HHS Public Access

Author manuscript

J Mol Biol. Author manuscript; available in PMC 2019 April 27.

Published in final edited form as:

J Mol Biol. 2018 April 27; 430(9): 1295–1310. doi:10.1016/j.jmb.2018.03.014.

Some surprising implications of NMR-directed simulations of substrate recognition and binding by cytochrome $P450_{cam}$ (CYP101A1)

Eliana K. Asciutto[a] and **Thomas C. Pochapsky**[b]

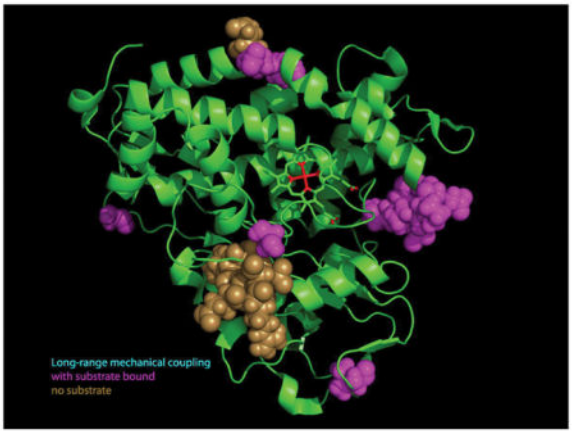
[a]School of Science and Technology, UNSAM and CONICET. Campus Migueletes, 25 de Mayo y Francia, Buenos Aires, Argentina

[b]Department of Chemistry and Rosenstiel Basic Biomedical Sciences Research Institute, MS 015, Brandeis University, 415 South St., Waltham, Massachusetts, USA

Abstract

Cytochrome $P450_{cam}$ (CYP101A1) catalyzes the stereospecific 5-*exo* hydroxylation of *d*-camphor by molecular oxygen. Previously, residual dipolar couplings (RDCs) measured for backbone amide ^{1}H-^{15}N correlations in both substrate-free and bound forms of CYP101A1 were used as restraints in soft annealing molecular dynamics (MD) simulations in order to identify average conformations of the enzyme with and without substrate bound. Multiple substrate-dependent conformational changes remote from the enzyme active site were identified, and site-directed mutagenesis and activity assays confirmed the importance of these changes in substrate recognition. The current work makes use of perturbation response scanning (PRS) and umbrella sampling MD of the RDC-derived CYP101A1 structures to probe the roles of remote structural features in enforcing the regio- and stereospecific nature of the hydroxylation reaction catalyzed by CYP101A1. An improper dihedral angle Ψ was defined and used to maintain substrate orientation in the CYP101A1 active site, and it was observed that different values of Ψ result in different PRS response maps. Umbrella sampling methods show that the free energy of the system is sensitive to Ψ, and bound substrate forms an important mechanical link in the transmission of mechanical coupling through the enzyme structure. Finally, a qualitative approach to interpreting PRS maps in terms of the roles of secondary structural features is proposed.

Graphical Abstract



Introduction

The activation of molecular oxygen, O_2, for reaction with singlet bonds is an important step in a multitude of biochemical pathways. Given the toxicity of active oxygen species (e.g., peroxide, superoxide and hydroxyl radicals), it is critical to restrict their in vivo production spatially and temporally. Nature often entrusts this job to cytochromes P450, a superfamily of heme-containing monooxygenases that have been found in every kingdom and genera of life. Over 450,000 genes have been assigned to the P450 superfamily in GenBank as of August 2017, and identified P450 substrates range from small terpenes and simple hydrocarbons to large complex molecules such as steroid hormones and macrolide antibiotics. Targets for P450-catalyzed oxidations are typically unactivated C-H and C=C bonds, and as such, most substrates present a multitude of possible reactive sites. Nevertheless, P450s involved in biosynthetic or catabolic pathways are usually highly regio- and stereoselective in their functions, implying that the enzymes bind their substrates in a single preferred orientation, presenting the appropriate site to the activated iron-oxo complex. Cytochrome $P450_{cam}$ (CYP101A1), originally isolated from the soil bacterium *Pseudomonas putida*, provides a good example of such selectivity. CYP101A1 catalyzes the first step in camphor catabolism by *P. putida*, the 5-*exo*-hydroxylation of *d*-camphor. The overall reaction is shown in Scheme 1.

Crystallographic structures of CYP101A1 provide a clear rationale for the observed regio- and stereoselectivity of the oxidation, placing C5 appropriately oriented for reaction with the reactive iron-oxo species (Figure 1). However, using nuclear magnetic resonance methods (NMR) we found that in solution, the camphor orientation differs from that seen in the crystal structure, and the binding of the iron-sulfur protein putidaredoxin (Pdx), the physiological redox partner and effector of CYP101A1, was required in order to attain the camphor orientation seen crystallographically [1, 2]. This observation suggested to us that the conformational ensembles available to CYP101A1 in solution might not be adequately represented by crystallographic structures, regardless of their resolution. On the other hand, NMR methods are ideally suited to describing such ensembles, assuming that sequence-specific resonance assignments are available. With such assignments in hand, NMR can be used to obtain *de novo* solution structures of small to medium sized proteins. We have

reported previously on our comprehensive assignments for backbone resonances of CYP101A1 in both oxidation states [2, 3].

However, the size of CYP101A1 (~47 kDa) and the paramagnetism of the oxidized form hinders the use of standard NMR structural techniques. Instead, we used residual dipolar couplings (RDCs) measured for assigned ^{1}H-^{15}N backbone amide correlations as restraints in molecular dynamics simulations of the folded enzyme, an approach that we call “soft annealing”. Unlike other NMR-derived restraints (nuclear Overhauser effects, chemical shifts, scalar couplings) that are local in nature, RDCs can be related to the orientation of the vector between the coupled spins in the molecular frame of reference for a single domain protein [4]. In particular, RDCs measured for backbone amide N-H pairs help orient the secondary structural features in which they reside with respect to the molecular frame. We have previously described the use of soft annealing methods to obtain solution structural ensembles of CYP101A1 with and without substrate *d*-camphor bound [5, 6], and demonstrated via a combined mutagenesis/activity study that the conformational changes in CYP101A1 that we observe upon substrate binding are not artifacts of the methodology, but reliable representations of substrate-dependent behavior in this enzyme [7]. We have recently extended the soft annealing methodology to another P450, MycG, in order to clarify differences between crystallographic and solution conformations of that multifunctional enzyme [8, 9]. Comparison of substrate-dependent conformational changes in both enzymes highlight significant similarities in their substrate binding and recognition behavior, suggesting that there is a common modality for substrate binding and orientation in the P450 superfamily, regardless of the nature of the substrate [10]. Furthermore, it was noted that those regions showing the greatest sequence conservation are also the least perturbed upon substrate binding in both enzymes. These include the J, K and L helices, the C-terminal loop and the β-meander containing the cysteine thiolate that ligates the heme Fe in all P450s (see Fig. 2). Alignment of these regions between CYP101A1 and MycG yields a 63% identity, 78% similarity and no gaps. This suggests that these regions form the “core” P450 structural elements required for the safe activation of O_2 [10].

One particularly interesting observation that arose from these efforts is that secondary structural features in both MycG and CYP101A1 maintain their integrity in the absence of substrate; that is, we do not see significant disordering or fraying of secondary structures in the substrate-free enzymes. Rather, secondary structural features (helices and sheets) move as units, reorienting or repositioning with respect to each other as substrate binds. The orientation of substrate in the active site is thus controlled not only by first-sphere interactions with side chains in the active site, but by more distant interactions between secondary structural features, which enforce a particular position of active site residues and substrate via mechanical coupling. This rationalizes the experimental observations by our group [7, 11] and others [12, 13] that mutations remote from the active site of cytochromes P450 can play a critical role is substrate selection and product specificity. Furthermore, it implies that mechanical coupling pathways exist through which substrate-dependent displacements are transmitted.

Results

In the current work, we characterize the RDC-based solution structure of *d*-camphor-bound CYP101A1 (Figure 2, PDB entry 2L8M, ref. [5]) using the anisotropic network model (ANM) [16]. Within this model, we calculate spring force constants and perturbation-response correlations extracted from molecular dynamics tracks [14]. With heat maps constructed from both, spring constants and response correlations, we first identified those sites in the enzyme that are most sensitive to the position and orientation of bound substrate *d*-camphor as it occurs in the original 2L8M structure, and compared these to the same maps obtained in the absence of substrate. These correlations are determined by the degree to which a motion at residue *i* correlates with a displacement at residue *j*. Initial simulations were performed with the substrate free to re-orient, followed by a series of simulations in which the substrate is restricted via harmonic restraints to a to a fixed orientation relative to the heme plane. We also quantified the changes in total free energy of the complex due to different orientations of bound substrate using umbrella sampling.

