Journal of Computer-Aided Molecular Design, **14:** 93–116, 2000.
KLUWER/ESCOM

Theoretical investigation of substrate specificity for cytochromes P450 IA2, P450 IID6 and P450 IIIA4

Francesca De Rienzo, Francesca Fanelli, M. Cristina Menziani & Pier G. De Benedetti*
Dipartimento di Chimica, Università di Modena e Reggio Emilia, Via Campi 183, I-41100 Modena, Italy

Received 29 December 1998; Accepted 11 May 1999

Key words: comparative modeling, cytochromes P450, P450 IA2, P450 IID6, P450 IIIA4, structure-specificity relationships, substrate docking

Summary

Three-dimensional models of the cytochromes P450 IA2, P450 IID6 and P450 IIIA4 were built by means of comparative modeling using the X-ray crystallographic structures of P450 CAM, P450 BM-3, P450 TERP and P450 ERYF as templates. The three cytochromes were analyzed both in their intrinsic structural features and in their interaction properties with fifty specific and non-specific substrates. Substrate/enzyme complexes were obtained by means of both automated rigid and flexible body docking. The comparative analysis of the three cytochromes and the selected substrates, in their free and bound forms, allowed for the building of semi-quantitative models of substrate specificity based on both molecular and intermolecular interaction descriptors. The results of this study provide new insights into the molecular determinants of substrate specificity for the three different eukaryotic P450 isozymes and constitute a useful tool for predicting the specificity of new compounds.

Abbreviations: 1A2, human cytochrome P450 IA2; 2D6, human cytochrome P450 IID6; 3A4, human cytochrome P450 IIIA4; P450s, cytochromes P450.

Introduction

The superfamily of P450 cytochromes provides one of Nature's most common and sophisticated tools to oxidize endogenous compounds in prokaryotic and eukaryotic organisms. They are also involved in the transformation and the metabolic excretion of xenobiotics, such as drugs or environmental and food contaminants, even activating some of them to toxins or carcinogens [1]. In particular, among a dozen or more P450 gene families, only the cytochromes P450 I, II, III and IV families either metabolize exogenous chemicals or are induced by them [2]. Three isozymes of these families, the cytochromes P450 IA2 (1A2), P450 IID6 (2D6) and P450 IIIA4 (3A4), that are bound to the human endoplasmic reticulum (ER) membrane are, besides their implication in the detoxification or activation of chemicals, of particular interest also because they can be involved in some adverse interactions [3].

Despite the large amount of information on the functional role of these enzymes, the knowledge of their three-dimensional (3D) structure is still lacking. To date, only four bacterial P450 structures have been solved by X-ray crystallography: P450 CAM (or P450 101), a camphor hydroxylase from *Pseudomonas putida* [4–6]; P450 BM-3 (or P450 102), a fatty acid monooxygenase from *Bacillus megaterium* [7]; P450 TERP (or P450 108), an α-terpinol monooxygenase from *Pseudomonas sp.* [8] and P450 ERYF (or P450 107A1), a 6-deoxyerythronalide B hydroxylase from *Saccaropolyspora erythrea* [9]. Comparisons of known structures indicate that bacterial P450s share a common fold, despite their low sequence similarities (20%). Moreover, although these bacterial enzymes display low (15–25%) amino acid sequence identity with eukaryotic P450s, certain regions of primary sequence, such as the heme and oxygen binding sites

*To whom correspondence should be addressed. E-mail: deben@unino.it.

as well as the site of interaction with redox partners, are quite conserved among different species, thus supporting the hypothesis of a general conservation of the tertiary structure. On the other hand, sequence comparisons predict that some of the structural peculiarities of eukaryotic P450s should reside in their membrane-bound portions [10–12]. In the last years, significant progress has been made in the understanding of the membrane topology of mammalian cytochromes, in particular of those belonging to ER. The most supported hypotheses predict these cytochromes as integral membrane proteins, anchored to the ER membrane by one or two transmembrane segments located at the N-terminus and probably also by some additional peripheral contacts [13–16]. Recently, molecular models of various mammalian isozymes were built by taking advantage of the structural homology among P450 cytochromes [17–29].

The theoretical investigation presented herein is focused on three particular P450 isozymes: the human microsomal cytochromes 1A2, 2D6 and 3A4.

Many known carcinogens are specific for cytochromes P450 I. In particular, polyaromatic and heteroaromatic hydrocarbons and their amino derivatives either induce these P450 family enzymes and/or are metabolized by them yielding reactive intermediates such as epoxides, diol epoxides, carbonium ions or nitrenium ions, which can interact with DNA, forming covalent adducts that result in miscoding, mutagenesis and carcinogenesis [2]. 1A2 substrates are neutral or basic, lipophilic, planar molecules with at least one putative hydrogen bond donating site [2, 30].

Differently from the substrates and the inducers of the P450 I family, those of the P450 II family generally have low toxicity [2]. Substrates of 2D6 include a set of lipophilic aryl-alkyl amines that possess the basic nitrogen atom at a distance close to 5 or 7 Å from the oxidation site [30–33].

Finally, cytochromes belonging to the P450 III family, such as 3A4, play a part in the metabolism of several steroid hormones. However, in most cases, this family mediates the detoxification of polyaromatic hydrocarbons and of a wide range of drugs covering all fields of pharmaceuticals [2]. 3A4 metabolizes a wide variety of compounds which are lipophilic, neutral or basic molecules with the site of oxidation often constituted by nitrogen atoms or allylic positions [30, 31].

Given the extremely important metabolic role of 1A2, 2D6 and 3A4, the understanding of the molecular determinants of substrate specificity, in connection with the structural and topological differentiation of cytochromes, is one of the major goals of current research in this field.

This work presents a theoretical investigation of the substrate specificity towards the 1A2, 2D6 and 3A4 isozymes. The semi-quantitative models obtained can be considered useful tools for predicting the specificity of new compounds.

Methods

Building and refinement of the 3D models

Sequences of human microsomal cytochromes 1A2, 2D6 and 3A4 were extracted from the SWISS-PROT protein sequences database. The 3D models of the three isozymes were built by means of restraint-based comparative modeling (Modeler [34]), using the X-ray crystallographic structures of bacterial cytochromes P450 BM-3, CAM, TERP and ERYF (PDB [35] entries: 2bmh, 3cpp, 1cpt and 1oxa, respectively) as templates. Sequence identity percentages between each of the target P450s and the templates P450 CAM, BM-3, TERP and ERYF are close to 20%.

Secondary structure predictions were obtained with the Rost and Sander method [36] at the Internet address PredictProtein@Embl-Heidelberg.de.

To the models of 1A2, 2D6 and 3A4 finally obtained, polar hydrogens were then added. Moreover, the heme molecule, with the heme iron in its ferric oxidation state, was extracted from the P450 BM-3 structure and fitted into the active site of each of the three cytochrome models.

The models were energy refined. Minimizations were carried out using 1500 steps of steepest descent followed by 5000 steps of conjugate gradient. A distance dependent dielectric term ($\epsilon = 4r$) and a 12 Å non-bonded cutoff distance were chosen. The CHARMm (Molecular Simulation Inc., Waltham, MA) united atom and all atom parametrizations were used for the protein and the heme molecule, respectively. The explicit hydrogen-bond term was used in the CHARMm energy calculation [37]. The protein backbone was fixed during the minimization procedure.

Docking procedures

The structures of most of the substrates were obtained from the Cambridge Structural Database (CSD) [38], while the structures of **11**, **12**, **13**, **15**, **34**, **45** and **47**

(Scheme 1) were built by slightly modifying homologous CSD structures, whereas those of **3**, **4**, **9**, **16**, **17**, **20** and **21** (Scheme 1) were built ex novo within the program QUANTA (release 96; Molecular Simulation Inc.).

The selection of the specific and non-specific substrates considered in this study has been driven by the experimental data available from the literature [2, 32].

Substrates were considered in their most populated prototropic form at physiological pH (pH $\cong$ 7.4), according to their pKa values (Scheme 1). These structures were optimized by means of AM1 [39] semi-empirical calculations. Molecular volumes and solvent accessible surfaces were computed within the QUANTA 96 package.

Substrates were docked into the active site of their metabolizing enzyme by using the automatic rigid body docking program DOCK 3.5 [40, 41]. The AMBER united-atom charges [42] and the Gasteiger and Marsili atomic charges [42] were used for the proteins and ligands, respectively. The AMBER-based van der Waals parameters were employed [42].

For each of the DOCK determined sphere clusters that fill the active sites of the three cytochromes, a subset of spheres close to the heme molecule was defined as the critical cluster. Chemical matching was additionally employed for the substrate/2D6 docking in order to favor the approaching of the positively charged nitrogen atom of the substrate and D301 of the enzyme. Contact scoring was employed for substrate/1A2 and substrate/3A4 docking, whereas the contact plus force field scoring was employed for the substrate/2D6 docking, performing the force field score optimization.

The most flexible substrates were also docked as flexible bodies by means of the upgraded version of the program (DOCK 4.0 [44]). Different conformational search options were tested. The anchor first search [45] was finally employed, setting the peripheral search and torsion drive. Minimization of torsions, reminimization of both anchor and ligand as well as optimization of the energy score were set.

The results of each docking were analyzed in order to select, among the many possible substrate orientations and conformations, the best one. The selection of the best docking mode was based, at least in part, on the DOCK scores and on the criteria that the substrate should direct its metabolic site towards the heme, approaching the free heme pyrrole A and a highly conserved Thr (corresponding to Thr 252 in P450 CAM) in HI. In fact, this threonine is supposed to be important for the oxidation reaction, as deduced from the analysis of X-ray solved complexes [4, 5, 46, 47]. Moreover, experimental data from Nuclear Magnetic Resonance (NMR) and site-directed mutagenesis were used to drive the filtering of the right orientations [19, 20].

The best orientation of each substrate in the active site of its metabolizing isozyme provided by automatic docking was manually modified, where necessary, to achieve consistency with the experimental data [4–6, 19, 20, 48–58].

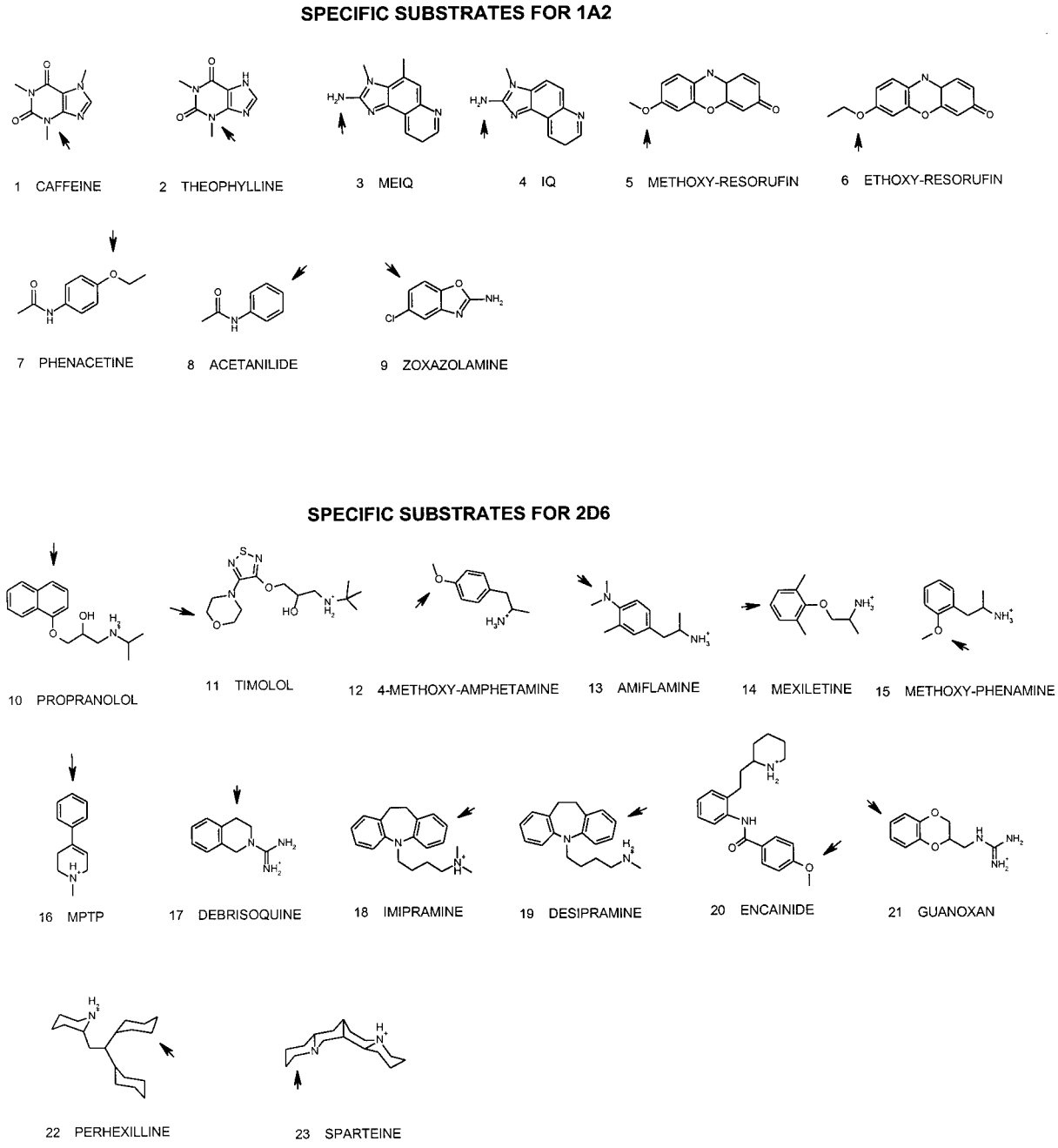
The complexes were energy refined employing the same minimization procedure above described, but allowing also the protein backbone to relax. The CHARMm all atom parametrization (Molecular Simulation Inc.) and the AM1 [39] charge distribution were used for the substrates. The interaction energies (*IE*) of the ligand/enzyme minimized complexes were computed according to the following formula: $IE = E_{\mathrm{COMPLEX}} - E_{\mathrm{ENZ}} - E_{\mathrm{LIG}}$, where E_{COMPLEX} is the energy of the ligand/enzyme minimized complex, E_{REC} and E_{LIG} are the energies of the enzyme and of the ligand, respectively, in the minimized complex. It is worth noting that the decomposition of the total interaction energy in the electrostatic, van der Waals and H-bonding contributions is an artifact made according to the different non bonding terms that constitute the CHARMm potential energy function [37].

