STRUCTURE OF THE HUMAN LUNG CYTOCHROME P450 2A13

Brian D. Smith, Jason L. Sanders, Patrick R. Porubsky, Gerald H. Lushington, C. David Stout, Emily E. Scott*

From the Department of Medicinal Chemistry and Molecular Graphics & Modeling Laboratory, University of Kansas, 1251 Wescoe Hall Dr., Lawrence, KS 66045, and the Department of Molecular Biology, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA, 92037

Running title: Structure of Human Lung Cytochrome P450 2A13

The human lung cytochrome P450 2A13 (CYP2A13) activates the nicotine-derived procarcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) into DNA-altering compounds that cause lung cancer. Another cytochrome P450, CYP2A6, is also present in human lung, but at much lower levels. Although these two enzymes are 93.5% identical, CYP2A13 metabolizes NNK with much lower $K_m$ values than does CYP2A6. To investigate the structural differences between these two enzymes the structure of CYP2A13 was determined to 2.35 Å by X-ray crystallography and compared to structures of CYP2A6. As expected, the overall CYP2A13 and CYP2A6 structures are very similar with an average r.m.s. deviation of 0.5 Å for the $Ca$ atoms. Like CYP2A6, the CYP2A13 active site cavity is small and highly hydrophobic with a cluster of Phe residues composing the active site roof. Active site residue N297 is positioned to hydrogen bond with an adventitious ligand, identified as indole. Amino acid differences between CYP2A6 and CYP2A13 at positions 117, 300, 301, and 208 relate to different orientations of the ligand plane in the two protein structures and may underlie the significant variations observed in binding and catalysis of many CYP2A ligands. In addition, docking studies suggest that residues 365 and 366 may also contribute to differences in NNK metabolism.

The two functional members of the human P450 2A gene family, CYP2A6 and CYP2A13, differ in only 32 of their 494 amino acids (Supplementary figure 1). Both metabolize substrates including coumarin, nicotine, and the nicotine-derived procarcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), the most abundant and potent lung carcinogen in tobacco. Additionally, CYP2A enzymes are reported to metabolize the mycotoxin aflatoxin B$_1$ (1,2). Both NNK and aflatoxin B$_1$ require activation to generate the reactive intermediates that form DNA adducts and ultimately result in carcinogenicity.

Several key findings indicate that 2A13, rather than 2A6, plays an important role in $in situ$ metabolism of NNK in the human respiratory tract. First, substantial amounts of CYP2A13 mRNA are generated in the respiratory tract, especially the nasal mucosa, trachea, and lung (3). By comparison, several studies reported either significantly lower CYP2A6 mRNA levels in respiratory tract tissues (3-5) or were unable to detect CYP2A6 mRNA at all (6). Recently a CYP2A13-specific antibody was used to confirm the presence of CYP2A13 protein in the epithelia of bronchus and trachea (7). Second, although CYP2A6 and CYP2A13 act on common substrates, rates of metabolism can differ significantly. Studies with recombinant enzymes have demonstrated that CYP2A13 is 30-215 times
more efficient at activating NNK into its carcinogenic metabolites than CYP2A6 (3,8,9). Although cytochromes P450 2E1, 2D6, and 3A4 have also been shown to metabolize NNK in vitro, their $K_m$ values are much higher than $K_m$ values for the 2A enzymes (10). Finally, a substantial reduction in lung adenocarcinoma has been associated with a genetic polymorphism of CYP2A13 (11) that is reported to cause a 2 to 3-fold reduction of catalytic efficiency for NNK (12). Thus, activation of NNK by CYP2A13 in the lung may be a potent determinant for lung adenocarcinoma in smokers, which is the leading cause of cancer death. CYP2A13 is also known to preferentially activate a number of other xenobiotics. CYP2A13 is more than four times more active than CYP2A6 in the metabolism of hexamethylyphosphoramide (HMPA), $N,N$-dimethylaniline (DMA), and 2'-methoxyacetophenone (MAP) (3). Additionally, a recent report demonstrated CYP2A13 metabolic activation of aflatoxin $B_1$ into both aflatoxin $B_1$ and aflatoxin $M_1$ 8,9-epoxides while CYP2A6 produced neither carcinogenic metabolite (1).

In order to better define the structural features of CYP2A enzymes that determine their substrate metabolism, and to specifically investigate the structural differences between human CYP2A6 and CYP2A13 enzymes, we have determined a structure of human CYP2A13 by X-ray crystallography. Careful comparison of the CYP2A13 active site with previously determined structures of CYP2A6 (13) allows identification of a set of amino acid residues that are likely to modulate the observed differences in human CYP2A function.

