Minireview

PHARMACOPHORE AND THREE-DIMENSIONAL QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP METHODS FOR MODELING CYTOCHROME P450 ACTIVE SITES

SEAN EKINS, MARCEL J. DE GROOT, AND JEFFREY P. JONES

Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, Indiana (S.E.); Molecular Informatics, Structure & Design, Pfizer Global Research and Development, Sandwich, Kent, United Kingdom (M.J.d.G.); and Department of Chemistry, Washington State University, Pullman, Washington (J.P.J.)

(Received March 2, 2001; accepted June 6, 2001)

This paper is available online at http://dmd.aspetjournals.org

ABSTRACT:

Structure activity relationships (SAR), three-dimensional structure activity relationships (3D-QSAR), and pharmacophores represent useful tools in understanding cytochrome P450 (CYP) active sites in the absence of crystal structures for these human enzymes. These approaches have developed over the last 30 years such that they are now being applied in numerous industrial and academic laboratories solely for this purpose. Such computational approaches have helped in understanding substrate and inhibitor binding to the major human CYPs 1A2, 2B6, 2C9, 2D6, 3A4 as well as other CYPs and additionally complement homology models for these enzymes. Similarly, these approaches may assist in our understanding of CYP induction. This review describes in detail the development of pharmacophores and 3D-QSAR techniques, which are now being more widely used for modeling CYPs; the review will also describe how such approaches are likely to further impact our active site knowledge of these omnipresent and important enzymes.

By the end of the 1990s, several reviews had characterized the active site details and physicochemical properties of substrates for the major cytochrome P450 (CYP) enzymes. These reviews had been gathered from analysis of physicochemical data (Smith and Jones, 1992), as well as an analysis of the literature for three-dimensional quantitative structure activity relationships (3D-QSAR), protein homology modeling and pharmacophore modeling (de Groot and Vermeulen, 1997), nuclear magnetic resonance, and site-directed mutagenesis techniques (Smith et al., 1997a,b). There has also been some suggestion of substrate rules defined in the form of a decision tree that describes determinants of CYP specificity (Lewis et al., 1998, 1999a). More recent reviews have addressed how a small number of physicochemical descriptors can explain the discriminatory abilities of the human CYPs (Lewis, 2000a,b) and provide a wealth of information for the interested reader. These types of data are important because there is still no crystal structure for a human membrane-bound CYP, although a crystal structure of the rabbit CYP2C5 is now available (Cosme and Johnson, 2000; Williams et al., 2000). In the face of this, these techniques and approaches have led to our most recent discoveries and insights into this large family of enzymes. Understanding the substrate specificity of the CYPs is essential because of their clinically important role in the metabolism of xenobiotics and endobiotics (Wrighton and Stevens, 1992). With the development of computational approaches, i.e., 3D-QSAR techniques, we are faced with a more visual description of how molecules may bind CYPs as either substrates and/or inhibitors. These techniques are still in the very early stages of development; as such, we have yet to see truly universal acceptance of this methodology. In an attempt to rectify this situation, it is worth considering that the foundation of these tools, mathematical approaches to understanding metabolism, is not entirely a modern endeavor, as these approaches began in the late 1960s (Hansch et al., 1968; Hansch, 1972). As early as 1972, Hansch had already suggested that quantitative relationships could be generated for the lipophilic character of drugs and metabolic parameters. Later QSARs for oxidation by CYPs were correlated with hydrophobic and steric parameters of substrates (Hansch and Zhang, 1993; Gao and Hansch, 1996). The publications by this group were perhaps a sign that CYPs could be modeled mathematically as a first step before computational analysis. However, we have now shifted to modeling specific CYPs and using tools with increased functionality available to us. To some extent these types of classical QSAR studies have been continued and expanded upon by relating the findings to docking of molecules in homology models of CYPs (Lewis, 1997; Lewis and Lake, 1998). The focus of this review, however, does not include such qualitative mammalian CYP homology models discussed in detail by others (Szklarz and Halpert, 1997b, 1998; Dai et al., 2000), which can be used to study enzyme function (Szklarz et al., 2000) or enable suggested semiquantitative models (De Rienzo et al., 2000). This review
intends to give an overview of the pharmacophore and 3D-QSAR models that have been used to describe P450s and indicate their varying degrees of success.

3D-QSAR and Pharmacophores

The development of computational tools has paralleled that of in vitro approaches to understanding and characterizing CYPs. One of the first visual 3D-QSAR computational approaches was comparative molecular field analysis (CoMFA) (Cramer et al., 1988), which enabled interpretation and understanding of enzyme active sites when a crystal structure was absent. However, it was not until in vitro drug-drug interaction studies were widely used (through the 1990s) in the pharmaceutical industry that a need arose for faster and cheaper approaches to determine this parameter. In this arena, computational approaches like CoMFA became useful in understanding requirements for CYP inhibitors (Fuhr et al., 1993; Strobl et al., 1993). Such computational 3D-QSAR techniques use relevant conformers of ligands to suggest functional groups, the geometry of structural features, and/or regions of electrostatic and steric interactions essential for activity or fit to the active site. The overall combination and 3D spatial distribution of physicochemical properties, the functional groups of the ligands, and a measure of binding site properties of an enzyme, such as the $K_{\text{m}}$ (apparent) (Nelsestuen and Martinez, 1997), $K_{i}$ (apparent), or $IC_{50}$ itself, define a pharmacophore. The methodologies used for generating 3D-QSAR applicable to CYPs have been reviewed in detail elsewhere (Ekins et al., 2000b,c).

Until recently, few CYP binding or active site models had been generated using enzyme kinetic data, and these focused primarily on inhibition. However, a considerable number of CYP pharmacophores have appeared in the literature, which presents us with the opportunity to review what is known about several CYPs based on such computational analyses.

