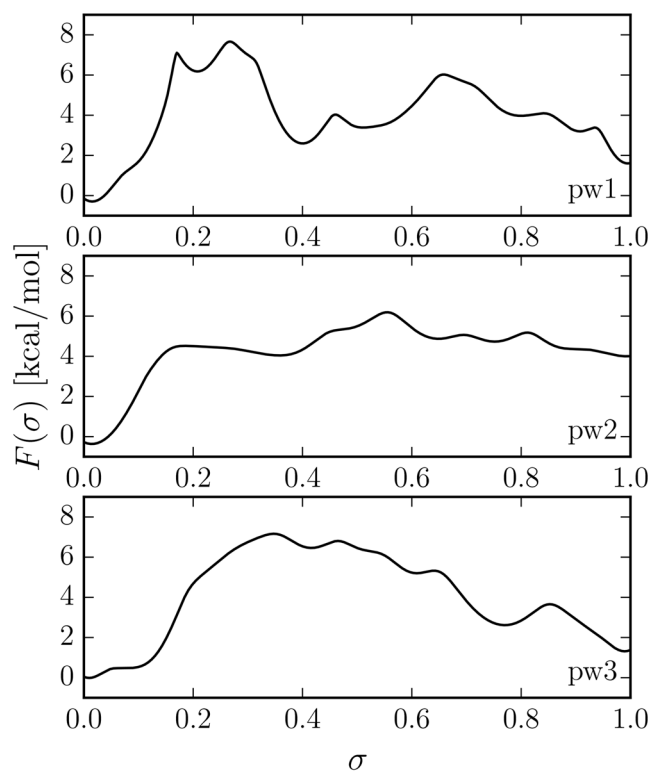


Figure 2. Free-energy profiles $F(\sigma)$ along the camphor diffusion pathways pw1-3. [The order parameter σ (cf. Methods) is introduced by scaling s].

helix (L) at the level of the distal helix (I). The initial position of camphor in all the metadynamics simulations is aligned by the B' helix (Tyr-96), B-B' loop (Phe-87) and I helix (Val-247 and Leu-244).

During the initial steps of the simulations the keto group of camphor accepts a hydrogen bond from the Tyr-96 hydroxyl group, which controls the position of camphor in the distal heme pocket. Camphor initiates its migration by an adjustment of the Tyr-96 hydroxyl group–hydrogen–camphor carbonyl oxygen angle. This change occurs on all the diffusion pathways, suggesting that it is a preferred position of camphor in the docking pocket. We run 10 ns unbiased MD simulations starting from this pose on each pathway confirming its thermodynamic stability (Table S2). Comparing to the crystallographic pose³³, the orientation of camphor toward the heme moiety is preserved although the distance between the Tyr-96 hydroxyl group oxygen and the carbonyl oxygen of camphor decreases from 2.9 Å to 2.5 Å, allowing camphor to closely interact with the B' helix in the active site. Frequently during the camphor diffusion, Tyr-96 adopted a new rotamer conformation, pointing out of the active site and vacating a space for the migrating camphor in the active site. This observation is corroborated by experimental studies in which Tyr-96 was shown to serve as a gateway to substrate binding and migration in the access channel³⁴. After the initial stage, camphor adopts a stable pose in the free-energy minimum. At this position, the residues from the relatively short B' helix (7 residues from 89 to 96) provide a mechanism of selective adjustment to camphor, contributing to free energy and substrate specificity. Next, the camphor diffusion leading to the dissociated state of camphor and cytochrome P450cam starts. The transition between the bound and dissociated states occurs through the multiple pathways facilitated by a number of nonnative transient hydrogen bonds, thus providing selective gating mechanisms.

pw1. The free-energy profiles along the camphor diffusion pathways pw1-3 are shown in Fig. 2. The migration of camphor along pw1 involves overcoming a region of small thermal motions. First, the camphor diffusion starts with an adjustment toward the I helix and the distal side of the heme group. The existence of the local maxima along pw1 was confirmed by running 50 short unbiased MD simulations for each maximum to check whether camphor diffuses to the neighboring local stable basins (Table S2). The highest free-energy barrier (≈ 8 kcal/mol at $\sigma \in \{0.17, 0.3\}$) comes from the passing of camphor between the I helix and heme and deforming Cys-242, Leu-245, Leu-246 and Val-247 (Table S1). Although the free-energy curve shows a shallow minimum ($\sigma \approx 0.2$), which may indicate a stable intermediate, we run a 10 ns unbiased MD simulation from this collective variable (CV) point and came to a conclusion that camphor migrates back to the active site. This feature in free energy is probably due to a relatively large error of free energy, which is the highest (< 1.2 kcal/mol, Fig. S6) in this CV range. After passing below the I helix camphor establishes a stable pose close to the C, D and E helices, which was confirmed by 10 ns unbiased MD simulation ($\sigma = 0.45 \pm 0.05$). The next free-energy local maximum (≈ 6 kcal/mol at $\sigma \approx 0.66$) is related to pushing past the D helix (Val-123, Lys-126, Leu-127 and Arg-130) at the verge of the enzyme structure. Next, camphor diffuses to solvent which is indicated by decreasing free-energy values.