**Experimental Procedures**

**Protein design, expression, and purification:** The cDNA for CYP2A13 was a gift from Dr. X. Ding (Wadsworth Center, Albany, NY). The full-length gene was altered to yield protein (2A13dH) that lacks the N-terminal transmembrane sequence (A2-23), has several charged residues substituted at the modified N-terminus (24-WRQRKSR-30 to 24-AKKTSSK-30) and has four histidine residues added at the C-terminus (Supplementary Figure 1). This altered gene was inserted into the pKK233-2 plasmid (Pharmacia, Uppsala, Sweden). 2A13dH was then expressed in *E. coli* TOPP-3 cells (Stratagene, La Jolla, CA) with an induction time of 48 hours as described (14), but without the addition of imidazole.

*E. coli* cells were harvested and disrupted as described (15). After centrifugation to pellet cell debris, 4.8 mM Cymal-5 detergent (Anatrace, Maumee, OH) and 0.3 M NaCl were added to the supernatant, followed by ultracentrifugation at 30,000 rpm for 60 minutes. The solubilized CYP2A13 lysate was applied to Ni$^{2+}$-agarose resin (Qiagen, Valencia, CA), which was subsequently washed with the loading buffer. The resin was washed first with 100 mM potassium phosphate buffer (100 mM potassium phosphate, pH 7.4, 20% glycerol, 200 mM NaCl, 10 mM 2-mercaptoethanol, 0.5 mM PMSF, 4.8 mM Cymal-5) and then with the same buffer including 6 mM histidine. CYP2A13 was eluted using 10 mM potassium phosphate buffer with NaCl reduced to 100 mM, and supplemented with 2 mM EDTA and 40 mM histidine. The purest P450-containing fractions were pooled, diluted 10X with 5 mM potassium phosphate, pH 7.4, 20% glycerol, 1 mM EDTA, 0.2 mM DTT, 1 mM PMSF, and 4.8 mM Cymal-5, and loaded onto a CM-sepharose CL-6B column. This column was washed with the previous buffer omitting the detergent. Purified CYP2A13 was eluted using the same buffer with potassium phosphate increased to 50 mM and including 500 mM NaCl. Purified protein was concentrated using centrifugal ultrafiltration.

**Protein crystallization, data collection, and structure determination:** Crystals were grown using hanging drop vapor diffusion equilibration. A 50 mg/ml solution of purified protein in 50 mM potassium phosphate buffer, pH 7.4, 20% glycerol, 1 mM PMSF, 0.2 mM DTT, 1 mM EDTA, 0.5 M NaCl, and 0.066 mM n-tridecyl-$\beta$-D-maltopyranoside was equilibrated against 0.3 M (NH$_4$)$_2$SO$_4$, 0.15 M sodium HEPES, pH 7.2, and 25% polyethylene glycol 2000 monomethyl ether at 20 degrees C. Plate-like crystals that grew in 10 days were immersed in mother liquor supplemented with 30% glycerol, and frozen in liquid nitrogen prior to data collection. A single native data set was collected on beamline 11-1 at the Stanford Synchrotron Radiation Laboratory (Stanford, CA). Data were recorded in two passes, a high resolution pass (180°, 1° oscillations, 10 s exposures) and a low
resolution pass (90°, 1° oscillations, 2 s exposures) using an ADSC Q315 detector. The data to 2.35 Å were integrated and processed using the programs Mosflm and Scala (16). Statistics are shown in Table 1.

The 2A13dH structure was solved by molecular replacement using the 4-(4-chlorophenyl)imidazole complex of cytochrome P450 2B4 (1SUO) as a search model and Phaser (17). The space group is P2, a very rare space group for proteins, but this assignment was based on the presence of 0k0 reflections with k odd. In addition, molecular replacement calculations gave the highest log likelihood (1443) for a solution with six molecules in the asymmetric unit that packed well without space for additional molecules. The initial experimental electron density map reinforced the space group assignment and the molecular replacement solution. The model was refined using CNS (18). Rigid body refinement was followed by iterative rounds of positional and isotropic B-factor minimization with constrained NCS. Xtalview (19) was used for model building into oA-weighted 2Fo – Fc electron density maps. During the last several iterations of refinement, NCS was constrained with decreasing weights, as validated by decreases in Rfree.

The final model contains residues 31-494, heme, indole, and 591 water molecules. The crystallographic R factor is 21.8 and the Rfree is 27.6. Residues in most favored regions of the Ramachandran plot include 86.8% of the structure, with 12.2% in additionally allowed regions, 0.8% in generously allowed regions, and 0.2% in disallowed regions. Coordinates have been deposited in the Protein Data Bank (PDB ID 2P85).

