Human Cytochrome P450 1A1 Structure and Utility In Understanding Drug and Xenobiotic Metabolism

Agnes A. Walsh¹, Grazyna D. Szklarz², and Emily E. Scott¹

¹From the Department of Medicinal Chemistry, The University of Kansas, 1251 Wescoe Hall Drive, Lawrence, KS 66045, USA

²Department of Basic Pharmaceutical Sciences, West Virginia University School of Pharmacy, P.O. Box 9530, Morgantown, WV 26506, USA

*Running title: Human CYP1A1 Structure and Use in Understanding Metabolism

SUMMARY

Cytochrome P450 (CYP) 1A1 is an extrahepatic monooxygenase involved in the metabolism of endogenous substrates and drugs, as well as the activation of certain toxins and environmental pollutants. CYP1A1 is particularly well known for its ability to biotransform polycyclic aromatic hydrocarbons (PAHs), such as benzo[a]pyrene in tobacco smoke, into carcinogens. CYP1A1 possesses functional similarities and differences with human CYP1A2 and CYP1B1 enzymes, but the structural basis for this has been unclear. We determined a 2.6 Å structure of human CYP1A1 with the inhibitor α-napthoflavone (ANF). ANF binds within an enclosed active site, with the planar benzochromen-4-one core packed flat against the I helix that composes one wall of the active site, and the 2-phenyl substituent oriented toward the catalytic heme iron. Comparisons with previously determined structures of the related cytochrome P450 1A2 and 1B1 enzymes reveal distinct features among the active sites that may underlie the functional variability of these enzymes. Finally, docking studies probed the ability of CYP1A1 structures to assist in understanding their known in vitro interactions with several typical substrates and inhibitors.

Cytochromes P450 are ubiquitous enzymes involved in the biotransformation of a vast array of drugs and other chemical compounds. Cytochrome P450 1A1 (CYP1A1) is a member of the human CYP family generally expressed in extrahepatic tissues, such as the epithelia of the lung, skin, and gastrointestinal tract, though it can also be detected in the placenta, fetus, and embryo (1-5). Although endogenous substrates include the inflammatory mediators arachidonic acid and eicosapentoic acid (6,7) and the hormones 17β-estradiol (8) and melatonin (9), CYP1A1 is best known as one of the most important enzymes in the bioactivation of procarcinogens to generate reactive metabolites (10). CYP1A1-mediated metabolism of environmental contaminants can lead to reactive metabolites that form DNA adducts which contribute to mutagenesis and ultimately tumor formation (11). Such CYP1A1 substrates include combustion and tobacco products, polycyclic aromatic hydrocarbons (PAHs), heterocyclic aromatic amines (HCA) found in charred meat, and industrial arylamines (12,13).

Constitutive CYP1A1 expression is generally low, but this enzyme can be greatly induced by many of its own substrates, particularly the PAHs,
via the arylhydrocarbon nuclear receptor (AhR) (14). For example, CYP1A1 mRNA levels are very high in the lung cells of smokers but typically undetectable in nonsmokers (15). CYP1A1 protein is also higher in smokers compared to nonsmokers (lung median levels of 16 vs. 6.5 pmol/mg microsomal protein, respectively) (16), as well as in other tissues (17,18).

This induction and procarcinogen activation make CYP1A1 a high-interest target for potential chemopreventive agents (19). Inhibitors such as dietary flavonoids can block the production of reactive metabolites and thus DNA damage. For example, the flavone acacetin can inhibit CYP1A1-mediated 7-ethoxyresorufin O-dealkylation with an IC50 of 80 nM (20), and the flavone apigenin can eliminate detectable DNA adduct formation by CYP1A1-activation of benzo[a]pyrene (21). In contrast, in certain malignancies that overexpress CYP1A1, another therapeutic approach utilizes this enzyme to activate anticancer prodrugs to reactive electrophilic species that inhibit tumor growth (22,23).

Recently the first X-ray structures have become available for the two other enzymes of the human CYP1 family, CYP1A2 (24) and CYP1B1 (25). Hepatic CYP1A2 is 72% identical in amino acid sequence to CYP1A1, and metabolizes caffeine, phenacetin, HCAs (26,27), and several clinically used drugs (28), as well as being involved in many drug-drug interactions (29). Like CYP1A1, CYP1B1 is an extrahepatic enzyme, but has low amino acid sequence identity with both CYP1A1 (38%) and CYP1A2 (37%). This is less than the 40% usually required to be classified to the same cytochrome P450 family, but the grouping was nonetheless assigned on the basis of their common induction by AhR and similar substrate specificities towards such compounds as PAHs, HCAs (30), and estradiol (31). However, there are also a number of functional differences among these three enzymes, and without a CYP1A1 structure, a basis for evaluating the features responsible for functional similarities and differences among human CYP1 enzymes has been lacking.

Herein we report a 2.6 Å structure of human CYP1A1 with the inhibitor α-naphthoflavone (ANF), which establishes structural elements for understanding ligand binding and thus may facilitate the design of new anticancer therapeutics. In addition, we compare and contrast the CYP1A1 active site residues and cavities with those of the other two members of the CYP1 family. Since other mammalian cytochrome P450 enzymes often have flexible active sites that depend on the ligand (32,33), we use docking to evaluate the practical question of whether this single structure of human CYP1A1 is sufficient to understand the binding of other ligands including substrates and inhibitors, or if further structures with different ligand scaffolds will be necessary to elucidate the full functional capacity of these enzymes. Finally, these docking outcomes were compared with docking of the same compounds to CYP1A2 to evaluate the implications of protein active site differences on ligand orientations and metabolism.

