Chemical Proteomic Tool for Ligand Mapping of CYP Antitargets: An NMR-Compatible 3D QSAR Descriptor in the *Heme-Based Coordinate System*

Huili Yao, Aurora D. Costache, and Daniel S. Sem*

Chemical Proteomics Facility at Marquette, Chemistry Department, P.O. Box 1881, Marquette University, Milwaukee, Wisconsin 53201

Received September 15, 2003

Chemical proteomic strategies strive to probe and understand protein−ligand interactions across gene families. One gene family of particular interest in drug and xenobiotic metabolism are the cytochromes P450 (CYPs), the topic of this article. Although numerous tools exist to probe affinity of CYP−ligand interactions, fewer exist for the rapid experimental characterization of the structural nature of these interactions. As a complement to recent advances in X-ray crystallography, NMR methods are being developed that allow for fairly high throughput characterization of protein−ligand interactions. One especially promising NMR approach involves the use of paramagnetic induced relaxation effects to measure distances of ligand atoms from the heme iron in CYP enzymes. Distances obtained from these T_1 relaxation measurements can be used as a direct source of 1-dimensional structural information or to restrain a ligand docking to generate a 3-dimensional data set. To facilitate such studies, we introduce the concept of the *Heme-Based Coordinate System* and present how it can be used in combination with NMR T_1 relaxation data to derive 3D QSAR descriptors directly or in combination with in silico docking. These descriptors should have application in defining the binding preferences of CYP binding sites using 3D QSAR models. They are especially well-suited for the biasing of fragment assembly and combinatorial chemistry drug design strategies, to avoid fragment or reagent combinations with enhanced affinity for CYP antitargets.

1. INTRODUCTION

The high cost of drug discovery and development is due largely to the failure of drugs in clinical trials due frequently to toxicity from binding to other protein targets ("antitargets") and poor ADMET (adsorption, distribution, metabolism, excretion, and toxicology) properties, causing either toxic effects or less than adequate drug efficacy.[1] Such adverse effects are often due to binding to the cytochromes P450 (CYPs) and can include drug−drug interactions. Thus, there has been a recent emphasis placed on predicting ADMET properties to avoid such interactions in the design of molecules from the outset. But, since there is currently no reliable way to predict this behavior a priori, there is a need for a database of empirically derived descriptors that are predictive of CYP binding. Although structures for mammalian CYPs in general[2,3] and particularly for human CYPs bound to drugs are few,[4] many studies have reported on the docking of drugs into homology modeled CYP binding sites.[5,6] Perhaps the prototype CYP enzyme for such studies is CYP2D6, which is responsible for the metabolism of 30% of all drugs. Computational approaches have combined homology modeling of CYP2D6 with the construction of pharmacophore models from docked substrates.[7,8] These and other strategies have recently been reviewed,[9] but it has been noted by deGroot and Ekins[10] that such studies would benefit from more in vitro data generated for more CYP enzymes. Methods such as NMR-based docking offer the possibility of creating such an expanded data set of docked structures to better enable 3D QSAR analyses. The inclusion of NMR structural data[11] to restrain the docking process is a recent improvement over purely in silico docking methods. Particularly exciting results have been obtained using NMR T_1 relaxation data to dock ligands and probe interactions with CYP2D6,[8,11] with the goal of improving ADMET properties for drugs that bind this enzyme.

Current Predictive ADMET Strategies. In silico strategies to predict ADMET have included the empirical matching of certain chemical properties (such as molecular weight) of drug candidates to existing drugs[12,13] as well as QSAR-based toxicity assessment programs that include CASE[14] and TOPKAT.[15] Rule-based strategies such as DEREK[16] are also widely used. But, as noted recently by Durham and Pearl,[17] current in silico toxicity prediction strategies are limited due in part to inadequate means for validating the accuracy of the models. Such validation would lead to more robust models and hopefully increase predictability, in line with the national toxicology program challenge.[18] Clearly, to truly enable prediction of drug metabolism and interactions, there is a need for a database of "receptor relevant"[19] molecular properties or descriptors, including experimentally validated binding affinitys and toxicophore maps for the major CYPs.

Herein we propose such a descriptor in the *Heme-Based Coordinate System*. The descriptor has the advantage of being (a) derived based on experimental data that can be generated easily for most CYPs, (b) easily expanded to a more detailed descriptor using docking data, (c) in a form that can be used to verify descriptors derived based purely on in silico docking, (d) readily converted to a binary form that can be used in genetic, clustering or other algorithms, to map CYP

* Corresponding author phone: (414)288-7859; e-mail: daniel.sem@marquette.edu.

10.1021/ci034208q CCC: $27.50
Published on Web 05/27/2004

Table 1. Constants Used in the Solomon−Bloembergen Equation

abbreviation	name	value[a]
μ_B	Bohr magneton	$9.27400899 \times 10^{-24}$ J T^{-1}
μ_o	magnetic permeability[b]	$4\pi \times 10^{-7}$ T^2 J^{-1}m^3
g_e	electron *g*-factor	−2.0023193043718
γ_I	proton gyromagnetic ratio[c]	2.67522205×10^8 s^{-1} T^{-1}

[a] Values for these physical constants are from Mohr and Taylor[23] and are summarized at http://physics.nist.gov/PhysRefData/contents.html. [b] More explicitly, the magnetic permeability of free space, also known as vacuum permeability. Units of T^2 J^{-1} m^3 are equivalent to units of N/A^2. This constant is often and erroneously left out of the equation in the biochemical literature. [c] If relaxation effects are being monitored on a nucleus other than the proton, then substitute the appropriate gyromagnetic ratio for that nucleus.

binding sites, and (e) useful in fragment assembly approaches and therefore useful for biasing the construction of combinatorial libraries against CYP binding.

2. METHODS

2.1. NMR-Derived Distances Using T_1 Relaxation Studies. Since mammalian CYP enzymes are not easily crystallized, the T_1-relaxation-based NMR strategy currently provides the only practical means of populating a large database with experimental structural data on binding interactions with CYPs. Such a database of descriptors would ultimately contain a large set of experimentally verified 3D shape descriptors for interactions with CYP enzymes and would therefore allow more accurate QSAR models to be built. This would enable better prediction of drug−drug interactions and metabolic potential, especially when combined with databases of metabolic pathways and metabolic rule dictionaries. Establishing binding and orientation in CYP substrate sites with NMR is typically based on T_1 relaxation effects induced by the paramagnetic heme iron.[11] The relationship between longitudinal relaxation rate ($1/T_{1,M}$) and distance is provided by the Solomon−Bloembergen equation[20−22]

$$1/T_{1,M} = (2/15)(\mu_o/4\pi)^2(\gamma_I{}^2 g_e{}^2 S(S+1)\mu_B{}^2/r^6) \times [3\tau_c/(1+\omega_I{}^2\tau_c{}^2) + 7\tau_c/(1+\omega_S{}^2\tau_c{}^2)] \quad (1)$$