Docking of NNK: Ligand structures were sketched and protonated in SYBYL (20) and were then structurally optimized using the Tripos Force Field (21) (Gasteiger-Marsili electrostatics (22) with an 8.0 Å nonbonding interaction cutoff; default convergence thresholds). The receptor structures were prepared from crystallographic coordinates of CYP2A6 [13] and CYP2A13 (reported herein) by protonating the structures and adding Gasteiger-Marsili charges in SYBYL, and aligning the two according to conserved backbone residues. The coumarin ligand in the CYP2A6 structure was then transposed into the CYP2A13 receptor, and the three crystallographic waters with close contacts to the transposed coumarin were deleted. Both CYP2A6 and CYP2A13 were then conditioned in the presence of bound coumarin via short (1 ps; 300K; Tripos Force Field) molecular dynamics simulations. The coumarin ligands were then removed from the receptor. Coumarin was then redocked into these two receptors for validation purposes via the FlexX program (23) taking into explicit consideration all receptor residues within 8.0 Å of the ligand (including the heme and remaining crystallographic waters, but neglecting the co-crystallized ligand) and using Gasteiger-Marsili charges for both ligand and receptor. Generation of docked conformers or poses was accomplished via the ligand fragment search and reassembly algorithm implemented in FlexX. The Asn297 side chain amide was specified as a critical H-bond acceptor site. No other preconditions were set on the ligand position or orientation. Among the 30 ligand docked conformers requested per ligand, 15 of the coumarin structures bound to CYP2A6 were in close qualitative agreement with the crystallographically determined position (root mean squared deviation <1.2 Å), with the carbonyl oxygen forming a hydrogen bond with the Asn297 side chain and the ring oxygen point up away from the heme. In the remaining 15 structures, the above H-bond was preserved, however the ring oxygen was found to point down toward the heme. The same docking methodology was then applied to NNK binding to CYP2A6 and CYP2A13.

Results and Discussion

X-ray Structure Determination and Overview

To better understand the differences between CYP2A6 and CYP2A13 ligand binding and metabolism, the structure of CYP2A13 was determined by X-ray crystallography. Truncated and His-tagged cytochrome P450 2A13dH crystallized in the P2 space group with six molecules in the asymmetric unit. The initial experimental maps were easily interpreted for most of the protein and the model refined using iterative model adjustment and refinement. Statistics for data collection and refinement are shown in Table 1.
The CYP2A13 structure follows the typical fold for mammalian cytochromes and is composed of 20 $\alpha$-helices and 4 $\beta$-sheets enclosing the heme prosthetic group (Figure 1), with no apparent open channels. The six molecules in the asymmetric unit are very similar (average C$\alpha$ r.m.s. deviation 0.4 Å). The overall CYP2A13 structure is very similar to that of the closely related human liver cytochrome P450 2A6 (Figure 1). The r.m.s. deviation for C$\alpha$ atoms in CYP2A13 vs. CYP2A6 (1Z10) is 0.5 Å, only slightly higher than the differences among the six different CYP2A13 molecules in the asymmetric unit or the four CYP2A6 molecules present in that asymmetric unit. Four of the six CYP2A13 molecules have the same packing as the four molecules in the CYP2A6 asymmetric unit. The secondary structure units are essentially identical between the two CYP2A enzymes (Supplementary figure 1).

Active site

The CYP2A13 active site is very hydrophobic and tightly packed. The cavity is small and planar, complementing the size and shape of many CYP2A substrates. The planar active site is oriented edge-on to the heme group. The active site of CYP2A13 is 307 Å$^3$, some 15-20% larger that that of the CYP2A6 structures and proportioned differently as a result of individual amino acid differences lining the cavity (Figure 2). The two CYP2A enzymes have the smallest active sites of the human cytochromes P450 whose structures have been determined. Amino acids lining the CYP2A13 active site include F107, F111, A117, F118, F209, L296, N297, F300, A301, E304, T305, M365, L366, L370, and F480.

Notably, six of the amino acids lining the active site are phenylalanine residues: F107, F111, F118, F209, F300, F480 (Figure 2). These phenylalanines cluster to form the “roof” of the active site. This cluster is even larger than the phenylalanine cluster in CYP2A6, as I300 in CYP2A6 is F300 in CYP2A13. The aromatic side chains substantially fill the active site and provide multiple opportunities for parallel-displaced $\pi$-$\pi$ stacking and edge-to-face interactions with aromatic ligands and with each other to stabilize the tertiary protein structure. Negishi and coworkers demonstrated that mutation at positions 209 or 481 substantially altered steroid hydroxylase activity in mouse 2A enzymes (24). Although F209 interacts directly with substrates in the active site cavity, residue 481 does not line the CYP2A13 active site. However, in CYP2A13, the side chain of A481 projects directly toward F209. In the CYP2A4 and CYP2A5 studies, mutation at 481 likely altered the packing of F209 and/or the adjacent F480, both of which do line the active site cavity, to alter metabolism indirectly.

