Crystal Structure of Human Cytochrome P450 2D6

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

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

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 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 has been found with previous bacterial enzyme structures and led quickly to a new generation of computational models (19–24). 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/sparetine 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 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 of...
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 Ndel 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 Ndel and XbaI, the PCR products were subcloned into the pCw expression vector (36). The sequence of all constructs was confirmed by automated dyeoxy-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 CTGAATGCTCTCCGTCMRNCTGCATATCC-CAGCCGTGCGCTG was used to introduce Asn, Lys, His, Gin, 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 initial absorbance 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 KPi, pH 7.6, 20% glycerol, 1 mM phenylmethylsulfonyl fluoride, and 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 TLA-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 was grown for 12 h at 37 °C with 0.2 mg/ml benzonase, and protease inhibitors) at 4 °C with vigorous shaking. The culture was harvested after a further 16 h at 30 °C by centrifugation at 5000 × g, 20 min, and the pellet was resuspended in 200 ml of Buffer A (400 mM KPi, 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: 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 KPi, 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 KPi, 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 KPi, 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 KPi, 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 KPi, 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 (Coomasie 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). A 10-fold dilution of 4-methylaminomethyl-7-methoxycoumarin (GlaxoSmithKline) was added to 10 µmol of 2D6 in the presence of 10 nmol of cumene hydroperoxide (as an electron donor), 1 mM dithiothreitol, 40 mM KPi, pH 7.5, and 20% glycerol. Readings were taken every 60 s for 10 min. The 2D6 L230D/L231R metabo-

**Cytochrome P450 2D6 Crystal Structure**
Cytochrome P450 2D6 Crystal Structure

### Crystallographic statistics

| Parameter | Value |
|-----------|-------|
| Resolution (Å) | 40–3.0 (3.11–3.00) |
| Observations | 94,11 |
| Unique reflections | 39,448 |
| Completeness (%) | 90.0 (90.7) |
| Average I/σI | 11.2 (1.6) |
| Rmerge | 0.107 (0.614) |

### Refinement statistics

| Parameter | Value |
|-----------|-------|
| Resolution (Å) | 40–3.0 |
| R factor | 0.230 |
| Rmerge (% of data) | 0.286 |
| Bond lengths (Å) | 0.016 |
| Bond angles (degrees) | 1.65 |
| Average B factors (Å²) | 56.9 (7240) |
| Protein side chain | 57.5 (6996) |
| Hemes | 52.9 (172) |
| Waters | 39.0 (11) |
| Sulfates | 53.9 (10) |

### Notes
- Data for the highest resolution shell are given in parentheses.
- Rmerge = Σ||Fobs|−⟨|Fobs⟩||/Σ|Fobs|.
- Rmerge (% of data) = Σ||Fobs|−⟨|Fobs⟩||/Σ|Fobs|.
- The number of model atoms is given in parentheses.

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 2Fo − Fc and Fo − Fc (where Fc represents the observed structure factor and Fc 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 Rfree values (0.230 and 0.286, respectively) compared with an equivalent non-TLS refinement cycle, where R and Rfree 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 alamines, 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 Ca 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% cytoplastic 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. 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 Ca 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

--- F helix----|loop----| helix? ----|----|loop-----| G helix--
2D6 Q E G L K E E S G F L R E Y L N A V P V S L H I P A L A G K V L R
2C5 N O V E R I L S S P W L Q Y 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 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 His6 tag and residues 2–10 of 2C3d inserted at the N terminus. Trunc6, 2D6 truncated at residue 25 with C-terminal His6 tag. Trunc7, 2D6 truncated at residue 32 with C-terminal His6 tag and residues 2–6 of 2C3d inserted at the N terminus.

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.

MARCH 17, 2006 • VOLUME 281 • NUMBER 11

JOURNAL OF BIOLOGICAL CHEMISTRY 7617

Cytochrome P450 2D6 Crystal Structure
structures (Protein Data Bank codes 1OG2 and 1OG5 versus 1R09), 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 \( \beta \) 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 \( \beta \) 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, \( \beta \) 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
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 sheet 1 (side chains of Val-370 and Met-374 and main chain atoms of Gly-373), and residues from the loop between the strands of 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 helix 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 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 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).

**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 sheet 4. In doing so, the substrate can make contact with Glu-216 and Phe-483, both of which have been...
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 Schleiniker 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.

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.

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. Hartkorn, 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 $\beta$-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$\alpha$-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$\alpha$-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 $\pi-\pi$ 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 B1-4. 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 overlap 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 Houlit 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 crystalization 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.