Results and discussion

So far, quantitative theoretical investigations of substrate specificity towards cytochromes P450 were mainly carried out by means of substrate-based approaches [2, 26].

Our aim was to build a semi-quantitative theoretical model of substrate specificity using intermolecular interaction descriptors computed on the substrate/cytochrome complexes, focusing on 1A2, 2D6 and 3A4 isozymes.

The results of this study are presented in terms of comparative analyses performed at three different levels. Firstly, the analysis of the 3D structures of the three isozymes as well as of fifty substrates in their unbound forms was instrumental for identifying the intrinsic structural features responsible for their potential specificity. Secondly, the molecular determinants of substrate specificity were inferred from the analysis of the substrate/enzyme interaction



Scheme 1. Structures of specific substrates for the cytochromes 1A2, 2D6 and 3A4, as well as non-specific substrates.

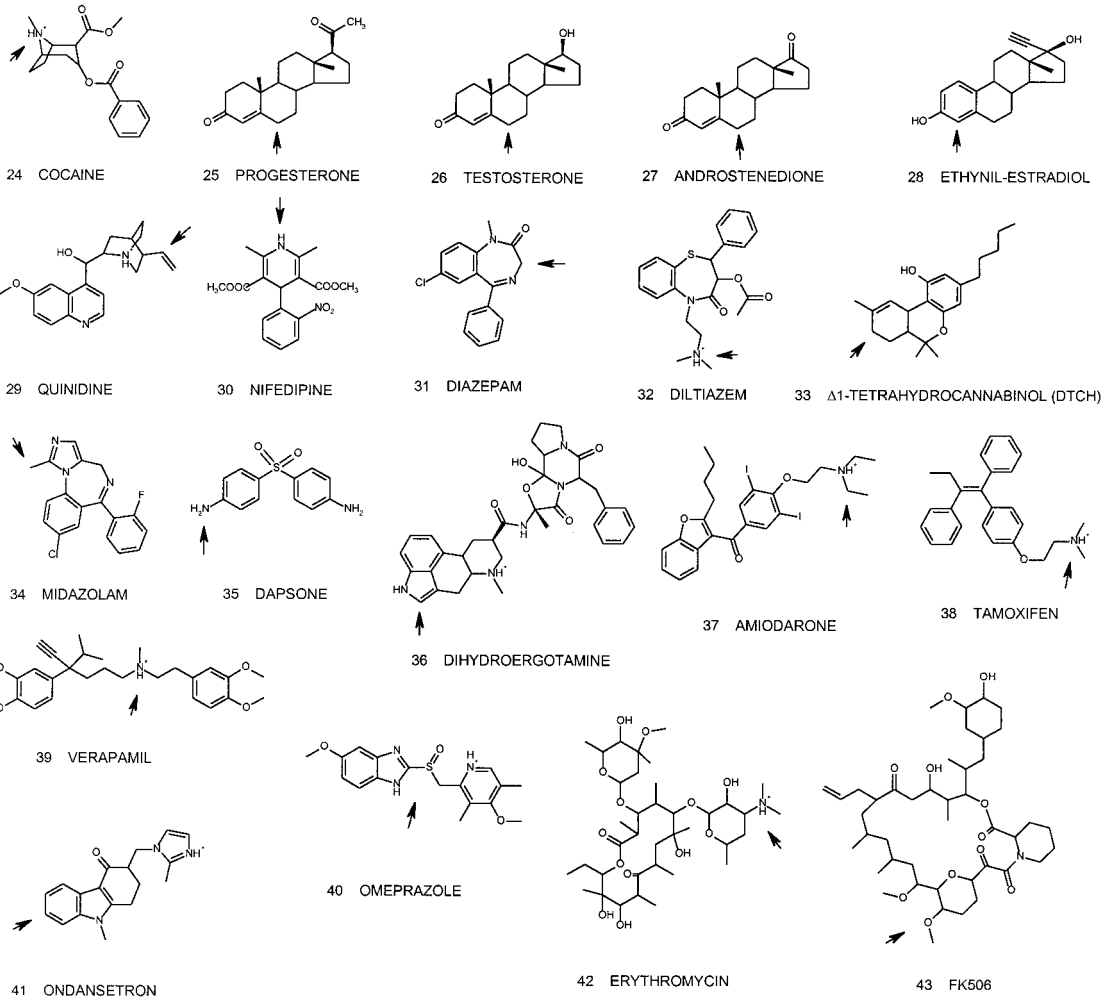
models.[1] Finally, theoretical descriptors computed on the substrate/protein complexes allowed the building of semi-quantitative models useful for predicting the specificity of new compounds.

[1]The coordinates of the models are available from the corresponding author. E-mail: deben@unimo.it

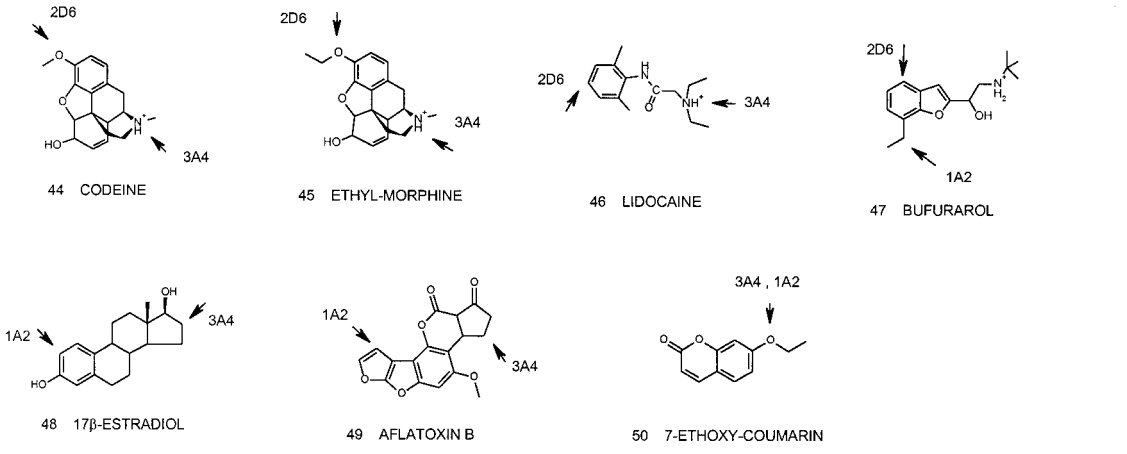
Comparative modeling of 1A2, 2D6 and 3A4

The 3D models of 1A2, 2D6 and 3A4 were built by means of restraint-based comparative modeling (Modeler [34]), using the X-ray crystallographic structures of bacterial cytochromes P450 BM-3, CAM, TERP and ERYF as templates.

SPECIFIC SUBSTRATES FOR 3A4



NON-SPECIFIC SUBSTRATES



Scheme 1. (continued).

Among the different alignments probed, including sequence-sequence and sequence-structure alignments, that proposed by Chang et al. [24] was finally employed (Figures 1–3). In fact, this alignment turned out to be quite similar, particularly in the structurally conserved regions (SCRs), to the sequence-structure alignment obtained by Modeler, but showing a lower number of insertions and deletions in the non-conserved regions.

Slight modifications of the alignment by Chang et al. were done in order to improve the quality of the final models obtained and to gain consistency with site-directed mutagenesis and NMR data [19, 20, 25–27, 48–58]. For the same reason, short stretches of amino acids in the structures of the templates (they are indicated by bold lowercase letters in the alignment) were deleted.

Protein regions are labeled according to the nomenclature by Hasemann et al. [8]. The first letter (H) in each capital letter pair stands for α-helix and the first letter (b) in each lowercase letter/number pair stands for β-strand.

The N-terminal stretches of amino acids ATELLLASAIFCLVFWVLKGLRPR (10–33) in 1A2, LVPLAVIVAIFLLLVDLMHRRQ (6–27) in 2D6 and AMETWLLLAVSLVLLYLYGTHS (7–28) in 3A4 were constrained in α-helical conformation during the comparative modeling, according to secondary structure predictions and predictions of the topology of membrane proteins. In the final models, the main axis of the N-terminal helix is nearly perpendicular to the heme plane. Similarly, α-helical restraints were imposed on the inserted segments FPILR (238–242) of 1A2, REVLNAVPVLL (221–231) of 2D6 and ITVFPFLIP (222–230) of 3A4, belonging to the HF/HG loop (Figures 1–3). For each cytochrome, thirty models were obtained by optimizing the molecular probability density function (pdf), starting from different initial conformations, and by setting the Cartesian coordinates of the atoms to random values between ±4 Å (random seed = −12312).

Finally, from each of the three sets of models, the best model showing the lowest violation of stereochemical restraints was selected. These models passed all criteria implemented in PROCHECK [59]. These criteria are Ramachandran plot, peptide bonds planarity, Cα tetrahedral distortion, non-bonded interactions, hydrogen bonding energies, and closeness of the side chain dihedral angles to ideal values. We also used the overall geometric factor (G-factor) provided by PROCHECK that is a measure of the overall 'normality' of the structure. It is obtained from an average of all the different G-factors of each residue in the structure. An overall G-factor below −1 may indicate properties that need to be investigated more closely.

The overall fold and side chain packing of the models passed also the test provided by the quality check performed by means of the program WHAT IF [60]. The quality check option in WHAT IF calculates an index that is a measure of the agreement between the distribution of the atoms around each residue fragment in the model and equivalent distributions derived from the database of known structures. In general, an average quality score below −5 indicates poorly packed structures. A score between −2.0 and −1.0 indicates proteins probably threaded correctly. The average quality of 200 highly refined structures was −0.5–−0.4.

The WHAT IF and PROCHECK scores are −1.278 and −0.15 for 1A2, −1.215 and −0.07 for 2D6, −1.195 and −0.10 for 3A4. These results are good in comparison to the same scores calculated on the template structures: −0.518 and −0.25 for 3cpp, −0.739 and 0.30 for 2bmh, −0.775 and −0.18 for 1cpt, −0.554 and 0.36 for 1oxa.

The selected models of the three cytochromes also fulfill the Protein Health analysis implemented in the molecular graphics package QUANTA. The worst modeled regions included the portion between the N-terminus and HA′, HE, HE/HF and HH/HI loops, the region between the meander and the Cys-pocket, and the C-terminus. 3D-Profile analysis implemented in QUANTA is not significant, due to the fact that lipids should constitute the environment of one or more portions of these membrane-bound proteins. In fact, embedding the N-terminal α-helix onto a phospholipid bilayer increased the 3D-Profile score.

Comparative analysis of the cytochrome models

Figures 4–6 show the refined models of 1A2, 2D6 and 3A4.

The 3D models of 1A2, 2D6 and 3A4 show very similar folds and are characterized by some similarities and differences. Similarities mainly involve the core of these three enzymes that is constituted of highly conserved regions close to the heme, such as HL and the Cys-pocket, on the proximal side of the porphyrin and HI, HJ and HK, on the distal one.