Although pw1 was suggested as a potential ligand passage in several articles^{15, 16, 23, 24}, we found that the free-energy barrier occurring on this path is the highest (≈ 8 kcal/mol) in comparison to pw2 and pw3 (≈ 6 kcal/mol and ≈ 7 kcal/mol respectively), which makes the migration along pw1 less favorable thermodynamically for a ligand as bulky as camphor.

pw2. The diffusion of camphor along pw2 incorporates migration through the SRS near the F/G loop, BC loop, B' helix and the β_1 strand. The SRS site is conserved in various cytochromes from the P450 family, showing evidence for substrate specificity³⁵. The mechanism of egress along pw2 leads to the free-energy barrier (≈ 4.5 kcal/mol) which has the lowest value to escape from the distal side of heme comparing to pw1 and pw3 (Fig. 2). This fact indicates a preference in choosing pw2 over the other pathways. Metadynamics simulations indicate that the highest free-energy value along pw2 comes from perturbing Tyr-96 and Phe-87 from the B' helix (Table S1), whose aromatic side-chains gate the camphor migration (≈ 4.5 kcal/mol at $\sigma \approx 0.17$). These findings are in agreement with Winn *et al.*³², showing that the side-chain torsions of Phe-87 and Phe-193 contribute to the camphor transport along pw2. This ratcheting/gating mechanism is thus twofold. First, the side-chains of the B' helix and B-B' loop provide steric barriers along pw2, and second, donor a hydrogen bond (Tyr-96) to the keto group of camphor.

On the other hand, the camphor egress along pw2 to solvent seems to be more complex than previously described. Namely, after reaching the free-energy maximum, camphor moves toward Ile-395 and Phe-193, and Thr-192 which also donates a hydrogen bond to camphor (Table S1). Before camphor exits to solvent, it is thermodynamically stabilized in a metastable state guarded by Tyr-29, which acts similar to the gates in the active site. From the above picture, one can see that Tyr-69, Phe-87 ($\sigma \approx 0.2$, an intermediate state for < 37 ps) and Tyr-29 ($\sigma \approx 0.84$, a transition point for < 578 ps) provide surprisingly complex gating mechanism. Namely, after camphor moves past Tyr-29, the side-chain of Tyr-29 rotates toward the leaving camphor, thus recurring on the diffusion pathway. This ratcheting movement suggests that the described gating mechanism is bidirectional. The values of free energy near $\sigma \approx 1$ are high comparing to the other diffusion pathways, since Asp-188, Thr-192, Glu-91, Asn-30 and Pro-31 located near the exit of pw2 are in close proximity of each other, creating a narrow gauge, and thus prohibiting camphor from leaving. At this value of σ camphor is still interacting with the neighboring amino acids of cytochrome P450cam.

The results show that salt link tethering plays a role in the preservation of the diffusion pathway. Namely, salt links stabilize the enzyme structure in the proximity of the camphor migration pathway, mostly on the helix-rich side of cytochrome P450cam, Asp-97–Lys-197, Asp-188–Lys-392 and Asp-251–Arg-186. The salt links loosen occasionally while disrupted by the migrating camphor, but relax to their initial positions after camphor moves toward the exit. We observed that the residues that participate in creating salt links establish three patches along pw2. The Asp-251–Arg-186 salt link is created between the I helix and F-G loop in close proximity to the active site. Next, the Asp-97–Lys-197 salt link stabilizes pw2 near the B'-C loop. Perhaps the most interesting is the Asp-188–Lys-392 salt link which is located near the exit route to solvent. It is the least deviated patch during diffusion (the mean distance of the salt link is ≈ 3.2 Å) that contributes to the formation of the narrow gauge at the exit.

Although there are calculations showing that the transport along pw2 may be divided into several subclasses depending on the exit chosen by camphor³⁵, during our metadynamics simulations we do not observe any tendency to diverge from the exit point of the reaction pathway. We suppose that this is strictly related to the algorithm used to uncover reaction coordinates. For instance, several pathways similar to pw2 (pw2a-c) were determined using random acceleration MD (in which camphor was accelerated in a purely random direction)¹⁵, which did not identify optimal reaction coordinates. Moreover, metadynamics simulations converge to the minimum free-energy pathway, thus providing direct quantitative values of free energy along the pathway.

pw3. The transport along pw3 begins from the adjustment of the camphor pose, as described in the case of pw1 and pw2. Subsequently, the side-chain of Tyr-96 rotates in the direction of Phe-87 causing an increase in free energy (≈ 7.2 kcal/mol). After Tyr-96 vacates a space for the movement along pw3, camphor establishes a hydrogen bond with Thr-101. Camphor occupies the place previously taken by the side-chain of Tyr-96 in the proximity of Arg-299, which results in a local decrease of free energy (at $\sigma \approx 0.5$). At this stage, the ligand is stabilized by the interaction with Thr-101 (Table S1). Next, the hydrogen bond with Thr-101 breaks, resulting in leaving the metastable state. At this point, camphor is in the vicinity of the heme propionate groups. Next, camphor reaches the proximal side of heme near Tyr-75 and Tyr-78 and migrates to solvent in the proximity of the C and L helices.

This pathway is noticeably shorter than the other pathways, which indicates that the number of transient hydrogen bonds between camphor and the aromatic side-chains along pw3 should be lower. In fact, this can be observed. On the other hand the rigidity of the heme group contributes to the free-energy barrier. The results show that the transport along pw3 is disfavored since the movement of camphor disrupts partially the hydrogen bonds between the heme propionate groups and Arg-299. Therefore, the tunnel leading from the active site to the proximal side of heme near the C and L helices is improbable for such a rigid and nonpolar ligand like camphor.

