Explicit Water Near the Catalytic I Helix Thr in the Predicted Solution Structure of CYP2A4

Anna Gorokhov,*† Masahiko Negishi,† Eric F. Johnson,‡ Lars C. Pedersen,† Lalith Perera,* Tom A. Darden,† and Lee G. Pedersen*†

*Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina 27599; †National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709; ‡The Scripps Research Institute, La Jolla, California 92037

ABSTRACT The solution structure of mouse cytochrome P450 2A4 (CYP2A4), a monooxygenase of deoxysteroids, was obtained using homology modeling and molecular dynamics. The solvent-equilibrated CYP2A4 preserves the essential features of CYP450s. A comparison of the models CYP2A4 and CYP2A4 with testosterone bound CYP2A4/T illustrates the changes induced by the binding of the substrate. Experimental evidence links four amino acid residues to the catalytic activity, substrate specificity, and regioselectivity of this enzyme. Three of the four amino acids are found within contact distance of the testosterone substrate, and therefore may control the binding of the substrate through direct interaction. Remarkably, a water complex previously observed in x-ray crystal structure forms near the bulge in the central I helix that contains a conserved Thr. The properties of the I helix are computed in the context of the presence or absence of ligand.

INTRODUCTION

Cytochrome P450 (CYP450) monooxygenases catalyze biological activation and transfer of oxygen during the synthesis of steroid hormones and the metabolism of exogenous compounds. Metabolism of xenobiotics by CYP450s usually reduces the toxicity and promotes excretion of these xenobiotics. However, CYP450 interactions may sometimes increase toxicity of drugs, with lethal outcomes (Guengerich, 2001). Therefore, biomedical researchers are greatly interested in understanding and controlling the interactions of CYP450s with novel drugs. The ability to screen out potentially harmful CYP450 interactions with drugs would be greatly improved by the availability of a three-dimensional model of CYP450. Potential drug candidates could then be tested by docking into the modeled active site. For this task, molecular modeling proves to be a valuable tool. Modeling not only allows one to visualize the fold of the enzyme, but also to examine closely the structural details underlying the substrate-enzyme interactions. The modeling work is facilitated by the general fold similarity that exists across the diverse CYP450 families, including both eukaryotic and prokaryotic CYP450s (Dai et al., 2000). In animals, more than 40 cytochrome P450 gene families give rise to a vast number of different CYP450s. However, certain structural elements, especially those involved in the catalytic activity of CYP450, are conserved throughout these families (Williams et al., 2000a). These conserved motifs make possible the construction of plausible molecular models of CYP450 enzymes for which structural data is not yet available.

Submitted May 21, 2002, and accepted for publication September 3, 2002.

Address reprint requests to Lee G. Pedersen, Dept. of Chemistry, University of North Carolina, Chapel Hill, NC 27599-3290. Tel.: 919-962-1578; Fax: 919-962-2388; E-mail: lee_pedersen@unc.edu.

0006-3495/03/01/57/12 $2.00

In this study, we develop an equilibrated solution structure of cytochrome P450 2A4 (CYP2A4), a 15-α-hydroxylase of 11-deoxysteroids (e.g., testosterone). CYP2A4, a mouse enzyme, although arguably of less immediate interest for the biomedical community than human CYP450s, has a rich associated experimental database. It shares 83% sequence identity with a human CYP2A6, a coumarin hydroxylase. CYP2A6 has been shown to activate many precarcinogens (Oscarson et al., 1998). Its murine counterpart, coumarin hydroxylase CYP2A5, has also been implicated in cancer (Kobliakov et al., 1993). CYP2A5 is closely related to CYP2A4, as the sequences of the two enzymes differ by only 11 amino acids. However, CYP2A5 exhibits a very different substrate specificity from that of CYP2A4. Pertinent experimental data is summarized in Table 1. Wild-type CYP2A5 binds and hydroxylates coumarin at position 7, and has no detectable activity toward testosterone. This diverse substrate specificity in two closely related enzymes illustrates a general theme in the functioning of CYP450 enzymes, which exhibit great substrate diversity as a group while individually maintaining high substrate specificity. The specificity and regioselectivity in a single enzyme are effected by a few mutations, most probably at the active site. Site-directed mutagenesis has been applied to identify the specific amino acids that control this specificity. Four residues (at positions 117, 209, 365, and 481) identified by these methods were proposed to be located in the heme-binding cavity, and possibly come into contact with the substrate (Lindberg and Negishi, 1989; Iwasaki et al., 1993a and 1993b; Negishi et al., 1996a; 1996b; 1996c). Most strikingly, a single mutation, Phe209Leu, is required to confer steroid 15-α-hydroxylase specificity to CYP2A5. We aim to develop a model that will enable us to examine the variations in the geometry of the binding pocket brought about by these mutations, and that lead to functional diversity between these two enzymes. The conservation of the functional importance of residues at positions 117, 209,

TABLE 1 Relative 2A4/2A5 substrate activity (Negishi et al., 1996b)

Amino acid position					Hydrolase activity*	
P450[†]	117	209	481	DHEA	Coumarin	Testosterone
mP450 2A5	Ala[‡]	Phe	Ala	7β-OH	Low	Low
sP450 2A5	Ala	Phe	Ala	—	Low	Low
dP450 2A5^{H}	Val	Phe	Ala	2α-OH	High	Low
dP450 2A5^{L}	Ala	Phe	Ala	7β-OH	Low	Low
dP450 2A4	Ala	Leu	Val	7α-OH	nd[§]	High (15α)

*High (H) testosterone C−15-α-hydrolase activity of dP450 2A4 is 30-fold greater than the low (L) activity in P450s 2A5.
[†]Strains described in Negishi, et al., 1996b.
[‡]The bold amino acids are the unique residue(s) mutation of which will change specificity.
[§]Activity not detected by experiment.

365, and 481 throughout the CYP2 family makes this study broadly applicable to CYP450 (Negishi et al., 1996b).

The model of CYP2A4 presented here is based mainly on two pieces of experimental evidence: the first mammalian x-ray crystal structure of CYP2C5, and extensive mutagenesis studies performed on CYP2A4 and CYP2A5 enzymes as described above. In the absence of an x-ray crystal structure of CYP2A4, anticipated fold similarities between members of CYP2 families are utilized to build a three-dimensional model of CYP2A4 in solution based on the recently solved x-ray crystal structure of CYP2C5 at 3.0 Å resolution (Williams et al., 2000a). CYP2C5 is a rabbit cytochrome, and the first mammalian CYP450 for which an x-ray crystal structure has been obtained. As a template for modeling other mammalian CYP450s, it is an improvement over the bacterial CYP450s, which were used before its publication. It has a higher sequence identity to other mammalian CYP450s than do bacterial CYP450s. Moreover, the fact that mammalian CYP450s are membrane bound whereas the bacterial enzymes are soluble accounts for some significant structural differences (Williams et al., 2000b). Despite these differences, bacterial CYP450 structures were useful in predicting the structure of a mammalian CYP450. Previous studies of bacterial CYP450s point to the recurring motifs in the three-dimensional structure of all CYP450s. Six substrate recognition sites common to bacterial and mammalian CYP450s were identified (Gotoh, 1992). Some important trends emerged, in particular, of three early bacterial CYP450s: $CYP450_{cam}$, $CYP450_{terp}$, and $CYP450_{bm3}$ (Hasemann et al., 1995). The most striking similarity first observed among the three bacterial structures and extended to the mammalian CYP450s is the conserved region in the core of the protein. The core contains the central and carboxy terminal part of the I helix, the L and J helices, and the residues around the cysteine that ligates the heme, referred to as the Cys-pocket. The most flexible parts of the structures are located closer to the surface of the protein, and include the F, G, A, and B helices. In particular, the FG loop is expected to be extremely flexible, inasmuch as it varies in length among CYP450s. The general features of CYP2C5 and the extent to which they were conserved in the CYP2A4 model will be discussed later. The most relevant feature (for the purposes of our work) that should be emphasized here is the geometry of the active site surrounding the heme in CYP2C5. In our sequence alignment of CYP2C5 and CYP2A4 (Fig. 1), four of the amino acids that line the active site in CYP2C5 (Ala-113, Val-205, Leu-358, and Val-474) map to the four residues in CYP2A4 that were experimentally determined to control substrate specificity (Ala-117, Val-209, Leu-365, and Val-481). This observation lends support to the hypothesis that the key amino acids control the specificity through direct interactions with the substrate.