Although no exogenous ligands were added to the purified CYP2A13, initial experimental maps indicated the presence of a ligand in the active sites of all six molecules of the asymmetric unit. During refinement, subsequent electron density maps reinforced the presence of this strong planar density. The modest resolution of the X-ray diffraction experiment did not allow identification of the bound molecule in the crystals. However, we have independently determined that the ligand is indole. CYP2A6 is one of several cytochromes P450 reported to metabolize indole into various indigoid pigments (25). In fact, when human CYP2A6 and the natural redox partner are overexpressed in E. coli, the bacterial cultures turn blue as a result of indigoid production from indole normally present in the E. coli cells. Likewise, when CYP2A13 is expressed in E. coli, the cultures also generate indigoid pigments, likely by accepting electrons from an endogenous E. coli reductase. We suspected that indole co-purified with the protein and constituted the ligand in the active site. First, early in the purification when large amounts of indigoid pigments are still present, UV/Vis spectra indicate Soret peaks at both 393 nm and 418 nm, indicating a mix of CYP2A13 molecules, some with a ligand and some with water in the active site, respectively. The spectrum of the purified protein has an absorbance maximum at 418 nm, with only a slight shoulder toward 393 nm, but indole does not cause a significant shift in the Soret at concentrations less than 40-fold over the CYP2A13 protein concentration. Second, the overall size, shape, and planarity of the bicyclic indole ring system correlate well with the observed electron density in the active site (Figure 3). Third, CYP2A13 purified for crystallization reacts with p-dimethylaminocinnamaldehyde to yield a product with an absorbance at 625 nm, indicative of the presence of indole and/or certain indole metabolites.
Finally, we were able to extract the 625 nm $p$-dimethylaminocinnamaldehyde product from purified CYP2A13 protein and identify it as a complex with unsubstituted indole by mass spectrometry. On the basis of these results, we have modeled the electron density in the present CYP2A13 structure as indole.

This strongly planar indole density in the CYP2A13 structure appears to adopt two alternate orientations in the active site. For most molecules, the predominant orientation overlaps the location of the coumarin molecule in CYP2A6 and similarly forms a direct hydrogen bond to N297 (indole position A, Figure 3a). However, one of the active sites has very strong density supporting a second orientation of indole with the benzene ring in the same location, but with the fused pyrrole ring more perpendicular to the heme plane. In this orientation indole also binds to Asn297, this time via a bridging water molecule (indole position B, Figure 3b). In addition to the electron density, this second orientation is also supported by molecular dynamics simulations at 100K that indicate the bridging arrangement is a stable one. Subsequent refinement of the occupancies of both ligands in each active site indicated that all active sites likely contain a mix of indole in orientations A and B ranging from 60/40 to 40/60 (Figure 3c). Dual ligand orientations have been observed previously in a complex of CYP2C5 with 4-methyl-N-nitro-2H-pyrazol-3-yl)benzenesulfonylamide (27). In both the CYP2A6 and CYP2A13 structures, N297 is the only polar residue in the active site and this amino acid likely plays a key role in ligand orientation. N297 is stabilized by a hydrogen bonding network including a water molecule and the backbone of residues in the loop between helices B' and C in CYP2A13 as well as CYP2A6. Asparagine is highly conserved at position 297 in CYP2A enzymes from many different organisms, but not in other subfamilies of human cytochromes P450. Thus, this hydrogen bond interaction may be a hallmark of CYP2A enzyme interactions with their ligands and responsible for orientation of many CYP2A ligands in the active site.

Although the indole in CYP2A13 binds in nearly the same location as coumarin in CYP2A6 with respect to the heme and N297, the plane of coumarin and indole differ by approximately 30°. When viewed end-on from opposite N297, coumarin is oriented along a northwest/southeast axis, while both of the CYP2A13 indole ligands are oriented along a northeast/southwest axis (Figure 4). Structural proximity implicates four amino acid residues in orientation of the ligand plane. In CYP2A6, I300 projects into the active site cavity from the northeast, limiting space available to the ligand in this quadrant. In CYP2A13, residue 300 is phenylalanine, but the larger side chain is torsioned away from the ligand, resulting in more space in the northwest quadrant of the CYP2A13 active site. Rotation of the aromatic ring away from the active site is likely permitted because of a “second shell” residue difference between CYP2A6 and CYP2A13 at the C-terminal end of the F helix. Located directly above F300 in the active site, the reduction in size from I208 in CYP2A6 to S208 in CYP2A13 both provides space for the aromatic ring torsion in CYP2A13 and repositions the aromatic ring of F209. The CYP2A13 S208I mutation has been shown to significantly decrease NNK metabolism to both metabolites (9), suggesting it might also be key in orienting other ligands. The identity of the residue at position 209 is key for determining metabolism with coumarin, steroids, and aflatoxin B$_1$ in mouse CYP2A enzymes (2,28,29).

At the base of the active site near the heme, residues 301 and 117 also act in concert to control the ligand orientation. In CYP2A13, the A301 side chain methyl occupies space that is available to ligands in CYP2A6, which has glycine at this position. On the opposite side of the active site, residue 117 is an alanine in CYP2A13 and occupies less space than the corresponding valine at this position in CYP2A6. Although there has been little exploration of the role of residue 301 in CYP2A enzymes, mutation at position 117 has been demonstrated to substantially impact the catalytic efficiency of CYP2A13 for coumarin (30), NNK (9), and aflatoxin B$_1$ (1). For coumarin and NNK the effects are primarily reflected in changes in $V_{\text{max}}$ rather than $K_m$, suggesting orientation of the ligand may be related to catalysis rather than ligand affinity. Additionally, variants at this position are responsible for significant differences in coumarin hydroxylation in mouse 2A enzymes (31).