Free energy response to changes in substrate orientation in 2L8M

To characterize the influence of substrate orientation in the active site on the overall structure of CYP101A1, umbrella sampling dynamics was performed, rotating the substrate around the improper dihedral angle Ψ defined between the heme Fe and NB pyrrole nitrogen and the camphor (substrate) carbon atoms C7 and C9, as shown in Figure 3. Rotation around Ψ results in changing the orientation of the substrate relative to the heme Fe atom, but does not significantly affect the interatomic distance between C7 and the heme Fe. (The restraint on Ψ does not include an explicit energetic restraint on the C7-Fe distance, which is likely maintained by steric interactions between camphor and side chains in the active site). The NMR-derived structure of CYP101A1 with camphor bound (PDB entry 2L8M) was used as starting structure for the simulations. Three umbrella sampling runs were performed, starting in each case from a different value of Ψ, while enforcing a full 360° rotation through the course of the dynamics. In each case, the global free energy minimum corresponded to a Ψ angle of −140°. Local minima were observed at Ψ= −50° and Ψ= 20°. Figure 4 shows the free energy profile of the system as a function of Ψ.

Effective force constant calculations

The Anisotropic Network Model (ANM) was used to obtain information on the relative mobility of the residues in CYP101A1 [15]. In ANM, the protein is represented as a network of interacting nodes. The C_α atoms of each residue provide the nodes, and interaction takes the form of a harmonic potential. (Heme and substrate atoms were not included). A distance cutoff is defined, so that only pair of nodes at a distance below the cutoff distance will interact. At equilibrium, the system has *3N-6* internal degrees of freedom, and hence it has *3N-6* normal modes of motion. The Hessian matrix in terms of the normal modes is:

$$H = \sum_{k=1}^{3N-6} \lambda_k v^k v^{kT} \quad (1)$$

with λ_k its eigenvalues and v^k its eigenvectors. The superscript T indicates transpose.

The eigenvalue λ_k represents the curvature of the potential along the normal mode *k*, in normal mode space, and hence can be associated with a spring constant in physical space. The inverse of the Hessian matrix is the covariance matrix, calculated using the spatial fluctuations. If an external perturbation matches the direction of a slow mode, there will be little resistance to the perturbation, corresponding to a small value of the spring constant. Therefore, it makes sense to define an effective spring constant between pairs of interacting residues in terms of projections onto normal modes. That is:

$$K_{eff} = \sum_{k}^{3N-6} \left| \overrightarrow{v}_k \overrightarrow{U} \right| (\lambda_k m) \quad (2)$$

where **U** is a column vector containing distances between residue i and residue j, with three components for each mode:

$$U^k = \begin{bmatrix} \Delta x_{ij}^k \\ \Delta y_{ij}^k \\ \Delta z_{ij}^k \end{bmatrix} \quad (3)$$

The K_{eff} matrix provides a mechanical response map that can be extracted from a molecular dynamics trajectory. Figure 5 shows the mechanical response map calculated from the 2L8M coordinates. Not surprisingly, helices and sheets are more resistant to deformation than unstructured elements. However, there are strong interactions between secondary structural features: In general, beta sheets show strong harmonic interaction with remote helices, e.g., β4 with the D and E helices, and β1 and β5 with helices D, E, G, and H. The K helix interacts strongly with almost all the other helices, and also with the β2 sheet. As might be expected, force constants in loop regions are weak (e.g., the C–D loop), suggesting low resistance to deformation. The largest force constants are observed between the lower region constituted by β1 sheet and the short helices B and B′. There also exists strong resistance to deformations between β4 with same helices. This result suggests tight mechanical coupling between the β-rich region (bottom portion of the structure in Figure 1) and helices that respond to substrate binding, as we have observed experimentally [10].

Perturbation response in 2L8M (substrate-bound) and 2LQD (substrate-free) CYP101A1

Once the mechanical response map has been established and the regions that are more easily deformable identified, the next step was to use perturbation response scanning (PRS) to examine how an external disturbance affects different conformations in the ensemble, that is, if the backbone C_α at residue ***a*** is perturbed, what is the effect at other residues ***b, c, d***? In order to establish a baseline for the response of the CYP101A1 structure to changes in camphor orientation, we first performed unrestrained MD as described in the experimental section on the solvated CYP101A1 structures with (2L8M) and without camphor (2LQD) bound. PRS matrices are calculated using the following protocol: An external perturbation resulting from a force **F** with random direction and unit magnitude is exerted on each residue *i*. The resulting displacement can then be calculated using the relation between the

covariance matrix as obtained from the trajectory and the force (**F** = **CΔR**). The response of each residue *k* to the given force constitute the elements of the PRS matrix:

$$\left\|\Delta R_k^i\right\|^2 = \left(\Delta \boldsymbol{r}_{\boldsymbol{kx}}^{\boldsymbol{i}}\right)^{\mathbf{2}} + \left(\Delta \boldsymbol{r}_{\boldsymbol{ky}}^{\boldsymbol{i}}\right)^{\mathbf{2}} + \left(\Delta \boldsymbol{r}_{\boldsymbol{kz}}^{\boldsymbol{i}}\right)^{\mathbf{2}} \quad (4)$$

This procedure is repeated *m* times to eliminate bias arising from the force being applied in different directions. The PRS matrix is composed of the average over *m*, $\left\|\Delta R_k^i\right\|^2$.

Next, the PRS matrix is normalized by dividing each element of the matrix by the diagonal element in the same column. After this normalization, row *i* of the matrix represents how a perturbation on residue *i* affects residue *j*, yielding an **influence profile** of residue *i* on the other residues, while column *j* describes the response of residue *j* to perturbations on the other residues.

Figure 6 shows the comparison between PRS maps obtained with and without bound substrate in CYP101A1. It is clear from the comparison in Figure 6 that, in the absence of substrate (structure 2LQD), long-range coupling of motions are considerably weakened, and different regions of the protein interact less than when substrate is bound. However, it should be emphasized that 2LQD does not show loss of secondary structure (fraying or disordering) in the absence of substrate, supporting our underlying assumption that secondary structures move as units upon substrate binding, with the loop regions between secondary structures exhibiting the most response. It is also worth noting that the effects are not symmetrical, that is, a perturbation at site *i* does not necessarily produce the same intensity effect at site *j* as is observed in reverse. Not surprisingly, the most sensitive residues are found at junctions between secondary structural features: Active site perturbations are more likely to give rise to more obvious effects at residues that have more freedom of movement those forming regular secondary structures. In other words, residues that are part of a regular secondary structural feature exert more influence than they experience, as expected based on the nature of the effective force constants K_{eff} described above.

Perturbation response in 2L8M with restraints on substrate orientation

We next characterized the response of the 2L8M structure to the imposition of a restraint on the orientation of substrate in the active site. The improper dihedral angle Ψ defined in Figure 3 was used as a harmonic restraint to define the relative positions of the heme iron, heme pyrrole nitrogen NB, and carbons C7 and C9 of camphor. In 2L8M, after equilibration, the starting value of Ψ = 88°. If Ψ is restricted harmonically to 88°, the heat map shown in Fig. 7 (right) is obtained, with the unrestrained heat map from Fig. 6 reproduced at the left for comparison. Clearly, the restraint affects correlated perturbations in discrete regions of the enzyme remote from the active site. The responses of Ser 346, which are markedly stronger in the absence of a restraint on Ψ (Fig. 7, left), and include responses to residues in the B–B′ and B′–C loops, are worth noting. Ser 346 is part of a hydrogen bonded network near the C-terminal end of the K′ helix, the importance of which in substrate binding and recognition in P450s has been described previously (see Figure 8) [10, 11]. Closer examination shows the backbone carbonyl and NH of Ser 346 forms a tight hydrogen

bonding pair with the side chain carbonyl and one amide $N_{\delta}H$ of Asn 332, the other $N_{\delta}H$ of which is within hydrogen bonding distance of Ser 325 carbonyl in the K′ helix (Figure 8). The adjacent residue, Arg 331, was found by the Arnold group to be important for changing substrate selectivity in CYP101A1 [16], while the Ser 325 amide NH shows a very large substrate-dependent chemical shift [11]. Furthermore, the region of the protein containing Ser 346 is considerably more flexible in the absence of substrate (Figure 6). This indicates that there is indeed mechanical coupling linking many regions of the enzyme to the active site that are important for maintaining a particular substrate orientation, while the substrate position is largely defined by the immediate active site contacts.

