Structure of a Substrate Complex of Mammalian Cytochrome P450 2C5 at 2.3 Å Resolution: Evidence for Multiple Substrate Binding Modes[†,‡]

Michael R. Wester,[§] Eric F. Johnson,*[,§] Cristina Marques-Soares,[||] Patrick M. Dansette,[||] Daniel Mansuy,[||] and C. David Stout*[,⊥]

Department of Molecular and Experimental Medicine, The Scripps Research Institute, 10550 North Torrey Pines Road, MEM-255, La Jolla, California 92037, Laboratoire de Chimie et Biochimie Pharmacologiques et Toxicologiques, UMR 8601 CNRS, Université Paris V45, Rue des Saints-Pères, 75270 Paris Cedex 06, France, and Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, MB8, La Jolla, California 92037

Received December 20, 2002; Revised Manuscript Received March 25, 2003

ABSTRACT: The structure of rabbit microsomal cytochrome P450 2C5/3LVdH complexed with a substrate, 4-methyl-*N*-methyl-*N*-(2-phenyl-2*H*-pyrazol-3-yl)benzenesulfonamide (DMZ), was determined by X-ray crystallography to 2.3 Å resolution. Substrate docking studies and electron density maps indicate that DMZ binds to the enzyme in two antiparallel orientations of the long axis of the substrate. One orientation places the principal site of hydroxylation, the 4-methyl group, 4.4 Å from the heme Fe, whereas the alternate conformation positions the second, infrequent site of hydroxylation at >5.9 Å from the heme Fe. Comparison of this structure to that obtained previously for the enzyme indicates that the protein closes around the substrate and prevents open access of water from bulk solvent to the heme Fe. This reflects a ∼1.5 Å movement of the F and G helices relative to helix I. The present structure provides a complete model for the protein from residues 27−488 and defines two new helices F′ and G′. The G′ helix is likely to contribute to interactions of the enzyme with membranes. The relatively large active site, as compared to the volume occupied by the substrate, and the flexibility of the enzyme are likely to underlie the capacity of drug-metabolizing enzymes to metabolize structurally diverse substrates of different sizes.

The P450 monooxygenases that metabolize xenobiotics constitute the majority of the more than 50 P450 genes in mammalian genomes. In contrast to the P450s that evolved to catalyze specialized biosynthetic pathways, the drug metabolizing P450s exhibit broad and overlapping substrate specificities. This metabolic diversity provides an effective means of oxidizing and eliminating foreign compounds such as drugs and environmental toxins. The xenobiotic metabolizing enzymes generally exhibit distinct substrate specificity profiles and often oxidize a structurally diverse range of substrates in a regiospecific manner. In some cases, multiple products arising from the oxidation of noncontiguous sites on the same substrate are obtained, suggesting multiple modes of substrate binding to the enzyme's active site or that the mobility of substrates in the active site of the enzyme is not highly restrained. The structural basis for the metabolic diversity and selectivity exhibited by mammalian xenobiotic metabolizing P450s is poorly understood, and on the basis of current knowledge, it is difficult to predict the substrate specificities and metabolic products for these enzymes. The present study addressed whether conformational changes occur when CYP2C5[1] binds substrates.

Crystallization of a modified form of rabbit microsomal CYP2C5 enabled determination of the first structure of a mammalian, microsomal P450 by X-ray crystallography (*1*) and provides a basis for experimental characterization of enzyme substrate complexes. The modified enzyme, P450 2C5/3LVdH, was produced by substitution of a short, positively charged N-terminal sequence for the native transmembrane leader sequence and by addition of a 4-histidine tag to the C-terminus. Additional amino acid substitutions that alter the surface of the protein in the vicinity of the F helix further improved the solubility and monodispersity of the enzyme in high salt buffers (*2*). The modified enzyme was used in this study to determine the first structure of a drug metabolizing P450 with a substrate bound in the active site.

The present manuscript reports on the structure of P450 2C5/3LVdH complexed with 4-methyl-*N*-methyl-*N*-(2-phenyl-2*H*-pyrazol-3-yl)benzenesulfonamide (DMZ) and discusses the implications of the observed structural changes

[†] This investigation was supported by NIH Grants GM31001 (E.F.J.) and GM59229 (C.D.S.) and by CNRS and the French Minister of Research (D.M.).

[‡] Structural coordinates have been deposited with the Protein Data Bank under accession code 1N6B.

* To whom correspondence should be addressed. (E.F.J.) Tel.: (858) 784-7918. Fax: (858) 784-7978. E-mail: johnson@scripps.edu. (C.D.S.) Tel.: (858) 784-8738. Fax: (858) 784-2857. E-mail: dave@scripps.edu.

[§] Department of Molecular and Experimental Medicine, The Scripps Research Institute.

[||] Université Paris V45.

[⊥] Department of Molecular Biology, The Scripps Research Institute.

[1] Abbreviations: CVFF, consistent valence force field; CYP, cytochrome P450; DMSO, dimethyl sulfoxide; DMZ, 4-methyl-*N*-methyl-*N*-(2-phenyl-2*H*-pyrazol-3-yl)benzenesulfonamide; DTT, dithiothreitol; EDTA, ethylenediamine tetraacetic acid; F_o, observed structure factor; F_c, calculated structure factor; HEPES, *N*-(2-hydroxyethyl)-piperazine-*N*′-(2-ethanesulfonic acid); PDB, Protein Data Base (http://www.rcsb.org/pdb/); rms, root-mean-square; SSRL, Stanford Synchrotron Radiation Laboratory.

10.1021/bi0273922 CCC: $25.00
Published on Web 05/08/2003

with regard to the substrate selectivity of mammalian microsomal P450s. DMZ is one of several analogues of sulfaphenazole, a potent and relatively selective inhibitor of human P450 2C9, that were synthesized to examine structure/activity relationships between the human 2C enzymes. In contrast to sulfaphenazole, DMZ proved to be a relatively general inhibitor of the human 2C enzymes (*3*−*5*). Subsequent studies described in the accompanying paper (*6*) showed that DMZ is a good substrate for 2C5dH, 2C5/3LVdH and the four human CYP 2C enzymes, and that DMZ primarily undergoes hydroxylation of its benzylic methyl substituent. On the basis of these observations, the structure of 2C5/3LVdH complexed with DMZ was determined to better understand structural determinants of substrate and inhibitor binding.