Less conserved regions such as b1-1, b1-2, b1-3, b1-4 and b2-2, HC, HE, HD, HG and the meander also belong to the core [7, 8]. HF and the HF/HG

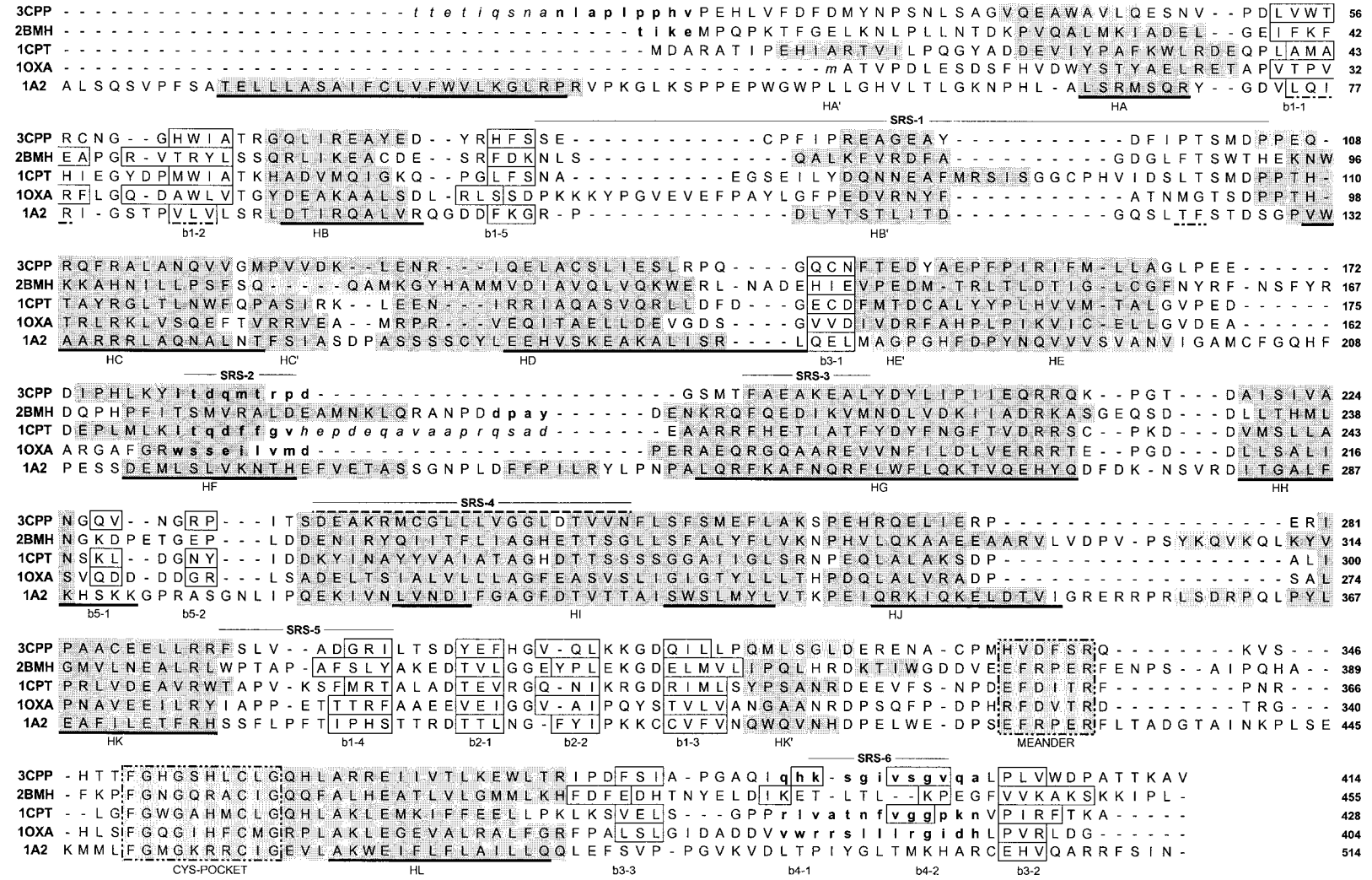


Figure 1. Alignments of the sequence of human microsomal cytochrome 1A2 with the structural alignments of the four templates P450 CAM (3cpp), P450 BM-3 (2bmh), P450 TERP (1cpt) and P450 ERYF (1oxa). Italic and bold lowercase letters label for amino acid residues not solved and deleted, respectively. Dark and light gray shadowing indicate regions found in α-helix or turn while boxes enclose amino acids found in β-strand by the protein analysis implemented in QUANTA 96. Continuous or dashed lines indicate the portions of the target sequence predicted in α-helix or β-strand, respectively. The different α-helix and β-strand regions of the proteins are labeled according to the nomenclature used by Hasemann et al. [8]. The first letter, H, in each capital letter pair, stands for α-helix. The first letter, b, in each lowercase letter/number pair stands for β-strand. Substrate recognition sites (SRS) proposed by Gotoh [61] are also reported.

Figure 2. Alignments of the sequence of human microsomal cytochrome 2D6 with the structural alignments of the four templates P450 CAM (3cpp), P450 BM-3 (2bmh), P450 TERP (1cpt) and P450 ERYF (1oxa). See the legend of Figure 1 for a description of this figure.

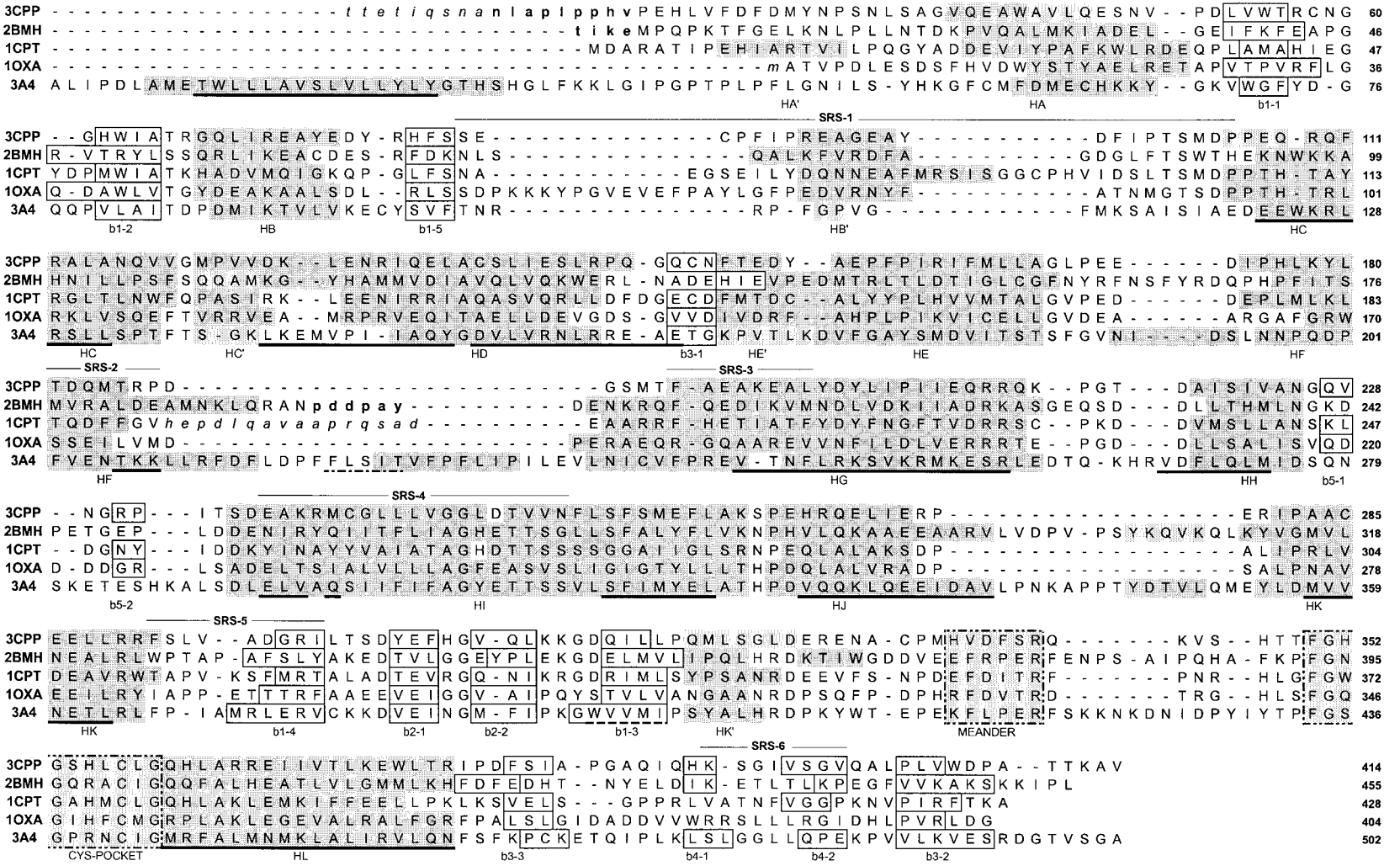


Figure 3. Alignments of the sequence of human microsomal cytochrome 3A4 with the structural alignments of the four templates P450 CAM (3cpp), P450 BM-3 (2bmh), P450 TERP (1cpt) and P450 ERYF (1oxa). See the legend of Figure 1 for a description of this figure.

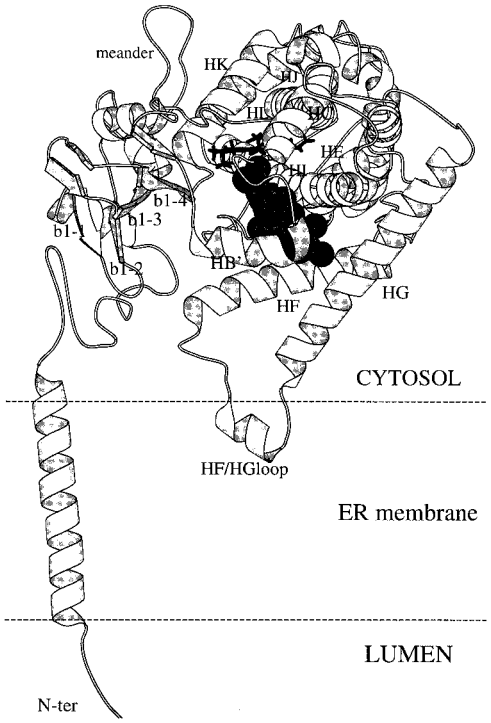


Figure 4. Cartoon representation of the 1A2 model. Labels indicate some of the most relevant regions, according to the nomenclature used by Hasemann et al. [8]. The heme molecule is drawn in black liquorice. Gray spheres indicate the active site cavity computed by the DOCK3.5 program. A hypothetical membrane topology is also shown in this figure. Drawings were performed using the MOLSCRIPT program [65].

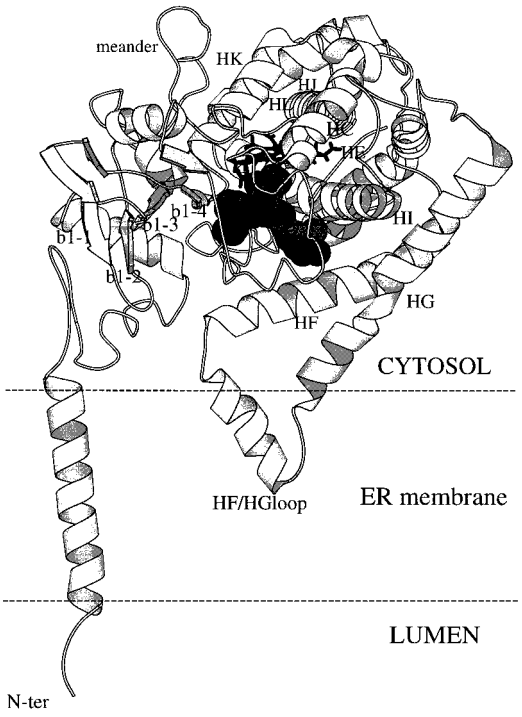


Figure 5. Cartoon representation of the 2D6 model. See Figure 4 for a description of this figure.

loop, the region between HB and HC, b4-1 and b4-2 are among the least structurally conserved regions and the most flexible ones: they are supposed to be involved in substrate recruitment and binding and are identified as substrate recognition sites (SRS) 1, 2 and 6, respectively (Figures 1–3, 4–6) [7, 61].

In the active sites of the three enzymes, the heme is sandwiched between the proximal HL and the distal HI, lying on two parallel planes. The heme iron and propionates contribute to stabilize the anchoring of this molecule into the enzymatic cavity by establishing, respectively, a covalent bond with a highly conserved cysteine (Cys-pocket) and electrostatic interactions with positively charged residues nearby.