EXPERIMENTAL PROCEDURES

Protein Engineering, Expression, and Purification- The cDNA (34) for human CYP1A1 was truncated to remove the coding sequence for the N-terminal transmembrane residues 2-34 and to substitute nucleotides coding for the hydrophilic amino acid sequence AKKTSS, a modification that has enabled the crystallization of many membrane cytochromes P450 to date, including CYP1B1. CYP1A2, however, was expressed with simply a truncation. For CYP1A1 nucleotides were added to code for a six-residue histidine tag at the C-terminus immediately prior to the stop codon. This modified cDNA was inserted into the pKK233-2 plasmid to generate the pKK1A1(MAK)H6 construct, which was transformed into tetracycline (tet)-resistant Escherichia coli TOPP-3 cells (Stratagene, La Jolla, CA) previously transformed with the pGro7 plasmid containing the gene for GroEL/ES (Takara Bio Inc., Shiga, Japan). Individual colonies were grown at 37°C overnight on Lysogeny Broth (LB) agarose plates containing 50 µg/ml ampicillin (amp) to select for the pKK1A1(MAK)H6 plasmid, which was transformed into tetracycline (tet)-resistant Escherichia coli TOPP-3 cells (Stratagene, La Jolla, CA) previously transformed with the pGro7 plasmid containing the gene for GroEL/ES (Takara Bio Inc., Shiga, Japan). Individual colonies were grown at 37°C overnight on Lysogeny Broth (LB) agarose plates containing 50 µg/ml ampicillin (amp) to select for the pKK1A1(MAK)H6 plasmid and 20 µg/ml chloramphenicol (chlor) to select for the pGro7 plasmid. A single colony was grown in 5 ml LB media supplemented with amp, chlor, and 25 µg/ml tet for ~12 hr at 37°C and 250 rpm. Fifty microliters of this culture were added to 200 ml LB containing amp, chlor, and tet and grown for ~12 hr at 37°C and 250 rpm. Fifteen milliliters from this 200 ml culture was added to each of 18 one-liter flasks containing 225 ml Terrific Broth.
(TB) medium, 25 ml TB salts, amp, chlor, and arabinose (0.5 mg/ml, to induce expression of GroES and GroEL chaperone proteins). Cells were then grown at 37°C, 250 rpm for ~2hrs or until the OD$_{600}$ reached 0.6. At this time 1 mM isopropyl-beta-D-1-thiogalactopyranoside and 80 mg/ml δ-aminolevulinic acid were added. E. coli were subsequently grown at 30°C, 190 rpm for ~72 hr, after which time they were harvested and the cell pellet was stored at -80°C until purification.

Protein purification was carried out at 4°C. The cell pellet from 2.25 liters of culture was thawed and resuspended in 200 ml 20 mM potassium phosphate buffer, pH 7.4 and 20% glycerol. Lysozyme (0.2 mg/ml) was added to the cell pellet suspension and stirred for 30 min. followed by the addition of 200 ml of cold water with stirring for 10 min. Spheroplasts were collected by centrifuging at 8,000 x g for 15 min. and the supernatant was discarded. The pellets were flash frozen in an ethanol/dry ice bath, thawed, and resuspended in 100 ml 500 mM potassium phosphate, pH 8.0, 300 mM NaCl, 20% glycerol. This solution was homogenized, divided into three 50 ml aliquots, sonicated three times for 30 sec. each with 1 min. on ice between each sonication, and centrifuged at 8,000 x g for 15 min. The ligand ANF (10 µM) and the detergent CHAPS (1% w/v) were added with stirring for 30 min. to stabilize and solubilize CYP1A1, respectively. Solubilized CYP1A1 was separated from membrane lipids by ultracentrifugation at 80,000 x g for 60 min.

The CYP1A1 protein was purified from the resulting supernatant by metal-affinity, ion-exchange, and size exclusion chromatography. Initially the ultracentrifuge supernatant was applied to a Ni$^{2+}$-nitrilotriacetic acid (Ni-NTA) column (Qiagen, Valencia, CA) that had been equilibrated with 100 mM potassium phosphate, pH 7.4, 200 mM NaCl, 20% glycerol, 0.5% CHAPS, and 10 µM ANF. The column was washed with two column volumes (CV) of equilibration buffer followed by twelve CV of equilibration buffer including 3.2 mM histidine. Elution was accomplished using two CV of 10 mM potassium phosphate, pH 7.4, 0.1 M NaCl, 20% glycerol, 0.5% CHAPS, 10 µM ANF, 2 mM EDTA, and 150 mM histidine. Eluted fractions containing significant red color were pooled, diluted five-fold with 5 mM potassium phosphate, pH 7.4, 20% glycerol, 0.5% CHAPS, 10 µM ANF, and 1 mM EDTA, and applied to a 5 ml CM- sepharose FF column (GE Healthcare, Uppsala, Sweden). The column was washed with six CV of dilution buffer and eluted with 2 CV of CM elution buffer (50 mM potassium phosphate, pH 7.4, 0.5 M NaCl, 20% glycerol, 0.5% CHAPS, 10 µM ANF, and 1 mM EDTA). Red fractions were pooled, concentrated to 1 ml, and loaded onto a Superdex 200 gel filtration column (GE Healthcare, Uppsala, Sweden) that had been equilibrated with CM elution buffer. The purest protein fractions were identified by the A$_{393}$/A$_{280}$ ratio (typically exceeding 1.1) and concentrated to 20 mg/ml by centrifugal ultrafiltration. During concentration, CYP1A1 was washed with CM elution buffer including 100 µM ANF and 0.4 M NH$_4$NO$_3$ to facilitate saturation with ligand and crystallization, respectively. Typical yields were 40-70 mmol/L E. coli culture.

Spectral Ligand Binding Assay—The spectral ligand binding assay was performed as described previously (35) with a protein concentration of 1 µM.

