THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 281, NO. 11, pp. 7614–7622, March 17, 2006

Crystal Structure of Human Cytochrome P450 2D6*[S]

Received for publication, October 16, 2005, and in revised form, November 22, 2005 Published, JBC Papers in Press, December 13, 2005, DOI 10.1074/jbc.M511232200

Paul Rowland[1,2], Frank E. Blaney[1], Martin G. Smyth, Jo J. Jones, Vaughan R. Leydon, Amanda K. Oxbrow, Ceri J. Lewis, Mike G. Tennant[3], Sandeep Modi, Drake S. Eggleston, Richard J. Chenery, and Angela M. Bridges[1]

From the Department of Discovery Research, GlaxoSmithKline, New Frontiers Science Park, Third Avenue, Harlow, Essex, CM19 5AW, United Kingdom

Cytochrome P450 2D6 is a heme-containing enzyme that is responsible for the metabolism of at least 20% of known drugs. Substrates of 2D6 typically contain a basic nitrogen and a planar aromatic ring. The crystal structure of human 2D6 has been solved and refined to 3.0 Å resolution. The structure shows the characteristic P450 fold as seen in other members of the family, with the lengths and orientations of the individual secondary structural elements being very similar to those seen in 2C9. There are, however, several important differences, the most notable involving the F helix, the F-G loop, the B′ helix, β sheet 4, and part of β sheet 1, all of which are situated on the distal face of the protein. The 2D6 structure has a well defined active site cavity above the heme group, containing many important residues that have been implicated in substrate recognition and binding, including Asp-301, Glu-216, Phe-483, and Phe-120. The crystal structure helps to explain how Asp-301, Glu-216, and Phe-483 can act as substrate binding residues and suggests that the role of Phe-120 is to control the orientation of the aromatic ring found in most substrates with respect to the heme. The structure has been compared with published homology models and has been used to explain much of the reported site-directed mutagenesis data and help understand the metabolism of several compounds.

The cytochromes P450[4] constitute a superfamily of heme-containing enzymes that catalyze the metabolism of a wide variety of endogenous and xenobiotic compounds. This is accomplished through the activation of molecular oxygen by the heme group, a process that involves the delivery of two electrons to the P450 system followed by cleavage of the dioxygen bond, yielding water and an activated iron-oxygen species (Compound 1), which reacts with substrates through a variety of mechanisms (1). In eukaryotic species, the electron source is a single flavoprotein, the FAD/FMN-containing cytochrome P450 reductase, which binds to the largely basic proximal face of the cytochrome through a number of salt bridges. Of the known human isoforms, cytochrome P450 2D6 is responsible for the metabolism of at least 20% of known drugs (2), with only 3A4 being responsible for a higher (50%) percentage.

The cDNA encoding human P450 2D6 has been characterized (3) and subsequently localized to chromosome 22 in the q13.1 region (4). A relatively large number of genetic polymorphisms have been described for 2D6, some of which can either result in rapid or very poor metabolism. One well characterized allelic variant is responsible for a condition known as debrisoquine/sparteine type polymorphism (5, 6). This arises as a result of various genetic mutations and affects a significant percentage of the Caucasian population (7). It results in the defective metabolism of a number of important drug molecules, including debrisoquine, from which the condition got its name. The inability of patients to turn over compounds such as debrisoquine eventually leads to toxic levels of the drug in the body. Binding of any drug to these allelic 2D6 variants can cause drug-drug interactions, which can lead to severe side effects and has resulted in early termination of several candidate drugs in development, refusal of regulatory approval, severe prescribing restrictions, and withdrawal of drugs from the market.

Whereas 3A4 exhibits a wide diversity in its substrate recognition, a fact that is often attributed to its large cavity size (8), 2D6 generally only recognizes substrates containing a (protonated) basic nitrogen and a planar aromatic ring. These features are found especially in a large number of central nervous system and cardiovascular drugs that act on the G protein-coupled receptor superfamily of proteins. For this reason, 2D6 is the most widely studied isoform, both from experimental site-directed mutagenesis (SDM) studies and from computational modeling. With regard to the latter, numerous pharmacophore models have been described (9–11), but it has become clear that no single model can account for the diversity observed in the regioselectivity of substrate metabolism. Likewise, various homology models have been constructed (10–24), but the sequence identity between 2D6 and the available x-ray crystal structures used is about 40% at best for the 2C isoforms and only 18% for 3A4. Thus, the availability of a crystal structure of 2D6 was anticipated to go a long way toward explaining the effects of polymorphism and the results of SDM studies and toward answering some of the questions raised by *in silico* modeling work. Such improved information would in time help achieve the ultimate goals of predicting the metabolic fate of drug compounds or predicting which compounds would inhibit the cytochromes and eventually lead to improved therapeutic ligand design.

