Computational Solvent Mapping Reveals the Importance of Local Conformational Changes for Broad Substrate Specificity in Mammalian Cytochromes P450†

Karl H. Clodfelter,‡ David J. Waxman,§ and Sandor Vajda*,‖

Bioinformatics Program and Departments of Biology and Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215

Received February 18, 2006; Revised Manuscript Received June 2, 2006

ABSTRACT: Computational solvent mapping moves small organic molecules as probes around a protein surface, finds favorable binding positions, clusters the conformations, and ranks the clusters on the basis of their average free energy. Prior mapping studies of enzymes, crystallized in either substrate-free or substrate-bound form, have shown that the largest number of solvent probe clusters invariably overlaps in the active site. We have applied this method to five cytochromes P450. As expected, the mapping of two bacterial P450s, P450 cam (CYP101) and P450 BM-3 (CYP102), identified the substrate-binding sites in both ligand-bound and ligand-free P450 structures. However, the mapping finds the active site only in the ligand-bound structures of the three mammalian P450s, 2C5, 2C9, and 2B4. Thus, despite the large cavities seen in the unbound structures of these enzymes, the features required for binding small molecules are formed only in the process of substrate binding. The ability of adjusting their binding sites to substrates that differ in size, shape, and polarity is likely to be responsible for the broad substrate specificity of these mammalian P450s. Similar behavior was seen at "hot spots" of protein−protein interfaces that can also bind small molecules in grooves created by induced fit. In addition, the binding of *S*-warfarin to P450 2C9 creates a high-affinity site for a second ligand, which may help to explain the prevalence of drug−drug interactions involving this and other mammalian P450s.

Cytochromes P450, a gene superfamily of heme protein monooxygenases found in eukaryotes, prokaryotes, and archaea, are notable both for the diversity of reactions that they catalyze and the range of chemically dissimilar substrates upon which they act (*1*). Many bacterial P450s catalyze specialized biosynthetic reactions or confer the ability to metabolize compounds used as a carbon source (*2*). Most extrahepatic mammalian P450s also catalyze specific steps in the biosynthesis of steroid hormones, cholesterol, prostanoids, and bile acids (*3*). However, the majority of hepatic P450s exhibit much broader substrate specificity, enabling them to metabolize endogenous and exogenous compounds, including drugs and other xenobiotics. Since all P450s share the same protein fold, and substrate binding can induce large conformational changes in both bacterial and mammalian P450s (*3*, *4*), it is not clear what structural features determine the range of substrates that can be hydroxylated by a particular enzyme.

In this paper we describe the application of computational solvent mapping (*5*−*7*), a powerful protein binding site analysis tool, to five P450s and show that these calculations provide new information on the potential origin of broad substrate specificity seen with many hepatic P450 enzymes. The idea of solvent mapping was first introduced by Ringe and co-workers (*8*, *9*), who determined protein structures in aqueous solutions of organic solvents and in each case found only a limited number of solvent molecules bound to the protein. When five or six structures of a protein determined in different solvents were superimposed, the organic molecules tended to cluster in the active site, forming "consensus" sites that delineate important subsites of the binding pocket (*8*). All other bound solvent molecules either are in crystal contact, occur only at high ligand concentration, or are in small, buried pockets, where only a few types of solvent molecules cluster compared to the active site (*7*).

We have developed an algorithm to perform solvent mapping computationally (*5*−*7*) rather than experimentally. The algorithm places small organic molecules ("probes") over the protein surface, finds the most favorable positions, clusters the conformations, and ranks the clusters on the basis of average free energy (*5*). The calculations reproduced the available experimental solvent mapping results (*5*), and in all our previous applications to enzymes, the consensus sites identified always corresponded to major subsites of the substrate-binding site (*5*−*7*). Moreover, when applied to peroxisome proliferator-activated receptor-γ, a nuclear receptor, computational solvent mapping identified important binding sites within the ligand-binding pocket, as well as

† This investigation was supported by Grant P42 ES07381 from the National Institute of Environmental Health Sciences and Grant GM64700 from the National Institutes of Health.

* To whom correspondence should be addressed. Tel: 617-353-4757. Fax: 617-353-6766. E-mail: vajda@bu.edu.

‡ Bioinformatics Program, Boston University.

§ Department of Biology, Boston University.

‖ Department of Biomedical Engineering, Boston University.

10.1021/bi060343v CCC: $33.50
Published on Web 07/15/2006

sites involved in stabilization of the ligand-binding domain and coactivator binding (*10*). We emphasize that, prior to mapping, all bound ligands and water molecules are removed. Thus, the results are based only on the structure of the protein, and hence the method can be used to identify and characterize binding sites even in proteins without any known ligand.

The cytochromes P450 examined in this study were selected from the structures available in the Protein Data Bank (PDB)[1] because of their high-resolution structures in substrate-bound and unbound conformations. As will be described, we have found that the results of mapping qualitatively differ between bacterial and mammalian P450s. The bacterial enzymes, represented here by P450 cam (CYP101) and P450 BM-3 (CYP102), behave as all other enzymes examined (*5*−*7*); i.e., they have a binding pocket that is already formed prior to substrate binding, and the mapping finds this site in both ligand-bound and ligand-free structures. In contrast, no probes cluster in the active site when mapping the ligand-free structures of the three mammalian P450s investigated here. Although P450s 2C5, 2C9, and 2B4 all have large binding channels that are readily apparent in their unbound states, these channels appear to lack the structural features required for anchoring the small molecules used as probes in the mapping procedure. However, the mapping finds the binding sites in the ligand-bound conformations of the same mammalian P450s. Thus, the appropriate grooves and crevices are formed in the process of substrate binding, indicating that these P450s adjust the shape and polarity of their binding sites to accommodate different substrates.

The fact that a binding site can be formed by induced fit is not particularly surprising. We have found it more interesting that these changes can be very small and yet essential for the recognition of the probes. For example, a comparison of the free and substrate-bound structures of P450 2C9 reveals only a few specific side-chain movements rather than large-scale conformational changes, but the mapping finds the binding site only in the substrate-bound structure. As will be discussed, a similar importance for ligand-induced conformational changes was observed in the binding of small molecules to "hot spots" of protein surface regions that primarily interact with other proteins. Although such regions are relatively flat, their plasticity enables the small molecules to create appropriate pockets in which they can bind with higher affinity (*11*, *12*).

MATERIALS AND METHODS

The five steps of computational solvent mapping were described previously (*5*−*7*) and only a summary is given here.

Step 1: Rigid Body Search. The protein structures are downloaded from the Protein Data Bank (PDB). All bound ligands and water molecules are removed (*13*). For each structure, eight small molecules (acetone, acetonitrile, *tert*-butyl alcohol, dimethyl sulfoxide, phenol, methanol, 2-propanol, and urea) are used as probes (*7*). For each probe, 2000 docked conformations are generated by the rigid body docking algorithm GRAMM (*14*). GRAMM (*g*lobal *ra*nge *m*olecular *m*atching) requires only the atomic coordinates of the two molecules; i.e., no a priori information on the binding site is used. The program places each molecule on a separate grid and performs an exhaustive six-dimensional search through the relative intermolecular translations and rotations using a very efficient fast Fourier transform (FFT) correlation technique and a simple scoring function that measures shape complementarity and penalizes overlaps (*14*). We have used a 1.5 Å grid step for translations and 15° increments for rotations. A total of 2000 docked conformations were retained.

Step 2: Minimization and Rescoring. The free energy of each of the 2000 complexes, generated in step 1, is minimized using the free energy potential

$$\Delta G = \Delta E_{\text{elec}} + \Delta E_{\text{vdw}} + \Delta G_{\text{des}} \quad (1)$$

where ΔE_{elec}, ΔE_{vdw}, and ΔG_{des} denote the electrostatic, van der Waals, and desolvation contributions to the protein−probe binding free energy (*6*). The sum $\Delta E_{\text{elec}} + \Delta G_{\text{des}}$ is obtained by the *a*nalytic *c*ontinuum *e*lectrostatic (ACE) model (*15*), as implemented in version 27 of CHARMM (*16*) using the parameter set from version 19 of the program. The model includes a surface area dependent term to account for the solute−solvent van der Waals interactions. The minimization is performed using an adopted basis Newton−Raphson method as implemented in CHARMM (*16*). During the minimization the protein atoms are held fixed, while the atoms of the probe molecules are free to move. At most 1000 minimization steps are allowed, although most complexes require far fewer steps to achieve convergence.

Step 3: Clustering and Ranking. The minimized probe conformations from step 2 are grouped into clusters based on Cartesian coordinate information. The method creates an appropriate number of clusters such that the maximum distance between a cluster's hub and any of its members (the cluster radius) is smaller than half of the average distance between all of the existing hubs. We have slightly modified this algorithm by introducing an explicit upper bound U on the cluster radius to account for the physical dimensions of the different probe molecules. U is set equal to 2.0 Å for methanol, while a value of 4.0 Å is used for the other ligands. Clusters with less than 10 members are excluded from consideration. For each retained cluster, we calculate the probability $p_i = Q_i/Q$, where the partition function Q is the sum of the Boltzmann factors over all conformations, $Q = \sum_j \exp(-\Delta G_j/RT)$, and Q_i is obtained by summing the Boltzmann factors over the conformations in the ith cluster only. The clusters are ranked on the basis of their average free energies $\langle \Delta G \rangle_i = \sum_j p_{ij} \Delta G_j$, where $p_{ij} = \exp(-\Delta G_j/RT)/Q_i$, and the sum is taken over the members of the ith cluster.

Step 4: Determination of Consensus Sites. Mapping is primarily used to find consensus sites at which many different probe molecules cluster. To find the consensus sites, we select the minimum free energy conformation in each of the five lowest average free energy clusters for each solvent. The structures are superimposed, and the position at which most probes of different types overlap is defined as the main consensus site. An additional clustering of probes close to the main consensus site is likely to indicate another subsite of the active site.

[1] Abbreviations: CYP, cytochrome P450; PDB, Protein Data Bank; FFT, fast Fourier transform; GRAMM, *g*lobal *ra*nge *m*olecular *m*atching; ACE, analytical continuum electrostatics; RMSD, root mean square deviation; DMZ, dimethylsulfaphenazole; DIF, diclofenac.