We next changed the restraint on ψ from the initial value of 88° to representative values over the full range of possible angles. Figure 9 shows several maps illustrating the interdependence of camphor orientation and the local perturbation responses of the enzyme. As can be seen, changes in the value of ψ result in almost completely different heat maps, with some interesting features. The B–B′ region shows sensitivity regardless the substrate orientation (See Figure 7 and Figure 9). The C–D loop shows almost no sensitivity for Ψ = 26°, but sensitivity is recovered for Ψ = −60° and is higher for Ψ = −100°. In the case of Ψ = −60°, high sensitivity is observed in the J helix region, while for Ψ = −100°, sensitivity in that region is almost lost and B, B′ region along with the C-terminal reactivate. These results indicate that different orientations of the substrate in the same location in the active site can activate (or deactivate) particular normal modes [17]. In turn, this suggests that the protein structure is evolved to exhibit normal modes that enforce the appropriate substrate orientation after binding.

Umbrella sampling and energy landscapes as a function of Ψ

Figure 4 shows the overall free energy of the structure and solvent as a function of Ψ in the course of an umbrella sampling protocol. Those changes are considerable: ΔG changes by ~4.5 kcal/mol in the course of the umbrella simulations, confirming that the CYP101A1 structure is highly sensitive to the orientation of camphor in the active site. Figure 10 compares the sensitivity profiles obtained with camphor free to reorient and restrained to Ψ = −140° (the value yielding the global minimum free energy by umbrella sampling, see Fig. 4). As Figs. 7 and 10 show, the B′– C loop is highly responsive if substrate is free to reorient, but considerably less responsive if the substrate is fixed (or absent). Given that the B′ helix is adjacent to the substrate binding pocket, this is not particularly surprising. On the other hand, Pro 215 at the C-terminal end of the G helix (Figure 10, right) becomes much more responsive with a restraint on Ψ, even though the G helix does not interact directly with bound substrate. It is possible that this sensitivity is due to indirect interactions involving Tyr 96 (B′ helix) and Phe 193 (G helix), although further work will be required to confirm this.

PRS responses with RDC restraints applied

As the starting structures 2L8M and 2LQD are both representatives of ensembles calculated using RDC restraints, it was important to characterize the impact of those restraints on the PRS maps, as RDCs result in a change in the net forces acting on the structure in the course of the simulation. Ideally, there should be little or no difference between the maps obtained

with and without RDC restraints active, as the energetic penalties on such restraints are relatively small and the RDCs sufficiently distributed over the structure so that no local structural distortions occur. Figure 11 compares the PRS maps obtained from a simulation of 2L8M with RDC restraints active with that obtained without RDCs (same as Figs. 6 and 7). The primary differences between the two are sensitivity, with the RDC-restrained simulations showing generally weaker responses. The exception lies in the B′–C loop, which is only marginally less responsive than in the unrestrained structure. Also, while the I–J loop is essentially unresponsive in unrestrained simulations, some weak response is noted in the presence of RDCs, while the β4 sheet loses what little response it shows in the absence of RDCs. Otherwise, the overall distribution of perturbation responses remains the same as in the unrestrained case. These results support the validity of the other PRS maps described here obtained in the absence of RDC restraints.

Discussion

The importance of remote (non-first-sphere) structural features in enzyme function is now an accepted concept [18–20]. The origins of these effects are often not obvious from crystallographic structures, especially if they manifest themselves dynamically, which is better probed by NMR and MD simulations [21–26]. Our observation that in at least two cases (MycG and CYP101A1), secondary structural features move as units upon substrate binding suggest that remote effects on enzyme function are transmitted via such movements: Much as a piston or cam transmits energy and information in a machine, so also do the coordinated motions of secondary structural features in an enzyme. While easy to imagine, the details of the coordination are more difficult to establish experimentally or computationally. Nevertheless, if rational redesign of enzymes is to become a reality, it will be necessary to envision (or predict) the interactions of disparate parts of the enzyme that result in observed function.

The goals of the current work are somewhat modest: We wished to probe the roles of various structural features in maintaining substrate orientation in a well-characterized enzyme, CYP101A1. We propose that the observed orientation of substrate *d*-camphor in the active site derived from RDC-directed simulations is the result of mechanical coupling between substrate contacts in the active site (which include side chains protruding from the B–B′ loop, B′ and I helices, β3 and β5 sheets, as well as the porphyrin of the heme itself) and remote features that maintain the “correct” substrate orientation by controlling the positions of the secondary structures from which local contacts project. If true, restraint of rotation around an improper angle that adequately represents the orientation of the substrate in the active site would aid in identification of those features involved in maintaining substrate orientation via PRS. Interestingly, the choice of improper angle Ψ was non-trivial. Because the *gem*-methyl groups (C8 and C9) are sterically prominent features of the otherwise rather spherical camphor molecule, it was found to be critical that one of those methyls be used in defining Ψ. Choices of other carbon atoms in the camphor skeleton resulted in little change in PRS maps as a function of Ψ.

Comparison of PRS in presence and absence of substrate

The heat maps comparing PRS in the presence and absence of bound substrate (Fig. 6) indicate that, while secondary structural features remain intact, mechanical coupling between disparate parts of the CYP101A1 structure is much weaker in the absence of substrate. Most notable is the almost complete absence of PRS in regions adjacent to the active site (B′–C), and a more diffuse PRS near the K′ helix and β-meander. It is clear that bound substrate provides a critical mechanical link between remote regions of the enzyme, which is to be expected if, in turn, the long-range mechanical couplings are responsible for maintaining substrate orientation in the active site.

Interpretation of PRS maps as a function of harmonic restraints on Ψ

The ultimate goal of this effort is to establish the roles played by multiple structural features of CYP101A1 in maintaining a particular orientation of substrate in the active site. While complete delineation of mechanical coupling pathways will require atomistic force distribution analysis [27], a qualitative interpretation of the results is worth examining here.

B–B′ loop

This region (Ser 82-Ile 88) contains a substrate contact (Phe 87) and has been implicated in the closure of the active site upon binding of redox partner putidaredoxin by *trans-cis* isomerization of the Ile 88-Pro 89 bond [2]. It shows modest sensitivity in the unrestrained simulations of 2L8M, and a slight specific sensitivity at Ser 83 with Ψ restrained at −60°. However, the lack of any large effects may be due to the remoteness of Phe 87 from the germinal methyl groups of camphor (C8 and C9). As such, regardless of the value of Ψ, contacts with Phe 87 are relatively unchanged.

B′–C loop

This loop is highly conserved in many P450s, with the sequence SMDPPEQR (Ser 102-Arg 109 at the beginning of the C helix) being essentially unchanged across of large number of aligned P450 sequences. Mutations in this region result in misfolding and heme loss in CYP101A1 [11]. Given the presence of two proline residues, it is expected that the loop would be displaced as a unit due to perturbations in the B′ helix, which provides a substrate contact with Tyr 96. However, likely more important here are interactions between the B′–C loop and the I helix, which provides several substrate contacts (Leu 244 and Val 247) with the camphor methyl carbon C8 in 2L8M. The I helix has previously been shown by NMR to be highly sensitive to substrate size and sterics [11], and thus appears to provide the mechanical linkage between substrate and the B′–C loop.

The B′–C loop is sensitive in unrestrained and RDC-restrained simulations of 2L8M, but becomes markedly less so with other camphor orientations enforced by restraints on Ψ. With Ψ = 26°, Val 247 still remains a contact for C8, but Leu 244 does not. Rather, Met 184, Thr 185 (F helix) and Phe 193 (F–G loop) become contacts for C8, leading to an increased response in the G–H (Pro 215) and H–I loops(Asn 225).

C–D loop

We have reported previously that residues in the C–D loop (Val 119–Val 123) are remarkably sensitive to the presence/size of substrate in CYP101A1, and that this region shows the largest displacement of any region of the CYP101A1 molecule upon substrate binding [7]. Site directed mutations and activity assays link this sensitivity to interactions between the I helix and E helix, which mechanically transmits active site perturbations to the C–D loop. PRS sensitivity is greatest in the unrestrained simulations, while RDCs and restraints on Ψ lead to considerable lessening of effects in the C–D loop. Again, this is likely to be due to reduced steric interactions with I helix residues in the case of restraints on Ψ.

F–G loop

This loop provides the canonical “cap” for the P450 active site, and for those enzymes with large substrates, often provides direct substrate contacts [28]. In no orientation that we examined in detail do residues of the F–G loop contact substrate, although F helix residues M184 and T185 contact C8 in R2 (Ψ = 26°). There is a distinct single residue sensitivity at Gly 189 in the F–G loop in R3 (Ψ = −60°) that shifts to Pro 187 in R4 (Ψ = −100°). Given that neither orientation shows direct contacts between the substrate methyl groups and residues in the F or G helices, the mechanical coupling pathway is not clear, but may involve contacts between the G helix and Phe 98 in the B′–C loop.