where S is the total spin present on the metal ion, ω_I and ω_s are the Larmor frequencies (in radians/second) at the particular field strength for protons and electrons, respectively, and other constants are as described in Table 1. This equation assumes that scalar interactions are negligible, which is a reasonable assumption in most cases when a molecule is not in the coordination sphere of the iron. The above equation is usually further simplified since $\omega_s > \omega_I$, which makes the second term in square brackets become insignificant. τ_c is the correlation time of the dipolar interaction, which is obtained for a given protein by measuring the frequency dependence of $1/T_1$. It should be noted that one can measure an accurate T_1 value only if ligand binding and release (exchange) is fast relative to the T_1 value. That is, the system needs to be in the "fast exchange" time regime with $\tau_m < T_{1,M}$, where $1/\tau_m$ is the exchange rate for ligand binding to the CYP enzyme. A simple test to see if the system is in the fast exchange regime is to determine temperature dependence of $T_{1,obs}$ (measured T_1 value). If $1/T_{1,obs}$ decreases with increasing temperature, then there is fast exchange.[21,22] Values for $T_{1,obs}$ are obtained using a standard inversion recovery pulse sequence, and fitting the exponential recovery of equilibrium magnetization (M_0), monitoring integrated intensity of the NMR signal (M_z) for the ligand proton of interest at various delay times (τ) allowing for magnetization recovery:

$$M_z = M_0(1 - 2\exp(-\tau/T_{1,obs})) \quad (2)$$

A nonlinear least-squares fit of measured M_z values at different values of τ then gives $T_{1,obs}$.

The measured value of longitudinal relaxation ($T_{1,obs}$) is not used directly in the Solomon−Bloembergen equation to obtain distances. One needs to first determine the paramagnetic contribution ($T_{1,p}$) to the longitudinal relaxation effect by (a) extrapolating to a lower ratio of [ligand]/[CYP] in a titration curve and (b) correcting for diamagnetic contributions to $T_{1,obs}$, by determining relaxation effects for a CYP enzyme that is in a diamagnetic state (reduced with dithionite and bound to CO)[11]

$$(1/T_{1,obs}) - (1/T_{1,diam}) = \{[CYP]/(K_d + [ligand])\}(1/T_{1,P} - 1/T_{1,free}) + 1/T_{1,free} \quad (3)$$

where free indicates for unbound ligand, diam indicates for diamagnetic control, [CYP] and [ligand] are total concentrations of CYP and ligand, respectively, and K_d is the dissociation constant for ligand binding to CYP. In practice, it is easiest to keep [CYP] constant at a low concentration like 10−50 μM and vary [ligand].

After simplification, the Solomon−Bloembergen equation can be rearranged to a form that provides a more transparent relationship between $T_{1,M}$ (equivalent to $T_{1,p}$) values and the distance (r) between a proton on the ligand and the paramagnetic iron atom:

$$r = C \cdot \{[S(S+1)] \cdot T_{1,p} \cdot [3\tau_c/(1+\omega_I{}^2\tau_c{}^2)]\}^{1/6} \quad (4)$$

with

$$C = [(2/15)(\mu_o/4\pi)^2(\gamma_I{}^2 g_e{}^2 \mu_B{}^2)]^{1/6} = 566\ \text{Å} \cdot \text{s}^{-1/3} \quad (5)$$

Thus, if one measures a value for $T_{1,p}$ as described above, all that is needed is to determine the spin-state of iron (5/2 if high spin in the complex with non-iron bound ligand) and the correlation time of the protein, then use the Larmor frequency for the particular field strength and the above value of C to calculate distance. As a cautionary note, in addition to the above assumptions made in obtaining this equation, there is an implicit assumption that spin density can be treated as entirely localized on the heme iron. But, if spin density is delocalized into orbitals on nearby atoms, then it will be necessary to calculate an electron spin density matrix and to consider distance as an effective distance (r_{eff}).[24]

2.2. Docking into Homology Modeled CYP Binding Sites. The distances obtained in the above NMR experiments permit restrained docking[11] into homology modeled CYP binding sites,[25,26] as was done for codeine in CYP2D6.[11] As a prelude to such studies, we have obtained a homology modeled CYP2D6 structure using SWISS-MODEL[27−29] with rabbit CYP2C5 as template. It has recently been reported that use of the CYP2C5 template offers a significant improvement over the bacterial templates.[30] Prior to our docking studies, both ligand and homology-modeled protein

were minimized using the MMFF94 force field.[31,32] Then, docking was performed using the MOE (Molecular Operating Environment) software package (Chemical Computing Group) and the same MMFF94 force field, with ligand flexible and protein kept rigid. We performed 25 docking runs using random starting orientations. Optimization was with simulated annealing, using an initial temperature of 1000 K and 6 cycles per run. Docking calculations included protein atoms within a 53 Å × 53 Å × 53 Å box surrounding the binding pocket above the heme. For the docking of codeine, the docking box with docked codeine was then further minimized under AMBER[33] with a distance restraint between the positively charge codeine amine and the Oδ of Asp301 (4.0 Å upper-bound) and between iron and the CH_3O- that gets oxidized (2.5 Å−3.5 Å). This restrained minimization was used to refine the docked codeine structure, which was initially docked with no restraints. Docking studies with combinatorial chemistry reagents that contained amines had the restraint to Asp301 included as part of the initial docking process, since they are less likely to have enough binding energy to orient in a relevant binding orientation without the added restraint. For one of the reagents (compound 1823), the docking was done with (1823) and without (1823′) this added restraint, for comparison purposes. Correct binding orientation awaits experimental verification with NMR. Although docking was done initially in the absence of NMR structural data, NMR-derived distances will be obtained later either to (a) validate a purely in silico docked structure or (b) serve as restraints to guide the docking process. Ligands that are docked in the CYP binding site can be used to generate a 3D shape descriptor. If descriptors are obtained for combinatorial chemistry reagents, they can be used to score which combinations of reagents in a combinatorial library could yield CYP binding. The joining of weak binding ligands is known to provide a tremendous boost in affinity through the "chelate effect",[34] and in an analogous manner complementary fragments with very low affinity to a given CYP enzyme might be expected to produce higher affinity for that CYP enzyme when joined. Thus, we hypothesize that just as the chelate effect could be used to build drugs from weak binding fragments, the same logic should allow us to predict which combinations of fragments should *not* be brought together, because they would likely generate high affinity for a CYP. Such biasing is based on the exact same chelate effect concept but now applied to antitargets.

2.3. NMR-Based 3D-QSAR Descriptor. The descriptor we are presenting herein is intentionally chosen to describe a docked structure in a way that is most consistent with the type of NMR distance data we are generating. That is, it views toxicophore features of a ligand as distances from the heme iron (raw NMR data) and in polar coordinates with the heme iron defining the origin (Figures 1 and 2). The descriptors can be derived from combinatorial chemistry reagents docked into CYP binding sites (3-dimensional) as well as from NMR T_1 relaxation data (1-dimensional). We are developing a database of these experimentally verifiable structural descriptors for various combinatorial chemistry building blocks from ChemDiv docked into CYPs, to provide a consistent and reliable toxicophore QSAR model. The 3D QSAR descriptors are of two types: (a) based on purely in silico docking and (b) based on docking restrained with the *NMR T_1 relaxation experimental* data described above. The former can be done on a large scale for thousands of compounds, and the latter will be performed experimentally with moderate throughput.