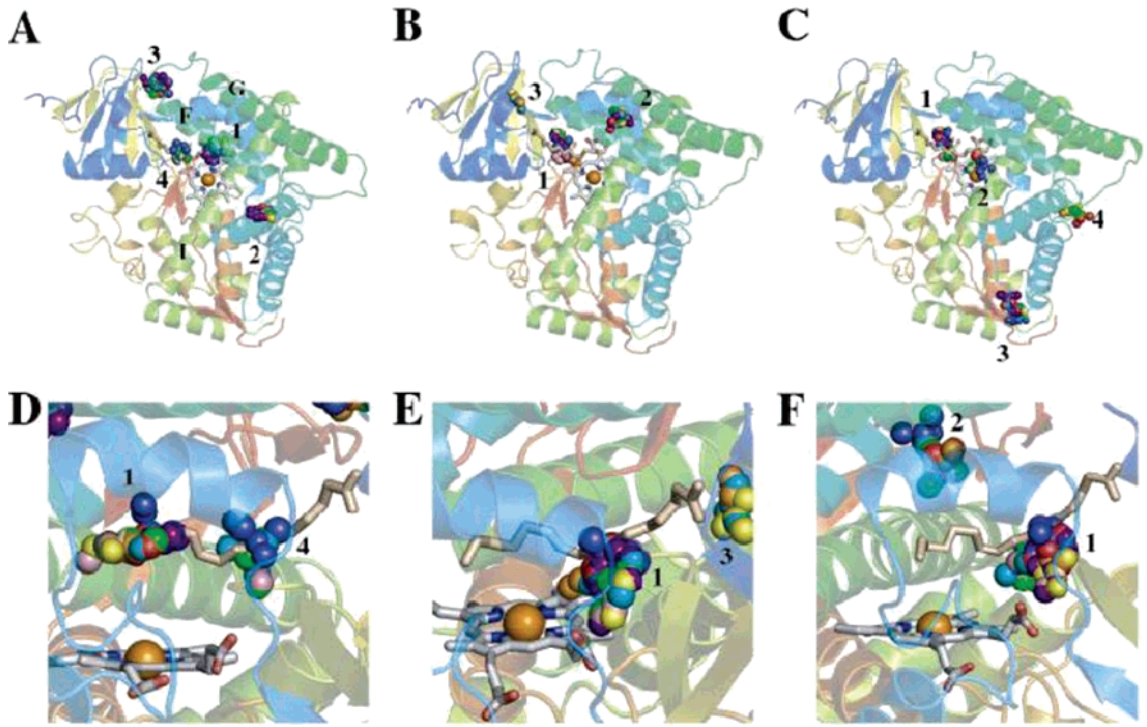

FIGURE 1: Mapping of P450 BM-3 structures. (A) P450 BM-3 structure 1FAG bound to palmitoleic acid. The F, G, and I helices are labeled. (B) Substrate-free P450 BM-3 in an open conformation (structure 1BU7, chain A). (C) Substrate-free P450 BM-3 in a closed conformation (1BU7, chain B). (D) Closer representation of the largest consensus site for 1FAG. The binding pocket is viewed from outside of the B−C loop toward helix I. (E) Closer representations of the largest consensus site for 1BU7 chain A. (F) Closer representations of the largest consensus site for 1BU7 chain B. The protein is shown in ribbon using PyMOL (*39*), with the color code from blue (N-terminus) to red (C-terminus). The heme is drawn with sticks, and iron is drawn as a sphere. The ligand (palmitoleic acid) is colored wheat and projected into the unbound structures based on its bound position. The organic molecules used as probes are colored as follows: acetonitrile (orange), acetone (teal), *tert*-butyl alcohol (blue), DMSO (red), phenol (purple), methanol (pink), 2-propanol (green), and urea (yellow). Each consensus site is labeled with its rank as annotated in Tables 1 and 3.

Ranking of the consensus sites is dependent on the number of different types of small molecules, with a maximum of eight, found within each consensus site; i.e., a consensus site containing seven of the eight different probe molecules is ranked better than a site with only four of the eight probe molecules. If two sites share the same number of different types, then duplicate types within the consensus site are considered in the count; i.e., a consensus site containing clusters from all eight different probe molecules and a second urea cluster is ranked better than a site containing a single cluster from all eight different probe molecules. Any sites that are still equivalent by these comparisons are then ranked by average energy rank of the probe clusters included at the site; i.e., a cluster where all eight probe clusters were rank 1, by average Boltzmann free energy for each probe type, is ranked better than a cluster where all eight probe clusters were rank 2 or lower.

Step 5: Subclustering and Determination of Nonbonded Contacts and Potential Hydrogen Bonds. For each ligand, the cluster at the consensus site is further divided into subclusters based on the RMSD between probe conformations in the cluster as well as differences in the free energies of the probe conformations in order to generate different subclusters for different mechanisms of binding at the same physical location (*5*). Each subcluster was represented by a single conformation with the lowest free energy. The LIGPLOT program (*17*, *18*) from Thornton and co-workers was used to find the nonbonded interactions and hydrogen bonds formed between each probe conformation and the protein. For each mapping result, the number of nonbonded interactions per residue was normalized by the number of total nonbonded interactions counted between all probes at the consensus site and the protein. Contacts between cocrystallized ligands and the protein were obtained from PRECISE (*pre*dicted and *c*onsensus *i*nteraction *s*ites in *e*nzymes) database at http://precise.bu.edu (*19*). The cocrystallized ligand contacts were normalized in the same manner as the consensus site results. Only residues with normalized percentages greater than 5% were considered for comparison between mapping and cocrystallized experimental results.

RESULTS

Mapping of Bacterial P450s. (A) Mapping of P450 BM-3. P450 BM-3 (CYP102) of *Bacillus megaterium* catalyzes NADPH-dependent oxidation of medium- and long-chain fatty acids, with optimum chain lengths of 14−16 carbons. Two high-resolution structures are available for the substrate-free P450 (*20*), one in an open conformation (*20*; PDB structure 1BU7, chain A) and one in a closed conformation (*20*; PDB structure 1BU7, chain B). Both substrate-free structures have a larger binding pocket than the palmitoleic acid-bound P450 BM-3 structure (*21*; PDB structure 1FAG). Computational mapping was applied to both 1BU7 chains and to the substrate-bound structure 1FAG after removal of the palmitoleic acid. Panels A, B, and C of Figure 1 show the largest consensus sites (i.e., the sites with the highest number of overlapping probe cluster representatives) for

Table 1: Free Energy Ranking of Probe Clusters Mapped to Bacterial P450s (CYP BM-3 and CYP CAM)

		solvent probe[b]							
protein structure	consensus site[a]	acetonitrile	acetone	*tert*-butyl alcohol	DMSO	phenol	methanol	2-propanol	urea
P450 BM-3 bound									
1FAG	**1**	**4 (8.23)**	**2 (7.98)**	**5 (6.47)**	**3 (7.81)**	**5 (7.35)**	**4 (8.09)**	**2 (7.70)**	**5 (7.96)**
	2	1 (18.77)	1 (19.12)		1 (18.89)	1 (18.17), 3 (18.92)	2 (19.32)	1 (18.96)	2 (19.86)
	3	3 (23.11)	4 (24.34)	1 (23.26)	4 (23.14)	4 (24.53)		4 (24.39)	
	4		**5 (8.96)**	**2 (8.61)**			**5 (9.07)**	**3 (8.81)**	
P450 BM-3 unbound									
1BU7 chain A, open conformation	**1**	**4 (4.91)**	**1 (9.37), 2 (6.56)**	**5 (7.79)**	**3 (7.14)**	**3 (9.30)**	**2 (10.09), 4 (8.68)**	**1 (8.97)**	**5 (8.63)**
	2		4 (23.26)		5 (21.88)	1 (22.38)		3 (23.25)	2 (21.46)
	3	3 (21.77)	5 (20.88)						3 (21.60)
1BU7 chain B, closed conformation	**1**	**4 (8.70)**	**3 (6.58)**	**4 (8.90)**	**3 (8.57)**	**3 (9.49)**	**2 (10.07), 5 (7.32)**	**1 (7.29)**	**3 (8.96)**
	2	2 (10.18)	1 (10.57)	2 (11.70)	1 (10.02)		1 (11.14)	2 (10.15)	
	3		2 (30.07)	1 (31.10)	2 (30.18)	1 (28.99), 2 (30.22)		3 (30.43)	5 (29.95)
P450 cam bound									
1DZ4	**1**	**1 (1.78)**	**2 (4.78)**	**1 (3.40)**	**1 (3.52)**	**1 (6.58)**	**5 (5.61)**	**1 (5.34)**	**5 (2.66)**
	2		1 (17.01)		3 (16.76)		1 (15.30)	2 (17.15)	3 (15.12)
	3			3 (20.49)	2 (20.53)			5 (20.55)	
	4		5 (19.46)		4 (18.83)		4 (18.25)		
P450 cam unbound									
1PHC	**1**	**1 (1.67)**	**1 (5.59)**	**1 (2.89)**	**1 (2.94)**	**1 (5.37)**	**1 (5.74)**	**1 (5.70)**	**5 (2.25)**
	2		5 (15.54)	2 (15.39)	3 (16.10), 2 (14.81)	5 (15.91)		5 (14.41)	
	3		2 (25.78)	5 (25.20)	4 (25.08)	3 (24.99)		2 (25.64)	

[a] Consensus site ranked according to the number of clusters contained. [b] Free energy rank of the probe cluster in the consensus site, where 1 denotes the lowest free energy cluster for each individual probe. The numbers in parentheses correspond to the distances in Å between the cluster center (the lowest free energy ligand conformation in the cluster) and the heme Fe atom. A cell with no entry indicates that the particular probe does not form a cluster at the consensus site. Entries shown in bold are within 5 Å of the ligand (see Table 3).

1FAG, 1BU7 chain A, and 1BU7 chain B, respectively, with a more detailed view of the heme and substrate-binding site presented in Figure 1D−F. The number of probe molecules clustered in each consensus site and their respective energy ranking are listed in Table 1. The distance from the center of each cluster to the heme iron atom is also shown (values in parentheses). For example, the first row of Table 1 shows that the largest consensus site (site 1) found by the mapping of 1FAG includes the fourth lowest free energy cluster of acetonitrile, located 8.23 Å from the iron atom, the second lowest free energy cluster of acetone, at 7.98 Å, and so on. A consensus site may include several clusters of the same probe, e.g., for phenol in consensus site 2 for 1FAG (Table 1). As shown in Table 1, the largest consensus site for solvent binding was localized to the substrate-binding pocket in all three P450 BM-3 structures, within 5 Å of the bound ligand position (bold entries in Table 1). Table 3 summarizes the consensus sites and shows the distances from the center of each consensus site to both the heme iron atom and the bound ligand. For ligand-free structures the distance to the ligand is calculated by superimposing a ligand-bound structure on the results of the mapping. In the case of 1FAG the fourth largest consensus site is also in the binding pocket (Figure 1D).

The nonbonded contacts and potential hydrogen bonds between the probe molecules of the largest consensus site and protein were calculated for each of the three mapping results for P450 BM-3, i.e., 1BU7 chain A, 1BU7 chain B, and 1FAG. The normalized percentages for nonbonded contacts per residue are shown in Figure 2. Percentages less than 5% of the total nonbonded contacts are not displayed. The nonbonded contacts for the two ligand-free conformations mapped are highly similar. The consensus sites for the two chains of 1BU7 contact the protein at residues S72, L75, and F87 on the helix B−helix C loop region as well as residues A328−S332 from the fourth β-sheet located on the opposite side of the binding pocket from the B−C loop. There are also contacts being made with L437 where the C-terminal loop extends into the binding pocket and the heme. The majority of the potential hydrogen bonds are to the side chain of S72, the backbone atoms of A330, and the heme (data not shown). Compared to the unbound conformations, there is a shift in the location of the consensus site in the bound conformation, 1FAG. The largest consensus site is located further into the pocket and thereby overlapping the site of enzymatic activity on the cocrystallized ligand (Figure 1D). This shift accounts for the change in contacts to include residues T260−A264 from helix I. The consensus site for 1FAG was also extended to include the probes composing the fourth largest consensus site. The contacts contributed by these additional probes are made with F87, A328, and A330 and thus agree with the same contacts from the mapping results for the unbound conformations. However, there are also a number of novel contacts made to V78, suggesting an increase in the hydrophobicity of the protein surface in the presence of the ligand.