Recently, the structures of a number of mammalian cytochromes have been solved by x-ray crystallography. The first of these, rabbit 2C5 (25), showed considerably higher homology to the human isoforms than had been found with previous bacterial enzyme structures and led quickly to a new generation of computational models (19–24). This has been followed by the reports of the crystal structures of the human 2C9 (26, 27), 2C8 (28), 3A4 (29, 30), and 2A6 (31) isoforms. This was made possible by truncation of the membrane-bound N-terminal domain and, in some cases, the introduction of some small mutations that helped to solubilize the protein. In recent years, we have embarked on a similar approach, and in this paper we report the x-ray crystal structure

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 2F9Q) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. 1–4.

[1] These authors contributed equally to this work.

[2] To whom correspondence should be addressed. Tel.: 44-1279-622997; Fax: 44-1279-627666; E-mail: Paul_2_Rowland@gsk.com.

[3] Present address: Pharmix Corp., 2000 Sierra Point Pkwy., Suite 500, Brisbane, San Francisco, CA 94005.

[4] The abbreviations used are: P450, a generic term for a cytochrome P450, with individual P450 enzymes denoted by a number/letter/number combination derived from the sequence identity; SRS, substrate recognition site; SDM, site-directed mutagenesis.

of the human cytochrome P450 2D6 at 3.0 Å resolution. The solubilizing mutations were designed using an early model that was based on homology with cytochrome P450 BM3 (32). This new structure has already proved to be valuable in understanding the metabolism of several compounds and the effects of many SDM studies.

MATERIALS AND METHODS

Early Molecular Modeling of 2D6—The initial sequence alignment was carried out using PSI-BLAST (33) against a nonredundant protein data base comprising over 400,000 sequences. A resultant 1200-sequence multiple sequence alignment was refined using the HMMER sequence profile alignment tools (34), and the 2D6 model was built using the homology modeling tools in ICM (35).

Generation of 2D6 Truncates—The 2D6 truncates were generated by PCR, introducing an XbaI site at the C terminus and an NdeI site at the N terminus plus various amino acid alterations. The N-terminal primers were designed to remove the extreme N terminus, containing the putative membrane spanning region, and polyhistidine tags were added at the N- or C terminus as described. After digestion with NdeI and XbaI, the PCR products were subcloned into the pCW expression vector (36). The sequence of all constructs was confirmed by automated dideoxy-DNA sequencing.

Generation of 2D6 Mutants—Mutations at residues Leu-230 and Leu-231 were introduced using the QuikChange mutagenesis kit (Stratagene, La Jolla, CA) according to the manufacturer's protocol. Multiple mutations were introduced in single PCRs using semirandom primers, with multiple bases inserted at exact positions during primer synthesis (*e.g.* the primer CTGAATGCTCTCCCCGTCMRNCTGCATATCCCAGCGCTGGCTG was used to introduce Asn, Lys, His, Gln, Arg, and Ser at residue 230).

Expression of 2D6 Truncates and Mutants to Test Solubility—Initial expression trials of all 2D6 constructs were performed using the host DH10B (Invitrogen). Cultures (5-ml scale) were grown overnight at 37 °C in LB broth supplemented with ampicillin. These cultures were used as a 1% inoculum in modified Terrific broth (100 ml), supplemented with ampicillin, 1 mM thiamine, 0.5 mM δ-aminolevulinic acid, and trace element solution (37). When the cultures reached an A_{600} of 0.6 at 30 °C, expression was induced by the addition of isopropyl 1-thio-β-D-galactopyranoside to a final concentration of 1 mM. The cells were harvested after a further 16 h at 30 °C by centrifugation at 5000 × *g* for 10 min at 4 °C.