The active site of 1A2, 2D6 and 3A4 is also characterized by the XT (where X is an acidic residue) motif of residues in the central portion of HI. This motif is supposed to be involved in the catalytic mechanism of P450s [46, 47]. It is noteworthy that this acidic residue is absent in the 2D6 isozyme, where the corresponding residue is Val 308. It was suggested that, in this case,

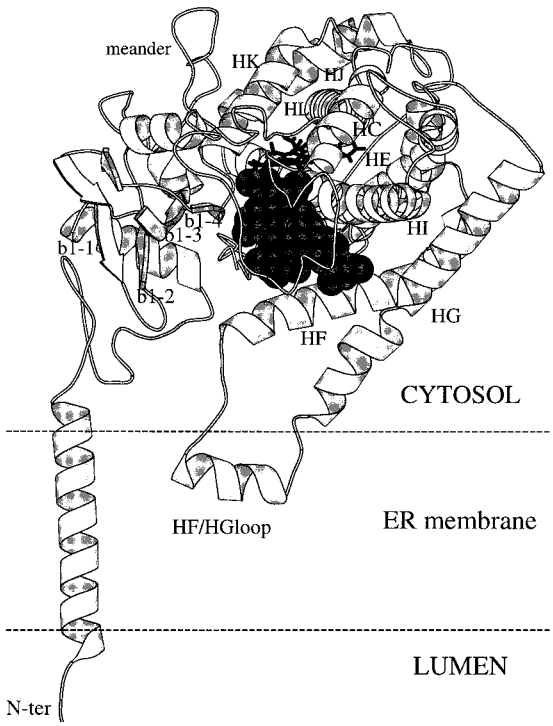


Figure 6. Cartoon representation of the 3A4 model. See Figure 4 for a description of this figure.

the requested proton might be transferred by a bound water molecule [20].

Another interesting common feature of the isozymes 1A2, 2D6 and 3A4 is the presence of a channel connecting the core binding site to the surface of the protein, which could be the access/exit channel through which the substrate is accepted or ejected. This channel is constituted by residues in HA′, the HA′/HA loop, HF, the HF/HG loop, b1-1 and b1-2 regions (Figures 4–6). The location of this channel in our models is in agreement with one of the three possible access/exit channels proposed by a recent theoretical study on P450 CAM [62].

Despite these similarities, the active sites of the three cytochromes considered in this study display also some structural peculiarities. In fact, in the substrate binding site of cytochrome 1A2, the heme is stabilized by interactions between the propionate of heme pyrrole D (D propionate) and amino acids W132, R136 (HC), R455 (Cys-pocket), on one hand, and between the propionate of heme pyrrole A (A propionate) and amino acids R455 (Cys-pocket), K105, R107 (b1-5), on the other one. The active site is occupied by amino acids from (a) HB′ (R107, Y111, T114, T117, D118, T123, F124); (b) HF (N221, T222, F225); (c) HI (N311, D312, G315, A316, F318, D319, T320); (d) HK/b1-4 loop (F380, L381) and (e) the C-terminus (I493 and Y494).

In 2D6, the heme is stabilized by interactions between the D propionate and R115 (C-terminal region of HB′), W128 and R132 (both in HC) and between the A propionate and R441 (Cys-pocket). The active site includes residues from (a) b1-5/HC loop (R101, P102, I106, L110, S116, P114, and Q117); (b) HI (D301, S304, A305, V308, and T309); (c) HK/b1-4 loop (V370) and (d) C-terminus (F483 and L484). In 2D6, differently from 1A2, HF lies on the boundary of the active site: i.e., L205 is the only residue in this helix that could be involved in substrate interaction.

In the 3A4 active site, the heme is stabilized by interactions between the D propionate and residues W125, R129 (HC) and R439 (Cys-pocket) and between the A propionate and residue R374 (b6-1). Among the residues from the b1-5/HC loop, those protruding in the active site cavity are P109, K114, A116, I117, while N103, R104, R105, F107, V110 are at the boundary of the site. The side chains of amino acids from (a) HI (I300, F303, A304, E307, T308), (b) HK/b1-4 (I368, R371, L371, E373) and (c) C-terminus (L481, L482) also protrude into the cavity. Amino acids in HF lie at the boundary of the site: only some of them could interact with large substrates.

Some of the structural differences among 1A2, 2D6 and 3A4 reside in the substrate binding sites that are characterized by different shapes and dimensions (Figures 4–6). The cavities of 1A2 and 2D6 are smaller than that of 3A4. Furthermore, the cavity of 1A2 shows an almost flat shape, according to the structural features of its substrates (Scheme 1). The different size of the three isozyme active sites is mostly due to the conformations assumed by the most variable regions, corresponding to HB′, HF, b4-1 and b4-2, in each isozyme. This finding is consistent with the hypothesis that these regions, being important in substrate recruitment and binding (SRS-1, 2 and 6), need to be highly adaptable to different substrates [7, 61].

In the last years, several 3D models of 1A2, 2D6 and 3A4 have been proposed [17–23]. Since they were built following the same approach (comparative modeling) employed in this work, they share with our models the same fold and the same topography of the active site.

However, despite these similarities, our models show several differences from those previously published that mainly concern the modeling of the domains that should allow these proteins to anchor the ER membrane.

Site directed mutagenesis and biochemical experiments suggest that the N-terminal portion of microsomal P450s works as a membrane insertion and anchoring region [13–16]. These hypotheses are also consistent with the results of the secondary structure predictions, suggesting that a stretch of 20–30 residues in the N-terminal portion of these proteins should form an α-helix, characterized by a length sufficient to span the phospholipid bilayer. According to a membrane topology model proposed for these isozymes, the anchoring to the membrane should be mediated by one N-terminal transmembrane segment and also by some peripheral contacts [13–16].

In our models, the N-terminal region was built in α-helical conformation with its main axis mainly perpendicular to the heme plane. A group of positively charged residues at the C-terminus of this membrane-spanning segment should interact with the phosphate head groups of the phospholipid bilayer. Furthermore, the heme plane is held in a position almost parallel to the ER membrane surface, while the access/exit channel connecting the active site with the surface of the protein faces the ER membrane. Our topological

model (Figures 4–6) is in agreement with the model recently published by Dai et al. for the microsomal rat P450 IIB1 [63].

Most of the modeling works published on 1A2, 2D6 and 3A4 did not deal with the problem of assigning structure and topology to the N-terminal region [17–21, 23], while Szklarz et al. [22] modeled the N-terminal α-helical segment of 3A4 in a different topology with respect to our model. Moreover, the HF/HG loop in our models is longer than in the others. We propose that this portion should perform peripheral contacts with the membrane.

Structural features of the substrates

Cytochromes 1A2, 2D6 and 3A4 display overlapping but different substrate specificities (Scheme 1). 1A2 is able to accommodate either neutral or basic molecules. In general, most of the 1A2 substrates are lipophilic and contain planar aromatic moieties and at least one putative hydrogen bonding donor. Most of them undergo an N- or O-dealkylation or aromatic ring hydroxylation reaction [2, 30].

All the substrates of 2D6 are aryl-alkyl amines sharing a basic nitrogen atom, that is protonated at physiological pH. They also show an almost fixed distance (close to 5 or 7 Å) between the protonated nitrogen atom and the oxidation site [30–33]. Often, the oxidation reaction site consists of aromatic ring or short alkyl side chain hydroxylation.

Substrates of 3A4 cover a wide range of lipophilic drugs that are characterized by different size and shape and are either neutral or basic. Some of them, such as **37**, **42** and **46**, undergo N-dealkylation while other substrates, such as the steroids (**25–28**, **48**) and **29**, are metabolized on their allyllic position [30, 31].

Some molecules can be substrates of more than one cytochrome. The opioids (**44**, **45**) and **43**, for example, are metabolized either by 3A4 or by 2D6, their oxidation site being different for the two isozymes (Scheme 1). Similarly, some steroids (i.e., **48**) are metabolized either by 3A4 or by 1A2; moreover, compound **47** is metabolized by both 2D6 and 1A2.

Specific substrates are instrumental for gaining insights into the molecular determinants of substrate specificity for the three isozymes under study.

Substrate/cytochrome interaction models: Qualitative analysis of substrate/1A2 complexes

Table 1 shows the interaction patterns for the substrate/1A2 complexes. Some of the residues involved in substrate binding belong to HB′ and the region around it: T123 and F124 (HB′/HC loop) establish van der Waals interactions while T114, T117 and D118 (HB′) perform hydrogen bonding and electrostatic interactions with the substrates. Other important residues belonging to HI (N311, D312, G315, A316, F318, D319, T320), HK/b1-4 loop (T380 and L381) and the C-terminus (I493 and Y494) interact with the substrates mainly through van der Waals attractive interactions (Table 1). Finally, residues in HF (N221, T222, F225) perform van der Waals or hydrogen bonding and electrostatic interactions with the substrates (Table 1).

In general, a cluster of aromatic residues defines the boundaries of the active site and stabilizes the interaction energy of the substrate/1A2 complexes. Most of these complexes show the substrate lying on a plane that is nearly perpendicular to the heme plane and intersects the heme pyrroles B and D. The oxidation site of the substrates is oriented towards the heme and residue T320 (Figure 7). Recently, several models of the complexes between 1A2 and some of the substrates considered in this study have been proposed [17, 18].

The orientation and the interaction pattern of compound **1** in our model is also in good agreement with that obtained by Lozano et al. [17]. In the case of compound **3**, where the substrate orientation is different in our and in Lozano's model [17], agreement is however achieved on the orientation of the ligand oxidation site with respect to the heme molecule and the type of residues that interact with the substrate.

Strong agreement has been found between the interaction models of compounds **1**, **5–8**, **49** and **50** obtained in this work and those proposed by Lewis and Lake [18].

Substrate/cytochrome interaction models: Qualitative analysis of substrate/2D6 complexes

Table 2 shows the amino acids involved in the stabilization of substrate/2D6 complexes.

It can be seen that a group of residues belonging to the region between b1-5 and HC interacts with most of the substrates: P102, I106 and L110 are involved in van der Waals interactions, while S116 and Q117 establish, in a few cases, hydrogen bonding and/or electrostatic interactions. Furthermore, many

Table 1. Amino acid residues that contribute, by attractive van der Waals (vdw), electrostatic (el) and H-bonding (hb) interactions, to the total interaction energy of substrate/1A2 complexes

sub.	ARG107	THR 114	THR 117	ASP 118	THR 123	PHE 124	GLU 214	ASN 221	THR 222	PHE 225	PHE 259	ASN 311	ASP 312	GLY 315	ALA 316	PHE 318	ASP 319	THR 320	PHE 380	LEU 381	ILE 493	TYR 494	LYS 499
1					vdw	vdw			vdw				vdw	vdw	vdw		el	vdw	vdw	vdw	vdw	vdw	
2		vdw			vdw	vdw			vdw	vdw				vdw	vdw		hb+el	vdw		vdw	vdw	vdw	
3		vdw	vdw	vdw	vdw	vdw		hb	vdw	vdw	vdw	vdw		el	vdw	vdw	el		vdw		vdw	vdw	
4		vdw	vdw	el		vdw		hb	vdw		vdw	vdw		el	vdw	vdw	el		vdw			vdw	
5	vdw				vdw	vdw			hb+el	vdw				vdw	vdw	vdw	vdw	vdw	vdw	vdw	vdw	vdw	
6	vdw				vdw	vdw			vdw					vdw	vdw	vdw	vdw		vdw	vdw	hb+el	hb	vdw
7		hb+el	el		vdw	vdw		vdw	hb+el	vdw				vdw	vdw			vdw	vdw	vdw			
8		vdw				vdw			hb+el					vdw			vdw			vdw	vdw	vdw	
9		hb+el			vdw	vdw			vdw					vdw	vdw			vdw	vdw	vdw	vdw	vdw	
47		el	vdw	el	vdw	vdw	vdw	vdw	hb+el	vdw		el	el	vdw	vdw	vdw	vdw	vdw		vdw	vdw	vdw	
48		hb	vdw	vdw	vdw	vdw		vdw	vdw	vdw		vdw	vdw	vdw	vdw		el	vdw	vdw	vdw	vdw	hb+el	
49	vdw	vdw		el	vdw	vdw			hb+el	vdw		vdw	vdw	vdw	vdw		vdw	vdw	vdw	vdw	vdw	vdw	
50				el	vdw	vdw			hb+el				el	vdw	vdw		vdw	vdw	vdw	vdw	vdw	vdw	