Protein Crystallization, Data Collection, and Structure Determination—CYP1A1 crystals were grown by the hanging drop vapor diffusion method. Twenty-four well crystallization plates were set up on ice with drops composed of 1 µl of the 20 mg/ml CYP1A1/ANF protein complex and 1 µl of well solution, which consisted of 0.1 M bis-Tris propane/HCl pH 8.5 (Emerald Biosystems, Bainbridge Island, WA),200 mM sodium malonate (Emerald Biosystems, Bainbridge Island, WA),20% w/v PEG 3350 (Hampton Research, Aliso Viejo, CA), and 14% glycerol. Drop:s were equilibrated against 1 ml of this well solution. Plates incubated at 4 °C for 7 days yielded triangular prism-shaped crystals that were subsequently immersed in cryoprotectant (well solution with 30% PEG 3350, 13% glycerol, and 100 µM ANF), harvested, and flash cooled in liquid nitrogen for storage until data collection.

A dataset was collected from a single crystal at the Stanford Synchrotron Radiation Lightsource (Stanford, CA) beam-line 9-2 at a wavelength of 0.979 Å. Data integration and scaling were carried out employing the programs MOSFLM (36) and SCALA (37). The structure was solved by molecular replacement using Phaser (37) with the CYP1A2 structure as a search model (24), (PDB 2HI4, molecule A) using the entire resolution...
range to 2.6 Å. The solution (P2_1_2_1 space group) had an asymmetric unit consisting of four molecules with LLG values of 880, 2878, 4206, and 6254. Iterative model building and refinement were carried out using COOT (38) and Refmac (39) in CCP4 (37). Statistics for data collection, refinement, and validation are shown in Table 1.

Structure analysis—The CYP1A1 structure herein was compared with the previously published CYP1A2 (PDB 2HI4) and CYP1B1 (PDB 3PM0) structures by overlapping all Cα atoms in each structure using Swiss-PDB Viewer (40). Secondary structures were defined using DSSP (41). Probe-occupied cavities were defined using VOIDOO (42) using a grid spacing of 0.33 and a probe radius of 1.4 Å. Figures were generated using PyMol (43).

Sequence analysis—Amino acid sequence identity was calculated using the full-length wild type sequences and EMBOSS Needle (44).

Molecular Modeling Simulations—All molecular modeling was performed on SGI workstation with Insight II/Discover software package (Accelrys, Inc., San Diego, CA) using consistent valence force field supplemented with parameters for heme (45,46). Substrates ethoxyresorufin, methoxyresorufin, benzo[a]pyrene, and phenacetin, and inhibitors resveratrol and α-naphthoflavone were built and energy optimized using the Builder module. The crystal structures of CYP1A1 and CYP1A2 (PDB 2HI4) (24) were refined by the addition of hydrogens and subsequent minimization (47). The ligands were initially placed manually into the active site of the enzyme in order to avoid steric overlaps, and then docked using the Affinity module of Insight II using default parameters, as described earlier (47-49). Residues within 10 Å of the docked ligand comprised the flexible region of the protein during all docking runs. Twenty distinct binding complexes obtained by the Monte Carlo search technique were subjected to simulated annealing prior to energy ranking, and ten lowest-energy binding orientations were evaluated.

RESULTS AND DISCUSSION

Spectral Characterization—When purified without the addition of ANF, the absolute spectrum of CYP1A1 exhibits an absorbance peak at 419 nm, suggesting the enzyme exists in a low spin state with water bound to the heme iron. However, when ANF is added during purification to stabilize the enzyme, the Soret peak occurs at 393 nm, consistent with the high spin state as ligand binding displaces the water. Monitoring this spectral shift upon titration of ligand-free CYP1A1 with ANF permitted estimation of the Kd~300 nM, but this estimate was limited by absorbance of the ANF itself at higher ligand concentrations. In contrast, CYP1A2 was reportedly isolated already in the high spin state and addition of ANF had no effect on the spin state (50), whereas purified CYP1B1 was isolated as a mixture of both high-spin and low-spin forms (51).

Overall CYP1A1 Structure—A 2.6 Å structure of human CYP1A1 determined in complex with the inhibitor α-naphthoflavone contained four molecules that were very similar, with an average Cα rmsd of 0.41 Å. While a short gap between helices H and I could not be modeled for molecules B-D, molecule A could be clearly modeled throughout this region (Table 1, Fig. S1) and thus served as the basis for CYP1A1 structural analysis herein. As expected, the global structure of CYP1A1 displayed the general cytochrome P450 fold (Fig. 1A). Canonical helices A through L are present, including short F’ and G’ helices typically thought to be buried in the membrane and often involved in active site access for hydrophobic ligands. Three of the four canonical β sheets are present, with the residues normally composing the C-terminal β_4-1 and β_4-2 strands having similar backbone placement, but slightly modified hydrogen bonding interactions. One unusual aspect of the secondary structure is a five-residue break in the middle of the F helix (Fig. 1), which has implications for ligand binding (vide infra).

Comparison of CYP1 Global Structures—CYP1A1 is the third and final member of the human CYP1 family to have its structure determined. Structures of CYP1A2 (24) and CYP1B1 (25) were previously determined, also as complexes with ANF, making comparisons among the enzymes more straightforward by eliminating structural differences that might arise in response to the binding of different ligands. Comparisons identify many similarities and differences among the three enzymes.

The overall structures of the three human CYP1 family enzymes are very similar, but have several differences in specific secondary structure elements (Fig. 1B, Fig. S1). First, the residues that normally compose the two strands of the β_4 system
in other human P450 enzymes do not adopt this secondary structure in either CYP1A1 or CYP1A2 but do in CYP1B1. The tip of the β₄ region lines the active site and has often been implicated in ligand contact and mobility associated with ligand access (52). Second, two short 2-residue beta strands in the HI loop of CYP1A2 are not present in the single molecule of CYP1A1 that could be modeled in this region. The other three molecules of CYP1A1 are substantially disordered in this region and could not be reliably modeled, which was also the case with CYP1B1. Thus, the β₄ region of CYP1A1 and CYP1A2 and the H/I loop in CYP1A1 and CYP1B1 may be less rigid. Third, there are minor differences in the backbone hydrogen bonding patterns of the B’ helix among the three structures (Fig. S1), but the overall structure in this region is very similar (Fig. 1B). This is key because the B’ helix forms a portion of the active site wall and has been proposed to be involved in substrate access/egress and regio- and stereoselective oxidation in other P450s (53,54). Close examination of the B’ helices herein will show that among the three human CYP1 enzymes this leads to an overall conservation of the active site dimensions, modulated primarily by side chain substitutions and orientations.