Additional active site differences between CYP2A6 and CYP2A13 occur at positions 365 and
366. These residues, along with T305, are located at the end of the active site opposite N297. In CYP2A6, the side chains of T305 and L366 approach each other to close off the end of the active site, with V365 just behind and shielded from the active site cavity. In CYP2A13, the larger methionine at 365 extends much closer towards the ligand and the flanking residues T305 and L366 are farther separated. As a result, the terminal atoms of M365 make up part of the active site wall. Although these residues are more distant from the ligand in this structure, they are likely to play a role in orienting slightly larger ligands. The identity of the side chain at position 365 in mouse CYP2A4 and CYP2A5 has a role in metabolism of coumarin, testosterone, and aflatoxin B1 (2.32).

Thus, both structure and function studies support the idea that a combination of the residues at positions 117, 208, 300, 301, 365, and possibly 366 modulate ligand binding and metabolism in CYP2A enzymes. Analysis of these residues in different CYP2A enzymes may help understand phenacetin metabolism. CYP2A13 has recently been reported (33) to metabolize phenacetin, normally a CYP1A2 marker substrate, more efficiently than CYP1A2 itself. Rabbit CYP2A10 and CYP2A11 also have this capacity, but CYP2A6 does not. Comparison of the active site residues at key positions where CYP2A13 and CYP2A6 differ (Table 2) suggests that in 2A enzymes residues S208, F300, A301, and M365 are compatible with phenacetin metabolism.

Heme Binding

CYP2A6 and CYP2A13 also differ in two of the residues responsible for interaction with the heme. In CYP2A6, S369 hydrogen bonds through its side chain hydroxyl group to the A ring propionate. Hydrogen bonding residues are generally conserved at this position in microsomal CYPs. CYP2A13, however, has a glycine at this position and is thus incapable of the hydrogen bonding interaction. In the place of the serine side chain, CYP2A13 has a water molecule ideally positioned to replace this interaction. The water molecule is stabilized by strong hydrogen bonds to the both the backbone amine and carbonyl of residue 370.

The residue at position 372 also interacts with the A ring propionate and differs between the two CYP2A proteins. This residue is histidine in 70% of the family 1 and family 2 cytochromes P450, including CYP2A13. The imidazole ring of H372 hydrogen bonds to the terminal propionate oxygens. In CYP2A6 and the remaining 30% of the family 1 and 2 enzymes, this residue is an arginine. In CYP2A6, the Nε of R372 takes the place of the imidazole nitrogen. From the structures, it appears that the two arrangements might be equivalent. Three site-directed mutagenesis studies, however, indicate that this residue is functionally important. The CYP2A13 mutation H372R disrupts the enzyme’s ability to form either of the carcinogenic aflatoxin 8,9 epoxide metabolites and substantially increased formation of the detoxification products aflatoxin P1 (1). In metabolism of NNK, the H372R mutation obliterates the ability to form the keto alcohol metabolite, and reduces the kcat for formation of the keto aldehyde by almost 130-fold compared to the CYP2A13 wild type enzyme (9), essentially converting the metabolite profile for NNK to that of CYP2A6. Additionally, CYP2A13 H372R has an 8-fold increase in the kcat for coumarin to near 2A6-levels, while introduction of the R372H mutation in CYP2A6 reduces its coumarin kcat 5-fold to ~2A13 levels (30).

Docking of NNK Into the Active Site

Of all human cytochrome P450s, the human lung-specific CYP2A13 has been reported as the most effective activator of nicotine-derived NNK. 2A13 activates NNK at both the methyl and methylene positions to form DNA pyridyloxobutylating and methylating intermediates that are ultimately detected as a keto alcohol and a keto aldehyde, respectively (32-34). CYP2A6 is much less efficient at NNK activation (Table 3). When NNK was docked into the CYP2A6 structure, all 30 docked conformers predicted a hydrogen bond between the NNK carbonyl oxygen and the Asn297 side chain and resulted in an orientation with the pyridine ring too close to the heme iron to permit close approach by either the methylene and methyl carbons (no approaches closer than 7.0 and 5.5 Å respectively, Figure 5a). This orientation is incompatible with generation of either carcinogenic metabolite. In contrast, when NNK was docked into the CYP2A13 structure, NNK generally oriented with the pyridine ring oriented away from the heme iron (26/30 docked conformers) with its nitrogen hydrogen bonded to Asn297. Nineteen of these
docked conformers placed the methyl carbon nearest to the heme iron (as close as 3.9 Å, Figure 5b). Another seven of the docked conformers placed the methylene carbon nearest to the heme iron (as close as 4.9 Å, Figure 5c). These orientations are suitable for hydroxylation by the activated oxygen intermediate into the keto alcohol and keto aldehyde metabolites observed experimentally. In the remaining four CYP2A13 docked conformers the nitroso oxygen hydrogen bonded to Asn297 and the pyridine ring was pointed toward the heme, similar to that observed in all of the CYP2A6 conformers. The energies of all CYP2A6 and CYP2A13 docking conformers were similar. Thus, subtle differences in the structures modify ligand orientation in docking experiments to correctly predict the metabolites observed (Table 3). In particular, residues at positions 366 and 365 differ between CYP2A13 and CYP2A6 and have very different interactions with NNK in CYP2A6 compared to CYP2A13. The combination and positioning of M365/L366 in CYP2A13 creates slight additional space adjacent to these residues that is not available adjacent to V365/I366 in CYP2A6 (Figure 2). When NNK is positioned for methyl hydroxylation in CYP2A13, the terminal nitrosamine functional group occupies this small pocket (Figure 5b). Rotation about the C-C bond adjacent to the nitrosamine would then yield the orientation of NNK suitable for methylene hydroxylation (Figure 5c). Neither of these two orientations appears to be available to NNK in CYP2A6, potentially due to the replacement of M365/L366 with V365/I366.