EXPERIMENTAL PROCEDURES

Structure Determination. DMZ was synthesized as described previously (*4*). P450 2C5/3LVdH was purified using a procedure that employed the detergent, CYMAL-5 (Anatrace) as described (*7*). Briefly, *Escherichia coli* expressed 2C5/3LVdH and was harvested from 3.0 L of culture by centrifugation, lysozyme treatment, and sonication. The P450 was purified by metal ion affinity column chromatography followed by CM-Sepharose ion exchange chromatography. CYMAL-5 detergent was introduced during the final wash and elution of the metal ion affinity column and was maintained in the buffers through the loading of the CM column. At this point, the column was washed extensively with detergent free buffer, and the P450 was eluted in 50 mM potassium phosphate buffer, pH 7.4 containing 500 mM NaCl, 1 mM EDTA, 0.2 mM DTT, and 20% glycerol. The purified protein was concentrated using a centrifugal concentrating device and combined with an equimolar amount of DMZ. The complex was crystallized by the vapor diffusion method using 2.5 μL hanging drops containing 0.24 mM P450, 0.24 mM DMZ, 2.4 mM CYMAL-5 detergent, 1.1 M ammonium sulfate, 0.05 M HEPES pH 7.5, 0.5% PEG400, 20 mM potassium phosphate pH 7.4, 200 mM NaCl, 0.4 mM EDTA, 0.08 mM DTT, and 8% glycerol. The drops were equilibrated against 2.2 M ammonium sulfate, containing 0.1 M HEPES, pH 7.5 and 1% PEG400 at 24 °C. Crystals were prepared for data collection by briefly soaking them in a cryoprotectant of 2.2 M ammonium sulfate, containing 0.1 M HEPES, pH 7.5, 1% PEG400, and 20% sucrose followed by flash-freezing in liquid N_2 (*7*). Crystals were then transferred to the cryo-stream, and data were collected at Stanford Synchrotron Radiation Laboratory (SSRL) beam line 9-2. Data analyzed here were collected at 100 K using a single crystal of dimensions 0.3 × 0.5 × 0.7 mm and were recorded using a Quantum4 CCD detector and 1° oscillations (90 frames, 45 s exposure). Low-resolution, saturated reflections were collected again using 4° oscillations (25 frames, 10 s exposure) to obtain accurate intensity data and merged with the high resolution data. The crystal did not decay noticeably during data collection and exhibited uniform mosaicity of ∼0.6°. The data were processed with CCP4 programs Mosflm and Scala (*8*, *9*), and a statistical analysis of the X-ray diffraction data is presented in Table 1.

The structure of P450 2C5 3LV/dH determined at 3.0 Å (PDB code 1DT6) was used as a starting point for crystallographic refinement. The model was positioned in the unit cell of the DMZ complex by rigid body refinement at 4.0, 3.5, and subsequently 3.0 Å resolutions. The model was subsequently refined against the 2.3 Å data by conjugate gradient least squares minimization, simulated annealing, and individual atomic, isotropic B-factor refinement using CNS (*10*).

Table 1: Data Collection and Refinement Statistics

P450 construct	2C5/3LVdH	
no. of crystals	1	
complex	DMZ	
space group	*I*222	
unit cell (*a*, *b*, *c*) (Å)	74.33, 134.29, 171.84	
Data Collection		
SSRL beam line	BL 9-2	
wavelength (Å)	0.979	
resolution range (Å)	50.0−2.30	
total observations	153 575	
unique reflections > 0.0 σ_F	37 253	
redundancy[a]	4.1 (2.4)	
completeness %[a]	96.6 (78.5)	
$\langle I/\sigma_I \rangle$[a]	13.2 (1.4)	
R_{symm} (*I*)[a]	0.058 (0.479)	
Refinement		
R-factor	0.257	
R_{free} (5% of data)	0.292	
rms deviation bonds (Å)	0.008	
rms deviation angles (deg)[b]	1.29	
Model		
residues/no. of atoms/av. B-factor (Å^2)		
protein[c]	3698	60.4
heme	43	39.5
DMZ[d]	46	80.9
water molecules	118	61.0
sulfate ions	10	81.0

[a] Values for the highest resolution shell in parentheses. Values of $\langle I/\sigma_I \rangle$ < 2.0 are accepted because of some anisotropy in the diffraction. [b] Ramachandran plot: 85.2% of residues in most favored regions; 13.3% in allowed regions; 0.7% in generously allowed regions; and 0.7% in disfavored regions. [c] Residues 27−488 of the 2C5 3LV/dH construct. [d] Includes AC1 and AC2 conformations.

Standard refinement protocols were employed, and the refinement proceeded normally through multiple cycles alternating with interpretation, editing, and adjustment of the model into σ_A-weighted $2|F_o| - |F_c|$ composite omit and $|F_o| - |F_c|$ electron density maps using the program Xfit/Xtalview (*11*). Altogether, the model was evaluated, edited, adjusted, and rebuilt nine times into successive composite omit maps. The increased resolution of the electron density allowed several improvements in the protein model to be made while also indicating a number of conformational changes resulting from substrate binding. While many smaller changes were made, significant adjustment (but not rethreading) of the model was necessary for residues 28−48 (N-terminal β-strand and turn, and A′-helix), 95−99 (Arg97 and Arg430 contacts to the heme propionates), 100−108 (B′-helix), 114−117 (active site side chains), 271−277 (polar surface loop), 412−417 (β-turn and meander region), and 470−474 (β-turn in the active site). In addition, strong electron density was apparent for the region between helices F and G allowing residues 208−211 and 223−228 to be adjusted and the intervening segment of residues 212−222 to be completed. Excluding these residues, the rms deviation of 405 Cα atoms between the starting model (1DT6), the substrate-free enzyme, and the refined DMZ complex (Table 1) is 0.66 Å. In the final stages, H_2O molecules were included

in the model, refined, and edited (Table 1). Strong difference electron density was initially apparent in the active site for DMZ. This density was modeled in the final stages of refinement and interpreted in terms of two alternate conformations suggested by automated substrate-docking studies (below). The occupancy of each conformation (AC1, AC2) was set to 0.5 based on comparison of refined B-factors for the two models of the substrate versus surrounding portions of the protein.

The model has good stereochemistry with only six residues in generously allowed or disallowed regions of the Ramachandran plot (Table 1). Of these, three are in weak density (Lys28, Glu272, Arg374) at the N-terminus or in loops, and three are well-defined in strong density (Ile38, Ala117, Ser426). The electron density defines the χ_1 and χ_2 torsion angles of most Val, Leu, and Ile side chains, whereas Thr, Asn, Gln, and His rotamers are additionally defined by hydrogen-bonding interactions. Coordinates for the DMZ complex of P450 2C5/3LVdH have been deposited in the PDB with accession code 1N6B.

Automated Docking of DMZ. Computer simulated automated docking studies were performed with an explicit hydrogen model generated from the model for the protein using REDUCE (*12*). Preliminary docking employed AUTODOCK 3.05, a grid based docking program (*13*), to examine probable binding sites for DMZ using a modified genetic algorithm that employs a local search to identify low energy binding sites and orientations of the probe molecule. A 50 $\times$ 50 $\times$ 50 grid with a spacing of 0.375 Å centered at -10.11, -14.19, and -26.68 Å that encompassed the active site was used. The parameters for the HC atom type in the AMBER parameter set, parm91 (*14*), were used to define the Leonard−Jones 12-6 pair potentials for the nonpolar hydrogens. The initial conformations of the molecule were based on the crystal structure of sulfaphenazole (*15*), which exhibits both a ring stacked and an extended conformation. The structure of DMZ was built in INSIGHT II (Accelrys) by substituting methyl groups for the amide hydrogen and the aromatic amino group of sulfaphenazole. Partial charges were assigned in INSIGHT II based on the Consistent Valence Force Field (CVFF). Flexible bonds were defined in the substrate to allow internal rotations around unconjugated bonds. The results of 50 randomly seeded runs were analyzed for each of the two conformations of DMZ. The results of each run were clustered when atomic coordinates of the models exhibited less than a 1 Å root-mean-square difference from each other. The results obtained were independent of the initial conformation of the substrate model. Representative structures of the two lowest energy clusters that closely matched the electron density map were used for subsequent crystallographic refinement as alternate conformers designated AC1 and AC2.