Induced-fit mechanism. In cytochrome P450cam, the mechanism for substrate binding depends on the size of the ligand³². According to Markwick *et al.*^{36, 37} larger ligands known to interact with the enzyme ingress the cytochrome P450cam active site and trap the ligand-bound system in an open conformation via a population shift mechanism³⁸. Small ligands (e.g., camphor) fully enter the binding pocket which causes the induced-fit mechanism, leading to the transition from the open to the thermodynamically-preferable closed state. Many X-ray structures of cytochrome P450cam in the ligand-bound state indicate that the conformational change triggered by the binding of substrates involves the movement of the H, G, F and B' helices, which are in close proximity to the SRS³². The SRS is highly conserved among the cytochrome family, showing the importance of this region for functioning of the enzyme³⁹, which includes the ligand diffusion within the enzymatic matrix. Our results

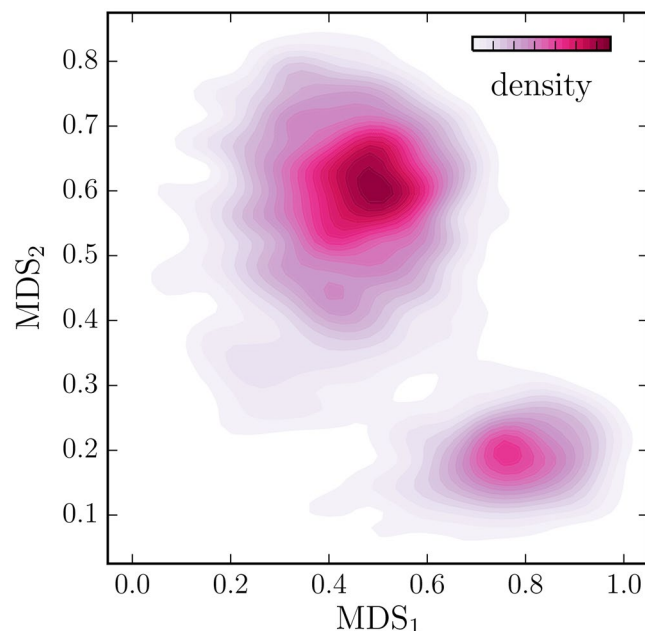


Figure 3. Conformational space of the camphor migration along pw2 in the proximity to the putative substrate recognition site of cytochrome P450cam retrieved using the multidimensional scaling. The conformations of the enzyme are split into the open ($\{0.5, 0.6\}$) and closed ($\{0.8, 0.2\}$) states. For the clustering of these conformers into the well-defined collections see Figs S2–S5 of the Supplementary Information.

indicate that the ligand diffusion pathway which engages in probing the SRS is pw2, and it is therefore crucial to investigate whether the conformations of the enzyme sampled along pw2 with metadynamics can be clustered into the open and closed states.

The cytochrome P450cam $C\alpha$ atoms of the conformations sampled by metadynamics along the three accessible ligand diffusion pathways subjected to the multidimensional scaling (MDS) indicate that the majority of the conformations occupy two basins (Fig. 3; for the full conformational space, we refer to Fig. S2 of the Supplementary Information). In order to assign the conformers to a particular basin, we performed clustering of the data embedded onto the two-dimensional space. As expected, our results show that the conformers can be separated into two states—open and closed. A majority of the conformations occupy the open basin (0.681) in comparison to the closed ligand-bound basin (0.213). For details concerning the DBSCAN clustering, see Figs S2–S5 in the Supplementary Information. The closed state is located at $\{0.8, 0.2\}$ and the open state at $\{0.5, 0.6\}$ as indicated in Fig. 3.

The metadynamics simulations start from probing the closed conformational basin in which camphor is bound within the enzymatic matrix. This basin is smaller than the open state basin, and the density of conformations is lower, suggesting that the diffusion of camphor within cytochrome P450cam is triggered by a jump of a trajectory from the closed to open state and its related conformational change. The closed basin is more structurally homogeneous comparing to the open basin, which comprises the need of a well-defined collection of conformers for the ligand-bound state. We show that the camphor binding selectively stabilizes the closed cytochrome P450cam conformers which agrees with IR vibrational echo spectroscopy experiments⁴⁰. This result is in agreement with the analysis of the crystal structures of cytochrome P450cam³⁶ and double electron-electron resonance study that demonstrates that the substrate-free enzyme prefers the open conformation and that the camphor binding results in conversion to the closed state^{41,42}. We observed that the main conformational change in the enzyme related to the diffusion of camphor along pw2 is an upward motion of the F, G and H helices, triggering salt-link tethering in close proximity of the SRS (Fig. 4) as shown previously using paramagnetic NMR spectroscopy⁴³. Also the B' helix undergoes an adjustment of its residues to the migrating camphor. We examined the X-ray structures of cytochrome P450cam with different ligands (PDB IDs: 1RE9, 1RF9, 3P6T, 3P6X) by calculating root-mean-square distance to conformations probed by the simulations. The X-ray structures fall into the closed state which is in agreement with a study of Markwick *et al.* and experimental studies⁴⁴. Moreover, we investigated the X-ray structure of cytochrome P450cam with two camphor molecules bound in the open state conformation (PDB ID: 4JX1)³⁴. This crystal structure falls into the open basin and depicts that a key feature in trapping a second camphor molecule in the pw2 exit route is a rotation of Tyr-96 in the opposite direction of the active site, which is also observed in our simulations.