CYP450 modeling is a vast and active area of research. The structure presented here is a homology model of a mammalian CYP450 based on the first experimental structure of another mammalian CYP450; however, it follows a long line of previous modeling work targeting CYP450s. Before the x-ray crystal structure of CYP2C5

```
CYP2A4 MLTSGLLLVAAVAFLSVLVLMSVWKQRKLSGKLPPGPTPLPFVGNFLQLNTEQMYNSLMK   60
CYP2A5 MLTSGLLLVAAVAFLSVLVLMSVWKQRKLSGKLPPGPTPLPFIGNFLQLNTEQMYNSLMK   60
CYP2C5 ----XXXXXXXXXXXXXXXXXXXXXXXXXXXPPGPTPFPIIGNILQIDAKDISKSLTK   56
                                                    h      hA

CYP2A4 ISQRYGPVFTIYLGSRRIVVLCGQEAVKEALVDQAEEFSGRGEQATFDWLFKGYGIAFSS  120
CYP2A5 ISQRYGPVFTIYLGPRRIVVLCGQEAVKEALVDQAEEFSGRGEQATFDWLFKGYGVVFSS  120
CYP2C5 FSECYGPVFTVYLGMKPTVVLHGYEAVKEALVDLGEEFAGRGSVPILEKVSKGLGIAFSN  116
                  b      b      hB

CYP2A4 GERAKQLRSFSIATLRDFGVGKRGIEERIQEEAGFLIDSFRKTNGAFIDPTFYLSRTVSN  180
CYP2A5 GERAKQLRRFSIATLRDFGVGKRGIEERIQEEAGFLIDSFRKTNGAFIDPTFYLSRTVSN  180
CYP2C5 AKTWKEMRRFSLMTLRNFGMGKRSIEDRIQEEARCLVEGLRKTNASPCDPTFILGCAPCN  176
            hC                          hD                    hE

CYP2A4 VISSIVFGDRFDYEDKEFLSLLRMMLGSLQFTATSMGQVYEMFSSVMKHLPGPQQQAFKE  240
CYP2A5 VISSIVFGDRFDYEDKEFLSLLRMMLGSFQFTATSMGQLYEMFSSVMKHLPGPQQQAFKE  240
CYP2C5 VICSVIFHNRFDYKDEEFLKLMESLHENVELLGTPWLQVYNNFPALLDYFPGIHKTLLKN  236
                              hF

CYP2A4 LQGLEDFITKKVEHNQRTLDPNSPRDFIDSFLIRMLEEKKNPNTEFYMKNLVLTTLNLFF  300
CYP2A5 LQGLEDFITKKVEHNQRTLDPNSPRDFIDSFLIRMLEEKKNPNTEFYMKNLVLTTLNLFF  300
CYP2C5 ADYIKNFIMEKVKEHQKLLDVNNPRDFIDCFLIKMEQEN---NLEFTLESLVIAVSDLFG  293
            hG                  hH

CYP2A4 AGTETVSTTLRYGFLLLMKYPDIEAKVHEEIDRVIGRNRQPKYEDRMKMPYTEAVIHEIQ  360
CYP2A5 AGTETVSTTLRYGFLLLMKHPDIEAKVHEEIDRVIGRNRQPKYEDRMKMPYTEAVIHEIQ  360
CYP2C5 AGTETTSTTLRYSLLLLLKHPEVAARVQEEIERVIGRHRSPCMQDRSRMPYTDAVIHEIQ  353
             hI                 hJ                h        hK

CYP2A4 RFADLIPMGLARRVTKDTKFRDFLLPKGTEVFPMLGSVLKDPKFFSNPKDFNPKHFLDDK  420
CYP2A5 RFADMIPMGLARRVTKDTKFRDFLLPKGTEVFPMLGSVLKDPKFFSNPKDFNPKHFLDDK  420
CYP2C5 RFIDLLPTNLPHAVTRDVRFRNYFIPKGTDIITSLTSVLHDEKAFPNPKVFDPGHFLDES  413
                b                  b    h              h

CYP2A4 GQFKKSDAFVPFSIGKRYCFGEGLARMELFLFLTNIMQNFHFKSTQAPQDIDVSPRLVGF  480
CYP2A5 GQFKKNDAFVPFSIGKRYCFGEGLARMELFLFLTNIMQNFHFKSTQAPQDIDVSPRLVGF  480
CYP2C5 GNFKKSDYFMPFSAGKRMCVGEGLARMELFLFLTSILQNFKLQSLVEPKDLDITAVVNGF  473
                              hL        b        h

CYP2A4 VTIPPTYTMSFLSR--  494
CYP2A5 ATIPPTYTMSFLSR--  494
CYP2C5 VSVPPSYQLCFIPIHH  489
                  b
```

FIGURE 1 Sequence alignment of CYP2A4, CYP2A5, and CYP2C5. Residues experimentally shown to modulate substrate specificity (Ala-117, Leu-209, Leu-365, and Val-481 in CYP2A4) are shown in gray (Negishi et al., 1996). Thr-305, conserved in bacterial and mammalian CYP450s, is also highlighted. The alignment was performed using Clustal X (Thompson et al., 1997). Helices and β-sheets are labeled with letters h and b, respectively. Additionally, helices A-L are identified, and residues comprising β-sheets are labeled in italics (as assigned by Williams et al., 2000a).

became available, molecular modeling studies of mammalian CYP450s were based primarily on their structural similarities to bacterial CYP450s. These studies aimed to map out the active site by building homology models to understand the structural causes underlying substrate specificity. Homology models made use of sequence alignment, and were challenged by the low sequence identity between mammalian and bacterial CYP450s (Dai et al., 2000). Such studies include the homology model of human CYP4A11 (Loew and Chang, 1999), the two subfamily (Lewis and Lake, 1995; Lewis et al., 1999) and one subfamily (Dai et al., 1998) to name a couple. Rather than focus exclusively on the active site, the model presented here is based on simulations under periodic boundary conditions on nanosecond time scales of all atoms in the system. Long-range interactions and solvation effects are accounted for, yielding a more complete, dynamical picture of the enzyme. CYP2A4 presented a special challenge to long time simulation because there are no stabilizing intramolecular disulfide bridges, and the heme is noncovalently bound.

COMPUTATIONAL PROCEDURE

Model building

The x-ray crystal structure of CYP450 2C53LV_dH (pdb code: 1DT6), which is CYP2C5 modified to increase solubility, served as a template for building CYP2A4 (Williams et al., 2000). The CYP2C5 x-ray crystal structure did not contain a ligand, nor was water detected in the substrate-binding cavity. A sequence alignment between CYP2C5 and CYP2A4 was performed (Fig. 1). The sequence of CYP2A5, which differs by only 11 residues from CYP2A4, was also included in the alignment. The sequence identity between CYP2C5 and CYP2A4 is 54% (for all residues). Before homology modeling, the CYP2C5 x-ray crystal structure was minimized in vacuum using AMBER (Case et al., 1999). Only side chains and hydrogen atoms (which were added to the crystal structure) were minimized for 10,000 steps, using a combination of steepest descent and conjugate gradient procedures. To preserve the initial backbone configuration of the x-ray crystal structure, no backbone atoms were allowed to move until the homology modeling was complete. The energy-minimized structure was gradually mutated into CYP2A4, a few residues at a time, using the Biopolymer module of SYBYL 6.5 (Tripos, St. Louis, MO). The mutated side chains were hydrogenated, and each set of mutations was followed by a local minimization of the mutated side chains and surrounding residues to alleviate steric conflicts. A total of 215 side chains were mutated. The fully mutated structure was superimposed on the x-ray crystal CYP2C5. The root mean square deviation (RMSD) of the backbone atoms was 0.1 Å, indicating that the perturbation of the backbone was negligible. At this point, additional residues were added to the model CYP2A4 using homology modeling. Specifically, amino acid residues 216-226 (FG loop) and 280-282 (gap in the sequence alignment) were modeled using the loop search algorithm in SYBYL 6.5. Finally the hydrogens and side chains of the model CYP2A4 were again minimized in vacuum for 100,000 steps, beginning with 5,000 steps of steepest descent minimization. All minimization cycles employed a nonbonded cutoff of 8 Å, with the nonbonded list updated at every step.

The Fe atom of the heme is hexavalent, coordinated by the sulfur atom of Cys-439 in the fifth axial position, and by an oxygen atom in the sixth axial position. The iron oxygen was absent from the x-ray crystal structure. Its addition to our CYP2A4 model was intended to guide the positioning of the ligand, and to reflect the active oxygen intermediate of the heme. The existing parameters and charges for this intermediate were used. These were previously developed to be consistent with the AMBER force field (Harris and Loew, 1995).

