THE PROTEIN SOCIETY

The dynamics of camphor in the cytochrome P450 CYP101D2

Shabana Vohra,[1] Maria Musgaard,[1] Stephen G. Bell,[2†] Luet-Lok Wong,[2] Weihong Zhou,[3*] and Philip C. Biggin[1*]

[1]Structural Bioinformatics and Computational Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, United Kingdom

[2]Inorganic Chemistry Laboratory, Department of Chemistry, University of Oxford, South Parks Road, Oxford, OX1 3QR, United Kingdom

[3]College of Life Sciences, Nankai University, Tianjin, 300071, China

Received 17 May 2013; Accepted 2 July 2013
DOI: 10.1002/pro.2309
Published online 5 July 2013 proteinscience.org

Abstract: The recent crystal structures of CYP101D2, a cytochrome P450 protein from the oligotrophic bacterium *Novosphingobium aromaticivorans* DSM12444 revealed that both the native (substrate-free) and camphor-soaked forms have open conformations. Furthermore, two other potential camphor-binding sites were also identified from electron densities in the camphor-soaked structure, one being located in the access channel and the other in a cavity on the surface near the F-helix side of the F-G loop termed the substrate recognition site. These latter sites may be key intermediate positions on the pathway for substrate access to or product egress from the active site. Here, we show via the use of unbiased atomistic molecular dynamics simulations that despite the open conformation of the native and camphor-bound crystal structures, the underlying dynamics of CYP101D2 appear to be very similar to other CYP proteins. Simulations of the native structure demonstrated that the protein is capable of sampling many different conformational substates. At the same time, simulations with the camphor positioned at various locations within the access channel or recognition site show that movement towards the active site or towards bulk solvent can readily occur on a short timescale, thus confirming many previously reported *in silico* studies using steered molecular dynamics. The simulations also demonstrate how the fluctuations of an aromatic gate appear to control access to the active site. Finally, comparison of camphor-bound simulations with the native simulations suggests that the fluctuations can be of similar level and thus are more representative of the conformational selection model rather than induced fit.

Keywords: CYP101D2; atomistic molecular dynamics; conformational selection

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Ministry of Science and Technology of China Project 973; Grant number: 2010CB530100 (to WZ); Grant sponsor: Natural Science Foundation of China; Grant number: 31170684 (to W.Z.). Grant sponsor: Oxford Centre for Integrative Systems Biology.

*Correspondence to: Weihong Zhou, College of Life Sciences, Nankai University, Tianjin 300071, China. E-mail: zhouwh@nankai.edu.cn. Philip C. Biggin, Structural Bioinformatics and Computational Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, United Kingdom.
E-mail: philip.biggin@bioch.ox.ac.uk
†Present address: School of Chemistry & Physics, University of Adelaide, Adelaide, 5005 Australia.

Introduction

The cytochrome P450 (CYP) superfamily of proteins are heme-containing monoxygenases involved in biosynthesis, biodegradation, and xenobiotic metabolism with significant pharmacological and biotechnological implications.[1–3] They usually catalyze the oxidation of a C–H bond in a wide range of organic substrates[4] and are present in both prokaryotes and eukaryotes forming a super family of enzymes with low sequence identity but an extremely conserved fold. Despite decades of research the mechanism of substrate recognition and entry into the active site is still poorly

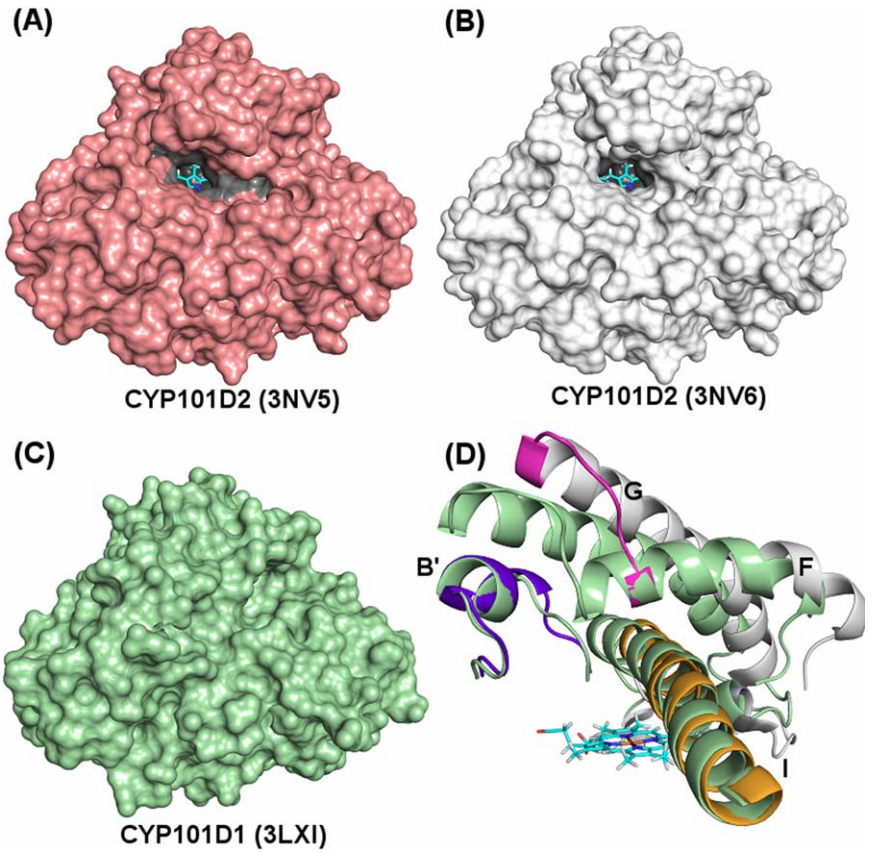

Figure 1. Surface representation for the native (A) and substrate (camphor) bound (B) structures of CYP101D2. Note that they are both "open" with a direct access route to the heme group (drawn in cyan liquorice representation). By comparison the camphor-bound structure of CYP101D1 (C) is clearly closed with no direct access to the heme group. The main reason for the difference is a shift in the orientation of helices F and G as shown in D. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

understood. X-ray analysis and molecular dynamics (MD) simulations have provided valuable insight into the structural features and variations in this family.[5,6] In CYP enzymes, the active site is buried deep inside the protein and isolated from the surrounding solvent. Protein dynamics is thus critical to substrate recognition, entry into the active site and exit of the product from the active site.[7]

In recent years, many crystal structures have been reported for various different CYPs, with the general trend that substrate-free (or native) structures are in an "open" state whereby the heme group appears to be accessible via a channel or tunnel to the bulk solvent, whereas substrate-bound structures tend to exhibit a "closed" structure in which the heme group is not directly accessible from bulk solvent.[8–14] As in previous studies, we use the term "native" to describe the protein when it is substrate free, but the heme group is still bound. Given the crystallographic information as a starting point, molecular simulation and NMR methods have suggested that the enzyme must undergo significant conformational changes to allow entry of the substrate and exit of the product from the active site. Further analyses demonstrated that the flexibility in the B, B′, F and G helices, F–G loops, and B–C loops leads to formation of different channels connecting the exterior solvent to the active site.[15–18]

It has been suggested that the formation of different channels/tunnels may provide specificity to a wide range of substrates and may influence the enzyme kinetics.[7,19,20] These channels have been classified via systematic analysis of previously determined structures.[21,27] Knowledge acquired from different studies suggests two theories for substrate binding and access into the active site. According to one model, the enzyme may adopt various conformations where substrate binding occurs by conformational selection[22–24] while the other suggests that the binding of the substrate induces conformational changes in the enzyme and the cavity opens up for the passage of the substrate.[25–27]