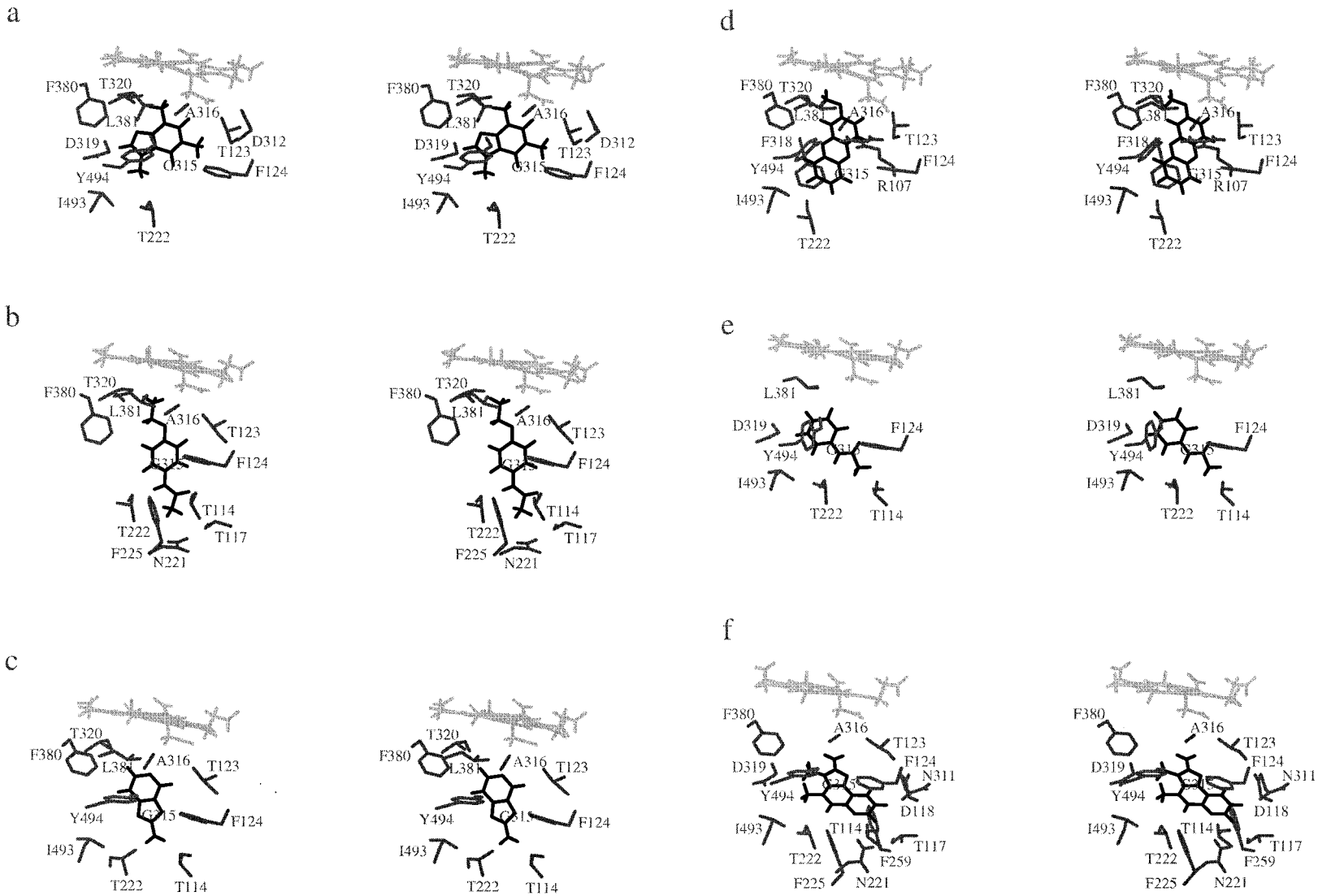


Figure 7. Stereoviews of some of the interactions characterizing (a) caffeine/, (b) phenacetine/, (c) zoxazolamine/, (d) methoxy-resorufin/, (e) acetanilide/ and (f) MEIQ/1A2 complexes. The heme molecule and the amino acid side chains are colored in light and dark gray, respectively. The substrate is colored in black. Drawings were performed with the MOLSCRIPT program [65].

Table 2. Amino acid residues that contribute, by attractive van der Waals (vdw), electrostatic (el) and H-bonding (hb) interactions, to the total interaction energy of substrate/2D6 complexes

sub.	ARG 101	PRO 102	ILE 106	ILE 109	LEU 110	SER 116	GLN 117	LEU 205	LEU 208	ALA 209	ASP 301	SER 304	ALA 305	VAL 308	THR 309	ILE 369	VAL 370	MET 374	PHE 483	LEU 484
10		vdw	vdw		vdw	vdw	vdw				hb+el	hb	vdw	vdw	vdw	vdw			vdw	vdw
11		vdw	vdw		vdw	vdw	vdw				hb+el	hb+el	vdw	vdw	vdw		vdw		vdw	
12		vdw	vdw		vdw	hb	vdw				hb+el	vdw	vdw		vdw		vdw		vdw	vdw
13		vdw	vdw		vdw	hb	vdw				hb+el	hb+el	vdw	vdw	vdw		vdw		vdw	vdw
14		vdw	vdw		vdw	hb	vdw				hb+el	vdw	vdw		vdw				vdw	vdw
15	vdw	vdw			vdw	hb	vdw				hb+el	hb+el	vdw	vdw	vdw				vdw	vdw
16		vdw	vdw		vdw	el	vdw				hb+el	vdw	vdw		vdw				vdw	vdw
17		vdw			vdw	el	el				hb+el	vdw	vdw				vdw		vdw	vdw
18	vdw	vdw	vdw		vdw	vdw	vdw				hb+el	vdw	vdw		vdw		vdw	vdw	vdw	vdw
19	vdw	vdw				vdw	vdw				hb+el	vdw	vdw		vdw		vdw		vdw	vdw
20	vdw	vdw	vdw	vdw	vdw	vdw	vdw		vdw	vdw	hb+el	hb+el	vdw				vdw		vdw	vdw
21	vdw	vdw	vdw			hb	vdw				hb+el	hb+el	vdw				vdw		vdw	vdw
22		vdw	vdw		vdw	hb	vdw	vdw			hb+el	vdw	vdw	vdw			vdw		vdw	vdw
23		vdw	vdw		vdw	vdw	vdw				hb+el	vdw	vdw	vdw	vdw		vdw			vdw
44	vdw	vdw	vdw		vdw	vdw	vdw				el	hb+el	vdw				vdw		vdw	vdw
45	vdw	vdw	vdw		vdw	vdw	vdw				hb+el	vdw	vdw	vdw			vdw		vdw	vdw
46		vdw	vdw		vdw	hb	el				hb+el	vdw	vdw		vdw					
47		vdw	vdw		vdw	hb	vdw				hb+el	hb	vdw	vdw	vdw		vdw		vdw	vdw

residues from HI are involved in intermolecular interactions. In fact, D301, that is always involved in charge-reinforced hydrogen bonds with the protonated nitrogen atom of the substrates, S304, that often interacts with the hydroxyl group of the substrate, as well as A305, V308 and T309 play important roles in determining the right orientation of the substrate in the active site. Moreover, V370 (HK/b1-4 loop), F483 and L484 (C-terminus) perform van der Waals attractive interactions with almost all substrates. The important role of F483 in substrate/2D6 interaction is consistent with recent findings showing that this residue can influence the catalytic selectivity of 2D6 [64].

In general: (a) residues in SRS-5 (370-374), SRS-6 (483-484) and HI (T309) interact with the substrate aryl moiety and with the atoms close to the oxidation site; (b) residues in HB′ interact with the substrate alkyl moiety and (c) D301 in HI performs charge-reinforced hydrogen bonding interactions with the protonated nitrogen atom of the substrate. Moreover, a small region in the proximity of residue L205 in HF (SRS-2) interacts, through van der Waals interactions, with the cyclic alkyl or aryl group of large substrates (Figure 8).

The active site of the 2D6 model obtained in this work shares the same topography with that proposed by Modi et al. [20]. This structural similarity allowed us to use the NMR restraint-based codeine/2D6 complex proposed by Modi et al. [19, 20] for selecting the final orientation and interaction pattern of codeine. Agreement is also found between the interaction models of compounds **17** and **47** proposed in this work and those proposed by Lewis et al. [21].

Substrate/cytochrome interaction models: Qualitative analysis of substrate/3A4 complexes

The substrate/3A4 interaction patterns are shown in Table 3. By analyzing this table, it can be seen that two large groups of residues contribute to stabilize the interaction energy of the complexes. In fact, two clusters of residues in HI (I300, F303, A304, E307, T308) and in the b1-5/HC loop (N103, R104, R105 P109, A116 and I117) establish hydrogen bonding, electrostatic or van der Waals interactions with the substrates. Residues in b1-4 (I368, R371, L372, E373, R374) and in the C-terminus (L481 and L482) also contribute to stabilize the substrate/3A4 complexes. Similarly to 2D6, some residues lying on HF, HI and b1-4, at the boundary of the active site cavity, can be reached only by a few particular substrates.

In general, at least for the hormones and for small flexible substrates, (a) the portion of the substrate around the oxidation site is directed under the heme towards T308, (b) the central portion of the substrate interacts with SRS-5 and 6, and (c) the side at the opposite end of the oxidation site of the substrate interacts with the HB/HC loop. It is worth noting that many of the residues protruding in the active site and interacting with the substrate belong to regions, such as 100–105 (HB/HC loop), 371-374 (b1-4) and HF, that should be characterized by conformational flexibility, consistently with published data [7, 61] (Figure 9).

Agreement is found between the orientation of compounds **26**, **30**, **38**, **40** and **41** obtained in this work and those proposed by Lewis et al. [23]. Most of the active site interactions are the same in ours and in Lewis' complexes as well as the interaction pattern involving the substrate and N103, notwithstanding the fact that N103 lies in the structurally variable b1-5 loop. On the other hand, F101, that also lies at the beginning of the b1-5 loop, is never involved in ligand interaction in our models, differently from the Lewis' ones.

Agreement is also found on the main orientation of the bulky compound **42** in our interaction model and in that proposed by Szklarz et al. [22]. Some of the residues that interact with compound **42** are shared by the two models, even if they perform different intermolecular interactions.

Substrate/cytochrome interaction models: Semi-quantitative analysis

The qualitative analysis of substrate/cytochrome interaction models highlighted some of the peculiar features characterizing the basic substrate binding requirements of 1A2, 2D6 and 3A4.

Substrate binding into the active site of 2D6 is mainly favored by hydrogen bonding and electrostatic interactions between the substrate and two residues in HI: D301 and S304. In particular, D301 performs charge-reinforced H-bonding interactions with the protonated nitrogen atom that characterizes the substrates of 2D6. On the contrary, van der Waals attractive interactions mainly contribute to stabilize the complexes involving both the isozymes 1A2 and 3A4. However, because of the smaller size of both the active site and the specific substrates of 1A2 with respect to those of 3A4, the number of intermolecular interactions that stabilize the substrate/1A2 complexes is, in

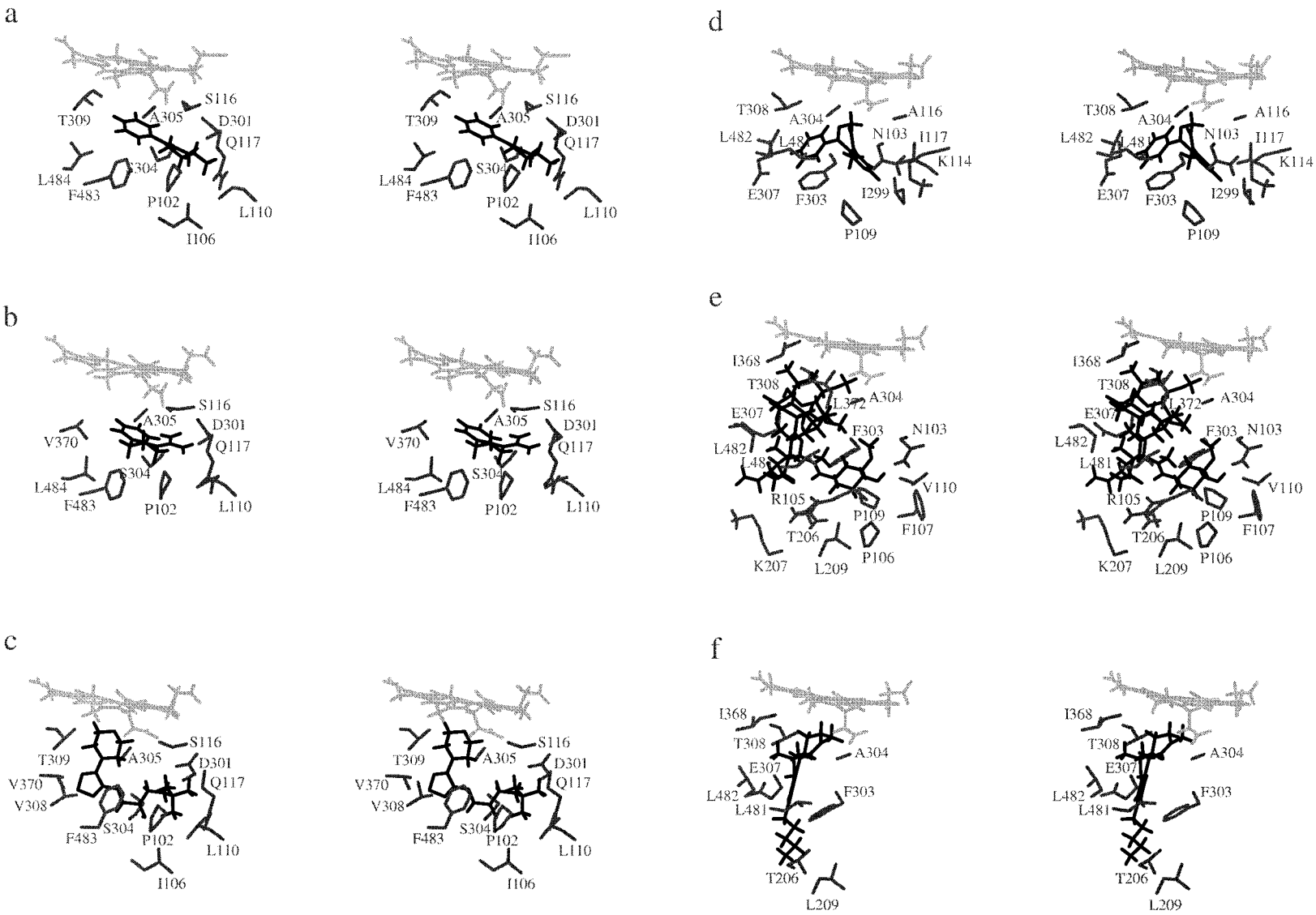


Figure 8. Stereoviews of some of the interactions characterizing (a) MPTP/, (b) debrisoquine/, (c) timolol/, (d) imipramine/, (e) guanoxan/ and (f) perhexilline/2D6 complexes. See Figure 7 for a description of this figure.