Molecular dynamics simulation

The energy-minimized model was evaluated with PROCHECK, which detected no steric conflicts. Subsequently, the system was solvated in a box of 16,867 TIP3P water molecules (Jorgensen et al., 1983). The walls of the box were initially located 12 Å from any protein atom. The initial size of the solvent box was 94.9Å $\times$ 90.8Å $\times$ 78.1Å. Three chloride counterions were added to keep the system electrically neutral during the simulation, bringing the total number of atoms in the system to 56,762. The simulation in solution was performed under periodic boundary conditions. The solvated CYP2A4 structure was energy minimized in solution. A stepwise energy minimization protocol was applied: the relaxation of the solvent molecules and counterions was followed by additional energy minimization of the solvent, counterions, and hydrogen atoms of the protein, and the side chains, for 30,000 steps. After minimization, the RMSD of the backbone atoms of the model with the CYP2C5 structure was 0.6 Å.

The system was then gradually heated from 0 K to 300 K in the course of constant-volume dynamics. Once the temperature reached 300 K, additional equilibration at constant volume and temperature was performed for 20 ps. During the entire constant volume run, the backbone was fixed by invoking the belly option in AMBER. However, when the production run at constant pressure was initiated, the belly option was removed as it sometimes may lead to stability difficulties when used with constant pressure algorithms. Instead, the backbone was kept fixed with Cartesian restraints, and then gradually relaxed by slowly reducing the force constant from 5.0 to 0 in the course of the initial 100 ps of dynamics.

This procedure was adopted after an earlier simulation, performed without the gradual relaxation of the backbone, resulted in a structure with major backbone deviations.

Although a degree of rearrangement in the backbone of the protein is expected for a homology model undergoing a long dynamical simulation, we found that the deviations observed in our model resulted in the loss of certain elements expected to be well- conserved. Most importantly, the heme exhibited a large motion (on the order of 4 Å) away from the I helix. This motion resulted in a partial unraveling of the L helix, situated directly under the heme. Additionally, the distances between the four catalytic amino acid residues and the heme differed greatly from the corresponding distances in the x-ray crystal CYP2C5. On the basis of this observation, the initial model of CYP2A4 was carefully reequilibrated with the procedure presented here. The resulting distances between the four catalytic residues and the heme, reported in our work, signify an improved model where the key elements are conserved.

After this point, the simulation was unconstrained, with the exception of the SHAKE algorithm employed to freeze out the vibrational motion of the hydrogen atoms (Ryckaert et al., 1977). The time step was 1 fs. The particle mesh Ewald summation was used to account for long-range interactions (Darden et al., 1993; Essman et al., 1995). The coordinate files were written and saved at every 5 ps. A trajectory of 3 ns was obtained.

Substrate docking and simulation of substrate-bound CYP2A4

The x-ray crystal structure of testosterone was obtained from the Cambridge Crystallographic Database (CSD reference code TESTON10). The geometry of the testosterone molecule was optimized with the Hartree-Fock method, employing the 6-31G* basis set in Gaussian 98 (Frisch et al., 1998). A set of RESP charges was fitted onto the optimized geometry of testosterone (Bayly et al., 1993). The AMBER force field parameters of testosterone were assigned by analogy. The testosterone ligand was manually docked into the binding site of CYP2A4. An equilibrated structure of CYP2A4 after 1 ns dynamics, with solvent molecules removed, was used. The ligand was initially placed in the active site so that the atoms of the ligand were no closer than 2.3 Å to any of the backbone atoms. This procedure ensured that no drastic perturbations to the backbone structure would be introduced as a result of ligand docking and subsequent energy minimization. The docking procedure placed the hydroxylated carbon C15 at 3.5 Å from the ferryl oxygen. Steric clashes between the ligand and protein atoms were alleviated with several minimization cycles in vacuum. Initially, the side chains of the active site residues (those within a 7 Å radius of the ligand) were energy minimized. In the second pass of energy minimization, all protein side chains were allowed to move. At this point, the backbone and the ligand remained fixed. The resulting structure was resolvated in a box of TIP3P water molecules. Three chloride ions were added to neutralize the charges of the system. The solvated system was once again equilibrated following the procedure described above, and a dynamics simulation was performed at constant pressure. During the heat-up and constant volume equilibration phases, the backbone atoms were kept fixed with belly constraints. For the constant pressure dynamics run, the belly constraints were released, and instead a Cartesian constraint was introduced for the backbone atoms and the ligand, for the initial 100 ps of the simulation. In the remainder of the simulation, the backbone was unconstrained. Instead a harmonic constraint was introduced between the atom C15 of the ligand and the ferryl oxygen of the heme. The constraint was used to gradually reduce the distance between these two atoms to 2.9 Å. This procedure allowed the testosterone ligand to explore rotational degrees of freedom while preserving the expected catalytic distance between the active oxygen and the activated carbon. The duration of the constant pressure simulation was 2 ns, at 300 K and with the time step of 1 fs. Additionally, a 1 ns simulation was performed, with no distance constraints on the ligand, and following the same protocol as described above.

The coordinate files for the 2A4 and 2A4/T models are available from the authors on request (lee_pedersen@unc.edu).

Helix analysis

The distortion of the I helix was analyzed with the secondary structure analysis program DSSP (Kabsch and Sander; 1983), by methods similar to those suggested in a study of protein α-helices (Kumar and Bansal, 1998) and from an algorithmic study of helices (Christopher et al., 1996). The central idea is that by defining a bisector of the angle formed by three consecutive α-carbons at two nearby locations, a local helix direction vector can be found from the cross product of the bisector vectors. Repeating this procedure at another location on the helix defines the helix direction at that location. The dot product of the two local helix direction vectors then provides the angle between the two directions. We chose eight sets of four C_αs at each end of the I helix to compute an average angle for the I helix.

RESULTS AND DISCUSSION

General evaluation of the model

The stability of the protein during dynamics was ascertained by calculating the RMSD of the backbone atoms during the course of the 3 ns simulation (Fig. 2). The initial energy-minimized CYP2A4 model served as a reference. As expected for a homology model with only 54% sequence identity (215 residues mutated), the structure evolved significantly from its starting configuration. However, the conserved core of helices I, J, L and the Cys pocket (residues 433-437) was expected to remain more stable than the rest of

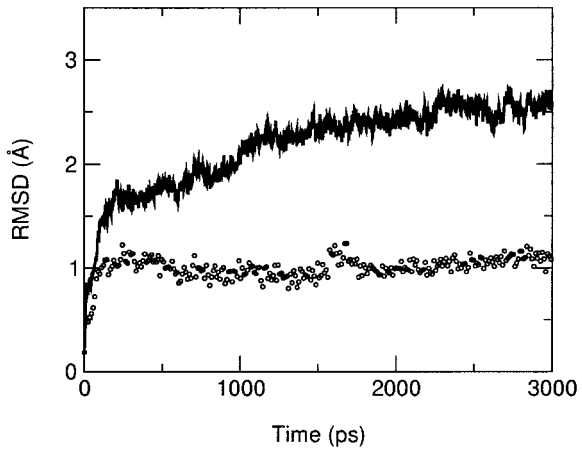


FIGURE 2 Root mean square deviations of backbone atoms of CYP2A4. The RMSD of all backbone atoms, including loops, converged at 2.5 Å (*solid*), whereas the RMSD of the conserved core (helices I, J, and L and the Cys-pocket) converged at 1 Å (*small circles*).

the protein throughout the simulation. The RMSD of the amino acid residues comprising the core of the enzyme was also evaluated. Whereas the RMSD of all backbone atoms leveled off at 2.5 Å, the core residues exhibit a RMSD of only 1 Å. Additionally, these core structural elements equilibrated much earlier in the simulation than the rest of the protein. Therefore, the enzyme core appears to be better structurally conserved than the remainder of the residues. This observation is in agreement with previous predictions (Hasemann et al., 1995).

The RMSD of individual helices was measured with the core elements aligned, over the course of the 3 ns simulation (results not shown). The largest difference was detected in helices A (average RMSD = 4.4 Å), B (average RMSD = 3.5 Å), G (average RMSD = 2.8 Å), and F (average RMSD = 1.8 Å). The interior helices are expected to have a lower deviation among various CYP450s than the external helices (Hasemann et al., 1995). Additionally, for the CYP2C5 x-ray crystal structure, the F helix and the N-terminus of the G helix form crystal contacts with a symmetry-related molecule (Williams et al., 2000a). The loss of these contacts upon solvation may account for an increased motion of the F helix and the N-terminus.