RESULTS

Improvements to the purification procedures for the enzyme that included the use of the detergent CYMAL-5 led to the identification of additional crystallization conditions that improved the limits of diffraction (*7*). Crystals were grown in the presence of equimolar concentrations of DMZ and protein. Although the enzyme crystallized in the absence of substrate, the crystals did not diffract as well. The initial concentrations of the enzyme and substrate were 240 μM in the drop, concentrations that greatly exceed the apparent binding constant of 20 μM estimated from substrate induced changes in the spin state of the enzyme. A data set collected for a crystal that diffracted to 2.3 Å resolution was used for model building and refinement. An initial model was obtained by molecular replacement and finalized by sequential rounds of building, fitting, and refinement. The final model exhibits an R value of 0.26 and an R_{free} value of 0.29 (Table 1).

As shown in Figure 1, the polypeptide fold for the structure of the enzyme substrate complex corresponds closely to that of the original structure, PDB code 1DT6, but significant conformation changes were observed in regions forming the distal portions of the substrate binding site that required extensive rebuilding (Experimental Procedures). These regions show larger B values than are seen for the structural core (Figure 1a). The Cα atoms of the F and G helices have moved roughly 1−1.5 Å relative to the axis of helix I (Figure 1c). Helix I is relatively straight in the earlier structure of 2C5/3LVdH, but it exhibits a distinct bend that displaces the N-terminal end of the helix in the current structure (Figure 1c).

The higher resolution and improved B values for the region between helices F and G, residues 205 to 230, allowed this portion of the enzyme to be modeled for the first time, Figure 2. As shown in Figures 1 and 2, the F−G loop contains two short helical segments, F′ and G′, separated by a turn. The G′ helix contributes to a hydrophobic surface on the catalytic domain near the N-terminus where it joins the transmembrane leader sequence. This region is thought to contribute to the membrane interactions exhibited by the truncated enzyme, 2C5dH (*1*). The G′ helix is largely hydrophobic with the exception of D224 and is flanked by aromatic residues. The helical structure of this region increases the likelihood that the hydrophobic tip of the protein can be buried in the membrane by reducing the need to hydrate the peptide backbone (*16*).

In addition, the region between helices B and C is better ordered revealing an additional helix, B′, which is generally seen in prokaryotic P450s (Figure 1a,b). The increased order in this region may be derived, in part, from substrate interactions with the B′, F, and G helices. As a result of the increased resolution (2.3 vs. 3.0 Å) and lower B values, the overall geometry of the model is improved relative to the previously published structure. In addition to overall better stereochemistry, β-strands are also better defined revealing a β-turn in the meander region, at residues 411−414, that is not evident in prokaryotic P450s.

Examination of the $|F_o| - |F_c|$ electron density map revealed significant electron density for DMZ in the active site of P450 2C5 (Figure 3). As the density did not uniquely define the orientation of the molecule, a computational approach was taken to model the interaction of DMZ with the refined model of 2C5/3LVdH. To model substrate interactions with the protein, an explicit hydrogen model of the protein was generated using the program REDUCE (*12*). REDUCE evaluates the orientations of the side-chain N and O atoms of Asn and Gln residues as well as of His χ_2 rotamers by selecting the best hydrogen bonding potentials while minimizing or eliminating clashes with other residues. The results of this minimization were also used to check for consistency with respect to the X-ray refinement. The