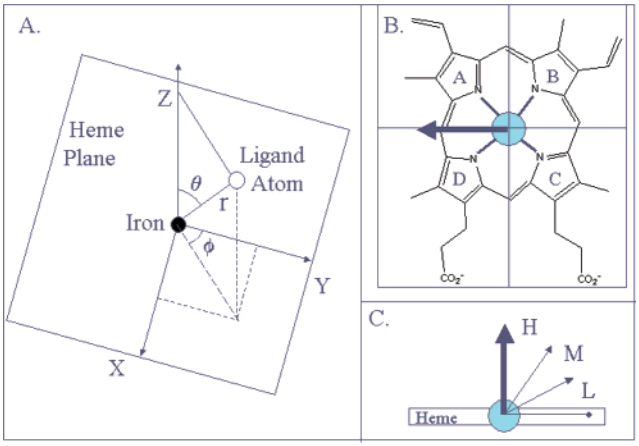
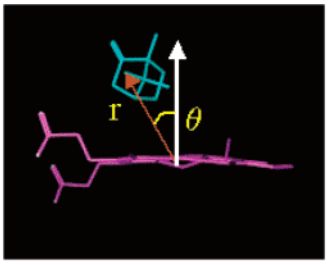


Figure 1. Description of the *Heme-based Coordinate System* (*HCS*). (A) Polar coordinates *r*, ϕ, and θ defined relative to the heme plane. (B) The angle ϕ in the *HCS* is relative to a vector pointing from the heme iron to the methine between pyrroles D and A and is binned into one of four quadrants defined as A (0−90°), B (90−180°), C (180−270°), or D (270−360°). Lettering for pyrrole rings is standard nomenclature.[39] (C) The angle θ is binned as H for high (0−30°), M for medium (30−60°), and L for low (60−90°). Distance *r* is from heme iron to the ligand atom and is binned downward (for example, "6" means anywhere from 6 to 6.9).

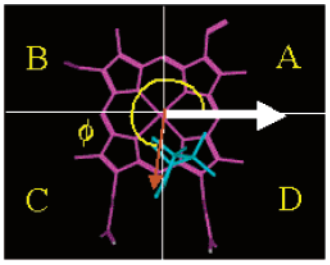
3. RESULTS AND DISCUSSION

3.1. The 3D QSAR Descriptor. The central feature of our studies is thus development of this novel 3D QSAR descriptor for defining toxicophores with greater accuracy than with purely in silico strategies. Our strategy builds on traditional QSAR[35,36] but uses shape features[37] that define "receptor relevant"[19] descriptors that can to be used in a principle component analysis (PCA) or related strategy to predict binding to CYP enzymes. What is unique about our descriptor is (a) its close integration of experimental (NMR) and computational (docking) methods, (b) that it gives access to *experimental* CYP structural data, and (c) that it will be based on a large and expanding data set where all data is obtained on well-defined protein systems with a consistent experimental format. The descriptor we propose is designed to have certain features, such as being (a) experimentally obtained, (b) able to be refined to higher resolution if the NMR data is used to dock the structure into a modeled CYP, (c) able to be adjusted so it can be used in commonly used 3D QSAR strategies such as CoMFA,[38] and (d) easily stored in a relational database and integrated with other data sources. Since the NMR T_1 relaxation data provides only distance from the paramagnetic iron for a given proton on a ligand, a simple vector could be created that is comprised of distances of all ligand protons from the iron, by specifying the atom type (toxicophore feature) for each in one of 15 distance bins (2−16 Å), as shown in Figure 3A (NMR TF T_1 Vector). If the NMR data were used to restrain the computational docking of a ligand, additional, albeit less reliable, information about location in the CYP binding site could be obtained (columns in Figure 3B). Thus, while NMR data provides iron-to-ligand atom distances (*r* in Figure 1),

A.



B.



C.

Ligand Atom	r, Å	θ	ϕ	TF	HCS (r)	HCS (θ)	HCS (ϕ)
H3n	6.1	46	256	L	6	M	C
H3x	6.6	48	276	L	6	M	D
H4	4.1	50	285	L	4	M	D
H5n	4.3	36	227	L	4	M	C
H5x	3.3	14	247	L	3	H	C
H6n	6.2	18	231	L	6	H	C
H6x	5.5	3.2	193	L	5	H	C
H8	6.6	41	325	L	6	M	D
H9	5.0	32	345	L	5	M	D
H10	7.6	14	326	L	7	H	D
O	8.1	29	264	A	8	H	C

D.

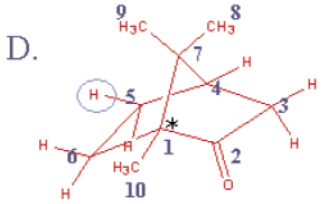


Figure 2. Illustration of the *HCS* using the cytochrome P450cam − camphor structure. (A) The heme and camphor ligand were extracted from the crystal structure of the Fe^{3+} complex[40] (pdb code 2cpp) and used to demonstrate measurement of distance *r* and angle θ in the *HCS*. Angle is relative to a vector projecting from the iron and orthogonal to the heme plane. (B) Demonstration of the measurement of the ϕ angle and binning (quadrant C for this atom). Note that the view is from above the face of the heme where substrate binds, which is opposite of that shown in Figure 1 (so lettering is in reverse). (C) Result of measuring *r*, θ, and ϕ for each atom in camphor, as numbered in (D). Each atom is classified with a toxicophore feature (Table 2) and binned in the *HCS* as described in Figure 1.

A.

	TF
2 Å	
3 Å	L
4 Å	L
5 Å	L
6 Å	L
7 Å	L
8 Å	A
9 Å	
10 Å	
11 Å	
12 Å	
13 Å	
14 Å	
15 Å	
16 Å	

Computational Docking

B.

	AL	AM	AH	BL	BM	BH	CL	CM	CH	DL	DM	DH
2 Å												
3 Å									L			
4 Å								L			L	
5 Å									L		L	
6 Å								L	L		L	
7 Å												L
8 Å									A			
9 Å												
10 Å												
11 Å												
12 Å												
13 Å												
14 Å												
15 Å												
16 Å												

Figure 3. Increasing the information content of the 3D QSAR descriptor derived from NMR data by computational docking. (A) In the vector descriptor based only on NMR data, all information is experimentally verified, but only distances of toxicophore features from the iron atom are captured. (B) Docking allows the expansion of structural data into two more dimensions in a polar coordinate system, represented here in a 2D matrix. Camphor atoms from the structure in Figure 2 are indicated in this matrix by white boxes and a letter indicating the toxicophore feature (Table 2).

restrained docking adds two polar coordinates (ϕ and θ in Figure 1) to place ligand atoms in 3-dimensional space in the *HCS*.