To validate the CYP2A4 model, it was compared not only to CYP2C5, but also to the bacterial $CYP450_{bm3}$, which is the bacterial enzyme nearest in three-dimensional structure to the mammalian CYP450 enzymes (Ravichandran et al., 1993). Williams et al., reported small deviations in the core regions of $CYP450_{bm3}$ and CYP2C5. We performed an alignment of the conserved core elements of the $CYP450_{bm3}$ and CYP2A4. The resulting RMSD were compared to the alignment of $CYP450_{bm3}$ with CYP2C5 (Table 2). The CYP2A4/$CYP450_{bm3}$ deviations are slightly larger than the corresponding CYP2C5/$CYP450_{bm3}$ deviations, owing perhaps to the expected increased mobility of the solvated CYP2A4 structure.

TABLE 2 Root mean square deviation (Å) of the enzymatic core (helices I, J, and L) among CYP2A4, CYP2C5, and $CYP450_{bm3}$, the bacterial enzyme most closely resembling mammalian CYP450s

$CYP450_{bm3}$ helix	RMSD with CYP2A4, Å	RMSD with CYP2C5, Å
I	0.79	0.73
L	0.84	0.53
J	0.40	0.53
I, J, L	1.50	1.05

A crucial feature for the proper functioning of CYP450s is the positioning of the heme group in the active site. Salt bridges between the active pocket residues and the propionate groups of the heme serve to accommodate these negatively charged groups in the largely hydrophobic environment of the pocket. Hydrogen bonding to the propionate groups of the heme was examined during the final 1 ns of the 3 ns simulation. Four amino acids (Arg-128, Arg-372, Ser-433, and Arg-437) maintain hydrogen bonding contacts to the propionate oxygens of the heme (Fig. 3). These contacts are conserved at corresponding positions in CYP2C5 (Williams et al., 2000a). A fifth hydrogen bond found in CYP2C5, between the propionate side chain of the D ring and the side chain of Trp-120, is not conserved in the CYP2A4 model. The residue at the corresponding position in CYP2A4 is Ala.

Another feature of the CYP2A4 structure that is conserved in CYP2C5, $CYP450_{cam}$, and $CYP450_{bm3}$ is the hydrogen bond between the side chain of Thr-305 and the carbonyl

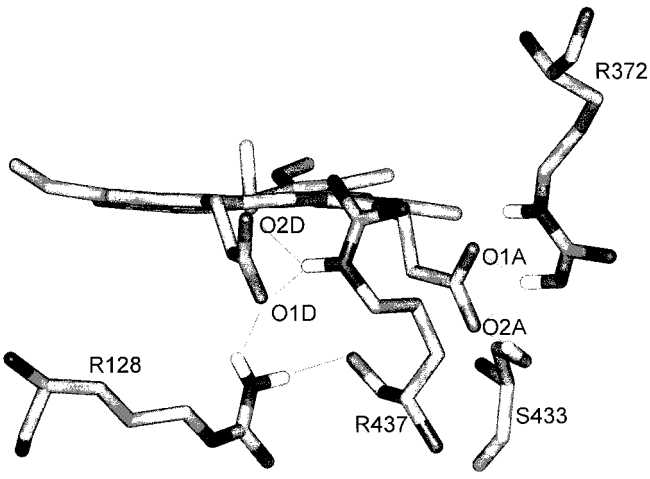


FIGURE 3 Hydrogen bonding contacts between protein residues and propionate oxygens of the heme. Arg-128, Arg-372, and Arg-437 form salt bridges to the propionate groups via main-chain and side-chain hydrogen atoms, and the side chain of Ser-433 forms a hydrogen bond to the propionate oxygen. Only hydrogen atoms involved in salt bridge/hydrogen bonding are shown.

oxygen of Ala-301. Thr-305 has been implicated in the delivery of a proton from the exterior of the protein to the reduced oxygen intermediates. This hypothesis was suggested by the bulge of the central I helix in the region containing the conserved Thr, observed in all three bacterial CYP450s (at position 252 in $CYP450_{cam}$). Along with this helical disruption, a hydrogen bond appeared invariably between the Thr and a residue four positions away, in the direction of the N-terminus to satisfy the lost hydrogen bonds in the helical deformation. In particular, a recent study of the $CYP450_{cam}$ active site showed that the mutation Thr-252-Ile decreased the monooxygenase activity 10-fold (Hishiki et al., 2000). The bulge of the central I helix is also observed in the CYP2A4 model, with Thr-305 located at the end of the resulting bulge. This helical distortion is, however, not evident in the CYP2C5 x-ray crystal structure. It has been proposed that the I helix bulge forms to accommodate the dioxygen bound to the heme iron. In the CYP2A4 model, only one oxygen atom is present. However, its presence in close proximity to, and the resulting interactions with, the surrounding residues of the I helix may account for the development of the I helical distortion in the CYP2A4 model. The possible functional significance of the I helical bulge will be revisited in a discussion on the role of water in the substrate binding cavity.

The deviation of C_α atoms of the solvent-equilibrated CYP2A4 from their original positions in the x-ray crystal CYP2C5 was used as a measure of structural difference between these two enzymes (Fig. 4). The regions of CYP2A4 that deviate most significantly from the x-ray crystal structure were identified. The backbone atoms of the conserved core of CYP2A4 and CYP2C5 were aligned for this comparison. As expected, the most mobile regions were the residues in the loops, and especially those comprising a turn. The N-terminus of the protein was particularly mobile, due to the fact that residues of CYP2C5 corresponding to residues 42-64 in CYP2A4 were involved in a crystal contact. A significant deviation was observed for the C-terminus of the A helix, the random coil containing SRS1, the N-terminus of the G helix and the F helix (also involved in a crystal contact (Williams et al., 2000a)), the FG loop, and several fragments of interhelical loops (120-122, 141-142, 337-338, 381-384, 401-408, 419-421, and 465-474). The residues in the Cys-pocket of CYP2A4 (residues 433-437) are located ~2.2 Å from the x-ray crystal structure positions of CYP2C5. This displacement is due in part to the mutation of Val-433 (CYP2C5) to a Phe-440 in CYP2A4. The mutation to a bulkier, more hydrophobic side chain, although removed by several residue positions from the region of greatest deviation, may have necessitated some backbone rearrangement that propagated to the neighboring residues. An alignment of several CYP450 family 2 proteins (Gotoh, 1992) shows that CYP2A4/2A5 enzymes are unique in having a Phe at position 440. In other enzymes, a residue with a smaller side chain (a Leu or Val) occupies this position. Therefore the subsequent rearrangement of the Cys-pocket may be a unique aspect of the CYP2A4/2A5 geometry. The secondary structure evaluation program DSSP was used to ascertain that the secondary structure elements (specifically the helices) were preserved in the 3 ns model. PROCHECK was used to evaluate the overall quality of the 3 ns structure. The overall G-factor score was −0.2, which falls within the range of expected values. The corresponding score for the x-ray crystal structure CYP2C5 was slightly better, at 0.01 (values of under −0.5 are considered poor) (Laskowski et al., 1993).

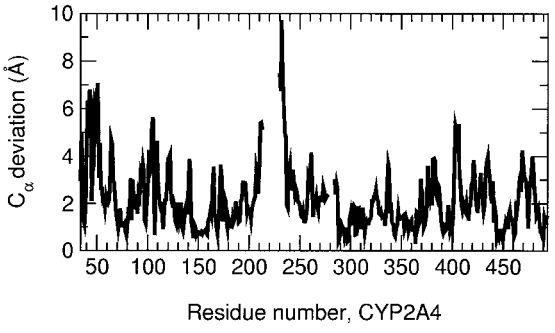


FIGURE 4 C_α deviation between x-ray crystal CYP2C5 and solvent equilibrated CYP2A4. The deviation of the C_α positions was calculated after backbone atoms of the core residues of CYP2C5 and CYP2A4 (snapshot after 3 ns of dynamics) were superimposed. The gaps in the plot correspond to the regions missing in the x-ray crystal structure: the unresolved segment of the FG loop (residues 216-226) and the gap resulting from the alignment with CYP2A4 (residues 280-282).