Subcellular Fractionation—*Escherichia coli* pellets were resuspended in TSE buffer (200 mM Tris acetate, pH 8.0, 500 mM sucrose, 1 mM phenylmethylsulfonyl fluoride, and 1 mg/ml lysozyme) and incubated on ice for 1 h. Cell pellets were recovered by centrifugation at 5000 × *g* for 10 min at 4 °C and resuspended in lysis buffer (0.5 M KP_i, pH 7.6, 20% glycerol, 1 mM phenylmethylsulfonyl fluoride, 1 mM dithiothreitol). Cells were lysed using sonication (four 30-s pulses with 5-min intervals at 50% of maximum power), and cell debris was pelleted by centrifugation at 5000 × *g* for 10 min at 4 °C. The supernatant was further clarified by centrifugation at 100,000 × *g* for 1 h at 4 °C in a Beckman TLN-100 rotor. The resultant supernatant was used as the cytosolic fraction (containing "soluble" P450), and the pellet contained the membrane fraction.

Expression of 2D6 at Large Scale—A shake flask containing 100 ml of LB was inoculated with a single colony taken from a fresh transformant previously plated out onto LB agar. The flask was incubated at 30 °C with constant shaking at 120 rpm for 14 h before transfer as a 2% inoculum to the final stage. The final stage consisted of 2.5 liters of MTB and 1% glycerol in a 3.6-liter laboratory scale fermenter (InforsAG, Switzerland). All media were supplemented with 100 μg/ml ampicillin. After inoculation, the culture was grown at 30 °C with an airflow rate of 1.5 liters air/min and an agitation speed of 680 rpm until induction. Induction was at A_{600} 0.8 with 0.5 mM isopropyl 1-thio-β-D-galactopyranoside and 0.5 mM δ-aminolevulinic acid. Following further incubation at the same conditions for 24 h, the cells were harvested by centrifugation. P450 concentration was determined by CO difference spectrum using whole cells.

Cell Lysis—Frozen pellets were resuspended in lysis buffer (100 mM Tris, pH 7.4, 500 mM sucrose, 0.5 mM EDTA, 0.1 mg/ml lysozyme, 1 μl/ml benzonase, and protease inhibitors) at 4 ml/g cell pellet. All manipulations were carried out at 4 °C. The lysate was stirred for 30 min and centrifuged (5000 × *g*, 20 min), and the pellet was resuspended in 200 ml of Buffer A (400 mM KP_i, pH 7.4, 20% glycerol, and 10 mM β-mercaptoethanol). The suspension was Dounce-homogenized on ice and left to stir for 30 min before being processed twice through a high pressure (10,000 p.s.i.) disruption system (Constant Systems Ltd., Northants, UK). The lysate was centrifuged (100,000 × g, 1 h), and the supernatant was retained.

Protein Purification—The protein was purified using an adaptation of a previously published method (38). All manipulations and purifications were carried out at 4 °C. Purification followed a three-step process: Ni^{2+}-nitrilotriacetic acid (Qiagen, Crawley, UK), hydroxyapatite II (Bio-Rad), and size exclusion (Superose 12, GE Healthcare, Chalfont, UK). Experiments showed that the protein was 95% pure after the first two stages. All purifications were carried out using an ÄKTA Purifier 100 instrument (GE Healthcare). The cleared lysate was applied to the nitrilotriacetic acid column in buffer A, washed (5 column volumes of wash buffer 400 mM KP_i, pH 7.4, 20% glycerol, 10 mM β-mercaptoethanol, 50 mM glycine, and 100 mM NaCl) and eluted using a linear gradient of 0–100% Buffer B (400 mM KP_i, pH 7.4, 20% glycerol, 10 mM β-mercaptoethanol, and 100 mM EDTA) over 10 column volumes. The P450 sample (post-Ni^{2+}-nitrilotriacetic acid) was dialyzed overnight in Buffer C (50 mM KP_i, pH 7.4, 20% glycerol, and 10 mM β-mercaptoethanol) and loaded onto the hydroxyapatite column. The column was washed (5 column volumes of buffer C), and the protein was eluted using a linear gradient of 0–100% Buffer D (400 mM KP_i, pH 7.4, 20% glycerol, and 10 mM β-mercaptoethanol) over 10 column volumes. The P450-containing fractions (posthydroxyapatite) were pooled and concentrated using a 15 ml Amicon centrifugal filter device (Millipore Corp., Watford, UK) to ~1.5 ml. The concentrated sample was applied to an S12 column equilibrated with Buffer E (50 mM KP_i, pH 7.4, 20% glycerol, 100 mM NaCl, and 10 mM β-mercaptoethanol). The resulting fractions were analyzed by SDS-PAGE (Cambrex, Nottingham, UK), peptide mass fingerprinting (Micromass, Manchester, UK), protein assay (Coomassie Plus; Pierce), and P450 content (CO difference assay).