We found that cytochrome P450cam exists in two well defined basins in the conformational space during the camphor diffusion, which agrees with the notion that the functional interplay between the enzyme conformers exists between the open and closed conformations due to the induced-fit mechanism^{38,43}. Surprisingly, these basins are sampled only by the conformers neighboring the ligand migration along pw2. This shows the

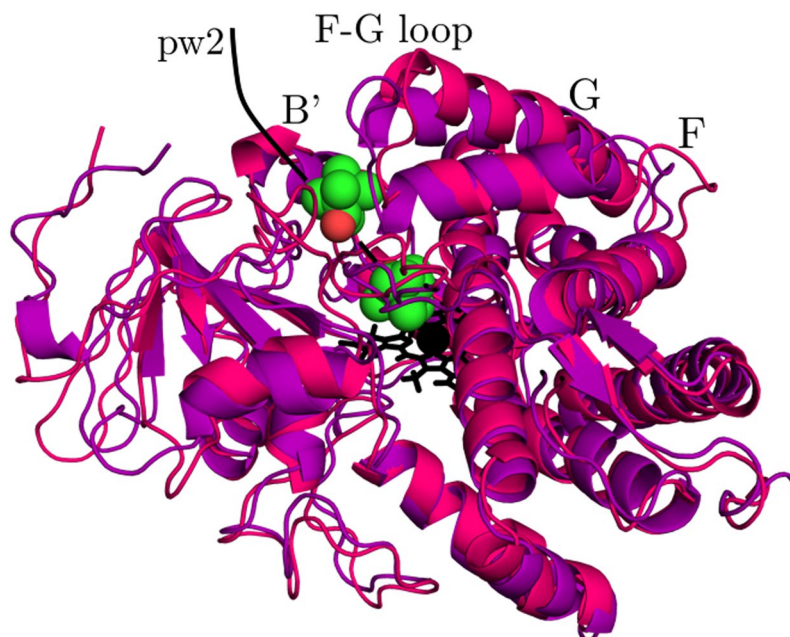


Figure 4. Representative conformations of cytochrome P450cam from the open and closed basins sampled during the diffusion along pw2 in the proximity of the putative substrate recognition site recovered using the multidimensional scaling. The open and closed conformations clustered into the open and closed states by DBSCAN are shown in pink and purple, respectively. The root-mean-square distance between these structures is 3.15 Å.

importance of pw2 in the diffusion of ligands and the functional motions of the enzyme indicating its molecular plasticity.

Conclusions

In conclusion, we introduced a novel computational methodology for characterizing complex structural transitions of ligands in enzymes that allowed for atomic-level reconstruction of the reaction coordinates of ligand diffusion and free-energy profiles along these pathways. The detailed information provided by the performed simulations can be used to design new experiments, for instance, to test potential inhibitors of the cytochrome family that are suggested to alter drug metabolism for individuals on certain medications. We note that our methodology is general and can be applied to other ligand-protein complexes.

We presented an investigation directed toward uncovering the ligand diffusion mechanism of camphor in cytochrome P450cam during both biased and unbiased all-atom MD simulations in explicit solvent. To calculate the reaction coordinates describing transitions between the active site of cytochrome P450cam and three alternative exit routes we used a novel method, which sampled the conformations via biased MD simulations by minimizing effective interaction energy between camphor and cytochrome P450cam^{23,24}. The configuration space was explored by biasing pathways defined based on the previously identified reaction coordinates by the well-tempered metadynamics with the adaptive Gaussians, which rendered the free-energy profiles along the reaction coordinates in a quantitative manner.

The results of MD simulations of P450cam suggest that the diffusion of camphor along pw2 near the SRS is thermodynamically preferred over pw1 and pw3. The camphor diffusion along pw1 deviates the heme group and the I helix, and along pw3 disrupts the hydrogen bonds between the heme propionate groups and cytochrome P450cam. For these reasons the free-energy barriers along pw1 and pw3 are higher compared to free-energy values met along pw2. The egress mechanism along each pathway involves a ratcheting movement of camphor in the proximity of gating side chains of aromatic residues, by either passing or pausing camphor. Furthermore, the conformational flexibility of the active site in the crystallographic structure enables the adjustment of camphor to Tyr-96 resulting in stronger interactions between the active site and the camphor carbonyl oxygen.

We found that the migration near the SRS triggers a transition from a heterogeneous collection of closed ligand-bound conformers to the basin comprising of the open conformations of cytochrome P450cam. The presence of the two basins indicates that these two basins consist of the structures of cytochrome P450cam engaging in the induced-fit mechanism. Upon camphor binding the enzyme sinks into the closed basin which is a collection of heterogeneous conformers, leading to the protein zipping up around the bound substrate. The conformational change accompanying the switch between the closed and open conformers is characterized by the retraction of the F and G helices and the disorder of the B' helix. These results are corroborated by experimental studies^{40,41,44} and potentially provide detailed insight into ligand migration and conformational behavior of the cytochrome family.

pathway	simulation time [μ s]	no. of frames	$\langle \text{MSD} \rangle^a$ [\AA^2]	λ [\AA^{-2}]	d^b [\AA]	$\langle \text{SE} \rangle^c$ [kcal/mol]
pw1	0.30	11	14.41	0.16	21	0.2
pw2	0.20	9	13.03	0.18	26	0.1
pw3	0.25	11	6.31	0.36	15	0.1

Table 1. Reaction coordinates representing 3 pathways of the camphor diffusion from the cytochrome P450cam active site. ^aMean-square deviation of the displacement set. ^bDistance between the center of mass of camphor and Fe. ^cMean standard error of free energy along the pathway after 100 ns of the simulation.