G–H loop

Pro 215 in the G–H loop shows singular sensitivity in unrestrained simulations as well as when Ψ is restrained to −140° (Fig. 10). This sensitivity is mirrored to some extent by Gly 226, which lies at the C-terminal end of the H helix. Again, this is likely due to steric interactions between the camphor *gem*-dimethyl groups (C8 and C9) and I helix, and transmitted via G–I helix contacts. The sensitivity is much reduced in other camphor orientations, where steric interactions between the I helix and substrate are less.

J helix

An interesting anomaly occurs in R3 (Ψ = −60°), with a general increase in the responsiveness of the J helix (Pro 268-Arg 277). Unlike most of the other sensitive regions, which are on the distal (active site) side of the heme plane, the J helix lies on the proximal side of the heme and, based on conservation of this region, might be considered a “core” feature of the P450 fold. That is, the J, K and L helices and the β-meander containing the axial heme ligand cysteine thiolate are highly conserved and exhibit few gaps or insertions from one P450 to another. Furthermore, the “core” region shows little or no perturbation upon substrate binding in either MycG or CYP101A1 by NMR, and RDC-refined structures of both enzymes show no substrate-dependent displacements of the core. Examination of the structure extracted from R3 shows that the gem dimethyl groups of camphor are oriented away from the heme, with direct contacts only to Asp 297 (β3) and Phe 98 (B′–C). We speculate that the interaction with Phe 98 may force the camphor molecule into closer contact with the heme porphyrin, and that it is this mechanical linkage (through the axial

cysteine ligand to the heme iron in the β-meander to the J and K helices) that increases the sensitivity of the J helix.

K′ - β-meander

In all of the calculations in which substrate is present, Ser 346 in the K′-meander loop shows a localized sensitivity (see Fig. 8). We have previously noted that the K′ helix is generally conserved in related P450 structures, and plays a role in positioning substrate contacts in the β3 sheet by shifting between α- and 3,10-helix hydrogen bonding patterns [10, 11]. Given that there is little variation of the sensitivity of Ser 346 with changing substrate orientations, it would seem that this response is not likely involved in driving a preferred orientation in the active site. Rather, it adjusts the overall volume of the active site in response to the volume of the bound substrate.

β4 and β5 strands

The β4 antiparallel sheet is actually an extension of the β3 strands adjacent to the active site, and responds weakly as a unit in the unrestrained calculations on 2L8M. Interactions between Asp 297 (β3) and the C8 methyl of camphor are likely responsible for the more specific sensitivity observed at Gly 309 in the loop connecting the β3 and β4 strands in R4 ($\Psi = -100°$).

In a similar manner, Gly 386 is found at the connection between the β5 sheet and the L helix, and appears to be responsive to interactions between substrate and Ile 395, which lies on a turn between the two strands of the β5 sheet in the active site. In this case, responsiveness is likely conducted via a net displacement of the β5 sheet along the long axis of the sheet, while the response at Gly 309 may be due to lateral movements against the β3 sheet. Gly 386 is not responsive in R2 ($\Psi = 26°$), where there is no interaction between substrate and Ile 395, but is seen in R3 and R4, both of which have substrate contacts with Ile 395.

Conclusions

The current work demonstrates that it is possible to identify discrete regions of enzyme structure that are sensitive not only to the presence of substrate, but to the orientation that the substrate assumes when bound. This supports our hypothesis that mechanical coupling between secondary structural features plays a vital role in enzyme function, and that evolutionary pressure is exerted not only on the active site, but on the entire enzyme structure in order to maintain a given function or evolve a new one. Where experimental data is available, particularly with respect to responses in the C–D loop and K′ helix of CYP101A1, the results of the PRS and experiment are complementary. The current work also suggests new avenues for exploring structure-function relationships in cytochromes P450, including the response of the CYP101A1 structure to the binding of non-native substrates.

Experimental

Residual dipolar coupling measurements and incorporation into molecular dynamics simulations

Sequential resonance assignments, methods for measurement of RDCs for CYP101A1 in two alignment media and their incorporation into structural calculations have been described in detail previously [5, 6]. RDC lists used in the current calculations are available in Supplementary Material.

Molecular dynamics (MD)

Much of this work was performed using the Extreme Science and Engineering Discovery Environment (XSEDE) node Comet at the San Diego Supercomputer Center [29]. Initial coordinates of CYP101A1 were taken from the RDC-based solution structure of *d*-camphor-bound CYP101A1 (PDB entry 2L8M, ref. [5]). Coordinates for CYP101A1 without camphor were taken from solution structure of substrate-free CYP101A1 (PDB entry 2LQD [6]. For each system, the enzyme was initially placed in a TIP3P water box containing 19772 molecules, KCl was added, resulting in a neutral solution [30]. The final size of the system was 94 ×91×93 $Å^3$. The Amber force field ff14SB was used to describe the enzyme [31]. This preparation was performed using *tleap* [32]. The energy was minimized first by steepest descent, followed by conjugate gradient minimization. The systems were equilibrated with constant NPT first, during 1.5 ns, using a Berendsen barostat with semi-isotropic pressure scaling, and then with constant NVT using a Langevin thermostat with collision frequency of 2 ps^{-1}. Production runs between 80–100 ns were then performed. Data was saved every 1 ps for later analysis.

Restraints on camphor

For the simulations where camphor was fixed, position restraints were used with a force constant of 300 kcal mol^{-1} $Å^{-2}$ applied only on camphor. For simulations where camphor was restrained at a specific angle, harmonic restraints were used centered on the dihedral angle Ψ defined between the heme Fe and NB pyrrole nitrogen and the camphor (substrate) carbon atoms C7 and C9 as in Fig. 3 with a force constant of 10 kcal mol^{-1} $Å^{-2}$. This same restraint was used in umbrella sampling calculations.

Umbrella Sampling (US)

Starting with a selected snapshot from the trajectory corresponding to 2L8M with camphor unrestrained, a US simulation was set up, in which the dihedral angle Ψ was restrained to force camphor rotation relative the heme. The US protocol consisted of 360 sampling windows, each separated by 10°, where the dihedral angle Ψ was varied using a harmonic restraint with a harmonic constant of 10 kcal mol^{-1} $Å^{-2}$. A 10 ns MD simulation was performed for each window using the same parameters as in the production runs. This protocol was performed three times, resulting in three independent US simulations. US results were analyzed with the Weighted Histogram Analysis Method, WHAM [33]. The error was calculated by constructing several profiles using trajectory lengths of 20000,

30000, and 40000 frames. Free energies were calculated and compared from each trajectory, yielding an error of +/−0.5 kcal/mol.

Effective spring constant map (K_{eff})

A set of coordinates taken from the MD trajectory corresponding to 2L8M with no restraints on camphor was modeled with ANM, using only alpha carbon atoms from the protein chain, a cutoff of 15 Å and corresponding eigenvalues and eigenvectors were calculated. The K_{eff} matrix was then calculated using Equation 2.

PRS heat maps

Covariance matrices were calculated using 20–30 ns segments of MD trajectories extracted after apparent equilibration was reached (that is, no significant changes in structure were observed). Calculations were performed using CPPTRAJ [34]. With covariance matrices, PRS matrices were calculated using Equation 4.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

TCP thanks Prof. R. Wade (EMBL Heidelberg) for helpful discussions. This work was supported in part by US NIH grant R01-GM44191 (TCP). EKA acknowledges support from Agencia Nacional de Promoción Científica y Tecnológica, under award PICT-2015-1706. TCP also acknowledges access to the NSF-supported XSEDE supercomputing consortium through a starter grant (TG-MCB160203).