Carbon Monoxide Difference Assay—Concentrations of P450 were estimated spectrophotometrically using a Cary Bio 100 instrument (Varian, Crawley, UK) from the difference spectra determined for the formation of the carbon monoxide complex with the protein after reduction with sodium dithionite (39). The specific activity of the 2D6 L230D/L231R protein was 11 nmol/mg.

Activity Assays—Activity assays were performed in a 384-well microplate using an LJL Biosystems Analyst HT (Sunnyvale, CA). Excitation and emission wavelengths were set at 405 and 530 nm (with a 425-nm diachronic mirror). The assay was performed using adaptations of previously published methods (40, 41). 0–128 μM 4-methylaminomethyl-7-methoxycoumarin (GlaxoSmithKline) was added to 10 μmol of 2D6 in the presence of 10 mmol of cumene hydroxyperoxide (as an electron donor), 1 mM dithiothreitol, 40 mM KP_i, pH 7.5, and 20% glycerol. Readings were taken every 60 s for 10 min. The 2D6 L230D/L231R metabo-

TABLE 1
Crystallographic statistics

Parameter	Value
Data collection statistics[a]	
Resolution (Å)	40–3.0 (3.11–3.00)
Observations	94,111
Unique reflections	39,448
Completeness (%)	90.0 (90.7)
Average $I/\sigma I$	11.2 (1.6)
R_{merge}[b]	0.107 (0.614)
Refinement statistics	
Resolution (Å)	40–3.0
R factor[c]	0.230
R_{free} (4% of data)	0.286
Root mean square deviations from ideality	
Bond lengths (Å)	0.016
Bond angles (degrees)	1.65
Average B factors (Å^2)[d]	
Protein main chain	56.9 (7240)
Protein side chain	57.5 (6996)
Hemes	52.9 (172)
Waters	39.0 (11)
Sulfates	53.9 (10)

[a] Data for the highest resolution shell are given in parentheses.
[b] $R_{merge} = \Sigma|I_j - \langle I_j\rangle|/\Sigma\langle I_j\rangle$.
[c] $R = \Sigma||F_o| - |F_c||/\Sigma|F_o|$.
[d] The number of model atoms is given in parentheses.

lizes 4-methylaminomethyl-7-methoxycoumarin with a K_m of 67.6 μM and a V_{max} of 2.85 ± 0.25 μM/min.

Crystallization and Data Collection—Crystals of the 2D6 L230D/L231R mutant construct were grown at room temperature (~20 °C) by free interface diffusion using Topaz XRAY chips (Fluidigm). A solution of the protein at 60 mg/ml in a buffer of 50 mM KP_i, pH 7.4, 100 mM NaCl, 20% glycerol, and 5 mM β-mercaptoethanol was loaded into the chip along with a range of solutions based on simple dilutions with water of a solution of 2.0 M ammonium sulfate, 0.1 M sodium citrate, pH 5.6, and 0.2 M potassium sodium tartrate, with the optimum dilution centering around an ammonium sulfate concentration of 1.48–1.52 M. Following protein and reagent loading, the chip interface line was opened and allowed to remain open for the duration of the crystallization process. All chips utilized water as the hydration fluid to maintain the environment of the chip at ~100% relative humidity, and the chips were all prehydrated for at least 24 h prior to the experiment. Crystals typically appeared after a few days and continued to grow for a further 5–10 days, usually growing out of a globular gel-like aggregate, which generally formed after about 24 h. The crystals are rectangular plates with dimensions that rarely exceed 80 × 20 × 10 μm. The crystals were harvested from the chips and stored in a solution of 1.8 M ammonium sulfate, 0.1 M sodium citrate, pH 5.6, 0.2 M potassium sodium tartrate, and 20% glycerol. The crystals were mounted directly from the harvesting solution and flash-frozen in liquid nitrogen. X-ray diffraction data were collected with a MarMosaic 225 CCD detector (Mar Research) at 100 K at a wavelength of 0.9538 Å using beam line ID23-1 at the European Synchrotron Radiation Facility. Due to the small size of the crystals, it was necessary to use exposure times on the order of 10–20 s per 0.5° oscillation, which resulted in considerable radiation damage during the data collection. The final native data set was assembled from 114 images collected from three crystals with refined mosaicities of 0.44, 0.51, and 0.66°. The data were processed and scaled using the HKL2000 suite of programs (42). Structure factors were derived from the reflection intensities using the CCP4 suite of programs (43). The crystals belong to space group $P2_12_12$ with unit cell dimensions a = 145.1 Å, b = 155.5 Å, c = 95.8 Å. Table 1 gives a summary of the data collection statistics.

