

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. Blaney¹, Martin G. Smyth, Jo J. Jones, Vaughan R. Leydon, Amanda K. Oxbrow, Ceri J. Lewis, Mike G. Tennant³, Sandeep Modi, Drake S. Eggleston, Richard J. Chenery, and Angela M. Bridges¹

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 P450⁴ 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 CTGAATGCTCTCCCCGTCMRNCTGCATATCC-CAGCGCTGGCTG 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 $5000 \times g$ 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 $5000 \times g$ 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 $5000 \times g$ for 10 min at 4 °C. The supernatant was further clarified by centrifugation at $100,000 \times g$ 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 μ g/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 μ l/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 ($5000 \times g$, 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 ($100,000 \times g$, 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: Ni²⁺-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-Ni²⁺-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 L/JL 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 μ M 4-methylaminomethyl-7-methoxycoumarin (GlaxoSmithKline) was added to 10 μ mol of 2D6 in the presence of 10 mmol of cumene hydroperoxide (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 $R_{\text{merge}} = \sum |I_j - \langle I_j \rangle| / \sum \langle I_j \rangle$.^c $R = \sum \|F_o\| - \|F_c\| / \sum \|F_o\|$.^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 $2.85 \pm 0.25 \mu\text{M}/\text{min}$.

Crystallization and Data Collection—Crystals of the 2D6 L230D/L231R mutant construct were grown at room temperature ($\sim 20^\circ\text{C}$) by free interface diffusion using Topaz XRAY chips (Fluidigm). A solution of the protein at 60 mg/ml in a buffer of 50 mM $\text{K}_2\text{P}_2\text{O}_7$, 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 $\sim 100\%$ 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 $80 \times 20 \times 10 \mu\text{m}$. 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 $P2_12_12_1$ with unit cell dimensions $a = 145.1 \text{ Å}$, $b = 155.5 \text{ Å}$, $c = 95.8 \text{ Å}$. 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 $\text{C}\alpha$ 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 17 α 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 suffi-

cient 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. 3A) 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. 3B). 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 C α 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

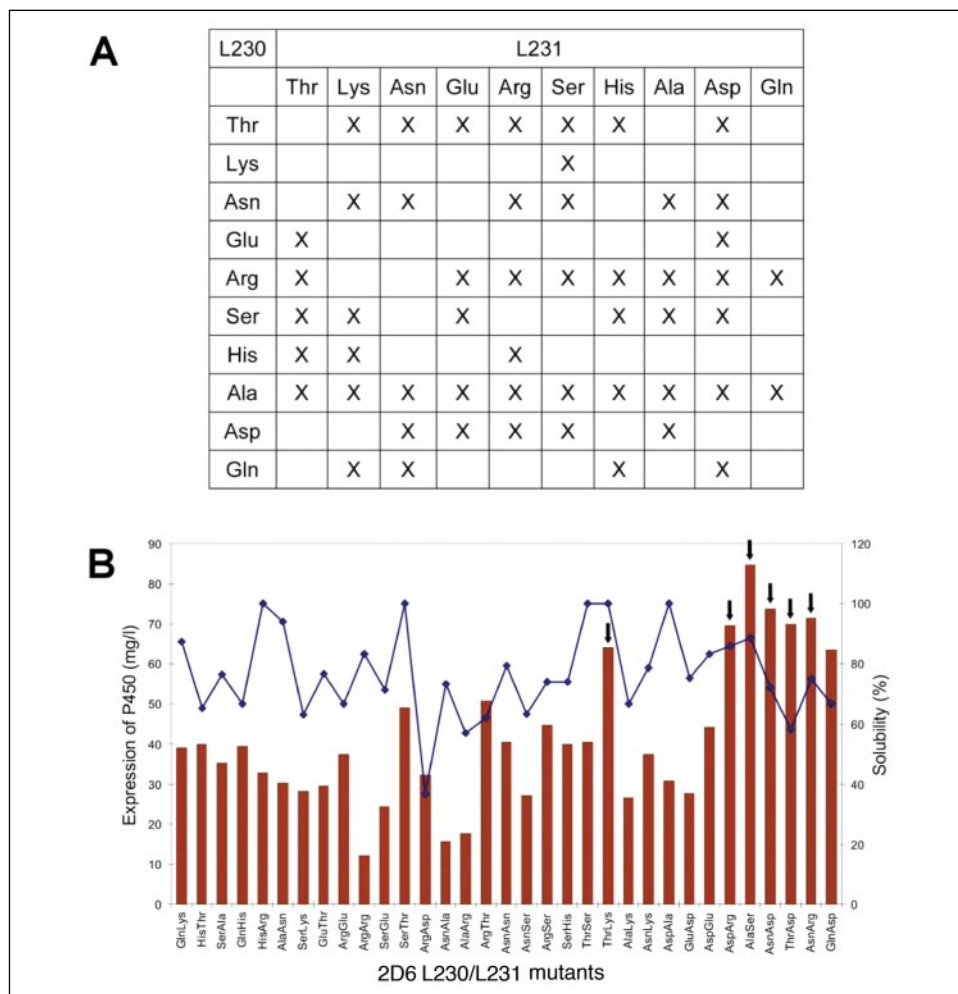
M--DLLIILGICLSCVLLS--LWKTHGKGLPPGPTPL	2C3 FL
M-----KKTSSKGLPPGPTPL	2C3d
M--DPVVVLVLGLCLLLLS--IWKQNSGRGLPPGPTPF	2C5 FL
M--DSLVLVLCLLCLLLLS--LWRQSSGRGLPPGPTPL	2C9 FL
MGLEALVPLAVIVAIFLLVLDLHRRQRWAARYPPGPLPL	2D6 FL
-----MAHHHHHSMHRRQRWAARYPPGPLPL	2D6 Trunc1
-----MAHHHHHSSKKTSSKGLPPGPLPL	2D6 Trunc2
-----MAHHHHHSRRQRWAARYPPGPLPL	2D6 Trunc3
-----MAHHHHHSRQVHSSWNLPPGPLPL	2D6 Trunc4
-----MAKKTSSKGLPPGPLPL	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 His₆ tag. *Trunc2*, 2D6 truncated at residue 34, with His₆ tag and residues 2–10 of 2C3d inserted at the N terminus. *Trunc3*, 2D6 truncated at residue 25, with N-terminal His₆ tag. *Trunc4*, 2D6 truncated at residue 34, with His₆ tag and residues 24–32 of 2E1 inserted at the N terminus. *Trunc5*, 2D6 truncated at residue 34, with C-terminal His₄ tag and residues 2–10 of 2C3d inserted at the N terminus. *Trunc6*, 2D6 truncated at residue 25, with C-terminal His₄ tag. *Trunc7*, 2D6 truncated at residue 32, with C-terminal His₄ 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.

