Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 26 (2008) 947–956
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Analysis of CYP2D6 substrate interactions by computational methods

Yuko Ito [a,*], Hiroki Kondo [a], Peter S. Goldfarb [b], David F.V. Lewis [b]

[a] *Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-City, Fukuoka 820-8502, Japan*

[b] *School of Biomedical and Molecular Sciences, University of Surrey, Guildford, Surrey GU2 7XH, UK*

Received 13 February 2007; received in revised form 18 July 2007; accepted 20 July 2007
Available online 27 July 2007

Abstract

Cytochrome P450 CYP2D6 is involved in the oxidation of well over 150 drugs and, in general, those which contain a basic nitrogen atom in the molecule. To clarify how the residues of CYP2D6 are utilized for orientating a wide range of its specific substrates and distinguishing them from a variety of other organic compounds, docking studies by AutoDock and molecular dynamics (MD) simulations were conducted. Specific ligands were docked to both the homology model and crystal structures optimally to estimate the site of reaction on the ligand molecule and the binding energy for the complex, which were generally in good agreement with the experimental data. MD simulation for the CYP2D6–propranolol complex was then carried out to reveal the amino acid residues interacting with the substrate at the active site. Phe-120, Glu-216, Asp-301, and Phe-483 are identified as the substrate-binding residues in agreement with previously reported site-directed mutagenesis data and the crystal structure reported recently (PDB code: 2F9Q). As well as these residues, our theoretical prediction suggests that Phe-219 and Glu-222 are also important residues for mediating oxidation of substrates, especially propranolol.

Keywords: Cytochrome P450 (CYP); Automated docking; Homology modelling; Molecular dynamics (MD); Site-directed mutagenesis

1. Introduction

Cytochromes P450 (CYP) constitute a large superfamily of heme-thiolate enzymes. They have been found in all five biological kingdoms, including mammals, possibly indicating that P450s may have evolved from a common ancestor during development of the biota [1,2]. P450s can usually metabolize a large number of structurally diverse endogenous and exogenous compounds due to a broad substrate specificity, and generally with a wide regio- and stereoselectivity [1,3]. Six substrate recognition sites (SRSs) are thought to be present on P450s which display distinct differences in their substrate-binding residues [4].

A total of 57 CYP enzymes have been identified thus far in humans. Recent estimates suggest that the following four human P450s, CYP1A2, CYP2C9, CYP2D6 and CYP3A4, are primarily responsible for the metabolism of the majority of pharmaceuticals in current clinical use. Approximately 34% of cytochrome P450-mediated drug oxidations are performed by CYP3A4, 19% by CYP2D6, 16% by CYP2C9 and 8% by CYP1A2 with smaller contributions from CYP2C19, CYP2B6, CYP2A6, CYP2E1 and CYP2C8 [5–7].

Among these drug-metabolizing human P450s, CYP2D6 bears unique structural features in its preferred substrates which usually contain a basic nitrogen and a planar aromatic ring. Another characteristic feature of CYP2D6 is related to the occurrence of genetic polymorphisms in human ethnogeographical populations [8–15]. The genetic factors modulating the catalytic competence of CYP2D6-mediated metabolism in humans behooves a considerably heightened awareness of the potential consequences for impaired clearance of CYP2D6 substrates in individuals with either ‘rapid-’ or ‘poor-metabolizer’ status. Consequently, there is current interest in the prediction of CYP2D6-mediated metabolism and selectivity, how CYP2D6 binds the substrate within the protein cavity and which residues are involved in the binding process. We have focused on using computational methods, including interactive docking studies and molecular dynamics (MD) simulation, to explore these features in some detail.

* Corresponding author. Tel.: +44 1483 68 6477; fax: +44 1483 57 6978.
E-mail addresses: yuko_ito77@hotmail.com, y.ito@surrey.ac.uk, d791002y@bio.kyutech.ac.jp (Y. Ito).

1093-3263/$ – see front matter
doi:10.1016/j.jmgm.2007.07.004

2. Methods

The three-dimensional (3D) coordinates for the CYP2D6 enzyme were constructed by homology modelling based on the rabbit CYP2C5 enzyme (PDB code: 1N6B), for which the crystal structure was available at high resolution [16,17]. When this study was in progress, however, the crystal structure of human CYP2D6 (PDB code: 2F9Q) became available, enabling comparison with our theoretical model [18]. Typical CYP2D6 substrates were then docked *in silico* to the resulting 3D enzyme structure to obtain substrate binding information for various compounds. Then MD simulation was performed on a selected docked structure, the propranolol–CYP2D6 structure complex, to reveal the details for important residues. These simulated data were compared with the published experimental data on the substrate affinity, the substrate metabolism, and site-directed mutagenesis.

2.1. Homology modelling

The three-dimensional coordinates for CYP2D6 were constructed by homology modelling based on the rabbit CYP2C5 crystal structure. All homology modelling and molecular mechanics of human P450 enzyme were performed using the Sybyl software package (Tripos Associates, St. Louis, MO) as implemented on a Silicon Graphics Indigo^2 10000 high impact graphic workstation operating under UNIX. Details of the homology modelling method have been reported previously [19–21].

2.2. Docking studies with AutoDock 3.05

Using the AutoDock 3.05 software package [22], docking studies were conducted on CYP2D6. The docking substrates were built and geometry optimized via the semi-empirical program MOPAC [23]. The search method used was a Lamarckian Genetic Algorithm (LGA) with 50 LGA runs [24]. The number of individuals in each population was set at 80. The maximum number of energy evaluations was set at 900,000 and the maximum number of generations was set at 50,000. The number of top docking orientations that automatically appear during the calculations was fixed at two. The rates of gene mutation and crossover were set at 0.02 and 0.80, respectively. We have found that these settings give consistent results in the distribution of the final top 50 structures. During this searching process, the ligands (i.e. substrates) were regarded as being flexible, while the enzyme was regarded as rigid and each of its atoms was approximated as a grid point. The box was fixed at the center of the heme moiety and the box size was normally set at 60, 60, and 60 Å (x, y, and z), though it is changed depending on the ligand size.

2.3. Molecular dynamics

MD simulation was carried out using the Amber 7.0 package including the Amber99 force-field, except for the atomic charges in P450 substrates which were assigned by the semi-empirical program MOPAC [23,25,26]. The heme parameters were taken from those compiled by Helms and Wade [27].

The MD simulation was conducted on the CYP2D6 model structure with the docked propranolol substrate. The system was solvated with TIP3 water molecules generated via a rectangular box. The number of solvated water molecules in each system is about 17,000 and the initial MD simulation cell dimension was about 95 Å × 80 Å × 100 Å and involved the complex being solvated by a layer of water molecules of at least 10 Å in all directions. Since the system has net positive charges, a correct number of counter ions as Cl^- ions were added to neutralize the box charges.

A periodic boundary condition was applied, and the pressure and temperature were kept constant in the system. Only bond lengths involving hydrogen atoms are constrained using the SHAKE algorithm on each 0.2 fs step. Particle Mesh Ewald was applied to Coulomb interactions as an approximation. Non-bonded van der Waals interactions were calculated by a cutoff method with the distance of the cutoff being set to 9 Å. The integration time-step of MD simulation used was 1 fs.