Previously, substrate-free and substrate-bound crystal structures of the bacterial cytochrome CYP101D2 from *Novosphingobium aromaticivorans* DSM12444 were reported.[28] Both structures were in similar open conformations (Fig. 1) with a root-mean squared deviation (RMSD) across all Cα atoms of 0.49 Å. Besides the camphor in the active site, the substrate-bound structure had two additional regions of weak electron density for camphor; one in the access channel near the B–C loop, F–G loop and

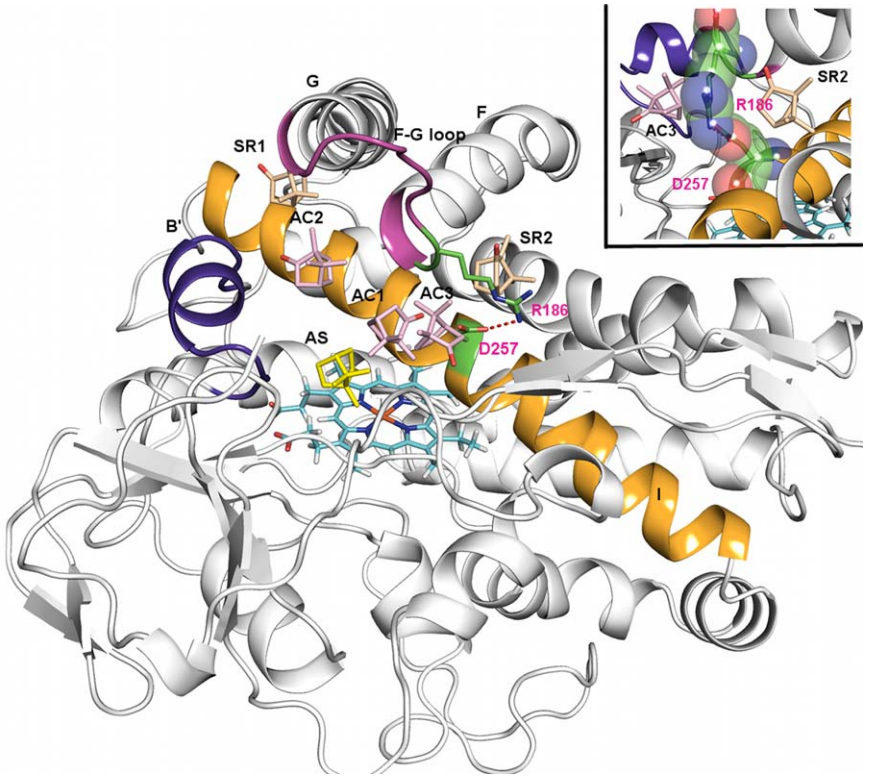

Figure 2. Cartoon representation of CYP101D2 illustrating the starting points of simulations with camphor starting in the active site (AS, camphor in yellow), in the access channel (AC1, AC2, and AC3, camphor in pink) and in the substrate recognition sites (SR1 and 2, camphor in pale yellow) respectively. The key parts of the protein with greatest flexibility are the B′ helix and the F-G loop (drawn in blue and magenta respectively). Helix I is drawn in gold for reference and the heme group is shown in cyan liquorice. The inset shows that a key salt-bridge between R186 and D257 blocks the SR2 site from the access-channel. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

I helix, and the other in the cavity on the surface near the F helix side of the F–G loop. Evidence of a second camphor binding site was reported in NMR studies of camphor-bound P450cam.[29] These intermediate locations may suggest the possible pathway for the substrate entry into, or product release from, the binding site. It was also noted that the orientation of the camphor molecule in the active site is not consistent with the formation of 5-*exo*-hydroxycamphor as the almost exclusive product.

Thus, questions that naturally arose from the crystal structure data were: (i) Is the substrate-bound open form of CYP101D2 conformationally stable or will it tend toward a closed conformation consistent with crystallographic studies of other P450 proteins such as the closely related CYP101D1?[28] (ii) Does the bound camphor reorient itself such that the C5 atom is over the heme iron which would be necessary for the formation of 5-*exo*-hydroxycamphor?, and (iii) What is the behavior of camphor in the intermediate sites? To address those questions, we used MD simulations.

Results

Previous studies of cytochromes have suggested that substrate-free structures may occur in open or closed states. However, substrate bound structures adopt a closed conformation for the oxidation of the substrate.[10,13] Both substrate-free and camphor-bound structures of the bacterial cytochrome, CYP101D2, were in an open conformation (Fig. 1). An obvious question is therefore whether this is a feature of the crystallographic conditions and whether the protein with substrate bound does prefer a closed conformation or whether this substrate-bound open conformation might be peculiar to this particular cytochrome. We therefore explored this through a series of MD simulations with camphor positioned at various locations within the protein (see "Materials and Methods" section for details and Fig. 2).

Simulation of the native structure

The native crystal structure was in an open state with a wide cleft connecting the active site to the protein surface [Fig. 1(A)]. As characterized previously,[6] the enzyme has a series of tunnels (see "Materials and Methods" section, Fig. 3 and Supporting Information) which change in dynamic fashion throughout the simulations and may be directly relevant to substrate access and product egress. In all the three simulation repeats, the F–G loop and B′ helix [Figs. 1(D) and S2] underwent large (of the order of a few Å) fluctuations, but the structure remained in an open, or partially open, state (Figs. 4 and 5). In repeat-1

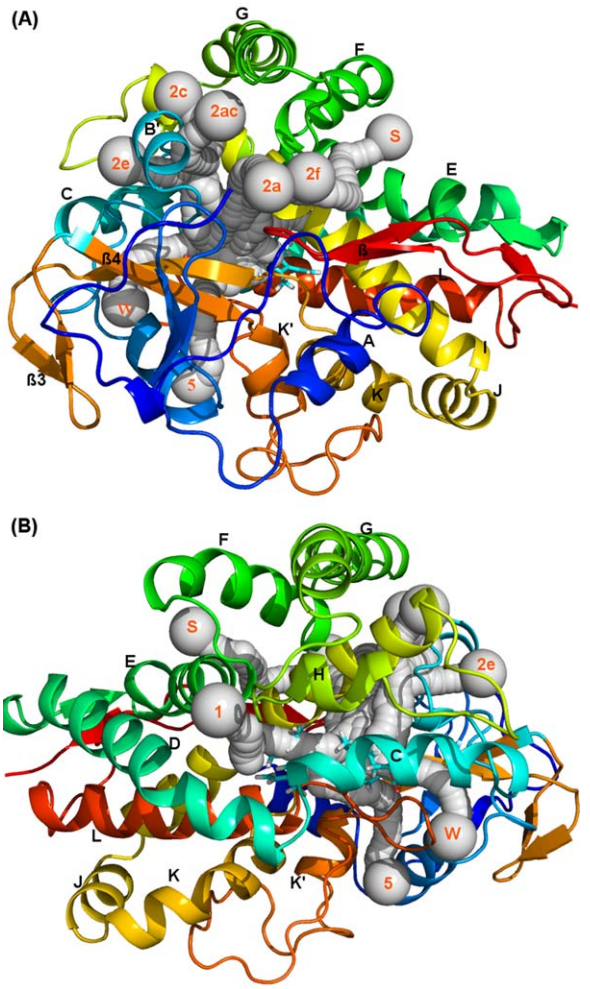

Figure 3. (A) CAVER analysis of the tunnels (coloured as a grey series of overlapping spheres) labeled according to the scheme from Cojocaru et al.[6] (B) Same analysis but with the view in A rotated about the long-axis of the page by 180°. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

and repeat-2, the structure sampled various open conformations (Fig. 5) with the F-G loop moving "outwards" forming a large tunnel wide enough to allow the entry of molecules like camphor (radius of ~2.1 Å). Tunnels 2a, 2f, and W appear to persist throughout both of these simulations (Fig. 3. and Supporting Information). In repeat-3, the structure visited both open and closed conformations. A hinge motion causes the F–G loop to form a lid over the cavity but the structure remained partially open via the 2f tunnel. Across the three repeats, the F–G loop and the B′ helix visited different conformations [Fig. 5(D)] forming various tunnels to the active site, suggesting that the native structure may adopt many conformational states for substrate recognition and entry into the active site.