Conclusions

Since lung cancer is the most common cancer and the leading cause of cancer-related death (37), understanding in situ activation of nicotine-derived NNK into carcinogens by CYP2A13 is an important goal. Comparison of the human CYP2A13 and CYP2A6 structures reveals several common features that may be hallmarks of CYP2A enzymes: 1) a relatively small, planar active site, 2) numerous active site phenylalanine residues for π-π with aromatic ligands, and 3) N297 oriented for hydrogen bonding with ligands. Differences between the two structures indicate that residues 117, 208, 300, and 301 act in concert to define the plane of ligands in the active site. Residue differences at positions 365 and/or 366 are also likely determinants of ligand binding and catalysis, especially for NNK, while the structure suggests residues at 369 and 372 might be more directly involved in heme binding. Most of the remaining differences between CYP2A6 and CYP2A13 are located outside the active site either in the N-terminal transmembrane region or on the surface of the protein. Mutagenesis results support the effects of residues at 117, 208, and 365, while the roles of residues at positions 297, 300, and 301 should be further explored by mutagenesis and functional studies.

Acknowledgments

We thank Dr. X. Ding (Wadsworth Center, Albany, NY) for the gift of CYP2A13 cDNA. We would like to acknowledge access to the facilities and support staff of SSRL. The SSRL is operated by the Department of Energy, Office of Basic Energy Sciences. The SSRL Biotechnology Program is supported by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and by the Department of Energy, Office of Biological and Environmental Research. We also thank Dr. Audrey Lamb for assistance with Phaser. This work was supported by the NIH grants RR016475 (JLS and PRP), GM61545 (CDS), RR017708 (EES), and GM076343 (EES).
References