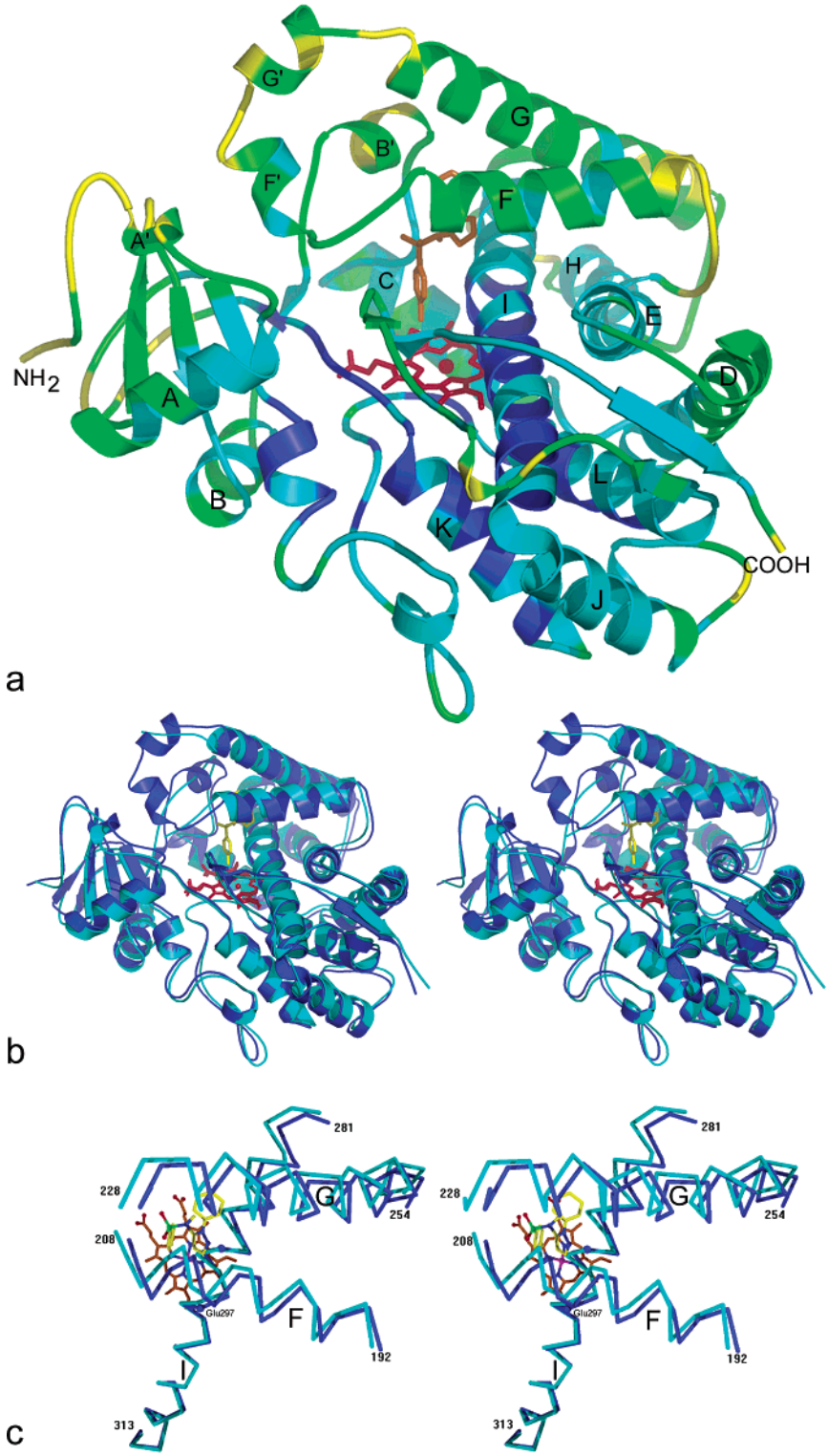

FIGURE 1: (a) Structure of cytochrome P450 2C5/3LVdH with DMZ bound. The principle helices and NH_2- and COOH-termini are labeled. The protein is colored according to average B-factor per residue: dark blue, <40 Å^2; cyan, 40−60 Å^2; green, 60−80 Å^2; yellow, 80−100 Å^2. The heme group is red, and the AC1 orientation of DMZ is orange. (b) Stereoview of the superposition of the DMZ bound (blue) and original (PDB 1DT6, cyan) structures of P450 2C5/3LVdH. The superposition is based on 405 Cα atoms as described in the Experimental Procedures, and the view is similar to that in panel a. (c) Stereoview of the superposition in panel b showing Cα atoms of helices F, G, and I with respect to the heme and DMZ. Helices F and G shift 1−1.5 Å in the DMZ complex, and helix I is distinctly bent at its N-terminal end with respect to the original structure of the enzyme. In addition, the turn in helix I at residues 293−297 expands in the DMZ bound form where two H_2O molecules are bound (Figure 7). Glu297 interacts with ordered H_2O molecules in the cleft between helices F and I; the bend in helix I originates at Gly293.

orientations of methyl hydrogens of methionine and of hydroxyl hydrogens were optimized to avoid clashes with other residues and, in the latter case, to promote hydrogen bonding.

Automated docking studies were performed using AUTODOCK 3.05 (*13*), which scans the active site for low energy binding sites using a modified genetic algorithm and precomputed grids for the evaluation of the interaction energy.

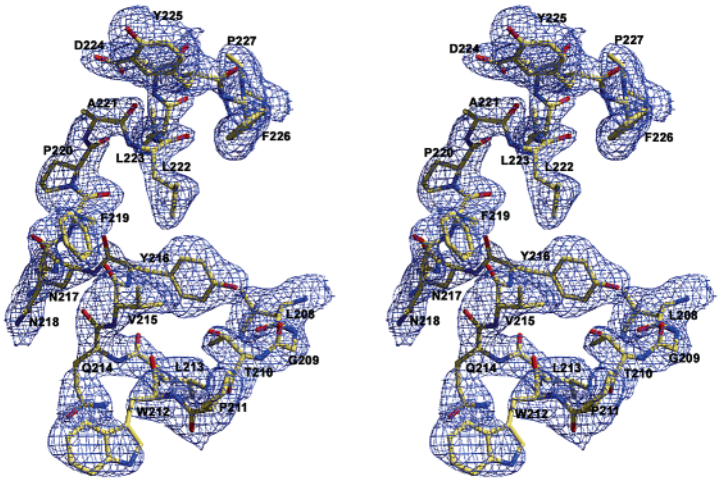

FIGURE 2: Stereo figure of the electron density for the FG region in the structure of P450 2C5/3LVdH with DMZ bound. The σ_A-weighted $2|F_o| - |F_c|$ map calculated with residues 208−227 omitted shown in the figure is contoured at 2σ. The polypeptide extends from the end of helix F at residues 207−209, exhibits a β-turn at residues 210−213, and then folds into a pair of short helical segments with 3_{10} and α conformation separated by Pro residues. Helix F′ is comprised of residues 214−219, and helix G′ is comprised of residues 220−226, with the chain changing direction at Pro211 preceding F′, Pro220 between F′ and G′, and Pro227 at the N-terminus of helix G.

The grids were calculated based on protein coordinates at the completion of refinement and prior to inclusion of DMZ in the model. Nonbonded interactions were computed based on the AMBER force field (*14*) as employed in version 2.5 of AUTODOCK (*17*). The results from 50 randomly seeded docking attempts were clustered when the results exhibited <1 Å rms deviation for their atomic coordinates. The initial conformations of the DMZ corresponded to either the extended or ring stacked conformation observed in crystal structures of sulfaphenazole (*15*). Most of the runs placed DMZ in the two lowest energy clusters. DMZ exhibited an extended conformation for each these clusters, and the results were independent of the initial conformation of the substrate model. The two clusters exhibited antiparallel orientations of the long axis of the molecule. Both models overlaid the electron density for the substrate, and each was refined against the X-ray diffraction data as alternate conformers with equal occupancy. The resulting structures for the substrate are shown together with the electron density observed for the substrate seen in a σ_A-weighted $|F_o| - |F_c|$ omit map (Figure 3a). The two locations of the substrate accounted for all of the density better than either of the conformations individually (Figure 3b,c). In addition, $|F_o| - |F_c|$ electron density maps indicated the presence of additional, positive electron density when only one conformer was included in the model.

Both conformations of the substrate reside in a tubular channel (Figure 4) that opens to bulk solvent between the loop between the B′ and the C helices and helices I and G in the original structure but that is closed by the hydrogen bonding of Lys241 with the backbone carbonyl of Val106 (Figure 5) in the DMZ complex. The tubular channel constrains the long axis of the substrate so that the molecule cannot reorient once the substrate enters the enzyme. Extra space is available in the active site when DMZ is bound in either location, and the two locations of DMZ are distinct and overlapping. There is additional space available to accommodate larger R1 and R2 groups in the AC2 location (Figure 4, lower panel) and to a lesser extent in the AC1 location. In addition, the portions of the protein bounding the extra space near the R1 and R2 groups in the AC2 location are very flexible. Results presented in the accompanying manuscript (*6*) indicate that analogues bearing larger groups are potent inhibitors of 2C5dH.

The opening to the active site is larger in the original structure of 2C5/3LVdH, PDB code 1DT6. Lys241 is not hydrogen bonded to the carbonyl of Val106 in this structure, and the B′ helix is shifted and exhibits a different conformation. The B′ helix has moved in the DMZ complex to maximize hydrophobic contacts with the substrate. This movement increases the contact with the G helix to close the access channel and facilitates hydrogen bonding to Lys241. Two factors are likely to make the B−C loop a flexible gateway for substrate access to the active site. The B′ helix is flanked by GlyXGly motifs (Figure 5) that are highly conserved in family 2 P450s. As the glycine residues do not have side chains and can readily adopt a wide range of conformations for the peptide backbone, these glycines provide torsional flexibility to the backbone. Second, helix B′ does not exhibit extensive van der Waals contacts with the rest of the structure (Figure 5). The single hydrogen bond between the side chain of Lys241 and the carbonyl of Val106 can be easily displaced by hydration of Lys241 to open the access channel. This flexibility potentially contributes to substrate access and adaptive conformation changes to different substrates.

A second solvent channel is seen in the original structure of the enzyme, 1DT6. The movement of helix F relative to helix I (Figure 1c) closes this exit channel for water between the F and the I helices near the conserved acidic residue, Glu297 (Figure 4). Closure of this channel could be important for control of proton access to reduced oxygen species formed at the heme iron during the catalytic cycle.