References

1. Wei JY, Pochapsky TC, Pochapsky SS. Detection of a high-barrier conformational change in the active site of cytochrome P450(cam) upon binding of putidaredoxin. Journal Of The American Chemical Society. 2005; 127:6974–6. [PubMed: 15884940]
2. OuYang B, Pochapsky SS, Dang M, Pochapsky TC. A functional proline switch in cytochrome P450(cam). Structure. 2008; 16:916–23. [PubMed: 18513977]
3. Asciutto EK, Madura JD, Pochapsky SS, OuYang B, Pochapsky TC. Structural and dynamic implications of an effector-induced backbone amide cis-trans isomerization in cytochrome P450(cam). Journal of Molecular Biology. 2009; 388:801–14. [PubMed: 19327368]
4. Prestegard JH, Bougault CM, Kishore AI. Residual dipolar couplings in structure determination of biomolecules. Chemical Reviews. 2004; 104:3519–40. [PubMed: 15303825]
5. Asciutto EK, Dang M, Pochapsky SS, Madura JD, Pochapsky TC. Experimentally Restrained Molecular Dynamics Simulations for Characterizing the Open States of Cytochrome P450(cam). Biochemistry. 2011; 50:1664–71. [PubMed: 21265500]
6. Asciutto EK, Young MJ, Madura JD, Pochapsky SS, Pochapsky TC. Solution structural ensembles of substrate-free cytochrome P450cam. Biochemistry. 2012; 51:3383–93. [PubMed: 22468842]
7. Colthart AM, Tietz DR, Ni Y, Friedman JL, Dang M, Pochapsky TC. Detection of substrate-dependent conformational changes in the P450 fold by nuclear magnetic resonance. Scientific Reports. 2016; 6:22035. [PubMed: 26911901]
8. Tietz DR, Podust LM, Sherman DH, Pochapsky TC. Solution Conformations and Dynamics of Substrate-Bound Cytochrome P450 MycG. Biochemistry. 2017; 56:2701–14. [PubMed: 28488849]
9. Li SY, Tietz DR, Rutaganira FU, Kells PM, Anzai Y, Kato F, et al. Substrate Recognition by the Multifunctional Cytochrome P450 MycG in Mycinamicin Hydroxylation and Epoxidation Reactions. Journal of Biological Chemistry. 2012:287.

10. Tietz DR, Colthart AM, Pochapsky SS, Pochapsky TC. Substrate recognition by two different P450s: Evidence for conserved roles in a common fold. Scientific Reports. 2017; 7:13581. [PubMed: 29051575]
11. Dang M, Pochapsky SS, Pochapsky TC. Spring-loading the active site of cytochrome P450cam. Metallomics. 2011; 3:339–43. [PubMed: 21186391]
12. Arnold FH. The nature of chemical innovation: new enzymes by evolution. Quarterly Reviews of Biophysics. 2015; 48:404–10. [PubMed: 26537398]
13. Lewis JC, Arnold FH. Catalysts on Demand: Selective Oxidations by Laboratory-Evolved Cytochrome P450 BM3. CHIMIA International Journal for Chemistry. 2009; 63:309–12.
14. Atilgan C, Atilgan AR. Perturbation-Response Scanning Reveals Ligand Entry-Exit Mechanisms of Ferric Binding Protein. PLOS Computational Biology. 2009; 5:e1000544. [PubMed: 19851447]
15. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. Biophysical Journal. 2001; 80:505–15. [PubMed: 11159421]
16. Joo H, Lin ZL, Arnold FH. Laboratory evolution of peroxide-mediated cytochrome P450 hydroxylation. Nature. 1999; 399:670–3. [PubMed: 10385118]
17. Dorner ME, McMunn RD, Bartholow TG, Calhoon BE, Conlon MR, Dulli JM, et al. Comparison of intrinsic dynamics of cytochrome p450 proteins using normal mode analysis. Protein Science. 2015; 24:1495–507. [PubMed: 26130403]
18. Yang G, Hong N, Baier F, Jackson CJ, Tokuriki N. Conformational Tinkering Drives Evolution of a Promiscuous Activity through Indirect Mutational Effects. Biochemistry. 2016; 55:4583–93. [PubMed: 27444875]
19. Francis K, Stojkovic V, Kohen A. Preservation of Protein Dynamics in Dihydrofolate Reductase Evolution. Journal of Biological Chemistry. 2013; 288:35961–8. [PubMed: 24158440]
20. Campbell E, Kaltenbach M, Correy GJ, Carr PD, Porebski BT, Livingstone EK, et al. The role of protein dynamics in the evolution of new enzyme function. Nature Chemical Biology. 2016; 12:944. [PubMed: 27618189]
21. Wade RC, Motiejunas D, Schleinkofer K, Sudarko, Winn PJ, Banerjee A, et al. Multiple molecular recognition mechanisms. Cytochrome P450 - A case study. Biochimica Et Biophysica Acta-Proteins and Proteomics. 2005; 1754:239–44.
22. Ludemann SK, Lounnas V, Wade RC. 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. Journal of Molecular Biology. 2000; 303:797–811. [PubMed: 11061976]
23. Hlavica P. Challenges in assignment of allosteric effects in cytochrome P450-catalyzed substrate oxidations to structural dynamics in the hemoprotein architecture. Journal of Inorganic Biochemistry. 2017; 167:100–15. [PubMed: 27919007]
24. Panek A, Swizdor A, Milecka-Tronina N, Panek JJ. Insight into the orientational versatility of steroid substrates-a docking and molecular dynamics study of a steroid receptor and steroid monooxygenase. Journal of Molecular Modeling. 2017:23.
25. Dodani SC, Kiss G, Cahn JKB, Su Y, Pande VS, Arnold FH. Discovery of a regioselectivity switch in nitrating P450s guided by molecular dynamics simulations and Markov models. Nature Chemistry. 2016; 8:419–25.
26. Palmer AG. Enzyme Dynamics from NMR Spectroscopy. Accounts of Chemical Research. 2015; 48:457–65. [PubMed: 25574774]
27. Stacklies W, Seifert C, Graeter F. Implementation of force distribution analysis for molecular dynamics simulations. BMC Bioinformatics. 2011; 12:101. [PubMed: 21501475]
28. Pochapsky TC, Kazanis S, Dang M. Conformational plasticity and structure/function relationships in cytochromes P450. Antioxidants & Redox Signaling. 2010; 13:1273–96. [PubMed: 20446763]
29. Towns J, Cockerill T, Dahan M, Foster I, Gaither K, Grimshaw A, et al. XSEDE: Accelerating Scientific Discovery. Computing in Science & Engineering. 2014; 16:62–74.
30. Jorgensen WL, Madura JD. Quantum and statistical studies of liquids. 25. Solvation and conformation of methanol in water. Journal of the American Chemical Society. 1983; 105:1407–13.

31. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. Journal of Chemical Theory and Computation. 2015; 11:3696–713. [PubMed: 26574453]

32. Case, DAJTB., Betz, RM., Cerutti, DS., Cheatham, TE., III, Darden, TA., Duke, RE., Giese, TJ., Gohlke, H., Goetz, AW., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, TS., LeGrand, S., Li, P., Luchko, T., Luo, R., Madej, B., Merz, KM., Monard, G., Needham, P., Nguyen, H., Nguyen, HT., Omelyan, I., Onufriev, A., Roe, DR., Roitberg, A., Salomon-Ferrer, R., Simmerling, CL., Smith, W., Swails, J., Walker, RC., Wang, J., Wolf, RM., Wu, X., York, DM., Kollman, PA. AMBER 16.0. University of California; San Francisco: 2016.

33. Grossfield, A. WHAM: the weighted histogram analysis method. v. 2.0.9.1 ed.

34. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. Journal of Chemical Theory and Computation. 2013; 9:3084–95. [PubMed: 26583988]

35. Raag R, Poulos TL. Crystal structure of the carbon monoxide-substrate-cytochrome P450CAM ternary complex. Biochemistry. 1989; 28:7586–92. [PubMed: 2611203]

Highlights

- Substrate orientation in P450cam is maintained by secondary structure interactions.
- Secondary structures re-organize upon substrate binding.
- Secondary structures maintain integrity in the absence of substrate.
- Mechanical coupling patterns change with changing substrate orientations.
- Substrate provides a focal point for mechanical couplings.

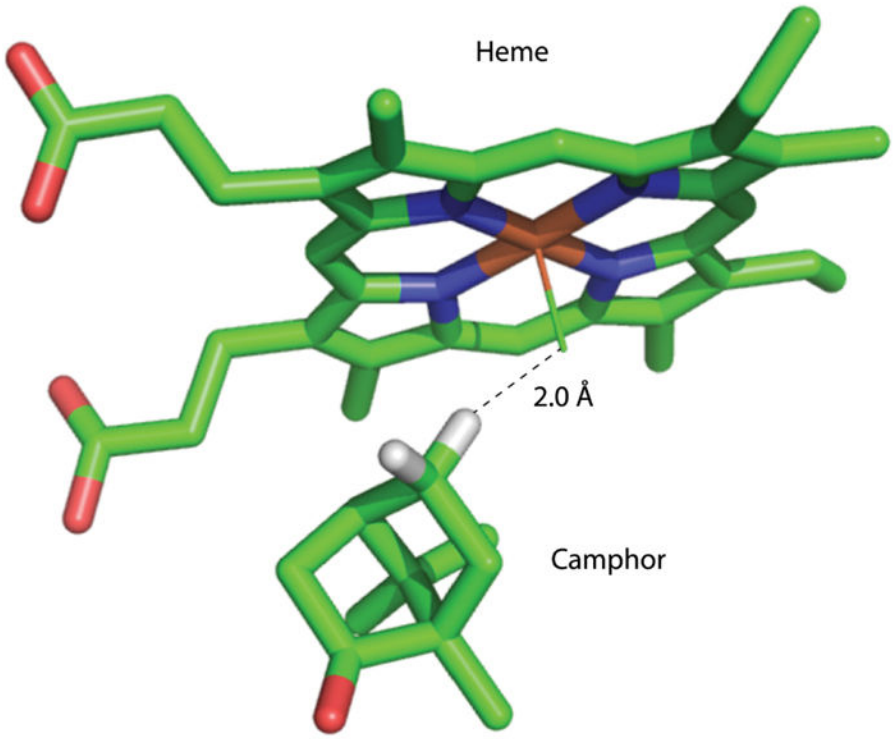


Figure 1.