FIGURE 3. Expression and solubility studies of the 2D6 mutants. A, the 52 L230/L231 double mutants. The mutants are marked (X). 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 Å³, as determined by VOIDOO

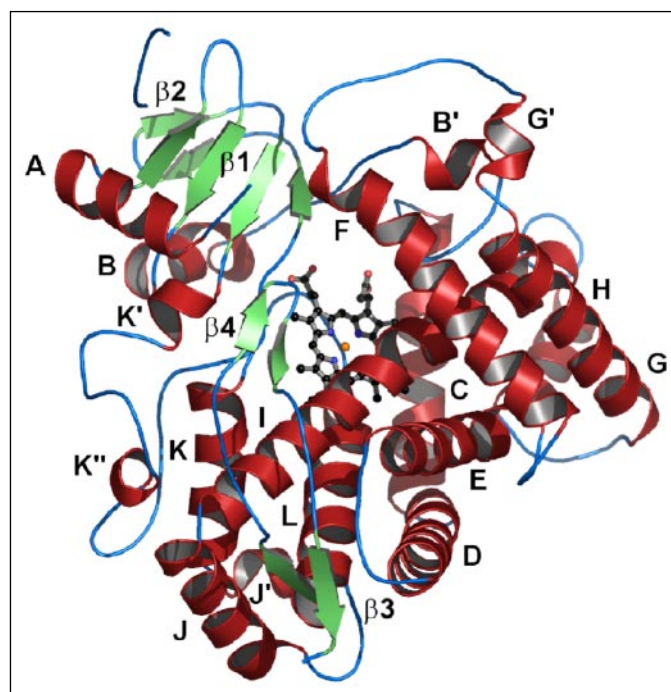


FIGURE 4. Ribbon diagram of the 2D6 structure.

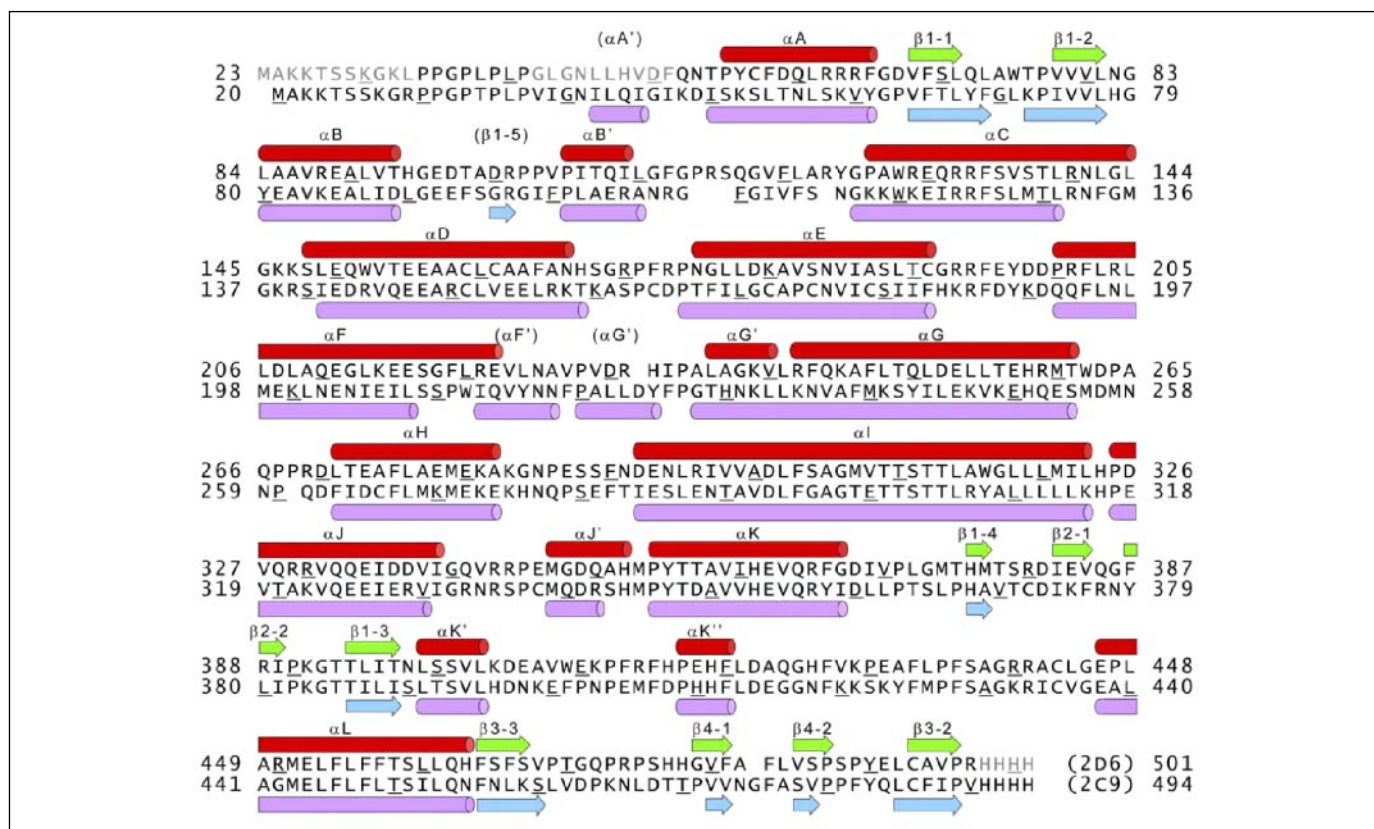


FIGURE 5. **Secondary structure assignments of 2D6 compared with 2C9.** The secondary structural elements are defined using assignments from PROCHECK with some manual adjustment based on hydrogen bonding patterns and are labeled according to established convention. The sequence alignment shows 2D6 above 2C9, with 2D6 helices and strands (red and green) shown above the alignment and 2C9 helices and strands (purple and cyan) shown below the alignment. The 2D6 sequence shown is that of the construct used to grow the crystals (Truncate 5 L230D/L231R) and uses the residue numbering of the wild-type protein. Regions of the sequence that are not included in the 2D6 model are shown in gray. Residues 229–239 (which includes the double mutation site) are built only as alanines in the model. The 2C9 sequence shown is that of the construct used in the 1OG2 structure and uses standard residue numbering. Residues that are multiples of 10 are underlined in both sequences. Secondary structural elements defined in 2C9 that are not defined in 2D6 are labeled in parentheses.

(57). It is bordered by the heme group and residues from the B' helix (side chain of Ile-106), the B'-C loop (side chains of Leu-110, Phe-112, Phe-120, and Leu-121 and main chain atoms of Gln-117, Gly-118, Val-119, and Ala-122), the F helix (side chains of Leu-213, Glu-216, Ser-217, and Leu-220), the G helix (side chains of Gln-244, Phe-247, and Leu-248), the I helix (side chains of Ile-297, Ala-300, Asp-301, Ser-304, Ala-305, Val-308, and Thr-309), the loop between helix K and β sheet 1 strand 4 (side chains of Val-370 and Met-374 and main chain atoms of Gly-373), and residues from the loop between the strands of β sheet 4 (side chains of Phe-483 and Leu-484) (Fig. 7).

The "heel" of the foot-shaped cavity lies above the heme, offset toward the propionate side. The foot "arch" is formed by the side chain of Phe-120. The "ball" of the foot is bordered by residues from the B'-C loop and the N-terminal end of the I helix. Additional residues in the I helix line the whole length of the right side of the foot. The "toe" area is bordered by residues from the B' and G helices. The upper part of the foot is bordered by residues in the F helix, which is perpendicular to the foot axis. The back of the heel is shaped by residues in the loop following the K helix. The "ankle" region marks a narrowing of the cavity and leads up to the outside of the protein and the cavity entrance. It is bordered by residues of the F helix at the front and residues of the I helix on the right, with the back of the ankle being defined entirely by residues from the loop between the strands of β sheet 4. The back, left side, and toe areas of the cavity are strongly hydrophobic in character. The upper part and right side of the foot has several important hydrophilic side chains (Glu-216 in helix F, Gln-244 in helix G, and Ser-304 in helix I). Under the ball of the foot lies Asp-301 (helix I). Above the ankle region, the cavity

entrance is bordered by a number of long charged/hydrophilic side chains from the F helix (Gln-210, Glu-211, Lys-214, and Arg-221) and residues from the region between the two strands of β sheet 4 (side chains of Ala-482 and Ser-486 and main chain atoms of Val-485), with the side chains of Asp-179 (helix E) and Thr-312 (helix I) also in the vicinity.