The procedure used in our simulations is shown in Fig. 1. (1) The potential energy minimization for solvent water molecules was performed first, and then only they were relaxed by MD simulation. (2) The system was then relaxed with sequential steps of energy minimization. (3) No restraints were applied at the final minimizing step. (4) The temperature of the whole system was gradually raised by heating to 300 K over 80 ps under MD. Sampling MD simulation was then performed over 1 ns at 300 K.

The convergence of the different simulations was assessed in terms of root-mean-square-deviation (R.M.S.D.) from the last minimized structure and root-mean-square fluctuation (R.M.S.F.) to estimate the *B*-factor.

3. Results and discussion

3.1. Initial model structures

Model structures for CYP2D6 were constructed as detailed under Section 2. There is a 42% protein sequence homology

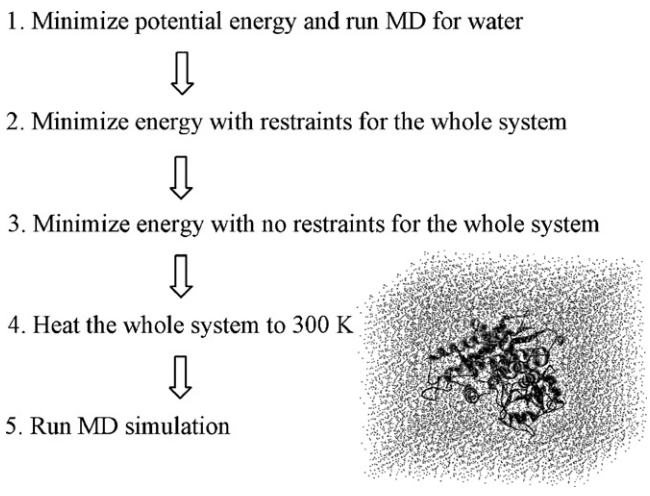

Fig. 1. Procedure for MD simulation.

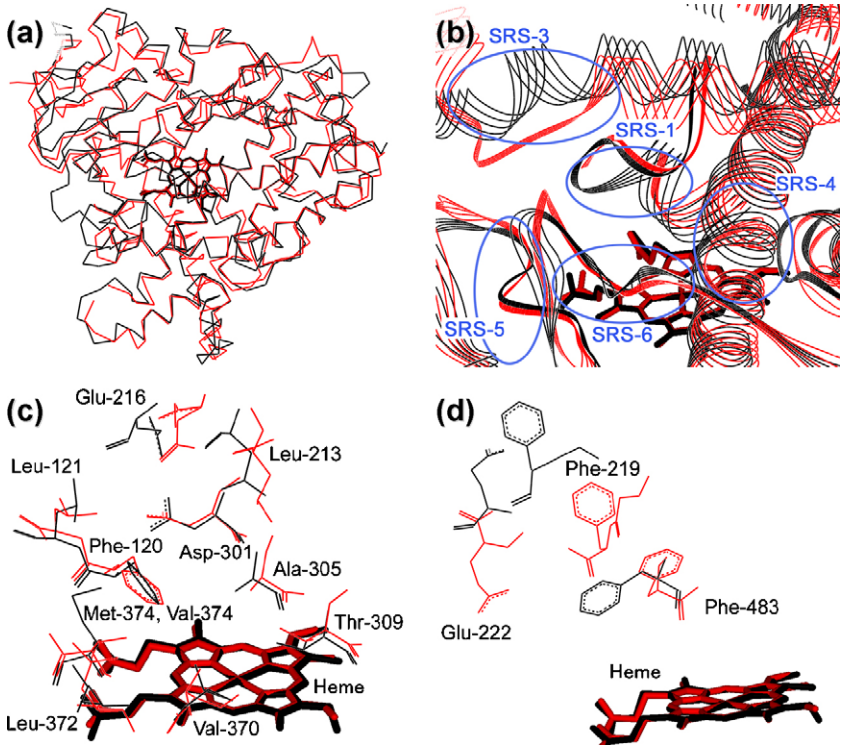

Fig. 2. Superimposed structures of the CYP2D6 crystal structure (PDB code: 2F9Q) and the model coordinates; α-carbon framework of the entire structures (a) and substrate binding (b), and substrate-binding residues of the same (c) and different positions (d). The crystal and model structures are depicted in black and red, respectively.

between the template structure, rabbit CYP2C5, and human CYP2D6. The energy-minimized homology model thus constructed was compared with the CYP2D6 (2F9Q) enzyme by structural overlay, as shown in Fig. 2a and b. The R.M.S.D. between the α-carbons in the constructed model and those in the crystal structure was found to be 2.77 Å. According to the criteria with respect to the sequence identity and structural R.M.S.D., the model structure can be thus categorized into the medium-accuracy class of comparative models [28,29].

Specifically, Phe-120, Leu-121 (SRS1), Leu-213, Glu-216 (the front part of SRS3), Asp-301, Ala-305, Thr-309 (SRS4), Val-370, Leu-372, and Met/Val-374 (SRS5) reside at almost the same position in the two structures (2F9Q is a mutated enzyme with methionine replaced for Val-374), as shown in Fig. 2c. On the other hand, the conformation of the three residues, Phe-219, Glu-222 (SRS3) and Phe-483 (SRS6), is different between the two structures, as shown in Fig. 2d. While the α-carbon of Glu-222 and Phe-483 is superimposed neatly between the model and crystal structures, the Glu-222 side-chain in the crystal is directed outwards from the cavity, which is totally opposite to that in the model. In addition, the aromatic ring of Phe-483 is slightly farther away from the catalytic center in the crystal structure.

With Phe-219 constituting SRS3, the location of the α-carbon itself is significantly different between the two structures. This residue is contained in the F-helix of the crystal structure, while a loop is formed at the same position in the model structure, as shown in Fig. 2b. In the CYP2C5 structure (1N6B), on which the current model is based, the sequence representing F-helix in CYP2D6 (2F9Q) is 10 amino acids shorter and it constitutes F–G′ loop rather than a helix. Accordingly, the corresponding sequence (SRS3) of CYP2D6 cannot form an α-helix in the model structure, thus making an extra space available in the catalytic cavity which is connected to another cavity. For this reason as well as the conformational difference in the side-chains described above, the cavity volume in the model is considerably larger than that in the crystal structure: the calculated volume for the substrate-binding cavity using VOIDOO is 620 $Å^3$ for the model and 412 $Å^3$ for the crystal structure (2F9Q) [30].

Although there are some differences between the model and crystal structures, their substrate-binding residues and the location of the α-carbons in the core region are comparable. In general, medium-accuracy class models are often useful in assessing the function of individual residues [29]. It is even possible to predict correctly certain features of the target protein which are not manifest in the template structure.