Table 3. Amino acid residues that contribute, by attractive van der Waals (vdw), electrostatic (el) and H-bonding (hb) interactions, to the total interaction energy of substrate/P450 IIIA4 complexes

sub.	ASN 103	ARG 104	ARG 105	PRO 106	PHE 107	PRO 109	VAL 110	LYS 114	ALA 116	ILE 117	ASN 205	THR 206	LYS 207	LEU 209
24			hb+el			vdw						vdw		vdw
25	hb		vdw	vdw		vdw	vdw		vdw	vdw				
26	hb		vdw			vdw	vdw		vdw	vdw		vdw		
27	hb					vdw			vdw	vdw				
28	vdw		vdw			vdw	vdw		vdw	vdw				
29						vdw	vdw	vdw		vdw				
30	hb		el			vdw	hb		vdw	vdw				vdw
31	vdw					vdw		vdw	vdw	vdw				
32	vdw	hb+el	vdw			el	vdw	vdw	vdw	vdw				
33												vdw		vdw
34	vdw							vdw	vdw	vdw				
35						vdw	vdw		vdw	vdw				vdw
36	vdw	el	vdw		vdw	vdw	vdw	vdw	vdw	vdw	vdw	vdw		
37	vdw	vdw			vdw	vdw	vdw		vdw	vdw				
38	vdw	vdw	vdw			vdw	vdw		vdw	vdw				
39						vdw	vdw	vdw	vdw	vdw				
40	hb					vdw	vdw			vdw				
41	hb	el	hb+el		vdw	vdw			vdw	vdw				
42			vdw	vdw		vdw	vdw					vdw	vdw	vdw
43	vdw		vdw	vdw	vdw	vdw	vdw					vdw	vdw	vdw
44	vdw	vdw	vdw			vdw	vdw		vdw	vdw				
45		vdw	vdw			vdw	vdw		vdw	vdw				
46	vdw		vdw			vdw			vdw	vdw				
48	hb		vdw		vdw	vdw			vdw	vdw				
49									vdw			vdw	el+hb	
50	hb					vdw	vdw		vdw	vdw				

sub.	ILE 299	ILE 300	PHE 303	ALA 304	GLU 307	THR 308	ILE 368	ARG 371	LEU 372	GLU 373	ARG 374	GLY 480	LEU 481	LEU 482
24			vdw	vdw	el	el	vdw	el	vdw			vdw	el	vdw
25		vdw	vdw	vdw		hb+el	vdw		vdw				vdw	vdw
26			vdw	vdw		hb	vdw		vdw	vdw			vdw	vdw
27			vdw	vdw		hb	vdw		vdw				vdw	vdw
28		vdw	vdw	hb+el		hb							vdw	
29		vdw	vdw			vdw	vdw	vdw	vdw	vdw			hb+el	vdw
30		vdw	vdw				vdw	vdw	vdw	vdw			vdw	
31	vdw		vdw	vdw	vdw	vdw							vdw	vdw
32		vdw	vdw	vdw			vdw		vdw	el	vdw	vdw	el	
33			vdw	vdw	hb+el	vdw	vdw		vdw			vdw	vdw	vdw
34		vdw	vdw	vdw		el	vdw						vdw	
35			vdw					vdw	vdw	hb			hb	vdw
36	vdw	vdw	vdw			hb+el	vdw	el	vdw	vdw		vdw	el	vdw
37				vdw	el	hb+el	vdw		vdw		vdw		vdw	vdw
38		vdw		vdw		vdw	vdw			vdw	vdw	vdw	hb+el	vdw
39		vdw	vdw	vdw	el	hb+el	vdw	vdw	vdw				vdw	vdw
40		vdw	vdw	vdw		vdw	vdw	hb+el	vdw	hb	vdw		el	vdw
41			vdw	vdw		vdw	vdw		vdw				vdw	vdw
42			vdw	vdw	hb+el	vdw	vdw		vdw			vdw	vdw	vdw
43			vdw	vdw	vdw	vdw	vdw	vdw	vdw				vdw	hb+el
44			el	vdw		vdw	vdw	vdw	vdw	hb			el	vdw
45			el			el	vdw	el	vdw	el			el	
46		vdw	vdw			vdw	vdw		vdw				vdw	vdw
48				vdw		vdw	vdw		vdw				vdw	
49		vdw	vdw		vdw	vdw	vdw						vdw	vdw
50		vdw	vdw			vdw							vdw	vdw

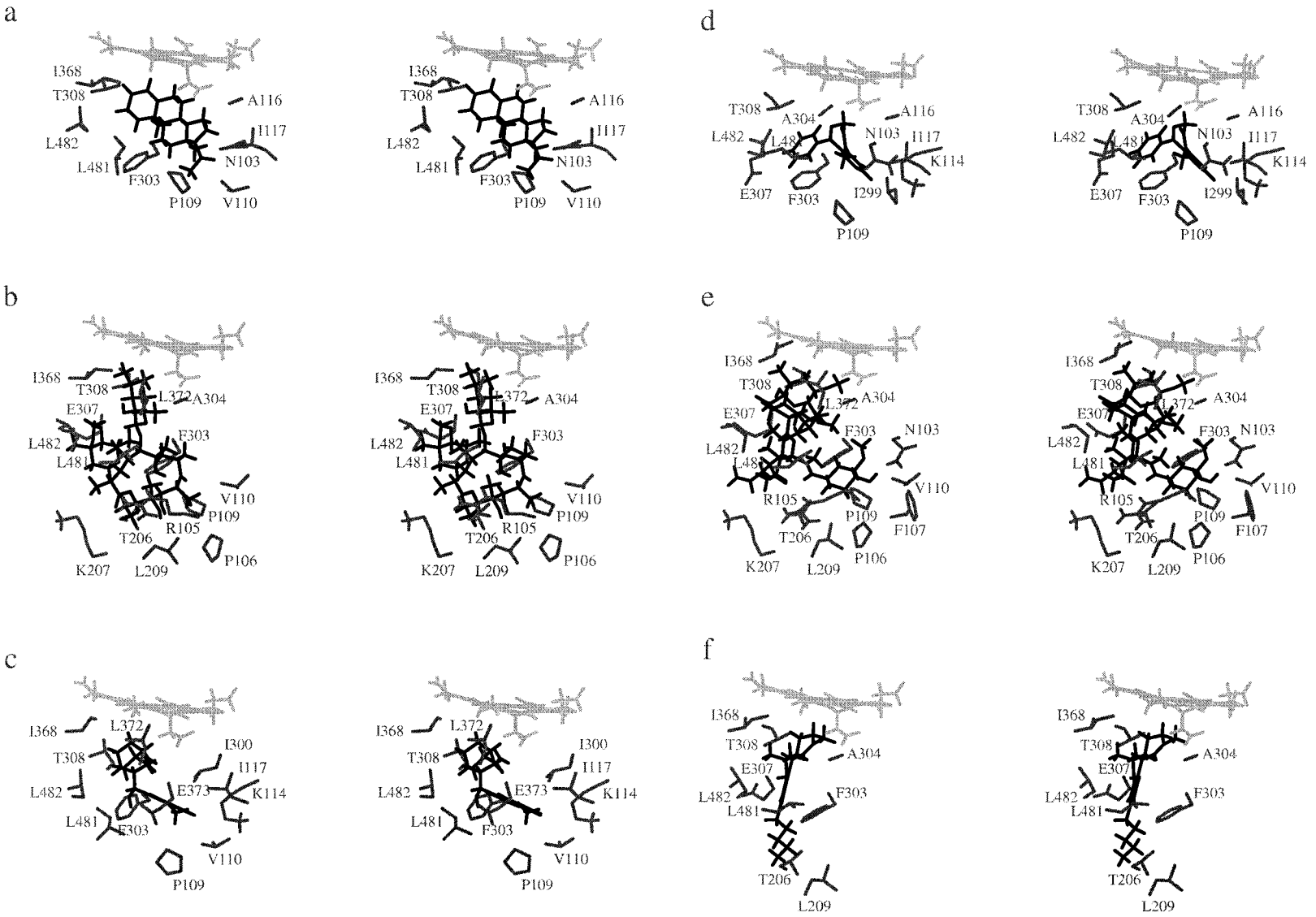


Figure 9. Stereoviews of some of the interactions characterizing (a) progesterone/, (b) erythromycin/, (c) quinidine/, (d) diazepam/, (e) FK506 / and (f) DTHC/3A4 complexes. See Figure 7 for a description of this figure.

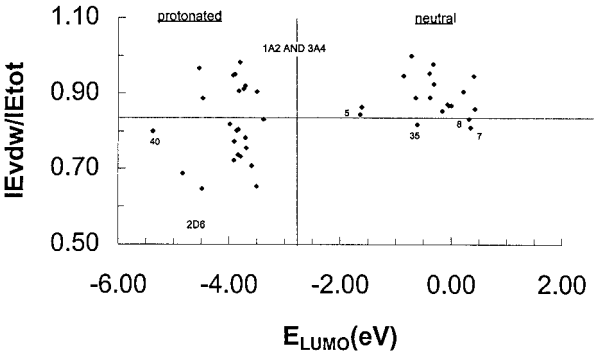


Figure 10. Plot of the ratio between the van der Waals contribution to the total interaction energy and the total interaction energy (IE_{vdw}/IE_{tot}) of the substrate/cytochrome complexes versus the energy of the lowest unoccupied molecular orbital (E_{LUMO}). Dashed circles enclose 2D6 specific substrates. Numbers label for the substrates whose specificity is not well described by either theoretical index.

general, lower than that stabilizing the substrate/3A4 complexes.

In order to translate the information obtained from the qualitative analysis of the substrate/cytochrome interaction models into semi-quantitative models of substrate specificity, we defined and tested several molecular and intermolecular-interaction descriptors handling the structural information derived from the analyses of the three cytochromes and of several substrates both in their free and bound forms. The descriptors selected are shown in Table 4, whereas the best semi-quantitative models are shown in Figures 10–12. It is worth noting that these models are based only on specific substrates.

Plotting the intermolecular interaction descriptor IE_{vdw}/IE_{tot} (ratio between the van der Waals contribution to the total interaction energy and the total interaction energy) versus the molecular descriptor E_{LUMO} (energy of the Lowest Unoccupied Molecular Orbital (LUMO)) (Figure 10) provided a model useful for differentiating the substrates specific for 2D6, which are all protonated, from those that are specific for the other two cytochromes. In fact, E_{LUMO} separates the substrates in two clusters: one constituted by the substrates that interact with the enzymes in their neutral form (E_{LUMO} close to 0 eV) and the other constituted by substrates that interact with the enzymes in their protonated form (E_{LUMO} around -4 eV). The first cluster contains all the neutral substrates specific for 1A2 and 3A4. In the second cluster, constituted by protonated substrates, all the 2D6 substrates, together with some 3A4 substrates, find place (Figure 10), according to the set of compounds considered in this study. Inside this cluster, the IE_{vdw}/IE_{tot} descriptor is able to separate the 2D6 specific substrates from the 3A4 specific ones. In fact, due to the nature of the intermolecular interactions characterizing these two cytochromes, this index is in the 0.60–0.83 range for the substrate/2D6 complexes, while it ranges from 0.80 to 1.00 for the substrate/3A4 complexes.