Camphor in the active site

Coordinates for camphor in these simulations were obtained from the substrate-bound crystal structure (PDB ID: 3NV6). We monitored the centre of mass distances between camphor and the heme group in all simulations (Fig. 6). In the crystal structure, camphor was in the active site close to the heme group with the C5 atom of camphor [Fig. 7(A), yellow circle] at a distance of 7.5 Å pointing away from the central iron of the heme group. In all the three repeats, varying degrees of motions in the F–G loop and the B′ helix leads to closure of the cleft which is open in the crystal structure (Supporting Information). Rearrangement of several active site residues helped camphor to reorient and position the C5 carbon closer to the heme group. This is important as camphor is stereo-selectively oxidized at C5 to form 5-*exo*-hydroxycamphor. The radial distribution of the camphor C5 carbon around the heme iron showed a maximum distribution of C5 at a distance of 5.4 Å for repeat-2 and repeat-3, an ideal distance to accommodate an oxygen molecule between the camphor C5 carbon and the heme iron to undergo oxidation[30] [Fig. 7(B)].

In all repeats, the aromatic ring of F87 flipped over and blocked the exit of tunnel 2a locking camphor into the active site [Fig. 7(D)]. The importance of opening of this aromatic gate for the exit of the ligand via 2a has also been reported in CYP2C9 simulations.[31,32] Furthermore, it is noticeable that the phenol side-chain of Y96, which is pointing outwards towards the solvent in the crystal structure, was capable of adopting several quite different conformations [Fig. 7(C)]. In repeat-1 Y96 continued to point outward whilst in repeat-3, it moved into the binding cavity to interact with camphor and formed a lid over the active site. In repeat-2, Y96 moved into the cavity in a sideways orientation, making π–π interactions with F87 to form an aromatic gate, separating the active site from bulk solvent which is believed to be important for catalysis [Fig. 7(D)]. In all three repeats, we also observed the formation of a tunnel on the proximal side of the heme group (Supporting Information). This tunnel has been identified as a water tunnel in previous studies.[6]

Camphor in the access channel

In the description of the crystal structure, Yang *et al.*[33] reported the presence of weak electron density in two locations within the access channel which might be suggestive of a potential pathway for camphor. MD simulations with camphor in three different positions (AC1, AC2, and AC3) within the access channel (Fig. 2 and Table I) were performed (three repeats) in an attempt to identify the routes camphor might take to enter or exit the buried active site. In the position labeled AC1, camphor was started at a distance of 10.6 Å from the centre of mass of heme forming interactions with residues in the active site and the access channel. In repeat-1, camphor reached the protein surface via tunnel 2a

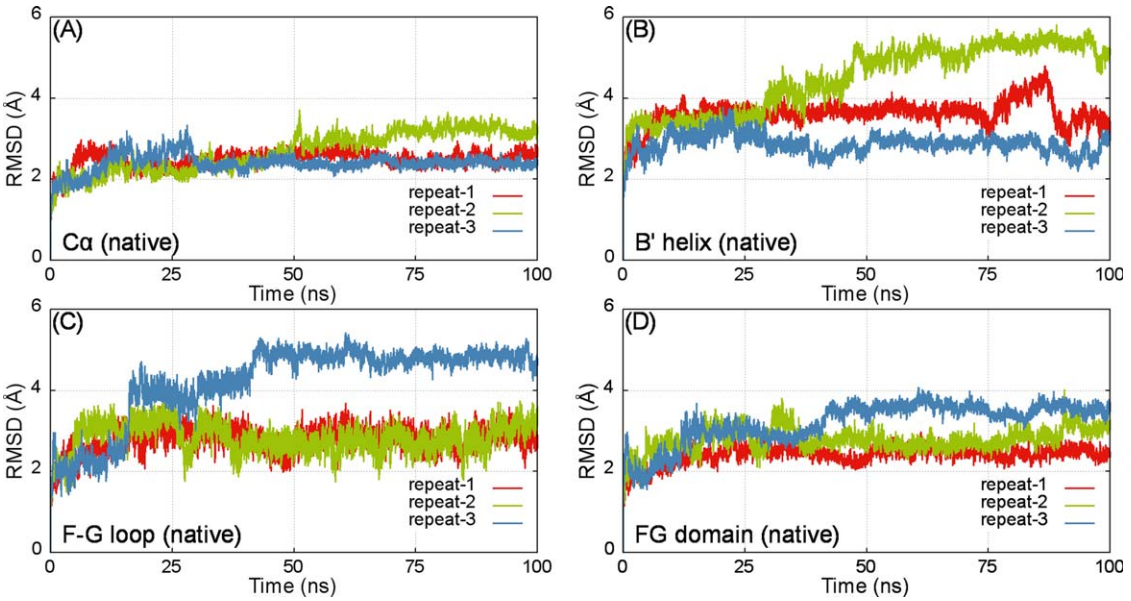

Figure 4. Root mean square deviations (RMSD) of (A) all Cα atoms, (B) B′ helix (residues 87 to 98), (C) F–G loop (residues 183 to 196), and (D) FG domain (residues 169 to 219) for the three native simulation repeats (red, green, and blue respectively). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

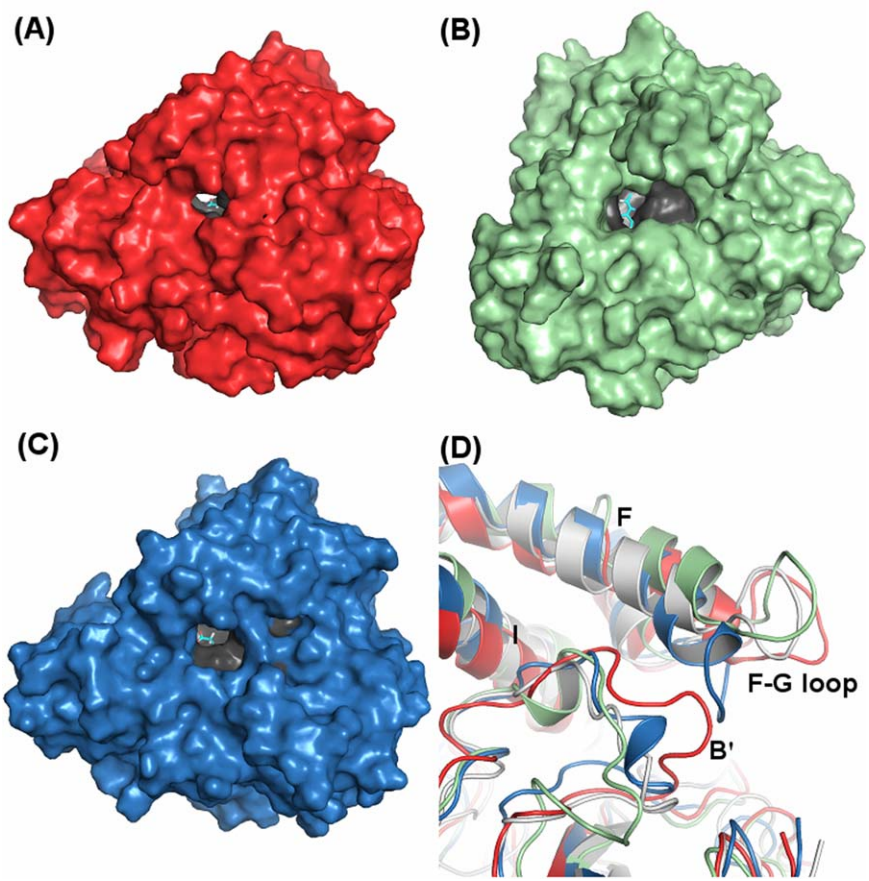

Figure 5. Surface representation of the final frame snapshot from repeat-1 (red), repeat-2 (green), and repeat-3 (blue) shown in A, B, and C respectively. Close-up cartoon view from the same snapshots of the F–G loop and B′ helix (D) with the reference structure (grey) for substrate free simulations. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