Camphor orientation in the active site of reduced and carbonmonoxy-bound CYP101A1 from crystal structure 3CPP [35]. Distance shown by dotted line (2.0 Å) is from the 5-*exo*-hydrogen of camphor to the carbon atom of bound carbon monoxide (approximately the same position as the active oxygen the Fe=O complex).

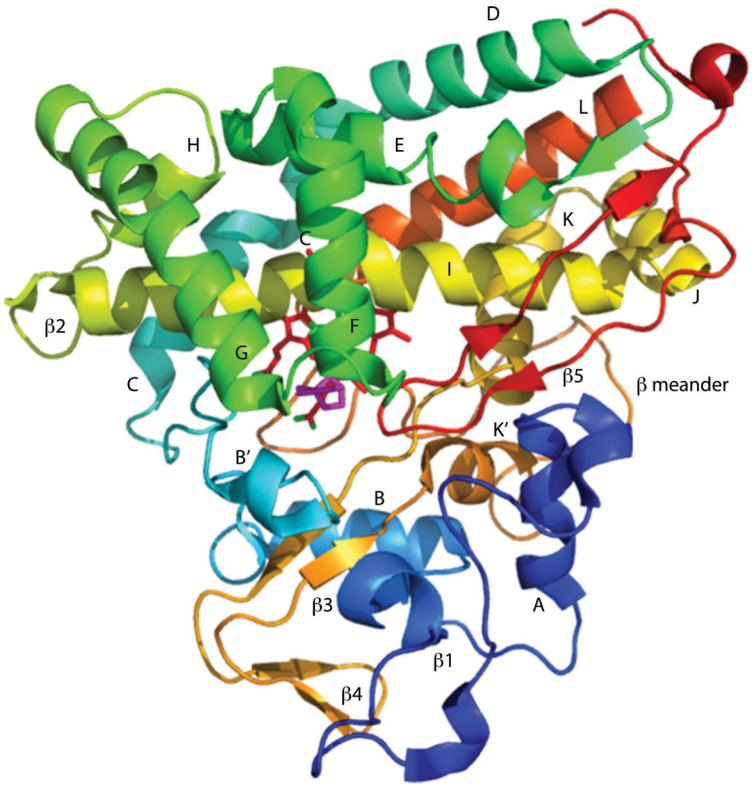


Figure 2.

Solution structure 2L8M of CYP101A1 [5], with secondary structural features referred to in the text labeled according to the scheme of Poulos [35]. Heme is shown as red sticks, camphor as magenta. Residue numbers are as follows: **A** helix, V38-E47; **β1**, L53-A65; **B** helix, R67-D77; **B′** helix, P89-E94; **C** helix, Q108-V119; **D** helix, D125-R143; **E** helix, F150-G168; **F** helix, P170-T185; **G** helix, T192-K214; **H** helix, A219-A224; **β2**, Q227-P232; **I** helix, T234-A265; **J** helix, S267-R277; **K** helix, R280-F292; **β3,** G298-L301 and Q317-L319; **β4,** Y305-L312; **K′** helix P321-L327; **β meander**, D328-L358, **L** helix, Q360-I378; **β5,** S382-V405.

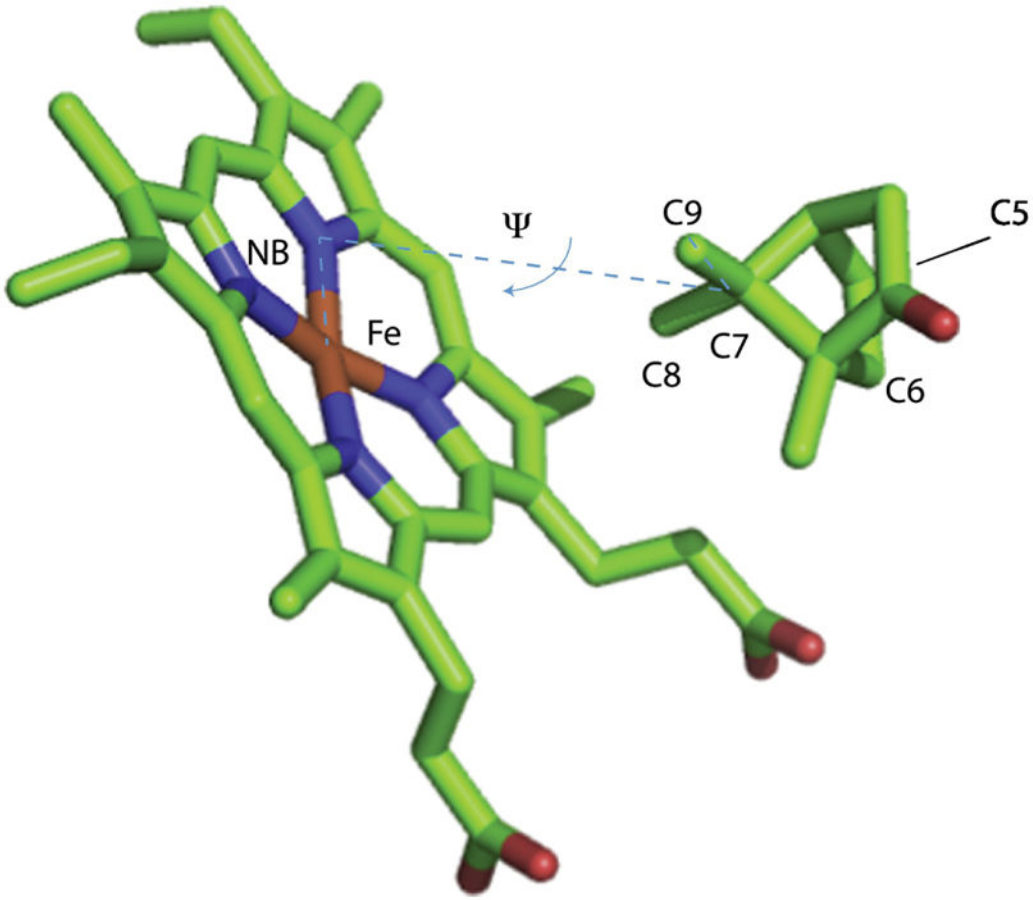


Figure 3.
Improper angle Ψ defined between heme Fe, NB, camphor C7 and C9 for restraint of camphor orientation in the active site of CYP101A1.

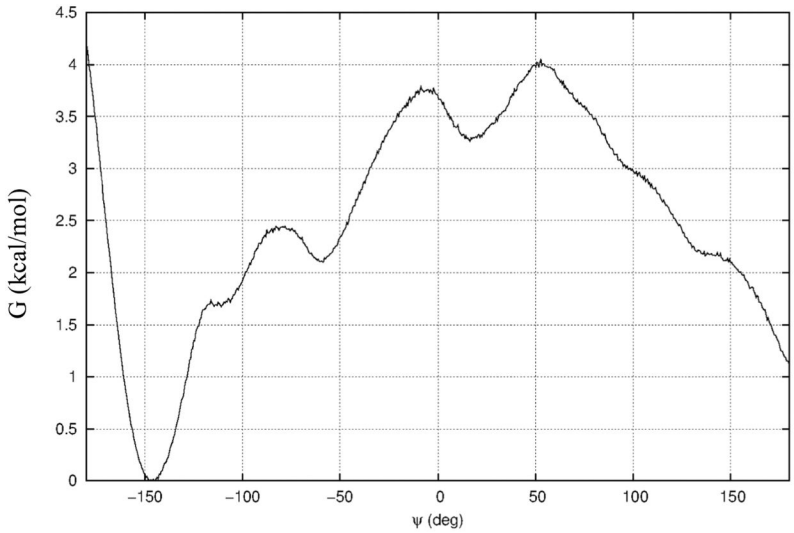


Figure 4.