Methods

Metadynamics simulations. Before running metadynamics simulations the camphor-cytochrome P450cam complex (PDB ID: 2CPP)³³ was minimized and equilibrated through 5 ns MD simulation under NPT conditions at 1 atm and 300 K (the velocity-rescaling thermostat⁴⁵ and Berendsen's barostat⁴⁶ using the CHARMM27 force field⁴⁷ as implemented in Gromacs-5.0.6⁴⁸. Standard parameters were used for camphor and the heme group^{2,49,50} including an explicit Fe-S bond between the heme group and Cys-357⁵¹. Point charges for camphor were collected from the article of Schöneboom *et al.*⁵². All the simulations were run using the TIP3P water model⁵³ in electrically neutralized environment with periodic boundary conditions. The long-range electrostatics was computed by the particle mesh Ewald (PME) method⁵⁴. Bonds were constrained holonomically by using the P-LINCS algorithm⁵⁵. Three arbitrary $C\alpha$ atoms of cytochrome P450cam (Met-261, Trp-374, Ser-141) were harmonically restrained in each direction in order to prevent the enzyme from movement caused by the diffusing camphor. They were not a part of any conformational change during the diffusion process. To prohibit camphor from migrating from the cytochrome P450cam matrix to solvent, additional constraints for the distance between Fe and the center of mass of camphor were added for each exit pathway (Table 1).

The time-independent estimator of free energy $F(\xi)$ at time t of the free-energy profiles $F(\xi, t)$ as a function of CVs ξ was determined by metadynamics⁵⁶ in the well-tempered variant⁵⁷ using $F(\xi, t) = -\frac{T+\Delta T}{\Delta T} V(\xi, t)$, where $V(\xi, t)$ is the biasing potential added to the system and T is the temperature used in the simulation. The difference between the temperature of the CV and the temperature of the simulation is denoted by ΔT . The bias potential is made by the sum of the Gaussians deposited along the trajectories of the CVs, $\xi(t)$. We also used the adaptive Gaussians⁵⁸. The free-energy profiles were reweighted by the time-dependent constant $c(t)$:

$$c(t) = \frac{1}{\beta} \ln \frac{\int d\xi e^{-\beta F(\xi)}}{\int d\xi e^{-\beta [F(\xi) + V(\xi, t)]}} \quad (1)$$

where β is the inverse temperature ($1/k_B T$), in order to achieve the time-independent estimator of free energy⁵⁹.

For a detailed description of metadynamics, we refer to several reviews^{60,61}. Metadynamics simulations were carried out in NVT ensemble using the PLUMED2⁶². During the course of metadynamics simulations, we used the following parameters. The adaptive Gaussians were set to be in the time domain and the width of the Gaussian was set to 0.25 ps. The rate of the Gaussian deposition was 1.92 kcal/mol per ps. The temperature in the well-tempered protocol was 300 K and the ΔT was 3300 K.

The time-independent estimator of free energy was used to ascertain the local quality of convergence across the collective variable space⁵⁹. Additionally, the convergence of the free-energy calculations was monitored by plotting the Gaussian height as a function of the simulation time. The free energy was considered partially converged when the Gaussian height decreased to zero⁵⁷. Moreover, the full convergence required multiple recrossings of ξ during which the diffusion pathways were sampled several times in the collective variables space. The residual height of the Gaussian was less than 1% of the initial height after the simulation for each exit route. Therefore, the convergence of the free-energy profiles reweighted by $c(T)$ (cf. Methods) was stated for those simulations in which the residual height was less than 1% of the initial height and no further refinement of the reweighted free-energy profiles was observed. The standard error of free energy along the diffusion pathways after 100 ns was also calculated (Fig. S6 in the Supplementary Information).

Conformations sampled during the metadynamics simulations ($C\alpha$ atoms) were subjected to the multidimensional scaling⁶³ (MDS) to embed the multidimensional simulation data into a 2-dimensional space and subsequently clustered using DBSCAN (density-based spatial clustering of applications with noise), which is an algorithm used to cluster data with noise based on densities⁶⁴. For details including the specific selection of input parameters, please see Figs S2–S5 of the Supporting Information.