3.2. Docking studies

Typical substrates were docked both in the homology model and the crystal structure. The theoretical binding energy was compared with the experimental binding affinity usually obtained from the K_m value (Michaelis constant) and, occasionally, the K_s value (spectroscopic binding constant), and the value is converted to experimental binding energy, ΔG_{bind}, by the equation $\Delta G_{bind} = RT \ln K$, where R is the gas constant, T the absolute temperature (310 K) and K is either K_m

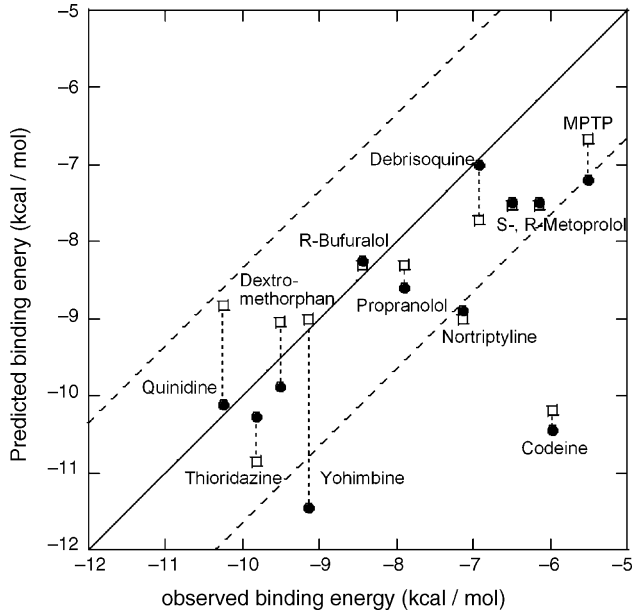

Fig. 3. Theoretical vs. observed binding energies for each of P450 docked structures (Table 1). The solid line represents a perfect fit, and the dotted lines show one standard deviation from ideality. (●) Model structure; (□) crystal structure.

or K_s. In addition, the theoretically preferred site of reaction on the substrate was deduced by the distance from the reaction center of the complex to the site of reaction in question.

The calculated energies are plotted against the observed values in Fig. 3, where the solid line represents the ideal line with a slope of unity and the dotted lines indicate the standard deviation from ideality. Of the twelve compounds tested, all but three, namely codeine, nortriptyline and yohimbine, fell inside of the dotted lines in the model structure, while two compounds, codeine and nortriptyline, are outside of the range in the crystal structure. Judging from the site of reaction on the substrates (Table 1) [31–41], the docked conformations seem to be also consistent with the experimental data in most cases. However, the docked conformation relative to the heme position failed to predict the site of reaction for yohimbine in the model structure, and debrisoquine and dextromethorphan in the crystal structure.

3.2.1. Codeine

Although AutoDock reproduces the conformation of codeine correctly in both the model and crystal structures, both of the docking energies are nearly 4.0 kcal/mol more stable than the experimental energy, as shown in Fig. 3. Their theoretical binding energies are −10.4 and −10.2 kcal/mol for the model and crystal structures, respectively, whereas the binding energy derived from the experimental binding affinity for codeine is −5.98 kcal/mol [38–40]. Such an inconsistency in the docking results can happen in AutoDock docking studies, even where the orientations are correctly predicted, as some of the interactions between the substrate and enzyme may be artifactual or missing in the inflexible protein structure. These kind of artifacts can occur, as the side-chains of amino acid residues are not allowed to move in the docking process.

3.2.2. Yohimbine

Despite the successful docking of yohimbine in the crystal structure, the docking result for yohimbine in the model structure is inconsistent with the experimental data in terms of docking energy and conformation. The binding energy is 2.45 kcal/mol lower than the experimental energy [41], and yohimbine lies above the heme in an orientation incompatible with metabolism. Compared with the model of CYP2D6, the crystal structure has a cavity which can accommodate ligands between the F-helix and the C-terminal loop located on the surface of the protein. Therefore, these loop and helix motifs appear to ‘sandwich’ yohimbine such that it binds to the heme perpendicularly. Whereas, the space constraints are much tighter in the model structure, thus hampering the binding of the

Table 1
Results of docking predictions on a dataset of 11 compounds

Ligand	K_m or K_s (μM)	ΔG_{exp} (kcal/mol)	ΔG_{calc} (kcal/mol) model/crystal	Preferred site of reaction	Orientation obtained by AutoDock model/crystal[a]
Propranolol	2.7	−7.90	−8.59/−8.30	4-Hydroxylation	○/○
Metoprolol	46(R) 26(S)	−6.15 −6.50	−7.49/−7.52	*O*-Demethylation	○/○
Thioridazine	0.1	−9.82	−10.27/−10.86	*S*-Oxidation	○/○
R-Bufuralol	1.1	−8.45	−8.25/−8.31	1′-Hydroxylation	○/○
MPTP	13	−5.51	−7.19/−6.67	*N*-Demethylation	○/○
Debrisoquine	13	−6.93	−7.00/−7.72	4′-Hydroxylation	○/△
Dextromethorphan	0.2	−9.51	−9.88/−9.04	*O*-Demethylation	○/△
Nortriptyline	47 15	−7.14	−8.89/−9.00	10-Hydroxylation	○/○
Codeine	330 150	−5.98	−10.44/−10.20	*O*-Demethylation	○/○
Quinidine	0.1	−10.24	−10.12/−8.82	Inhibition	–/–
Yohimbine	0.4	−9.14	−11.46/−9.00	Inhibition (11-Hydroxylation)	△/○

[a] (○) The reacting atom is oriented toward the heme. (△) The ligand is located close to the heme, but the reacting atom is not oriented toward the heme. (–) Not a substrate but an inhibitor.

substrate in the right orientation. Generally, where there is a loop(s) near the substrate-binding site, it is difficult to construct a correct model because of the flexibility of the loop. In the homology modelling methods used, several loops were generated using loop optimization in some cases and several candidates can then be selected from those found in the protein database by Sybyl software, but finally one loop conformation needs to be chosen by the user on a subjective basis. Accordingly, such a problem is difficult to avoid with the methodology currently in use.

3.2.3. Quinidine

This compound is unique and attracted special attention, as it is not metabolized by CYP2D6 but is a potent competitive inhibitor. AutoDock gives binding energies of −10.1 and −8.82 kcal/mol for the model and crystal structures, respectively. These energies do not differ much from the value of −9.3 kcal/mol obtained by the previous docking studies, using GOLD, and agree satisfactorily with the experimental value of −10.24 kcal/mol [42–44]. Site-directed mutagenesis and homology modelling studies have revealed that Phe-120, Glu-216 and Asp-301 are important for quinidine binding in CYP2D6 [42]. Thus, Phe-120 exerts a direct effect on quinidine to bind in an unproductive fashion, and Glu-216 plays a role in determining the mode of binding, whereas Asp-301 is involved in the binding indirectly through interactions with the backbone of the B′–C loop containing Phe-120 [42]. In fact, these interactions are also observed in quinidine binding for our current docking study using AutoDock.

