QSAR Modeling of in Vitro Inhibition of Cytochrome P450 3A4*†

Boryeu Mao,*,‡ Rafael Gozalbes,§ Frédérique Barbosa,§ Jacques Migeon,‡ Sandra Merrick,‡ Kelly Kamm,‡ Eric Wong,‡ Chester Costales,‡ Wei Shi,‡ Cheryl Wu,‡ and Nicolas Froloff*,§

Cerep, 15318 NE 95th Street, Redmond, Washington 98052, and Cerep, 19 avenue du Québec, 91951 Courtaboeuf Cedex, France

Received March 15, 2006

We report the QSAR modeling of cytochrome P450 3A4 (CYP3A4) enzyme inhibition using four large data sets of in vitro data. These data sets consist of marketed drugs and drug-like compounds all tested in four assays measuring the inhibition of the metabolism of four different substrates by the CYP3A4 enzyme. The four probe substrates are benzyloxycoumarin, testosterone, benzyloxyresorufin, and midazolam. We first show that using state-of-the-art QSAR modeling approaches applied to only one of these four data sets does not lead to predictive models that would be useful for in silico filtering of chemical libraries. We then present the development and the testing of a multiple pharmacophore hypothesis (MPH) that is formulated as a conceptual extension of the traditional QSAR approach to modeling the promiscuous binding of a large variety of drugs to CYP3A4. In the simplest form, the MPH approach takes advantage of the multiple substrate data sets and identifies the binding of test compounds as either proximal or distal relative to that of a given substrate. Application of the approach to the in silico filtering of test compounds for potential inhibitors of CYP3A4 is also presented. In addition to an improvement in the QSAR modeling for the inhibition of CYP3A4, the results from this modeling approach provide structural insights into the drug−enzyme interactions. The existence of multiple inhibition data sets in the BioPrint database motivates the original development of the concept of a multiple pharmacophore hypothesis and provides a unique opportunity for formulating alternative strategies of QSAR modeling of the inhibition of the in vitro metabolism of CYP3A4.

INTRODUCTION

In vitro screening of biological activities plays an important role in modern drug discovery. Results from such screening assays help identify tangible leads for further optimization and for clinical development into drug candidates. At the same time, the assay data become a source of information useful to medicinal chemists for the derivation of structure−activity relationships that are key factors in the optimization of desirable properties for new drugs. Another important impact of in vitro screening assays in the drug discovery process is that, in addition to the optimization of potency against the primary biological target, more attention is being paid to absorption, distribution, metabolism, excretion, and toxicity (ADME-Tox)-related properties that have been traditionally reserved for later stages in the development of drug candidates.

Cytochrome P450 enzymes (CYPs) constitute a superfamily of heme proteins involved in the metabolism of endogenous and exogenous substrates in living organisms. Included in this enzyme family is the dominant group of enzymes that controls the oxidative metabolism and clearance of drug molecules and other xenobiotics.[1] Of all CYP isoforms, CYP2D6 and CYP3A4 account for the metabolism of more than half of the known drugs.[2,3] CYP3A4 is also one of the most abundant isoforms of the CYP enzymes in the liver.[4,5] Interactions of drug molecules with these enzymes could lead to altered pharmacokinetic parameters, such as the half-life and the maximum of serum concentration, or the total systemic exposure (the "area under the curve" of the serum concentration integrated over time), of other drug molecules that might be coadministered. Thus, knowing how a drug molecule interacts with these two enzymes is essential for understanding not only its own metabolic fate but also how it might adversely influence the metabolism of other drugs through CYP-based drug−drug interactions.[6−8] For CYP3A4, the large variety of structurally diverse drugs that are metabolized suggests a broad enzyme binding pocket (and thus a broad range of possible intermolecular interactions) for this enzyme;[9] in addition, the noncompetitive inhibition of CYP3A4 shows that multiple substrates or inhibitors can bind simultaneously in the enzyme binding pocket.[9−16] The three-dimensional structure of CYP3A4 recently determined by X-ray crystallography confirms a large binding site near the heme group.[17,18]

To reduce the attrition of otherwise promising lead candidates because of toxicity liabilities from drug−drug interactions, in vitro screening tests of drug candidates for the inhibition of drug-metabolizing CYPs are now carried out early in drug discovery. The evaluation of these biological activities of new chemical structures can be moved to an even earlier stage, especially for large combinatorial libraries,

* Corresponding author fax: (425) 896-8668 (B.M.); e-mail n.froloff@cerep.com (N.F.), b.mao@cerep.com (B.M.).