Heme Binding Site—The heme is anchored in the binding site by hydrogen bonding interactions with the side chains of Arg-101, Trp-128, Arg-132, His-376, Ser-437, and Arg-441 in a close approximation to the situation seen in 2C9. The heme iron is pentacoordinated with Cys-443, there being no visible sign of a water molecule in the sixth coordination position in the electron density maps. There is a small area of residual electron density about 5–6 Å above the heme group, which could not be identified (supplemental Fig. 2). It is not particularly close to any active site residues but is nearest to the side chain of Phe-120. Since there are no hydrogen bonding residues nearby, it is unlikely to be a water molecule, but since a peak is present in all four 2D6 molecules, it may be significant. A similar situation has also been seen with 2C9 (26). The highly conserved Thr-309 in the I helix is in an ideal position to hydrogen-bond to the water molecule formed from the cleavage of the dioxygen bond of the heme-hydroperoxy intermediate during the P450 cycle (58).

Access Channels—Access to the cavity is almost certainly through the normal solvent channel, which, using the nomenclature of Gotoh (59), passes between the second substrate recognition site (SRS2) in the F-helix and the SRS6 turn region of β sheet 4. In doing so, the substrate can make contact with Glu-216 and Phe-483, both of which have been

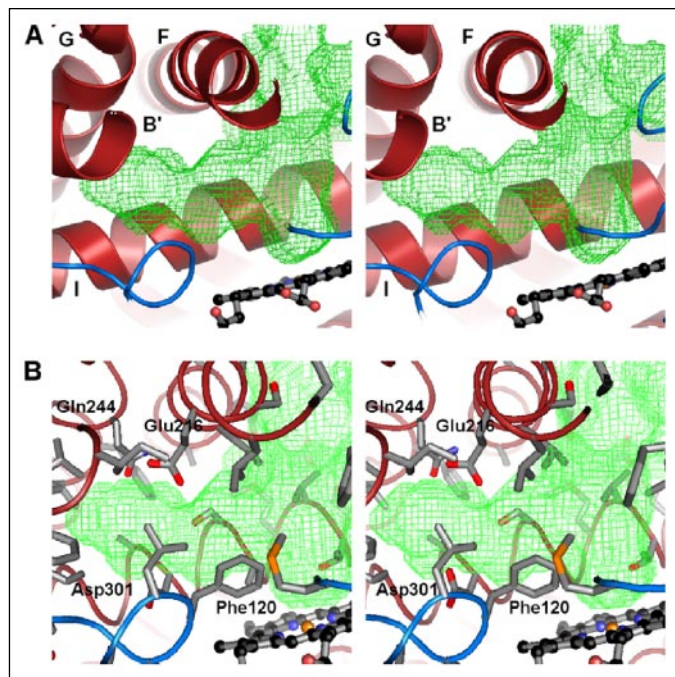


FIGURE 6. **Active site cavity.** *A*, ribbon diagram of the "right foot"-shaped active site cavity, showing the location of the B', F, G, and I helices (labeled). Also shown are the B'-C loop (lower left), the loop between β 4-1 and β 4-2 (upper right), and the loop between helix K and β 1-4 (lower right). The cavity shape was generated using a 1.4-Å radius probe occupied volume. *B*, close up of the residues surrounding the cavity, with some of the key residues labeled.

shown to be important from SDM studies (23, 60–63). A potential second egress channel can also be seen, which is similar to the PW2c pathway described by Schleinkofer *et al.* (64) for the 2C5 structure. This is quite hydrophobic in nature, initially passing Phe-120 before entering the channel itself, which is lined by Leu-248, Leu-110, and Phe-112. The product then exits through a basic opening gated by Lys-245 (G helix) and Arg-296 (I helix). Slight movements of the B'-C and SRS6 loops are necessary for free passage of both substrates and products in these two channels. However, such movements would normally be expected in such loops during the time scale of diffusion processes. The temperature factors of the residues in the first strand of β sheet 4 increase along the strand before reaching a maximum at Ala-482 that supports some degree of flexibility of the SRS6 region, whereas any effect is less noticeable in the B'-C loop.

Comparison of the Crystal Structure with Computational Models and SDM—Undoubtedly because of its importance in the metabolism of central nervous system drugs, 2D6 is one of the most widely studied P450 isoforms in terms of molecular modeling. As early as 1993, Koymans *et al.* (12) published a preliminary model based on the crystal structure of P450 101. This model only contained the active site regions but was the first to implicate Asp-301 as a residue necessary for catalytic activity (12). Since then, numerous other models have appeared, based both on bacterial enzymes (10–18) and more recently on the mammalian 2C5 isoform (19–24). Despite the low sequence identity between the structures used as the basis for homology and 2D6 itself, all of these models have a broadly similar topology, differing mainly in the loop regions as expected. Comparisons between the different models have raised some interesting questions regarding the explanations for experimental results such as SDM data, substrate regio- and stereoselectivity, and inhibitor binding. The availability of the 2D6 crystal structure has now been able to answer a number of these questions.

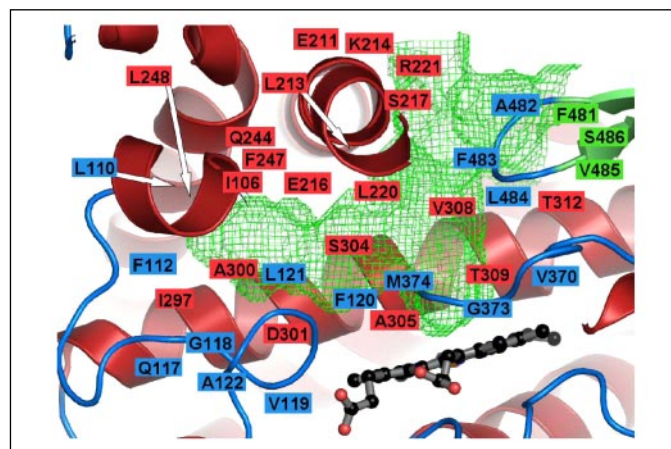


FIGURE 7. **Schematic diagram of the residues around the cavity.** The approximate locations of the amino acid side chains are indicated with labels, color-coded according to the secondary structure (red, α -helix; green, β -strand; blue, loop). The loops and strands all lie "in front of" the cavity, whereas many of the side chains labeled in the I helix, running from left to right, are "behind" the cavity.

Role of Asp-301 and Glu-216—The cavity contains the two negatively charged residues, Asp-301, in the I helix at the base of the cavity, and Glu-216, which lies on the underside of the F helix and points down into the cavity space. The carboxylate oxygens of the two residues are about 6 Å apart. Using SDM, Ellis *et al.* (65) showed in 1995 that Asp-301 played a key role in the binding of substrates to 2D6. Mutation of this residue to anything other than glutamate had a severe detrimental effect on substrate oxidation. The positioning of Asp-301 in the various models studied showed that it could readily explain the so-called 5–7-Å pharmacophore model (9) (*i.e.* that the primary binding nitrogen was 5–7 Å distant from the site of metabolism). However, the existence of numerous substrates, such as metoprolol, which are metabolized at sites further from this nitrogen, gave rise to a different 10-Å pharmacophore (10) and led Lewis to suggest that Glu-216 was the primary binding residue (16). Kirton *et al.* (22) carried out some automated docking of various ligands using the GOLD program and came to the conclusion that Glu-216 was the more likely binding residue. They and, independently, Hanna *et al.* (66) concluded that Asp-301 played a structural role in hydrogen bonding to a backbone NH of the B'-C loop. The crystal structure clearly shows that Asp-301 does indeed form two hydrogen bonds with the backbone NH groups of Val-119 and Phe-120. It is interesting to note, however, that whereas mutation of either Glu-216 or Asp-301 to Asp and Glu, respectively, can alter the rate and regioselectivity of hydroxylation of debrisoquine, mutation of either residue to a neutral amino acid results in loss of activity.⁵