In addition to the probe−protein interactions, Figure 2 shows the nonbonded contacts between the palmitoleic acid and the P450 BM-3 in the ligand-bound crystal structure normalized by the total number of contacts [experimental (1FAG)]. The intermolecular hydrogen bonds were also determined (not shown). The ligand extends across a wide

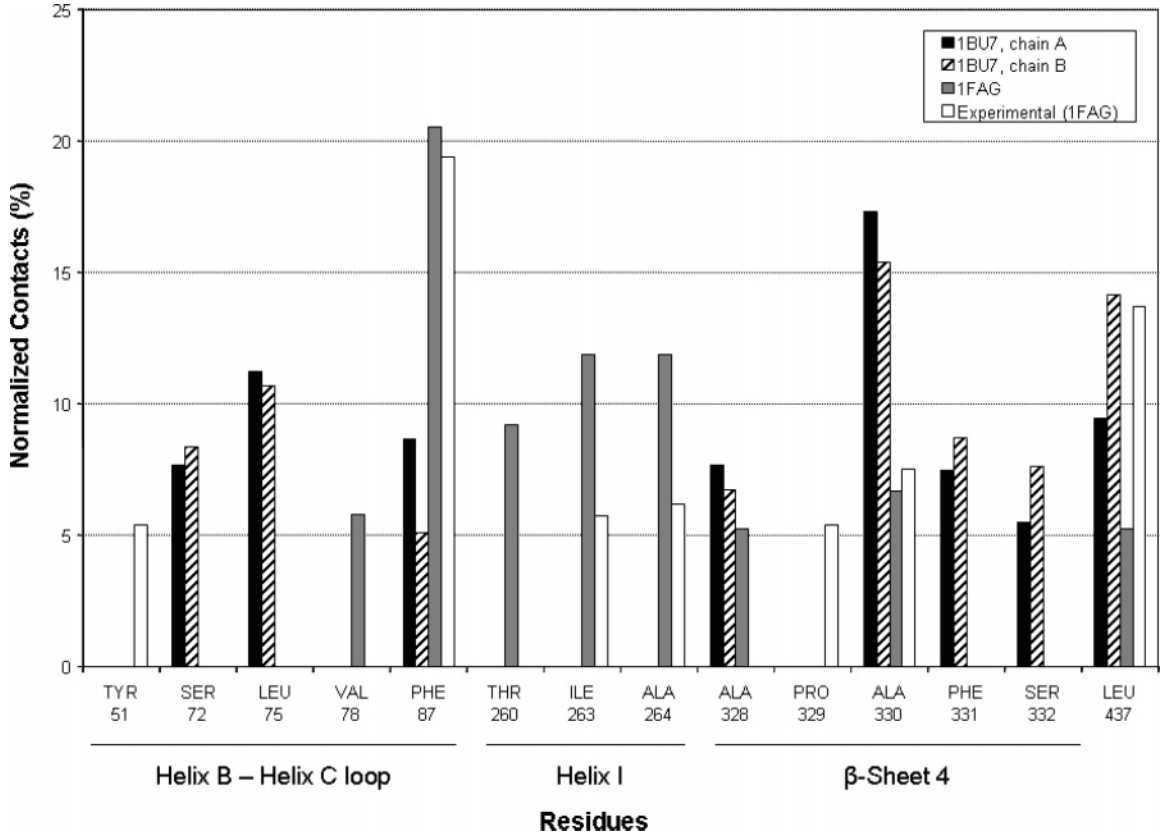

FIGURE 2: Distribution of normalized intermolecular nonbonded contacts for P450 BM-3 structures. The mapping results are based on the interactions found between the probes in the consensus site and the residues of the protein. The interactions are normalized on the basis of the total number of interactions for each mapping result, and only residues making at least 5% of the total contacts are shown (see Materials and Methods). The best consensus site is considered for the unbound P450 BM-3 structures 1BU7 chain A and 1BU7 chain B. The best consensus site and fourth best consensus site are considered for the ligand-bound conformation 1FAG without ligand present. The fourth best consensus site was considered because of its proximity to the ligand position and best consensus site position (Figure 1D). The nonbonded contacts for palmitoleic acid in the bound structure 1FAG were obtained from the database PRECISE (see Materials and Methods) and are shown for comparison to the mapping results. The structural references below the residues were included to help in orientation of the interactions within the binding pocket. L437 is located in the C-terminal loop. The figure shows the four distributions of interactions determined for P450 BM-3.

region of the binding pocket, from 6 to 7 Å above the iron atom of the heme toward Y51 on β-sheet 2, where it forms a hydrogen bond between the side chain of Y51 and the carboxylic acid headgroup of the fatty acid. Nonbonded interactions are formed with a large number of residues including Y51, F87, I263, A264, P329, A330, and L437 (Figure 2). Palmitoleic acid makes less than 5% of its total contact with the protein at 18 other residues (data not shown). F87 accounts for nearly one-fifth of the contacts with the ligand and is in close proximity to the double bond between the ω-6 and ω-7 carbon atoms, as well as the ω-1 to ω-3 carbon atoms where enzymatic activity takes place on the ligand. I263 and A264 on the I helix also contact the ω-1 to ω-3 carbon atoms from the opposite side of the binding pocket in relation to F87.

The consensus sites for both chains of the unliganded protein structure, 1BU7, heavily overlap in position within the binding pocket despite some conformational differences between the two chains. The probes contact the B−C loop region as well as β-sheet 4 backbone atoms, placing the consensus site halfway down the binding channel of the protein. F87 acts as a wall preventing the probes from reaching further into the active site. This is in sharp contrast to the consensus site and natural ligand positions found in the bound structure, 1FAG. In the latter structure, the χ_2 dihedral angle of F87 is increased by 101°, rotating the side chain into a conformation nearly parallel to the plane of the heme. The ligand, as well as the consensus site for 1FAG, is then able to fit aside F87, next to the I helix, and closer to the point of enzymatic activity above the heme iron atom. The position of the consensus sites seen in the two unbound structures remains occupied in 1FAG (consensus site 4 in Figure 1D and Table 1), even though the top consensus site has relocated deeper into the pocket. The probes appear to form this deeper consensus site in the bound but not in the unbound structure because ligand binding results in a narrower pocket that enables the probes to find favorable van der Waals interactions with the I helix, while also maintaining favorable contacts with the residues around the mid-region of the binding site as previously seen in the unbound structures (Figure 1).

As shown by the mapping results, the consensus sites do not necessarily cover the entire ligand-binding pocket, which is not surprising for BM-3 which can accommodate very large and elongated ligands. More generally, our experience shows that the consensus sites frequently identify the most important regions of the binding pocket and that this information can be utilized for the characterization of hot spots in the site (*7*). For the purposes of this paper, however, the most important observation is that the mapping reveals

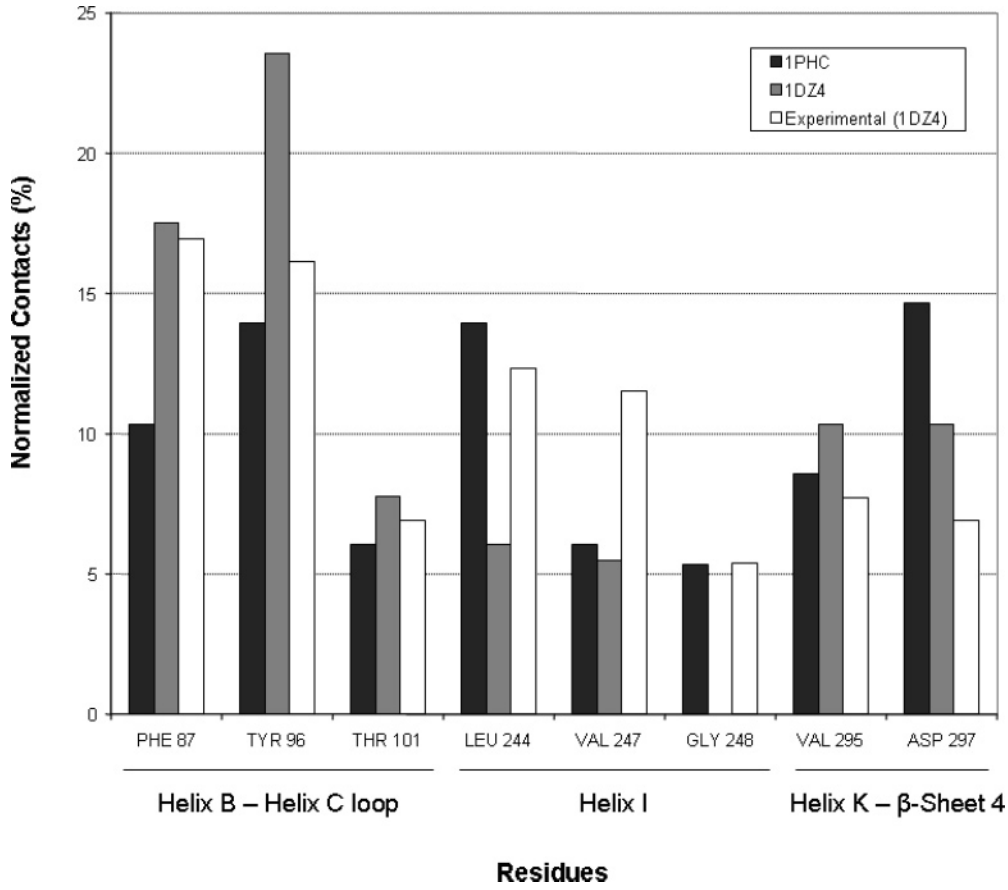

FIGURE 3: Distribution of normalized intermolecular nonbonded contacts for P450 cam structures. The mapping results are based on the interactions found between the probes in the consensus site and the residues of the protein. The interactions are normalized on the basis of the total number of interactions for each mapping result, and only residues making at least 5% of the total contacts are shown (see Materials and Methods). The best consensus site is considered for the unbound P450 cam structure 1PHC. The best consensus site is considered for the ligand-bound conformation 1DZ4 without ligand present. The nonbonded contacts for *d*-camphor in the bound structure 1DZ4 were obtained from the database PRECISE (see Materials and Methods) and are shown for comparison to the mapping results. The structural references below the residues were included to help in orientation of the interactions within the binding pocket. The figure shows the union of the three distributions of interactions determined for P450 cam.

well-defined, albeit not identical, binding sites in both the bound and unbound conformations of BM-3.

(B) Mapping of P450 cam. P450 cam is a camphor monooxygenase found in *Pseudomonas putida*, which enables the bacteria to use *d*-camphor as the sole carbon source. The natural ligand cocrystallizes in the active site 3.8 Å above the heme iron atom. In this position, the ligand forms nonbonded contacts with a number of hydrophobic residues, F87, L244, V247, G248, and V295 (Figure 3). It also finds favorable contacts with hydrophilic patches on the protein surface created by residues Y96, T101, and N297, and a hydrogen bond is formed between the oxygen atom of *d*-camphor and the side chain of Y96. Apart from a small repositioning of the F87 side chain, all of these residues show little rotation between the ligand-bound and unbound conformations of the protein. Thus, the binding pocket surface is stable and unchanging.