Finally, and most notably at the overall structure level are differences in the F helix. The three-residue disruption of helix F previously observed only in CYP1A2 and CYP1B1 is clearly extended to five residues in CYP1A1 (Fig. 2). In addition, in CYP1A1 the backbone across this break is significantly more distorted from a helical conformation compared to both CYP1A2 and CYP1B1. Like the B’ helix, the F helix has been associated with substrate access/egress and regio- and stereoselective substrate binding (52,55). In contrast to the B’ region, the differences in the F helix backbone do alter interactions that modulate the active sites of the three enzymes, but do so while retaining a critical π-π stacking interaction between a conserved Phe in helix F and ANF (vide infra). Thus, the 5-residue break in the F helix of CYP1A1 represents a significant difference between 1A1 and the other human family 1 enzymes and could allow for greater flexibility in substrate movement, binding, and orientation.

**Active Site of CYP1A1—** CYP1A1 binds ANF (Fig. 3, pink sticks) adjacent to the heme within an enclosed active site. The active site cavity volumes are similar among the four CYP1A1 molecules with small differences due to slight torsions of a few bordering side chains (Table 1), yielding an average of 524 Å³. The heme forms the “floor” of the cavity (Fig. 3, grey sticks), which is flanked by “walls” consisting of the I helix (yellow), the B’ helix region (blue), the loop between helix K and β₁₄ (orange), and residues at the tip of what is usually the β₄ region (red). Opposite the heme, the “roof” of the active site is composed primarily of helices F and G (green). The ANF density is clearest in molecules A and B and supports ANF binding in a single orientation with the planar benzo[h]chromen-4-one system packed flat against the I helix (Fig. 3B). In particular, the planar peptide bond between Gly316 and Ala317 (not shown) complements the flat ANF surface. The 2-phenyl substituent is directed toward the heme iron at a distance of 4.5 Å (Fig. 2A). On the opposite side from the I helix, ANF forms a very complementary π-π stacking interaction with the side chain of Phe224, adjacent to the break in the F helix. This interaction with Phe224 and the packing of the hydrophobic residues Phe123, Ile115, L496, Val382, and Ile386 constrain the front and sides of the ANF binding pocket. Additional hydrophobic residues, Phe319, Phe258, and Leu312 form the rear of the active site. Polar residues within 5 Å of ANF include Ser116, Ser120, Ser122, Asn222, Asn255, Asp313, Asp320, the I helix Thr at position 321, and Thr497, but they appear to serve more structural roles within the active site rather than forming direct interactions with ANF.

Thus, the only direct interaction of ANF with CYP1A1 appears to be π-π stacking with Phe224. The orientation of this Phe is likely facilitated by the unusual break in the middle of the F helix, which encompasses Asn219 to Asn223 and is highly conserved among all four CYP1A1 molecules. This break also directs Asn222 into the active site, where it is part of an extensive hydrogen bonding network which includes Tyr187, Asp320, Thr324, Thr497, Lys499, a water molecule, and the amide nitrogen of Gly225 (Fig. 4A). This network lies at the confluence of four discrete structural elements, likely stabilizing interactions of the E, F, and I helices and the loop containing T497 and K499.

On the opposite side of the active site a second hydrogen bonding network secures the B’-C region to the I helix (Fig. 4B). Asp313, the acid of the “acid-alcohol pair” often found in the CYP I helix

---

**Human CYP1A1 Structure and Use in Understanding Metabolism**

---

---
over the active site, is within hydrogen bonding distance to the side chain hydroxyls of a trio of B/C region serines (116, 120, and 122), the amide nitrogen of Ser122, and a water molecule located just outside the active site. This water molecule also interacts with Asn309 (I helix), Arg134 (C helix), and the amide nitrogen of Met121 (B’-C loop). Although similar among all four CYP1A1 molecules, the details of these two hydrogen bonding networks vary somewhat, with fewer such interactions in molecules C and D.

Finally, in CYP1A1 molecules A and B the $|F_o|-|F_c|$ electron density map includes a small spherical area of positive density between ANF and Ser122, which could be interpreted as indicating the presence of a weakly bound water molecule (not shown). However the interaction distances were too long for hydrogen bonds and refinement did not support a water molecule or an alternate conformation of ANF. This density is much less clear for CYP1A1 molecules C and D, as is the density for ANF in those two active sites.

Comparison of Human CYP1A1 with other CYP1 Active Sites—Within the three human CYP1 enzyme active sites the ANF ligand occupies the same plane, packing against the I helix in each structure. The orientations in CYP1A1 and CYP1A2 are very similar (Fig. 5A), while in CYP1B1 the molecule is flipped by 180° about the long axis of the ligand so that the carbonyl is pointing in the opposite direction (Fig. 5B). While ANF in CYP1A1 appears to be primarily oriented by sterics and stacking with the side chain of F224, the CYP1A2 active site additionally contains a water molecule hydrogen bonded to the carbonyl of ANF and to the carbonyl of Gly316. Detailed comparisons of the bordering residues are discussed below, with the CYP1A1 residue or number given first and the corresponding CYP1A2 or CYP1B1 residue listed second when there is a difference in either the amino acid or numbering.