Structure Determination—The structure was solved by Molecular Replacement using PHASER (44) with a 2C9 crystal structure (Protein Data Bank code 1OG2) as a search model. The search model included all protein atoms for the 2C9 monomer from Pro-30 to Val-490 and excluded the heme group and water molecules. The sequence identity between the 2C9 and 2D6 sequences was 40.7% in a protein sequence alignment covering this region, matching 187 of 459 residues. A convincing molecular replacement solution comprising four molecules in the asymmetric unit was found showing 222 symmetry (consistent with peaks observed in the self-rotation function). The resulting set of phases was used to calculate $2F_o - F_c$ and $F_o - F_c$ (where F_o represents the observed structure factor and F_c is the calculated structure factor) electron density maps. The presence of large positive peaks in positions corresponding to the heme iron atoms confirmed that the molecular replacement solution was correct.

Model Building and Refinement—The crystal structure was built using multiple cycles of model building with the molecular graphics program COOT (45), followed by structure refinement with REFMAC (46). Due to the limited resolution of the diffraction data, very tight noncrystallographic symmetry restraints were imposed throughout the refinement. In the last cycle, these restraints were relaxed for a small number of protein residues, which appeared to exhibit some differences in conformation between the individual molecules. The last cycle of refinement also incorporated TLS parameters (47) to model anisotropic displacements, specifically utilizing one TLS group per molecule. This resulted in significantly better R and R_{free} values (0.230 and 0.286, respectively) compared with an equivalent non-TLS refinement cycle, where R and R_{free} values were 0.253 and 0.311, respectively. The final refinement statistics are given in Table 1. The final model contains 14,429 atoms and comprises four 2D6 molecules, two sulfate ions, closely involved in crystal contacts, and 11 water molecules. The protein model includes the residues 52–497 (full-length protein sequence numbering) and additionally a short stretch of the proline-rich N-terminal region (residues 34–41). Residues 229–239 of the F-G loop have been built only as alanines, since the electron density maps were insufficiently clear to unambiguously assign the correct residue side chains. In a Ramachandran plot, 84.7% of residues are in the most favored regions as defined by PROCHECK (48), with 12.9 and 1.6% of residues in the additionally allowed and generously allowed regions, respectively, and 0.8% in disallowed regions. The four protein molecules are related by 222 symmetry and are all essentially the same with only minor differences, the root mean square deviations between Cα atoms for all possible molecule pairs being as follows: molecules A/B, 0.10 Å; A/C, 0.14 Å; A/D, 0.14 Å; B/C, 0.12 Å; B/D, 0.13 Å; and C/D, 0.14 Å. The protein figures were drawn using PYMOL (49).

Recent Modeling of 2D6—All molecular dynamics simulations mentioned here were performed with the CHARMm program (50) on a Silicon Graphics 48xR12000 processor Origin server. Visualizations were carried out with a Silicon Graphics Octane work station using the QUANTA program (51). The debrisoquine ligand was built, and *ab initio* charges (3-21G* natural atomic orbital) were calculated using the SPARTAN program (52). Substrate dockings were carried out manually by placing the compound in a number of plausible starting poses and then minimizing them in the protein using CHARMm, with 500-step Steepest Descent followed by 5000-step Adopted Basis Newton Raphson. A distance constraint was used to keep the iron-oxygen atom within reacting distance to the site of metabolism. For the heme group, optimization of a "picket fence" porphyrin, containing an iron-bound oxygen atom on the distal side and a thiomethyl group on the proximal side, was carried out at the unrestricted Hartree-Fock level, using a 6-31G* basis set. The charges used were natural atomic orbital charges. The iron-cysteine bond was formed in CHARMm using a patch RTF file

written according to the standard CHARMm protocol. The heme model was least-squares-fitted to the crystal structure but otherwise was kept rigid during the simulation, thus alleviating the need for special parameters for the octahedral iron complex, which is not readily handled by conventional force fields.