Solvent mapping was applied to substrate-free (*22*; PDB structure 1PHC) and substrate-bound (*23*; PDB structure 1DZ4) structures of P450 cam. The two structures are highly similar, with less than 1 Å root mean square deviation (RMSD). Solvent mapping yielded similar results for the two structures (Tables 1 and 3). In both cases the largest consensus site for probe binding was around 5 Å from the heme iron, directly overlapping with the ligand in the cocrystallized structure. These consensus sites were formed by clusters of all of the eight probes. Moreover, for most probes these clusters had the lowest average free energy. The nonbonded contacts and potential hydrogen bonds between the consensus sites in both conformations highly correlate with the contacts made between the natural ligand and protein as well (Figure 3). The probes make the greatest number of contacts with the heme (data not shown) but also contacted F87, Y96, T101, L244, V247, V295, and N297. The only residue where the probes showed a subtle difference in contact with the protein is in the absence of interactions between the consensus site and G248 from the mapping of the bound conformation 1DZ4. Even in this case, some contact with the probes (~2% of the normalized contacts) was observed, but the signal was not strong enough to meet the threshold for consideration. Thus, for P450 cam the small molecule probes localize to the substrate-binding site of the ligand-free P450 structure, indicating that the latter is well formed prior to substrate binding.

Mapping of Mammalian P450s. (A) Mapping of P450 2C5. P450 2C5 is a mammalian enzyme which catalyzes the oxidation of substrates that are structurally diverse and differ widely in size and functionality (*24*−*26*). This P450 has been

Table 2: Free Energy Ranking of Probe Clusters Mapped to Mammalian P450s (CYP 2C5, CYP 2C9, and CYP 2B4)

		solvent probe[b]							
protein structure	consensus site[a]	acetonitrile	acetone	*tert*-butyl alcohol	DMSO	phenol	methanol	2-propanol	urea
P450 2C5 bound									
1NR6 DIF	**1**		**3 (10.68)**	**1 (10.75), 3 (10.41)**	**3 (12.31)**	**4 (12.50)**	**5 (11.85)**	**3 (11.45), 5 (11.48)**	
	2		5 (20.75)	4 (21.17)	5 (20.79)	3 (19.58)			3 (19.72)
	3		1 (28.43)		2 (28.32)	1 (28.88)		1 (28.56)	5 (28.28)
	4	**1 (1.68)**		**2 (4.25)**	**4 (4.32)**				
1N6B DMZ	**1**	**5 (14.67)**	**4 (14.34)**	**3 (10.62)**	**4 (14.04), 5 (10.93)**	**5 (13.76)**		**4 (15.39)**	
	2		1 (12.86)		1 (11.81)	4 (11.90)	1 (13.48)	1 (13.09)	5 (12.73)
	3	**1 (1.76)**	**5 (4.66)**		**2 (3.94)**				
P450 2C5 unbound									
1DT6	1		5 (12.32)		2 (11.51)		2 (13.19)	3 (12.11)	4 (12.47)
	2		3 (23.33)	1 (21.59)		4 (23.65)		2 (23.20)	
	3	1 (19.36)	1 (19.53)		1 (19.02)		4 (19.89)		
P450 2C9 bound									
1R9O (FLP)	1	5 (18.78)	3 (15.82), 5 (19.31)	2 (17.32)	5 (18.33)	2 (18.08), 4 (16.10)		2 (18.87)	
	2			**1 (9.98)**	**1 (9.32)**		**2 (14.97)**	**3 (11.06)**	
	3	1 (18.47)	1 (17.49)						1 (16.24)
1OG5 (SWF absent)	**1**	**1 (11.46)**	**1 (12.40)**		**5 (9.64)**	**2 (11.02)**		**1 (11.95)**	
	2			1 (15.99)	1 (16.05)	1 (15.24)		3 (15.29)	
	3	4 (14.47)	3 (14.36)				2 (15.10)		3 (13.33)
	4	3 (9.59)	2 (9.56)			3 (9.14)	4 (8.99)		
	5	2 (8.47)	4 (7.53)				1 (9.18)		1 (7.64)
1OG5X (SWF present)	**1**	**1 (14.87)**	**1 (13.47)**	**2 (9.50)**	**1 (11.76), 5 (13.28)**	**2 (12.17), 3 (12.11)**		**1 (12.76)**	**2 (14.42)**
	2	4 (9.56)	2 (9.55)			4 (9.12)	4 (8.97)		
	3	2 (8.47)	3 (7.52)				1 (9.16)		1 (7.70)
	4			1 (16.00)	2 (17.37)	1 (14.99)		2 (15.29)	
P450 2C9 unbound									
1OG2	1	2 (16.28)	1 (17.08)		1 (15.02)	2 (15.49)		1 (16.19)	4 (18.07)
	2	5 (13.81)	3 (14.77)				1 (12.28)		2 (12.01)
	3	3 (13.48)			3 (13.55)		3 (12.77)	3 (12.90)	
	4			5 (10.59)	4 (11.22)	4 (10.21)		5 (12.03)	
P450 2B4 bound									
1SUO CPZ	**1**	**4 (6.14)**	**5 (6.40)**	**2 (3.71)**	**2 (3.67)**	**3 (5.85)**		**5 (6.70)**	
	2	2 (8.62)	4 (8.63)	1 (8.02)	3 (8.53)			3 (7.93)	
	3		2 (9.59)	3 (9.64)	4 (9.67)	1 (9.34)		2 (9.43)	
	4			5 (23.42)	5 (22.68)	5 (21.49)		4 (22.07)	5 (20.10)
P450 2B4 unbound									
1PO5	1		1 (13.82)	1 (13.06)	1 (14.14)	5 (13.66)	1 (15.71)	1 (13.79)	
	2		3 (26.31)	2 (25.96)	4 (26.00)			2 (26.49)	

[a] Consensus site ranked according to the number of clusters contained. [b] Free energy rank of the probe cluster in the consensus site, where 1 denotes the lowest free energy cluster for each individual probe. The numbers in parentheses correspond to the distances in Å between the cluster center (the lowest free energy ligand conformation in the cluster) and the heme Fe atom. A cell with no entry indicates that the particular probe does not form a cluster at the consensus site. Entries shown in bold are within 5 Å of the ligand (see Table 3).

crystallized in the absence of ligand (*24*; PDB structure 1DT6), bound to diclofenac (DIF; see ref *25*; PDB structure 1NR6), and bound to dimethylsulfaphenazole (DMZ; see ref *26*; PDB structure 1N6B). The two ligands partially overlap in the active site, with both located proximal to the heme within the binding pocket and extending alongside the I helix and terminating just below the G helix, between the B′ helix of the B−C loop and the F helix. The size of the ligands allows them to contact a range of residues throughout the binding pocket, including multiple aliphatic residues in the B−C loop, I helix, and C-terminal loop. It has been observed that binding significantly changes the size and hydrophilicity of the binding site (*25*), as several polar patches are rearranged to accommodate either DIF or DMZ that substantially differ both in size and in polarity (*25*, *26*). In particular, D290 on the I helix, known to be involved in a network of water molecules responsible for binding DIF, shows a χ_2 bond rotation of ~140° allowing it to point into the binding pocket in the presence of DIF but not in the unbound structure or the DMZ-bound structure. On helix F, N204 acts in a similar fashion by rotating to point toward the bound DIF but points toward helix G in the absence of ligand or in the presence of the larger ligand, DMZ (rotations of ~60° for both the χ_1 and χ_2 dihedral angles). Furthermore, when P450 2C5 is bound to DMZ, the side chain of N204 spins along the χ_2 dihedral angle to place the amide nitrogen in interaction with the π orbitals of the phenyl group of DMZ.

The ligands were removed for the two ligand-bound structures, and all three structures were mapped. For the two ligand-bound structures, the largest consensus sites were found in the binding pocket (Tables 2 and 3) and were superimposable with the distal part of the bound ligands (Figure 4A,B). These consensus sites formed nonbonded contacts with the protein along the F and I helices as well as the B−C loop (Figure 5). The consensus site from the DMZ-bound structure 1N6B primarily contacts the amide

Table 3: Summary of the Consensus Sites of Small Molecule Clusters

protein structure[a]	cons site[b]	no. of clusters[c]	distance to Fe[d]	distance to ligand[e]
bacterial P450 BM-3 bound				
1FAG (PAM)	**1**	**8**	**8.46**	**0.81**
	2	7 (8)	19.97	17.31
	3	6	25.09	9.44
	4	**4**	**9.82**	**0.60**
bacterial P450 BM-3 unbound				
1BU7 chain A, closed	**1**	**8 (10)**	**9.03**	**2.05**
	2	5	23.68	16.10
	3	3	22.61	6.22
1BU7 chain B, open	**1**	**8 (9)**	**9.57**	**2.49**
	2	6 (7)	31.88	33.07
	3	6	12.22	6.19
	4	4	26.41	25.19
bacterial P450 cam bound				
1DZ4 (CAM)	**1**	**8**	**5.08**	**0.55**
	2	5	17.97	10.50
	3	3	20.30	20.70
	4	3	22.01	20.30
bacterial P450 cam unbound				
1PHC	**1**	**8**	**4.97**	**0.77**
	2	5 (6)	16.87	14.12
	3	5	26.67	28.81
mammalian P450 2B4 bound				
1SUO (CPZ)	**1**	**6**	**6.21**	**0.72**
	2	5	11.07	12.72
	3	5	9.67	11.64
	4	5	23.57	14.19
mammalian P450 2B4 unbound				
1PO5	1	6	15.15	12.40
	2	4	27.25	18.73
mammalian P450 2C5 bound				
1NR6 (DIF)	**1**	**6 (8)**	**12.55**	**1.85**
	2	5	21.87	23.34
	3	5	29.57	31.94
	4	**3**	**5.23**	**1.21**
1N6B (DMZ)	**1**	**6 (7)**	**14.87**	**2.95**
	2	6	13.75	7.44
	3	**3**	**4.84**	**4.46**
mammalian P450 2C5 unbound				
1DT6	1	5	13.25	16.70
	2	4	24.37	13.12
	3	4	20.52	19.03
mammalian P450 2C9 bound				
1R9O (FLP)	1	6 (8)	18.87	9.09
	2	**4**	**12.74**	**0.71**
	3	3	18.82	15.27
1OG5 (SWF absent)	**1**	**5**	**12.31**	**0.36**
	2	4	17.43	13.11
	3	4	15.40	9.41
	4	4	10.58	17.15
	5	4	8.85	11.75
1OG5X[f] (SWF present)	**1**	**7 (9)**	**14.51**	**4.66**
	2	4	10.56	17.15
	3	4	8.84	11.74
	4	4	17.39	13.17
mammalian P450 2C9 unbound				
1OG2	1	6	17.91	6.50
	2	4	14.43	7.92
	3	4	14.40	15.18
	4	4	12.21	8.99

[a] Ligands of each protein are shown in parentheses: PAM, palmitoleic acid; CAM, *d*-camphor; CPZ, 4-(4-chlorophenyl)imidazole; DIF, diclofenac; DMZ, dimethysulfaphenazole; FLP, fluriprofen; SWF, *S*-warfarin. [b] Consensus sites ranked according to the number of clusters contained. [c] The number of different probe clusters. Numbers in parentheses denote the total number of clusters, including multiple clusters of the same probe. The consensus sites in bold are within 5 Å of the ligand. [d] Distance from the geometric center of the consensus site to the heme iron. [e] Distance from the geometric center of the consensus site to the closest atom of the bound ligand. The bound ligand position is superimposed to calculate this distance for PDBs of unbound structures. Both available ligand orientations are used in the case of 1N6B. [f] Mapped in the presence of a bound *S*-warfarin molecule.

nitrogen of the N204 side chain, including a large number of potential hydrogen bonds between the side chain and acceptor groups on the probes. The same contacts with N204 are made by the consensus site mapped to the DIF-bound structure 1NR6, but the probes extend deeper in the pocket (2−4 Å closer to the heme) and thereby also form contacts with residues S289−G293 in the I helix. A separate, smaller consensus site also overlaps the end of the ligands closest to the heme in both 1NR6 and 1N6B (sites 4 and 3, respectively, in panels D and E of Figure 4). These latter interactions are not shown in Figure 5.