Reaction coordinates. Each diffusion pathway is atomistically characterized by the reaction coordinate that consists of conformations $S(1), \dots, S(N)$, where N is the number of conformations describing the transition between the bound and unbound states, and S is a conformation of a ligand-protein complex. Each reaction coordinate was computed in our previous article²⁴ (Fig. S1) by means of MS²³. In MS, a biasing force is used to guide the ligand along the diffusion pathway, toward the k th conformation (Fig. S7). The biasing force is maintained during m steps of MD. The k th conformation is determined on-the-fly by minimizing the functional Λ which describes effective interaction energy between the protein and ligand. We used Λ of the following form:

$$\Lambda = \sum_{i < j} h_{ij} e^{-\frac{r_{ij}^2}{2w^2}} + \gamma, \quad (2)$$

where i and j indicate atoms indices of the ligand and the protein, respectively. In Eq. 2 the distance between the atoms is denoted by r_{ij} . The height of the Gaussian is defined as $h_{ij} = s_i v_j + s_j v_i$, where s and v are the partial volume and the solvation coefficient of the given atoms. The width of the Gaussian is given by w . Additionally, we added the term γ describing the van der Waals interactions and electrostatics between the ligand and protein to Λ . The algorithm for finding k th conformation $S(k)$ (given $S(k-1)$ is known) involves the following steps:

1. The k th ligand conformation in a local neighborhood of the $(k-1)$ th ligand conformation is found by minimizing Λ . [The local domain is assured by sampling the ligand conformations inside a sampling sphere centered at the $(k-1)$ th conformation];
2. The ligand is pulled by the biasing force in m steps of the MD simulation in the direction of the k th ligand conformation;
3. After m steps of an MD simulation, the conformation of the system consisting of the k th ligand conformation and the protein that is the course of the MD simulation adjusted to the moving ligand, is saved as the k th conformation $S(k)$.

Details concerning minimizing the functional Λ can be found in refs 23, 24. The conformations defining the three exit routes of camphor (Fig. S1) were used to perform metadynamics simulations in a path collective variable (PCV) space.

Path collective variables. The camphor diffusion pathways in cytochrome P450cam were studied using PCVs⁶⁵. These pathways were represented in the space of the mean-square deviation of the camphor and P450cam coordinates. PCVs require two sets of atoms for the metadynamics simulations. The alignment set was used to superimpose the reaction coordinates with the current state of the system, whereas the displacement set is employed to compute the mean-square displacement between the aligned conformations. The sets for alignment and displacement consisted of the C α atoms of the I, G and C helices, and camphor, respectively. These sets were modeled after on previous calculations of the ligand diffusion^{20, 66, 67}.

The reactions coordinates describe 3 pathways between the bound and unbound states. To follow these pathways, $s(\mathbf{R})$ and $z(\mathbf{R})$ PCVs were used^{20, 65}:

$$s(\mathbf{R}) = \frac{\sum_{k=1}^N k e^{-\lambda \|S(\mathbf{R}) - S(k)\|^2}}{\sum_{k=1}^N e^{-\lambda \|S(\mathbf{R}) - S(k)\|^2}} \quad (3)$$

and

$$z(\mathbf{R}) = -\frac{1}{\lambda} \ln \sum_{k=1}^N e^{-\lambda \|S(\mathbf{R}) - S(k)\|^2}, \quad (4)$$

where $S(\mathbf{R})$ is a reduced representation of a configuration \mathbf{R} . The progress along the reaction coordinate is computed by $s(\mathbf{R})$; the distance (deviation) from the reaction coordinate is measured by $z(\mathbf{R})$. The λ values for each reaction coordinate are shown in Table 1. Typically, with a reasonable guess of the reaction coordinate values of $z(\mathbf{R})$ should be small which means that the system does not deviate much from the reaction coordinate. We performed metadynamics only in the space of $s(\mathbf{R})$ while $z(\mathbf{R})$ was constrained to $z(\mathbf{R}) < 6\text{\AA}^2$. This choice gives the possibility for the system to explore conformations different from S , however maintaining at the same time the system reasonably close to the chosen reaction coordinates.

References

1. Thomas, L. P. Heme enzyme structure and function. *Chem. Rev.* **114**, 3919–3962 (2014).
2. Ilme Schlichting, J. *et al.* The catalytic pathway of cytochrome P450cam at atomic resolution. *Science* **287**, 1615–1622 (2000).
3. Sheel, C. D. *et al.* Discovery of a regioselectivity switch in nitrating P450s guided by molecular dynamics simulations and Markov models. *Nat. Chem.* **8**, 419–425 (2016).
4. Nair, P. C., McKinnon, R. A. & Miners, J. O. Cytochrome P450 structure–function: Insights from molecular dynamics simulations. *Drug Metab. Rev.* **48**, 434–452 (2016).
5. Peter Guengerich, F. Cytochrome P450 and chemical toxicology. *Chem. Res. Toxicol.* **21**, 70–83 (2007).
6. Munro, A. W., Girvan, H. M., Mason, A. E., Dunford, A. J. & Kirsty, J. M. L. What makes a P450 tick? *Trends Biochem. Sci.* **38**, 140–150 (2013).
7. Poulos, T. L. *et al.* The 2.6-Å crystal structure of *Pseudomonas putida* cytochrome P-450. *J. Biol. Chem.* **260**, 16122–16130 (1985).
8. Cojocar, V., Winn, P. J. & Wade, R. C. The ins and outs of cytochrome P450s. *Biochim. Biophys. Acta* **1770**, 390–401 (2007).
9. Rydzewski, J. & Nowak, W. Ligand diffusion in proteins via enhanced sampling in molecular dynamics. *Phys. Life Rev.* doi: [10.1016/j.plev.2017.03.003](https://doi.org/10.1016/j.plev.2017.03.003) (2017).
10. Baron, R. & McCammon, J. A. Molecular recognition and ligand association. *Annu. Rev. Phys. Chem.* **64**, 151–175 (2013).
11. Elber, R. Ligand diffusion in globins: Simulations versus experiment. *Curr. Opin. Struct. Biol.* **20**, 162–167 (2010).
12. Voß, B., Seifert, R., Kaupp, U. B. & Grubmüller, H. A quantitative model for camp binding to the binding domain of MloK1. *Biophys. J.* **111**, 1668–1678 (2016).
13. Grubmüller, H., Heymann, B. & Tavan, P. Ligand binding: Molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* **271**, 997–999 (1996).