CYP Models

CYP1A2. CYP1A2 is an inducible member of the CYP superfamily, which can be inhibited by some selective serotonin reuptake inhibitors (Brosen et al., 1993) and antiarrhythmics (Kobayashi et al., 1998). One of the first human CYP1A2 3D-QSAR studies was performed with SYBYL and ALCHEMY II software with 44 energy-minimized quinolone antibacterials (Fuhr et al., 1993), using the percentage of inhibition of caffeine 3-demethylation as the CYP1A2 specific activity probe (Fuhr et al., 1993). All 44 molecules were assumed by the authors to be competitive inhibitors, and the percent inhibition data were suggested to be similar to that which would be obtained if $K_{i}$ or $IC_{50}$ values were calculated. This study showed that there was a lack of correlation between lipophilicity and inhibitory effect. Four pharmacophore features were identified and suggested to be involved in the binding of CYP1A2 inhibitors in the active site. These were two positive potentials and two negative potentials that were thought to be necessary for binding of the quinolone type CYP1A2 inhibitors in the active site (Fuhr et al., 1993). However, these authors did not state interfeature dimensions or try to predict the inhibitory potential of molecules excluded from the model. In contrast, a later theoretical study suggested that the methyl group present at the 8-position on the xanthine ring of some tri- and di-substituted xanthine inhibitors is important for CYP1A2 inhibition. This would also be within 3 Å of one of the three regions of negative electrostatic potential (Sanz et al., 1994). A QSAR analysis using INSIGHT II with data for the inhibition of caffeine N3-demethylation by 16 naturally occurring flavonoids showed that the volume to surface area ratio descriptor had the greatest effect. In addition, the sigma factor (Hammett coefficient of the B ring) and electron densities at the C4’ and C3 atoms were also influential (Lee et al., 1998). Planar molecules with a small volume to surface area were the most potent inhibitors, while the number of hydroxyl groups was inversely proportional to inhibition, and glycosylation of the free hydroxyls decreased inhibition (Lee et al., 1998). A QSAR analysis of 12 heterocyclic amines using the COMBINE and GRID/GOLPE approaches enabled a comparison of structures of the ligand-protein complex with the structures of the ligands alone, respectively (Lozano et al., 2000). Both methods produced similar results and provided useful insights into positions of likely hydrogen bonding or placement of hydrophobic or bulky groups (Lozano et al., 2000).

With regard to predicting substrates for CYP1A2, one study has suggested that they are generally neutral or protonated and that they possess a total interaction energy greater than $-40$ kcal/mol and a molecular volume lower than 200 Å$^3$ (De Rienzo et al., 2000). However, this analysis was essentially a comparison of CYP1A2, CYP2D6, and CYP3A4 substrates docked in their respective active sites, so the interaction energy calculation is highly model-dependent. As yet, there has been no reported CYP1A2 substrate pharmacophore or QSAR used for quantitative predictions of a test set of molecules.

CYP2A6. To date there has been no published CYP2A6 QSAR; however, the related mouse form, CYP2A5, has been studied. One group analyzed substrate requirements using a graphical method and concluded that bicyclic ring systems with an electron-rich moiety were essential for the 11 molecules analyzed (Tehtmeyer and Legrum, 1994). A CoMFA study with 16 substrates and inhibitors of CYP2A5 generated from literature data suggested the importance of negative electrostatic potential close to a lactone moiety, steric effects around the methoxy group on methoxsalen, and positive electrostatic potential para to the methoxy group (Poso et al., 1995). When partial least-squares molecular surface weighted holistic invariant molecular (PLS MS-WHIM), an alignment-independent approach, was taken to modeling this same data set, size, positive molecular electrostatic potential, and hydrogen bond acceptor were important in producing a statistically significant PLS model (Bravi and Wikle, 2000). As yet, there have been few studies reporting large amounts of $K_{m}$ or $IC_{50}$ data for the related human CYP2A6 (Draper et al., 1997). Until a data set for CYP2A6 becomes available, we are unlikely to see further 3D-QSAR for this enzyme; however, these models will probably be similar to those derived for CYP2A5.

CYP2B6. Many examples of xenobiotics metabolized in part by CYP2B6 have been identified and described in more detail (Ekins and Wrighton, 1999). Recently a Catalyst hypothesis was generated from 16 $K_{m}$ (apparent) values either generated by the authors or obtained from the literature. Four features were suggested to be necessary for substrate binding in the active site, namely, three hydrophobes and one hydrogen bond acceptor (Ekins et al., 1999c). The three hydrophobic regions present in the pharmacophore were located at distances of 5.3, 3.1, and 4.6 Å from the hydrogen bond acceptor, with intermediate angles of 72.8° and 67.6°. The Catalyst pharmacophore demonstrated a good correlation of observed versus estimated $K_{m}$ values ($r = 0.85$). A preliminary version of this pharmacophore was independently suggested as being consistent with the homology model constructed by another group (Lewis et al., 1999b). In addition, the review of numerous QSAR studies on CYP2B proteins suggested the importance of hydrophobic and electronic properties in the binding of substrates (Lewis et al., 1999b), in further agreement with the aforementioned Catalyst pharmacophore. A second 3D-QSAR approach using molecular surface descriptors was also used (PLS MS-WHIM). This produced a model with a leave-one-out $q^2$ value of 0.67 and a more realistic five random groups repeated up to 100 times $q^2$ value.
of 0.61. Interestingly, the molecular surface properties, hydrophobicity, and hydrogen bond acceptor capacity were selected as important alongside unweighted and positive electrostatic potential (Ekins et al., 1999c). Both of these computational approaches were also used to predict a test set of five compounds. In four of five cases, the $K_i$ value was predicted within 1 log residual. These results in themselves are important as Catalyst and PLS MS-WHIM represent two quite distinct modeling approaches. Within Catalyst, molecules are described in terms of pharmacophoric features, i.e., hydrogen bond donor or acceptor atoms, hydrophobic moieties, and positively or negatively charged groups. The final result is represented by the 3D spatial disposition of essential features for the activity. In contrast, MS-WHIM descriptors are alignment-independent and are aimed at capturing global 3D chemical information at the molecular surface level. The shape and the electronic properties of a molecule are, therefore, characterized in 3D space and condensed into a single numerical vector. A structure-activity relationship is derived by means of PLS, and the final model is presented in the form of a linear equation. The fact that both of these approaches selected similar features suggests some degree of concordance or intermethod validation, although both methods failed to predict a different molecule in the test set, which suggests a possible advantage of using multiple and different 3D-QSAR approaches with data derived for CYPs. To date there have been no reported pharmacophore models for inhibitors of CYP2B6.