RESULTS AND DISCUSSION

Expression of 2D6 Truncates—P450 2D6 expresses at high levels in *E. coli* if the extreme N terminus containing the putative membrane-spanning region is replaced with different signal sequences, the OmpA signal sequence (53), or the sequence used to express P450 17α in *E. coli* (54). However, 2D6 protein generated with these N-terminal sequences associates with *E. coli* membranes and is not amenable to crystallization. Generation of other soluble P450 isozymes has been successfully performed by removal of this hydrophobic N-terminal sequence (*e.g.* 2C3 (38), 2C5 (25), 2C8 (28), 2C9 (26, 27), and 3A4 (29, 30)). 2C3, unlike many other P450s, contains several hydrophilic residues between the membrane-spanning region and the PPGP motif, which is required for heme incorporation (55), and a simple truncation was sufficient to obtain almost 100% cytoplasmic expression of 2C3 (38) (2C3d; see Fig. 1). Soluble expression of 2C5 required the introduction of residues 2–6 of 2C3d to obtain similar levels of soluble protein (25). A simple truncation approach attempted with 2D6 was unsuccessful (2D6 truncate 1; see Fig. 1). Since the sequence preceding the PPGP motif is quite hydrophobic in 2D6, additional truncates were generated (truncates 2–7; see Fig. 1), which contained fusions between the N termini of 2C3d and a variety of different residues in the N terminus of 2D6. Histidine tags were used both N- and C-terminally. The most successful expression of 2D6 was obtained using truncate 5, but this truncation was not sufficient to obtain 100% soluble P450, so it was decided that mutation of hydrophobic surface residues on 2D6 would be required to obtain increased solubility, using truncate 5 as a template.

Generation of 2D6 Mutants with Improved Solubility—A model of 2D6 was derived from a multiple-sequence alignment and based on the crystal structure of bacterial cytochrome P450 BM3. The sequence identity of 2D6 to BM3 is low, and although some regions of the 2D6 model can be considered reliable, the quality of the 2D6 model around the putative substrate access channel is less good. However, careful analysis of the model revealed a patch of hydrophobic residues in the loop region between the F and G helices (Fig. 2), which was proposed to be situated on the surface of the protein and could possibly contribute to protein aggregation and membrane association. Two of these residues, Leu-230 and Leu-231, were selected for mutagenesis to hydrophilic residues using primers randomized at specific nucleotides. Fifty-two mutants of 2D6 were generated (Fig. 3*A*) and tested for expression of holo-P450. The solubility of those mutants expressing holo-P450 was then tested by partial purification in high salt buffers (Fig. 3*B*). Six of the mutants showing the highest solubility and expression of holo-P450 were selected for larger scale growth, further purification, and crystallization trials (L230T/L231K, L230D/L231R, L230A/L231S, L230N/L231D, L230T/L231D, and L230N/L231R). Of these, 2D6 L230D/L231R appeared to give the most reliable yields and the best crystals and was used in all subsequent studies.

Crystal Structure of 2D6—The 3.0 Å crystal structure of 2D6 shows the characteristic P450 fold as seen in other members of the family (Figs. 4 and 5). The lengths and orientations of the individual secondary structural elements in 2D6 are very similar to those seen in 2C9 (supplemental Fig. 1). A structural alignment of 2D6 with 2C9 using the program LSQMAN (56) gave a root mean square distance of 1.16 Å for 389 aligned Cα atoms (an equivalent alignment with 3A4 gave a root mean square distance of 1.82 Å for 334 atoms). Despite the similarities between 2D6 and 2C9, there are six main areas where significant differences can be found. Two of these areas are located on the proximal face of the protein. In 2D6, there is an extra turn at the end of helix C, resulting in a shorter loop between it and helix D (the differences spanning residues 139–148). Although the total number of residues is the same as in 2C9, this short loop substantially reduces the interactions between the C-D connection region and the G-H loop in 2D6 (as evidenced by the observed differences between the G-H loop conformations in the two structures). A second difference is observed in the area of β sheet 2 (residues 380–392), where there is a considerable shift in position of the two strands relative to sheet 1. In 2D6, these strands bend up toward the underside of sheet 1 much more than in the case of 2C9, this closer packing between sheets being facilitated by 2D6 having more small hydrophobic side chains than 2C9 in this interface.