Comparison of CYP2A4 and CYP2C5

Once the overall stability of the CYP2A4 model in solution was ensured, the differences between the active sites of CYP2A4 and CYP2C5 were assessed, initially in the absence of the ligand. The substrate binding site in CYP450s is hemispherical, with the heme forming the base and the I helix one side. In the course of the simulation, the heme in the model CYP2A4 has moved by ~1.5 Å from its original position in the CYP2C5 crystal structure, in a direction away from the center of the I helix. Once the structure is equilibrated in solution, the amplitude of this displacement actually fluctuates between 1 Å and 2 Å throughout the simulation. As described above, this displacement is not sufficient to break the hydrogen bonding contacts that anchor the heme in the active site. However, it induces some rearrangement in the surrounding amino acid positions. We focused specifically on the four residues that play a role in substrate binding and activation: Leu-209, located at the C-terminus of the F helix; Ala-117, in the BC loop; Val-481 in the turn between beta strands; and Leu-365 in the region between the K helix and a β-sheet. Additionally, the position of Thr-305, located in the bulge of the I helix, was considered, because of that residue's possible role in proton

transfer. To assess the positions of these amino acids relative to the heme, the distances between their C_α and Fe of the heme were compared to the corresponding distances in CYP2C5 (Table 3). We find that the largest deviation (less than 2 Å) is consistent with the displacement of the heme in CYP2A4 relative to its position in CYP2C5.

Comparison of the substrate-free and substrate-bound forms of CYP2A4

The solvent equilibrated substrate-free CYP2A4 model may be compared to the equilibrated model of CYP2A4 with the substrate (testosterone) bound (CYP2A4/T). The RMSD of all backbone atoms between the two final structures was 2.7 Å. The core helices I, J, and L were found to deviate with 1.4 Å RMSD between the two model structures. The cause of the deviations is apparent when the binding pocket is examined, as the substrate binding induces some backbone and side-chain rearrangement.

The residues within a contact distance (4 Å) from any of the testosterone atoms were considered. Snapshots of the dynamical trajectory were examined at each 50 ps during a 500-ps time period. The following residues were consistently identified to be within contact distance of the ligand: Arg-101, Glu-103, Phe-107, Ala-117, Phe-118, Leu-209, Thr-212, Ala-213, Asn-297, Phe-300, Ala-301, Glu-304, Thr-305, Leu-370, Ala-371, Arg-437, Phe-480, and Val-481. The backbone atoms of these residues were then used to superimpose CYP2A4 and CYP2A4/T (RMSD = 1.7 Å). When this superimposition was performed, the heme units in the two molecules were overlayed almost exactly. This configuration therefore gives an indication of the motion of the ligand binding residues with respect to the heme and the ligand. The C_α deviations of those amino acids in contact with the ligand were measured to determine how much displacement was caused by the presence of the ligand. The two active sites are compared in Fig. 5. Several side chains found within 4 Å of testosterone that experienced significant change are shown. Additionally, the backbone of residues 101-103 shifts nearer to the ligand. The C_α deviations for these three residues from their positions in the ligand-free form were, respectively, 1.5 Å, 1.9 Å, and 2.6 Å. Glu-103 has moved toward the interior of the cavity (although its negatively charged side chain still points away from the cavity). On the contrary, the hydrophobic Phe-107 side chain undergoes a rotation toward the ligand (the C_α deviation of Phe-107 was 1.0 Å). Residues 117 and 118, located in the BC loop, are especially interesting because Ala-117 influences the catalytic activity of the CYP2A4 enzyme (Table 1, Negishi et al., 1996c). The backbone atoms of these two residues deviated from their position in CYP2A4 by 1.8 Å, moving nearer the substrate in the ligand-bound form. This motion was most likely caused by steric repulsion, because the bulky side chain of Phe-118 was originally situated close to the interior of the active site, partially blocking the substrate binding cavity. The addition of the ligand and the subsequent energy minimization led to a rotation of that side chain away from the cavity interior.

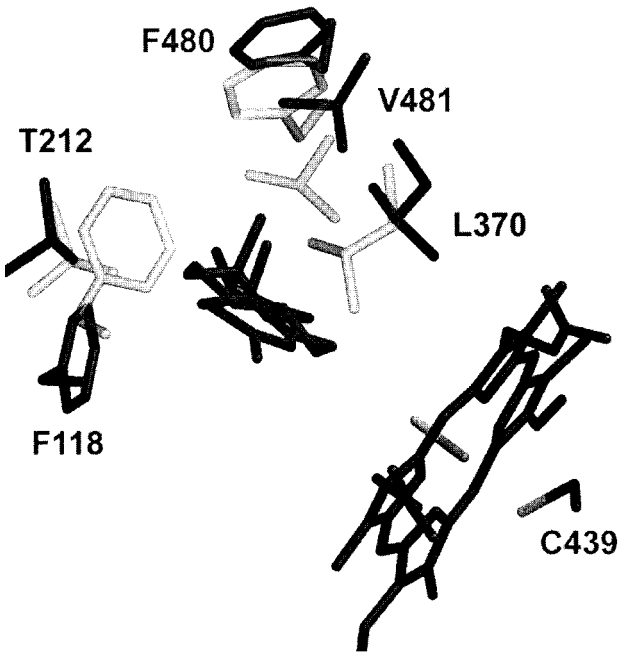


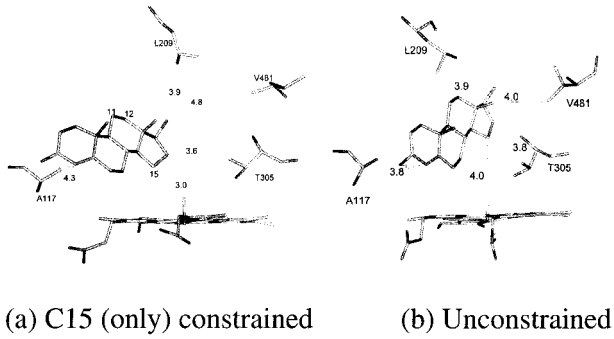
FIGURE 5 The active site of solvent equilibrated CYP2A4 (*light*) and CYP2A4/T (*dark*). Only the residues lining the active site (within 4 Å of the docked testosterone in CYP2A4/T) that showed significant change are shown. The hemes of the two structures were overlaid; only the CYP2A4/T heme is shown.

During the course of the simulation, Ala-117, Leu-209, Thr-305, and Val-481 directly interact with the ligand (Fig. 6). The binding pocket for the simulations with a single constraint (Fig. 6 *A*) and without constraints (Fig. 6 *B*) are shown. The side chain of Ala-117 is near atoms C4-C6, including the keto group of testosterone. Leu-209 is near the hydroxyl group of the ligand. The position of the Leu-209 side chain fluctuates in the course of the simulation; however, for a significant period its side chain is within 4 Å of C12 of the ligand. The side chain methyl group of Thr-305 interacts with testosterone via a hydrophobic contact with C16, whereas Val-481 is positioned so as to

TABLE 3 Location of functionally critical amino acids in the active sites of the solvent equilibrated model CYP2A4, CYP2A4/T (unconstrained ligand), and the x-ray crystal structure of CYP2C5 (Williams et al., 2000a)

Residue	Distance from Fe in CYP2C5, Å	Distance from Fe in CYP2A4, Å	Distance from Fe in CYP2A4/T, Å
Ala-117	9.2	10.7	10.7
Leu-209	13.5	13.4	17.3
Thr-305	6.8	8.6	7.2
Leu-365	10.5	10.6	11.9
Val-481	11.3	10.2	11.7

The distances in Å between C_α and the heme iron are listed.



(a) C15 (only) constrained (b) Unconstrained

FIGURE 6 The location of residues Ala-117, Leu-209, Thr-305, and Val-481 with respect to the ligand in (*a*), the simulation with the distance between C15 of testosterone and the ferryl oxygen constrained, and (*b*), the simulation with no distance constraints. The distances, in angstroms, between side-chain carbons of these residues and the heavy atoms of testosterone, are indicated. The keto group of testosterone points toward Ala-117. C11, C12, and C15 are labeled in (*a*).

make a hydrophobic contact with one of the testosterone methyl groups (C18). Leu-365 is the only catalytic amino acid with no direct interaction with the ligand. However, its neighboring residue, Ile-366, occasionally moves within 4 Å of the ligand in the course of the simulation, suggesting the possibility of an indirect effect of Leu-365 on the binding of the ligand. These hydrophobic contacts are also observed in the simulation with no distance constraints on the ligand (Fig. 6 *B*). As the ligand moves away from the heme (the equilibrium distance between C15 and the ferryl oxygen is now 4.0 Å), the distance between Leu-209 and C12 remains unchanged whereas that of Val-481 and C18 decreases. Although the constrained simulation preserves the favorable orientation between the ferryl oxygen and C15 of testosterone, the position of the ligand obtained in the unconstrained simulation agrees better with previous experimental results. Iwasaki et al. (1993), showed that the combination of the residue 209 and the substituent at position C11 of the ligand alter substrate binding specificity at the active sites of CYP2A4 and CYP2A5. Their results indicate that the interaction of 11-β-hydroxysteroids inhibits testosterone 15-α-hydroxylase activity when the residue at position 209 is Asn, although 11-deoxysteroids inhibits testosterone 15-α-hydroxylase activities regardless of the nature of the residue at position 209. In the unconstrained simulation (Fig. 6 *B*), the side chain of Leu-209 is located nearer to the C11 position than in the constrained simulation (Fig. 6 *A*), and can potentially interact with an 11-β-hydroxy substituent on the ligand. Additionally, the location of the ligand farther from the heme, as the unconstrained simulation indicates, may reflect the more favorable configuration of the substrate upon binding to the enzyme, but before the hydroxylation reaction. A shift of the substrate position toward the heme may occur as a consequence of the electron transfer to CYP2A4 from its redox partner, NADPH-binding reductase, which CYP2A4 requires for catalysis. A movement of the bound substrate 6 Å in the direction of the heme, which positions it for the hydroxylation reaction, was observed in $CYP450_{bm3}$ complexed with reductase as a consequence of the first electron transfer (Modi et al., 1996).