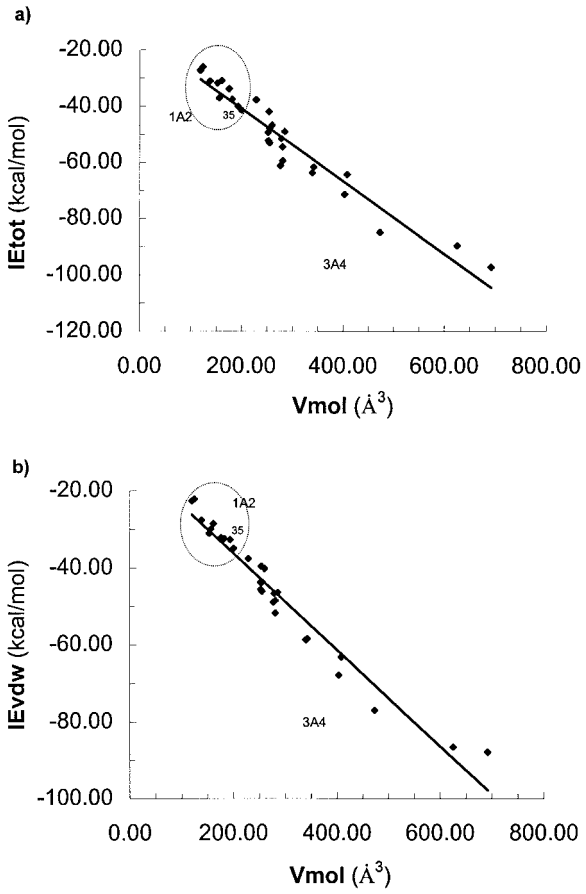


Figure 11. (a) Correlation between the total interaction energy (IE_{tot}) of the substrate/enzyme complexes and the molecular volume (Vmol) of the substrates. Linear regression equation is $IE_{tot} = -14.924\ (\pm\ 2.100)\ -0.129\ (\pm\ 0.007)$ Vmol; $n = 29$; $r = 0.964$; $s = 4.922$. (b) Correlation between the van der Waals contribution to the total interaction energy (IE_{vdw}) and the molecular volume (Vmol) of the substrates. Linear regression equation is $IE_{vdw} = -11.174\ (\pm 1.594)\ -0.125\ (\pm\ 0.005)$ Vmol; $n = 29$; $r = 0.978$; $s = 3.735$; where n is the number of compounds, r is the correlation coefficient, s is the standard deviation and the numbers in parentheses are the 95% confidence ranges. Dashed circles enclose 1A2 specific substrates.

The 1A2 and 3A4 substrates (**5**, **7**, **8**, **35** and **40**) that, according to IE_{vdw}/IE_{tot} values, lie at the borderline between the 1A2-3A4 and 2D6 areas, are

Table 4. Molecular (Elumo, Vmol, Stot, Sphi, Sphi/Stot) and intermolecular interaction (IEtot, IEvdw, IEvdw/IEtot) descriptors of substrates and substrate/cytochrome complexes

sub.	cyto	Elumo	Vmol	Stot	Sphi	Sphi/Stot	IEtot	IEvdw	IEvdw/IEtot
1	1A2	-0.3247	153.25	179	24	0.134	-31.74	-31.01	0.98
2	1A2	-0.3811	138.50	162	29	0.179	-31.09	-27.64	0.89
3	1A2	-0.3907	175.75	199	19	0.095	-33.73	-32.14	0.95
4	1A2	-0.3114	161.25	187	19	0.102	-30.86	-28.54	0.92
5	1A2	-1.6116	182.00	211	27	0.128	-37.47	-32.39	0.86
6	1A2	-1.6383	200.38	228	26	0.114	-41.32	-34.89	0.84
7	1A2	0.3506	156.50	190	15	0.079	-36.99	-29.93	0.81
8	1A2	0.3243	119.63	145	12	0.083	-27.21	-22.64	0.83
9	1A2	-0.1592	124.38	148	21	0.142	-26.02	-22.20	0.85
10	2D6	-3.6882	232.13	268	9	0.034	-56.98	-43.03	0.76
11	2D6	-3.8356	255.50	299	26	0.087	-61.51	-45.32	0.74
12	2D6	-3.7918	154.13	185	8	0.043	-39.31	-28.79	0.73
13	2D6	-3.5901	183.75	212	4	0.019	-54.70	-38.66	0.71
14	2D6	-3.9114	167.63	196	4	0.020	-45.24	-32.66	0.72
15	2D6	-3.5050	153.25	183	3	0.016	-51.17	-33.39	0.65
16	2D6	-3.9016	161.38	193	0	0.000	-38.73	-29.91	0.77
17	2D6	-4.8405	152.25	181	13	0.072	-41.96	-28.85	0.69
18	2D6	-3.8589	254.00	284	0	0.000	-52.26	-41.89	0.80
19	2D6	-3.9816	239.75	271	1	0.004	-52.24	-42.73	0.82
20	2D6	-3.8287	294.38	337	15	0.045	-58.75	-47.25	0.80
21	2D6	-4.4859	169.13	202	24	0.119	-50.59	-32.70	0.65
22	2D6	-3.7024	265.5	289	1	0.003	-57.14	-44.71	0.78
23	2D6	-3.3774	210.38	227	0	0.000	-40.36	-33.53	0.83
24	3A4	-3.7347	255.38	276	16	0.058	-47.88	-43.67	0.91
25	3A4	-0.0450	280.88	282	23	0.082	-59.52	-51.67	0.87
26	3A4	0.0027	255.13	262	21	0.080	-53.02	-46.01	0.87
27	3A4	-0.0622	252.88	257	25	0.097	-52.28	-45.53	0.87
28	3A4	0.4267	259.75	272	18	0.066	-46.70	-40.16	0.86
29	3A4	-3.4941	278.75	300	14	0.047	-51.52	-46.57	0.90
30	3A4	-0.6380	280.88	302	34	0.113	-54.45	-48.39	0.89
31	3A4	-0.7136	228.75	255	14	0.055	-37.62	-37.61	1.00
32	3A4	-3.9250	342.50	372	16	0.043	-61.68	-58.45	0.95
33	3A4	0.4105	285.13	320	10	0.031	-49.01	-46.35	0.95
34	3A4	-0.8581	253.75	279	9	0.032	-41.82	-39.58	0.95
35	3A4	-0.6078	193.25	221	34	0.154	-39.95	-32.65	0.82
36	3A4	-3.8182	473.50	482	24	0.050	-84.90	-76.94	0.91
37	3A4	-3.8936	403.50	423	13	0.031	-71.38	-67.83	0.95
38	3A4	-3.7067	339.63	371	2	0.005	-63.77	-58.66	0.92
39	3A4	-3.7995	408.50	413	24	0.058	-64.30	-63.13	0.98
40	3A4	-5.3741	276.75	314	29	0.092	-61.16	-48.81	0.80
41	3A4	-4.4721	252.38	277	12	0.043	-49.28	-43.70	0.89
42	3A4	-4.5405	625.75	555	46	0.083	-89.53	-86.51	0.97
43	3A4	0.2195	692.38	643	51	0.079	-97.13	-87.85	0.90
44	2D6	-3.8173	249.00	256	16	0.063	-53.04	-45.32	0.85
44	3A4	-3.8173	249.00	256	16	0.063	-46.18	-42.85	0.93
45	2D6	-3.7755	261.38	271	12	0.044	-53.27	-44.78	0.84
45	3A4	-3.7755	261.38	271	12	0.044	-46.66	-43.72	0.94
46	2D6	-3.9814	217.25	246	7	0.028	-55.70	-41.78	0.75
46	3A4	-3.9814	217.25	246	7	0.028	-41.98	-41.84	1.00
47	1A2	-3.7279	233.88	275	7	0.025	-54.12	-46.22	0.85
47	2D6	-3.7279	233.88	275	7	0.025	-57.71	-43.67	0.76
48	1A2	0.4396	238.13	238	18	0.076	-52.09	-44.35	0.85
48	3A4	0.4396	238.13	238	18	0.076	-50.68	-42.52	0.84
49	1A2	-1.3638	234.25	256	38	0.148	-47.30	-42.85	0.91
49	3A4	-1.3638	234.25	256	38	0.148	-41.74	-38.66	0.93
50	1A2	-0.8610	158.13	186	22	0.118	-34.04	-29.51	0.87
50	3A4	-0.8610	158.13	186	22	0.118	-34.93	-30.89	0.88

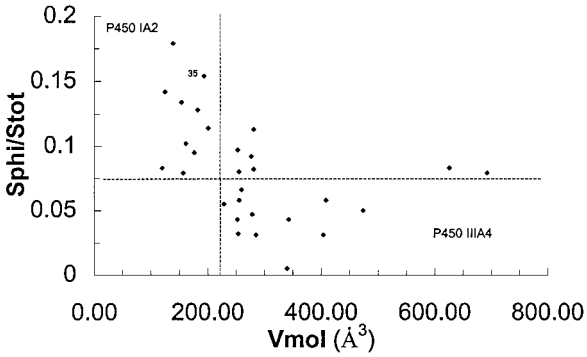


Figure 12. Graph of the ratio between the hydrophilic component and the total solvent accessible surface (Sphi/Stot) plotted versus the molecular volume (Vmol).

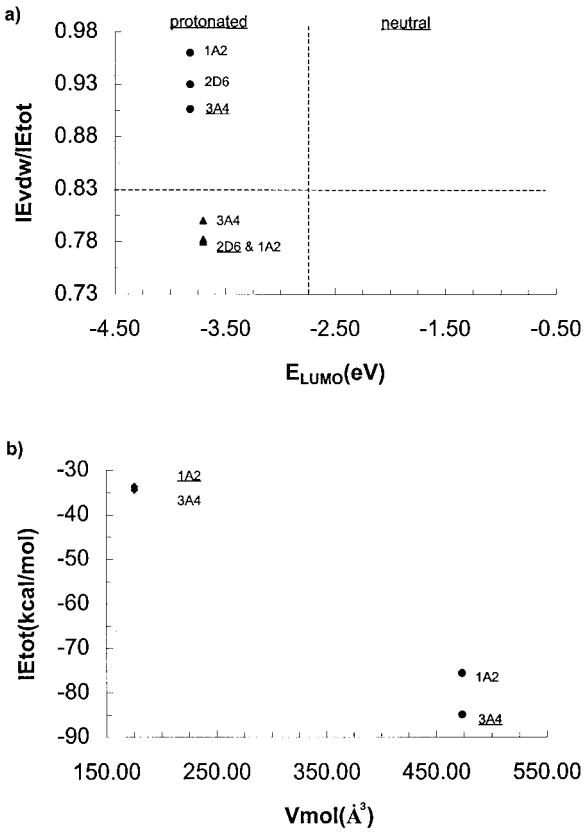


Figure 13. This figure shows the model response of three substrates specific for 1A2 (compound **3**, labeled by filled rhombs), 2D6 (compound **22**, labeled by filled triangles) and 3A4 (compound **36**, labeled by filled circles) when they are docked into their specific enzyme (underlined labels) as well as into the other two cytochromes (non underlined labels). The models employed in (a) and (b) are those depicted in Figures 10 and 11, respectively.

however well discriminated by E_{LUMO} because, differently from all the 2D6 substrates, they are neutral molecules.

The behavior of the 3A4 substrate omeprazole (**40**) is not well described neither by the IE_{vdw}/IE_{tot} nor by the E_{LUMO} indexes. In fact, the intermolecular interaction descriptor has a value typical of 2D6 specific substrates. On the other hand, the molecular descriptor E_{LUMO} accounts for the unique electronic features of this compound.

The cluster of neutral substrates (E_{LUMO} close to 0 eV) is formed by substrates that are specific either for 3A4 or for 1A2. In fact, since all the 1A2 and 3A4 complexes with their specific substrates are characterized by similar values of the descriptor IE_{vdw}/IE_{tot}, the model shown in Figure 10 is not able to separate the substrates which are specific for 1A2 or 3A4. In this case, the total interaction energy (IE_{tot}) works better in differentiating the two classes of substrates. The ability of IE_{tot} to discriminate between 1A2 and 3A4 substrates is due to the marked difference between the sizes of the active sites (Figures 4 and 6) as well as of the substrates of these two enzymes (Scheme 1). In fact, due to the smaller size of both the active site and the specific substrates of 1A2 with respect to 3A4, the number of stabilizing interactions is lower and the total interaction energy is higher for the substrate/1A2 complexes than for the substrate/3A4 complexes.

Thus, interaction energy (IE_{tot}) holds the structural information about the size of the interacting substrates. In fact, this index, similarly to the van der Waals component of the total interaction energy (IE_{vdw}), is highly correlated with the molecular volume (Vmol) of the substrates (Figure 11). Moreover, the combined use of IE_{tot} or IE_{vdw} and the molecular descriptor Vmol provides a theoretical model able to discriminate between substrates specific for 1A2 and 3A4 (Figure 11). 1A2 substrates are, in fact, characterized by higher values of IE_{tot} (or IE_{vdw}) and lower values of Vmol with respect to the 3A4 substrates. Another molecular descriptor able to separate substrates of 1A2 from those of 3A4 is the ratio between the hydrophilic component and the total solvent accessible surface (S_{phi}/S_{tot}) (Figure 12). In fact, due to the flat shape characterizing the typical 1A2 substrates, the relative hydrophilic surface exposed to the solvent is larger for these substrates than for those of 3A4. However, Vmol works better than S_{phi}/S_{tot} in differentiating the 1A2 specific substrates from the 3A4 specific ones (Figure 12).