Free energy profile extracted from umbrella sampling starting from 2L8M, with the reaction coordinate being the change in the improper dihedral angle Ψ as defined in Figure 3.

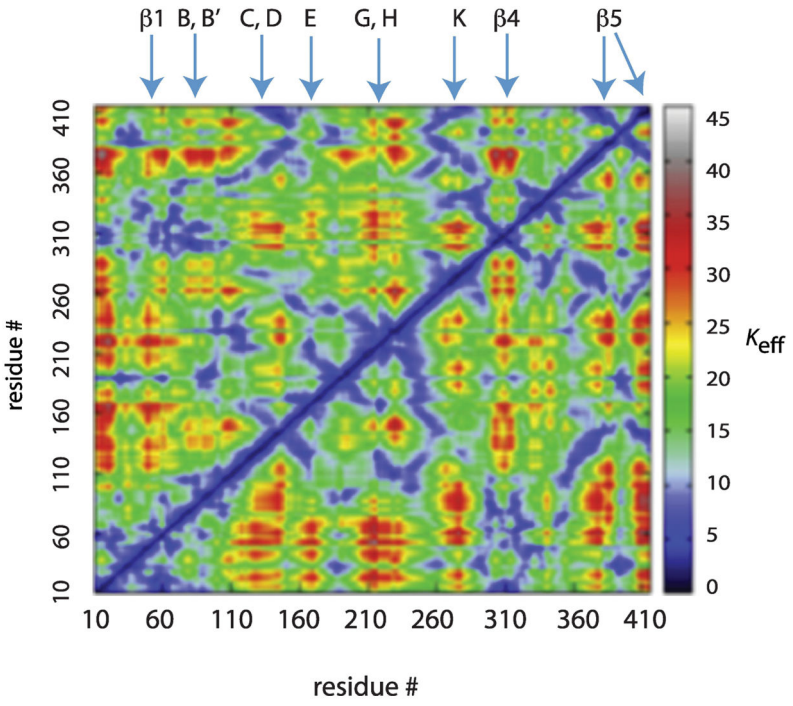


Figure 5.

Effective force constant (K_{ij}) map for 2L8M in arbitrary units. Secondary structural features are labeled at the top of the map as in Figure 1.

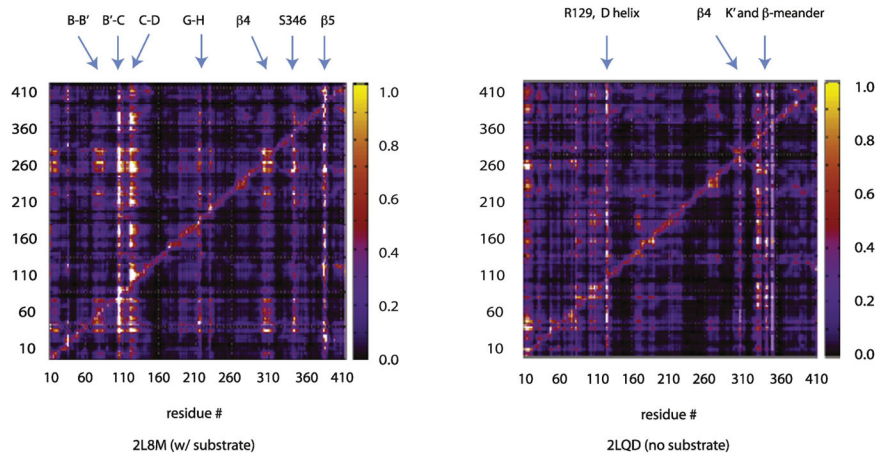


Figure 6.

Heat maps of Cα perturbation response scanning (PRS) of RDC-derived solution structures of CYP101A1 with (2L8M) and without (2LQD) *d*-camphor bound. The stronger the motional correlation, the “hotter” the cross-peak between positions. The most correlated regions in each map are labeled with reference to secondary structural features from Fig. 2.

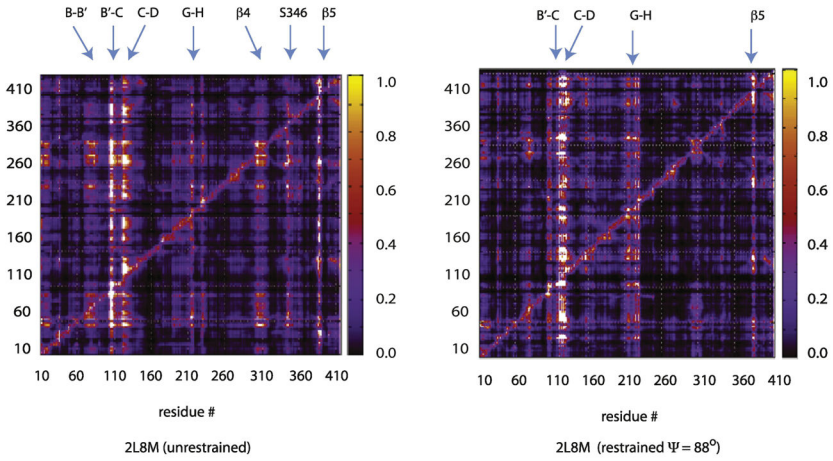


Figure 7.

Comparison of Cα PRS heat maps between unrestrained Ψ angle (left) and Ψ restrained to the starting angle from 2L8M after equilibration (88°). See Figure 2 for definition of Ψ. Axes correspond to residue number in both dimensions. Lettering corresponds to structural features shown in Figure 1. The intense feature labeled β5 in both maps corresponds to Gly 386 near the end of one strand of the β5 sheet.

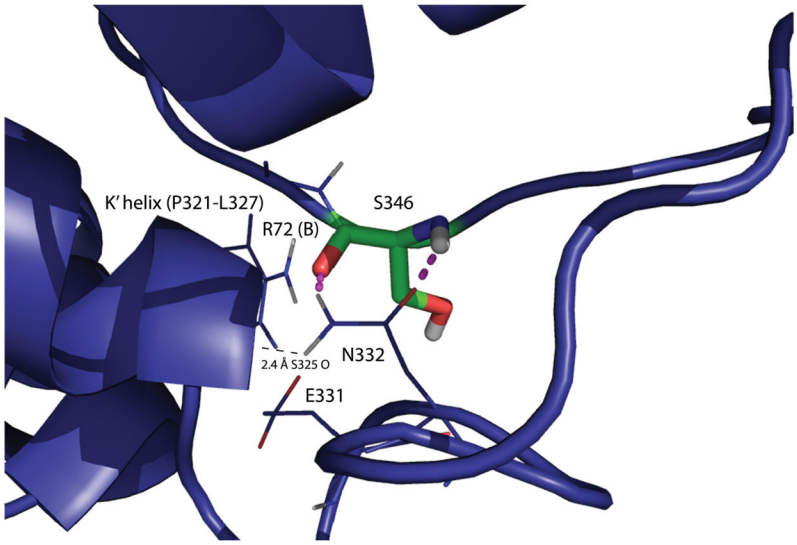


Figure 8.

Hydrogen bonding network involving Ser 346 (β-meander). Ser 346 is sensitive to motions in the B–B′ and B′–C loop in the presence of substrate (Fig. 7). The backbone amide and carbonyl of Ser 346 forms β-type hydrogen bonds with the side chain of Asn 332 (red dotted lines), which in turn is within hydrogen bonding distance of Ser 325 C=O in the K′ helix. The amide NH resonance of Ser 325 is strongly perturbed upon substrate binding [11], and Glu 331 was found by directed evolution to affect substrate selectivity in CYP101A1 (Ref. [16]).

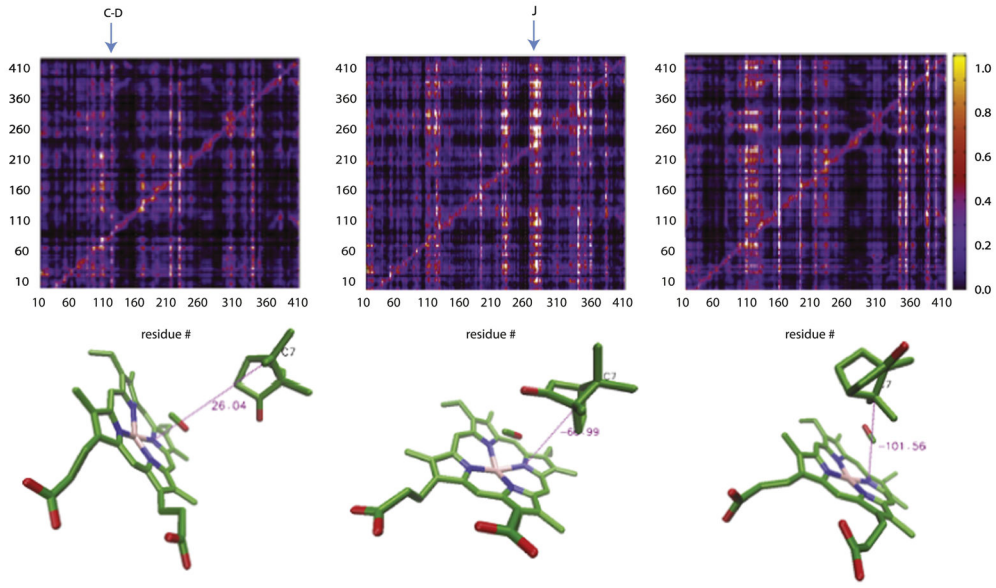


Figure 9.