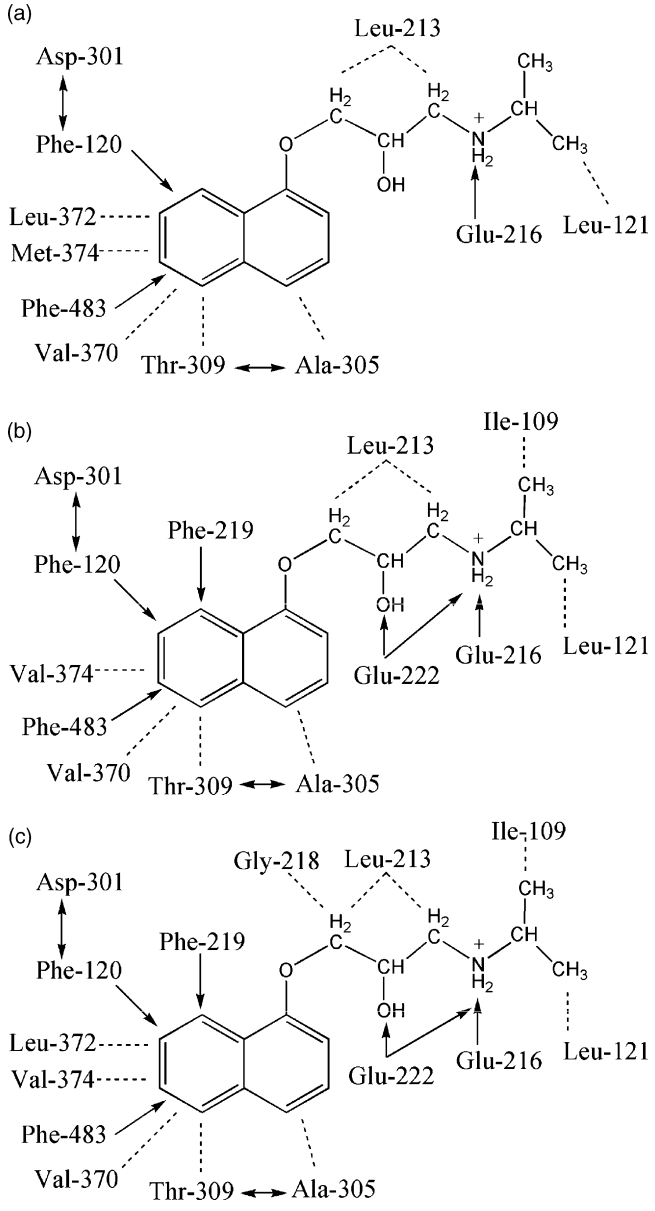

Fig. 4. Schematic representation of the contacts between propranolol and the substrate-binding residues in docking simulations by AutoDock for crystal structure (a), homology model (b), and for the final structure of MD simulation on the model (c). (→) Hydrogen bonds and electrostatic interactions; (- - -) hydrophobic interaction.

One of the differences between the results on quinidine docking by GOLD and AutoDock lies in a hydrogen bond between the hydroxyl group of quinidine and the carboxyl side-chain of Glu-216. Thus, the nitrogen of the quinuclidinyl group is predicted to orientate toward Glu-216 such that there is an electrostatic interaction in the AutoDock study on both the model and crystal structures, whereas a hydrogen bond between Glu-216 and the amino group of the substrate is observed in all of our docked structures except quinidine. Presumably, Glu-216 participates in the ligand binding electrostatically when the nitrogen in question is cationic, but through a hydrogen bond when the nitrogen is neutral.

In addition to Phe-120, Glu-216 and Asp-301, Phe-483 was reported to take part in the binding of quinidine and other CYP2D6 substrates by π–π stacking with the ligands [42]. Phe-483 seems to interact with them from the direction opposite to that of Phe-120. Although the substrates in the model structure have two π–π stacking interactions with Phe-120 and Phe-483, the crystal structure has only one interaction from Phe-120, and the side-chain of Phe-483 is located in a place unsuitable for substrate binding in the crystal structure. Moreover, in the AutoDock study, quinidine is positioned 9 Å away from the heme for catalytic turnover, even though it binds to the active site. The docking by GOLD also yielded the same position incompatible with metabolism. It is hence concluded that the competitive inhibitor, quinidine, seems to serve as such by binding in a place distant from the reaction center at the active site of CYP2D6.

3.2.4. Debrisoquine and dextromethorphan

Docking debrisoquine and dextromethorphan in the crystal structure did not give a conformation consistent with the experimental data (Table 1). These docked conformations in the crystal structure seem to be affected most by the position of Phe-483. Since Phe-483 has been reported as one of the important residues for substrate binding [45], Phe-483 in the crystal structure would need to come somewhat closer to the catalytic center to bind substrates, especially for these two compounds, debrisoquine and dextromethorphan.

3.2.5. Propranolol

Since we used propranolol for the MD simulation described below, its binding to CYP2D6 was explored in more depth

(Fig. 4a and b). When docked to either the model or crystal structure, the alkyl group of propranolol lies between SRS4 (I-helix) and SRS1 (B′-helix and B′–C loop) and the naphthyl group is surrounded by SRS3 (F-helix), SRS4 (I-helix), SRS5 (the loop after K-loop) and SRS6 (C-terminal loop). In the crystal structure docking, the alkyl group of propranolol is oriented slightly toward the I-helix, therefore the hydrophobic interaction between propranolol and Ile-109 (B′-helix) is lacking, unlike in the model docking result. Furthermore, Leu-372 and the mutated Met-374 may undergo hydrophobic interactions with the naphthyl group of propranolol in the crystal structure, while only Val-374 interacts with that portion in the model structure. As Ile-109 and Leu-372 are located in the SRS1 and SRS5 sequences, respectively, it is possible that they undergo hydrophobic interactions with propranolol or at least they are important residues at the substrate-binding cavity, regardless of their function.

Since the side-chain of Glu-222 and Phe-219 in the crystal structure is orientated outwards from the catalytic cavity, the hydrogen bond and π–π stacking interactions are also absent in the docked crystal structure. According to the multiple sequence alignments for SRS2 shown in the Supplementary material S1, Glu-216, as well as Phe-219 and Glu-222, is not conserved in other CYP2 family enzymes. Hence, they are likely to have evolved specifically in CYP2D6 for the binding of basic substrates. In fact, a kinetic analysis with bufuralol, whose conformation is quite similar to that of propranolol, reveals that the K_m of mutant E222A for bufuralol is increased markedly [46,47]. It is hence suggested that the carboxyl side-chain of Glu-222 is also essential for substrate binding as well as Glu-216, which is regarded as one of the most important residues for substrate binding in CYP2D6.

Another unique residue Phe-219 undergoes direct interactions with propranolol in the model structure, presumably because the F-helix containing this residue was not predicted properly. The majority of CYP2 subfamily enzymes have proline at position 219 instead of phenylalanine in CYP2D6. It is possible, therefore, that Phe-219 is specifically required in CYP2D6 for binding substrates containing a planar aromatic ring close to the site of reaction. Although Phe-219 does not interact with the CYP2D6 substrates directly in such a way as that observed for Phe-219 in the crystal structure, we believe that the side-chain of Phe-219 is involved in the control of electron transfer with another aromatic residue such as Phe-120 or Phe-483 for catalysis.