CYP2C9. Smith and Jones (1992) originally proposed that CYP2C9 might be a target for pharmacophore modeling in 1992. This suggestion was followed by the first manually superimposed pharmacophore model of CYP2C9 in 1993, which speculated that the active site contained a hydrogen bond donor or acceptor (Jones et al., 1993). A more detailed model based on the overlay of eight substrates and one inhibitor indicated that a hydrogen bond donor heteroatom was around 7 Å from the site of metabolism (Jones et al., 1996a). While this model was not statistically validated, it did indicate that overlays of substrates may provide a method to probe the active site of this enzyme.

This challenge was answered by Mancy et al. (1995), who developed a manually superimposed pharmacophore model using BIOSYM (cvff forcefield) software, based on 10 2-arylthiophenes and 10 additional CYP2C9 substrates. These molecules ranged in affinity ($K_i$) from 6 to 140 μM. This model defines a distance between the hydroxylation site and the anionic site to be 7.8 Å, a distance not significantly different from the Smith and Jones hypothesis. The angle between the protein cationic site, the substrate anionic site, and the site of hydroxylation was predicted to be around 82 degrees (Mancy et al., 1995). This model was extended in 1996 to reflect studies done with sulfaphenazole, a very potent inhibitor of CYP2C9 (Mancy et al., 1996). The new model included a hydrophobic zone between the hydroxylation site and the cationic site on the protein, and it indicated that the tight binding of sulfaphenazole was due to a ligand interaction between the aniline nitrogen of sulfaphenazole and the heme-iron. While these models were not quantitative structure activity relationships, they suggested the possible rationale for the overlay of substrates that could then be validated by statistical methods. These authors speculated that it was important that a CYP2C9 substrate contain an anionic site. As more substrates for the enzyme have become known, it is evident that neutral compounds can bind to this enzyme with high affinity (Miners and Birkett, 1998). In fact, the tightest binding compound to CYP2C9 is benz bromarone, with a $K_i$ value of 10 nM, which is not an anion at physiological pH (Takahashi et al., 1999). However, it has been observed that when two very similar compounds, such as warfarin (ring closed and not an anion at physiological pH) and phenprocoumon (an anion at physiological pH), bind to CYP2C9, the anion is the tightest binder (He et al., 1999).

The first quantitative QSAR for CYP2C9 was published in 1996 (Jones et al., 1996b). This work overlaid 27 compounds using the active-analog 9(S)-11(R)-cyclocoumarol as a relatively rigid template. This overlay assumed an aromatic binding region and a sterically forbidden region along the axis of the 9(S) carbon-hydrogen bond of the cyclocoumarol. The resulting PLS model, based on the CoMFA program, had a standard error of the estimate of 0.17 log units. Unlike the previous models for this enzyme, two cationic enzyme-binding sites were predicted to be important, along with the aromatic binding region and a steric region near the site of metabolism that was assumed to be the I-helix. While the 3D pharmacophore for this enzyme is complex, a 2D pharmacophore that emphasizes the most important binding regions can be constructed. This pharmacophore indicates favorable interactions when the substrate has a partial negative charge 10 Å from the oxidation site, with a second position of favored negative charge on the substrate 6 Å from the oxidation site and 35° clockwise from the first electrostatic site. The aromatic binding site is about equidistant from each of the electrostatic sites and above the plane defined by the oxidation site and the two electrostatic sites. The two electrostatic sites are thus around 6 Å apart.

This model was validated by testing 14 new compounds that had $K_i$ values ranging from 0.1 to 48 μM (Rao et al., 2000). While the initial training set contained mostly coumarin-containing compounds, this validation set contained mostly sulfonamides. Interestingly, the initial model predicted the affinity of the validation set reasonably well, predicting 13 of the 14 compounds within 1 log residual. Finally, when these compounds were included in the training set, the pharmacophore remained essentially the same.

In separate experiments, conducted at the same time as the validation study described above, pharmacophore and PLS predictive models where constructed using Catalyst and PLS MS-WHIM, respectively (Ekins et al., 2000a). This study used three training sets for Catalyst. The first predicted two important hydrophobic regions, one hydrogen bond acceptor and one hydrogen bond donor. The second, one hydrophobic region, two hydrogen bond donors, and one hydrogen bond acceptor. The third training set produced one hydrophobe and two hydrogen bond donors. Thus, all of the models had at least one hydrophobic region and one hydrogen bond acceptor. All three models indicate that the hydrogen bond acceptor and the hydrogen bond donor/acceptor are 3.4 to 5.7 Å apart, with the hydrophobic region 3 to 5.8 Å from the hydrogen bond acceptor. These results are in complete agreement with those presented above for the CoMFA model. In addition, the PLS MS-WHIM approach yielded several models that appeared to be internally predictive using leave-one-out and five random groups cross-validation. All of these descriptor models possessed similar features identified in the three Catalyst models generated with the same data sets.