1. He, X. Y., Tang, L., Wang, S. L., Cai, Q. S., Wang, J. S., and Hong, J. Y. (2006) Int. J. Cancer. 118, 2665-2671
2. Pelkonen, P., Lang, M. A., Negishi, M., Wild, C. P., and Juvonen, R. O. (1997) Chem. Res. Toxicol. 10, 85-90
3. Su, T., Bao, Z., Zhang, Q. Y., Smith, T. J., Hong, J. Y., and Ding, X. (2000) Cancer Res. 60, 5074-5079
4. Mace, K., Bowman, E. D., Vautravers, P., Shields, P. G., Harris, C. C., and Pfeifer, A. M. (1998) Eur. J. Cancer 34, 914-920
5. Crawford, E. L., Weaver, D. A., DeMuth, J. P., Jackson, C. M., Khuder, S. A., Frampton, M. W., Utell, M. J., Thilly, W. G., and Willey, J. C. (1998) Carcinogenesis 19, 1867-1871
6. Raunio, H., Hakkola, J., Hukkanen, J., Pelkonen, O., Edwards, R., Boobis, A., and Anttila, S. (1998) Arch. Toxicol. Suppl. 20, 465-469
7. Zhu, L. R., Thomas, P. E., Lu, G., Reuhl, K. R., Yang, G. Y., Wang, L. D., Wang, S. L., Yang, C. S., He, X. Y., and Hong, J. Y. (2006) Drug Metab. Dispos. 34, 1672-1676
8. Jalas, J. R., Ding, X., and Murphy, S. E. (2003) Drug Metab. Dispos. 31, 1199-1202
9. He, X. Y., Shen, J., Ding, X., Lu, A. Y., and Hong, J. Y. (2004) Drug Metab. Dispos. 32, 1516-1521
10. Patten, C. J., Smith, T. J., Murphy, S. E., Wang, M. H., Lee, J., Tynes, R. E., Koch, P., and Yang, C. S. (1996) Arch. Biochem. Biophys. 333, 127-138
11. Wang, H., Tan, W., Hao, B., Miao, X., Zhou, G., He, F., and Lin, D. (2003) Cancer Res. 63, 8057-8061
12. Zhang, X., Su, T., Zhang, Q. Y., Gu, J., Caggana, M., Li, H., and Ding, X. (2002) J. Pharmacol. Exp. Ther. 302, 416-423
13. Yano, J. K., Hsu, M. H., Griffin, K. J., Stout, C. D., and Johnson, E. F. (2005) Nat. Struct. Mol. Biol. 12, 822-823
14. Scott, E. E., Spatzenegger, M., and Halpert, J. R. (2001) Arch. Biochem. Biophys. 395, 57-68
15. Wester, M. R., Stout, C. D., and Johnson, E. F. (2002) Methods Enzymol. 357, 73-79
16. CCP4. (1994) Acta Cryst. D 50, 760-763
17. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C., and Read., R. J. (2005) Acta Cryst. D 61, 458-464
18. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Gross-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Acta Cryst. D 64, 905-921
19. McRee, D. E. (1999) J. Struct. Biol. 125, 156-165
20. Associates', T. T. (2006) SYBL 7.2. In., St. Louis, MO
21. Clark, M., III, R. D. C., and Opdenbosch, N. V. (1989) Journal of Computational Chemistry 10, 982-1012
22. Gasteiger, J., and Marsili, M. (1980) Tetrahedron 26, 3219-3228
23. Rarey, M., Kramer, B., and Lengauer, T. (1997) J Comput Aided Mol Des 11, 369-384
24. Iwasaki, M., Darden, T. A., Pedersen, L. G., and Negishi, M. (1995) Biochemistry 34, 5054-5059
25. Gillam, E. M., Notley, L. M., Cai, H., De Voss, J. J., and Guengerich, F. P. (2000) Biochemistry 39, 13817-13824
26. Turner, J. M. (1961) *Biochem. J.* **78**, 790-792
27. Wester, M. R., Johnson, E. F., Marques-Soares, C., Dansette, P. M., Mansuy, D., and Stout, C. D. (2003) *Biochemistry* **42**, 6370-6379
28. Iwasaki, M., Lindberg, R. L., Juvonen, R. O., and Negishi, M. (1993) *Biochem. J.* **291**, 569-573
29. Uno, T., Mitchell, E., Aida, K., Lambert, M. H., Darden, T. A., Pedersen, L. G., and Negishi, M. (1997) *Biochemistry* **36**, 3193-3198
30. He, X. Y., Shen, J., Hu, W. Y., Ding, X., Lu, A. Y., and Hong, J. Y. (2004) *Arch. Biochem. Biophys.* **427**, 143-153
31. Lindberg, R. L., Juvonen, R., and Negishi, M. (1992) *Pharmacogenetics* **2**, 32-37
32. Lindberg, R. L., and Negishi, M. (1989) *Nature* **339**, 632-634
33. Fukami, T., Nakajima, M., Sakai, H., Katoh, M., and Yokoi, T. (2006) *Drug Metab. Dispos.*
34. Wang, S. L., He, X. Y., Shen, J., Wang, J. S., and Hong, J. Y. (2006) *Toxicol. Sci.* **94**, 38-45
35. Scott, E. E., White, M. A., He, Y. A., Johnson, E. F., Stout, C. D., and Halpert, J. R. (2004) *J. Biol. Chem.* **279**, 27294-27301
36. 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) *Proc. Natl. Acad. Sci. USA* **100**, 13196-13201
37. Society, A. C. (2004) Cancer Facts and Figures, 2004. In: Office, A. C. S. N. H. (ed). *Cancer Facts and Figures*, American Cancer Society, Atlanta, GA
38. Kleywegt, G. J., and Jones, T. A. (1994) *Acta Crystallogr.* **D50**, 178-185
39. Kabsch, W., and Sander, C. (1983) *Biopolymers* **22**, 2577-2637
Footnote: NNK, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone; PMSF, phenylmethanesulfonyl fluoride; DTT, DL-dithiothreitol.
**Figure Legends**

Figure 1. CYP2A13 structure (blue) overlaid with the CYP2A6 structure (grey). Heme is shown in red sticks and sphere (Fe).

Figure 2. Comparison of active site volumes of CYP2A6 (yellow mesh) and CYP2A13 (blue mesh) surrounded by active site residues of CYP2A13 (blue ribbons and sticks). Heme, red sticks. The solvent-occupied surface of the active site cavities were calculated using VOIDOO (38) with a 1.4 Å probe and a mesh of 0.33.

Figure 3. Indole binding in the CYP2A13 active site. Panel A, indole in position A in CYP2A13 molecule C of the asymmetric unit. Panel B, indole in position B in CYP2A13 molecule F of the asymmetric unit. Panel C. The indole density could be best fit by partial occupancy of indole in two positions with the benzene rings overlapping (CYP2A13 molecule A shown). Indole in position A, pink sticks; indole in position B, cyan sticks; composite omit $\sigma_A$-weighted $2|F_o|-|F_c|$ electron density contoured at 1.0 $\sigma$ around indoles, blue mesh; heme, red sticks and spheres; components of active site walls and active site residues colored by tertiary structure component: B’ helix and adjacent loops, blue; I helix, yellow; F helix, orange; loop between K helix and $\beta_1-4$, green; $\beta_{4-1}/\beta_{4-2}$ turn, magenta.