In contrast to the two bound structures, mapping of the unbound P450 2C5 structure did not yield any consensus site overlapping the position of the ligand. The largest consensus site (site 1 in Figure 4C,F) contained only five of the eight probe molecules and was found in an exterior pocket 10 Å from the cysteine-coordinated side of the heme, far from the active site. None of the other consensus sites localized to the substrate-binding pocket. The second best consensus site (site 2 in Figure 4C,F) was near the F and G helices but was not at any substrate recognition site (*27*).

(B) Mapping of P450 2B4. The X-ray structure of mammalian P450 2B4 cocrystallized with the specific inhibitor 4-(4-chlorophenyl)imidazole in the active site was recently determined (*28*; PDB structure 1SUO). Solvent mapping of 1SUO (after removing the ligand) placed the largest consensus site above the heme and within the binding pocket, overlapping with the position of the inhibitor (Tables 2 and 3). The probe molecules that constitute this consensus site make contacts primarily with residues in helix I. The crystal structure of unbound P450 2B4 (*29*; PDB structure 1PO5) reveals a large open cleft that extends from the protein surface directly to the heme iron between the α-helical and β-sheet domains. Since the active site is wide open, it is not surprising that the mapping of this structure did not place any probe molecules within the active site (Tables 2 and 3). The largest consensus site is located between helix I and the C-terminal loop.

(C) Mapping of P450 2C9. The crystal structures of a human P450 2C9 have been determined unbound (*30*; PDB structure 1OG2) and in complex with *S*-warfarin (*30*; PDB structure 1OG5) and with flurbiprofen (*31*; PDB structure 1R9O). The first two structures are based on a chimeric enzyme containing seven amino acid substitutions (*30*) and have short helices B′ and G′, respectively, in the B−C and F−G loops. In contrast, the flurbiprofen-bound structure 1R9O is based on the native protein (*31*) and does not have the short helices in the loop regions, resulting in a substantially broader entrance channel to the substrate-binding site than in the other two structures. In the other bound structure 1OG5, *S*-warfarin is located in a distal subpocket of the P450 2C9 binding pocket. It contacts residues in the B−C loop, such as I99, F100, A103, and F114, as well as L366 and P367 just prior to β-sheet 4 and F476 on the C-terminus loop (Figure 7). The interactions with the phenylalanine side chains F114 and F476 are of particular interest because of the π−π orbital stacking available between F476 and the phenyl ring of the *S*-warfarin structure, as well as F114 and the fused ring group of the *S*-warfarin.

The largest consensus site found by solvent mapping of the warfarin-bound structure 1OG5 overlaps the warfarin-binding site (Tables 2 and 3 and Figure 6A). Since the bound

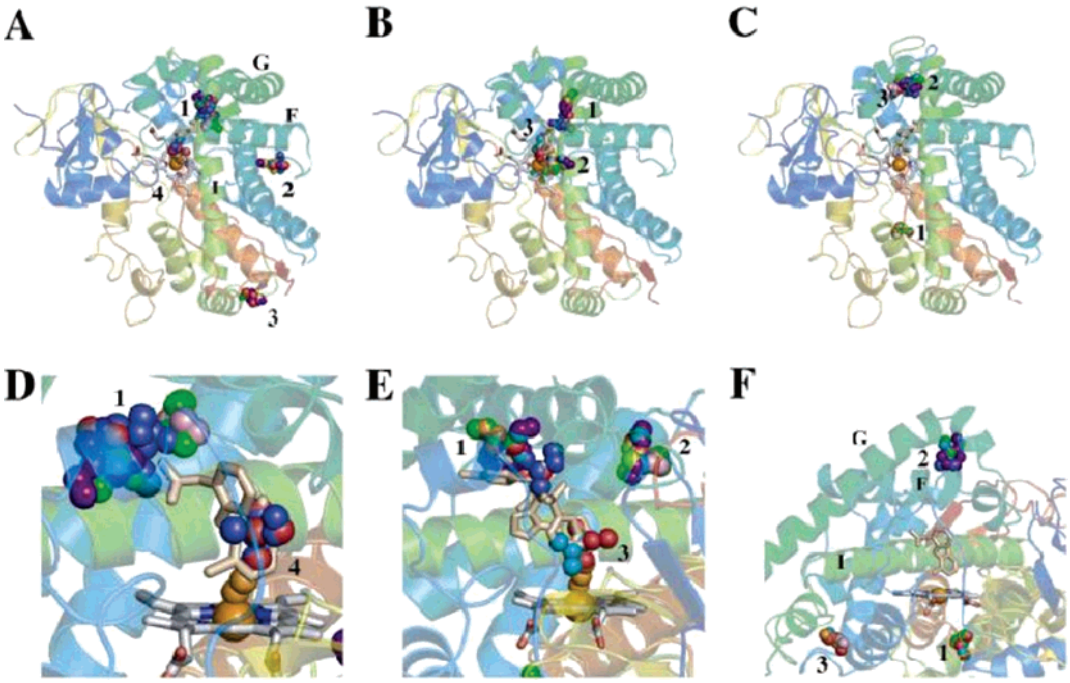

FIGURE 4: Mapping of P450 2C5 structures. (A) P450 2C5 bound to diclofenac (structure 1NR6). The F, G, and I helices are labeled. (B) P450 2C5 bound to DMZ (1N6B). (C) Substrate-free P450 2C5 (1DT6). (D) Closer view of the 1NR6 binding pocket. The binding pocket is presented from outside of the B−C loop toward helix I. (E) Closer view of the 1N6B binding pocket. Both orientations of the ligand available in the PDB file are shown. (F) Closer representation of the 1DT6 binding pocket showing that no consensus site overlaps the ligand. The F, G, and I helices are labeled. The protein, ligand, and small molecules are colored and labeled as in Figure 1.

S-warfarin is likely to be too distant for hydroxylation to occur (*30*), this pocket may not be part of the catalytic site. Nevertheless, it provides a site for binding various molecules, such as *S*-warfarin and five of the eight organic solvents used as probes. The solvent probes primarily interacted with the B′ helix and nearby residues in the B−C loop. Specifically, F114 forms the greatest number of nonbonded contacts with the consensus site (Figure 7). There is also substantial contact between the probes and residues R97−I99 and A103 in the B−C loop. Furthermore, there are nonbonded contacts to residues L366 and P367 next to β-sheet 4 on the opposite side of the binding pocket. There is substantial overlap between the contacts made by the ligand and consensus site. F100 and F476 stand out as exceptions where the *S*-warfarin molecule is making contacts, but it is not present in the consensus site contact list. Both of these phenylalanines are contacting parts of the molecule other than the fused ring group. The consensus site directly overlaps only the fused ring (Figure 6). The consensus site is also finding good contact with aliphatic residues in the B−C loop where less substantial contact is made with the *S*-warfarin. Mapping of the flurbiprofen-bound structure 1R9O placed the largest consensus site at the entrance of the binding channel between the B−C and F−G loops. It is important to note that in the chimeric protein (1OG5) a large fraction of this region is occupied by the B′ and G′ helices (Figure 6C). However, the second largest consensus site for 1R9O is in the flurbiprofen-binding pocket, and thus, solvent mapping still finds the ligand-binding site.

In contrast to the bound structures, no consensus site was located near the binding pocket for the unbound P450 2C9 (1OG2) (Figure 6B), despite the close similarity between this structure and the structure with the bound *S*-warfarin (1OG5). The largest consensus site in the unbound structure was adjacent to the F−G loop, and a smaller consensus site was located in a position where it can make contacts with the F and I helices. The potential origin of this qualitative difference between the mapping results for bound and unbound structures will be discussed further in the paper.

(D) Mapping of P450 2C9 with S-Warfarin Bound. As mentioned, although *S*-warfarin is a substrate for P450 2C9, *S*-warfarin bound to P450 2C9 structure 1OG5 is too distant from the heme for hydroxylation to occur. To explain this discrepancy, *S*-warfarin was proposed to move from the primary binding site shown in the X-ray structure toward the heme (*30*). Alternatively, the relatively large (1833 Å^3) binding pocket of this P450 may simultaneously accommodate multiple ligand molecules, with the tightly bound *S*-warfarin seen in the crystal structure interacting with the P450 to form a second, more heme-proximal substrate-binding site (*30*). To test the latter hypothesis, we mapped the bound P450 2C9 structure in the presence of *S*-warfarin (structure identified as 1OG5X in Tables 2 and 3). Inclusion of the ligand introduces changes in the protein residues in the binding pocket available for interactions with probe molecules. The resulting largest consensus site is comprised of seven of the eight probe molecules, including two DMSO clusters and two phenol clusters, and is located between the F and F′ helices, the C-terminal loop, and the bound *S*-warfarin molecule. Nonbonded contacts between the probes and the protein occur with residues L208 and Q214 in the F−G helix−loop−helix and N474, F476, and A 477 in the C-terminal loop. However, the primary point of contact is the *S*-warfarin molecule (Figure 7). The location of the consensus site is more orthogonal to the heme than the *S*-warfarin-binding site in the X-ray structure (Figure 6D), and it has higher propensity for substrate binding (i.e., binds a larger number of probe clusters) than the best consensus