In conclusion, our docking dataset reveals a tendency or a difference in the mode of ligand binding between the model and crystal structures. Interestingly, the docking studies based on the model structure are as accurate as the crystal structure docking studies, even where the modelled sequence has only 42% homology with the template. However, since the crystal structure reported (2F9Q) does not contain a bound substrate, test dockings for the CYP3A4 crystal structure with (1W0F) and without substrate (1W0E) were carried out as a benchmark. It yielded nearly the same result as for the binding mode and energy for various substrates, as shown in Table S1 in the Supplementary material, suggesting that the docking result is not affected by whether the crystal structure contains a substrate or not.

It is generally thought that the crystal structure presents the best paradigm for delineating enzymatic reactions and functions. However, in P450 with broad substrate specificity its active site conformation alters flexibly and subtly for different substrates, and a single crystal structure such as 2F9Q alone is not sufficient for predicting the docked conformation of each substrate precisely. Some substrates, such as debrisoquine and dextromethorphan, can be fitted to the binding cavity more properly in the model than in the crystal structure. Although the docking study provides a lot of clues to elucidating the substrate-binding conformation which cannot be obtained from just the crystal structure alone, especially where there is no substrate bound, it is difficult to predict docking of all compounds precisely on the basis of a single snapshot structure.

3.3. MD simulations

To gain further insight into the substrate-binding mechanism in CYP2D6 time-dependent atomic motions were explored by molecular dynamics (MD) simulations. MD is a time-consuming exercise, so only propranolol was selected as a representative substrate for the MD simulation studies. The total energy, R.M.S.D., R.M.S.F., and the distance between the site of reaction on propranolol and the heme iron in a 1 ns MD simulation on the CYP2D6 model structure are presented in Fig. 5. The total energy decreases rapidly at first, and in 400 ps it goes up and down around -1.357×10^5 kcal/mol until 1 ns (Fig. 5a), when propranolol is bound at about 7 Å away from the heme (Fig. 5b). This distance of 7 Å is apparently too far for the reaction to take place between the two. This happens because the structural change covered by MD simulation represents a relatively small one over a relatively short time span (1 ns, for example). As a result, the final structure deduced heavily depends on the initial structure adopted. Nevertheless, analysis of time-dependent atomic motions provides an effective means to explore further the substrate-binding interaction, which is difficult to obtain from static structures alone, such as a crystal structure and that used in inflexible docking studies.

The R.M.S.D. for the whole structure rises to 2.6 Å at the end of the whole process, after 1 ns (Fig. 5c). As the R.S.M.D. values for the individual helices and loops indicate, the loops' flexibility is the dominant factor to dictate the whole R.M.S.D. In fact, large peaks in R.M.S.F. are observed for the loop portions, especially those located on the protein surface, as shown in Fig. 5d. The F–G loop fluctuated as much as 2.4 Å, and the H–I loop up to 3 Å. These loops in the model structure assume conformations different from those in the crystal structure even at the initial state. Probably, they are more flexible in the model than real, and they fluctuate in the Amber force field dynamically with various conformations interconverting readily. Incidentally, the prediction of loop conformations is one of the big issues in theoretical biology [45,48–59]. In contrast to the loops, the core part of the structure, especially the substrate-binding residues, barely fluctuate during the 1 ns

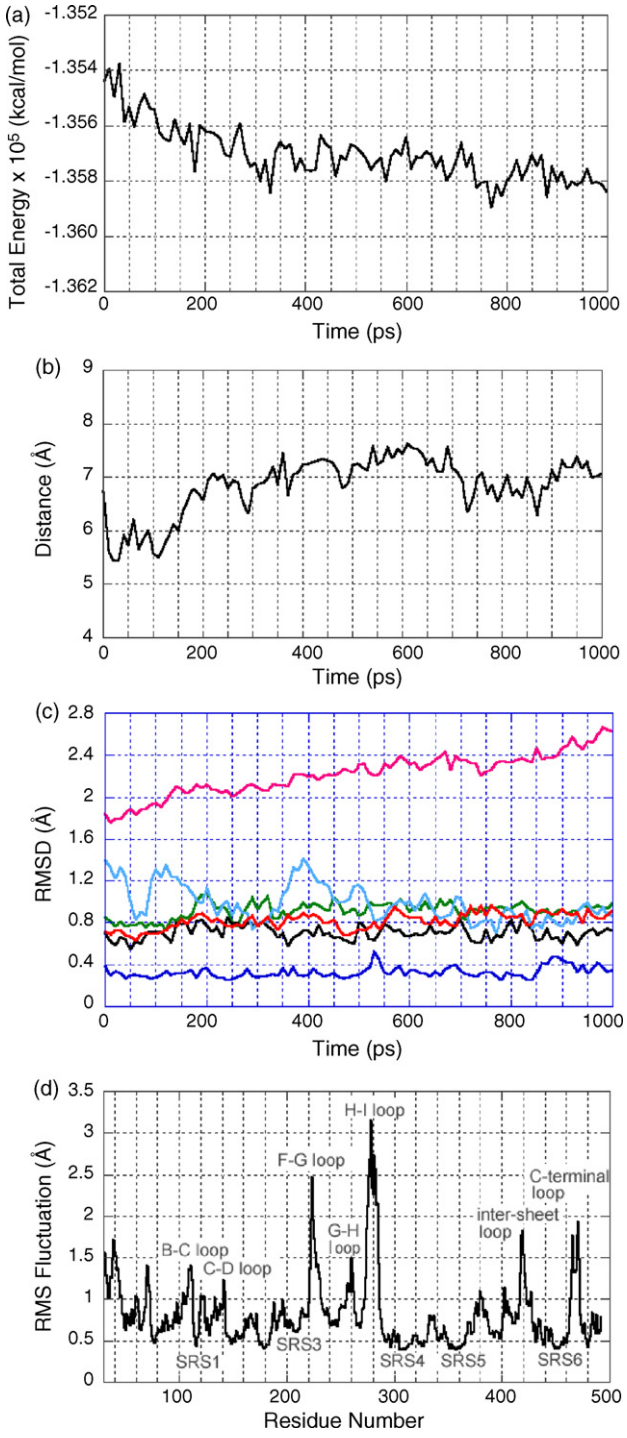

Fig. 5. (a) Total energy obtained by MD calculation for the CYP2D6 model with propranolol bound over 1 ns. (b) The distances from the heme iron to the hydrogen atom susceptible to metabolism. (c) R.M.S.D. for the whole structure (pink), B-helix (blue), C-helix (green), F-helix (black), G-helix (light blue) and I-helix (red). (d) RMS fluctuations (*B*-factors) for the backbone residues calculated over 1 ns.

MD simulation, as shown in R.M.S.D. for helices in Fig. 5c and for the R.M.S.F. in Table 2 and Fig. 5d. The residues changed considerably are only those involved in creating a cavity to accommodate the propranolol molecule. Eventually, the cavity volume became reduced from 1260 to 800 $Å^3$ at the end of the run.

Table 2
Distance and fluctuation values for the residues involved in the propranolol binding

Residue	Distances (Å)		R.M.S.F. (Å)
	Initial	Final	
Ile-109	4.412	3.825	1.356
Phe-120	4.369	3.740	0.427
Leu-121	5.270	4.127	0.531
Leu-213	5.540	3.715	0.606
Glu-216	1.696	2.734	0.777
Gly-218	7.500	3.790	0.883
Phe-219	5.031	4.233	0.838
Glu-222	1.936	2.606	0.789
Asp-301	6.126	8.363	0.559
Ala-305	3.430	4.967	0.547
Thr-309	3.481	4.914	0.573
Val-370	3.137	3.983	0.468
Leu-372	8.444	3.865	0.592
Val-374	3.321	6.041	0.708
Phe-483	4.363	3.550	0.535

The distance indicates the length between the residue in question and the heme group before and after MD simulation.