† Originally submitted for the partial special issue "Seventh International Conference on Chemical Structures, Noordwijkerhout, The Netherlands, June 2005", published in the March/April, 2006, issue of *J. Chem. Inf. Model.* (Vol. 46, No. 2).

‡ Cerep, Redmond, Washington.

§ Cerep, Courtaboeuf Cedex, France.

10.1021/ci0600915 CCC: $33.50
Published on Web 08/17/2006

if accurate in silico methods are available. Ideally suited for developing in silico methods by QSAR modeling is BioPrint, a pharmacoinformatics platform which covers structural, physicochemical, in vitro pharmacological, and ADME-Tox-related properties of currently over 2400 marketed drugs, drug-like molecules, and reference compounds[19,20] (http://www.cerep.fr/cerep/users/pages/Collaborations/Bioprint.asp). In particular, among the in vitro ADME-Tox-related properties that have been systematically tested and collected in the BioPrint database are inhibition data on CYP enzymes; these data sets are ideally suited for the modeling of CYP enzyme inhibition by QSAR methods, similar to those widely reported in the literature.[21−24] The inhibition data set in BioPrint for CYP2D6 has previously been used for generating a robust and validated QSAR model.[25] We report here a QSAR study of the inhibition of the in vitro metabolism of four different probe substrates of CYP3A4 on the basis of the full data set in BioPrint.

We first show evidence that the inhibition of the CYP3A4 metabolism of a coumarin substrate by compounds in the BioPrint collection is not adequately predicted by the QSAR model derived by Wanchana et al.[26] from an experimentally comparable but smaller data set. Nor does the inhibition data set lead to a model of acceptable statistical parameters (correlation coefficient r^2 or mean-squared error MSE) by our own state-of-the-art QSAR modeling methods in BioPrint (which had previously proven successful in modeling the in vitro inhibition of CYP2D6[25]). Thus, seeking a rationale for the unique difficulty in modeling enzyme−inhibitor interactions in CYP3A4, and motivated by the knowledge of a broad binding pocket for the enzyme, we develop an alternative hypothesis of multiple pharmacophores for the enzyme inhibition. We present the motivation and the support for the multiple pharmacophore hypothesis (MPH) on the basis of the data in BioPrint. The hypothesis in the simplest form is tested on the BioPrint in vitro inhibition data sets for four chemically distinct CYP3A4 substrates by a QSAR strategy that is formulated conceptually within the established multilinear modeling techniques. The results show that MPH improves the QSAR modeling of the in vitro inhibition of CYP3A4 and provides structural insights into the proximity and the possible overlap of substrate binding sites in the CYP3A4 enzyme. The implication of multiple binding for the QSAR strategy of building "local models" is discussed. Last, as the confirmation that MPH is a valid approach for the filtering of chemical libraries for potential inhibitors of CYP3A4, we demonstrate the application of MPH-derived models to the in silico prediction of the CYP3A4 inhibition of test compounds.

MATERIALS AND METHODS

In Vitro Assay of Cytochrome P450 Enzyme Inhibition. The measurement of CYP3A4 inhibition in BioPrint consists of four in vitro assays as a part of a larger panel for a total of nine CYP isozymes in the ADME-Tox section of the database. While the inhibition of most other isoforms of CYP enzymes in the BioPrint database is assayed using a single substrate, for CYP3A4, the inhibition is measured in four assays, each with a specific substrate: 7-benzyloxy-4-(trifluoromethyl)-coumarin (BFC), 7-benzyloxy-resorufin (BzR), testosterone (TES), and midazolam (MID) (Figure 1). A microplate-based assay procedure coupled with either a fluorometric or a high-performance liquid chromatography/UV−vis detection is used to assay the inhibition of cDNA-expressed human CYP enzymes by test compounds.[13,27,28] The inhibition of the CYP3A4 activity for each substrate is first assayed as the percent reduction in the control activity due to the presence of each BioPrint compound at a 10 μM concentration. For compounds that show more than 30% inhibition in the primary screen, the determination of IC50 values is carried out by measurements at eight concentrations in the secondary, follow-up assay. Typically, about 2% of the IC50 determinations for CYP3A4 fail because of experimental conditions (such as the solubility limit at high concentrations precluding a complete IC50 curve). All assays are run in duplicate, with standard reference inhibitors as controls.