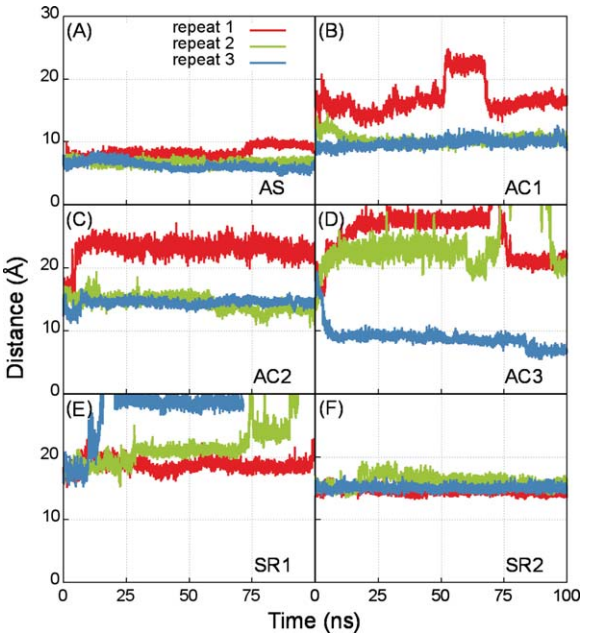

Figure 6. Distance of the camphor centre of mass from the heme centre of mass for (A) the AS simulations, (B–D) the AC1, AC2, and AC3 simulations, and (E,F) the SR1 and SR2 simulations, respectively. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

(Fig. 3) 60 ns into the simulation [Fig. 6(B)]. It remained on the surface for 20 ns before it started to drift back into the access channel [Fig. 6(B)].

In repeat-2 and 3, camphor remained in the access channel [Fig. 6(B)] and reoriented itself to sample different conformations. In repeat-2, camphor moved outward for about 8 ns but then moved back into the access channel close to the starting position. Y96 formed a π–π interaction with F87 and these two residues along with I401 and V402 formed a gate blocking the exit of camphor as previously observed in the simulations for camphor in the active site. In repeat-3, camphor moved towards the active site and the average RMSD for Cα atoms is high (between 3–4 Å) due to large movements in the B′ helix and the F–G loop (Supporting Information). Tunnel 2a remained open during the simulation (Supporting Information) and often opened wide enough for camphor to exit, however hydrophobic packing between numerous residues in this region (I86, F87, Y96, M98, T101, L250, L253, and I401) prevented its egress.

Similar to the simulations of camphor in the active site, Y96 adopted different orientations in all three repeats. In repeat-1, it remained on the surface interacting with residues in helix G. In repeat-2, it moved into the cavity in a sideways orientation making π–π interactions with F87 and in repeat-3, it moved into the cavity to interact with camphor [similar to movements in Fig. 7(C)].

In the second of the access channel positions, AC2, camphor was positioned at 12.6 Å from the heme centre. In repeat-1, camphor moved towards the surface via tunnel 2a within the first 5 ns of the simulation [Figs. 6(C), 8(B), and 9]. Side-chain rotations in F87, Y96, and backbone movement in the F–G loop provide a pathway for camphor to exit. Y29, P89, P192, M195, A196, L199, and I401 made interactions with camphor as it moved to the protein surface. In repeat-2, camphor moved away from the heme towards the surface for the first 10 ns but the hinge motion of the F–G loop appeared to drag it back into the access channel. It is clear that exit (via tunnel 2a) is once again impeded by the formation of the aromatic gate. In repeat-3, camphor moved towards the B′ helix within the first 7 ns but remained enclosed in a cavity formed by hydrophobic residues. Interactions between E94 and K102 also formed a barrier to prevent the exit of camphor via tunnel 2e. Exit via the 2a tunnel was hindered by the F87 and Y96, respectively.

In the final of the access channel positions, AC3, camphor was placed at 15.8 Å from the heme centre below the R186-D257 salt bridge that separates recognition site SR2 from the access channel (Fig. 2). In repeat-1 and 2, camphor moved to the surface into solvent but in repeat-3 it entered into the active site [Fig. 8(C)]. In repeat-1, camphor escaped to the surface through an exit route between the B′ helix and the G helix (via entrance of tunnel 2ac) at 7.2 ns into the simulation before jumping into solvent at 70 ns. It then moved towards recognition site SR2 and remained there for the last 25 ns [Fig. 10(A)]. In repeat-2, camphor drifted to the surface towards the entrance of tunnel 2f before it diffused into the solvent after 75 ns. A flipping movement of H398 directed the exit of camphor towards the surface and it formed new interactions with I396 and E406 and other residues at the entrance of the 2f tunnel, before moving into solvent and then moving to the SR1 site for the last 6 ns of the simulation [Fig. 10(B)]. In repeat-3, contrary to

Table I. *Summary of CYP101D2 Simulations*

Name	Description	Distance[a]
Native	Substrate free	N/A
AS	Camphor in the active site	6.2 Å
AC1	Camphor docked to the first access channel site.	10.6 Å
AC2	Camphor docked to the second access channel site.	12.6 Å
AC3	Camphor docked to the third access channel site.	15.8 Å
SR1	Camphor docked to the first substrate recognition site.	16.3 Å
SR2	Camphor docked to the second substrate recognition site.	15.6 Å

Three repeats of 100 ns duration were performed for all simulations (see Fig. 2 for diagram of positions).
[a] Distance between the centers of mass of the heme group and camphor at the start of the simulations.