To gain confidence in the pharmacophores generated for CYP2C9, an attempt was made to determine the specific amino acid residues that might be involved in establishing the pharmacophore. Initial docking of the 9(R)-11(S)-cyclocoumarol and visualizing the CoMFA field in a CYP2C9 homology model indicated that two phenylalanine residues, Phe$^{110}$ and Phe$^{114}$, might be involved in the aromatic binding region. Arg$^{105}$ may be the cationic binding site that binds one anionic substrate site, and Asp$^{193}$ may be the second electrostatic site. Mutagenesis experiments confirmed that Phe$^{114}$ but not Phe$^{110}$ have the phenotypic variation consistent with an aromatic binding region (Haining et al., 1999). Mutagenesis in the region of Arg$^{105}$ indicates that this region is not an important anionic binding site, but that Arg$^{97}$ is a more likely candidate (Ridderstrom et al., 2000). However, this residue is not predicted to be in the active site by either homology modeling or by analogy to the crystal structure available for CYP2C5.
(Cosme and Johnson, 2000; Ridderstrom et al., 2000). The importance of Arg100 was also demonstrated by the removal of this amino acid, which dramatically impacts formation of 4'-hydroxydiclofenac compared with the CYP2C9 wild type (>100-fold difference) (Ridderstrom et al., 2000).

**CYP2C19.** One group has focused on obtaining substrate structure activity relationships for the polymorphic CYP2C19 using inhibitors of omeprazole 5-hydroxylation (Lock et al., 1998a,b). Using mainly benzodiazepines which are N-dealkylated and 3-hydroxylated, it was suggested that these sites and the carbonyl group were important for inhibition. Electron-withdrawing groups were found to further decrease inhibition. As yet, the data for the 14 compounds used in these two studies have not been used to produce a published 3D-QSAR.

**CYP2D6.** Human CYP2D6 is a polymorphic member of the CYP superfamily and is absent in 5 to 9% of the Caucasian population as a result of a recessive inheritance of gene mutations (Mahgoub et al., 1977; Eichelbaum et al., 1979; Armstrong et al., 1994). This results in deficiencies in drug oxidations known as the debrisoquine/sparteine polymorphism, which affect the metabolism of numerous drugs. A diminished metabolism of these drugs is found in poor metabolizers, which have two nonfunctional CYP2D6 alleles, compared with extensive metabolizers with at least one functional allele. A relatively large number of small-molecule models has been derived for this particular human CYP isoenzyme, using a variety of substrates or inhibitors (Wolff et al., 1985; Meyer et al., 1986; Islam et al., 1991; Strobl, 1991; Koymans et al., 1992).

The first substrate models were manual alignments based on substrates containing a basic nitrogen atom at either 5 Å (Wolff et al., 1985) or 7 Å (Meyer et al., 1986) from the site of oxidation, and on aromatic rings which were coplanar (Wolff et al., 1985; Meyer et al., 1986). In the 5-Å model, no substrates were actually fitted onto each other (Wolff et al., 1985). The main problem with both of these initial models was that neither of the two models could explain the other group of substrates.

Islam et al. (1991) derived an extended model, which indicated a distance between a basic nitrogen atom and the site of oxidation between 5 and 7 Å. This small-molecule model also contained the heme moiety from the crystal structure of CYP101 [P450cam (Poulos et al., 1986)] above which the template molecule (debrisoquine) was positioned arbitrarily (Islam et al., 1991), in a manner resembling the orientation of camphor in the CYP101 crystal structure (Poulos et al., 1985). The model also included the iron-oxygen complex involved in the hydroxylation, and a set of 15 compounds was fitted onto debrisoquine.

Another small-molecule model for CYP2D6 was derived by Koymans et al. (1992). This model suggested that a hypothetical carboxylate group within the protein was responsible for a well-defined distance of either 5 or 7 Å between basic nitrogen atom and the site of oxidation within the substrate. This model used debrisoquine and dextromethorphan as templates for the 5- and 7-Å compounds, respectively. The oxidation sites of the two templates were superimposed and the areas next to the sites of oxidation were fitted coplanar, while the basic nitrogen atoms were placed 2.5 Å distant, interacting with different oxygen atoms of the postulated carboxylate group in the protein. The model consisted of 16 substrates, accounting for 23 metabolic reactions with their sites of oxidation and basic nitrogen atoms fitted onto the site of oxidation of the templates, and one of the basic nitrogen atoms of the template molecules, respectively. The model was verified by predicting the metabolism of four compounds giving 14 possible CYP-dependent metabolites. In vivo and in vitro metabolism studies on these substrates indicated that 13 of 14 predictions were correct (Koymans et al., 1992), indicating the relatively high predictive value of the model. The predictive value of the model was recently further confirmed, as two metabolites of 1-[2-[bis-(4-fluorophenyl)methoxy]-ethyl]-4-(3-phenyl-propyl)-pipera
dine (GBR 12909) were also correctly predicted and shown to be formed by heterologous expressed CYP2D6 (de Groot et al., 1995), and the model was extended with more substrates (de Groot et al., 1997b).