Contacts formed by the protein with the substrate are predominantly hydrophobic for both binding locations. Residues within 5 Å of DMZ are Leu103, Val106, Ala113, Phe114, Leu201, Asn204, Val205, Leu208, Leu213 (AC1 only), Ala237, Ile240, Ser289 (AC2 only), Asp290, Gly293, Ala294, Thr298, Leu359, Leu363, Phe473, and Val474 (AC2 only). These are illustrated for the AC1 location in Figure 6. Most of these residues identified in this study were predicted to be substrate contact residues based on the

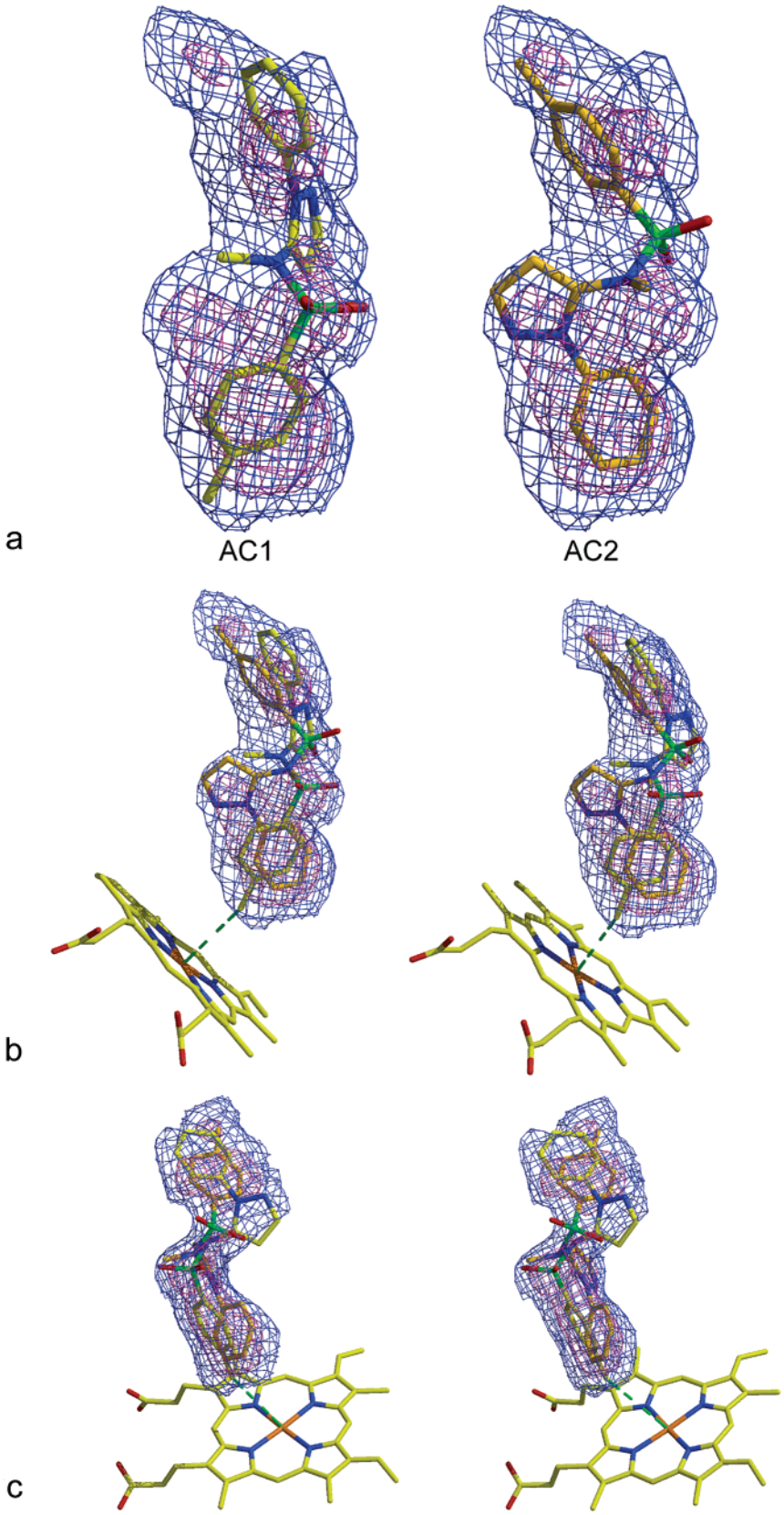

FIGURE 3: Stereo figures of the electron density for DMZ in the structure of P450 2C5/3LVdH. The map is calculated with DMZ omitted from the model using σ_A-weighted $|F_o| - |F_c|$ coefficients and is contoured at 3σ and 6σ. In panel a, the productive AC1 orientation of DMZ is shown in the left panel, and the less productive AC2 orientation is shown in the right panel, illustrating how either orientation alone is insufficient to account for the electron density. In panels b and c, the AC1 and AC2 orientations are superposed, and the difference electron density is shown with respect to the heme in orthogonal views. On the basis of refinement of B-factors, the occupancy of the two orientations is approximately equal. The productive AC1 orientation places the benzylic methyl group 4.4 Å from the heme Fe (green dashed line), while the less productive AC2 orientation places this group toward the entrance to the active site access channel. The *N*-phenyl ring carbon closest to the heme in the AC2 orientation is 5.9 Å from the Fe. Atoms of DMZ are colored yellow or gold for C, blue for N, red for O, and green for S.

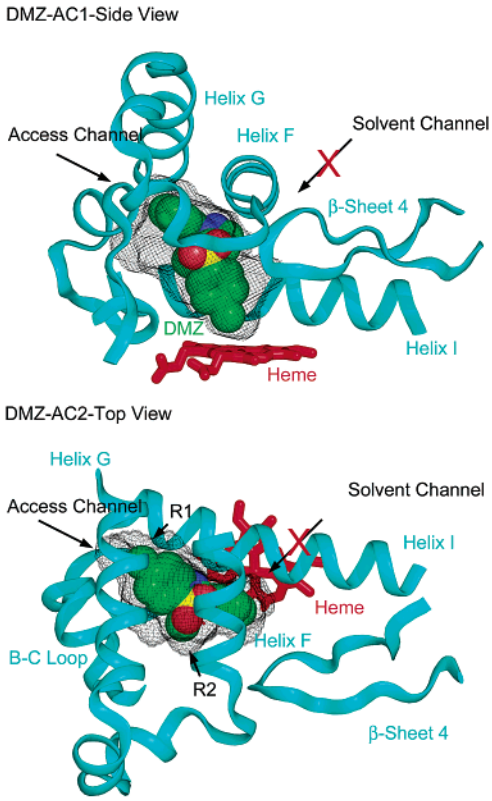

FIGURE 4: Solvent accessible surface of the substrate binding cavity in the DMZ complex of P450 2C5/3LVdH is represented by a black mesh. The AC1 and AC2 orientations of DMZ are shown as CPK models. The upper panel depicts the AC1 location of DMZ from a side view opposite helix I. Extra space is evident that is utilized when DMZ binds to the AC2 location. The lower panel depicts the AC2 location of DMZ from above the heme. Extra space is evident that can accommodate larger R1 and R2 groups. The substrate binding cavity is bounded by residues flanking helix B′ (B−C loop); helices B′, F, G, and I; and β-sheet 4. The substrate binding cavity extends from the active site to the enzyme surface between helices B′, I, and G, suggesting that the substrate accesses the active site via this channel with concomitant conformational change in the B′ and FG regions. At the same time, a solvent access channel between helices F and I is blocked in this conformation.