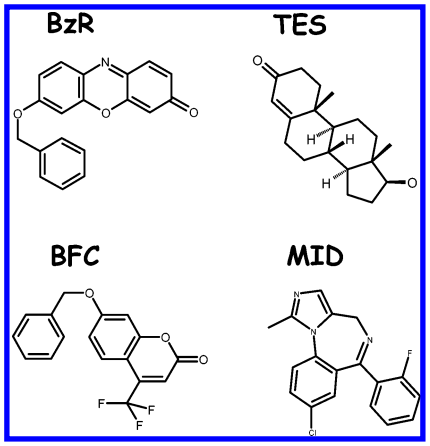

Figure 1. Chemical structures of the four probe substrates used in the CYP3A4 in vitro inhibition assays: 7-benzyloxy-resorufin (BzR), 7-benzyloxy-4-(trifluoromethyl)-coumarin (BFC), testosterone (TES), and midazolam (MID).

QSAR Modeling. Web-based software developed in-house at Cerep, the "BioPrint Program Manager" (BPPM[19]), is used for the QSAR modeling in this study. The BPPM includes a proprietary set of pharmacophore-based bipolar (FBPA), surface area (PTA), and field (EFO) descriptors for molecular structures;[29,30] these are essential pharmacophore-derived molecular features that are relevant to molecular properties and biological activities. The pharmacophore descriptors are augmented with topological descriptors based on molecular connectivity and the electrotopological state[31−34] and fragment descriptors of common functional groups (see, for example, Yoshida and Topliss[35]). Within the BPPM, the structure−activity relationship is obtained by a genetic algorithm (GA)-driven descriptor selection procedure for deriving linear and neighborhood behavior[36,37] models, followed by a synergistic combination of model estimates of both types. The synergistic combination, leading to a "synergy" model, is a linear combination of a selected multivariate linear model estimate and that of a selected neighborhood behavior model for a given property or activity. A confidence indicator provided by the neighborhood behavior model is used in deriving the coefficients for linear and neighborhood behavior estimates in the synergy model. The standard QSAR model-building process in the BPPM includes an optional built-in step in which a learning and a validation set are constructed according to the distribution of the activity range of the input compound data set. The

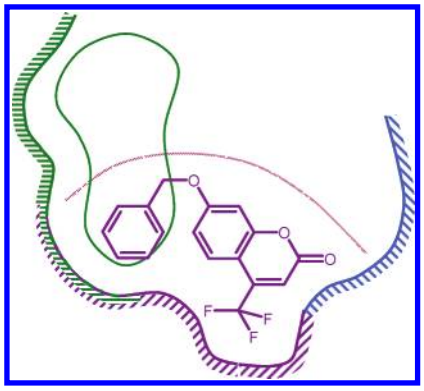

Figure 2. Schematic diagram for the "multiple pharmacophore hypothesis" for the binding of substrates and inhibitors to CYP3A4. The pharmacophore set for the recognition of BFC (structure drawn in purple) is marked by purple hashes on the surface, which represents the enzyme binding pocket, while the pharmacophore set for a second substrate is indicated by green hashes. Some pharmacophore features are shared by both sets of pharmacophores (i.e., the overlap of the purple and the green hashes). Note that the three-dimensional arrangement of two pharmacophore sets might bring the geometrical location of the two binding sites together in the enzyme binding pocket, a special case within the general scope of MPH. Depending on its relative location, the binding of an inhibitor may be labeled as either proximal or distal to BFC, respectively in the two regions below and above an "ideal" boundary marked by the dashed line. Proximal and distal binding regions can be alternatively labeled with respect to other substrates. In our current implementation of MPH, proximal and distal binders are identified for a given substrate, and each subgroup probes a subset of the multiple pharmacophores for binding to the CYP3A4 enzyme. The geometric interpretation of MPH is supported by the descriptor analysis of QSAR models (Table 3).

GA search for the best QSAR models is guided by the cross-validated correlation coefficients with respect to the target activity or property. Once built, QSAR models are accessible within the same software interface for model application and prediction. Full details of the proprietary descriptor set and the original QSAR modeling approach have been described elsewhere.[19,20,25,29,30,36,37]

For testing the QSAR model of CYP3A4 inhibition by Wanchana et al.,[26] the calculation of topological indices and other descriptors required by the QSAR equation is implemented as prescribed,[31−34] and the descriptor values are calculated for compounds in the BioPrint collection.