The actual positions of the heme moiety and the I-helix containing Asp101 [derived from a protein homology model of CYP2D6 (de Groot et al., 1996)] have been added to this small-molecule model, thereby incorporating some steric restrictions and orientational preferences into this model (de Groot et al., 1997a). In this refined small-molecule model, an aspartic acid residue was coupled to the basic nitrogen atoms, thus enhancing the small-molecule model with the direction of the hydrogen bond between the aspartic acid in the protein and the (protonated) basic nitrogen atoms (de Groot et al., 1997a). The site of oxidation above the heme moiety was one of the two possible sites of oxidation as suggested by a protein model for CYP2D6 (de Groot et al., 1996), located above pyrrole ring B of the heme moiety. In this model, the substrate sites of oxidation were fitted onto the defined oxidation site above pyrrole ring B of the heme moiety, while the Cα and Cβ atoms of the attached aspartic acid moiety were fitted onto the Cα and Cβ atoms of Asp101, respectively (de Groot et al., 1997a). A variety of substrates fitted in the original substrate model for CYP2D6 (Koymans et al., 1992; de Groot et al., 1995, 1997b) were successfully fitted into the refined substrate model. This indicated that the refined substrate model for CYP2D6 accommodates the same variety in molecular structures as the original substrate model. Recently, this model was used successfully to design a novel and selective CYP2D6 substrate suitable for high-throughput screening (Onderwater et al., 1999) and to investigate the hydroxylation of debrisoquine (Lightfoot et al., 2000).

Recently, a combined pharmacophore and homology model for CYP2D6 has been derived (de Groot et al., 1999a,b). This model consists of a set of two pharmacophores (one for O-dealkylation and oxidation reactions and a second one for N-dealkylation reactions catalyzed by CYP2D6) embedded in a protein homology model based on bacterial CYP crystal structures (de Groot et al., 1999a,b). This model for the first time combines the strengths of pharmacophore models (atom-atom overlap and reproducible starting points) and homology models (steric interactions and the possibility to identify amino acids involved in binding). This model correctly predicted the metabolism of a wide variety of compounds (de Groot et al., 1999a,b).

An inhibitor model for CYP2D6 has also been derived. The template of this model was derived by fitting six strong reversible inhibitors of CYP2D6 onto each other (Strobl et al., 1993). The basic nitrogen atoms were superimposed and the aromatic planes of these inhibitors were fitted coplanar. Consecutively, other inhibitors, such as derivatives of ajmalicine and quinidine, were fitted onto the derived template. The derived preliminary pharmacophore model consisted of a tertiary nitrogen atom (protonated at physiological pH) and a flat hydrophobic region. There also appeared to be two regions in which functional groups with lone pairs were allowed. In one of these regions, these groups caused enhanced inhibitory potency, which was not the case in yet another region (Strobl et al., 1993). The overall criteria derived for this inhibitor-based small-molecule model (Strobl et al., 1993) were very similar to the criteria for the proposed substrate models of CYP2D6 (Islam et al., 1991; Koymans et al., 1992; de Groot et al., 1997a). The pharmacophore consists of a tertiary nitrogen atom, a flat hydrophobic region, a region in which additional functional groups with lone pairs enhance inhibitory potency, and a region in which hydrophobic groups are allowed but cause no enhanced inhibitory effect (Strobl et al., 1993). For these reasons the substrate-based and inhibitor-based small-molecule models can probably be merged.
A set of 3D/4D-QSAR pharmacophore models has also been created for competitive inhibitors of CYP2D6 in a manner similar to that described for CYP2B6 and CYP2C9 using Catalyst (Ekins et al., 1999a). The first model for 20 inhibitors of bufuralol 1'-hydroxylation gave a good correlation between observed and predicted IC$_{50}$ values ($r = 0.75$), while a second model using 31 literature-derived IC$_{50}$ values provided a correlation between observed and predicted IC$_{50}$ values of $r = 0.91$. Both pharmacophores were capable of predicting IC$_{50}$ values for 9 to 10 of 15 CYP2D6 inhibitors within 1 log residual (Ekins et al., 1999a). Recently, a QSAR analysis identified a correlation between IC$_{50}$ values and lipophilicities in a series of close analogs designed as specific CYP2D6 substrates (Venhorst et al., 2000).

**CYP2E1.** CYP2E1 is involved in the metabolism of many toxic and carcinogenic molecules such as low molecular weight solvents and anesthetics. Early on, it was suggested that the active site was restricted due to the limited size of known substrates. A graphical model of the active site topology was derived from reactions of human CYP2E1 with phenyldiazene, 2-naphthyl, and p-biphenylyldiazene. This work indicated that the active site was open above the pyrrole rings A and D of the heme for a height of 10 Å (Mackman et al., 1996). A CoMFA study using rat in vivo data derived for 12 chlorinated volatile organic compounds essentially related to metabolism by CYP2E1 and CYP2B1 suggested the importance of gaining access to the active site relative to ligand-enzyme complementarity (Waller et al., 1996). This work was presented as an easily visualized schematic consisting of a long channel through which the substrate must progress before reaching the active site. The importance of electrostatic (long-range recognition), steric (substrate size), and hydrophathy (substrate surface interaction) was demonstrated in the resulting QSAR. A recent study has described dissociation constants for ethanol, halothane, isoniazid, and pyrazole with a range from 0.011 to 167 mM; this study also modeled arachidonic acid in CYP2E1 (Smith et al., 2000), which seemed to agree with the previous hypotheses of a long hydrophobic access channel.

**CYP3A4.** Smith et al. have described in detail the CYP3A4 active site characteristics (as well as those of the other major mammalian CYPs) based on homology models built using soluble bacterial CYP structures as a template (Smith et al., 1997a). They also suggested that the binding of CYP3A4 substrates in the active site is due to lipophilic forces, based on evidence from octanol and cyclohexane partition coefficients (Smith et al., 1997b). Additionally, these authors have suggested that the flexibility of the conformation of the CYP3A4 active site indicated by many researchers may contribute to the overall diverse substrate selectivity of CYP3A4 (Smith et al., 1997b). Many other papers have focused on the active site of this CYP using homology models and site-directed mutagenesis. Structural requirements of CYP3A4 substrates have been suggested to include a hydrogen bond acceptor atom 5.5 to 7.8 Å from the site of metabolism and 3 Å from the oxygen molecule associated with the heme (Lewis et al., 1996).