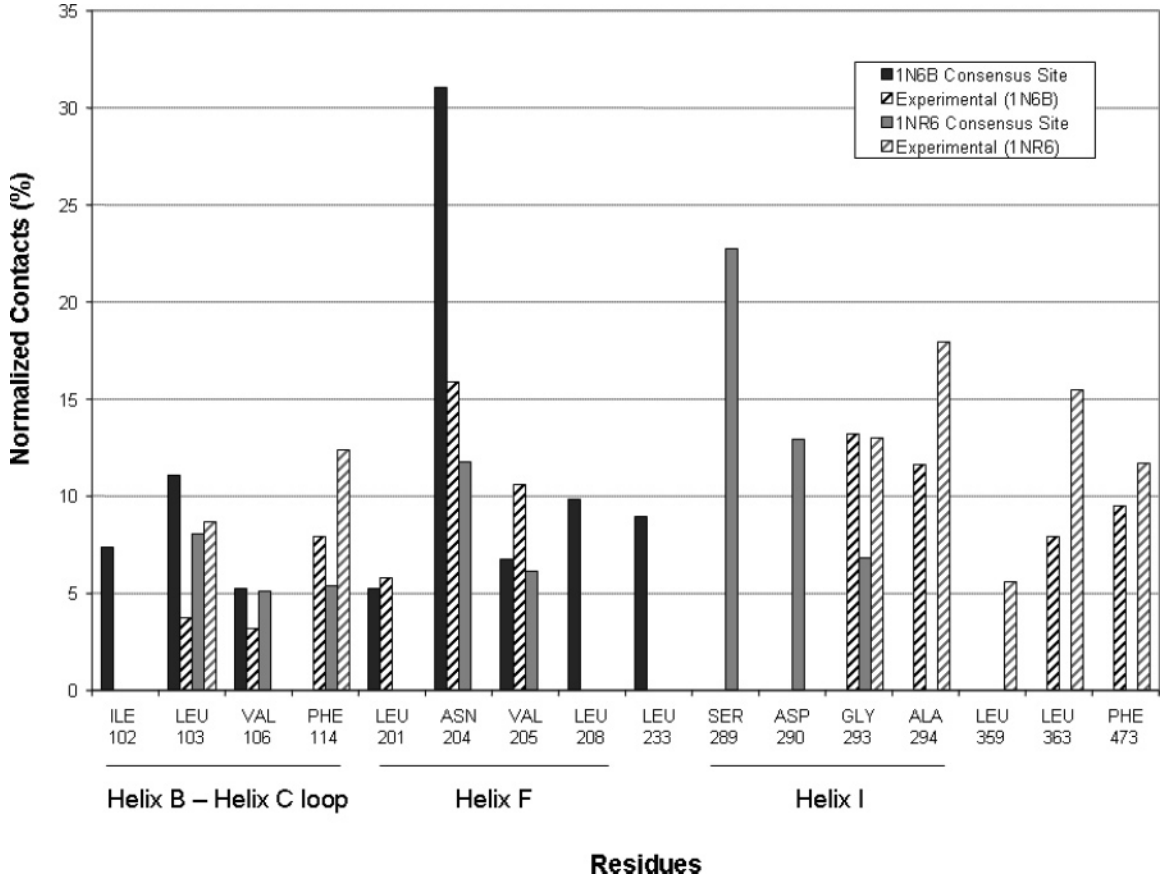

FIGURE 5: Distribution of normalized intermolecular nonbonded contacts for P450 2C5 structures. The mapping results are based on the interactions found between the probes in the consensus site and the residues of the protein. The interactions are normalized on the basis of the total number of interactions for each mapping result, and only residues making at least 5% of the total contacts are shown (see Materials and Methods) and are shown for comparison to the mapping results. The best consensus site is considered for the dimethylsulfaphenazole-bound P450 2C5 structure 1N6B without ligand present. The best consensus site is considered for the diclofenac-bound conformation 1NR6 without ligand present. The nonbonded contacts for dimethylsulfaphenazole and diclofenac in the bound structures, 1N6B and 1NR6, respectively, are shown with a striped bar of the same shade and to the right of the consensus site interactions for the respective structures. The contacts for the ligands were obtained from the database PRECISE (see Materials and Methods). The structural references below the residues were included to help in orientation of the interactions within the binding pocket. L233 is located on the G helix. L359 and L363 are located between helix K and β-sheet 4. F473 is located on the C-terminal loop. The figure shows the union of the four distributions of interactions determined for P450 2C5.

site obtained for the bound structure 1OG5 without *S*-warfarin present in the pocket. The bound *S*-warfarin not only creates a better defined pocket above the heme but itself has a number of interactions with the probes.

DISCUSSION

Computational solvent mapping was applied to two bacterial P450 proteins that act on relatively narrow classes of substrates, P450 cam and P450 BM-3, and to three mammalian P450s (P450s 2C5, 2B4, and 2C9), each capable of oxidizing a wide range of substrates. Qualitatively different results were obtained for the two groups of P450s. In the case of the two bacterial P450s, the probes clustered in the substrate-binding site when mapping was carried out using either ligand-bound or unliganded P450 structures, despite the fact that substrate binding is associated with a substantial large-scale conformational change of the F−G loop in the structure of P450 BM-3 (24). In contrast, for P450s 2C5, 2B4, and 2C9, the binding site was only found in the ligand-bound P450 structures; the unliganded proteins did not retain any probe clusters in the region of the binding site. This is particularly interesting for 2C9 because its bound and unbound structures are very similar (*30*). Thus, despite the presence of a visible cavity in the unliganded structures of the mammalian P450s, the site is unable to bind small molecules without adjusting its shape.

P450s with Narrow Substrate Specificity. The similarity of mapping results for bound and unbound structures is not surprising for P450 cam, because this P450 has a well-defined camphor-binding pocket, and no conformational changes result from substrate binding other than a small repositioning of a F87 side chain that contacts the *d*-camphor substrate (*22*). In contrast, P450 BM-3 has an open substrate access channel near the loop between helices F and G in the substrate-free structure 1BU7 (20). According to the Pocket-Finder algorithm (see http://www.bioinformatics.leeds.ac.uk/pocketfinder/), the volumes of this channel are 2155 and 2010 Å^3 in the two chains of 1BU7. The binding of the elongated fatty acid substrate leads to major conformational changes in some of the loop regions (*21*) and reduces the volume of the binding channel to 933 Å^3 in 1FAG. Despite this large change, P450 BM-3 behaves as P450 cam and any typical enzyme in the sense that solvent mapping places the consensus site with the highest number of probe clusters at

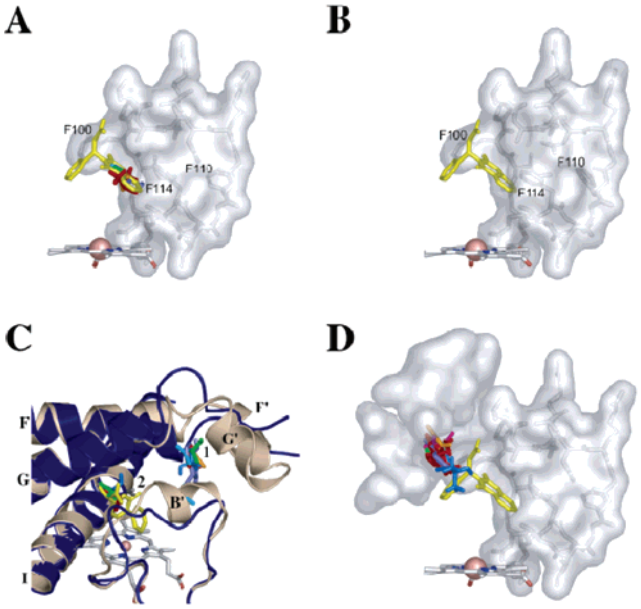

FIGURE 6: The binding pockets of P450 2C9. (A) P450 2C9 bound to *S*-warfarin (PDB structure 1OG5). (B) Substrate-free P450 2C9 (1OG2). (C) Comparison of unbound P450 2C9 (1OG2) and P450 2C9 bound to flurbiprofen (1R9O). The unbound structure is shown in tan, and the bound structure is shown in blue. The major helices shown are labeled, and the consensus sites are numbered by their rank in Table 2. The heme is shown as a gray stick figure with the heme iron atom as a sphere. Flurbiprofen is shown as a yellow stick figure. The B′−C′ loop and F′−G′ loop are shown, along with portions of the F, G, and I helices. (D) P450 2C9 bound and mapped in the presence of *S*-warfarin (1OG5X). In panels A, B, and D, residues 96−115 of the P450's B−C loop are shown with gray surface and stick representations. Panel D also includes P450 2C9 residues 209−217 and 474−478 in surface representation. The heme is represented by sticks and the heme iron by a sphere. *S*-Warfarin is shown inserted into the active site in yellow stick representation. In all four panels, the small molecules are shown in stick representations colored as in Figure 1, except that urea is colored magenta.

the substrate-binding sites in both the ligand-free and ligand-bound structures. As described earlier, the only difference is that the narrowing of the binding channel in the bound structure allows the main consensus site to move deeper toward the heme, overlapping with the proximal end of the palmitoleic acid (Figure 1D). However, a consensus site exists around the mid-region of the channel in all three structures (Figure 1).

Our earlier analyses indicate that the clustering of solvent probe molecules depends on three conditions (*7*). First, the probes cluster only in pockets that are deep and tight enough to surround the probes, providing a substantial number of probe−protein interactions. Second, the binding pockets must be partially nonpolar, because the hydrophobic interactions provide important contributions to the binding free energy. Third, the presence of several polar patches in the binding site is very important, because it enables the probes to bind in a number of rotational and translational states (*5*). The size condition is clearly satisfied in P450 cam, and the binding of the probes in the ligand-free P450 BM-3 indicates that, around the mid-region of the large binding channel, there is an appropriate site that provides the required probe−protein interactions. A number of hydrophobic residues have been identified in P450 BM-3 as contacting the probes, such as L75, F87, A328, A330, and L437, and polar patches are established in the binding site by residues S72 and S332. The same is true for P450 cam where hydrophobic residues F87, L244, V247, and V295 surround the pocket and residues Y96, T101, and D297 provide the polar patches, Y96 also forming hydrogen bonds with the probes. Since the probes cluster in the binding sites of P450 cam and P450 BM-3, we conclude that these proteins satisfy all three conditions even in their unbound states.

P450s with Broad Substrate Specificity. Our mapping of P450s 2C5, 2B4, and 2C9 gave distinctly different results for ligand-free and ligand-bound structures. For the latter, the largest consensus site was always in the substrate-binding site, at a position that overlaps with the bound ligand. However, for the unliganded structures the solvent probes did not cluster in the active site. While this implies that the binding pocket does not satisfy one or more of the conditions for solvent binding outlined in the previous section, it is of great interest to identify the particular conformational changes between bound and unbound structures that may be responsible for the inability of the latter to bind small organic molecules. We suggest that, prior to ligand binding, the binding sites of these broad substrate specificity P450s are very large and have relatively few distinct structural features and thus do not provide pockets suitable for binding the relatively small solvent ligands. It is not at all surprising that the active site of the unbound P450 2B4 (1PO5) does not bind probes, as the structure reveals a large open cleft rather than a well-defined binding pocket around the heme (*29*). The origin of nonbinding is more difficult to identify in the two other mammalian P450s.