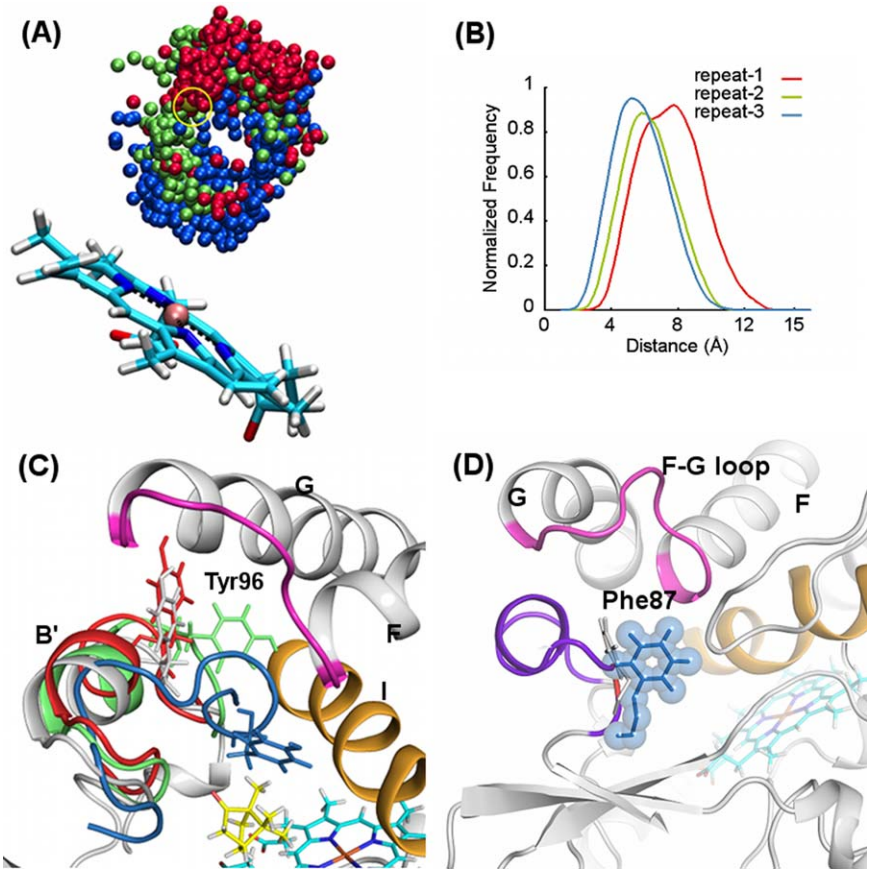

Figure 7. (A) The distribution (for the three repeats red, green and blue) of the C5 atom (which is oxidized during the reaction) with respect to the central iron of the heme group. The starting position is represented by the yellow circle. (B) The frequency distribution of the distance between C5 and the heme iron for the three repeats. While repeats 2 and 3 show a distribution more compatible with potential oxidation, camphor in repeat 1 moves away from the heme group. (C) The role of Y96 as illustrated by the final snapshot for the AS1 simulations from repeat-1 (red), repeat-2 (green) and repeat-3 (blue) compared to the crystal structure (grey). (D) Y96 (grey spheres and sticks) along with F87 (blue spheres and sticks) forms an aromatic gate that seals off the active site from the bulk solvent. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the other two repeats, camphor gradually moved into the active site close to the heme centre. The maximum of the radial distribution of the camphor C5 carbon is at 7.2 Å from the heme iron [Fig. 8(C)]. Y96 and F87 were observed to play a role in directing and locking camphor into active site. In the last 20 ns, camphor moved closer to the heme and sampled various orientations placing the C5 atom at a distance between 5 and 6 Å from the heme which would be the ideal distance for catalytic oxidation.

Camphor in the recognition sites

We docked camphor into two recognition sites close to what has been identified previously as substrate recognition site 1 (SR1) and substrate recognition site 2 (SR2)[34] (Fig. 2). Electron density was also observed near SR2 in a camphor-bound structure.[33] In the SR1 simulations, camphor was located about 16 Å from the heme centre at the entrance of tunnel 2c in the cavity formed by the G helix/B′ helix and the B–C loop. In all three repeats, camphor remained close to the surface and also made excursions into the bulk solvent. In the starting (crystal) structure, tunnel 2c connects to the access channel. However, this opening closed early in the simulations due to hydrophobic packing between residues in the G-helix and B′ helix obstructing the access of camphor to the access channel.

In the SR2 simulations, camphor was located 15.6 Å from the heme centre. In all three repeats, camphor remained in the recognition site interacting with many residues, some of which have previously been reported in the recent crystal structure.[33] Camphor reoriented around the starting position but was not able to move towards the active site because of a salt-bridge formed by D257 and R186 (Fig. 2) which prevented access. It is clear that this salt-bridge would have to break to allow camphor to approach the active site as has been discussed previously.[35,36]

Discussion

In agreement with previous studies of cytochrome P450 enzymes, simulations of the native protein demonstrated that the protein is capable of sampling

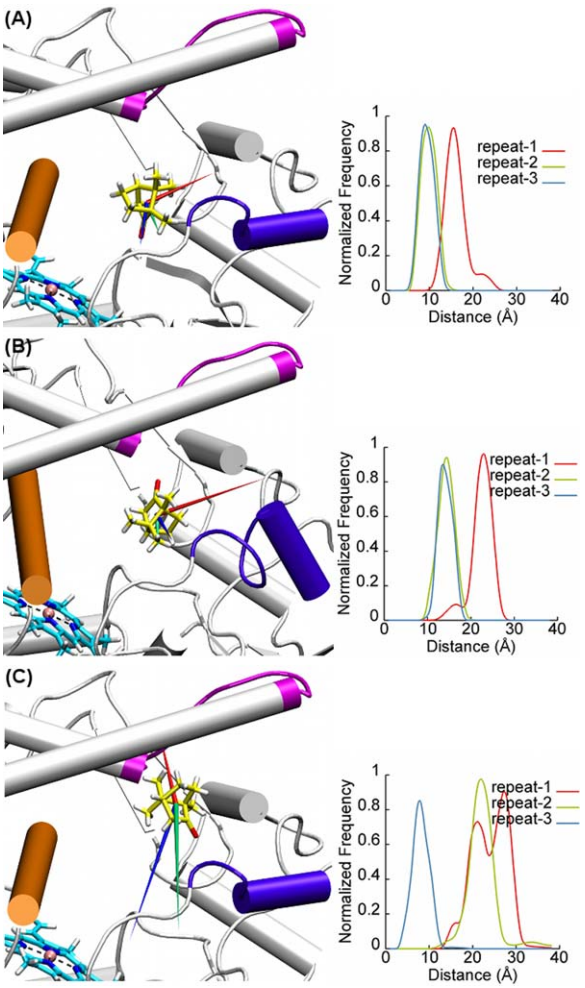

Figure 8. Substrate pathways and the corresponding heme-camphor centre-of-mass distance frequency distributions, for the AC1 (A), AC2 (B), and AC3 (C) simulations. Paths are depicted as a linear vector from the start to a final position. Repeats 1–3 are colored red, green, and blue respectively. Camphor is drawn as yellow sticks and heme as cyan sticks. The F–G loop, the B′ helix and helix I are colored magenta, blue, and gold, respectively. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

many different conformational sub-states. In particular, large movements in helices B′, F, G, and H and the F–G, G–H, and H–I loops lead to the formation of different tunnels, which merge reversibly during the simulations (Supporting Information). The greatest degree of motion occurs in the B′ helix and F–G loop (Fig. 4), suggesting this region has a prominent role in substrate recognition and entry. The potential movement of the F/G region has long-been recognized.[37] When comparing the native simulation with the simulations where camphor is included, the movement of the B′ helix in particular appears to be little-effected by the presence of camphor, which is perhaps what one might expect if the protein is to be as efficient as possible in allowing the various tunnels to form to allow access to the active site.

Tunnel 2f is formed in all three repeats for the native state. This is particularly interesting because in the recent crystal structure, electron density was observed at the entrance of the tunnel 2f suggesting that this tunnel in particular may be a key entry route for camphor. The simulations support previous interpretations of conformational selection made for the CYP119 protein from *Sulfolobus acidocaldarius*.[38] Taken together these simulations suggest that the fluctuations that allow substrates to enter and exit from the active site occur on tens of nanoseconds timescale and that entry/exit may be possible from a variety of different routes. This aspect may be a way that P450 ensures maximal efficiency in the balance between ensuring tight control of the active site environment whilst maintaining the fastest possible way to get molecules into and out of the protein.