All the descriptors employed fail in classifying dapsone (**35**). This is probably due to the small size

and the strong hydrophilic character of this compound that make it an 'atypical' 3A4 substrate, within our theoretical models.

According to the semi-quantitative models proposed, prediction of substrate specificity for 1A2, 2D6 and 3A4 should follow the protocol listed below.

(1) For a substrate to be predicted specific for 2D6, it must carry a protonated nitrogen atom as primary molecular feature. Its intermolecular interaction peculiarity should be a van der Waals contribution to the total interaction energy for its complex with 2D6 significantly lower than 0.80, while docking the same substrate into 1A2 and 3A4 should give IE_{vdw}/IE_{tot} values close to or lower than 0.83. As an example, the behavior of a specific substrate (**22**) of 2D6 is clearly displayed in Figure 13a, which shows that docking **22** with each of the three cytochromes produces IE_{vdw}/IE_{tot} indices lower than 0.83 (IE_{vdw}/IE_{tot} values for the complexes of **22** with 1A2, 2D6 and 3A4 are, respectively, 0.78, 0.78 and 0.80), thus placing this compound always in the 2D6 area, independently of the interacting cytochrome considered (Figure 13a).

(2) Specific substrates for 1A2 and 3A4 may be neutral or protonated (Scheme 1).

(a) Protonated substrates specific for either 1A2 or 3A4 should display values for their complexes with each of the three cytochromes close to or higher than 0.83 (Figure 10). In fact, docking the protonated specific 3A4 substrate **36** in each of the three cytochromes always produces IE_{vdw}/IE_{tot} indices significantly higher than 0.83 (IE_{vdw}/IE_{tot} values for the complexes of **36** are respectively 0.96, 0.93, 0.91) (Figure 13a). As a consequence, this substrate lies in the typical zone of protonated 1A2 or 3A4 substrates independently of the interacting cytochrome (Figure 13a).

However, since IE_{vdw}/IE_{tot} is not able to differentiate 1A2 from 3A4 specific substrates, the procedure outlined below in (b) and (c) must be followed, in order to predict whether they are specific for only one of the two isoforms. The following procedure works also for neutral substrates.

(b) A neutral or protonated substrate may be predicted specific for 1A2 when the total interaction energy (IE_{tot}) of its complexes with 1A2 and 3A4 is higher than -40 kcal/mol (Vmol lower than 200 Å^3, according to Figure 11). In fact, from Figure 13b it can be seen that the 1A2 specific substrate **3** falls in the typical area of specific 1A2 substrates (Figures 11 and 13b), not only because it gives IE_{tot} values higher than -40 kcal/mol, upon docking into both 1A2 and 3A4 (IE_{tot} values of the complexes of **3** with 1A2 and 3A4 are, respectively, -33.73 and -34.42 kcal/mol), but also because this molecule is characterized by a volume lower than 200 Å^3. On the other hand, a specific substrate for 3A4 should be characterized by a total interaction energy of its complexes with 1A2 and 3A4 lower than -50 kcal/mol (Vmol higher than 275 Å^3, according to Figure 11). The example relative to this statement is also shown in Figure 13b where it can be seen that the specific 3A4 substrate **36** always falls in the typical area of specific 3A4 substrates because of its volume as well as of the IE_{tot} values that are always significantly lower than -50 kcal/mol, independently of the interacting cytochrome (IE_{tot} values of the complexes of **36** with 1A2 and 3A4 are, respectively, -75.52 kcal/mol and -84.90 kcal/mol).

(c) Moreover, size and shape molecular descriptors like Vmol and S_{phi}/S_{tot} may be additionally used to improve the predictive power of 1A2/3A4 specificity models.

(3) Where the above conditions are not unequivocally fulfilled, the new compound should be predicted to be non-specific.

These observations are satisfactorily consistent with the molecular and intermolecular interaction features displayed by the non-specific compounds considered in this study (Scheme 1; Table 4).

In conclusion, the combined study of the structural features of the three eukaryotic P450 cytochromes as well as of the specific substrates in their free and bound forms provides new insights into the molecular determinants of substrate specificity for three different isozymes and was instrumental to build semi-quantitative models useful for predicting the specificity of new compounds.

Acknowledgements

Financial support from CNR and technical support from CICAIA (University of Modena) are acknowledged.

References

1. Woggon, W.D., Top. Curr. Chem., 184 (1997) 39, and references therein.
2. Lewis, D.F.V., In Lipkowitz, K.B. and Boyd, D.B. (Eds.), Reviews in Computational Chemistry, VCH Publishers, Inc. New York, NY, 1992, pp. 173–222, and references therein.

3. Guenguerich, F.P., FASEB J., 6 (1992) 745.
4. Poulos, T.L., Finzel, B.C. and Howard, A.J., Biochemistry, 25 (1986) 5314.
5. Poulos, T.L., Finzel, B.C. and Howard, A.J., J. Mol. Biol., 195 (1987) 687.
6. Raag, R. and Poulos, T.L., Biochemistry, 30 (1991) 2674.
7. Ravichandran, K.G., Boddupalli, S.S., Hasemann, C.A., Peterson, J.A. and Deisenhofer, J., Science, 261 (1993) 731.
8. Hasemann, C.A., Ravichandran, K.G., Peterson, J.A. and Deisenhofer, J., J. Mol. Biol., 236 (1994) 1169.
9. Cupp-Vickery, J.R. and Poulos, T.L., Struct. Biol., 2 (1995) 144.
10. Nelson, D.R. and Strobel, H.W., J. Biol. Chem., 263 (1988) 6038.
11. Nelson, D.R. and Strobel, H.W., Biochemistry, 28 (1989) 656.
12. Edwards, R.J., Murray, B.P., Boobis, A.R. and Davies, D.S., Biochemistry, 28 (1989) 3762.
13. Black, S.D., FASEB J., 6 (1992) 680 and references therein.
14. Pernecky, S.J., Larson, J.R., Philpot, R.M. and Coon, M.J., Proc. Natl. Acad. Sci. USA, 90 (1993) 2651.
15. Black, S.D. and Martin, S.T., Biochemistry, 33 (1994) 6945.
16. Van den Broek, P.J.A., Barroso, M. and Lechner, M.C., Experientia, 52 (1996) 851.
17. Lozano, J.J., Lopez-de-Brinas, E., Centeno, N.B., Guigò, R. and Sanz, F., J. Comput.-Aided Mol. Design, 11 (1997) 395.
18. Lewis, D.F.V. and Lake, B.G., Xenobiotica, 26 (1996) 723.
19. Modi, S., Gilham, D.E., Sutcliffe, M.J., Lian, L.-Y., Primrose, W.U., Wolf, C.R. and Roberts, G.C.K., Biochemistry, 36 (1997) 4461.
20. Modi, S., Paine, M.J., Sutcliffe, M.J., Lian, L.-Y., Primrose, W.V., Wolf, C.R. and Roberts, G.C.K., Biochemistry, 35 (1996) 4540.
21. Lewis, D.F.V., Eddershaw, P.J., Goldfarb, P.S. and Tarbit, M.H., Xenobiotica, 27 (1997) 318.
22. Szklarz, G.D. and Halpert, J.R., J. Comput.-Aided Mol. Design, 11 (1997) 265.
23. Lewis, D.F.V., Eddershaw, P.J., Goldfarb, P.S. and Tarbit, M.H., Xenobiotica, 26 (1996) 1067.
24. Chang, Y.-T., Stiffelmann, O.B., Vakser, I.A., Loew, G.H., Bridges, A. and Waskell, L., Protein Eng., 10 (1997) 119.
25. Szklarz, G.D. and Halpert, J.R., Life Sci., 61 (1997) 2507 and references therein.
26. De Groot, M.J. and Vermeulen, N.P.E., Drug. Metab. Rev., 29 (1997) 747 and references therein.
27. Lewis, D.F.V. and Lake, B.G., Xenobiotica, 27 (1997) 443.
28. Lewis, D.F.V., Bird, M.G. and Parke, D.V., Toxicology, 118 (1997) 93.
29. Lewis, D.F.V., Dickins, M., Weaver, R.J., Eddershaw, P.J., Goldfarb, P.S. and Tarbit, M.H., Xenobiotica, 28 (1998) 235.
30. Smith, D.A., Ackland, M.J. and Jones, B.C., Drug Discov. Today, 2 (1997) 479.
31. Smith, D.A. and Jones, B.C., Biochem. Pharmacol., 44 (1992) 2089, and references therein.
32. Wolff, T., Distlerath, L.M., Warthington, M.T., Groopman, J.D., Hammons, G.J., Kadlubar, F.F., Prough, R.A., Martin, M.V. and Guengueric h, F.P., Cancer Res., 45 (1985) 2116.
33. Meyer, U.A., Gut, J., Kronback, T., Skoda, C., Meier, U.T., Catin, T. and Dayer, P., Xenobiotica, 16 (1986) 449.
34. Sali, A. and Blundell, T.L., J. Mol. Biol., 234 (1993) 799.
35. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., J. Mol. Biol., 112 (1977) 535.
36. Rost, B. and Sander, C., Proteins, 19 (1994) 55.
37. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M., J. Comput. Chem., 4 (1983) 187.
38. Allen, F.H., Kennard, O. and Taylor, R., Acc. Chem. Res., 16 (1983) 146.
39. Dewar, M.J.S., Zoebisch, E., Healy, E.F. and Stewart, J.J.P., J. Am. Chem. Soc., 107 (1985) 3902.
40. Schoichet, B.K., Bodian, D.L. and Kuntz, I.D., J. Comput. Chem., 13 (1992) 380.
41. Meng, E.C., Schoichet, B.K. and Kuntz, I.D., J. Comput. Chem., 13 (1992) 505.
42. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta Jr., S. and Weiner, P.A., J. Am. Chem. Soc., 106 (1984) 765.
43. Gasteiger, J. and Marsili, M., Tetrahedron, 36 (1979) 3219.
44. Makino, S. and Kuntz, I.D., J. Comput. Chem., 18 (1994) 1812.
45. Leach, A.R. and Kuntz, I.D., J. Comput. Chem., 13 (1992) 730.
46. Mackman, R., Tschirret-Guth, R.A., Smith, G., Hayhurst, G.P., Ellis, S.W., Lennard, M.S., Tucker, G.T., Wolf, C.R. and Ortiz de Montellano, P.R., Arch. Biochem. Biophys., 331 (1996) 134.
47. Yeom, H., Sligar, S.G., Li, H., Poulos, T.L. and Fulco, A.J., Biochemistry, 34 (1995) 14733.
48. Counts Gerber, N. and Sligar, S.G., J. Biol. Chem., 269 (1994) 4260.
49. Oscarson, M., Hidestrand, M., Johansson, I. and Ingelman-Sundberg, M., Mol. Pharmacol., 52 (1997) 1034.
50. He, Y.A., He, Y.Q., Szklarz, G.D. and Halpert, J.R., Biochemistry, 36 (1997) 8831.
51. Domansky, T.L., Liu, J., Harlow, G.R. and Halpert, J.R., Arch. Biochem. Biophys., 350 (1998) 223.
52. Wang, H., Dick, R., Yin, H., Licad-Coles, E., Kroetz, D.L., Szklarz, G., Harlow, G., Halpert, J.R. and Correia, M.A., Biochemistry, 37 (1998) 12536.
53. Hiroya, K., Murakami, Y., Shimizu, T., Hatano, M. and Ortiz de Montellano, P.R., Arch. Biochem. Biophys., 310 (1994) 397.
54. Hishigooka, M., Shimizu, T., Hiroya, K. and Hatano, M., Biochemistry, 31 (1992) 1528.
55. Furuya, H., Shimizu, T., Hirano, K., Hatano, M. and Fujii-Kuriyama, Y., Biochemistry, 28 (1989) 6848.
56. Mayuzumi, H., Sambongi, C., Hiroya, K., Shimizu, T., Tateishi, T. and Hatano, M., Biochemistry, 32 (1993) 5622.
57. Krainev, A.G., Shimizu, T., Hiroya, K. and Hatano, M., Arch. Biochem. Biophys., 298 (1992) 198.
58. Mayuzumi, H., Shimizu, T., Sambongi, C., Hiroya, K. and Hatano, M., Arch. Biochem. Biophys., 310 (1994) 367.
59. Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M., J. Appl. Crystallogr., 26 (1983) 283.
60. Vriend, G., J. Mol. Graphics, 8 (1990) 52.
61. Gotoh, O., J. Biol. Chem., 267 (1991) 83.
62. Luedemann, S.K., Carugo, O. and Wade, R.C., J. Mol. Model., 3 (1997) 369.
63. Dai, R., Pincus, M.R. and Friedman, F.K., J. Protein Chem., 17 (1998) 121.
64. Smith, G., Modi, S., Pillai, I., Lian, L.-Y., Sutcliffe, M.J., Pritchard, M.P., Friedberg, T., Roberts, G.C.K. and Wolf, C.R., Biochem. J., 331 (1998) 783.
65. Kraulis, P.J., J. Appl. Crystallogr., 24 (1991) 946.