As we discussed, the binding sites of the three P450 2C5 structures (unbound, DIF-bound, and DMZ-bound) show substantial differences, as a number of residues are rearranged to accommodate the two ligands that substantially differ both in size and in polarity (*25*, *26*). The side chain of N204 is particularly important both for the binding of the ligands and for the binding of the probes. N204 not only maintains a different conformation for each mapped structure but also appears to be the most similar point of contact between the three structures in terms of the consensus sites. As mentioned, this residue accounts for the greatest number of contacts with the probes in the DMZ-bound structure 1N6B and fewer but still a substantial number of contacts in the DIF-bound structure 1NR6 (Figure 5). In 1N6B the amide nitrogen of N204 points toward the ligand and due to its partial positive charge appears to interact with the π orbitals of the phenyl group in DMZ. The carbonyl oxygen of N204 couples to the amide nitrogen of another asparagine, N236. The coupling of N204 and N236 side chains fixes the former in a stable conformation and creates a well-defined subsite with hydrophobic walls containing the side chains of L201, V205, and L208 from the F helix, L233, A237, and I240 from the G helix, and a positively charged bottom with the nitrogen of N204. A similar coupling is maintained in the diclofenac- (DIF-) bound structure, although in this instance the two asparagine residues rotate along the χ_1 dihedral to enter further into the binding pocket. This allows the N204 side chain to maintain contacts with DIF which is smaller than DMZ, as well as with a network of water molecules, but yields a shallower pocket which appears to be less favorable for probe binding than the deeper site in

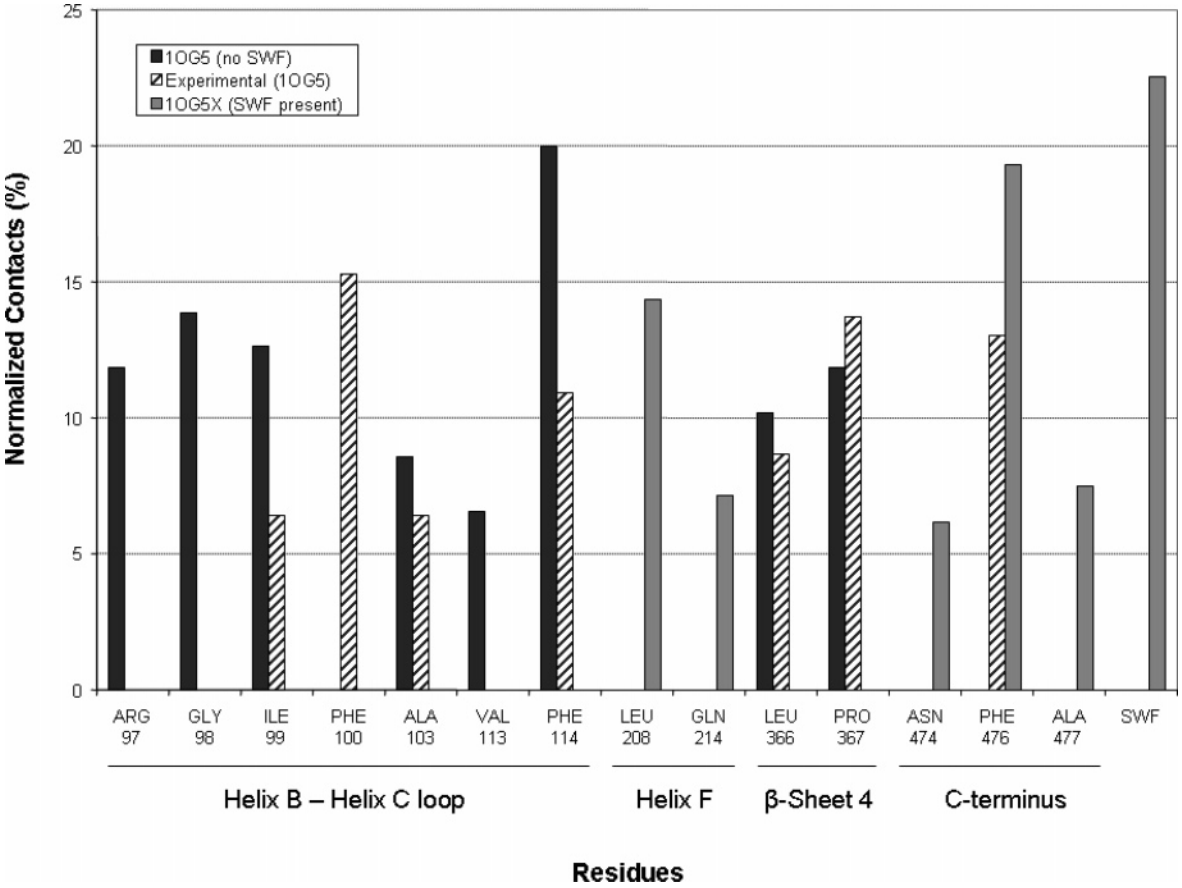

FIGURE 7: Distribution of normalized intermolecular nonbonded contacts for P450 2C9 structures. The mapping results are based on the interactions found between the probes in the consensus site and the residues of the protein. The interactions are normalized on the basis of the total number of interactions for each mapping result, and only residues making at least 5% of the total contacts are shown (see Materials and Methods). The best consensus site is considered for the *S*-warfarin-bound P450 2C9 structure 1OG5, without ligand present (1OG5) and with ligand present (1OG5X). The nonbonded contacts for *S*-warfarin in the bound structure 1OG5 are obtained from the database PRECISE (see Materials and Methods) and are shown for comparison to the mapping results. The structural references below the residues were included to help in orientation of the interactions within the binding pocket. SWF denotes the intermolecular contacts found between the probes and *S*-warfarin for 1OG5X, where *S*-warfarin was retained as part of the protein. The figure shows the union of the three distributions of interactions determined for P450 2C9.

the DMZ-bound structure 1N6B (Figure 5). Thus, it is likely to be important that in the unbound bound structure both N204 and N236 side chains rotate nearly 180° around their χ_2 dihedral angles relative to their bound forms, placing now the carbonyl oxygen of N204 rather than the amide nitrogen in the pocket. The negatively charged oxygen does not form interactions with the π orbitals of aromatic groups, and it also appears to be unable to form hydrogen bonds with the appropriate donor groups of the polar probes while the probes interact with the hydrophobic walls of the subsite, and thus N204 does not interact with any probe in the unbound structure.

Several other side chains show differences between unbound and bound structures. Hydrophobic residues L103, F114, and F473 are displaced the most in the binding site in the DMZ-bound conformation as compared to the other two crystal structures. More so, L103 is the hydrophobic residue with the greatest contact to the consensus site resulting from the mapping of 1N6B. Interactions with the ligand also seem to bring a number of polar groups deeper into the pocket, which yields a smaller and less hydrophobic cavity in the bound state. Indeed, the volumes of the binding pockets in DIF-bound (1NR6) and in DMZ-bound (1N6B) structures of 2C9 are 1233 and 847 Å^3, whereas in the unliganded 2C9 structure (1DT6) the volume of the largest pocket above the heme is 2290 Å^3, although the unusually large volume may be partially due to some unordered regions in this structure. Thus, we conclude that ligand binding induces substantial conformational changes, although the binding site in the unbound structure is large enough to accommodate substrates without any change. In fact, the site is so large that each of the substrates occupies only a portion of it, with the remainder of the cavity filled by ordered networks of water molecules (*26*), and it is clear that the volume also has to be reduced for effective ligand binding.

In contrast to P450 2C5, ligand binding to P450 2C9 does not reduce the volume of the channel, and we even see a slight increase, from 1635 Å^3 in chain A of the unbound structure 1OG2 to 1646 Å^3 in chain A of the warfarin-bound structure 1OG5. However, the comparison of the two structures emphasizes the importance of the induced fit mechanism for creating subpockets within the large binding site, since such subpockets appear to be necessary for binding the small ligands, including the probe molecules used in this study. This phenomenon is illustrated in Figure 6A,B for the P450 2C9 binding site, which is almost unchanged upon ligand binding (*30*) and yet binds organic solvents only in the ligand-bound conformation. Although the cavity of the

unliganded P450 2C9 is already so large that it could accommodate two *S*-warfarin molecules simultaneously, amino acid residue F114 is oriented in the unbound structure such that its phenyl ring would overlap with the bound ligand (Figure 6B). The φ angle of F114 is rotated 23.3° and the χ_1 dihedral angle is rotated 17.4° from the unbound conformation to the bound conformation. The rotation of F114 causes the largest displacement in the subpocket surface and creates a local pocket that binds the fused ring group of *S*-warfarin (Figure 6A). Moving the F114 side chain away from the binding channels and into the protein structure requires some cooperativity and is accompanied by substantial χ_2 displacement (30.7° between bound and unbound conformation) of F110 in the B−C loop. A third phenylalanine, F476 on the C-terminal loop, also moves substantially to contact the ligand in the bound conformation (23° of φ displacement and 42° of χ_1 displacement).

The Adaptive Nature of the Binding Site in P450s with Broad Substrate Specificity. Since the two bacterial P450s investigated here catalyze specialized enzymatic reactions (*2*), their binding pockets have evolved for binding relatively narrow classes of substrates. In contrast, the three mammalian liver P450s hydroxylate structurally diverse endogenous and foreign chemical substrates (*3*) and hence may have acquired the ability to adapt their binding sites to substrates that differ in size, shape, and polarity. Indeed, our mapping results show that the ligand-bound mammalian P450 structures contained a more defined binding pocket surface than the unliganded ones, suggesting that broad substrate specificity is at least partially achieved by an induced fit mechanism that optimizes the properties of the pocket for particular substrates and other ligands. The need for the structural rearrangements may be responsible for the fact that the oxidation rates are frequently orders of magnitude lower in mammalian P450s than in bacterial ones.

As mentioned, for almost all other enzymes mapped in previous studies, the largest number of probe clusters occurred in the active site when mapping both ligand-free and ligand-bound structures with little difference between the two results (*7*). Thus, the clustering of the probes in the active site of the bound but not of the unbound structures of the three mammalian P450s is highly unusual. In previous applications of mapping we have seen a similar result for the Fc fragment of human immunoglobulin G (M. Silberstein, personal communication). The Fc fragment interacts with at least four different natural protein scaffolds (domain B1 of protein A, domain C2 of protein G, rheumatoid factor, and neonatal Fc receptor) (*11*). All four proteins bind to an overlapping region at the C_{H2}/C_{H3} interface. DeLano et al. (*11*) used in vitro selection to isolate a 13-residue peptide that binds to the Fc fragment with substantial affinity. The X-ray structure of the Fc fragment with this peptide shows that the latter also binds at the same C_{H2}/C_{H3} interface. The mapping of the peptide-bound conformation of the Fc fragment places the largest consensus site at this region, indicating a well-defined peptide binding pocket. However, no probes clustered at this site when we mapped the protein-bound Fc structures. Thus, the protein−protein interface does not have the structural features necessary for retaining the small ligands used as probes, but these features are created by an induced fit mechanism in the process of peptide binding (*11*, *12*). As discussed by DeLano et al. (*11*), this behavior reveals a hot spot with a highly adaptive interface that promotes cross-reactive binding. Another example of a hot spot is the region of cytokine IL-2 interacting with the IL-2 α receptor, which also binds a small molecule that buries into a groove not seen in the free structure of the protein (*12*).

The results of the present paper suggest that some mammalian P450s use an adaptive mechanism to attain broad substrate specificity. We emphasize that the occurrence of localized conformational changes required for the recognition of small ligands is largely independent of the overall magnitude of the conformational changes that occur upon substrate binding. On one extreme, the mapping finds the binding site in both ligand-free and ligand-bound structures of P450 BM-3, despite the large deviation between the two structures. On the other extreme, the ligand-free structure of P450 2C9 is unable to retain any of the small molecules used as probes, although the bound and free structures are very similar. It was recently shown by Gutteridge and Thornton (*32*) that the majority of enzymes do not undergo large-scale rearrangements of the active site on substrate binding, and in many cases it is hard to distinguish motion due to substrate binding from experimental uncertainty. Our mapping results show that even very small conformational changes may be extremely important for ligand recognition and that for the three mammalian P450s studied here these changes occur in the process of ligand binding.