This descriptor describes toxicophore feature atoms (Table 2) on a ligand within a *Heme-based Coordinate System* (*HCS*). One way to view the *HCS* is as a polar coordinate system with iron at the origin and position defined by distance from iron (obtained experimentally with NMR) as well as two angles defined using the heme plane (obtained by computational docking). These two angles (Figure 1A) are measured relative to (a) a vector pointing toward the methine between pyrrole rings D and A (Figure 1B) and (b) a vector orthogonal to the heme plane (Figure 1C). For expediency we bin the angles in rather large bins, since they will be less accurately known than the NMR-derived distances. The angle within the heme plane is 0−90° (heme quadrant A), 90−180° (heme quadrant B), 180−270° (heme quadrant C), or 270−360° (heme quadrant D). The angle above the heme plane defined relative to the orthogonal vector is 0−30° (H for high), 30−60° (M for medium), and 60−90° (L for low). The descriptor has distances from heme iron in 2−16 Å bins for the rows and pairs of polar coordinate angles defining the columns (Figure 3B). The entry in each matrix position is one (or more) of six

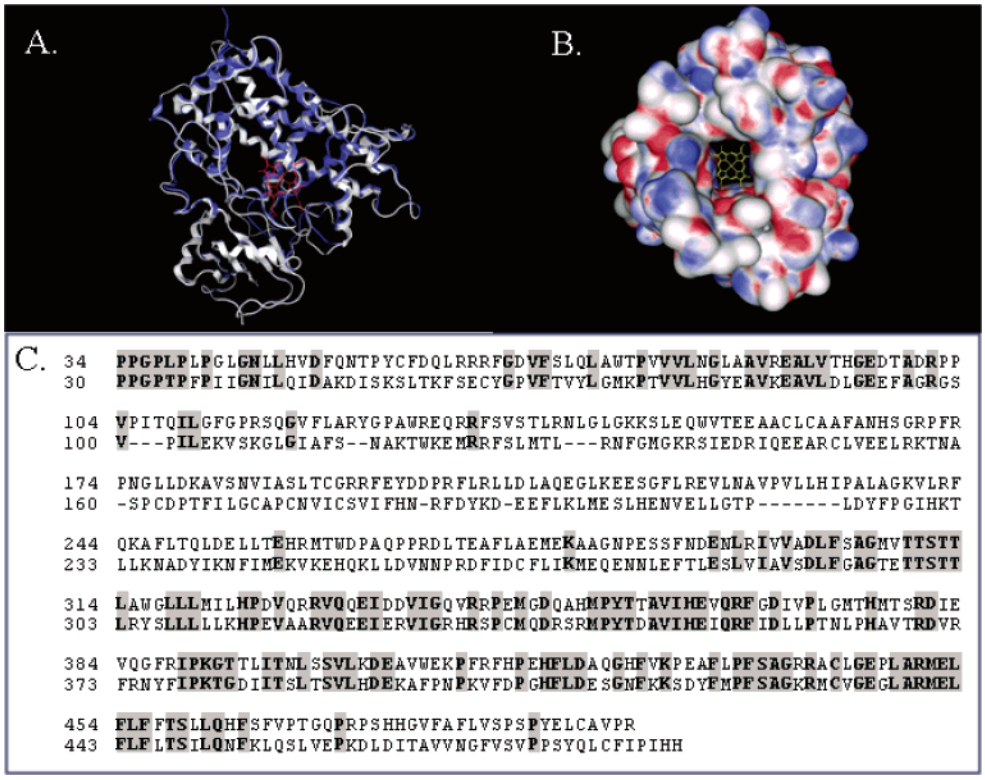


Figure 4. (A) Homology model of CYP2D6 (white) overlaid on the crystal structure of rabbit CYP2C5 (blue: pdb code 1dt6).[2] Model was constructed with SWISS− MODEL.[27−29] (B) Charge surface (red = negative; blue = positive) of the modeled CYP2D6 prepared using MOE software (Chemical Computing Group), with a rectangular slice through the protein removed to make the heme prosthetic group visible. The heme defines the *HCS*, as described in Figures 1 and 2. (C) Sequence alignment of homology modeled human CYP2D6 on the rabbit CYP2C5 template (pdb code 1dt6) used to generate the model. Alignment is according to the structural overlay in Figure 4A.

Table 2. Toxicophore Features[a]

abbreviation	feature description
N	negative charge[b]
P	positive charge[b]
L	aliphatic
R	aromatic
D	hydrogen bond donor[b] (neutral)
A	hydrogen bond acceptor[b] (neutral)

[a] When protons are not clearly resolved by NMR, as with methyl protons, pseudoatoms are used. In the case of methyl protons, the carbon of the methyl group serves as the pseudoatom. [b] Toxicophore feature is within two bonds of the hydrogen atom.

toxicophore features for the atom, classified as described in Table 2. A descriptor matrix such as this can be created either with the benefit of NMR data or based on purely in silico docking.

For the purpose of illustration, the *HCS* is used to define a 3D QSAR descriptor from the crystal structure of the camphor-P450cam complex. Distance r and angle θ are determined as illustrated for a camphor atom in Figure 2A, and the angle ϕ is determined as in Figure 2B. The values for r, θ, and ϕ are then binned in the HCS as described in Figure 1. The binned values are given in the table in Figure 2C for all camphor atoms, as defined in Figure 2D. After classifying each camphor atom as the appropriate toxicophore feature (Table 2), the HCS coordinates in Figure 2C are used to construct the 3D QSAR matrix in Figure 3B. If we had only NMR T_1 data rather than a full structure, we could have constructed the condensed vector shown in Figure 3A or a variation of that vector for each individual camphor atom. Extension of the NMR data with docking would then have allowed generation of a 3D QSAR descriptor like that in Figure 3B. Such NMR studies are currently underway.

3.2. Homology Modeled CYP2D6. Although P450cam is an excellent model system due to the wealth of structural and mechanistic information available for it, our primary interest from a predictive ADMET perspective is with the human drug metabolizing CYPs. Homology modeled structures have been reported for the major drug metabolizing CYPs, including CYP1A2,[41] CYP3A4,[42] CYP2C,[43] CYP2D6,[44] and CYP2E1.[45] Such studies were often focused on developing predictive rules for the binding of drugs. The CYP structures used as templates for homology modeling included initially P450cam[46] and later P450BM3,[47] although the recent solution of the first mammalian CYP structure[2] offered an even better template. Our studies using CYP2D6 relied on a homology model constructed using SWISS-MODEL.[27,28,29] The homology modeled CYP2D6 in Figure 4A has a Cα backbone RMSD difference compared to rabbit CYP2C5[2] of only 1.5 Å. Such structural similarity is not unexpected since sequence identity is 32%. This structure is used in the following docking studies. The heme located at the center of the CYP2D6 structure (Figure 4B) will serve as the coordinate system for defining how docked ligands are oriented in the binding site above the heme, according to the *HCS* definitions (Figures 1 and 2). A portion of the CYP2D6 structure has been removed, so the heme reference can be more easily visualized.

3.3. In Silico Docking into CYP2D6. A number of studies have been reported on homology modeling of CYP enzymes and the use of these models to predict drug binding and metabolism.[48−51] But, only limited studies have been reported