Differences in the substrate-binding residues between AutoDock and MD simulation studies, depicted in Fig. 4b and c, are compared by tracking the distances of those binding residues from propranolol over 1 ns (Table 2). Gly-218, Asp-301, Leu-372, and Val-374 change their location relative to propranolol by more than 2 Å during the MD simulation. The reorganization of the substrate-binding cavity makes Gly-218 and Leu-372 come closer to propranolol and Val-374 moves away from the substrate somewhat. Asp-301 interacts with the backbone of Phe-120 by a hydrogen bond and moves about 2 Å from propranolol, together with a conformational change of Phe-120. In other words, Asp-301 plays a role to immobilize Phe-120 by forming a hydrogen bond throughout the entire 1 ns simulation. The other residues do not undergo extensive conformational changes upon propranolol binding.

Phe-120, Glu-216, Asp-301 and Phe-483 in CYP2D6 have been regarded as the key residues for substrate binding in previous studies including our own docking study [46,47,51–59]. The results obtained above on the motion of substrate-binding residues by the MD simulation (Fig. 4c) lend credence to this notion. In addition to these important residues, the dynamics simulation on CYP2D6 with propranolol points to the roles of Phe-219 and Glu-222 more clearly, which are regarded as substrate-binding residues in the docking studies. Phe-219 is located 4.2 Å away from propranolol and, moreover, this residue is positioned 4.4 Å away from Phe-120. Since the location of α-carbons in this region is equivocal at the initial state, the direct interaction with propranolol might be artifactual. Nonetheless, as Phe-120, Phe-483 and the aromatic group of the substrate are clustered around Phe-219, it is reasonable to assume that Phe-219 serves as a conduit of electron flow for oxidation of substrates by cooperating with the other aromatic groups.

Another possible binding residue, Glu-222 makes a hydrogen bond with the amino group of propranolol at a

2.9 Å distance. This hydrogen bond is formed from the direction opposite to that of the Glu-216 hydrogen bond. Moreover, Glu-222 makes another hydrogen bond to the hydroxyl group of propranolol at a 2.6 Å distance, as shown in Fig. 4c. By reference to other experimental data on bufuralol binding, Glu-222 could be a binding residue in cooperation with Glu-216. However, the conformation of Glu-222 on the F-helix in the crystal structure, that is, it is pointing outwards from the cavity, and the supposition that F-helix is located in the substrate entrance, suggest that Glu-222 brings the substrate into the cavity by a hydrogen bond. In fact, Glu-222 is at the end of the F-helix and there is enough space for the Glu-222 side-chain to swing toward the center of the cavity from outside. Hence, we suppose that Glu-222 plays a dual role; it may guide the substrate to the catalytic center and place it above the heme. If valid, this mechanism and the function of Phe-219 and Glu-222 will apply not only to propranolol but also to other substrates.

To assess the validity of the results on propranolol, a control MD simulation was carried out on the same model without substrates. It was expected that a binding energy close to the experimental one is obtained from the difference in the relative energy between the models with and without substrate. It turned out, however, that this attempt was futile, as it is difficult to compare reliably these two MD simulations for such a small amount of energy. In the simulation, the total energy decreases significantly at the beginning and after that the average energy goes up and down around 25 kcal/mol in every 1 fs. Moreover, as the two systems contain a different number of solvent water molecules, which contribute to the total energy, a small energy difference associated with substrate binding is easily masked by it. The other data and the R.M.S.F. for the simulation in the absence of substrate point to the same result as the propranolol MD simulation; the surface loops are flexible and the core parts are rigid. Interestingly, the α-carbons of the SRS regions hardly fluctuate even in the enzyme in the absence of substrate, as shown in the Supplementary material S2. In summary, the SRS structure itself seems to remain fairly unchanged, though the size and shape of the active site cavity changes in adapting to various substrates.

4. Conclusions

Although it is true that X-ray crystallography and nuclear magnetic resonance spectroscopy are powerful techniques for elucidating the three-dimensional (3D) structures of proteins to provide clues for clarifying their function, they are usually time-consuming and laborious. If alternative or computational methods can offer equally useful information for proteins whose 3D structure is not available, it would be beneficial in developing novel drugs which can regulate the function of target proteins. This concept has been tested in this work with the P450 enzyme as a model to prove that a plausible structure can be deduced by homology modelling, and that a specific ligand can be docked reasonably accurately for the most part, as validated by binding energies and the sites of reaction on the substrate molecule. This, in turn, has provided useful information on the protein residues interacting with the substrate, allowing one to assess the likely function of these individual residues. Although the information obtained in this way needs to be verified experimentally, it provides a good starting point for delineating the function of proteins, P450 in particular. It would be helpful in the future if computational methods could provide guidelines as to what chemical manipulations of ligands will be useful to control protein function. In conclusion, substrate-binding prediction using docking studies and molecular dynamics may contribute to the identification of metabolic routes for drugs in development and thus aid in the design of new drugs.

Acknowledgements

Yuko Ito would like to thank KIT and the Japan Foundation for the award of a visiting scientist scholarship as part of a PhD programme. David Lewis acknowledges the financial support of ExxonMobil, the Daiwa Anglo-Japanese Foundation and the British Technology Group.

Appendix A. Supplementary data

A part of the multiple sequence alignment of human CYP2 family; compared R.M.S.F. data between propranolol and non-substrate in MD simulation; and a benchmark dockings for CYP3A4 crystal structure with substrate (1W0F) and non-substrate (1W0E). These materials are available free of charge via the Internet at doi:10.1016/j.jmgm.2007.07.004. Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2007.07.004.

References

[1] D.F.V. Lewis, Guide to Cytochromes P450 Structure and Function, Taylor & Francis, London, 2001.
[2] T. Omura, Y. Ishimura, Y. Fujii, Cytochromes P450, Kodansha, Tokyo, 2003.
[3] D.F.V. Lewis, 57 varieties: the human cytochromes P450, Pharmacogenomics 5 (2004) 305–318.
[4] O. Gotoh, Substrate recognition sites in cytochrome P450 family 2 (CYP2) protein inferred from comparative analyses of amino acid and coding nucleotide sequences, J. Biol. Chem. 267 (1992) 83–90.
[5] A.K. Daly, Pharmacogenetics of the cytochromes P450, Curr. Top. Med. Chem. 4 (2004) 1733–1744.
[6] M.J. de Groot, S.B. Kirton, M.J. Sutcliffe, In silico methods for predicting ligand binding determinants of cytochromes P450, Curr. Top. Med. Chem. 4 (2004) 1803–1824.
[7] D.F.V. Lewis, Essential requirements for substrate binding affinity and selectivity toward human CYP2 family enzymes, Arch. Biochem. Biophys. 409 (2003) 32–44.
[8] R.J. Bertz, G.R. Granneman, Use of in vitro and in vivo data to estimate the likelihood of metabolic pharmacokinetic interactions, Clin. Pharmacokinet. 32 (1997) 210–258.
[9] M.F. Fromm, H.K. Kroemer, M. Eichelbaumm, Impact of P450 genetic polymorphism on the first-pass extraction of cardiovascular and neuroactive drugs, Adv. Drug Deliv. Rev. 27 (1997) 171–199.
[10] L. Bertilsson, M.L. Dahl, Polymorphic drug oxidation: relevance to the treatment of psychiatric disorders, CNS Drugs 5 (1996) 200–223.
[11] U.A. Meyer, The molecular basis of genetic polymorphisms of drug metabolism, J. Pharm. Pharmacol. 46 (1994) 409–415.