original structure and correspond to residues implicated in substrate specificity by mutagenesis studies of family 2 enzymes (*1*). Additional residues identified in this study include Leu213, which is present in the newly completed portion of the structure between the F and the G helices, and S289, which only contacts the less productive AC2 orientation of the substrate. Ala237 and Ile240 are located in the SRS3 region predicted by Gotoh (*18*). Residues in SRS3 had not been implicated in substrate binding by earlier mutagenesis studies of CYP2 enzymes, and this is the first demonstration that these residues in SRS3 contribute to substrate binding interactions in mammalian P450s. Amino acid contacts with the sulfonyl moiety of DMZ are hydrophobic. It should be noted that methylation of the sulfonamide group to reduce its polarity plays a significant role in the conversion of sulfaphenazole into a substrate for this enzyme (*6*). In addition, the two orientations of the substrate do not overlay closely, indicating that excess space is present in the active site cavity that is not utilized by a single conformation of the bound substrate (Figure 3). Consistent with this, the *B* values for DMZ are larger than those of the surrounding residues suggesting that the substrate is not

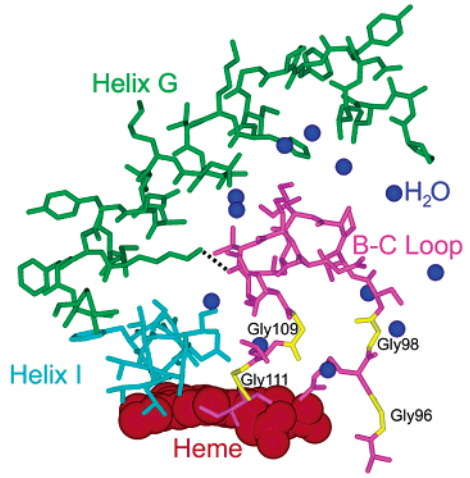

FIGURE 5: Features in the structure of P450 2C5/3LVdH with DMZ bound indicate flexibility of the B−C loop consistent with conformational change required for substrate access. The B′ helix is flanked by two conserved GlyXGly motifs. Contacts between helices G and I and the C-terminal end of helix B′ are mediated primarily by intervening H_2O molecules with only one direct contact involving the carbonyl of Val106 and Lys241.

tightly constrained in the active site. Hence, DMZ is able to bind in two alternative locations that involve favorable hydrophobic contacts within the active site.

Oxidation of DMZ mainly occurs on its 4-methyl group, and to a very minor extent (<1%), on its *N*-phenyl ring (*6*). The most productive orientation of DMZ (AC1) places the 4-methyl group 4.4 Å from the heme iron (Figure 3b,c). This distance is similar to that seen for substrates in the active sites of P450 cam (*19*, *20*) and P450 eryF (*21*), and it is likely that a carbon−hydrogen bond will be appropriately oriented and positioned for reaction with the reactive oxygen intermediate formed on the heme Fe during catalysis. The less productive orientation, AC2, places the *N*-phenyl ring of DMZ 5.9 Å from the Fe, such that it is less likely to interact productively with the oxyferrylporphyrin cation radical intermediate in the catalytic cycle (Figure 3b,c).

Incubation of P450 2C5dH with DMZ produces only small changes (20−30%) in the visible absorption spectrum of the enzyme suggesting that the binding of the substrate does not completely displace water bound to the heme Fe (*6*). The AC2 location of the substrate would not conflict with the presence of a heme bound water molecule, and although the distance between the benzylic methyl group of DMZ in the AC1 orientation and the heme bound water is less than ideal (ca. 3 Å), the presence of the substrate in this location might not lead to extensive displacement of water from the iron. Examination of $|F_o| - |F_c|$ and $2|F_o| - |F_c|$ electron density maps does not reveal significant positive, residual density above the heme that would suggest that a water molecule is bound to the iron. It is possible that partial occupancy of H_2O and the resolution of the data do not permit electron density for a water molecule to be clearly resolved from the density for the heme iron.

Although a heme bound water molecule could not be clearly identified, a number of ordered water molecules were identified near the substrate binding site (Figure 7). Two H_2O molecules bind in the expanded groove of helix I between the amides of Glu297 and Thr298 and the carbonyls of Gly293 and Ala294 that results from the increased deviation

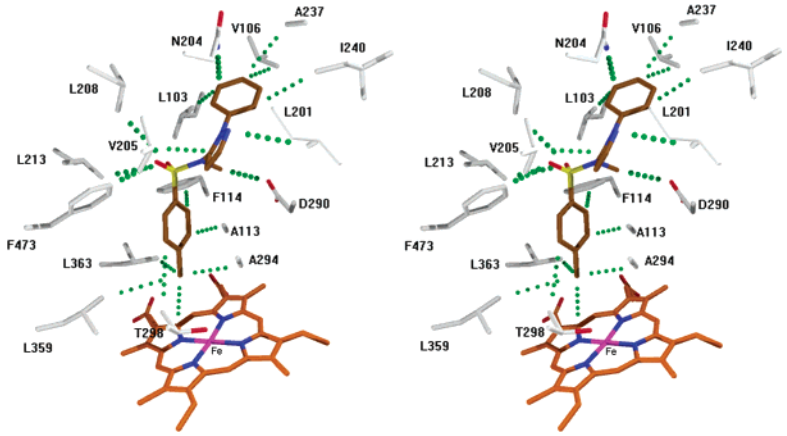

FIGURE 6: Stereo figure showing interatomic distances less than 5 Å between the more productive AC1 orientation of DMZ and amino acid side chains in 2C5/3LVdH. Contacts occur with side chains of 17 residues. Additional contacts occur with G293, which does not have a side chain, and two with the heme: one of these is the 4.4 Å distance between the C4-methyl group and Fe (not shown here; see Figure 3); the other involves a pyrrole carbon atom. The side chains of two additional residues, S298 and V474, that contact the AC2 conformation of DMZ, are not shown in the figure. The majority of the contacts involve nonpolar atoms on both the substrate and the enzyme.

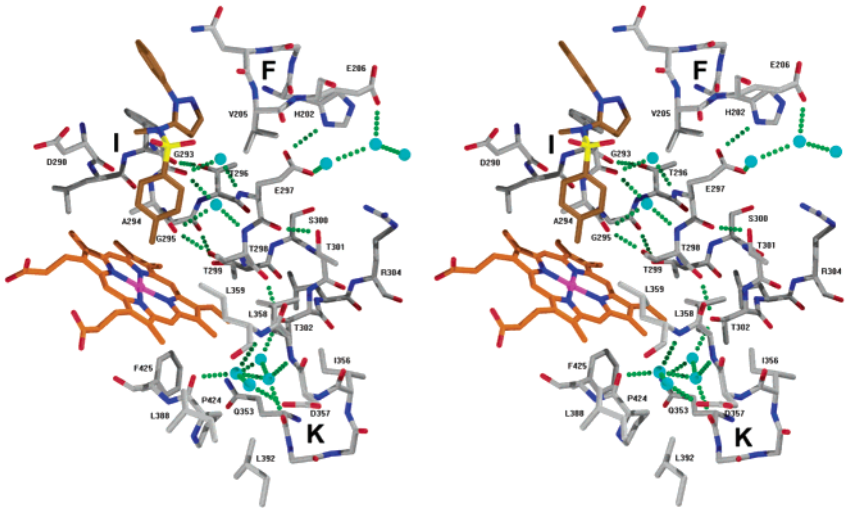

FIGURE 7: Three sites for ordered H_2O molecules in the 2.3 Å structure of the DMZ complex of P450 2C5/3LVdH are shown. H_2O molecules are displayed as cyan spheres with hydrogen bonds as green dotted lines; hydrogen bonds involving helix I side chains are also shown. H_2O molecules are observed in an expanded grove in helix I, in the cleft between helices F and I, and within a cavity adjacent to the heme (see text). The productive AC1 orientation of DMZ is shown as in Figure 6.