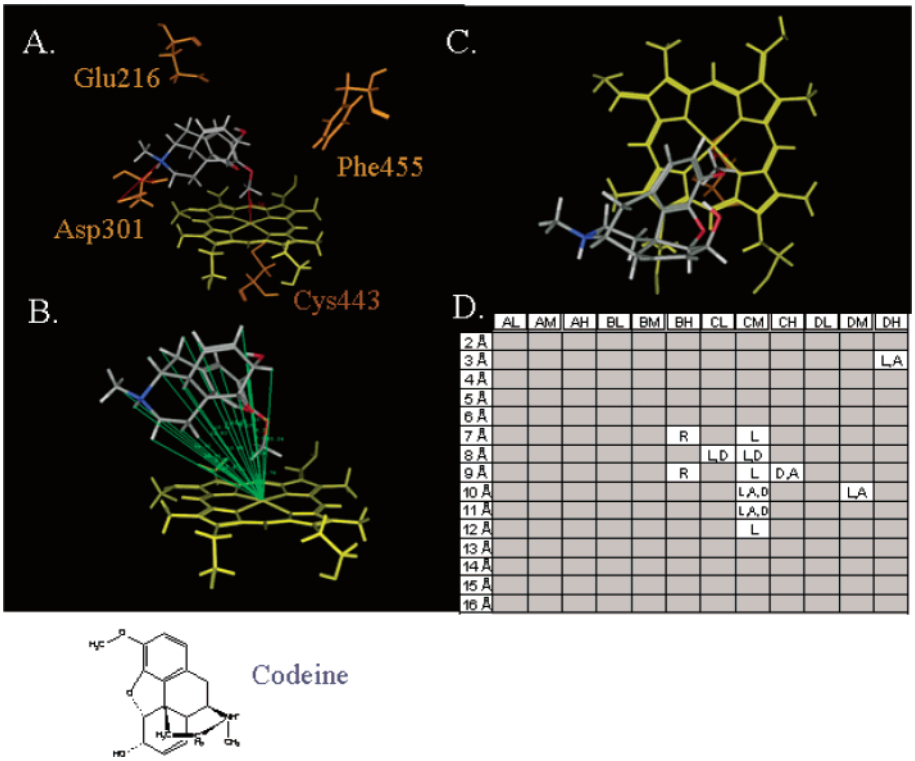


Figure 5. Computational docking of codeine into the homology modeled CYP2D6 structure. (A) An expansion of the active site region is shown, with proximity of the methyl group of codeine to the iron indicated as well as key binding site residues. Distances (*r*) of various atoms of the ligand are indicated relative to the heme iron in (B), and the projection of ligand onto the heme is shown in (C). These data were used to generate the 3D QSAR matrix shown in panel (D).

which rely on experimental structural data like NMR.[8,11] Eventually, our NMR T_1 data can be used to get distances of each ligand proton from the paramagnetic heme iron, which will be used to restrain the docking of drugs and drug building blocks (combinatorial chemistry reagents from ChemDiv) into CYP2D6, as was done by Modi et al. with codeine.[11] We have also docked codeine as well as 3 secondary amine combinatorial chemistry reagents into the modeled structure for CYP2D6. We are ultimately going to use NMR T_1 data to restrain the docking of additional drugs and drug building blocks and generate descriptors from these docked structures. As a prelude to these studies, and in order to refine our computational strategy, we have derived the descriptor for the purely in silico docked codeine (Figure 5). This descriptor will later be validated with NMR T_1 relaxation studies. Likewise, we have docked three combinatorial chemistry reagents from ChemDiv into CYP2D6 and calculated descriptors in Figures 6−8. As had been done previously by other labs performing docking studies with CYP2D6, we restrained the positively charged amines on the ligands to be within 4.0 Å of Oδ of Asp301. This residue is known by a number of studies[52,53] to play a role in stabilizing the positive charge on CYP2D6 substrates and inhibitors.

Given the tendency of CYP2D6 to interact with positively charged ligands, potential for CYP2D6 binding should be considered in the enumeration of any combinatorial library. We anticipate that CYP2D6 ionic interactions could occur frequently with compounds produced in combinatorial libraries, since a survey of over 650 000 compounds from the screening collections from three of the major suppliers indicates that there are many amines which are predicted to be positively charged at pH 7.4. Using the pK_a prediction algorithm contained in Pipeline Pilot (SciTegic, Inc.), we have found that these collections contained 0.6−2.5% primary amines, 5−6% secondary amines, and 5−20% tertiary amines, which would be positively charged at pH 7.4. Clearly it would be advantageous to have prescreened the combinatorial building block reagent combinations to avoid CYP2D6 binding. Using the NMR-derived 3D QSAR descriptor described herein, combinatorial reagents could be computationally screened to see which combinations are likely to yield enhanced CYP2D6 binding, and these combinations could be avoided in the enumeration process.

The three representative positively charged amine reagents we docked are compared in Figure 9A, and we have included a potential alternative docking/binding mode for one of these reagents, obtained upon dropping the restraint to Asp301 (1823′). NMR T_1 data is needed to determine exact binding mode of this and other potential ligands.

3.4. Applicability of the Matrix for QSAR Studies. The 15 × 12 matrix in Figures 5−8 can also be converted to a 180 × 1 matrix (now a vector) where columns contain all three (rather than just two) polar coordinates. This form of the descriptor allows one to compare ligands, as is commonly done using the field matrix of ComFA and other 3D QSAR tools. This form of the 3D QSAR descriptor could be created for each compound (Figure 9B). But for simplicity and purpose of illustration, we have shown only the 3−13 Å region of the C-quadrant. Furthermore, rather than indicating toxicophore features (Table 2), Figure 9C is a binary form for only one (L = aliphatic) of the six toxicophore features, with "1" indicating presence of the feature and "0" indicating absence. Rather than indicating specific features, it may on

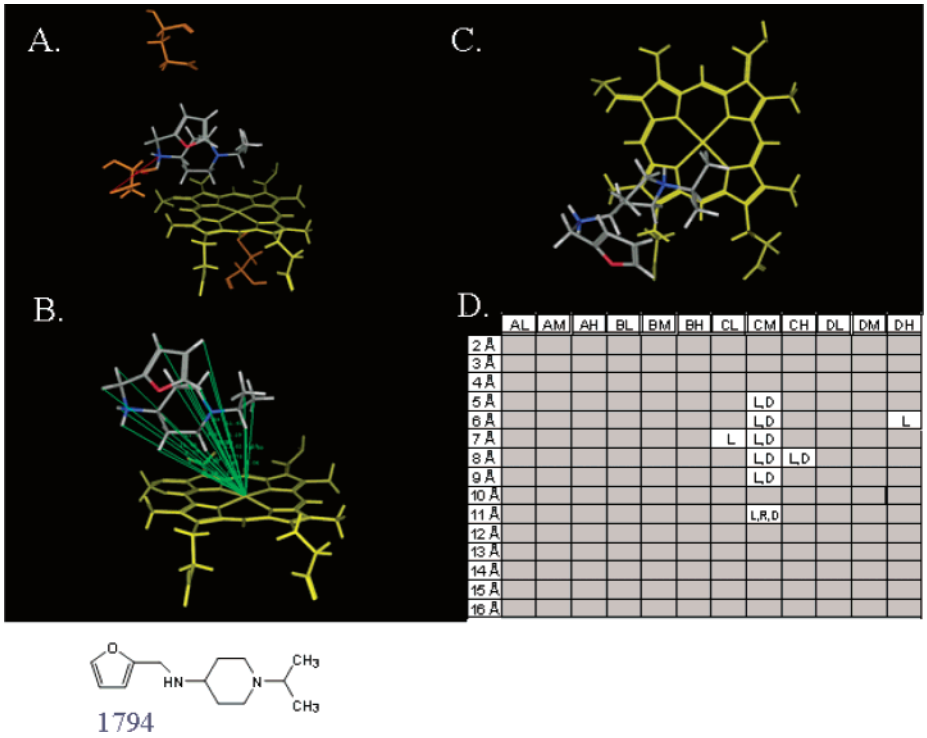


Figure 6. Computational docking of a combinatorial chemistry reagent (1794) into the homology modeled CYP2D6 structure. (A) An expansion of the active site region is shown, with key binding site residues indicated as in Figure 5. Distances of various atoms of the ligand are indicated relative to the heme iron in (B), and the projection of ligand onto the heme is shown in (C). These data were used to generate the 3D QSAR matrix shown in panel (D).