The importance of ligand-induced conformational changes for recognition suggests that it may be far from easy to dock small molecules to the X-ray structures of unliganded mammalian P450s using the currently available docking methods that keep the protein rigid. Although there is tremendous interest in the high-throughput virtual screening of compound libraries for binding to particular cytochromes P450, the problem is very challenging because most P450s of major pharmaceutical interest (e.g., 3A4 and 2D6) have broad substrate specificity. In view of our results, the binding sites of such P450s are likely to be highly adaptive and adjust their shapes for the recognition of specific ligands. This shows the need for improved docking methods that account for side chain and possibly backbone flexibility.

It should be noted that some mammalian P450s exhibit substrate specificities more narrow than the three *CYP2* family enzymes considered here. Examples include P450s involved in bile acid biosynthesis (*33*) and P450 1A2, which preferentially binds planar, aromatic structures (*34*). Mapping of ligand-free structures of these P450s would be predicted to place the largest consensus site in the active site, similar to the bacterial P450s. Unfortunately, no X-ray structures are presently available for any narrow substrate specificity mammalian P450s to test this hypothesis. Spectroscopic methods that measure the flexibility of the binding site may provide some information (*35*). Although such spectroscopic methods have not been applied to the mammalian P450s considered here, these methods revealed a very flexible binding site in the case of P450 3A4, which also exhibits a very broad substrate specificity, whereas a rigid active site was indicated for P450 1A2 (*35*). Application of this method to P450 BM-3 also revealed a flexible, but rather stable, active site (*34*).

Binding of a Second Ligand. As we described, the binding sites of P450s with broad substrate specificity are too large

and featureless to bind small molecules, and each incoming ligand has to create its own groove by induced fit. Solvent mapping of P450 2C9 in the presence of *S*-warfarin revealed that a bound ligand may also introduce new binding surfaces that facilitate simultaneous binding of a second ligand within the binding pocket. When P450 2C9 accompanied by the bound ligand was mapped, the largest consensus site shifted to a position previously suggested to bind a second substrate molecule (*35*). The greatest amount of contact made by this consensus site is between the small solvent molecules and *S*-warfarin (Figure 7). A large number of contacts are also made to F476, a residue known to be involved in regioselectivity of the enzymatic products (*36*). This finding of apparent cooperativity between two substrate molecules simultaneously present in the binding pocket may explain higher order kinetics, such as allosteric activation, observed with P450 2C9 (*37*, *38*). The positional shift and the increased number of probe clusters in the best consensus site upon inclusion of *S*-warfarin are consistent with other observations regarding activation of this P450 and drug cooperativity in the active site (*37*, *38*).

ACKNOWLEDGMENT

We thank Michael Silberstein for providing mapping results for the Fc fragment prior to publication.

REFERENCES

1. Nelson, D. R., Koymans, L., Kamataki, T., Stegeman, J. J., Feyereisen, R., Waxman, D. J., Waterman, M. R., Gotoh, O., Coon, M. J., Estabrook, R. W., Gunsalus, I. C., and Nebert, D. W. (1996) P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature, *Pharmacogenetics 6*, 1−42.
2. Pylypenko, O., and Schlichting, I. (2004) Structural aspects of ligand binding to and electron transfer in bacterial and fungal P450s, *Annu. Rev. Biochem. 73*, 991−1018.
3. Stout, C. D. (2004) Cytochrome P450 conformational diversity, *Structure 12*, 1921−1922.
4. Graham, S. E., and Peterson, J. A. (2002) Sequence alignments, variabilities, and vagaries, *Methods Enzymol. 357*, 15−28.
5. Dennis, S., Kortvelyesi, T., and Vajda, S. (2002) Computational mapping identifies the binding sites of organic solvents on proteins, *Proc. Natl. Acad. Sci. U.S.A. 99*, 4290−4295.
6. Kortvelyesi, T., Dennis, S., Silberstein, M., Brown, L., III, and Vajda, S. (2003) Algorithms for computational solvent mapping of proteins, *Proteins 51*, 340−351.
7. Silberstein, M., Dennis, S., Brown, L., III, Kortvelyesi, T., Clodfelter, K., and Vajda, S. (2003) Identification of substrate binding sites in enzymes by computational solvent mapping, *J. Mol. Biol. 332*, 1095−1113.
8. Mattos, C., and Ringe, D. (1996) Locating and characterizing binding sites on proteins, *Nat. Biotechnol. 14*, 595−599.
9. Allen, K. N., Bellamacina, C. R., Ding, X., Jeffery, C. J., Mattos, C., Petsko, G. A., and Ringe, D. (1996) An experimental approach to mapping the binding surfaces of crystalline proteins, *J. Phys. Chem. 100*, 2605−2611.
10. Sheu, S.-H., Kaya, T., Waxman, D. J., and Vajda, S. (2005) Exploring the binding site structure of the PPAR-γ ligand binding domain by computational solvent mapping, *Biochemistry 44*, 1193−1209.
11. DeLano, W. L., Ultsch, M. H., de Vos, A. M., and Wells, J. A. (2000) Convergent solutions to binding at a protein−protein interface, *Science 287*, 1279−1283.
12. Arkin, M. R., Randal, M., DeLano, W. L., Hyde, J., Luong, T. N., Oslob, J. D., Raphael, D. R., Taylor, L., Wang, J., McDowell, R. S., Wells, J. A., and Braisted, A. C. (2003) Binding of small molecules to an adaptive protein−protein interface, *Proc. Natl. Acad. Sci. U.S.A. 100*, 1603−1608.
13. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, I. N., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank, *Nucleic Acids Res. 28*, 235−242.
14. Vakser, I. A., Matar, O. G., and Lam, C. F. (1999) A systematic study of low-resolution recognition in protein−protein complexes, *Proc. Natl. Acad. Sci. U.S.A. 96*, 8477−8482.
15. Schaefer, M., and Karplus, M. A. (1996) A comprehensive analytical treatment of continuum electrostatics, *J. Phys. Chem. 100*, 1578−1599.
16. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, *J. Comput. Chem. 4*, 187−217.
17. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) LIGPLOT: a program to generate schematic diagrams of protein−ligand interactions, *Protein Eng. 8*, 127−134.
18. McDonald, I. K., and Thornton, J. M. (1994) Satisfying hydrogen bonding potential in proteins, *J. Mol. Biol. 238*, 777−793.
19. Sheu, S.-H., Lancia, D. R., Jr., Clodfelter, K. H., Landon, M. R., and Vajda, S. (2005) PRECISE: a database of predicted and consensus interaction sites in enzymes, *Nucleic Acids Res. 33*, D206−D211.
20. Sevrioukova, I. F., Li, H., Zhang, H., Peterson, J. A., and Poulos, T. L. (1999) Structure of a cytochrome P450-redox partner electron-transfer complex, *Proc. Natl. Acad. Sci. U.S.A. 96*, 1863−1868.
21. Li, H., and Poulos, T. L. (1997) The structure of the cytochrome P450 BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid, *Nat. Struct. Biol. 4*, 140−146.
22. Poulos, T. L., Finzel, B. C., and Howard, A. J. (1986) Crystal structure of substrate-free *Pseudomonas putida* cytochrome P-450, *Biochemistry 25*, 5314−5322.
23. 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) The catalytic pathway of cytochrome P450cam at atomic resolution, *Science 287*, 1615−1622.
24. Williams, P. A., Cosme, J., Sridhar, V., Johnson, E. F., and McRee. D. E. (2000) Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity, *Mol. Cell 5*, 121−131.
25. Wester, M. R., Johnson, E. F., Marques-Soares, C., Dijols, S., Dansette, P. M., Mansuy, D., and Stout, C. D. (2003) Structure of mammalian cytochrome P450 2C5 complexed with diclofenac at 2.1 Å resolution: evidence for an induced fit model of substrate binding, *Biochemistry 42*, 9335−9345.
26. Wester, M. R., Johnson, E. F., Marques-Soares, C., Dansette, P. M., Mansuy, D., and Stout, C. D. (2003) Structure of a substrate complex of mammalian cytochrome P450 2C5 at 2.3 Å resolution: evidence for multiple substrate binding modes, *Biochemistry 42*, 6370−6379.
27. 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.
28. Scott, E. E., White, M. A., He, Y. A., Johnson, E. F., Stout, C. D., Halpert, J. R. (2004) Structure of mammalian cytochrome P450 2B4 complexed with 4-(4-chlorophenyl)imidazole at 1.9-Å resolution: insight into the range of P450 conformations and the coordination of redox partner binding, *J. Biol. Chem. 279*, 27294−27301.
29. Scott, E. E., He, Y. A., Wester, M. R., White, M. A., Chin, C. C., Halpert, J. R., Johnson, E. F., and Stout, C. D. (2003) An open conformation of mammalian cytochrome P450 2B4 at 1.6-Å resolution, *Proc. Natl. Acad. Sci. U.S.A. 100*, 13196−13201.
30. Williams, P. A., Cosme, J., Ward, A., Angove, H. C., Vinkovic, D. M., and Jhoti, H. (2003) Crystal structure of human cytochrome P450 2C9 with bound warfarin, *Nature 424*, 464−468.
31. Wester, M. R., Yano, J. K., Schoch, G. A., Yang, C., Griffin, K. J., Stout, C. D., and Johnson, E. F. (2004) The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-Å resolution, *J. Biol. Chem. 279*, 35630−35637.
32. Gutteridge, A., and Thornton, J. (2005) Conformational changes observed in enzyme crystal structures upon substrate binding, *J. Mol. Biol. 346*, 21−28.
33. Kagawa, N., and Waterman, M. R.(1995) Regulation of steroidogenic and related P450s, in *Cytochrome P450: Structure, Mechanism, and Biochemistry* (Ortiz deMontellano, P. R., Ed.) 2nd ed., pp 419−442, Plenum Press, New York.

34. Lewis, D. F., Dickins, M., Eddershaw, P. J., Tarbit, M. H., and Goldfarb, P. S. (1999) Cytochrome P450 substrate specificities, substrate structural templates and enzyme active site geometries, *Drug Metab. Drug Interact. 15*, 1−49.
35. Anzenbacher, P., and Hudecek, J. (2001) Differences in flexibility of active sites of cytochromes P450 probed by resonance Raman and UV-Vis absorption spectroscopy, *J. Inorg. Biochem. 87*, 209−213.
36. Melet, A., Assrir, N., Mansuy, D., et al. (2003) Substrate selectivity of human cytochrome P450 2C9: importance of residues 476, 365, and 114 in recognition of diclofenac and sulfaphenazole and in mechanism-based inactivation by tienilic acid, *Arch. Biochem. Biophys. 409*, 80−91.
37. Hutzler, J. M., Hauer, M. J., and Tracy, T. S. (2001) Dapsone activation of CYP2C9 mediated metabolism: evidence for activation of multiple substrates and a two-site model, *Drug Metab. Dispos. 29*, 1029−1034.
38. Hutzler, J. M., Wienkers, L. C., Wahhlstrom, J. L., Carlson, T. J., and Tracy, T. S. (2003) Activation of cytochrome P450 2C9-mediated metabolism: mechanistic evidence in support of kinetic observations, *Arch. Biochem. Biophys. 410*, 16−24.
39. DeLano, W. L. (2002) The PyMol User's Manual, DeLano Scientific, San Carlos, CA.

BI060343V