The crystal structure of the camphor-bound CYP101D2 was found to be in an open conformation. We were interested to know whether this structure could adopt a closed-conformation or whether this state was perhaps unique to this particular P450. In all simulations with camphor in the active site, we observed the protein to adopt either a fully closed or partially closed state through the movement of helices F, G, and B′. This tends to suggest that the open-state of the crystal structure may be artificially stabilized by the crystal lattice. However, it is clear that there is still a lot of plasticity in the protein with the constant appearance and disappearance of tunnels throughout the structure (Supporting Information). Another feature of the crystal structure was that the camphor appeared to be orientated in such a way that was not compatible with oxidation at the C5 carbon. During the course of the simulations, camphor was able to reorient such that the C5 carbon was in an optimal position for oxidation by the central heme group.

In several of the simulations where we positioned camphor in the access channel, we observed direct egress to bulk solution (Figs. 9 and 10) via the 2a, 2ac, or 2f channels. In bacterial CYPs, 2a has been the predominant substrate access pathway[26] and based on our simulations, we suggest that the 2a route is the preferred pathway for the exit of camphor in CYP101D2. The role of the Y96/F87 hydrophobic gate, which we report here, has been recently discussed for P450cam.[39] Interestingly, it has also been discussed in the context of CYP2C9,[6,40,41] and it may play a critical role in the control of entry of substrates and exit of products. The dynamic motions in the protein allow the formation of different tunnels. It has been observed that these different tunnels may serve as the entry/exit pathways but it is not clear if different ligands use different tunnels guided by their properties or if the same ligand may employ different routes to enter

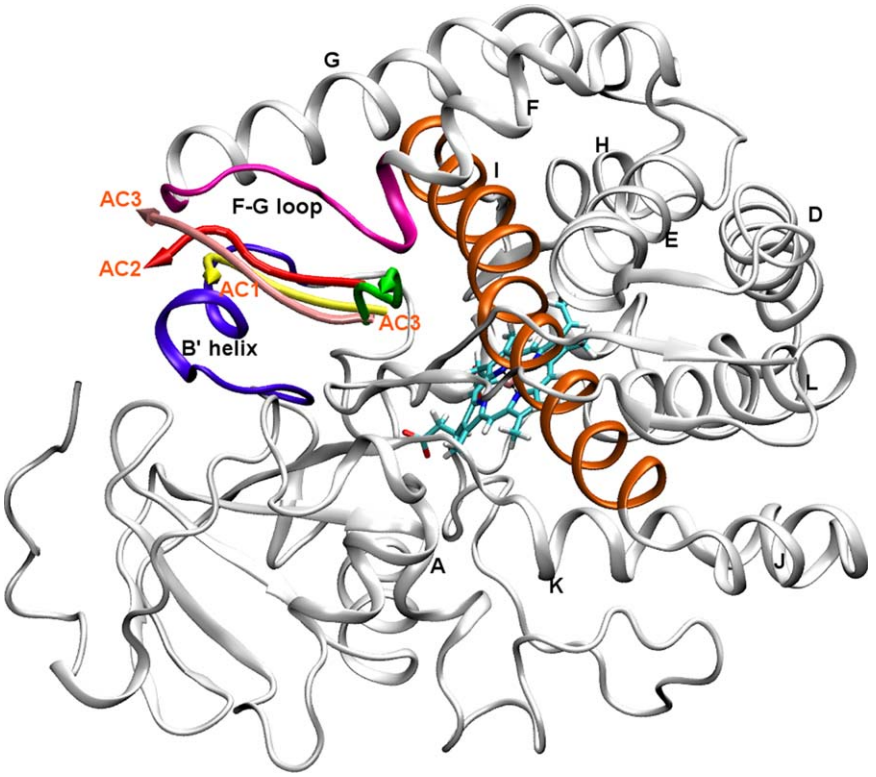

Figure 9. Exit pathways for camphor from the access channels as represented by colored arrows (AC1, repeat-1: yellow arrow, within the 2a tunnel, AC2, repeat-1: red within the 2a tunnel, AC3, repeat-1: pink within the 2ac tunnel, and AC3, repeat-2: green within the 2f tunnel). Secondary structure elements are colored as in previous figures and the heme group is drawn in cyan stick representation. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

and exit the active site. In our simulations, the 2f tunnel, which can merge reversibly with other tunnels, appears to be the prevalent tunnel, but as stated above we observe the most frequent exit from the 2a channel.

In all simulations, there is an opening at the location of the so-called "water channel." It is clear that the passage of a molecule the size of camphor cannot move through this tunnel due to the presence of the heme group. Furthermore, it was not seen as a route for substrate egress in random accelerated MD simulations.[7] It may have a role in the passage of small molecules like water and molecular oxygen. It was identified in P450cam filled with a cluster of water molecules near the heme propionate.[8,42] Since this channel opens towards the proximal side of the heme group it may have a role in the electron transport system as the electron transport protein binds to the cytochrome at the surface proximal to the heme group. Alternatively, it may be a route for molecular oxygen to access the active site required for the reaction to occur, or the tunnel may be involved in the transport of protons via a network of ordered water molecules.

Conclusions

Atomistic unbiased MD simulations have demonstrated that despite the camphor-bound crystal structure adopting an open state, the underlying dynamics of CYP101D2 appear to be similar to other CYP enzymes with the protein being able to readily form many different possible access routes and tunnels for substrate entry and product exit. Simulations with the camphor positioned at various points in the access channel and recognition sites show that movement towards the active site or towards bulk solvent is possible on a relatively short (tens of nanosecond) timescale. Comparison to native simulations supports the notion that the mechanism of movement through the protein is closer to a conformational selection rather than induced fit model.

Materials and Methods

Molecular dynamics simulations

We simulated the crystal structures of both conformations of CYP101D2, one without the substrate (PDB ID: 3NV5, 2.4 Å resolution) and the other with camphor bound in the active site (PDB ID: 3NV6, 2.20 Å resolution). Both structures were in an open state with a heme group in the active site and a large cavity connecting the exterior to the active site [Fig 1(A,B)]. The 3NV5 (native) structure had some residues missing in the B′ helix (residues 92–95) and F–G loop (residues 189–191). The numbering scheme is that for the mature protein. MODELLER

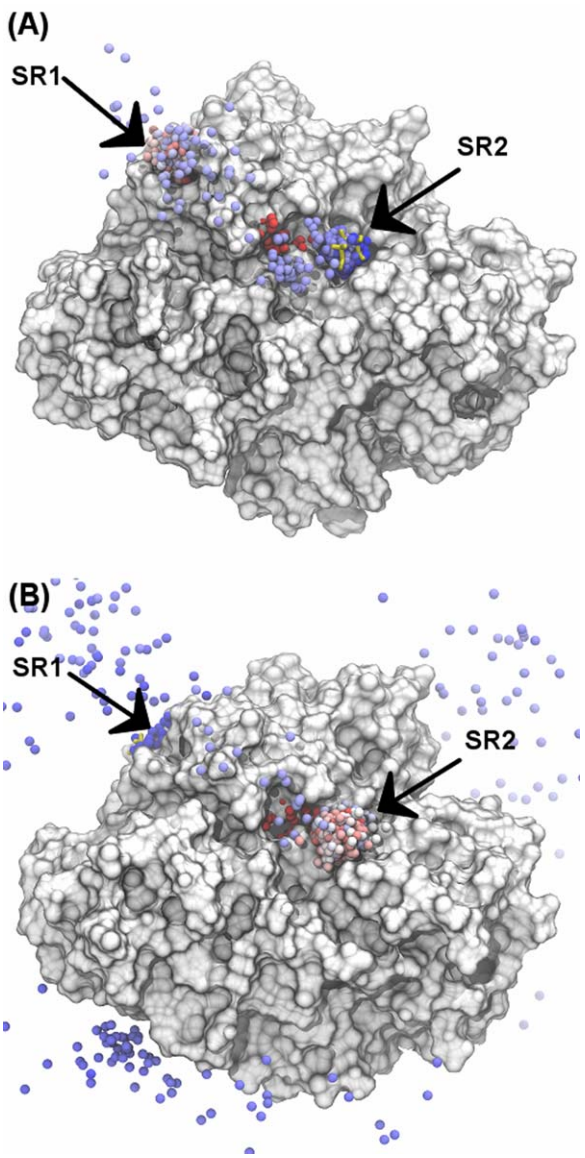

Figure 10. Movement of camphor when placed in the AC3 site in repeat-1 (A) where the camphor exits through tunnel 2ac, diffuses through the solvent and moves to the SR2, and movement of the camphor when placed in AC3 (repeat-2) where camphor exits into bulk solvent via tunnel 2f before moving to SR1 (B). The movement of camphor over time is shown by spheres (representing the centre of mass), which change from red to blue as a function of time through the simulation. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