In the equilibrated constrained ligand simulation, the ligand is hydrogen bonded at the keto end (to the backbone oxygen of Phe-118). In the equilibrated unconstrained ligand simulation, however, this hydrogen bond is broken. Instead, the keto group becomes hydrogen bonded to a water molecule that itself interacts with a heme propionate group and with the side chains of R117 and R437. The hydroxyl end of the ligand, interacts hydrophobically with L209, A213, and E304 (on the I helix). Overall, testosterone is anchored in the active site/binding pocket by hydrophobic interactions.

A measure of the degree of rearrangement of the binding pocket/active site on ligand binding is seen (Table 3) from the changes in the distances of the α-carbon of key residues from the Fe atom. The presence of ligand does not affect A117 but L209, V365, and V481 are at larger distances. T305 actually moves nearer (by 1.4 Å) to the Fe.

The FG loop

The loop connecting the helices F and G has been attributed a key role in controlling the substrate access into the active site of CYP450s. In bacterial CYP450s, the FG region has very high B-factors compared to the rest of the structure, suggesting that the flexibility of this loop allows it to close and open the entrance to the substrate binding site (Hasemann et al., 1995). In CYP2C5, a 10-residue portion of the loop had ambiguous electron density, but was predicted, due to the crystal lattice constraints, to fill the space between the N-terminal β-sheets and F and G helices (Williams et al., 2000a). In the solvent-equilibrated CYP2A4, in both substrate-free and substrate-bound forms, the FG loop assumes the predicted conformation, fitting between the ends of the F and G helices and the system of β-sheets. The mobility of the FG loop is restricted by hydrogen bonding contacts to the residues of the β-sheet. In the CYP2A4/T (both constrained and unconstrained models), Gln218 makes a strong hydrogen bond to Ala-371 and Glu-390 (Fig. 7). In the substrate-free CYP2A4, Gln-218 instead hydrogen bonds to Arg-372. These hydrogen bonds anchor the FG loop, and as a result, it does not exhibit extensive motion in the course of the simulation. The fluctuations of the FG loop were evaluated by calculating an average C_α deviation for the residues of this loop while helices F, G, and the conserved core were aligned. A comparison of a snapshot early in the simulation with the final snapshot of CYP2A4 revealed an average C_α deviation of 1.56 Å (Fig. 8 *A*). The average C_α deviation of the FG loop of solvent equilibrated CYP2A4/T from the early snapshot of CYP2A4 was 2.75 Å (Fig. 8 *B*). A recent study of $CYP450_{cam}$ demonstrated that the opening of the substrate access channel in this enzyme is

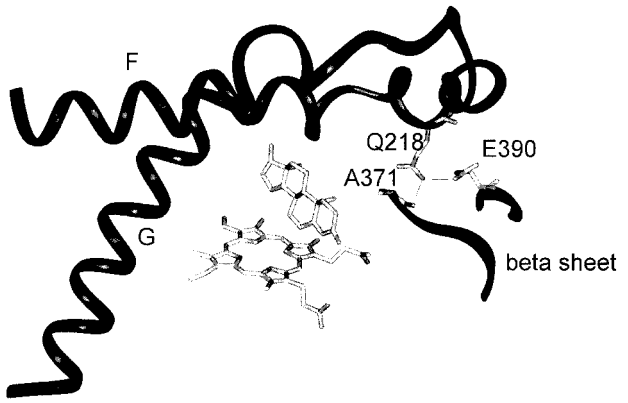


FIGURE 7 Position of the FG loop in solvent equilibrated CYP2A4/T. The FG loop is anchored between the ends of helices F and G and the β-sheet by strong hydrogen bonds formed by residues Gln-218, Ala-371, and Glu-390.

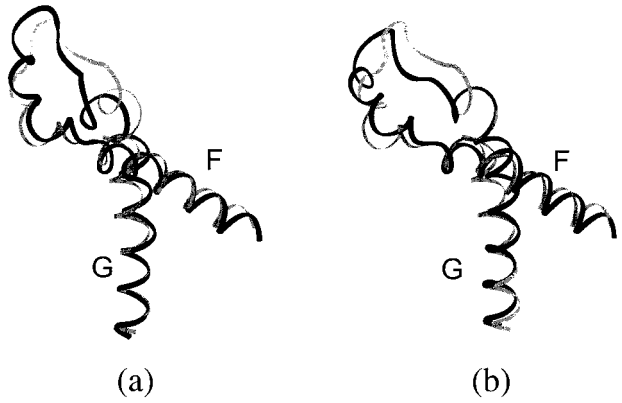


FIGURE 8 Motion of the FG loop. The coordinates of (*a*), the final CYP2A4 model (*dark*), and (*b*), the final CYP2A4/T model (*dark*), were superimposed onto the coordinates of CYP2A4 (*light*) early in the simulation. Helices F, G and the conserved core were used for the alignment. The FG loops exhibit moderate fluctuations in both cases.

facilitated by the motion of the F and G helices, which retracts the FG loop from the β-sheet domain, and induces some conformational changes in the I helix (Dunn et al., 2001). The configuration of the FG loop in the model CYP2A4 and CYP2A4/T resembles the closed state of the $CYP450_{cam}$. The corresponding open state, which would have been signaled by the FG loop retraction away from the β-sheet domain, was not observed in our simulation. Partly this is due to the induced fit methodology employed in docking the substrate. Testosterone was placed directly in the active site, and the surrounding protein residues were relaxed to accommodate the ligand. A dynamical simulation with the substrate guided into the interior of the protein through an entrance channel, similar to those performed by Lüdemann et al., may be necessary to induce the transition from the closed to the open conformations of CYP2A4 (Lüdemann et al., 2000a,b).

Water in the interior of solvent equilibrated CYP2A4

Because water has been implicated in the mechanism of activation of substrates by P450s, it is imperative to examine the water molecules in the active site/binding pocket region for our simulations. The x-ray crystal structure (Williams et al., 2000a) was not of sufficient resolution (3.0 Å) to ''see'' water molecules that were likely present, although some residual density in the region was thought to be due to an unidentified ligand.

At the end of the 3 ns simulation of CYP2A4 in solution, four waters were detected in the interior of the enzyme and less than 10 Å away from the oxygen of the heme. The water diffusion from bulk to the final location in the interior of CYP2A5 was tracked by mapping consecutive snapshots of these four water molecules in the course of the simulation. The resulting paths identify at least three entrance channels for water. As the water molecules are much smaller than any CYP450 substrate, the water entrance pathway may not be indicative of a substrate access channel. Water molecule 1 enters the enzyme between helix F and residues 476-486, which form the C-terminal segment after helix L. Two water molecules (2 and 3) are located near one another, in the proximity of the C ring of the heme. Water molecule 2 forms hydrogen bonds to the backbone oxygen of Val-178 of the E helix, and to the side chain of Glu-448 of the L helix. The neighboring water molecule 3 also hydrogen bonds to the side chain of Glu-448, and additionally, the two waters are sufficiently close to bond to each other. Water molecule 2 and water molecule 3 follow the same path, entering between the L helix and the E helix. Finally, water molecule 4 resides close to the propionate groups of the heme at the end of the 3 ns simulation, where it forms hydrogen bonds with the backbone oxygen of Leu-370 and the side chain nitrogen of Arg-437. Water molecule 4 appears to enter the interior of the protein between helix C and the BC loop. Only one of the interior water molecule (water molecule 1) enters the active site cavity. It is located 6.5 Å from the heme oxygen, near the bulge of the I helix, where it forms hydrogen bonds to the backbone nitrogen atoms of Thr-305 and Glu-304 of the I helix, and the backbone oxygen atoms of Ala-235 and Phe-234 of the H helix (Fig. 9). The importance of a water molecule in this position has been described for the bacterial CYP450 enzymes, where the conserved Thr, the neighboring acidic residue, and the water in the groove of the I helix are believed to comprise the essential elements of a proton transfer network (Hasemann et al., 1995). Specifically, in $CYP450_{bm3}$, the proton was proposed to be transferred to the heme from the side chain of the conserved Thr, which was protonated by a water molecule. The acidic residue neighboring the conserved Thr could in turn protonate the water molecule (Ravichandran et al., 1993). In $CYP450_{cam}$, the side chain of the conserved Thr is oriented so as to make