More recently, a pharmacophore for inhibitors of CYP3A4-mediated midazolam 1'-hydroxylation was developed that consisted of four features necessary for the inhibition of CYP3A4 (Ekins et al., 1999b). These features included three hydrophobes at distances of 5.2 to 8.8 Å from a hydrogen bond acceptor and resulted in a pharmacophore with an excellent correlation of observed versus estimated IC$_{50}$ (apparent) values ($r = 0.91$). Literature IC$_{50}$ values obtained from studies on the inhibition of CYP3A4-mediated cyclosporin A metabolism that covered 3 orders of magnitude (Pichard et al., 1990, 1996) were also modeled. A pharmacophore derived from this data produced five features, namely, three hydrophobes at distances of 4.2 to 7.1 Å from a hydrogen bond acceptor and a further 5.2 Å from another hydrogen bond acceptor (Ekins et al., 1999b). This pharmacophore demonstrated a reasonable correlation of observed versus estimated IC$_{50}$ (apparent) ($r = 0.77$). A second literature-derived data set was used in which IC$_{50}$ values obtained from the inhibition of CYP3A4-mediated quinine hydroxylation covered 4 orders of magnitude (Zhao and Ishizaki, 1997). This IC$_{50}$ pharmacophore consisted of four features including one hydrophobe at distances from 8.1 to 16.3 Å from the two furthest of three hydrogen bond acceptors (Ekins et al., 1999b). This Catalyst IC$_{50}$ pharmacophore demonstrated an excellent correlation of observed versus estimated IC$_{50}$ ($r = 0.92$). Comparing both IC$_{50}$ (apparent) pharmacophores should reveal more detail regarding the features necessary for inhibition. In this case, the two hydrophobes 5.2 and 7 Å away from the hydrogen bond acceptor in the first IC$_{50}$ pharmacophore overlap with the three hydrophobes indicated in the model derived from the data of Pichard et al. (1990, 1996). Merging these models results in two hydrophobic regions separated by hydrogen bond acceptors, suggesting similar domains and sites of hydrogen bond donation in the CYP3A4 active site.

To evaluate these 3D-QSAR models, the activity of molecules excluded from the training set was predicted and then compared with those observed by means of a 1 log residual. Eight molecules were selected from the literature with IC$_{50}$ (apparent) values. Both of the CYP3A4 IC$_{50}$ (apparent) Catalyst models predicted the IC$_{50}$ values similarly. Seven of eight best fit predictions were within 1 log unit residual, for both models, and the correct rank ordering of three protease inhibitors was observed (Ekins et al., 1999b).

Using the same commercially available software, a Catalyst hypothesis for 38 CYP3A4 substrates was generated using literature IC$_{50}$ data (Ekins et al., 1999d). The pharmacophore possessed two hydrogen bond acceptors, one hydrogen bond donor, and one hydrophobic region. The interhydrogen bond acceptor distance was 7.7 Å, while the hydrophobe and hydrogen bond donor were 6.6 and 6.4 Å from one of the hydrogen bond acceptors, respectively. The selected pharmacophore also demonstrated a good correlation of observed versus estimated IC$_{50}$ (apparent) values using the fast fit function ($r = 0.67$). This model produced estimates within 1 log unit residual for 12 test set molecules not included in the model (Ekins et al., 1999d). The presence of hydrogen bond acceptors in this CYP3A4 substrate pharmacophore was found to agree with the suggestion that hydrogen bonding of many substrates and inhibitors occurs with residue Asn 74, identified by docking molecules in the homology-modeled CYP3A4 active site (Lewis et al., 1996). This contrasts with an earlier theory put forward by Ferenczy and Morris (1989), who indicated that hydrogen bonds between the enzyme and substrate were not major interactions. The pharmacophore also confirms the findings of other groups because it includes a hydrophobic feature, implying an interaction with one or more of the suggested multiple hydrophobic domains in the active site of CYP3A4, as identified by alignment with bacterial CYPs (Ferenczy and Morris, 1989; Szklarcz and Halpert, 1997a).

Analyses of the likely features of activators of CYP3A4 have also been undertaken, as three substrates (carbamazepine, nifedipine, and testosterone) within the 38-molecule training set used in the CYP3A4 pharmacophore were known CYP3A4 autoactivators. A common features analysis of these molecules using the HipHop function within Catalyst generated a pharmacophore illustrating three hydrophobic areas and one hydrogen bond acceptor. The hydrophobic areas were located 4.4 to 7.6 Å from the hydrogen bond acceptor feature, and the sites of metabolism were colocated. Therefore, hydrophobic interactions with the CYP3A4 active site may be more important than hydrogen bonding for these same CYP3A4 substrates (Ekins et al., 1999d).