We propose here that both Asp-301 and Glu-216 can act as binding residues for substrates and inhibitors of 2D6. However, the two rotameric states, *trans* and *gauche*—, of the aspartate can account for all the various pharmacophoric models, and therefore Glu-216, which sits at the top of the active site cavity, is more likely to act as a recognition residue that attracts basic ligands to the pocket and forms an intermediate binding site prior to the substrate migrating to a "reactive" position within the cavity. Docking studies of the substrate, debrisoquine, into the 2D6 crystal structure do show that it can fit readily into the pocket at the entrance channel mouth between Glu-216 and Phe-483 (supplemental Fig. 3A), without approaching a suitable distance for a reaction

⁵ S. W. Ellis, M. C. Anderson, R. C. Hartkoorn, J. R. Harlow, J. E. Chowdry, B. Mather, F. E. Blaney, and G. T. Tucker, poster presented at Molecular Graphics & Modelling Society meeting on Absorption, Distribution, Metabolism, Elimination, & Toxicity problems, April 2–4, 2003, Oxford, UK.

to occur. A similar intermediate binding pocket was observed for warfarin in its complex with 2C9 (26). Movement of the substrate toward a second docked position, where it interacts with Asp-301, could then allow the reaction to occur (supplemental Fig. 3B). Despite the difference in interpretation, most homology models show an almost identical positioning of Asp-301 and Glu-216 when compared with the crystal structure. Further resolution of the 216/301 debate must await the results of co-crystallization of substrates in the enzyme, and this is currently ongoing.

Role of the Active Site Aromatic Residues—Here again there has been some debate about the positioning of certain aromatic residues within the active site. In SRS6, which is the loop between the two β -strands of the fourth sheet region, there are two phenylalanines, Phe-481 and Phe-483. Ellis and co-workers (60, 68) have mutated these two residues both individually and together and shown that both have similar detrimental effects on the oxidation of debrisoquine. Examination of various homology models has shown that, in general, Phe-483 is oriented into the cavity, whereas Phe-481 is far removed from any possible interaction that would allow metabolism to occur. Interestingly, an exception to this is the model of de Groot *et al.* (10), which has Phe-481 pointing directly toward the heme group, with Phe-483 in an inaccessible site. The crystal structure clearly shows that Phe-483 is in the binding site, whereas Phe-481 is located remotely, in keeping with these models. It is still difficult, therefore, to explain the results of the Phe-481 mutation based on the crystal structure. However, by using constrained molecular dynamics simulations, we can show that this loop can alter its conformation so that both phenylalanines are oriented into the cavity (supplemental Fig. 3C). However, in this orientation, whereas both residues can be involved in substrate recognition, only Phe-483 can be involved in the reaction site binding.

Another critical residue identified from recent SDM studies is Phe-120, situated in the B'-C loop. It is generally found that homology models based on BM3 have this phenylalanine in the active site cavity, whereas those based on the more recent 2C5 structure have it quite distant. This arises as a result of aligning the phenylalanines present in each B'-C loop. An alternative alignment of the Phe-120 with the adjacent alanine in 2C5 does, however, place it within the cavity (21). The 2D6 crystal structure clearly places this residue within the active site in a position similar to that found in BM3. The crystal structure suggests that the role of Phe-120 is to control the orientation of substrates with respect to the heme. Almost all known 2D6 substrates contain an aromatic ring, and this is likely in many cases to form π - π interactions with Phe-120 and/or Phe-483. In docking debrisoquine into a reactive position in the crystal structure, it is prevented from adopting a coplanar orientation with respect to the porphyrin ring, and the resulting orthogonal approach of the substrate to the heme (supplemental Fig. 3B) easily explains the exclusive enantioselectivity (the main product is 4S-hydroxydebrisoquine). The aromatic hydroxylation of substrates such as debrisoquine (69) can also be better explained by the crystal structure. Bathelt *et al.* (70) have recently shown, using high level density functional calculations, that the barrier to orthogonal oxygen insertion from Compound 1 is lower than with a coplanar transition state conformation. Phe-120 plays a major role in aligning the substrate in this orientation. It is interesting to note, however, that this phenylalanine is not conserved in the rat orthologues, 2D4 and 2D2, where it is replaced by a valine. Phe-483 is also notably absent in these orthologues. 2D4 is essentially inactive toward debrisoquine, although metabolism by 2D2 can proceed smoothly (71). Venhorst *et al.* (21) have attributed this difference in 2D2/4/6 metabolism to a large negative electrostatic potential in the cavities of 2D2 and 2D6, which is quite different in 2D4. A number

of SDM studies have been carried out recently on Phe-120. Flanagan *et al.* (72) showed that the relative rates of *O*- versus *N*-demethylation of dextromethorphan were altered by the mutation of Phe-120 to alanine. Furthermore, a previously unknown metabolite, 7-hydroxydextromethorphan, was identified with this mutant (72). Keizers *et al.* (73) found that the same mutant abolished the *O*-demethylation of 7-methoxy-4-(aminomethyl)-coumarin, whereas bufuralol metabolism was unaffected. Very recently, McLaughlin *et al.* (74) reported that the F120A mutation changed the role of quinidine from being a potent inhibitor to becoming a substrate. The position of Phe-120 in the crystal structure is in clear agreement with these studies.

Other Binding Site Residues—Several of the other active site residues have been previously studied or implicated in binding by SDM and modeling. In the I helix, Ser-304 has been mutated (75) and found to have no effect on substrate turnover. In our structure, it makes a hydrogen bond with the backbone carbonyl of Ala-300. On the other side of the cavity, Val-370 and Met-374 form a hydrophobic cleft on the loop between the K helix and β 1-4. Ellis *et al.* (67) have shown that mutation of the latter to valine has a substantial effect on the stereoselectivity of metoprolol metabolism, although the kinetics of metabolism of many compounds are unaffected by this mutation. In fact, Val-374 is believed to be the actual residue in the wild-type protein. In the pocket around the entrance channel, several authors have implicated Leu-484 in their models, although no mutation work has been reported on this residue (18).

Reductase Binding Region—It has generally been found that the C-terminal one-third of the residues from various cytochromes P450 are more highly conserved, and it was found that the overlay between our most recent homology model and the crystal structure in this domain was generally very good. This region contains a number of basic residues, which form an interface with the cytochrome reductase protein. We recently described a modeling and SDM study in which interactions in the proximal face of 2D6 were studied in detail (24). In particular, a rare allelic variant in which Arg-440 is replaced by histidine is known to be inactive, and it was found that this residue formed much of the essential binding to the reductase. It was gratifying to see (supplemental Fig. 4) that there is an excellent overlap between the crystal structure and our homology model in this region.

Conclusions—In this paper, we describe the first crystal structure of human cytochrome P450 2D6 at a resolution of 3.0 Å. This shows a fold similar to other recently solved mammalian structures, although some notable differences are apparent. The structure has been compared with published homology models and has been used to explain much of the reported SDM data. Further work is now under way to determine the structures of various bound substrates and inhibitors, which should give necessary additional information on the details of ligand recognition and specificity as well as providing a structural basis on which to study the effects of various polymorphs on substrate metabolism.