Figure 4. The CYP2A13 protein (blue ribbons and sticks), heme (red sticks), and indole A (cyan sticks) are shown overlaid with selected CYP2A6 residues (green sticks) that reorient plane of coumarin (green sticks) in the active site. CYP2A13 indole B is not shown for clarity, but adopts a very similar plane to indole A.

Figure 5. NNK docking into CYP2A6 and CYP2A13. Panel A, NNK docked into the CYP2A6 structure assumes a nonproductive orientation with the methylene and methyl carbons no closer than 7.0 and 5.5 Å from the heme iron, respectively. In contrast, NNK docked into the CYP2A13 structure assumes an orientation with either the methyl (Panel B) or methylene (Panel C) oriented for hydroxylation into the two carcinogenic metabolites observed experimentally. Black circles highlight methyl and methylene carbons hydroxylated to form carcinogenic metabolites. Residues and helices shown as in Figure 3.
Table 1. Data collection and refinement statistics.

| Crystal |  |
|---------|---------|
| Construct | 2A13dH |
| Space group | P2 |
| Unit cell (a, b, c) (Å) | 121.32, 110.28, 142.07 |
| (α, β, γ) (°) | 90, 110.29, 90 |

| Data collection |  |
|----------------|---------|
| SSRL beam line | 11-1 |
| Wavelength (Å) | 0.98 |
| Resolution range (Å) | 29.34 – 2.35 |
| Total/unique observations | 660,337/142,896 |
| Completeness (%)<sup>a</sup> | 97.3 (97.3) |
| <I/σ(I)><sup>a</sup> | 14.5 (2.9) |
| Multiplicity<sup>a</sup> | 4.7 (3.7) |
| R<sub>sym</sub><sup>a</sup> | 0.073 (0.431) |

| Refinement |  |
|------------|---------|
| R (%) | 21.8 |
| R<sub>free</sub> (%) | 27.6 |
| r.m.s. deviations |  |
| Bonds (Å) | 0.009 |
| Angles (°) | 1.37 |
| Model statistics |  |
| No. atoms | Average B factor (Å<sup>2</sup>) | Occupancy in position A/B<sup>c</sup> |
| Protein | 22,584 | 45.7 |  |
| Heme | 258 | 30.0 |  |
| Water | 597 | 42.6 |  |
| Molecule A indoles | 18 | 31.5 | 60/40 |
| Molecule B indoles | 18 | 31.5 | 40/60 |
| Molecule C indoles | 18 | 31.5 | 60/40 |
| Molecule D indoles | 18 | 31.5 | 50/50 |
| Molecule E indoles | 18 | 31.5 | 50/50 |
| Molecule F indoles | 18 | 31.5 | 40/60 |

<sup>a</sup>Values for the highest-resolution shell in parenthesis. <sup>b</sup>B values were defined as the average of the B values for the amino acid residues immediately surrounding the ligands. <sup>c</sup>Position A is defined as direct hydrogen bonding of the indole to Asn297. Position B is defined as hydrogen bonding of indole to Asn297 via a bridging water molecule.
Table 2. Comparison of key active site residues in 2A enzymes.

| CYP  | 117 | 208 | 300 | 301 | 365 | 366 |
|------|-----|-----|-----|-----|-----|-----|
| 2A6  | V   | I   | I   | G   | V   | I   |
| 2A10 | V   | S   | F   | A   | M   | I   |
| 2A11 | A   | S   | F   | A   | M   | I   |
| 2A13 | A   | S   | F   | A   | M   | L   |
Table 3. Comparison of CYP2A13 and CYP2A6 kinetic parameters and docking orientations.

|                      | NNK α-methyl hydroxylation |                      | NNK α-methylene hydroxylation |                      |
|----------------------|-----------------------------|----------------------|-------------------------------|----------------------|
|                      | \(K_m\) (µM)\(^{a}\) | \(V_{max}\) \(^{a}\) | \(V_{max}/K_m\) \(^{a}\) | # compatible docking conformers | \(K_m\) (µM)\(^{a}\) | \(V_{max}\) \(^{a}\) | \(V_{max}/K_m\) \(^{a}\) | # compatible docking conformers |
| CYP2A6               | 141.0                       | 0.4                  | 0.003                         | 0/30                 | 118.0                       | 1.0                  | 0.008                         | 0/30                 |
| CYP2A13              | 13.1                        | 1.2                  | 0.09                          | 19/30                | 11.3                        | 4.1                  | 0.36                          | 7/30                 |

\(^{a}(3)\)
Figures

Figure 1.
Figure 3.
Figure 4.
Figure 5.