**CYP19 (Aromatase).** The importance of CYPs that metabolize endogenous substrates can be demonstrated by aromatase, which catalyzes the metabolism of androstenedione to estrone, 16α-hy-
droxyandrostenedione to estriol, and testosterone to estradiol via the aromatization of the A ring and the removal of the C19 methyl group (Oprea and Garcia, 1996). Some early QSAR on this enzyme used molecular mechanics and quantum chemical calculations to show that the C4 to C5 double bond on 46 derivatives of 4-androstene-3,17-dione and 5α-androstane-3,17-dione enables the adoption of a readily hydroxylated conformation for these inhibitors of human placental aromatase (Bohl et al., 1988). Later work using 3D-QSAR with substituted dichlorophenyl aromatase inhibitors generated two possible pharmacophores resulting from molecular descriptors and multidimensional linear regression (Nagy et al., 1994). Both pharmacophores used the two aromatic rings to define a rigid, nonpolar molecular shape in space and also contained a hydrogen bond acceptor (Nagy et al., 1994). In 1996, two groups described CoMFA studies for steroidal (Oprea and Garcia, 1996) and nonsteroidal (Recanatini and Cavalli, 1996) inhibitors of aromatase. The former study analyzed 50 steroidal inhibitors using CoMFA and GOLPE to suggest two hydrophobic binding pockets in the C6 region of the steroid (Oprea and Garcia, 1996). One of these is large and located in the α-face while another is smaller and located in the β-position. The later paper also used CoMFA for 29 molecules related to fadrozole and aligned on 5α-fadrozole. This model suggests two regions around the reference molecule and defines the importance of the α-region; it also presents an idea of the size of the hydrophobic site (Recanatini, 1996). This study was later expanded to include tetralone derivatives to provide a training set of 49 inhibitor molecules and a further test set of eight molecules (Recanatini and Cavalli, 1998). It was found necessary to use two different aromatase inhibitors for alignment, suggesting that there may be multiple binding orientations for different classes of nonsteroidal aromatase inhibitors (Recanatini and Cavalli, 1998). Further diverse inhibitors of aromatase have been synthesized, such as the aryl-substituted pyrrolizine and indolizine derivatives (Sonnet et al., 2000), but it must also be considered that such molecules may be inhibitors for other CYPs.

A 3D-QSAR approach may be strengthened by using a pharmacophore in the design of inhibitors for this bacterial CYP. Applying computational approaches like those discussed as well as homology models [using mammalian CYP51 alignments (Ji et al., 2000)] may result in more selective therapeutic agents that do not inhibit the host CYP. This potential for inhibition may result in interference with meiosis-activating sterol production and could possibly affect mammalian reproduction (Debeljak et al., 2000). Alternatively, inhibition of human CYP51 may be therapeutically useful (Harwood et al., 1999); therefore, modeling this may be valuable. Clearly, as more in vitro data are generated, pharmacophore models for this human CYP will not be far behind.

**Computational Models for CYP Inducers**

Not only is the prediction of CYP-mediated inhibitory drug-drug interactions important for the pharmaceutical industry, but CYP induction represents an additional drug-drug interaction that should be avoided. Computational modeling of CYP inducers is complicated because most of the induction literature relates to animal in vivo data, which in some instances can be modeled using QSAR analysis (Rekka and Kourounakis, 1996) or in vitro data (van der Buught et al., 1999). The appearance of data sets from human in vitro models such as those for CYP3A4 (Scheutz et al., 1996), nuclear hormone receptors pregnane X receptor (Kliewer et al., 1998; Lehmann et al., 1998; Moore et al., 2000), and CAR (Honkakoski et al., 1998) offers potentially larger data sets. This would suggest that we might be able to define physicochemical properties or molecular features necessary for induction in the latter cases, binding to one or more nuclear hormone receptor site(s). However, the narrow range of fold induction with these data sets applicable to humans may be a limitation with the presently available modeling approaches such as those described in this review. The ability to computationally predict CYP3A4 induction, for instance, would be a great advance in throughput over the currently available in vitro methods using hepatocytes and functional and binding assays.

**Discussion**

The first review to discuss pharmacophores applied to CYPs reviewed the literature up to 1993 and suggested the likely difficulty in producing active site templates for other CYPs that might not possess specific ionic interactions (Korzekwa and Jones, 1993). However, the present review and previous discussions on predicting drug-drug interactions in silico (Ekins et al., 2000b) point to the increased...
TABLE 1
Summary of key pharmacophore features for xenobiotic-metabolizing human CYPs