Acknowledgments—Full-length 2D6 cDNA and the pCW vector were obtained from Prof. Roland Wolf (University of Dundee). This work could not have progressed without the input and assistance of Stephen Clarke and Tanya Coleman, and we greatly appreciate the help of Diana Hoult in sequencing all of the DNA constructs described. We also thank Sheena Patel for early work on the protein purification, David Earnshaw for help in obtaining activity data, Jim Nelson and Fabrice Gorrec for help with crystallization experiments, Ed Mitchell for help with data collection, Ben Bax for help in the structure determination process, and Mike Hann for helpful discussions. We acknowledge Rob Cooke, Kate Smith, Owen Jenkins, and Danuta Mossakowska for strong support of this work.

REFERENCES

- Guengerich, F. P. (2001) *Chem. Res. Toxicol.* **14**, 611–650
- Bertz, R. J., and Granneman, G. R. (1997) *Clin. Pharmacokinet.* **32**, 210–258
- Gonzalez, F. J., Skoda, R. C., Kimura, S., Umeno, M., Nebert, D. W., Gelboin, H. V., Hardwick, J. P., and Meyer, U. A. (1998) *Nature* **331**, 442–446
- Gough, A. C., Miles, J. S., Spurr, N. K., Moss, J. E., Gaedigk, A., Eichelbaum, M., and Wolf, C. R. (1990) *Nature* **347**, 773–776
- Lennard, M. S. (1990) *Pharmacol. Toxicol.* **67**, 273–283
- Meyer, U. A., Skoda, R. C., and Zanger, U. M. (1990) *Pharmacol. Ther.* **46**, 297–308
- Daly, A. K., Leathart, J. B., London, S. J., and Idle, J. R. (1995) *Hum. Genet.* **95**, 337–341
- Smith, D. A., Ackland, M. J., and Jones, B. C. (1997) *Drug Discov. Today* **2**, 406–414
- Koymans, L., Vermeulen, N. P., van Acker, S. A., te Koppele, J. M., Heykants, J. J., Lavrijsen, K., Meuldermans, W., and Donne-Op den Kelder, G. M. (1992) *Chem. Res. Toxicol.* **5**, 211–219
- de Groot, M. J., Ackland, M. J., Horne, V. A., Alex, A. A., and Jones, B. C. (1999) *J. Med. Chem.* **42**, 1515–1524
- de Groot, M. J., Ackland, M. J., Horne, V. A., Alex, A. A., and Jones, B. C. (1999) *J. Med. Chem.* **42**, 4062–4070
- Koymans, L. M. H., Vermeulen, N. P. E., Baarslag, A., and Donne-Op den Kelder, G. M. (1993) *J. Comput. Aided Mol. Des.* **7**, 281–289
- De Graaf, C., Vermeulen, N. P. E., and Feenstra, K. A. (2005) *J. Med. Chem.* **48**, 2725–2755
- Lewis, D. F. V. (1995) *Xenobiotica* **25**, 333–366
- Modi, S., Paine, M. J., Sutcliffe, M. J., Lian, L.-Y., Primrose, W. U., Wolf, C. R., and Roberts, G. C. K. (1996) *Biochemistry* **35**, 4540–4550
- Lewis, D. F. V., Eddershaw, P. J., Goldfarb, P. S., and Tarbit, M. H. (1997) *Xenobiotica* **27**, 319–339
- de Groot, M. J., Vermeulen, N. P. E., Kramer, J. D., van Acker, F. A. A., and Donne-Op den Kelder, G. M. (1996) *Chem. Res. Toxicol.* **9**, 1079–1091
- De Rienzo, F., Fanelli, F., Menziani, M. C., and De Benedetti, P. G. (2000) *J. Comput. Aided Mol. Des.* **14**, 93–116
- Lewis, D. F. V. (2002) *Xenobiotica* **32**, 305–323
- Snyder, R., Sangar, R., Wang, J., and Ekins, S. (2002) *QSAR & Combinatorial Sci.* **21**, 357–368
- Venhorst, J., ter Laak, A. M., Commandeur, J. N. M., Funae, Y., Hiroi, T., and Vermeulen, N. P. E. (2003) *J. Med. Chem.* **46**, 74–86
- Kirton, S. B., Kemp, C. A., Tomkinson, N. P., St. Gallay, S., and Sutcliffe, M. J. (2002) *Proteins Struct. Funct. Bioinf.* **49**, 216–231
- Guengerich, F. P., Hanna, I. H., Martin, M. V., and Gillam, E. M. J. (2003) *Biochemistry* **42**, 1245–1253
- Allorge, D., Breant, D., Harlow, J., Chowdry, J., Lo-Guidice, J.-M., Chevalier, D., Cauffiez, C., Lhermitte, M., Blaney, F. E., Tucker, G. T., Broly, F., and Ellis, S. W. (2005) *Proteins Struct. Funct. Bioinf.* **59**, 339–346
- Williams, P. A., Cosme, J., Sridhar, V., Johnson, E. F., and McRee, D. E. (2000) *Mol. Cell* **5**, 121–131
- Williams, P. A., Cosme, J., Ward, A., Angove, H. C., Vinković, D. M., and Jhoti, H. (2003) *Nature* **424**, 464–468
- Wester, M. R., Yano, J. K., Schoch, G. A., Yang, C., Griffin, K. J., Stout, C. D., and Johnson, E. F. (2004) *J. Biol. Chem.* **279**, 35630–35637
- Schoch, G. A., Yano, J. K., Wester, M. R., Griffin, K. J., Stout, C. D., and Johnson, E. F. (2004) *J. Biol. Chem.* **279**, 9497–9503
- Williams, P. A., Cosme, J., Vinković, D. M., Ward, A., Angove, H. C., Day, P. J., Vonnrhein, C., Tickle, I. J., and Jhoti, H. (2004) *Science* **305**, 683–686
- Yano, J. K., Wester, M. R., Schoch, G. A., Griffin, K. J., Stout, C. D., and Johnson, E. F. (2004) *J. Biol. Chem.* **279**, 38091–38094
- Yano, J. K., Hsu, M.-H., Griffin, K. J., Stout, C. D., and Johnson, E. F. (2005) *Nat. Struct. Mol. Biol.* **12**, 822–823
- Ravichandran, K. G., Boddupalli, S. S., Hasermann, C. A., Peterson, J. A., and Deisenhofer, J. (1993) *Science* **261**, 731–736
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nuc. Acids Res.* **25**, 3389–3402
- Eddy, S. R. (1996) *Curr. Opin. Struct. Biol.* **6**, 361–365
- Abagyan, R. A., Totrov, M. M., and Kuznetsov, D. A. (1994) *J. Comput. Chem.* **15**, 488–506
- Barnes, H., Arlotto, M., and Waterman, M. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 5597–5601
- Gillam, E. M. J., Baba, T., Kim, B. R., Ohmori, S., and Guengerich, F. P. (1993) *Arch. Biochem. Biophys.* **305**, 123–131
- von Wachenfeldt, C., Richardson, T. H., Cosme, J., and Johnson, E. F. (1997) *Arch. Biochem. Biophys.* **339**, 107–114
- Omura, T., and Sato, R. (1964) *J. Biol. Chem.* **239**, 2379–2385
- Zhang, Y., and Pernecky, S. J. (1999) *Biochem. Biophys. Res. Commun.* **258**, 32–38
- Venhorst, J., Onderwater, R. A. C., Meerman, J. H. N., Commandeur, J. N. M., and Vermeulen, N. P. E. (2000) *Drug Metab. Dispos.* **28**, 1524–1532
- Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307–326
- CCP4 (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50**, 760–763
- Storoni, L. C., McCoy, A. J., and Read, R. J. (2004) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 432–438
- Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132
- Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53**, 240–255
- Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57**, 122–133
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283–291
- DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA
- Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) *J. Comput. Chem.* **4**, 187–217
- Accelrys Inc. (1999) *Quanta98*, Version 98.1111, Accelrys Inc., San Diego, CA
- Wavefunction Inc. (1999) *SPARTAN SGI*, Version 5.1.3, Wavefunction Inc., Irvine, CA
- Pritchard, M. P., Ossetian, R., Li, D. N., Henderson, C. J., Burchell, B., Wolf, C. R., and Friedberg, T. (1997) *Arch. Biochem. Biophys.* **345**, 342–354
- Richardson, T. H., Hsu, M. H., Kronbach, T., Barnes, H. J., Chan, G., Waterman, M. R., Kemper, B., and Johnson, E. F. (1993) *Arch. Biochem. Biophys.* **300**, 510–516
- Szczesna-Skorupa, E., Straub, P., and Kemper, B. (1993) *Arch. Biochem. Biophys.* **304**, 170–175
- Kleywegt, G. J. (1996) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **52**, 842–857
- Kleywegt, G. J., and Jones, T. A. (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50**, 178–185
- Schlichting, I., Berendzen, J., Chu, K., Stock, A. M., Maves, S. A., Benson, D. E., Sweet, R. M., Ringe, D., Petsko, G. A., and Sligar, S. G. (2000) *Science* **287**, 1615–1622
- Gotoh, O. (1992) *J. Biol. Chem.* **267**, 83–90
- Chowdry, J., Pucci, M. R., Harlow, J., Tucker, G. T., and Ellis, S. W. (2002) *Br. J. Clin. Pharmacol.* **53**, 443–444
- Paine, M. J. I., McLaughlin, L. A., Flanagan, J. U., Kemp, C. A., Sutcliffe, M. J., Roberts, G. C. K., and Wolf, C. R. (2003) *J. Biol. Chem.* **278**, 4021–4027
- Ellis, S. W., Keenan, L., Harlow, J., Chowdry, J., Gregory, A., Modi, S., Eddershaw, P. J., and Tucker, G. T. (2002) *Br. J. Clin. Pharmacol.* **53**, 444–445
- Lussenburg, B. M. A., Keizers, P. H. J., de Graaf, C., Hidestrand, M., Ingelman-Sundberg, M., Vermeulen, N. P. E., and Commandeur, J. N. M. (2005) *Biochem. Pharmacol.* **70**, 1253–1261
- Schleinkofer, K., Sudarko, Winn, P. J., Luedemann, S. K., and Wade, R. C. (2005) *EMBO Rep.* **6**, 584–589
- Ellis, S. W., Hayhurst, G. P., Smith, G., Lightfoot, T., Wong, M. M. S., Simula, A. P., Ackland, M. J., Sternberg, M. J. E., Lennard, M. S., Tucker, G. T., and Wolf, C. R. (1995) *J. Biol. Chem.* **270**, 29055–29058
- Hanna, I. H., Kim, M.-S., and Guengerich, F. P. (2001) *Arch. Biochem. Biophys.* **393**, 255–261
- Ellis, S. W., Hayhurst, G. P., Lightfoot, T., Smith, G., Harlow, J., Rowland-Yeo, K., Larsson, C., Mahling, J., Lim, C. K., Wolf, C. R., Blackburn, M. G., Lennard, M. S., and Tucker, G. T. (1996) *Biochem. J.* **316**, 647–654
- Hayhurst, G. P., Harlow, J., Chowdry, J., Gross, E., Hilton, E., Lennard, M. S., Tucker, G. T., and Ellis, S. W. (2001) *Biochem. J.* **355**, 373–379
- Lightfoot, T., Ellis, S. W., Mahling, J., Ackland, M. J., Blaney, F. E., Biljoo, G. J., de Groot, M. J., Vermeulen, N. P. E., Blackburn, G. N., Lennard, M. S., and Tucker, G. T. (2000) *Xenobiotica* **30**, 219–233
- Bathelt, C. M., Ridder, L., Mulholland, A. J., and Harvey, J. N. (2004) *Org. Biomol. Chem.* **2**, 2998–3005
- Hiroi, T., Chow, T., Imaoka, S., and Funae, Y. (2002) *Drug Metab. Dispos.* **30**, 970–976
- Flanagan, J. U., Marechal, J.-D., Ward, R., Kemp, C. A., McLaughlin, L. A., Sutcliffe, M. J., Roberts, G. C. K., Paine, M. J. I., and Wolf, C. R. (2004) *Biochem. J.* **380**, 353–360
- Keizers, P. H. J., Lussenburg, B. M. A., de Graaf, C., Mentink, L. M., Vermeulen, N. P. E., and Commandeur, J. N. M. (2004) *Biochem. Pharmacol.* **68**, 2263–2271
- McLaughlin, L. A., Paine, M. J. I., Kemp, C. A., Maréchal, J.-D., Flanagan, J. U., Ward, C. J., Sutcliffe, M. J., Roberts, G. C. K., and Wolf, C. R. (2005) *J. Biol. Chem.* **280**, 38617–38624
- Ellis, S. W., Rowland, K., Ackland, M. J., Rekka, E., Simula, A. P., Lennard, M. S., Wolf, C. R., and Tucker, G. T. (2000) *Biochem. J.* **345**, 565–571