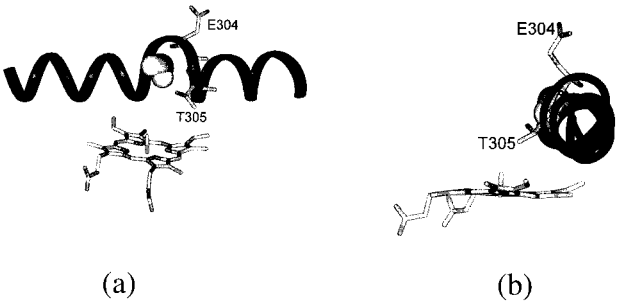


FIGURE 9 Water in the active site of the solvent equilibrated CYP2A4. Two views shown illustrate the proximity of the water molecule (rendered) to the I helix, and the residues in the I helical bulge, Thr-305 and Glu-304.

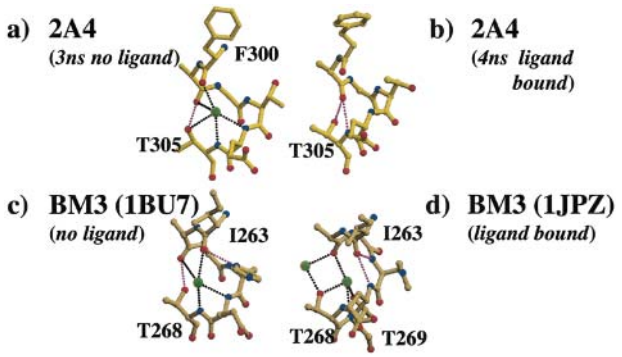


FIGURE 10 Local structure near the I helix bulge for a) 2A4 (no ligand), b) 2A4/T (ligand unconstrained), c) BM3 (1BU7, no ligand), and d) BM3 (1JPZ, ligand).

a hydrogen bond to the sixth ligand of the heme. This, however, is not the case in $CYP450_{bm3}$ or $CYP450_{terp}$, where instead a carbonyl oxygen is located closest to the sixth axial ligand (Hasemann et al., 1995). In the solvent equilibrated CYP2A4, the water reaches its position near the I helical bulge in the latter third of the 3 ns simulation.

It is clearly of interest to relate, if possible, the degree of the bulge in the I helix to the presence or absence of ligand and/or to the water molecule with binds near the bulge. Such a possible relationship was suggested by Haines et al. 2001, who noted in their x-ray crystal work the difference in water coordination in a ligand-bound BM3 (pdb = 1JPZ) and the corresponding ligand-free BM3 (pdb = 1BU7). Fig. 10 (*a* and *b*) show the disposition of the T305 locality of the I helix for 2A4 and 2A4/T (unconstrained), respectively. There is no water bound in the ligand unconstrained case. Fig. 10 (*c* and *d*) show the same view for BM3 (no ligand, 1BU7) and BM3 (ligand bound, 1JPZ). Remarkably, the nature of the H_2O binding in BM3 (no ligand, 1BU7) and the water molecule in 2A4 (no ligand) are similar. The evolution of this water molecule path in 2A4 (no ligand) from initially outside the protein shell to a binding pattern recognizable from that of the BM3 (1BU7) suggests that the overall force field is doing very well in representing the local dynamics of water and its interactions with proteins. Different ligands are present in 2A4/T (ligand bound (*b*)) and in BM3 (ligand bound, 1JPZ(*d*)). This difference perhaps accounts for the unlike structures present around the T305 site: BM3 (1JPZ) has two molecules of water bound in a form different than seen in Fig. 10 *a* or *c* and 2A4/T (ligand bound(*b*)) has no water bound.

Table 4 summarizes structural parameters for the I helix for the 2C5 x-ray crystal structure, for the simulation structures (bound and unbound) and for three BM3 crystal structures. The helix angle (see Methods), the classification of nonhelical residues by the standard DSSP program (Kabsch and Sander, 1983), the Fe-T305 distance, and the distance between the ends of the I helix are computed. The helix angle for the ligand bound cases (2A4/T, 1JPZ) is slightly larger than the ligand-free (2A4, 1BU7 and 1HP D) cases. An opposing correlation was reported by Haines et al. 2001, for the BM3 case (bound 1JPZ versus unbound 1BU7)—the difference may depend on what sections of the helix are used to compute the angle. We used residues at either end of the helix; if residues on either side (and near) the T305 site are used, angles near those reported by Haines et al. 2001, are found. Our protocol yields a helix angle of 8.1° for the 2C5 x-ray crystal structure. Unfortunately, we do not know if water is bound at the bulge site or what ligand (if any) is bound. We also computed these same properties for a set of four P450cam structures (pdb: 5CP4, 5CPP, 1K2O, and 1PHC) (results not shown). Three of these structures have water bound at the bulge site and have ligand bound, although one (1PHC) has neither. All have helix angles in the

TABLE 4 Comparison of I helix parameters from structural analysis of bacterial, mammalian, and simulation pdb files

Structure	Ligand	Water Complex	Helix Angle*	DSSP†	FeT(N)‡	C_α-C_α§, Nres¶
Mammalian and simulation						
1DT6‖	yes	not resolved	8.1 ± 2.7	0	9.43	46.6, 31
2A4(sim)	no	yes	13.9 ± 6.4	2	9.10	45.4, 30
2A4/T(sim)	yes	(no)**	22.5 ± 13.1	0	7.71	45.0, 30
Bacterial-BM3						
1BU7##	no	yes	15.3 ± 4.5	0	7.76	45.8, 31
1JPZ§§	yes	yes	17.3 ± 4.6	2	7.64	47.1, 31
1HPD¶¶	no	yes	15.4 ± 4.2	0	7.76	47.0, 31

*See Methods: averaged over 11 C_α coordinates from each end of helix, in degrees.
†Kabsch and Sander, 1983: number of residues with nonhelix classification by DSSP in I helix.
‡Heme Fe-conserved T backbone nitrogen distance (in Å).
§Distance between C_αs at end of I helix.
¶Number of residues in I helix defined by DSSP.
‖Williams et al., 2000a: ligand not seen due to resolution but may be present.
**Water not seen in I helix bulge region after 4 ns total simulation.
##Sevrioukova et al., 1999.
§§Haines et al., 2001.
¶¶Ravichandran et al., 1993.

8-11° range; all are identified by DSSP as having a nonhelical segment near the catalytic threonine. Because the calculation of the helix angle is so dependent on where the defining stretches are chosen, we are reluctant to associate any property, including water complex structure or even function, to its magnitude.

Thus, although our simulations are provocative in that they allow us to obtain a reasonable glimpse of the pre-transition state structure, they are obviously not the last word. Ultimately, quantum mechanical calculations that accommodate the key components of the active site will be required to unravel the sequence of events that take place in P450 activation.

CONCLUSION

Mammalian cytochromes CYP450 present a challenge for molecular modelers due to their great diversity and the relative scarcity of experimental structural data. Though they share a similar fold, the details of each structure, especially the active site, are unique and contribute to the great versatility of these enzymes. Previous modeling work has been based on bacterial CYP450s, the sequences of which are quite different from those of the mammalian CYP450s. In this work, a model of mouse CYP2A4 based on an x-ray crystal structure of another mammalian enzyme (CYP2C5) is presented. The structure is modeled in solution on a multiple-nanosecond time scale. Although the initial structure of CYP2A4 requires extensive homology modeling, the core of the enzyme containing helices I, J, and L remains well-conserved throughout the simulation. The heme in the solvent equilibrated CYP2A4 is located in such a way that the salt bridges and hydrogen bonds between the heme and the protein, observed in CYP2C5, are maintained in our model. The docking of testosterone, a primary substrate of CYP2A4, induced moderate rearrangements in the active site. Mainly hydrophobic interactions guide the binding of the ligand. A remarkable result is the natural evolution of a water complex near the conserved I helix Thr. A similar complex had been identified previously in a BM3 structure. The local structure (including water) near the center of the I helix and between it and the catalytic Fe depends on the presence or absence of ligand.