Comparison of the active site residues of CYP1A1 to CYP1B1 reveals a number of similarities and differences as well (Fig. 5B). There are two sets of phenylalanines that overlap nearly identically: Phe123/134 and Phe224/231, which sterically constrain the ANF plane. Other identical residues that occupy virtually the same space in CYP1A1 and CYP1B1 are Ile115/Val126, Ser122/Ala133, Leu312/Thr325, and Tyr319/Gln332. Interestingly, the overall positioning of side chains lining the active site is more similar for CYP1A1 and CYP1B1 than for CYP1A1 and CYP1A2, despite the higher sequence identity between the latter two enzymes.

Comparison of the cavity sizes and topologies of the three CYP1 enzymes reveals an overall similarity of shape and size. All three are narrow cavities, consistent with the planar ligand topology favored by this family of P450 enzymes (Fig. 6). Comparison of the probe-occupied voids using a probe radius of 1.4 Å reveals that the volumes range from an average of 524 Å³ for CYP1A1 to 469 Å³ for CYP1A2 and 441 Å³ for CYP1B1. This size ranks CYP1A cavities intermediate between those observed for CYP2E1 and CYP2A enzymes (typically 190-300 Å³) and those of CYP2C8, CYP2C9, CYP2D6, and CYP3A4 (typically 800-1,500 Å³).

The differences between the CYP1A1 and CYP1A2 cavities (Fig. 6A and 6B) arise from several instances in which the CYP1A1 side chains lining the active site are smaller (S116/Thr118, S122/Thr124, Gly225/Val227, Leu254/Phe256, and Val382/Leu382) and allow for a greater active site cavity volume. The most significant difference in the CYP1A cavity proportions is an extension of the CYP1A2 active site over the I helix in CYP1A2, which is not present in CYP1A1 (Fig. 6B). The reason for this is the rotations of two conserved phenylalanines in CYP1A1 compared to CYP1A2 (Phe319 and Phe258/260) in opposite directions.
Overall, the detailed topology of the CYP1A1 active site is more similar to the CYP1B1 cavity than to the cavity of CYP1A2 (Fig. 6C and 6D). The CYP1A1 cavity is slightly larger than the CYP1B1 cavity in part due to side chain reorientation of L254/264 on the G helix, a slight shift of the B’ helix near residues S116/127, and a torsion of Asp320/333 (Fig. 6C). The CYP1B1 cavity is larger than CYP1A1 in one area where Leu496/509 are in slightly different positions (Fig. 6D).

The closer similarity between the CYP1A1/ANF active site and the CYP1B1/ANF compared to the CYP1A2/ANF active site may contribute to the observed functional diversity among these three enzymes. Although all three share substrates and inhibitors, CYP1A1 and CYP1B1 could be said to have a generally more similar substrate and inhibitor profile than do CYP1A1 and CYP1A2. For example, CYP1A1 and CYP1B1 are better at O-dealkylating ethoxyresorufin than methoxyresorufin, while CYP1A2 has the opposite preference (56).

Additionally, CYP1A1 and CYP1B1 have more similar oxidation profiles for a variety of substrates (57), as well as inhibition by flavonoids (58) and hydroxystilbenes (59) than do CYP1A1 and CYP1A2. Moreover, CYP1A1 and CYP1B1 can activate the 7,8-diol of benzo[a]pyrene (60) and PAH diols (61) more efficiently than CYP1A2.

**Ligand Docking as a Means to Compare Human CYP1A1 and CYP1A2 Metabolism**—Structures of a number of other human cytochrome P450 enzymes involved in xenobiotic metabolism have demonstrated that P450 active site topology can be very different for the same enzyme in the presence of different ligands. Thus it is important to determine if the active site topology experimentally defined by CYP1A enzyme structures can be used to rationalize the binding of other known CYP1A ligands. To test this idea, we have focused on the two human CYP1A enzymes. A chemically diverse set of ligands were docked into the experimentally-determined active sites, while allowing for active site flexibility, and the resulting orientations compared to the known sites of metabolism.

First, as a control docking experiment, ANF was removed from each of the CYP1A/ANF X-ray structures and then docked back into each of the empty active sites. For CYP1A1, inspection of the ten lowest energy conformations revealed that they were all very similar in energy (610-637 kcal/mol; <2.4% variation) and that the lowest energy pose and three others were in essentially the same orientation observed in the experimental structure (Fig. 7A). However other orientations of ANF were also observed, in which its structure was flipped by 180° along either the ligand long (C4’-C6) or short (C3-O1) axes (Fig. 1A), but in all cases the plane of the ligand was the same as that of ANF in the experimental structure. ANF was also docked into the empty experimental CYP1A2 structure and the ten lowest energy poses (404-421 kcal/mol; 4% variation) evaluated. Three of the resulting poses, including the lowest energy pose, were also very similar to that observed in the CYP1A2 X-ray structure (Fig. 7B). As in the case of CYP1A1, other orientations of ANF were also observed in which the ligand was flipped by 180° along either of the ligand major axes, but the ligand plane was always the same as in the experimental structure. One difference, however, was that for CYP1A1 two of these poses oriented C6 towards the heme, whereas this orientation was not observed with CYP1A2. This is consistent with the slow metabolism of ANF by CYP1A1 at the 5,6 position, while for CYP1A2 very little or no ANF metabolism was detected at this position (62). However any of the docked orientations would be consistent with the high spin states observed in solution for both enzymes in the presence of ANF. Although one binding orientation is observed in the X-ray structure of each enzyme, it does not exclude the possibility that at least some of these other orientations might also occur under different conditions.