Supplementary Fig. 1. Stereo views of the 2D6 structure.

A. The 2D6 C α trace. *B.* Superposition of the 2D6 C α trace (blue) with that of 2C9 (magenta), with the heme groups both shown in green. Both stereo views are shown in the same orientation as Fig. 4.

Supplementary Fig. 2. Electron density in the vicinity of the heme environment.

The final $2F_o - F_c$ map is shown at a 1σ contour level (blue) with the corresponding $F_o - F_c$ map at contour levels 3σ (green) and -3σ (red).

Supplementary Fig. 3. Docking and molecular dynamics studies.

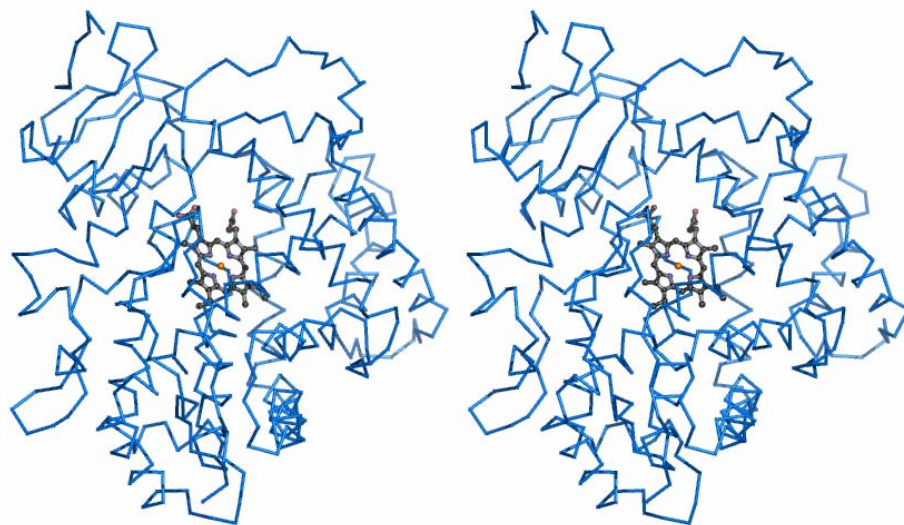
A. Debrisoquine docked into the crystal structure cavity at the mouth of the entrance channel. It forms a clear salt bridge between the substrate's guanidine group and Glu-216 with orthogonal π - π stacking with Phe-483. Phe-481 is too far away to make any contact with the ligand. In this binding pocket debrisoquine is a long distance (12.2 Å) from the oxygen of Compound 1. *B.* Debrisoquine now docked into the crystal structure's reaction site with the substrate bound between Phe-120 and Phe-483. The salt bridge is formed with Asp-301 in its alternative rotameric state to that found in the crystal. It can be seen that the *pro-S* hydrogen is 2.5 Å from the oxygen whereas the *pro-R* hydrogen is much further away at 3.5 Å. This easily explains the preference for the stereoselective hydrogen abstraction which gives rise to the 4*S*-hydroxy product via a radical rebound mechanism. *C.* Molecular dynamics simulation showing movement of the β sheet 4 loop. In yellow is the crystal structure where Phe-481 and Phe-483 are far apart. Overlaid in grey is the loop structure following the simulation where Phe-481 has now moved into the cavity.