of the N-terminal end of the I helix when DMZ binds. These H_2O molecules are expected to play a role in proton transport during catalysis (*19*). Three H_2O molecules are observed between Glu297 and Arg304 on helix I and His202 and Glu206 on helix F (Figure 7). This polar environment on the protein surface also includes Ser300 and Thr301 on helix I, Asn471, and Ser479 on β-sheet 4 and Asp165 and Thr167 on helix E. These ordered H_2O molecules are present in the end of the solvent channel external to the active site. The solvent channel is closed by conformation changes that occur when DMZ binds to the enzyme (Figure 4). A chain of four H_2O molecules with multiple hydrogen bonding interactions is located between the heme edge, residues of helices I and K, and a β-bulge following helix K. These H_2O molecules interact with the amides of Leu358, Leu359, the carbonyls of Gln353 and Pro424, and the side chains of Thr302, Gln353, and Asp357 within a cavity defined by the heme and the side chains of Ile356, Leu358, Leu359, Leu388, Leu392, Pro424, and Phe425 (Figure 7). While being ~10 Å from the Fe, these H_2O molecules are linked to helix I via a hydrogen bond with Thr302. H_2O molecules also bind between the carboxylate of the highly conserved helix I residue Asp290, the amide groups of Ile112, Ala113, and Phe114, and the side chain of Ser107 on helix B′, forming a network of interactions.

DISCUSSION

The structure of DMZ bound CYP2C5/3LVdH provides a number of interesting differences between the DMZ complex and the original structure of the enzyme. Our results indicate that the flexible regions of the protein adapt for substrate binding. These conformation changes contribute to favorable binding interactions and are also likely to control subsequent access of protons to reduced oxygen species formed at the binding site. However, the substrate appears to bind in two distinct, overlapping locations of which only one leads to extensive oxidation. Multiple substrate binding orientations and locations are likely to contribute to the ability of family 2C enzymes to form multiple products from a single substrate. The multiple binding locations reflect extra space in the active site that also contributes to the recognition of structurally distinct substrates by these enzymes.

Comparisons of the original structure and the DMZ-bound structure of CYP2C5 indicate that substrate binding results in significant shifts in the flexible portions of the structure that form the substrate-binding cavity. The largest movement is seen for the B′ helix and the region between helix F and the N-terminal end of helix I. The original structure was determined for CYP2C5/3LVdH that was crystallized in the absence of substrate. However, there was unexplained electron density in the active site suggestive that the protein may have bound ligands present in the *E. coli* host (*1*) and that the original structure may not reflect the structure of the substrate/ligand free enzyme. A comparison of electron density maps for the original and DMZ complex suggests that if a ligand is present in the original structure the occupancy is relatively low and poorly defined. However, the crystals may selectively incorporate a relatively closed form of the enzyme, and more open conformations may exist in solution. Such open conformations would increase the accessibility of substrates to the active site.

Substrate binding to CYP102 also elicits a rearrangement of the region from helix F through the N-terminal end of helix I in CYP102, reflecting a lateral movement of the F and G helices across the surface of the I helix corresponding to one helical turn. This closes an open cleft above the substrate binding site (*22*, *23*). However, the apparent entrances to the CYP102 and CYP2C5 active sites are located on opposite sides of the flexible B′ helix. The positions of the F and G helices along the axis of helix I in CYP2C5/3LVdH are similar to those observed for the substrate bound structure of CYP102, and the movement of the F and G helices in CYP2C5 when DMZ binds reflect >1 Å translations normal to the axis of helix I when compared to the original structure. Differences in the conformation of helix F are also seen when the inhibitor bound (*24*) and substrate-free structures (*25*) of CYP119 are compared that reflect the contraction of the protein around the bound inhibitor to maximize favorable contacts and the displacement of water from the substrate binding site. Translation of helix F when substrate binds closes a water channel between the F and I helices in both CYP102 (*22*) and CYP2C5. Closure of this channel may facilitate control of the protonation of reduced oxygen intermediates that are bound to the heme iron during catalysis.

Substrate docking studies suggested that the extended conformation of DMZ binds in two locations with antiparallel orientations of the long axis of the substrate. Both binding locations coincided with the electron density for DMZ. Refinement of the two models indicated that neither model accounted for all of the density and that the two models together did. At this resolution, it is not possible to independently refine the occupancies and *B* values for the substrate, and the occupancies of the two models of the bound substrate were arbitrarily set to 50% each. The refined *B* values for the two alternative locations of the substrate were lower than those obtained if either of the two models was refined with full occupancy. The refined *B* values at 50% occupancy are also more similar to those of active site residues with that of the AC1 orientation being somewhat greater than the AC2 orientation. The AC1 orientation places the predominant site of hydroxylation close to the heme.

DMZ approaches the heme more closely when bound to the AC1 location with the principal site of hydroxylation positioned 4.4 Å from the heme. The respective substrates for prokaryotic P450s cam (*19*, *20*) and eryF (*21*) are positioned such that the carbon that is hydroxylated resides between 4.5 and 4.8 Å from the Fe. These substrates are also oriented so that the hydrogen that is abstracted during the reaction is predicted to be within 2 Å of the oxygen of the oxy-perferryl intermediate. The AC1 location of DMZ is consistent with efficient hydrogen abstraction from the benzylic methyl group by the oxy-perferryl intermediate, and this is the predominant site of hydroxylation. In contrast, the AC2 orientation of the substrate places the *N*-phenyl moiety closest to the heme. Hydroxylation of this ring accounts for <1% of the product (*6*). Consistent with this, the closest approach of the phenyl ring is 5.9 Å from the heme iron, and the aromatic hydrogen atoms are not directed toward the iron. Occupation of the active site by the substrate in the AC2 orientation is likely to reduce the effective k_{cat} of the enzyme for oxidation of the benzylic methyl group by competing for occupancy of the enzyme with the substrate in the more productive AC1 location.

Substrate binding to CYP101 and CYP102 leads to a conversion of these enzymes from a low spin to high spin heme protein suggesting that a water molecule is displaced from the sixth coordination site of the heme iron. In CYP102, the bound substrate is positioned more than 7 Å from the heme iron, and the displacement of the water reflects a change in conformation of the I helix that results in the movement of water away from the iron because of altered hydrogen bonding between water molecules and amino acids in the I helix (*22*). In contrast, the binding of DMZ to wild-type CYP2C5 or the modified form that was crystallized produces only a small increase in high spin character (*6*). A broad survey of mammalian P450s suggests that substrate elicited spin-state changes differ greatly in magnitude and are not clearly correlated with rates of oxidation (*26*). Well-resolved density for a significantly occupied water bound to the sixth coordination site of the heme iron was not evident at this resolution for the DMZ complex of CYP2C5. It is of interest to note that the less productive orientation of the substrate AC2 would not conflict with water bound to the iron because the substrate is located too far away. However, the more productive orientation of the substrate would place the benzylic methyl group of DMZ ∼3 Å from a H_2O bound to the iron potentially reducing the occupancy of water bound to the heme iron. It seems unlikely that the extent of the spin state change and the relative occupancy of the two orientations of the substrate are strictly correlated. Therefore, the apparent absence of bound H_2O for the less productive binding mode of DMZ remains to be explained with respect to the small spin state change observed with DMZ binding.