Knowing the site(s) of metabolism should assist in distinguishing between likely and unlikely binding orientations. Thus, the CYP1A substrates phenacetin, ethoxyresorufin, methoxyresorufin, and benzo[a]pyrene (Fig. 7I) were similarly docked into the empty X-ray CYP1A1 and CYP1A2 structures. When the O-deethylated substrate phenacetin (Fig. 7I) was docked into CYP1A1, the set of ten lowest energy orientations (15-32 kcal/mol) had four poses, including the lowest energy one, in which phenacetin was oriented in a manner consistent with the experimentally observed metabolism (Fig. 7C). The remaining six poses essentially swapped the positions of the acetamide and the ethoxy groups, resulting in non-productive binding orientations.
In all cases, the substrate was docked in the ANF plane. Likewise, in CYP1A2, the lowest energy pose for phenacetin also revealed binding in the same plane as ANF with an orientation suitable for the experimentally observed O-deethylation (Fig. 7D). Overall, six poses were observed with the ethoxy group oriented towards the heme and four poses with the ethoxy group directed away from the heme, but the differences in energy among these poses were minimal but higher than CYP1A1 (~5%; 411 to 433 kcal/mol). However the reported K_d values for phenacetin binding are 57.4 and 17.9 µM for CYP1A1 and CYP1A2, respectively (47). Interestingly, NMR experiments performed in solution detect an average phenacetin position at much greater distance from the heme iron (63), but with the oxidation site of the substrate closer to heme iron in CYP1A2 then in CYP1A1. However, since NMR T1 experiments represent only the average binding orientation, it is possible that multiple binding modes are available for phenacetin (63).

Docking of the O-dealkylated resorufin substrates (Fig. 7I) 7-methoxyresorufin (134-150 kcal/mol) and 7-ethoxyresorufin (73-105 kcal/mol) to CYP1A1 (Fig. 7E) similarly revealed two types of binding orientations. For methoxyresorufin, six of the poses, including the lowest energy pose, had the methoxy substituent oriented away from the heme, thereby representing non-productive binding (Fig. 7E, black). However, the methoxy group was directed toward the heme in the remaining four poses, consistent with the observed O-dealkylation. The same two types of orientations were observed for ethoxyresorufin docked into CYP1A1, but in this case six poses, including the lowest energy one, were oriented as would be expected for O-deethylation (Fig. 7E, grey), while the remaining four had the alkoxy group oriented away from the heme. The number of poses in each orientation and the lowest energy poses for each substrate are in agreement with the known preference of CYP1A1 for ethoxyresorufin metabolism over methoxyresorufin (64). As in the case of CYP1A1, both resorufins docked into CYP1A2 in the same plane as ANF and with a bimodal distribution of the alkoxy group, either directed toward the heme suitable for O-dealkylation, or away from the heme and toward the opposite end of the cavity. For methoxyresorufin, the lowest energy pose (Fig. 7F, black) and four others were consistent with O-dealkylation, but the remaining 50% of the time the alkoxy group was directed away from the heme. However, all binding orientations had very similar energies (395-408 kcal/mol; 3.2% variation). By comparison, ethoxyresorufin docked into CYP1A2 (501-515 kcal/mol; 2.7% variation) with seven of the poses oriented for metabolism (including the lowest energy pose, Fig. 7F, grey) and three in a nonproductive orientation. These ~100 kcal/mol differences in docking energies for methoxyresorufin vs. ethoxyresorufin are consistent with preferential oxidation of methoxyresorufin by CYP1A2 (64).

Two other ligands were selected for docking into CYP1A1. The first was benzo[a]pyrene, which is known to be metabolized by CYP1A1 at many different sites around the fused ring structure (Fig. 7I) (62). In contrast to phenacetin and alkoxysresorufins, docking of benzo[a]pyrene yielded ten poses with virtually the same binding energies (657-670 kcal/mol; <2% variation), but presenting many different carbon atoms to the heme iron (Fig. 7G). For the ten lowest energy poses, the C1, C2, C3, C7, C8, or C9 atoms (Fig. 7I) were variously positioned closest to the heme, all of which are consistent with the experimentally observed formation of the multiple benzo[a]pyrene metabolites. By comparison, earlier work that involved docking benzo[a]pyrene into a CYP1A1 homology model based on the CYP2C5, the first mammalian P450 structure, yielded additional nonproductive orientations (65), likely due to an active site topology that is much less planar than that of the current experimental structure.

Most of the ligands docked into CYP1A1 thus far consisted of multiple fused rings and adopted an orientation in which a planar core docked in essentially the same plane as ANF. On the other hand, resveratrol was selected because it has a different central structure, a trans-stilbene core. Docking results revealed that all ten lowest energy poses still assumed the same plane as observed with the other ligands (Fig. 7H), with either the dihydroxyphenyl ring or the monohydroxyphenyl ring directed toward the heme. These results are consistent with the metabolism by CYP1A1 (Fig. 7I) to piceatannol and 3,4,5,4′-tetrahydroxystilbene (66), resulting from oxidation of the opposite ends of the molecule. Interestingly, the binding energies for resveratrol (-122 to -97 kcal/mol) were the lowest observed for CYP1A1 in these studies, perhaps because of
multiple hydrogen bonding opportunities in the CYP1A1 active site. Of the ten lowest energy poses, each could form at least one hydrogen bond, and half could form two to Ser122, Asn255, Asn222, Asp313, Asp320 and/or T497.

Overall, when the X-ray structures of CYP1A1 or CYP1A2 were used to dock a variety of substrates and inhibitors, all poses of all ligands aligned with the plane observed for ANF in the experimental structures. The narrow, planar active site thus appears to strictly constrain ligand orientation even when side chains are allowed to reposition. In general, ligands docked either with the known site of oxidation oriented directly toward the heme iron or inverted so that the site of oxidation was most distant from the iron. Both docking and functional evidence suggests that steric properties play a key role in ligand orientation beyond controlling the plane in which the ligand binds. The residue at position 382 is known to be important in determining the functional differences between CYP1A1 and CYP1A2 (64,65). While CYP1A1 has a preference for ethoxy- over methoxyresorufin, CYP1A2 is the opposite. However, exchanging the Leu found in CYP1A2 and the Val found in CYP1A1 interconverts the functional preference for the two alkoxyresorufins (34,64). This is consistent with a difference in volume between CYP1A1 and CYP1A2 adjacent to the 382 side chain (Fig. 6A) as well as the docking studies finding that the lowest energy pose of ethoxyresorufin, but not methoxyresorufin, was oriented for metabolism in CYP1A1. However in CYP1A2 the lowest energy poses for both alkoxyresorufins are consistent with oxidation. The CYP1A2 L382V mutation also has a significant impact on phenacetin metabolism, increasing catalytic efficiency ($k_{cat}/K_m$) by more than 26-fold (47), consistent with the smaller 382 residue facilitating ligand placement