The authors acknowledge the availability of the resources of the North Carolina Supercomputing Center; an IRTA position (A.G.) and access to resources (L.G.P.) at the National Institute of Environmental Health Sciences, Research Triangle Park, NC; National Institutes of Health support through grant HL-06350 (L.G.P.) and GM31001 (E.F.J.); National Science Foundation grant ITR/AP/DMR-0121361 (L.G.P.); and D.L. Lewis for helpful discussions.

REFERENCES

Bayly, C. I., P. Cieplak, W. D. Cornell, and P. A. Kollman. 1993. A well-behaved electrostatic potential based method using charge restraints for determining atom-centered charges: the RESP model. *J. Phys. Chem.* 97:10269–10280.

Case, D. A., D. A. Pearlman, J. W. Caldwell, T. E. Cheatham, W. S. Ross, C. Simmerling, T. A. Darden, K. Mertz, R. Stanton, A. Cheng, J. Vincent, M. Crowley, D. M. Ferguson, R. Radmer, G. L. Seibel, U. C. Singh, P. Weiner, and P. A. Kollman. 1999. AMBER (5.0). University of California, San Francisco, California.

Christopher, J. A., R. Swanson, and T. O. Baldwin. 1996. Algorithms for finding the axis of a helix. *Computers Chem.* 20:339–345.

Dai, R., M. R. Pincus, and F. K. Friedman. 2000. Molecular modeling of mammalian cytochrome P450s. *Cell. Mol. Life Sci.* 57:487–499.

Dai, R., S. Zhai, X. Wei, M. R. Pincus, R. E. Vestal, and F. K. Friedman. 1998. Inhibition of human cytochrome P450 1A2 by flavones: a molecular modeling study. *J. Protein Chem.* 17:643–650.

Darden, T. A., D. York, and L. G. Pedersen. 1993. Particle mesh Ewald: An N*log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.

Dunn, A. R., I. J. Dmochowski, A. M. Bilwes, H. B. Gray, and B. R. Crane. 2001. Probing the open state of cytochrome P450cam with ruthenium-like substrates. *Proc. Natl. Acad. Sci. USA.* 98:12420–12425.

Essman, U., L. Perera, M. L. Berkowitz, T. A. Darden, H. Lee, and L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8592.

Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. Robb, J. Cheeseman, V. Zakrzewski, J. Montgomery, R. Stratmann, J. C. Burant, B. Dapprich, J. M. Millam, A. D. Daniels, K. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennuci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Lui, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle, and J. A. Pople. 1994. Gaussian 98 (Revision A.1).Gaussian, Pittsburgh.

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.

Guengerich, P. F. 2001. Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem. Res. Tox.* 14:611–650.

Haines, D. C., D. R. Tomchick, M. Machius, and J. A. Peterson. 2001. Pivotal role of water in the mechanism of P450 BM-3. *Biochemistry.* 40:13456–13465.

Harris, D. L., and G. Loew. 1995. Prediction of regiospecific hydroxylation of camphor analogs by cytochrome $P450_{cam}$. *J. Am. Chem. Soc.* 117:2738–2746.

Hasemann, C. A., R. G. Kurumbail, S. S. Boddupalli, J. A. Peterson, and J. Deisenhofer. 1995. Structure and function of cytochrome P450: a comparative analysis of three crystal structures. *Structure.* 15:41–62.

Hishiki, T., H. Shimada, S. Nagano, T. Egawa, Y. Kanamori, R. Makino, S. Park, S. Adachi, Y. Shiro, and Y. Ishimura. 2000. X-ray crystal structure and catalytic properties of Thr252Ile mutant of cytochrome P450cam: roles of Thr252 and water in the active center. *J. Biochem.* 128:965–974.

Iwasaki, M., R. L. Lindberg, R. O. Juvonen, and M. Negishi. 1993a. Site-directed mutagenesis of mouse steroid 7α-α-hydroxylase (cytochrome P-$450_{7\alpha}$): role of residue-209 in determining steroid-cytochrome P-450 interaction. *Biochem. J.* 291:569–573.

Iwasaki, M., T. A. Darden, L. G. Pedersen, D. G. Davis, R. O. Juvonen, T. Sueyoshi, and M. Negishi. 1993b. Engineering mouse P450coh to a novel corticosterone 15α-α-hydroxylase and modeling steroid-binding orientation in the substrate pocket. *J. Biol. Chem.* 268:759–762.

Jorgensen, W. L., J. Chandrasekhar, J. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure. *Biopolym.* 22:2577–2637.

Kobliakov, V., L. Kulikova, D. Samoilov, and M. A. Lang. 1993. High expression of cytochrome P450 2a-5 (coumarin 7-hydroxylase) in mouse hepatomas. *Mol. Carcinog.* 7:276–280.

Kumar, S., and M. Bansal. 1998. Geometric and sequence characteristics of alpha-helices in globular proteins. *Biophys. J.* 75:1935–1944.

Laskowski, R. A., M. W. MacArthur, D. S. Moss, and J. M. Thornton. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26:283–291.

Lewis, D. F. V., and B. G. Lake. 1995. Molecular modeling of members of the P4502A subfamily: application to studies of enzyme specificity. *Xenobiotica.* 25:585–598.

Lewis, D. F. V., B. G. Lake, M. Dickins, P. J. Eddershaw, M. H. Tarbit, and P. S. Goldfarb. 1999. Molecular modelling of CYP2B6, the human CYP2B isoform, by homology with the substrate-bound CYP102 crystal structure: evaluation of CYP2B6 substrate characteristics, the cytochrome b_5 binding site and comparisons with CYP2B1 and CYP2B4. *Xenobiotica.* 29:361–393.

Lindberg, R. L., and M. Negishi. 1989. Alteration of mouse cytochrome P450coh substrate specificity by mutation of a single amino-acid residue. *Nature.* 339:632–634.

Loew, G., and Y. Chang. 1999. Homology modeling and substrate binding study of human CYP4A11 enzyme. *Proteins.* 34:403–415.

Lüdemann, S. K., V. Lounnas, and R. C. Wade. 2000a. 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.

Lüdemann, S. K., V. Lounnas, and R. C. Wade. 2000b. How do substrates enter and products exit the buried active site of cytochrome P450cam? 12. Steered molecular dynamics and adiabatic mapping of substrate pathways. *J. Mol. Biol.* 303:813–830.

Modi, S., M. J. Sutcliffe, W. U. Primrose, L. Y. Lian, and G. C. Roberts. 1996. The catalytic mechanism of cytochrome P450 BM3 involves a 6 Å movement of the bound substrate on reduction. *Nat. Struct. Biol.* 3:414–417.

Negishi, M., M. Iwasaki, R. O. Juvonen, T. Sueyoshi, T. A. Darden, and L. G. Pedersen. 1996a. Structural flexibility and functional versatility of cytochrome P450 and rapid evolution. *Mutat. Res.* 350:43–50.

Negishi, M., T. Uno, T. A. Darden, T. Sueyoshi, and L. G. Pedersen. 1996b. Structural flexibility and functional versatility of mammalian P450 enzymes. *FASEB J.* 10:683–689.

Negishi, M., T. Uno, P. Honkakoski, T. Sueyoshi, T. A. Darden, and L. G. Pedersen. 1996c. The roles of individual amino acids in altering substrate specificity of the P450 2a4/2a5 enzymes. *Biochimie.* 78:685–694.

Oscarson, M., H. Gullsten, A. Rautio, M. L. Bernal, B. Sinues, M. Dahl, J. Stengard, O. Pelkonen, H. Raunio, and M. Ingelman-Sundberg. 1998. Genotyping of human cytochrome P450 2A6 (CYP2A6), a nicotine C-oxidase. *FEBS Lett.* 438:201–205.

Ravichandran, K. G., S. S. Boddupalli, C. A. Hasemann, J. A. Peterson, and J. Deisenhofer. 1993. Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. *Science.* 261:731–736.

Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23:327.

Sevrioukova, I. F., H. Li, H. Zheng, J. A. Peterson, and T. L. Poulos. 1999. Structure of a cytochrome P450-redux partner electron-transfer complex. *Proc. Nat. Acad. Sci.USA.* 96:1863–1868.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25:4876–4882.

Williams, P. A., J. Cosme, V. Sridhar, E. F. Johnson, and D. E. McRee. 2000a. Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Mol. Cell.* 5:121–131.

Williams, P. A., J. Cosme, V. Sridhar, E. F. Johnson, and D. E. McRee. 2000b. Microsomal cytochrome P450 2C5: comparison to microbial P450s and unique features. *J. Inorg. Biochem.* 81:183–190.