The other four areas showing large differences between 2D6 and 2C9 are situated on the distal face of the protein. Three of them are directly involved in defining the shape and character of the protein's active site. The most obvious difference is the position of the F helix and the F-G loop. Although there are substantial differences between the F-G loops in the literature 2C9

```
M--DLLIILGICLSCVVLLS--LWKKTHGKGKLPPGPTPL   2C3 FL
M--------------------------KKTSSKGKLPPGPTPL   2C3d
M--DPVVVLVLGLCCLLLLS--IWKQNSGRGKLPPGPTPF   2C5 FL
M--DSLVVLVLCLSCLLLLS--LWRQSSGRGKLPPGPTPL   2C9 FL
MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPL   2D6 FL
------------MAHHHHHHSMHRRQRWAARYPPGPLPL   2D6 Trunc1
------------MAHHHHHHSKKTSSKGKLPPGPLPL   2D6 Trunc2
------------MAHHHHHHSRRQRWAARYPPGPLPL   2D6 Trunc3
------------MAHHHHHHSRQVHSSWNLPPGPLPL   2D6 Trunc4
----------------------MAKKTSSKGKLPPGPLPL   2D6 Trunc5
----------------------MARRQRWAARYPPGPLPL   2D6 Trunc6
----------------------MAKKTSSRYPPGPLPL   2D6 Trunc7
```

FIGURE 1. **N-terminal sequence alignment of the 2D6 truncates with other P450 isozymes.** The full-length (*FL*) N-terminal sequences of 2C3, 2C5, 2C9, and 2D6 are shown with the soluble 2C3d sequence and the 2D6 truncates. *Trunc1*, 2D6 wild-type sequence, truncated at residue 23 to remove the membrane-spanning region, with N-terminal His_6 tag. *Trunc2*, 2D6 truncated at residue 34, with His_6 tag and residues 2–10 of 2C3d inserted at the N terminus. *Trunc3*, 2D6 truncated at residue 25, with N-terminal His_6 tag. *Trunc4*, 2D6 truncated at residue 34, with His_6 tag and residues 24–32 of 2E1 inserted at the N terminus. *Trunc5*, 2D6 truncated at residue 34, with C-terminal His_4 tag and residues 2–10 of 2C3d inserted at the N terminus. *Trunc6*, 2D6 truncated at residue 25, with C-terminal His_4 tag. *Trunc7*, 2D6 truncated at residue 32, with C-terminal His_4 tag and residues 2–6 of 2C3d inserted at the N terminus.

```
      --- F helix----|-loop-|---- helix? ----|----loop------|- G helix--
2D6   Q E G L K E E S G F L R E V L N A V P V L L H I P A L A G K V L R
2C5   N E N V R I L S S P W L Q V Y N N F P A L L D Y F P G I H K T L L
BM3   L D E A M N K L Q R A N . . . . . P D D P . A Y D E N . K R Q F
      ---- F helix-----|---------- loop----------|----- G helix------
```

FIGURE 2. **Predicted secondary structural elements for human 2D6 in the region of the F and G helices.** The crystal structure of P450 BM3 was used as the basis of our early 2D6 modeling. The secondary structure of BM3 is shown *below* the alignment, with the predicted 2D6 structural elements shown *above*. Also included is the rabbit 2C5 sequence, (all three sequences are wild-type). The 2D6 residues Leu-230 and Leu-231 are *highlighted*.

A

L230	L231									
	Thr	Lys	Asn	Glu	Arg	Ser	His	Ala	Asp	Gln
Thr		X	X	X	X	X	X		X	
Lys						X				
Asn		X	X		X	X		X	X	
Glu	X								X	
Arg	X			X	X	X	X	X	X	X
Ser	X	X		X			X	X	X	
His	X	X			X					
Ala	X	X	X	X	X	X	X	X	X	X
Asp			X	X	X	X		X		
Gln		X	X				X		X	