Multiple Pharmacophore Hypothesis (MPH). For most enzymes, the QSAR modeling of the recognition of inhibitors is based on a "simple pharmacophore binding" hypothesis. The pharmacophore set (namely, the descriptors characterizing the structure−activity relationship with respect to the inhibitory activity) is considered "simple" in this case because the key interactions with the enzyme are generally shared by the probe substrate and the inhibitors on account of the enzyme specificity. In contrast, CYP3A4 recognizes and metabolizes multiple drugs, as shown schematically in Figure 2. In addition to BFC, for example, the presence of a secondary substrate binding site in this enzyme (namely, those complementary structural features of the enzyme that recognize the secondary substrate) means (a) that the binding sites might share little or no overlap such that some inhibitors may interfere with the binding and the subsequent metabolism of only one of the substrates or (b) that the binding sites share significant overlap such that a molecule might show inhibitory activity against a substrate without possessing its full set of "pharmacophore features" (the case depicted schematically in Figure 2). The structure−activity relationship becomes more complex because each substrate probes a subset of the multiple binding pharmacophores in the catalytic binding pocket of CYP3A4.

Table 1. Pairwise Comparison of IC50 Values[a]

	BFC			BzR		
	%inh	IC50 [nM]	MPH group	%inh	IC50 [nM]	MPH group
compound 1	87	3080	d	67	160	p
compound 2	23			43	13000	p
compound 3	83	3500	p	46	9500	d
compound 4	70	3840	d	72	540	p
compound 5	38	51400	p	39		
compound 6	58	16200	p	34	17000	d
compound 7	45	27800	d	91	3700	p
compound 8	99	77	d	98	0.35	p
compound 9	56	12200	d	91	390	p
compound 10	29			63	4900	p
compound 11	63	11500	d	95	340	p
compound 12	64	16900	d	58	1200	p
compound 13	40	14700	d	65	3000	p
compound 14	69	5920	d	92	50	p
compound 15	−6			41	13000	p
compound 16	98	458	d	94	8	p
compound 17	54	10100	d	61	3500	p

[a] Test compounds are assigned to the proximal (p) or the distal (d) subgroups for two substrates, BFC and BzR, according to their relative potency. For example, compound 1 is assigned to group p for BzR because the IC50 is stronger against this substrate. Without an IC50 value for the substrate BFC, compound 2 is assigned to neither group; by default, it is assigned to group p for BzR. The IC50 column is left blank for a compound if the primary assay result (%inh) is below the threshold of 30% or if the experimental conditions preclude an IC50 determination.

Our initial QSAR modeling results provide the motivation and the supporting evidence for the multiple pharmacophore hypothesis, as presented in the Results and Discussion section. These results also suggest the following strategy for incorporating MPH in the QSAR modeling for the enzyme inhibition of CYP3A4. For a pair of substrates (e.g., BFC and BzR), compounds that bind proximally to one substrate would be expected to generally have a higher inhibitory activity against its metabolism than that of the other substrate to which it is located distally. Thus, to the very first approximation, one can estimate the relative proximity of the location of a compound with respect to the locations of BFC and BzR by the relative activity against the two substrates, as shown in Table 1. The IC50 data set for each substrate is therefore divided into two subgroups, approximating respectively the proximal binding (group p) and the distal binding (group d), each subgroup subjected to an independent search of linear QSAR models. For reference, a QSAR model is also derived for the combined data set (group R). [N.B., Although the terms "proximal" and "distal" traditionally are references to locations relative to the heme group in CYP enzymes, in the context of MPH, they refer conceptually to the geometric relationship with respect to the binding of probe substrates.]