The observation of multiple binding orientations for the substrate is consistent with observations that P450s can produce multiple products arising from the oxidation of distal sites on the same substrate molecule. For example, the closely related enzyme, P450 2C3, hydroxylates progesterone in both the 16α and the 6β positions (*27*). Although P450 2C5 generally exhibits a single product, it is interesting to note that 21-hydroxy pregnenolone, the product of pregnenolone hydroxylation by P450 2C5, is also a substrate that is converted by P450 2C5 to 21-hydroxy-progesterone. This is thought to arise from the formation of a gem diol at C3 that subsequently rearranges to form the Δ4,3-keto product

and a molecule of water (*28*), indicating that steroid substrates can enter the enzyme and position the A ring proximal to the heme. P450 2C5 also hydroxylates the A ring of 17β-estradiol to form catechol estrogens (*29*).

For some P450s or in those cases where the substrate is small, a spectrum of products is often obtained, reflecting oxidation of the substrate at different sites. This could reflect the ability of the substrate to reorient in the active site, and the product distribution often reflects the relative chemical reactivity of the sites of oxidation. In other P450s, the sites of hydroxylation are likely to reflect steric constraints that direct the site of oxidation (*30*). The shape of the active site cavity of 2C5 does not permit the substrate to be inverted about the long axis once it is bound. More significant steric constraints are likely to occur near the heme iron where the protein structure exhibits the greater rigidity resulting from well-defined secondary structure and internal packing contacts. The docking results, and a general increase in *B* values away from the heme and toward the entrance to the active site, suggest that the active site is less restrictive further from the iron, and this clearly contributes to the substrate binding in different orientations.

The picture of the active site for P450 2C5 that emerges is one where a tubular access channel to the heme accommodates a range of substrates in different orientations. Polar atoms of the substrate can remain solvated near the channel entrance without specific, compensatory protein interactions, while hydrophobic atoms displace H_2O molecules within the active site. At the same time, the protein undergoes conformation changes that maximize hydrophobic interactions with the substrate and restrict access of water to the active site. This architecture may be a general one underlying the capacity of drug metabolizing P450 enzymes to bind and oxidize a wide range of xenobiotic and endobiotic substrates.

ACKNOWLEDGMENT

The authors thank the staff of the Stanford Synchrotron Radiation Facility for their generous assistance and James Zuberi for assistance with figures.

REFERENCES

1. Williams, P. A., Cosme, J., Sridhar, V., Johnson, E. F., and McRee, D. E. (2000) *Mol. Cell 5*, 121−132.
2. Cosme, J., and Johnson, E. F. (2000) *J. Biol. Chem. 275*, 2545−2553.
3. Mancy, A., Dijols, S., Poli, S., Guengerich, F. P., and Mansuy, D. (1996) *Biochemistry 35*, 16205−16212.
4. Ha-Duong, N. T., Dijols, S., Marques-Soares, C., Minoletti, C., Dansette, P. M., and Mansuy, D. (2001) *J. Med. Chem. 44*, 3622−3631.
5. Ha-Duong, N. T., Marques-Soares, C., Dijols, S., Sari, M. A., Dansette, P. M., and Mansuy, D. (2001) *Arch. Biochem. Biophys. 394*, 189−200.
6. Marques-Soares, C., Dijols, S., Macherey, A.-C., Wester, M. R., Johnson, E. F., Dansette, P. M., and Mansuy, D. (2003) Sulfaphenazole derivatives as tools for comparing cytochrome P450 2C5 and human cytochrome P450 2Cs: Identification of a new high affinity substrate common to those enzymes, *Biochemistry 42*, 6363−6369.
7. Wester, M. R., Stout, C. D., and Johnson, E. F. (2002) *Methods Enzymol. 357*, 73−79.
8. Leslie, A. G. W. (1996) in *A data collection strategy option in MOSFLM*, CCP4 Newletter.
9. CCP (1994) *Acta Crystallogr. D50*, 760−763.
10. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D54 (Pt 5)*, 905−921.
11. McRee, D. E. (1999) *J. Struct. Biol. 125*, 156−165.
12. Word, J. M., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (1999) *J. Mol. Biol. 285*, 1735−1747.
13. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998) *J. Comput. Chem. 19*, 1639−1662.
14. Weiner, S. J., Kollman, P. A., Nguyen, D. T., and Case, D. A. (1985) *J. Comput. Chem. 7*, 230−252.
15. Patel, H. C., and Singh, T. P. (1987) *Acta Crystallogr. C43*, 1131−1134.
16. White, S. H., and Wimley, W. C. (1998) *Biochim. Biophys. Acta 1376*, 339−352.
17. Morris, G. M., Goodsell, D. S., Huey, R., and Olson, A. J. (1996) *J. Comput. Mol. Des. 10*, 293−304.
18. Gotoh, O. (1992) *J. Biol. Chem. 267*, 83−90.
19. Schlichting, I., Berendzen, J., Chu, K., Stock, A. M., Maves, S. A., Benson, D. E., Sweet, R. M., Ringe, D., Petsko, G. A., and Sligar, S. G. (2000) *Science 287*, 1615−1622.
20. Poulos, T. L., Finzel, B. C., and Howard, A. J. (1987) *J. Mol. Biol. 195*, 687−700.
21. Cupp-Vickery, J. R., and Poulos, T. (1995) *Nature Struct. Biol. 2*, 144−153.
22. Haines, D. C., Tomchick, D. R., Machius, M., and Peterson, J. A. (2001) *Biochemistry 40*, 13456−13465.
23. Li, H. Y., and Poulos, T. L. (1997) *Nature Struct. Biol. 4*, 140−146.
24. Yano, J. K., Koo, L. S., Schuller, D. J., Li, H., Ortiz de Montellano, P. R., and Poulos, T. L. (2000) *J. Biol. Chem. 275*, 31086−31092.
25. Park, S. Y., Yamane, K., Adachi, S., Shiro, Y., Weiss, K. E., and Sligar, S. G. (2000) *Acta Crystallogr. D56*, 1173−1175.
26. Guengerich, F. P. (1983) *Biochemistry 22*, 2811−2820.
27. Hsu, M.-H., Griffin, K. J., Wang, Y., Kemper, B., and Johnson, E. F. (1993) *J. Biol. Chem. 268*, 6939−6944.
28. Trant, J. M., Lorence, M. C., Johnson, E. F., Shackleton, C. H. L., Mason, J. I., and Estabrook, R. W. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 9756−9760.
29. Schwab, G. E., and Johnson, E. F. (1985) *Arch. Biochem. Biophys. 237*, 17−26.
30. Jones, J. P., Shou, M., and Korzekwa, K. R. (1996) *Adv. Exp. Med. Biol. 387*, 355−360.

BI0273922