version 8v2[43] was used to add missing residues using the 3NV6 structure as template. In order to understand the access of camphor to the binding site, we docked the camphor molecule at different positions in the access channel and in the recognition site closed to the residues flanking the positions of observed electron density (where no coordinates had been modeled in). Camphor was docked in three different positions in the access channel (AC1, AC2, and AC3) and two substrate recognition sites (SR1 and SR2) using AutoDock4.[44] Three 100 ns simulations were performed for each location along with simulations of the native structure and the structure with camphor in the active site. The details of the simulations and initial distance between the centre of mass of the heme group and camphor in all six simulations are summarized in Table I (Fig. 2)

All MD simulations were performed using the GROMACS[45] suite of tools (version 4.5.4) with the OPLS-AA force field[46,47] under periodic boundary conditions. The heme parameters (including the Fe–S parameters) were provided by Gogonea[48] with atomic charges obtained from the wave function of the optimized-geometry using the Merz–Kollman[49] method as implemented in the Gaussian98 program.[50] We used the topobuild package in GROMACS to generate the camphor topology and manually checked the parameters.

All crystallographic waters from the initial PDB coordinate files were removed and the system solvated in a cubic box with the explicit TIP4P water model.[51] The water box was 15 Å from the protein on all sides. Sufficient amounts of Na^+ and Cl^- ions were added to make up a neutral solution of 150 m*M* NaCl. The Particle Mesh Ewald method[52] was used to evaluate the electrostatic interactions with a real space cut-off of 10 Å. Temperature was coupled by Nose–Hoover heat bath[53] with a tau value of 0.5 ps and the pressure coupling was performed with the Parrinello–Rahman barostat[54] with a tau of 1.0 ps. The LINCS[55] algorithm was used to constrain the bonds involving hydrogen. Energy minimization was performed using the steepest descent method with positional restraints on protein heavy atoms and a force tolerance of 1000 kJ mol^{-1} nm^{-2}. The system was equilibrated for 200 ps with positional restraints on heavy atoms using a harmonic potential with a force constant of 1000 kJ mol^{-1} nm^{-2}. Different production runs were assigned different initial velocities at this point. Finally, a 100 ns unrestrained production simulations were performed in an NPT ensemble maintained at a temperature of 300 K and 1 bar pressure. The time step was 0.5 fs during equilibration and 2 fs during production. The set of simulations described above were repeated three times with different initial velocities. Snapshots from the trajectories were saved every 5 ps for analysis. Trajectories were visualized using VMD[56] and images were generated using both PYMOL[57] and VMD.

Analysis of tunnel formation

We used the standalone version of newly released CAVER 3.0[58] to explore the tunnels connecting the protein surface to the (buried) active site. We used a probe radius of 1.4 Å (equivalent to the radius of water) to identify the tunnels formed during the simulations using 500 snapshots obtained at 200 ps intervals along each simulation. Different thresholds

were used for individual simulations to cluster the tunnels. The tunnels were assigned to one of the 10 tunnels as described in Figure 3 and listed in Supporting Information and were as originally described by Wade and colleagues.[6] The bottleneck radius (the smallest radius along the tunnel) for each tunnel along the simulation was plotted to see if the tunnel was open for the passage of water or camphor.

Acknowledgments

We thank Jerome Ma and Phillip Stansfeld for helpful discussions. SV thanks the Oxford Centre for Integrative Systems Biology for support.

References

1. Anzenbacher P, Anzenbacherova E (2001) Cytochromes P450 and metabolism of xenobiotics. Cell Mol Life Sci 58:737–747.
2. Bernhardt R (2006) Cytochromes P450 as versatile biocatalysts. J Biotech 124:128–145.
3. Guengerich FP (2006) Cytochrome P450s and other enzymes in drug metabolism and toxicity. AAPS J 8: E101–E111.
4. Cryle M, Stok JE, De Voss JJ (2003) Reactions catalyzed by bacterial cytochromes P450. Aust J Chem 56: 749–762.
5. Anzenbacher P, Anzenbacherova E, Lange R, Skopalik J, Otyepka M (2008) Active sites of cytochromes P450: what are they like? Acta Chim Slov 55:63–66.
6. Cojocaru V, Winn PJ, Wade RC (2007) The ins and outs of cytochrome P450s. Biochim Biophys Acta 1770:390–401.
7. Ludemann SK, Lounnas V, Wade RC (2000) 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.
8. Poulos TL, Finzel BC, Howard AJ (1987) High-resolution crystal structure of cytochrome P450cam. J Mol Biol 195:687–700.
9. Scott EE, He YA, Wester MR, White MA, Chin CC, Halpert JR, Johnson EF, Stout CD (2003) An open conformation of mammalian cytochrome P4502B4 at 1.6-angstrom resolution. Proc Natl Acad Sci USA 100: 131396–113201.
10. Bell SG, Xu F, Forward I, Bartlam M, Rao Z, Wong L-L (2008) Crystal structure of CYP199A2, a para-substituted benzoic acid oxidizing cytochrome P450 from *rhodopseudomonas palustris*. J Mol Biol 383:561–574.
11. Lee YT, Wilson RF, Rupniewski I, Goodin DB (2010) P450CAM visits an open conformation in the absence of substrate. Biochemistry 49:3412–3419.
12. Pochapsky TC, Kazanis S, Dang M (2010) Conformational plasticity and structure/function relationships in cytochromes P450. Antioxid Redox Signal 13:1273–1296.
13. Bell SG, Yang W, Tan ABH, Zhou R, Johnson EOD, Zhang A, Zhou W, Rao Z, Wong L-L (2012) The crystal structures of 4-methoxybenzoate bound CYP199A2 and CYP199A4: structural changes on substrate binding and the identification of an anion binding site. Dalton Trans 41:8703–8714.
14. Wilderman PR, Halpert JR (2012) Plasticity of CYP2B enzymes: structural and solution biophysical methods. Curr Drug Metab 13:167–176.
15. Markwick PR PL, Goodin DB, McCammon JA (2011) Adaptive accelerated molecular dynamics (AD-AMD) revealing the molecular plasticity of P450cam. J Phys Chem Lett 2:158–164.
16. Lee YT, Glazer EC, Wilson RF, Stout CD, Goodin DB (2011) Three clusters of conformational states in P450cam reveal a multistep pathway for closing of the substrate access channel. Biochemistry 50:693–703.
17. Stoll S, Lee YT, Zhang M, Wilson RF, Britt RD, Goodin DB (2012) Double electron-electron resonance shows cytochrome P450cam undergoes a conformational change in solution upon binding substrate. Proc Natl Acad Sci USA 109:12888–12893.
18. Dunn AR, Dmochowski IJ, Bilwes AM, Gray HB, Crane BR (2001) Probing the open state of cytochrome P450cam with ruthenium-linker substrates. Proc Natl Acad Sci USA 98:12420–12425.
19. Lampe JN, Floor SN, Gross JD, Nishida CR, Jiang Y, Trnka MJ, Ortiz de Montellano PR (2008) Ligand-induced conformational heterogeneity of cytochrome P450 CYP119 identified by 2D NMR spectroscopy with the unnatural amino acid 13c-p-methoxyphenylalanine. J Am Chem Soc 130:16168–16169.
20. Zhao YH, White MA, Muralidhara BK, Sun L, Halpert JR, Stout CD (2006) Structure of microsomal cytochrome P4502B4 complexed with the antifungal drug bifonazole: insight into P450 conformational plasticity and membrane interaction. J Biol Chem 281:5973–5981.
21. Nishida CR, de Montellano PRO (2005) Thermophilic cytochrome P450 enzymes. Biochem Biophys Res Commun 338:437–445.
22. Ekroos M, Sjögren T (2006) Structural basis for ligand promiscuity in cytochrome P450 3A4. Proc Natl Acad Sci USA 103:13682–13687.
23. Li HY, Poulos TL (2004) Crystallization of cytochromes P450 and substrate-enzyme interactions. Curr Top Med Chem 4:1789–1802.
24. Park SY, Yamane K, Adachi S, Shiro Y, Weiss KE, Maves SA, Sligar SG (2002) Thermophilic cytochrome P450 (CYP119) from sulfolobus solfataricus: high resolution structure and functional properties. J Inorg Biochem 91:491–501.
25. Krishnamoorthy N, Gajendrarao P, Thangapandian S, Lee Y, Lee K (2010) Probing possible egress channels for multiple ligands in human CYP3A4: a molecular modeling study. J Mol Model 16:607–614.
26. Winn PJ, Lüdemann SK, Gauges R, Lounnas V, Wade RC (2002) 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.
27. Wade RC, Winn PJ, Schlichting E, Sudarko X (2004) A survey of active site access channels in cytochromes P450. J Inorg Biochem 98:1175–1182.
28. Yang W, Bell SG, Wang H, Zhou W, Hoskins N, Dale A, Bartlam M, Wong L-L, Rao Z (2010) Molecular characterization of a class i P450 electron transfer system from Novosphingobium aromaticivorans DSM12444. J Biol Chem 285:27372–27384.
29. Yao H, McCullough CR, Costache AD, Pullela PK, Sem DS (2007) Structural evidence for a functionally relevant second camphor binding site in P450cam: model for substrate entry into a P450 active site. Prot Struc Func Genet 69:125–138.