[12] G. Alván, P. Bechtel, L. Iselius, U. Gundert-Remy, Hydroxylation polymorphisms of debrisoquine and mephenytoin in European populations, Eur. J. Clin. Pharmacol. 39 (1990) 533–537.
[13] A.K. Daly, J. Brockmöller, F. Broly, M. Eichelbaum, W.E. Evans, F.J. Gonzalez, J.D. Huang, J.R. Idle, M. Ingelman-Sundberg, T. Ishizaki, E. Jacqz-Aigrain, U.A. Meyer, D.W. Nebert, V.M. Steen, C.R. Wolf, U.M. Zanger, Nomenclature for human CYP2D6 alleles, Pharmacogenetics 6 (1996) 193–201.
[14] R. Tyndale, T. Aoyama, F. Broly, T. Matsunaga, T. Inaba, W. Kalow, H.V. Gelboin, U.A. Meyer, F.J. Gonzalez, Identification of a new variant CYP2D6 allele lacking the codon encoding Lys-281: possible association with the poor metabolizer phenotype, Pharmacogenetics 1 (1991) 26–32.
[15] I. Johansson, M. Oscarson, Q.Y. Yue, L. Bertilsson, F. Sjöqvist, M. Ingelman-Sundberg, Genetic analysis of the Chinese cytochrome P4502D locus: characterization of variant CYP2D6 genes present in subjects with diminished capacity for debrisoquine hydroxylation, Mol. Pharmacol. 46 (1994) 452–459.
[16] M.R. Wester, E.F. Johnson, C. Marques-Soares, P.M. Dansette, D. Mansuy, C.D. Stout, Structure of a substrate complex of mammalian cytochrome P450 2C5 at 2.3 Å resolution: evidence for multiple substrate binding modes, Biochemistry 42 (2003) 6370–6379.
[17] M.R. Wester, E.F. Johnson, C. Marques-Soares, S. Dijols, P.M. Dansette, D. Mansuy, C.D. Stout, Structure of mammalian cytochrome P450 2C5 complexed with diclofenac at 2.1 Å resolution; evidence for an induced fit model of substrate binding, Biochemistry 42 (2003) 9335–9345.
[18] P. Rowland, F.E. Blaney, M.G. Smyth, J.J. Jones, V.R. Leydon, A.K. Oxbrow, C.J. Lewis, M.G. Tennant, S. Modi, D.S. Eggleston, R.J. Chenery, A.M. Bridges, Crystal structure of human cytochrome P450 2D6, J. Biol. Chem. 281 (2006) 7614–7622.
[19] D.F.V. Lewis, Homology modelling of human CYP2 family enzymes based on the CYP2C5 crystal structure, Xenobiotica 32 (2002) 305–323.
[20] D.F.V. Lewis, Modelling human cytochromes P450 involved in drug metabolism from the CYP2C5 crystallographic template, J. Inorg. Biochem. 91 (2002) 502–514.
[21] D.F.V. Lewis, M. Dickins, B.G. Lake, P.S. Goldfarb, Investigation of enzyme selectivity in the human CYP2C subfamily: homology modelling of CYP2C8, CYP2C9 and CYP2C19 from the CYP2C5 crystallographic template, Drug Metab. Drug Interact. 19 (2003) 257–285.
[22] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function, J. Comput. Chem. 19 (1998) 1639–1662.
[23] J.J.P. Stewart, MOPA.C. A semiempirical molecular orbital program, J. Comput. Aid. Mol. Des. 4 (1990) 1–105.
[24] H.R. Mashhadi, H.M. Shanechi, C. Lucas, A new genetic algorithm with Lamarckian individual learning for generation scheduling, IEEE Trans. Power Syst. 18 (2003) 1181–1186.
[25] D.A. Case, D.A. Pearlman, J.W. Caldwell, T.E. Cheatham III, J. Wang, W.S. Ross, C. Simmerling, T. Darden, K.M. Merz, R.V. Stanton, A. Cheng, J.J. Vincent, M. Crowley, V. Tsui, H. Gohlke, R. Redmer, Y. Duan, J. Pitera, I. Massova, G.L. Seibel, U.C. Singh, P. Weiner, P.A. Kollman, AMBER7, University of California, San Francisco, CA, 2002.
[26] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.R. Ross, T.E. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel, P. Kollman, AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules, Comp. Phys. Commun. 91 (1995) 1–41.
[27] V. Helms, R.C. Wade, Thermodynamics of water mediating protein–ligand interactions in cytochrome P450cam: a molecular dynamics study, Biophys. J. 69 (1995) 810–824.
[28] D. Vitkup, E. Melamud, J. Moult, C. Sander, Completeness in structural genomics, Nat. Struct. Biol. 8 (2001) 559–566.
[29] D. Baker, A. Sali, Protein structure prediction and structural genomics, Science 294 (2001) 93–96.
[30] G.J. Kleywegt, T.A. Jones, Detection, delineation, measurement and display of cavities in macromolecular structures, Acta Crystallogr. D 50 (1994) 178–185.
[31] S.W. Ellis, K. Rowland, M.J. Ackland, E. Rekka, A.P. Simula, M.S. Lennard, C.R. Wolf, G.T. Tucker, Influence of amino acid residue 374 of cytochrome *P*-450 2D6 (CYP2D6) on the regio- and enantio-selective metabolism of metoprolol, Biochem. J. 316 (1996) 647–654.
[32] R.S. Obach, A.E. Reed-Hagen, Measurement of Michaelis constants for cytochrome P450-mediated biotransformation reactions using a substrate depletion approach, Drug Metab. Dispos. 30 (2002) 831–837.
[33] D.A. Smith, B.C. Jones, Speculations on the substrate structure–activity relationship (SSAR) of cytochrome P450 enzymes, Biochem. Pharmacol. 44 (1992) 2089–2098.
[34] X.-Q. Li, T.B. Andersson, M. Ahlstrom, L. Weidolf, Comparison of inhibitory effects of the proton pump-inhibiting drugs omeprazole, esomeprazole, lansoprazole, pantoprazole, and rabeprazole on human cytochrome P450 activities, Drug Metab. Dispos. 32 (2004) 821–827.
[35] R.L. Walsky, R.S. Obach, Validated assays for human cytochrome P450 activities, Drug Metab. Dispos. 32 (2004) 647–660.
[36] D.F.V. Lewis, S. Modi, M. Dickins, Quantitative structure–activity relationships (QSARs) within substrates of human cytochromes P450 involved in drug metabolism, Drug Metab. Drug Interact. 18 (2001) 221–242.
[37] D.F.V. Lewis, S. Modi, M. Dickins, Structure–activity relationship for human cytochrome. P450 substrates and inhibitors, Drug Metab. Rev. 34 (2002) 69–82.
[38] S. Modi, M.J. Paine, M.J. Sutcliffe, L.Y. Lian, W.U. Primrose, C.R. Wolf, G.C. Roberts, A model for human cytochrome P450 2D6 based on homology modeling and NMR studies of substrate binding, Biochemistry 35 (1996) 4540–4550.
[39] P. Dayer, J. Desmeules, R. Striberni, In vitro forecasting of drugs that may interfere with codeine bioactivation, Eur. J. Drug Metab. Pharmacokinet. 17 (1992) 115–120.
[40] M. Oscarson, M. Hidestrand, I. Johansson, M. Ingelman-Sundberg, A Combination of mutations in the CYP2D6*17 (CYP2D6Z) allele causes alterations in enzyme function, Mol. Pharmacol. 52 (1997) 1034–1040.
[41] T. Inaba, M. Jurima, W.A. Mahon, W. Kalow, In vitro inhibition studies of two isozymes of human liver cytochrome P-450. Mephenytoin *p*-hydroxylase and sparteine monooxygenase, Drug Metab. Dispos. 13 (1985) 443–448.
[42] L.A. McLaughlin, M.J.I. Paine, C.A. Kemp, J.D. Maréchal, J.U. Flanagan, C.J. Ward, M.J. Sutcliffe, G.C.K. Roberts, C.R. Wolf, Why is quinidine an inhibitor of cytochrome P450 2D6? The role of key active site residues in quinidine binding, J. Biol. Chem. 280 (2005) 38617–38624.
[43] G. Jones, P. Willett, R.C. Glen, Molecular recognition of receptor sites using a fenetic algorithm with a description of desolvation, J. Mol. Biol. 245 (1995) 43–53.
[44] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, J. Mol. Biol. 267 (1997) 727–748.
[45] G. Smith, S. Modi, I. Pillai, L.Y. Lian, M.J. Sutcliffe, M.P. Pritchard, T. Friedberg, G.C. Roberts, C.R. Wolf, Determinants of the substrate specificity of human cytochrome P-450 CYP2D6: design and construction of a mutant with testosterone hydroxylase activity, Biochem. J. 331 (1998) 783–792.
[46] K. Masuda, K. Tamagake, T. Katsu, F. Torigoe, K. Saito, N. Hanioka, S. Yamano, S. Yamamoto, S. Narimatsu, Roles of phenylalanine at position 120 and glutamic acid at position 222 in the oxidation of chiral substrates by cytochrome P450 2D6, Chirality 18 (2006) 167–176.
[47] K. Masuda, K. Tamagake, Y. Okuda, F. Torigoe, D. Tsuzuki, T. Isobe, H. Hichiya, N. Hanioka, S. Yamamoto, S. Narimatsu, Change in enantioselectivity in bufuralol 1′-hydroxylation by the substitution of phenylalanine-120 by alanine in cytochrome P450 2D6, Chirality 17 (2005) 37–43.
[48] J. Wojcik, J.P. Mornon, J. Chomilier, New efficient statistical sequence-dependent structure prediction of short to medium-sized protein loops based on an exhaustive loop classification, J. Mol. Biol. 289 (1999) 1469–1490.
[49] J.M. Kwasigroc, J. Chomilier, J.P. Mornon, A global taxonomy of loops in globular proteins, J. Mol. Biol. 259 (1996) 855–872.