PRS matrices for selected camphor orientations. Ψ= 26° (left), Ψ= −60° (center), and Ψ= −100° (right).

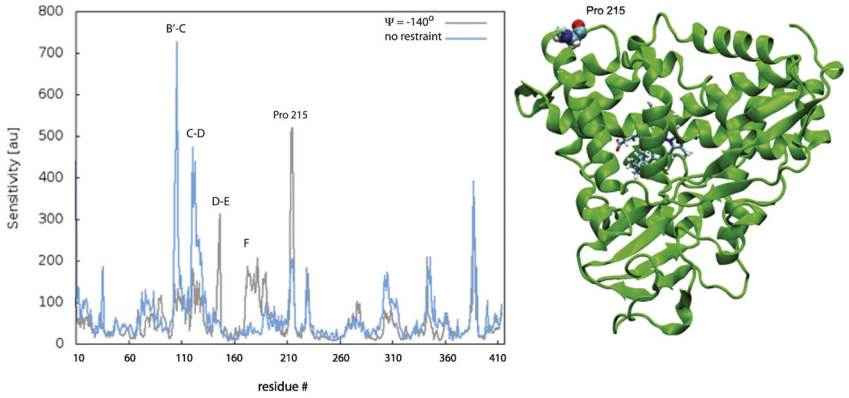


Figure 10.

Left: Sensitivity profile in arbitrary units corresponding to the dihedral angle ψ= −140° (black line), compared with sensitivity profile with no restrictions applied to camphor (blue line). **Right**: position of Pro 215 at the end of the G helix, showing an increased sensitivity when ψ= −140°.

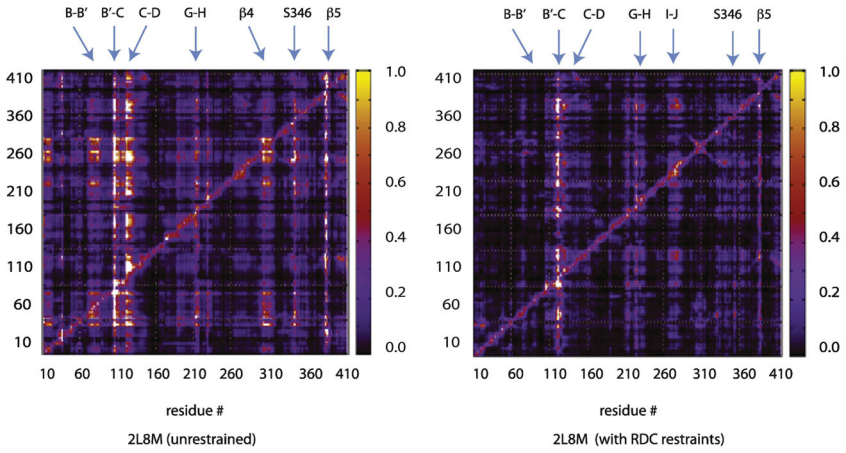


Figure 11.

PRS heat map obtained for 2L8M with RDC restraints applied (right) compared with the map for unrestrained 2L8M (from Figs. 6 and 7).

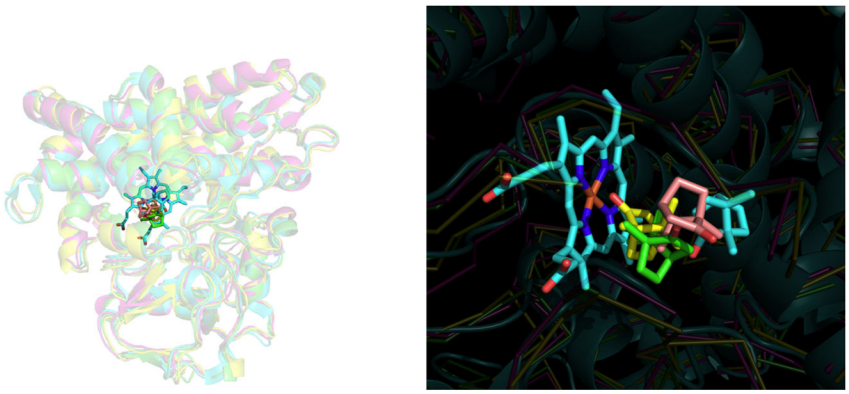


Figure 12.

Left, superposition of 2L8M on structures extracted from dynamics tracks with restraints on Ψ. Right, close-up of camphor orientations from the superposition: 2L8M (green), Ψ = 26° (cyan), Ψ −100° (salmon), Ψ = −60° (yellow).

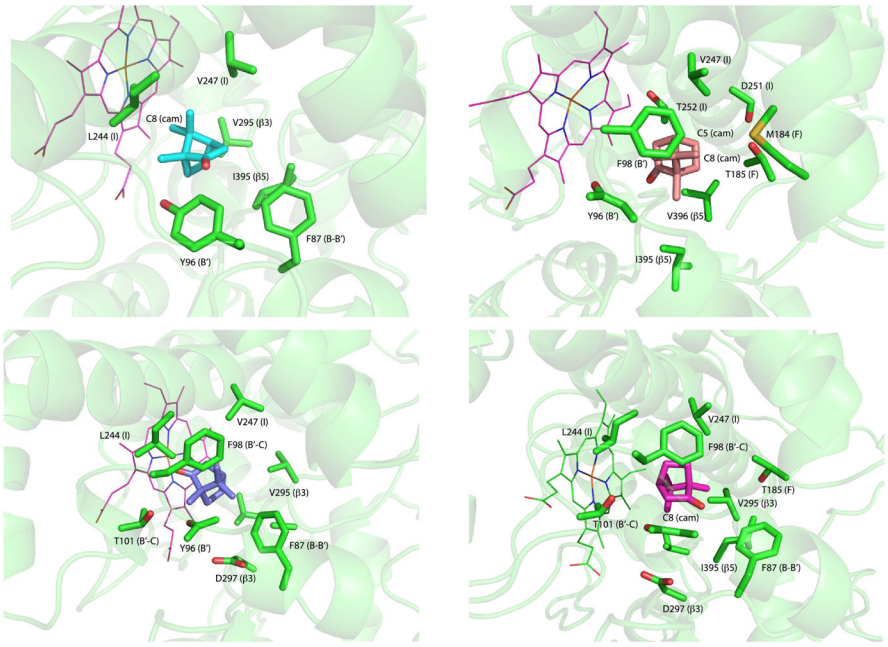
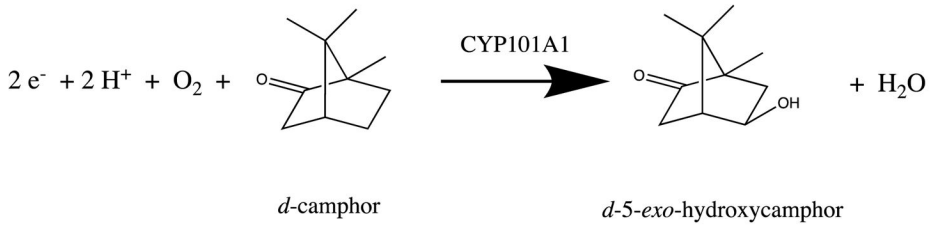


Figure 13.

First sphere contacts in CYP101A1 active site as a function of Ψ. **Top left**: 2L8M. **Top right**, R2: Ψ = 26 °. **Bottom left,** R3, Ψ = −60 °, bottom right Ψ = −100°. Methyl carbon C8 in substrate camphor is identified. In 2L8M, C8 contacts L244 and V247 on the I helix. In R2, C8 contacts V247, M184 and T185 (F). In R3, D297 is only major contact for C8, although F87 (B–B′) is within 5 Å. In R4, C8 contacts V295, D297 (β3) and I395 (β5).



d-camphor *d*-5-*exo*-hydroxycamphor

Scheme 1.