Supplementary Fig. 4. Overlay of the crystal structure (yellow) with our homology model (grey) in the proximal region of the protein.

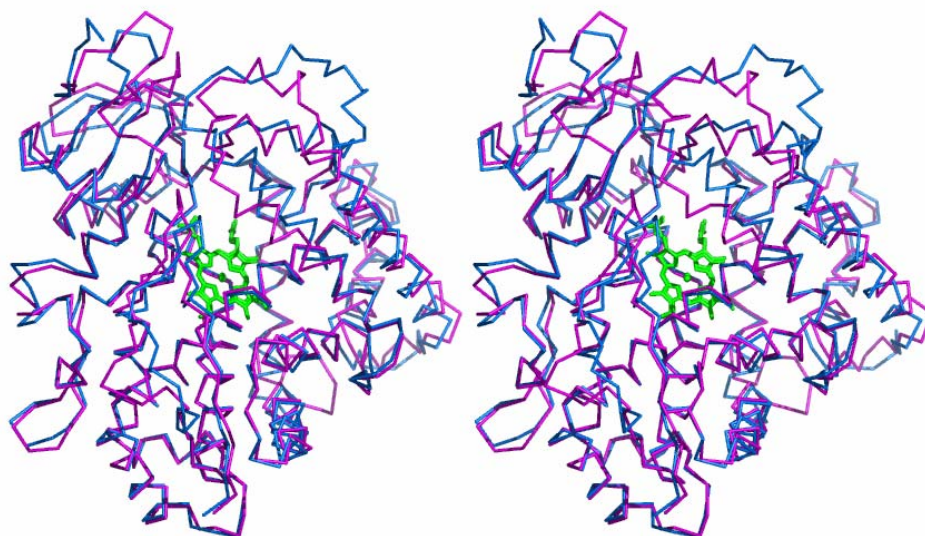
There is an almost perfect juxtaposition of Arg-440 which is the key residue involved in the reductase binding.