Similarly, the active site residue S122 in CYP1A1, corresponding to T124 in CYP1A2, alters the active site volume (Fig. 6A). The CYP1A1 S122T mutation increases the O-dealkylation of both alkoxyresorufins, while the CYP1A2 T124S mutation decreases the metabolism of both resorufin substrates (64). CYP1B1 has A133 at the corresponding position, with even lower rates of alkoxyresorufin metabolism (56). This progression suggests that Thr is optimal in this position with the smaller Ser and Ala reducing O-dealkylation.

These residues and other steric differences among the CYP1 active sites should be investigated to determine how general their roles are in sterically directing metabolism and inhibition by other ligands. For example, resveratrol has shown over 50-fold higher selectivity for inhibiting alkoxyresorufin O-dealkylation by CYP1A1 over CYP1A2 (67). Additionally, CYP1A2 metabolizes resveratrol to piceatannol and 3,4,5,4'-tetrahydroxystilbene at rates 6-fold and 5-fold higher than CYP1A1 and CYP1B1, respectively (66). CYP1A1 and CYP1B1 have the same valine (V382 and V395, respectively) at the position where CYP1A2 has leucine (L382) and both prefer ethoxy- over methoxyresorufin O-dealkylation (56).

The new crystal structure of CYP1A1 can also be used as a tool to predict the metabolism of various compounds, including likely oxidation sites as well as various kinetic parameters. For example, in earlier studies using a homology model of CYP1A1, binding free energies were calculated for methoxyresorufin, ethoxyresorufin and benzo[a]pyrene, and were found to be close to experimental values (65). Similar calculations were also performed for the CYP1A1 V382A and V382L mutants with alkoxyresorufin substrates, where little or no change in binding energy correlated well with similar $K_m$ values obtained experimentally (34). Regiospecificity of CYP1A1 oxidations was studied with benzo[a]pyrene and arachidonic and eicosapentaenic acids using a homology model of the enzyme and molecular dynamics simulations to determine product profiles (48). These and other computational methods can now be applied to CYP1A1 crystal structure, likely providing more accurate data concerning the metabolism of many xenobiotics.

In summary, the structure of CYP1A1 with $\alpha$-naphthoflavone reveals a narrow, enclosed active site, well suited for the binding and metabolism of polycyclic aromatic hydrocarbons and other planar molecules. Comparisons with other CYP1 family enzyme structures suggest similarities and differences among the three active sites that reflect their varied substrate specificities and correlate well with the functional results of mutagenesis between CYP1A1 and CYP1A2. Docking studies reveal ligand orientations constrained to the plane observed for ANF with sites of metabolism oriented either toward the heme or flipped by ~180 degrees so that the site of metabolism is directed...
Human CYP1A1 Structure and Use in Understanding Metabolism

toward the far end of the cavity, orientations in general agreement with many of the functional studies of CYP1A1. Thus this structure appears to provide a useful framework for understanding CYP1A1 interactions with phenacetin, alkoxyresorufins, benzo[a]pyrene, and resveratrol. It remains to be evaluated whether this active site topology is also informative with regard to the binding of other chemically diverse ligands such as steroids and fatty acids to CYP1A1 or if additional structures will be required; in other words, where does the CYP1A1 active site fall on the continuum of active site flexibility vs. rigidity?
REFERENCES


Acknowledgments—Portions of this research (RR001209) were carried out at the Stanford Synchrotron Radiation Lightsource, a Directorate of SLAC National Accelerator Laboratory and an Office of Science User Facility operated for the U.S. Department of Energy Office of Science by Stanford University. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the NIH, NIGMS (including GM103393), and NCRR (RR001209). Crystals were grown using the facilities of the Protein Structure Laboratory at the University of Kansas which is supported by NCRR (RR017708) and NIGMS (GM103420). Molecular modeling studies were performed at the Computational Chemistry and Molecular Modeling Laboratory, Department of Basic Pharmaceutical Sciences, School of Pharmacy, West Virginia University, Morgantown, WV. Primary research funding was provided by NIGMS (GM076343 to EES). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS, NCRR, or NIH.

FOOTNOTES
* This work was supported by NIH grant R01 GM076343 to EES.

† To whom correspondence may be addressed: Department of Medicinal Chemistry, The University of Kansas, 1251 Wescoe Hall Dr., Lawrence, KS 66045, USA, Tel: (785) 864-5559; Fax (785) 864-5326; e-mail: eescott@ku.edu

The atomic coordinates and structure factors (code 4I8V) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

Abbreviations: CYP, cytochrome P450; ANF, alpha-naphthoflavone; PAH, polycyclic aromatic hydrocarbons; HCA, heterocyclic aromatic amines; LB, lysogeny broth; TB, terriffic broth; CV, column volume; rmsd, root mean square deviation.

FIGURE LEGENDS

Figure 1. Overall structures. Human CYP1A1 colored from blue N-terminus to red C-terminus with heme as grey sticks and ANF as pink sticks. ANF structure and numbering inset. B. Comparison between CYP1A1 (blue), CYP1A2 (yellow), and CYP1B1 (green) aligned by Ca atoms.

Figure 2. Comparison of the break in the F helix. A. CYP1A1. B. CYP1A2. C. CYP1B1. ANF ligand in grey sticks.