30. de Graaf C, Oostenbrink C, Keizers PH, van der Wijst T, Jongejan A, Vermeulen NP (2006) Catalytic site prediction and virtual screening of cytochrome P450 2D6 substrates by consideration of water and rescoring in automated docking. J Med Chem 49:2417–2430.
31. Cojocaru V, Winn PJ, Wade RC (2012) Multiple, ligand-dependent routes from the active site of cytochrome P450 2C9. Curr Drug Metab 13:143–154.
32. Cojocaru V, Balali-Mood K, Sansom MSP, Wade RC (2011) Structure and dynamics of the membrane-bound cytochrome P450 2C9. PLoS Comp Biol 7:e1002152.
33. Yang W, Bell SG, Wang H, Zhou W, Bartlam M, Wong LÄ, Rao Z (2011) The structure of CYP101D2 unveils a potential path for substrate entry into the active site. Biochem J 433:85–93.
34. Gotoh O (1992) 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.
35. Deprez E, Gerber NC, Di Primo C, Douzou P, Sligar SG, Hui Bon Hoa G (1994) Electrostatic control of the substrate access channel in cytochrome P-450cam. Biochemistry 33:14464–14468.
36. Lounnas V, Wade RC (1997) Exceptionally stable salt bridges in cytochrome P450cam have functional roles. Biochemistry 36:5402–5417.
37. Li H, Poulos TL (1995) Modeling protein-substrate interactions in the heme domain of cytochrome P450BM-3. Acta Cryst D51:21–32.
38. Lampe JN, Brandman R, Sivaramakrishnan S, de Montellano PRO (2010) Two-dimensional NMR and all-atom molecular dynamics of cytochrome P450 CYP119 reveal hidden conformational substates. J Biol Chem 285:9594–9603.
39. Tripathi S, Li H, Poulos TL (2013) Structural basis for effector control and redox partner recognition in cytochrome P450. Science 340:1227–1230.
40. Melet A, Assrir N, Jean P, Pilar Lopez-Garcia M, Marques-Soares C, Jaouen M, Dansette PM, Sari M-As, Mansuy D (2003) Substrate selectivity of human cytochrome P450 2C9: importance of residues 476, 365, and 114 in recognition of diclofenac and sulfaphenazole and in mechanism-based inactivation by tienilic acid. Arch Biochem Biophys 409:80–91.
41. Seifert A, Tatzel S, Schmid RD, Pleiss J (2006) Multiple molecular dynamics simulations of human P450 monooxygenase CYP2C9: the molecular basis of substrate binding and regioselectivity toward warfarin. Prot Struc Func Gen 64:147–155.
42. Oprea TI, Hummer G, Garcia AE (1997) Identification of a functional water channel in cytochrome P450 enzymes. Proc Natl Acad Sci USA 94:2133–2138.
43. Šali A, Sánchez R, Badretdinov A (1997) MODELLER: a program for protein structure modelling. Release 4. New York: The Rockefeller University.
44. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) Autodock 4 and autodocktools 4: automated docking with selective receptor flexibility. J Comp Chem 30:2785–2791.
45. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simuation. J Chem Theory Comput 4:435–447.
46. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118:11225–11236.
47. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the opls-aa force field for proteins via comparison with accurate quantum chemical calculations on peptides. J Phys Chem B 105:6474–6487.
48. Gogonea V, Shy JM, Biswas PK (2006) Electronic structure, ionization potential, and electron affinity of the enzyme cofactor (6r)−5,6,7,8-tetrahydrobiopterin in the gas phase, solution, and protein environments. J Phys Chem B 110:22861–22871.
49. Besler B, Merz KM, Kollman P (1990) Atomic charges derived from semiempirical methods. J Comp Chem 11: 431–439.
50. Gaussian Inc., Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JJA, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone H, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukada R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Maliek DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Revision B, Gaussian Inc., Wallingford CT.
51. Jorgensen WL, Chandresekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79: 926–935.
52. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an n.Log(n) method for Ewald sums in large systems. J Chem Phys 98:10089–10092.
53. Nose S (1984) A molecular dynamics method for simulations in the canonical ensemble. Mol Phys 52:255–268.
54. Parinello M, Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. J Appl Phys 52:7182–7190.
55. Hess B, Bekker J, Berendsen HJC, Fraaije JGEM (1997) Lincs: a linear constraint solver for molecular simulations. J Comp Chem 18:1463–1472.
56. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14:33–38.
57. DeLano WL (2004) The pymol molecular graphics system.San Carlos, CA, UDA: DeLano Scientific LLC.
58. Chovancova E, Pavelka A, Benes P, Strnad O, Brezovsky J, Kozlikova B, Gora A, Sustr V, Klvana M, Medek P, Biedermannova L, Sochor J, Damborsky J (2012) Caver 3.0: a tool for the analysis of transport pathways in dynamic protein structures. PLoS Comput Biol 8:e1002708.