14. Martinez, L., Polikarpov, I. & Skaf, M. S. Only subtle protein conformational adaptations are required for ligand binding to thyroid hormone receptors: Simulations using a novel multipoint steered molecular dynamics approach. *J. Phys. Chem. B* **112**, 10741–10751 (2008).
15. Lüdemann, S. K., Lounnas, V. & Wade, R. C. How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms. *J. Mol. Biol.* **303**, 797–811 (2000).
16. Lüdemann, S. K., Lounnas, V. & Wade, R. C. How do substrates enter and products exit the buried active site of cytochrome P450cam? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways. *J. Mol. Biol.* **303**, 813–830 (2000).
17. Cohen, J., Kim, K., King, P., Seibert, M. & Schulten, K. Finding gas diffusion pathways in proteins: Application to O₂ and H₂ transport in Cpi [FeFe]-hydrogenase and the role of packing defects. *Structure* **13**, 1321–1329 (2005).
18. Czereminski, R. & Elber, R. Computational studies of ligand diffusion in globins: I. Leghemoglobin. *Proteins* **10**, 70–80 (1991).
19. Elber, R. & Karplus, M. Enhanced sampling in molecular dynamics: Use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *J. Am. Chem. Soc.* **112**, 9161–9175 (1990).
20. Fidelak, J., Juraszek, J., Branduardi, D., Bianciotto, M. & Gervasio, F. L. Free-energy-based methods for binding profile determination in a congeneric series of CDK2 inhibitors. *J. Phys. Chem. B* **114**, 9516–9524 (2010).
21. Gervasio, F. L., Laio, A. & Parrinello, M. Flexible docking in solution using metadynamics. *J. Am. Chem. Soc.* **127**, 2600–2607 (2005).
22. Provati, D., Bortolato, A. & Filizola, M. Exploring molecular mechanisms of ligand recognition by opioid receptors with metadynamics. *Biochemistry* **48**, 10020–10029 (2009).
23. Rydzewski, J. & Nowak, W. Memetic algorithms for ligand expulsion from protein cavities. *J. Chem. Phys.* **143**, 124101 (2015).
24. Rydzewski, J. & Nowak, W. Machine learning based dimensionality reduction facilitates ligand diffusion paths assessment: A case of cytochrome P450cam. *J. Chem. Theory Comput.* **12**, 2110–2120 (2016).
25. Lounnas, V. & Wade, R. C. Exceptionally stable salt bridges in cytochrome P450cam have functional roles. *Biochemistry* **36**, 5402–5417 (1997).
26. Schleinkefer, K., Sudarko, Winn, P. J., Lüdemann, S. K. & Wade, R. C. Do mammalian cytochrome P450s show multiple ligand access pathways and ligand channelling? *EMBO* **6**, 584–589 (2005).
27. Cojocar, V., Balali-Mood, K., Sansom, M. S. P. & Wade, R. C. Structure and dynamics of the membrane-bound cytochrome P450 2C9. *PLoS Comput. Biol.* **7**, e1002152 (2011).
28. Li, W., Shen, J., Liu, G., Tang, Y. & Hoshino, T. Exploring coumarin egress channels in human cytochrome P450 2A6 by random acceleration and steered molecular dynamics simulations. *Proteins* **79**, 271–281 (2011).
29. Vohra, S. *et al.* The dynamics of camphor in the cytochrome P450 CYP101D2. *Protein Science* **22**, 1218–1229 (2013).
30. Yu, X., Cojocar, V. & Wade, R. C. Conformational diversity and ligand tunnels of mammalian cytochrome P450s. *Biotechnol. Appl. Biochem.* **60**, 134–145 (2013).
31. Magistrato, A., Sgrignani, J., Krause, R. & Cavalli, A. Single or multiple access channels to the CYP450s active site? An answer from free energy simulations of the human aromatase enzyme. *J. Phys. Chem. Lett.* **8**, 2036–2042 (2017).
32. Winn, P. J. *et al.* Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine. *Proc. Natl. Acad. Sci. USA* **99**, 5361–5366 (2002).
33. Poulos, T. L., Finzel, B. C. & Howard, A. J. High-resolution crystal structure of cytochrome P450cam. *J. Mol. Biol.* **195**, 687–700 (1987).
34. Tripathi, S., Li, H. & Poulos, T. L. Structural basis for effector control and redox partner recognition in cytochrome P450. *Science* **340**, 1227–1230 (2013).
35. Wade, R. C., Winn, P. J., Schlichting, I. & Sudarko A survey of active site access channels in cytochromes P450. *J. Inorg. Biochem.* **98**, 1175–1182 (2004).
36. 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* **50**, 693–703 (2011).
37. Hays, A.-M. A. *et al.* Conformational states of cytochrome P450cam revealed by trapping of synthetic molecular wires. *J. Mol. Biol.* **344**, 455–469 (2004).
38. 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.* **2**, 158–164 (2011).
39. 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.* **267**, 83–90 (1992).
40. Thielges, M. C., Chung, J. K. & Fayer, M. D. Protein dynamics in cytochrome P450 molecular recognition and substrate specificity using 2D IR vibrational echo spectroscopy. *J. Am. Chem. Soc.* **133**, 3995–4004 (2011).
41. Stoll, S. *et al.* Double electron–electron resonance shows cytochrome P450cam undergoes a conformational change in solution upon binding substrate. *Proc. Natl. Acad. Sci. USA* **109**, 12888–12893 (2012).
42. Liou, S.-H., Mahomed, M., Lee, Y.-T. & Goodin, D. B. Effector roles of putidaredoxin on cytochrome P450cam conformational states. *J. Am. Chem. Soc.* **138**, 10163–10172 (2016).
43. Skinner, S. P. *et al.* Delicate conformational balance of the redox enzyme cytochrome P450cam. *Proc. Natl. Acad. Sci. USA* **112**, 9022–9027 (2015).
44. Lee, Y.-T., Wilson, R. F., Rupniewski, I. & Goodin, D. B. P450cam visits an open conformation in the absence of substrate. *Biochemistry* **49**, 3412–3419 (2010).
45. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
46. Tuckerman, M. *Statistical mechanics: Theory and molecular simulation* (Oxford University Press, 2010).
47. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *J. Comp. Chem.* **31**, 671–690 (2010).
48. Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19–25 (2015).
49. Miao, Y. & Baudry, J. Active-site hydration and water diffusion in cytochrome P450cam: A highly dynamic process. *Biophys. J.* **101**, 1493–1503 (2011).
50. Zheng, J., Altun, A. & Thiel, W. Common system setup for the entire catalytic cycle of cytochrome P450cam in quantum mechanical/molecular mechanical studies. *J. Comp. Chem.* **28**, 2147–2158 (2007).
51. Morozov, A. N., D’Cunha, Cassian, Alvarez, C. A. & Chatfield, D. C. Enantiospecificity of chloroperoxidase-catalyzed epoxidation: Biased molecular dynamics study of a cis- β -methylstyrene/chloroperoxidase-compound I complex. *Biophys. J.* **100**, 1066–1075 (2011).
52. Schöneboom, J. C. *et al.* The elusive oxidant species of cytochrome P450 enzymes: Characterization by combined quantum mechanical/molecular mechanical (QM/MM) calculations. *J. Am. Chem. Soc.* **124**, 8142–8151 (2002).
53. 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.* **79**, 926–935 (1983).
54. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
55. Hess, B. P-lincs: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **4**, 116–122 (2008).
56. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA* **99**, 12562–12566 (2002).

57. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100**, 020603 (2008).
58. Branduardi, D., Bussi, G. & Parrinello, M. Metadynamics with adaptive gaussians. *J. Chem. Theory Comput.* **8**, 2247–2254 (2012).
59. Tiwary, P. & Parrinello, M. A time-independent free energy estimator for metadynamics. *J. Phys. Chem. B* **119**, 736–742 (2014).
60. Laio, A. & Gervasio, F. L. Metadynamics: A method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.* **71**, 126601 (2008).
61. Valsson, O., Tiwary, P. & Parrinello, M. Enhancing important fluctuations: Rare events and metadynamics from a conceptual viewpoint. *Annu. Rev. Phys. Chem.* **67**, 159–184 (2016).
62. Gareth, A. *et al.* Plumed 2: New feathers for an old bird. *Comp. Phys. Comm.* **185**, 604–613 (2014).
63. Rohrdanz, M. A., Zheng, W. & Clementi, C. Discovering mountain passes via torchlight: Methods for the definition of reaction coordinates and pathways in complex macromolecular reactions. *Annu. Rev. Phys. Chem.* **64**, 295–316 (2013).
64. Ester, M. *et al.* A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD* **96**, 226–231 (1996).
65. Branduardi, D., Gervasio, F. L. & Parrinello, M. From A to B in free energy space. *J. Chem. Phys.* **126**, 054103 (2007).
66. Paloncýová, M., Navrátilová, V., Berka, K., Laio, A. & Otyepka, M. Role of enzyme flexibility in ligand access and egress to active site: Bias-exchange metadynamics study of 1, 3, 7-trimethyluric acid in cytochrome P450 3A4. *J. Chem. Theory Comput.* **12**, 2101–2109 (2016).
67. Limongelli, V. *et al.* Molecular basis of cyclooxygenase enzymes (COXs) selective inhibition. *Proc. Natl. Acad. Sci. USA* **107**, 5411–5416 (2010).

Acknowledgements

JR acknowledges funding (grants 2015/19/N/ST3/02171 and 2016/20/T/ST3/00488) from National Science Centre, Poland. The authors thank Michele Parrinello, Enrico Caldarulo and Ferruccio Palazzesi for helping at the initial stage of this project as well as suggestions from an anonymous reviewer that substantially improved this work. We are indebted to Swiss National Supercomputing Centre at Eidgenössische Technische Hochschule Zürich and Interdisciplinary Centre for Modern Technologies at Nicolaus Copernicus University in Torun for providing computer facilities.

Author Contributions

J.R. designed the research; J.R. performed the research; J.R. analyzed the data; J.R. and W.N. wrote the paper.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-07993-0

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017