<table>
<thead>
<tr>
<th>CYP</th>
<th>Key Features of Pharmacophore</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A2</td>
<td>Two positive potentials and two negative potentials are required for quinolone inhibitors.</td>
<td>Fuhr et al., 1993</td>
</tr>
<tr>
<td></td>
<td>Planar molecules with a small volume to surface area are potent inhibitors.</td>
<td>Lee et al., 1998</td>
</tr>
<tr>
<td></td>
<td>Identified positions of hydrogen bonding and hydrophobic groups for inhibitors</td>
<td>Lozano et al., 2000</td>
</tr>
<tr>
<td>2A6</td>
<td>Potent CYP2A6 inhibitors do not include a lactone constituent.</td>
<td>Poso et al., 2001</td>
</tr>
<tr>
<td>2B6</td>
<td>Substrates possess at least three hydrophobes and one hydrogen bond acceptor.</td>
<td>Ekins et al., 1999c</td>
</tr>
<tr>
<td>2C9</td>
<td>Hydrogen bond donor 7 Å from site of metabolism in substrates</td>
<td>Jin et al., 1996a</td>
</tr>
<tr>
<td></td>
<td>The distance between the hydroxylation site and the anionic site is 7.8 Å. There is a hydrophobic</td>
<td>Maney et al., 1995</td>
</tr>
<tr>
<td></td>
<td>zone between the hydroxylation site and the cationic site on the protein.</td>
<td>Maney et al., 1996</td>
</tr>
<tr>
<td></td>
<td>Inhibitors display two cationic enzyme binding sites that were predicted to be important, along</td>
<td>Jones et al., 1996b</td>
</tr>
<tr>
<td></td>
<td>with the aromatic binding region, and a steric region. Favorable interactions occur when the</td>
<td>Rao et al., 2000</td>
</tr>
<tr>
<td></td>
<td>substrate has a partial negative charge 10 Å from the oxidation site, with a second position of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>favored negative charge on the substrate 6 Å from the oxidation site and 35 degrees clockwise</td>
<td></td>
</tr>
<tr>
<td></td>
<td>from the first electrostatic site.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inhibitor models had at least one hydrophobic region and one hydrogen bond acceptor. The</td>
<td>Ekins et al., 2000a</td>
</tr>
<tr>
<td></td>
<td>hydrophobic bond acceptor and the hydrophobic bond donor/acceptor are 3.4 to 5.7 Å apart, with</td>
<td></td>
</tr>
<tr>
<td></td>
<td>the hydrophobic region 3 to 5.8 Å from the hydrogen bond acceptor.</td>
<td></td>
</tr>
<tr>
<td>2D6</td>
<td>Substrates possess a basic nitrogen atom at either 5 or 7 Å from the site of oxidation, and</td>
<td>Wolff et al., 1985;</td>
</tr>
<tr>
<td></td>
<td>aromatic rings that are coplanar.</td>
<td>Meyer et al., 1986;</td>
</tr>
<tr>
<td></td>
<td>Substrates possess a carboxylic group within the protein responsible for a well defined</td>
<td>Islam et al., 1991</td>
</tr>
<tr>
<td></td>
<td>distance of either 5 or 7 Å between basic nitrogen atom and the site of oxidation within the</td>
<td>Koymans et al., 1992</td>
</tr>
<tr>
<td></td>
<td>substrate. An aspartic acid residue was coupled to the basic nitrogen atoms.</td>
<td>de Groot et al., 1997a</td>
</tr>
<tr>
<td></td>
<td>Two substrate pharmacophores (one for O-dealkylation and oxidation reactions and a second one</td>
<td>de Groot et al., 1999a,b</td>
</tr>
<tr>
<td></td>
<td>for N-dealkylation reactions catalyzed by CYP2D6) were generated.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inhibitor model contains a tertiary nitrogen atom (protonated at physiological pH) and a flat</td>
<td>Strobl et al., 1993</td>
</tr>
<tr>
<td></td>
<td>hydrophobic region plus two regions in which functional groups with lone pairs are allowed.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inhibitor pharmacophores contain a hydrogen bond acceptor and a hydrogen bond donor and two to</td>
<td>Ekins et al., 1999a</td>
</tr>
<tr>
<td></td>
<td>three hydrophobic regions.</td>
<td></td>
</tr>
<tr>
<td>2E1</td>
<td>Long hydrophobic access channel</td>
<td>Walser et al., 1996</td>
</tr>
<tr>
<td>3A4</td>
<td>Substrate model includes hydrogen bond acceptor atom 5.5 to 7.8 Å from the site of</td>
<td>Lewis et al., 1996</td>
</tr>
<tr>
<td></td>
<td>metabolism and 3 Å from the oxygen molecule associated with the heme.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inhibitor pharmacophores include three hydrophobes at distances of 5.2 to 8.8 Å from a hydrogen</td>
<td>Ekins et al., 1999b</td>
</tr>
<tr>
<td></td>
<td>bond acceptor, three hydrophobes at distances of 4.2 to 7.1 Å from a hydrogen bond acceptor,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>or one hydrophobe at distances from 8.1 to 16.3 Å from the two furthest of three hydrogen bond</td>
<td></td>
</tr>
<tr>
<td></td>
<td>acceptors. Substrate pharmacophore possessed two hydrogen bond acceptors, one hydrogen bond</td>
<td></td>
</tr>
<tr>
<td></td>
<td>donor, and one hydrophobic region.</td>
<td></td>
</tr>
</tbody>
</table>

interest and relative ease in modeling all the major human CYPs. In conclusion, we have shown that various 3D-QSAR models can be generated and that these models could be used to classify molecules for their likely ability to be CYP substrates or inhibitors (Table 1). The models reviewed here show that although there are common features present in all P450s modeled so far, the relative contributions vary dramatically between different P450 isoenzymes (e.g., relative contributions of electrostatic versus hydrophobic interactions and the importance of flexibility of the active site). In the future, CYP induction will probably be modeled to the same extent as we now attempt to predict drug-drug interactions for these enzymes. This pharmacophoric/3D-QSAR approach would naturally be a very useful tool for virtual drug discovery once the models were suitably refined to account for some of the poor predictions presently observed. Such an iterative approach to model building would provide CYP prototype models that could be widely applied in discovery alongside other in silico models for absorption, distribution, metabolism, and excretion/toxicology properties and bioactivity to predict molecules likely to avoid attrition during development.

**Notes Added in Proof.** A CoMFA and GOLPE study of CYP2A6 and CYP2A5 inhibitors suggests the active site of the latter is larger due to observed larger steric regions. Furthermore, the most potent CYP2A6 inhibitors do not include a lactone constituent (Poso et al., 2001).

**Acknowledgments.** S.E. acknowledges Steven A. Wrighton, James Wikel, Gianpaolo Bravi, Patrick Murphy, and colleagues in Computational Chemistry and Molecular Structure Research, In Vitro groups, and other collaborators at Eli Lilly. In addition, S.E. is grateful to my past colleagues R. Scott Obach and Dayna Mankowski at Pfizer Inc. (Groton, CT) for their collaborations, constructive criticism in writing, and modeling.

**References**
de Groot MJ, Ackland MF, Horne VA, Alex AA and Jones BC (1999b) Novel approach to...


Fujita T (1997) Recent success stories leading to commercializable bioactive compounds with the value of comparing a three dimensional model of cytochrome P450 with the most important biotransformation enzymes. Drug Metab Rev 29:747–799.