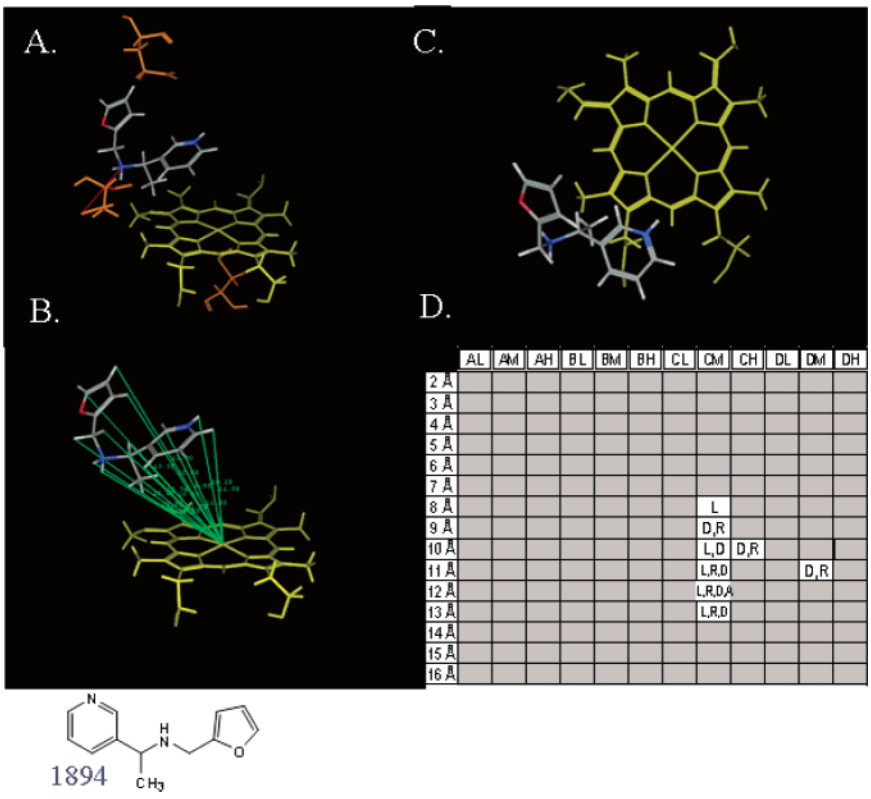
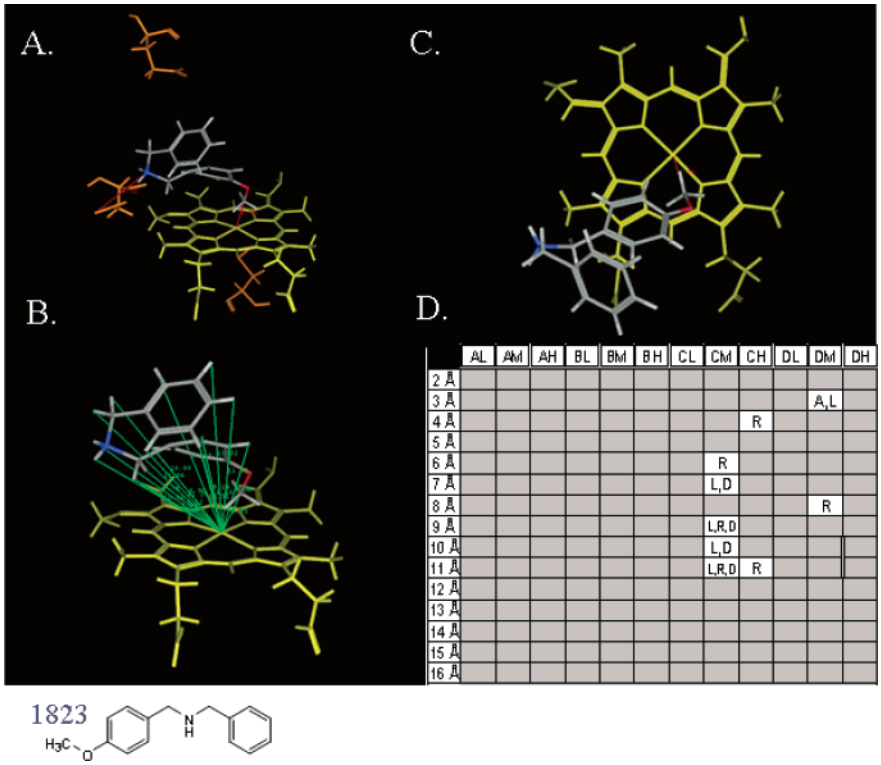


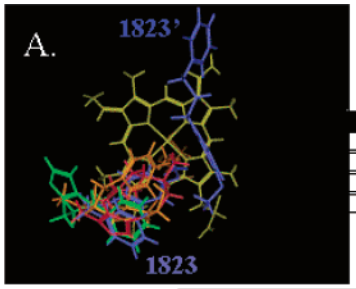
Figure 7. Computational docking of a combinatorial chemistry reagent (1894) into the homology modeled CYP2D6 structure. (A) An expansion of the active site region is shown, with key binding site residues indicated as in Figure 5. Distances of various atoms of the ligand are indicated relative to the heme iron in (B), and the projection of ligand onto the heme is shown in (C). These data were used to generate the 3D QSAR matrix shown in panel (D).

occasion be more useful to let the "1" indicate the presence of any feature, thereby capturing spatial information (irrespective of feature type). These binary forms of the 3D QSAR descriptor have many advantages, including the ability to use different combinations of them in a PCA analysis, depending on which is more predictive of CYP binding. Also,



	AL	AM	AH	BL	BM	BH	CL	CM	CH	DL	DM	DH
2 Å												
3 Å											A,L	
4 Å									R			
5 Å												
6 Å								R				
7 Å								L,D				
8 Å											R	
9 Å								L,R,D				
10 Å								L,D				
11 Å								L,R,D	R			
12 Å												
13 Å												
14 Å												
15 Å												
16 Å												

Figure 8. Computational docking of a combinatorial chemistry reagent (1823) into the homology modeled CYP2D6 structure. (A) An expansion of the active site region is shown, with key binding site residues indicated as in Figure 5. Distances of various atoms of the ligand are indicated relative to the heme iron in (B), and the projection of ligand onto the heme is shown in (C). These data were used to generate the 3D QSAR matrix shown in panel (D).



B.

	AL2	AL3	...	AL16	AM2	AM3	...	AM16	...	...	D16
Codeine											
Cpd 1794											
Cpd 1894											
Cpd 1823											

C.