[50] R. Sanchez, A. Sali, Advances in comparative protein–structure modeling, Curr. Opin. Struct. Biol. 7 (1997) 206–214.
[51] B.M.A. Lussenburg, P.H.J. Keizers, C. de Graaf, M. Hidestrand, M. Ingelman-Sundberg, N.P.E. Vermeulen, J.N.M. Commandeur, The role of phenylalanine 483 in cytochrome P450 2D6 is strongly substrate dependent, Biochem. Pharmacol. 70 (2005) 1253–1261.
[52] M.J.I. Paine, L.A. McLaughlin, J.U. Flanagan, C.A. Kemp, M.J. Sutcliffe, G.C.K. Roberts, C.R. Wolf, Residues glutamate 216 and aspartate 301 are key determinants of substrate specificity and product regioselectivity in cytochrome P450 2D6, J. Biol. Chem. 278 (2003) 4021–4027.
[53] G.P. Hayhurst, J. Harlow, J. Chowdry, E. Gross, E. Hilton, M.S. Lennard, G.T. Tucker, S.W. Ellis, Influence of phenylalanine-481 substitutions on the catalytic activity of cytochrome P450 2D6, Biochem. J. 355 (2001) 373–379.
[54] F.P. Guengerich, I.H. Hanna, M.V. Martin, E.M. Gillam, Role of glutamic acid 216 in cytochrome P450 2D6 substrate binding and catalysis, Biochemistry 42 (2003) 1245–1253.
[55] S.W. Ellis, G.P. Hayhurst, G. Smith, T. Lightfood, M.M.S. Wong, A.P. Simula, M.J. Ackland, M.J.E. Sternberg, M.S. Lennard, G.T. Tucker, C.R. Wolf, Evidence that aspartic acid 301 is a critical substrate-contact residue in the active site of cytochrome P450 2D6, J. Biol. Chem. 270 (1995) 29055–29058.
[56] S.W. Ellis, G.P. Hayhurst, T. Lightfoot, G. Smith, J. Harlow, K. Rowland-Yeo, C. Larsson, J. Mahling, C.K. Lim, C.R. Wolf, M.G. Blackburn, M.S. Lennard, G.T. Tucker, Evidence that serine 304 is not a key ligand-binding residue in the active site of cytochrome P450 2D6, Biochem. J. 345 (2000) 565–571.
[57] J.U. Flanagan, J.D. Marechal, R. Ward, C.A. Kemp, L.A. McLaughlin, M.J. Sutcliffe, G.C. Roberts, M.J. Paine, C.R. Wolf, Phe120 contributes to the regiospecificity of cytochrome P450 2D6: mutation leads to the formation of a novel dextromethorphan metabolite, Biochem. J. 380 (2004) 353–360.
[58] F.P. Guengerich, G.P. Miller, I.H. Hanna, M.V. Martin, S. Leger, C. Black, N. Chauret, J.M. Silva, L.A. Trimble, J.A. Yergey, D.A. Nicoll-Griffith, Diversity in the oxidation of substrates by cytochrome P450 2D6: lack of an obligatory role of aspartate 301-substrate electrostatic bonding, Biochemistry 41 (2002) 11025–11034.
[59] I.H. Hanna, M.S. Kim, F.P. Guengerich, Heterologous expression of cytochrome P450 2D6 mutants, electron transfer, and catalysis of bufuralol hydroxylation: the role of aspartate 301 in structural integrity, Arch. Biochem. Biophys. 393 (2001) 255–261.