Supplementary Figure 1

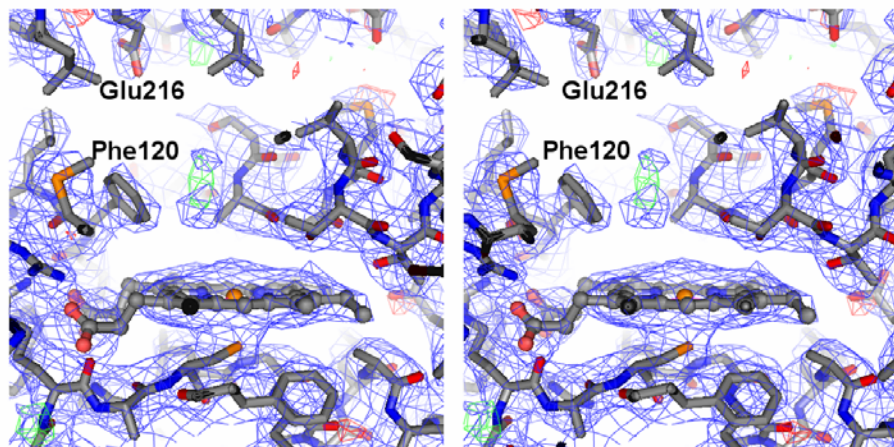
A



B

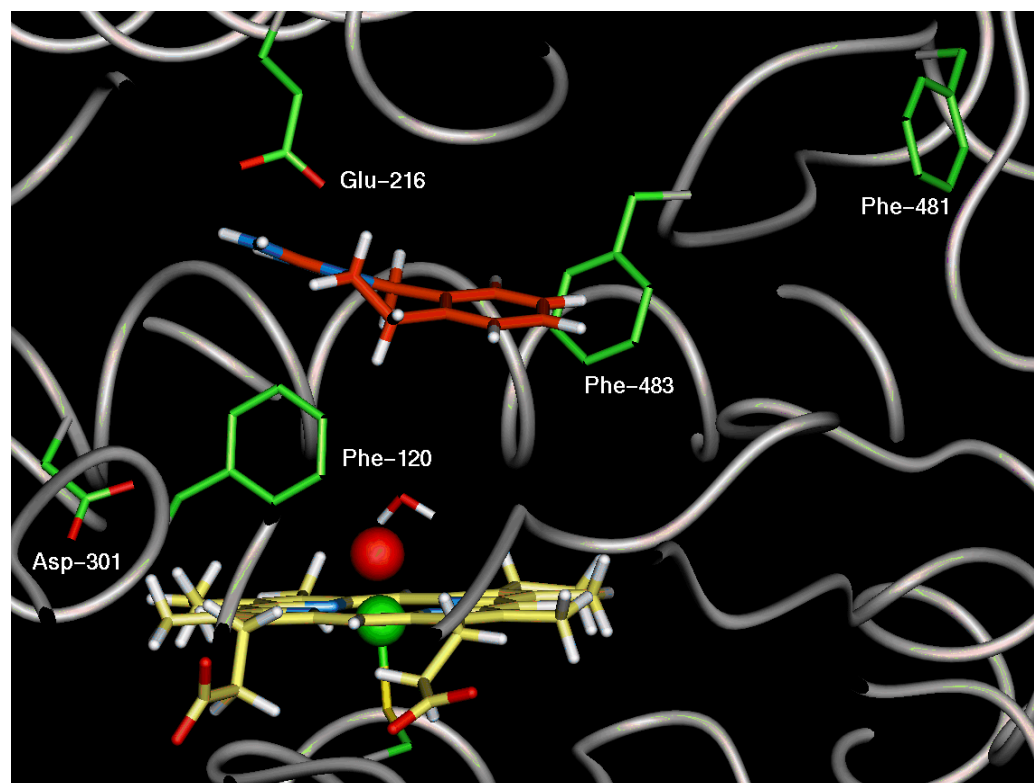


Supplementary Figure 2

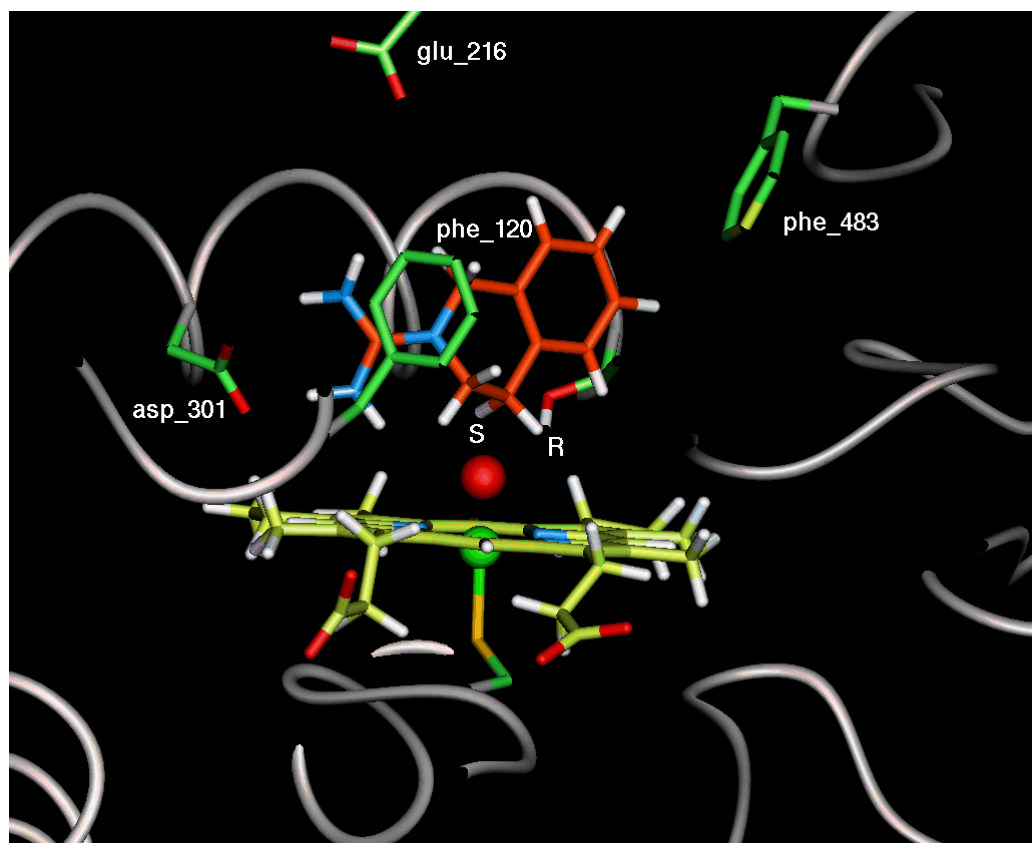


Supplementary Figure 3

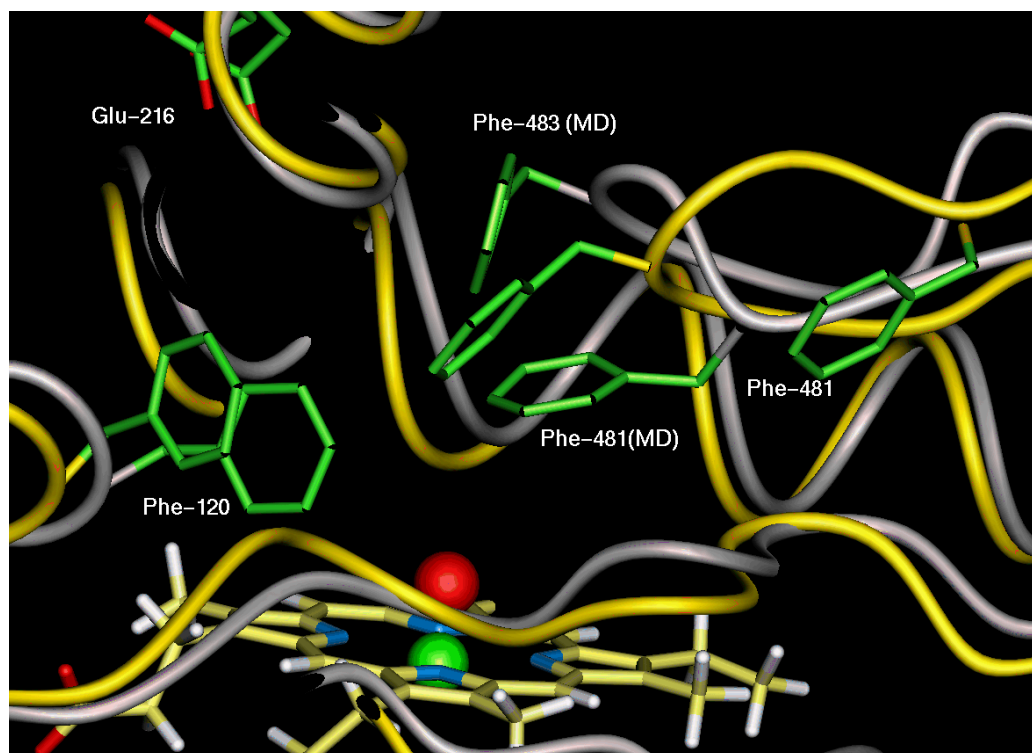
A



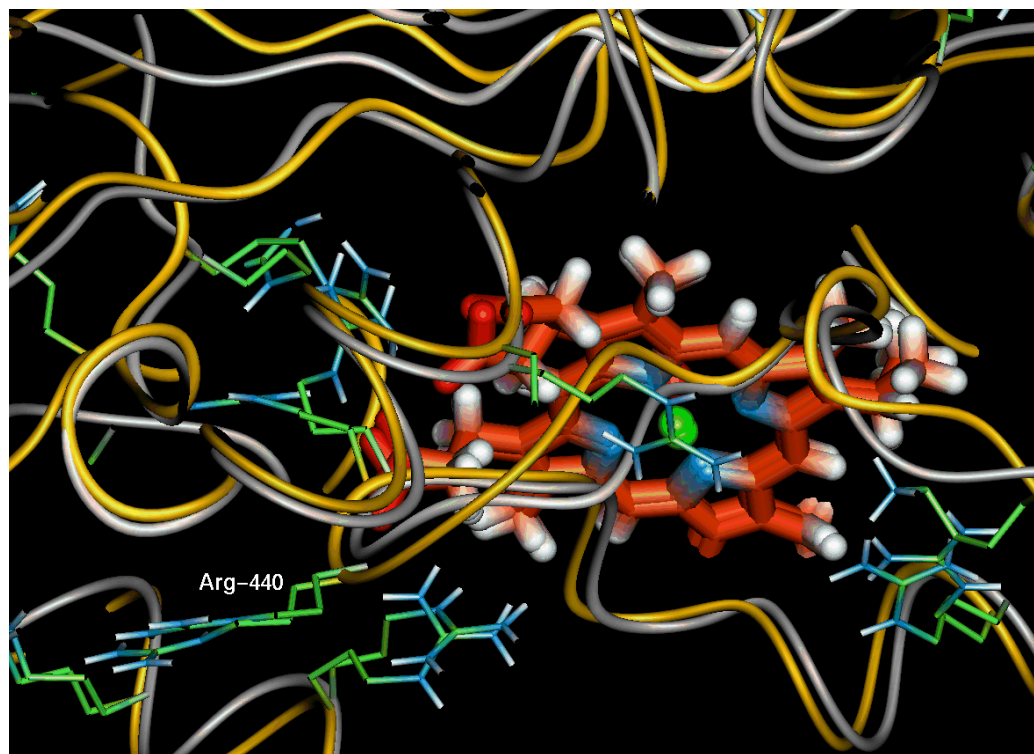
B



C



Supplementary Figure 4



Crystal Structure of Human Cytochrome P450 2D6

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

J. Biol. Chem. 2006, 281:7614-7622.

doi: 10.1074/jbc.M511232200 originally published online December 13, 2005

Access the most updated version of this article at doi: [10.1074/jbc.M511232200](https://doi.org/10.1074/jbc.M511232200)

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/2005/12/14/M511232200.DC1.html>

This article cites 72 references, 16 of which can be accessed free at
<http://www.jbc.org/content/281/11/7614.full.html#ref-list-1>