	CL 3	CL 4	CL 5	CL 6	CL 7	CL 8	CL 9	CL 10	CL 11	CL 12	CL 13	CM 3	CM 4	CM 5	CM 6	CM 7	CM 8	CM 9	CM 10	CM 11	CM 12	CM 13	CH 3	CH 4	CH 5	CH 6	CH 7	CH 8	CH 9	CH 10	CH 11	CH 12	CH 13
Codeine	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Cpd 1794	0	0	0	0	1	0	0	0	0	0	0	0	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Cpd 1894	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
Cpd 1823	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Cpd 1823'	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 9. Summary of docked combinatorial chemistry reagents looking down on the heme (A). Another potential docking mode of one reagent is also indicated (1823′), and codeine was included as a reference molecule. (B) An abbreviated form of the matrix is presented where the 2D matrix representation shown in Figures 5−8 has been converted to a vector representation so that ligands can be easily compared. (C) The vector in (B) has been converted to binary form for the C quadrant only (the full vectors are too large to show) and for the L feature (aliphatic) only. There will exist corresponding vectors for each feature present in a molecule, so each could be used in a QSAR analysis.

and most importantly, columns represent spatial location in a binding site with proximity of columns roughly correlated with proximity in a binding site. Thus, it is easily determined which combinatorial chemistry reagents occupy different regions of the binding site with no or minimal overlap. Such combinations would be scored against in a library enumeration, since they would have enhanced likelihood of binding with higher affinity to a given CYP. The combination of a pair of combinatorial reagents is treated like a crossover event in a genetic algorithm,[54] and one then optimizes for the minimum number of "1"s, with crossover events only permitted into regions of "0"s, since synergistic CYP ligand pairs cannot occupy the same region of space. This could also be achieved by adding the bit strings but not adding combinations that give "2"s (occupy same part of binding sites) and selecting for the minimum number of "1"s. For example, in Figure 9C combining 1823′ and 1894 in the same molecule might produce enhanced binding to CYP2D6, since

they have "1"s in nonoverlapping regions of their bit strings. But, 1823 and 1894 would not produce enhanced affinity upon linkage, since they occupy overlapping regions in the *HCS*. This highlights the need for NMR data to distinguish binding modes (1823 versus 1823′). Conversely, genetic algorithm-based optimization of fragments that *do* bind CYPs could be used to define the binding site toxicophore. This toxicophore model could then be used for scoring in cases where NMR data are not available for the combinatorial reagent in question.

4. CONCLUSIONS

Since all CYP enzymes possess a heme prosthetic group at the base of their substrate binding site, we propose that a *Heme-Based Coordinate System* (*HCS*) may be the best way to describe the location of ligand atoms in a CYP binding site. The *HCS* provides a uniform way of viewing the CYP binding site, as described in Figures 1 and 2, and can be used to prepare a matrix that provides a complete description of the location of ligand atoms, classified as toxicophore features. This matrix can be used as is or in a condensed form that can be experimentally verified with simple NMR experiments that are capable of reasonably high throughput studies. Distances from the heme iron are easily obtained from these NMR T_1 relaxation studies, using the following condensed form of the Solomon−Bloembergen equation:

$$r = 566\ \text{Å} \cdot s^{-1/3}\{[S(S+1)] \cdot T_{1,\text{p}} \cdot [3\tau_c/(1+\omega_\text{I}^2\tau_\text{c}^2)]\}^{1/6} \quad (6)$$

These distances for all ligand atoms define a vector of toxicophore features (Figure 3A) or can be used to restrain the docking of a ligand into a CYP binding site to get a more complete (3-dimensional) description of ligand orientation in the *HCS* (Figure 3B). The NMR data thus enables the construction of an experimentally verifiable 3D QSAR descriptor. This descriptor can be used to construct QSAR models to bias against the assembly of molecular fragments that are likely to produce binding to a given CYP antitarget. The descriptor should be helpful in fragment assembly strategies for inhibitor design as well as the enumeration of combinatorial libraries. For these applications, it is most useful if the descriptor is converted to a binary vector form (Figure 9C).

In summary, the 3D QSAR descriptor described herein is (a) tailored to the CYP gene family, (b) easily integrated with NMR studies and therefore experimentally verifiable, and (c) a flexible tool for biasing fragment assembly strategies to avoid binding to CYP antitargets. Its advantage over other descriptors is that it is based on data that are experimentally derived and structural in nature. Most other methods rely heavily on purely in silico docked structures and/or measured protein−ligand affinities. Furthermore, this descriptor can easily be applied to combinatorial fragment assembly problems when it is expressed as a bit string.

ACKNOWLEDGMENT

We thank Yuri Nikolsky of ChemDiv for providing the combinatorial chemistry building block reagents and James Kincaid and Mohammed Ibrahim for helpful discussions.

REFERENCES AND NOTES

(1) Sem, D. S.; Pellecchia, M. NMR in the acceleration of drug discovery *Curr. Opin. Drug Discuss. Devel.* **2001**, *4*, 479−493.
(2) Williams, P. A.; Cosme, J.; Sridhar, V.; Johnson, E. F.; McRee, D. E. Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Mol. Cell.* **2000**, *5*, 121−131.
(3) Wester, M. R.; Johnson, E. F.; Marques-Soares, C.; Dansette, P. M.; Mansuy, D.; Stout, C. D. Structure of a substrate complex of mammalian cytochrome P450 at 2.3 Å resolution: evidence for multiple substrate binding modes. *Biochemistry* **2003**, *42*, 6370−6379.
(4) Williams, P. A.; Cosme, J.; Ward, A.; Angove, H. C.; Matak Vinkovic, D.; Jhoti, H. Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* **2003**, *424*, 464−468.
(5) Lewis, D. F.; Eddershaw, P. J.; Goldfarb, P. S.; Tarbit, M. H. Molecular modeling of cytochrome P4502D6 (CYP2D6) based on an alignment with CYP102: structural studies on specific CYP2D6 substrate metabolism. *Xenobiotica* **1997**, *27*, 319−339.
(6) Lewis, D. F. Homology modelling of human cytochromes P450 involved in xenobiotic metabolism and rationalization of substrate selectivity. *Exp. Toxicol. Pathol.* **1999**, *61*, 369−374.
(7) de Groot, M. J.; Ackland, M. J.; Horne, V. A.; Alex, A. A.; Jones, B. C. Novel approach to predicting P450-mediated drug metabolism: developments of a combined protein and pharmacophore model. *J. Med. Chem.* **1999**, *42*, 1515−1524.
(8) de Groot, M. J.; Ackland, M. J.; Horne, V. A.; Alex, A. A.; Jones, B. C. A novel approach to predicting P450 mediated drug metabolism. CYP2D6 catalyzed N-dealkylation reactions and qualitative metabolite predictions using a combined protein and pharmacophore model for CYP2D6. *J. Med. Chem.* **1999**, *42*, 4062−4070.
(9) Vermeulen, N. P. Prediction of drug metabolism: the case of cytochrome P450 2D6. *Curr. Top. Med. Chem.* **2003**, *3*, 1227−1239.
(10) de Groot, M. J.; Ekins, S. Pharmacophore modelling of cytochromes P450. *Adv. Drug Deliv. Rev.* **2002**, *54*, 367−383.
(11) Modi, S.; Paine, M. J.; Sutcliffe, M. J.; Lian, L. Y.; Primrose, W. U.; Wolf, C. R.; Roberts, G. C. A model for human cytochrome P450 2D6 based on homology modeling and NMR studies of substrate binding. *Biochemistry* **1996**, *35*, 4540−4550.
(12) Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235−249.
(13) Lipinski, C. A.; Lombardo, F.; Dominy, B. M.; Feeny, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3−26.
(14) Klopman, G. Artificial intelligence approach to structure−activity studies. *J. Am. Chem. Soc.* **1984**, *106*, 7315−7321.
(15) TOPKAT, version 5.01. Oxford Molecular Group, Inc. www.oxmol.com.
(16) Sanderson, D. M.; Earnshaw, C. G. Computer prediction of possible toxic action from chemical structure. *Hum. Experimental Toxicol.* **1991**, *10*, 261−273.
(17) Durham, S. K.; Pearl, G. M. Computational methods to predict drug safety liabilities. *Curr. Opin. Drug Discov. Devel.* **2001**, *4*, 110−115.
(18) Benigni, R. The first US National Toxicology Program exercise on the prediction of rodent carcinogenicity: definitive results. *Mutat. Res.* **1997**, *387*, 35−45.
(19) Pearlman, R. S. Metric validation and the receptor relevant subspace concept. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 28−35.
(20) Solomon, I.; Bloembergen, N. *J. Chem. Phys.* **1999**, *25*, 261.
(21) Krugh, T. R. Spin-Label-Induced Nuclear Magnetic Resonance Relaxation Studies of Enzymes. In *Spin Labeling: Theory and Applications*; Berlinger, L. J., Ed.; Academic Press: New York, 1976.
(22) Jardetzky, O.; Roberts, G. C. K. In *NMR in Molecular Biology*; Academic Press: New York, 1981.
(23) Mohr, P. J.; Taylor, B. N. CODATA recommended values of the fundamental physical constants: 1998. *Rev. Modern Phys.* **2000**, *72*, 351−495.
(24) Wilkens, S. J.; Xia, B.; Volkman, B. F.; Weinhold, F.; Markley, J. L.; Westler, W. M. Inadequacies of the point-dipole approximation for describing electron−nuclear interactions in paramagnetic proteins: hybrid density functional calculations and the analysis of NMR relaxation of high-spin iron(III) rubredoxin. *J. Phys. Chem. B* **1998**, *102*, 8300−8305.
(25) Graham-Lorence, S. E.; Petersonm, J. A. Structural alignments of P450s and extrapolations to the unknown. *Methods Enzymol.* **1996**, *272*, 315−326.
(26) Jean, P.; Pothier, J.; Dansette, P. M.; Mansuy, D.; Viari, A. Automated multiple analysis of protein structures: application to homology modeling of cytochromes P450. *Proteins* **1997**, *28*, 388−404.
(27) Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* **2003**, *31*, 3381−3385.
(28) Guex, N.; Peitsch, M. C. SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modelling. *Electrophoresis* **1997**, *18*, 2714−2723.