B

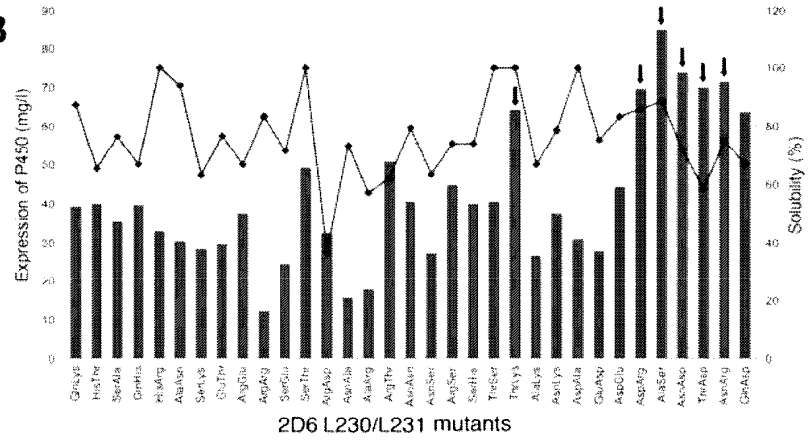

FIGURE 3. **Expression and solubility studies of the 2D6 mutants.** *A*, the 52 L230/L231 double mutants. The mutants are marked (×). *B*, expression levels and solubility. The expression levels are represented by the *bars* (*scale* on the *left axis*), and the solubility is shown by the *line plot* (*scale* on the *right*). The six mutants selected for large scale study are *highlighted* with *arrows*.

structures (Protein Data Bank codes 1OG2 and 1OG5 *versus* 1R9O), the situation with 2D6 is clearly different from 2C9. The F helix in 2D6 has two additional turns and arcs down much more closely over the heme pocket toward the N-terminal end of strand 2 of β sheet 1. This difference in the length of the F helix correlates strongly with an observed shift in the position of strands 1 and 2 of β sheet 1 (from residue 71 to 78). These adopt a very different conformation from that seen in 2C9. At the end of the F helix, the F-G loop lies across the side of the B′ helix, thereby enclosing the side that is completely open in 2C9. The F-G loop then rejoins the G helix, which adopts approximately the same orientation as that of 2C9, except that there is a significant shift along the helical axis, such that the turns do not align with each other. The quality of the electron density maps for the F-G loop region were unfortunately not high enough to give a completely satisfactory model for this important part of the structure, and for this reason, only a polyalanine trace was built for part of this loop. There is no sign of an F′ helix, although the backbone of a short G′ helix does seem to be present. The two remaining differences between the 2D6 and 2C9 structures are also related to the F helix shift. The B′ helix in 2D6 is pushed out away from the heme pocket, and there are an additional three residues in the loop immediately following it (residues 101–118). Similarly, on the opposite side of the F helix from the B′ helix, β sheet 4 (residues 468 – 487) adopts a shift in conformation in the same direction as the F helix shift.

Active Site Cavity—The 2D6 structure has a well defined cavity above the heme (Fig. 6), which could be described as having the shape of a "right foot." The volume is about 540 Å^3, as determined by VOIDOO

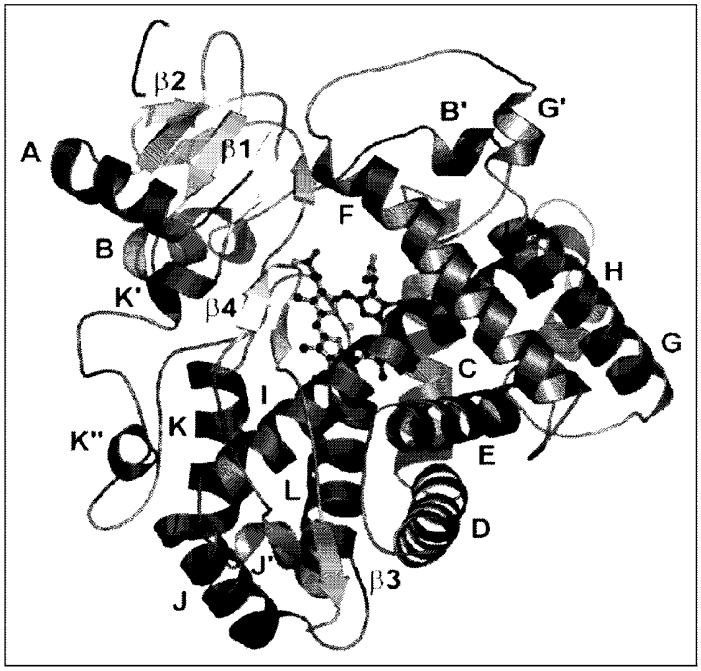

FIGURE 4. ***Ribbon diagram*** **of the 2D6 structure.**

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.