Figure 3. CYP1A1/ANF active site colored as in Fig. 1A showing 2|Fo|-|Fc| electron density map contoured at 1σ (blue mesh) for ANF. A. View face on. B. View edge on. Structural elements surrounding ANF include the I helix (yellow), the F and G helices (green), the B' region (blue), the K/β1-4 loop (orange), and the residues that normally comprise the β4 loop (red).

Figure 4. CYP1A1 (colored as in Fig. 1A) hydrogen bonding (dashed lines) networks around ANF. A. Interactions between residues in the E and F helices (green), I helix (yellow), and the β4 loop (red) above ANF. B. Interactions between the B'/C region (blue) and I helix residues (yellow).

Figure 5. Wall-eye stereo comparisons between the active site residues. A. CYP1A1 (blue) vs. CYP1A2 (yellow). B. CYP1A1 (blue) vs. CYP1B1 (green). Residue numbering is CYP1A1/CYP1A2 and CYP1A1/CYP1B1, respectively. Heme is shown as dark grey sticks with rust sphere for iron.
Figure 6. Comparisons between the active site cavity topologies. A (front view) and B (side view) of CYP1A1 (blue) and CYP1A2 (yellow) cavities. C (front view), D (side view) of CYP1A1 (blue) and CYP1B1 (green) cavities.

Figure 7. Ligands docked into CYP1A1 and CYP1A2 X-ray structures. A. CYP1A1/ANF experimental structure (blue) and CYP1A1 with ANF docked lowest energy pose (grey). B. CYP1A2/ANF experimental structure (yellow) and CYP1A2 with ANF docked lowest energy pose (grey). C. CYP1A1 (blue) with lowest energy pose of phenacetin (black). D. CYP1A2 (yellow) with lowest energy pose of phenacetin (black). E. CYP1A1 (blue) with lowest energy poses of methoxyresorufin (black) and ethoxyresorufin (grey). F. CYP1A2 (yellow) with lowest energy poses of methoxyresorufin (black) and ethoxyresorufin (grey). G. CYP1A1 (blue) with representative poses for B(a)P (black and grey). H. CYP1A1 (blue) with representative poses for resveratrol (black and grey). I. Structures of ligands docked into CYP1A enzymes and sites of metabolism (arrows).
### Table 1. X-ray data collection, refinement, and validation statistics.

<table>
<thead>
<tr>
<th>CYP1A1/α-naphthoflavone</th>
<th><strong>Data collection</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space group</strong></td>
<td>P2₁2₁2₁</td>
<td></td>
</tr>
<tr>
<td><strong>Cell dimensions a, b, c (Å)</strong></td>
<td>65.00, 195.49, 235.86</td>
<td></td>
</tr>
<tr>
<td><strong>Resolution (Å)</strong></td>
<td>38.18 – 2.60 (2.67-2.60)</td>
<td></td>
</tr>
<tr>
<td><strong>Total Reflections</strong></td>
<td>684,437 (47,718)</td>
<td></td>
</tr>
<tr>
<td><strong>Unique Reflections</strong></td>
<td>93,519 (6,776)</td>
<td></td>
</tr>
<tr>
<td><strong>Redundancy</strong></td>
<td>7.3 (7.0)</td>
<td></td>
</tr>
<tr>
<td><strong>Rpim</strong></td>
<td>0.051 (0.508)</td>
<td></td>
</tr>
<tr>
<td><strong>Mn(I/sd)</strong></td>
<td>12.6 (2.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
<td>99.9 (100.0)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Refinement</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Resolution (Å)</strong></td>
<td>38.24-2.60</td>
</tr>
<tr>
<td><strong>No. reflections</strong></td>
<td>88,726</td>
</tr>
<tr>
<td><strong>Rwork / Rfree</strong></td>
<td>18.8/24.3</td>
</tr>
<tr>
<td><strong>Residues modeled for molecules A-D</strong></td>
<td></td>
</tr>
<tr>
<td>A: 36-512</td>
<td></td>
</tr>
<tr>
<td>B: 34-292, 300-512</td>
<td></td>
</tr>
<tr>
<td>C: 37-292, 301-512</td>
<td></td>
</tr>
<tr>
<td>D: 37-291, 301-512</td>
<td></td>
</tr>
<tr>
<td><strong>Wilson B factor</strong></td>
<td>74.6</td>
</tr>
<tr>
<td><strong>Number of atoms / B factor</strong></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>15,083 / 74.2</td>
</tr>
<tr>
<td>Ligand</td>
<td>84 / 83.9</td>
</tr>
<tr>
<td>Heme</td>
<td>172 / 51.2</td>
</tr>
<tr>
<td>Water</td>
<td>181 / 55.2</td>
</tr>
<tr>
<td>Nitrate</td>
<td>12 / 85.0</td>
</tr>
<tr>
<td><strong>R.m.s deviations</strong></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.014</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.666</td>
</tr>
<tr>
<td>Coordinate error, Luzzati plot (Å)</td>
<td>0.325</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Structure analysis</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ramachandran (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Most favored</td>
<td>88.5</td>
</tr>
<tr>
<td>Additional allowed</td>
<td>11.2</td>
</tr>
<tr>
<td>Generously allowed</td>
<td>0.4</td>
</tr>
<tr>
<td>Disallowed</td>
<td>0</td>
</tr>
<tr>
<td>Average RMSD between molecules (Å)</td>
<td>0.41</td>
</tr>
<tr>
<td>Void volumes (Å³)</td>
<td>516, 515, 537, 538</td>
</tr>
</tbody>
</table>

*Highest resolution shell in parenthesis
FIGURE 1.
FIGURE 2.
FIGURE 3.
FIGURE 5.
FIGURE 6.
Human Cytochrome P450 1A1 Structure and Utility In Understanding Drug and Xenobiotic Metabolism
Agnes A. Walsh, Grazyna D. Szklarz and Emily E. Scott

J. Biol. Chem. published online March 18, 2013

Access the most updated version of this article at doi: 10.1074/jbc.M113.452953

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2013/03/18/M113.452953.DC1