(29) Peitsch, M. C. ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling. *Biochem. Soc. Trans.* **1996**, *24*, 274−279.
(30) Kirton, S. B.; Kemp, C. A.; Tomkinson, N. P.; St-Gallay, S.; Sutcliffe, M. J. Impact of incorporating the 2C5 crystal structure into comparative models of cytochrome P450 2D6. *Proteins* **2002**, *49*, 216−231.
(31) Halgren, T. A. Merck molecular force field. I Basis, form, scope, parametrization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.
(32) Halgren, T. A. Merck molecular force field. II van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, *17*, 520−552.
(33) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.* **1995**, *177*, 5179−5197.
(34) Page, M. I.; Jencks, W. P. Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68*, 1678−1683.
(35) Hansch, C.; Fujita, T. rho-sigma-pi. A Method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616.
(36) Hansch, C.; Leo, A. In *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995.
(37) Hopfinger, A. J.; Burke, B. J. In *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, C. A., Eds.; Wiley: New York, 1990; p 173.
(38) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on the binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.
(39) Voet, D.; Voet, J. G. In *Biochemistry*, 2nd ed.; Wiley: Sumerset, NJ, 1995.
(40) Poulos, T. L.; Finzel, B. C.; Howard, A. J. High-resolution crystal structure of P450cam. *J. Mol. Biol.* **1987**, *195*, 687−700.
(41) Lewis, D. F. V.; Lake, B. G.; George, S.; Dickins, M.; Beresford, A. P.; Eddershaw, P. J.; Tarbit, M. H.; Goldfarb, P. S.; Guengerich, F. P. Molecular modeling of CYP1 family isoforms CYP1A1, CYP1A2, CYP1A6 and CYP1B1 based on sequence homology with CYP102. *Toxicology* **1999**, *139*, 53−79.
(42) Lewis, D. F. V.; Eddershaw, P. J.; Goldfarb, P. S.; Tarbit, M. H. Molecular modeling of CYP3A4 from an alignment with CYP102: identification of key interactions between putative active site residues and CYP3A-specific chemicals. *Xenobiotica* **1996**, *26*, 1067−1086.
(43) Lewis, D. F. V.; Dickins, M.; Weaver, R. J.; Eddershaw, P. J.; Goldfarb, P. S.; Tarbit, M. H. Molecular modelling of CYP2C subfamily enzymes CYP2C9 and CYP2C19: rationalization of substrate specificity and site-directed mutagenesis experiments in the CYP2C family. *Xenobiotica* **1998**, *28*, 235−268.
(44) Lewis, D. F. V.; Eddershaw, P. J.; Goldfarb, P. S.; Tarbit, M. H. Molecular modeling of cytochrome P4502D6 (CYP2D6) based on alignment with CYP102: structural studies on specific CYP2D6 metabolism. *Xenobiotica* **1997**, *27*, 319−340.
(45) Lewis, D. F. V.; Lake, B. G.; Bird, M. G.; Dickins, M.; Eddershaw, P. J.; Tarbit, M. H.; Goldfarb, P. S. Molecular modeling of human CYP2E1 by homology with the CYP102 hemoprotein domain: investigation of the interactions of substrates and inhibitors within the putative active site of the CYP2E1 isoform. *Xenobiotica* **2000**, *50*, 1−25.
(46) Poulos, T. L. Modelling of mammalian P450s on the basis of P450cam X-ray crystal structure. *Methods Enzymol.* **1991**, *206*, 11−30.
(47) Modi, S.; Primrose, W. U.; Boyle, J. M. B.; Gibson, C. F.; Lian, L.-Y.; Roberts, G. C. K. NMR studies of substrate binding to cytochrome P450BM3: Comparisons to cytochrome P450cam. *Biochemistry* **1995**, *34*, 8982−8988.
(48) Graham-Lorence, S. E.; Peterson, J. A. Structural alignments of P450s and extrapolations to the unknown. *Methods Enzymol.* **1996**, *272*, 315−326.
(49) Jean, P.; Pothier, J.; Dansette, P. M.; Mansuy, D.; Viari, A. Automated multiple analysis of protein structures: application to homology modeling of cytochromes P450. *Proteins* **1997**, *28*, 388−404.
(50) Lewis, D. F. Homology modeling of human cytochromes P450 involved in xenobiotic metabolism and rationalization of substrate selectivity. *Exp. Toxicol. Pathol.* **1999**, *61*, 369−374.
(51) Keseru, GM A virtual high throughput screen for high affinity P450cam substrates. Implications for *in silico* prediction of drug metabolism. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 649−657.
(52) Islam, S. A.; Wolf, C. R.; Lennard, M. S.; Sternberg, M. J. E. A three-dimensional molecular template for substrates of human cytochrome P450 involved in debrisoquine 4-hydroxylation. *Carcinogenesis* **1991**, *12*, 2211−2219.
(53) Ellis, S. W.; Smith, G.; Wolf, C. R.; Tucker, G. T.; Woods, H. F. Evidence that Aspartate 301 is a critical substrate-contact residue in the active site of cytochrome P450 2D6. *J. Biol. Chem.* **1995**, *270*, 29055−29058.
(54) Goldberg, D. E. In *Genetic Algorithms*; Addison-Wesley: New York, 1989.

CI034208Q