Group Classifier for MPH Models. QSAR modeling according to the multiple pharmacophore hypothesis develops multiple MPH models for specific subgroups with respect to a substrate (see Results and Discussion). To test the QSAR models developed from MPH, and to develop a virtual

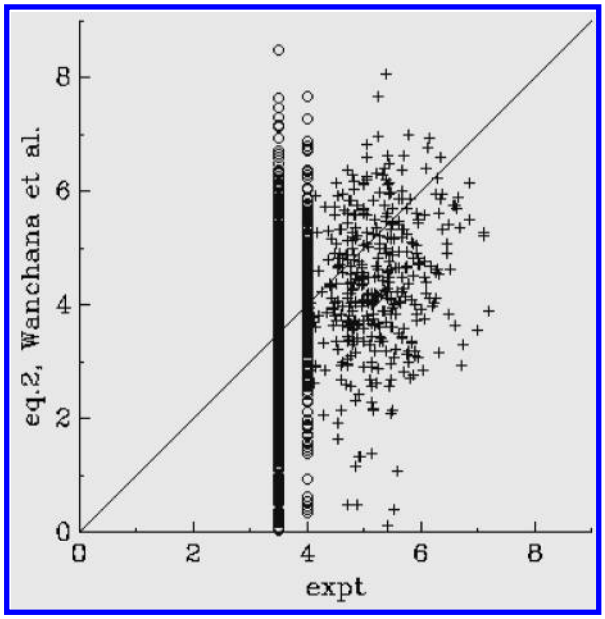

Figure 3. Comparison of experimental IC50 values and those predicted by the QSAR model of Wanchana et al.[26] for 2100 compounds in the BioPrint database. The experimental (expt) and the calculated activity values are represented in pIC50's defined as −log(IC50 in M). Compounds without experimental IC50 values are given an "estimated" experimental IC50 of 316 μM (pIC50 of 3.5) if the primary assay result is below 30% and a value of 100 μM (pIC50 of 4.0) if the primary assay is above 30% (circles).

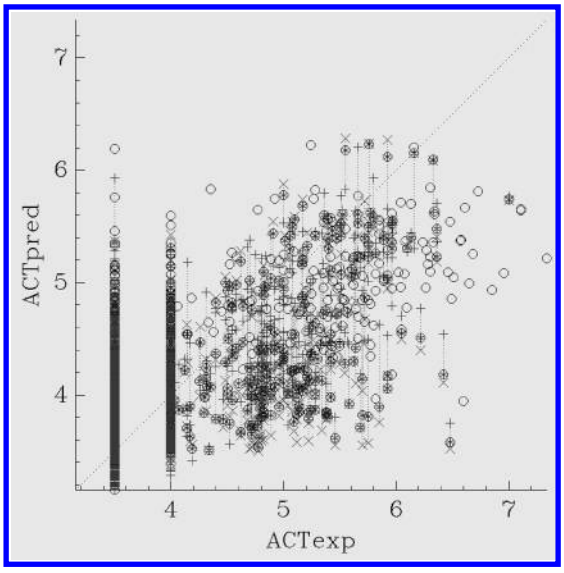

Figure 4. Comparison of experimental pIC50 values [pIC50 = −log(IC50 in M), ACTexp] and calculated pIC50 values using a "standard" BPPM synergy QSAR model (ACTpred) developed for the inhibition of CYP3A4 (BFC substrate). Compounds with "estimated" experimental pIC50 values appear at ACTexp values of 3.5 (316 μM) and 4.0 (100 μM) (cf. Figure 3). Activity values predicted using a synergy approach are marked by circles spiked by an asterisk and bracketed by linear (+) and neighborhood (×) predicted values, whereas values predicted from linear equations only are marked by hollow circles.

screening application of these models, it is necessary to determine in advance whether a compound is likely to belong to which specific subgroup for a given substrate (i.e., either the proximal or the distal binding group in our MPH strategy). Because the MPH models were developed independently for the subgroups, and therefore the QSAR equations for the subgroups do not differentiate compounds from out groups, a group classifier is needed for determining the proximal or distal binding of a compound for which the activity is to be predicted. We use a support vector machine (SVM) algorithm[38] to construct the classifier from the combined descriptor set of the compounds in the two subgroups for each substrate. Similar to other cheminformatics applications of support vector machines,[39,40] the SVM classifier is sufficiently robust and achieves an overall 95% accuracy for the training set, with a 5-fold cross-validation accuracy of about 75%.

RESULTS AND DISCUSSION

Initial QSAR Studies of the Inhibition of CYP3A4 Metabolism of BFC. Because of the identical experimental procedures for the measurement of enzyme inhibition, the model for CYP3A4 inhibition derived from 50 diverse drug molecules by Wanchana et al.[26] is applied to the BFC data set in BioPrint. As shown in Figure 3, the CYP3A4 inhibitory activity of the compound set in BioPrint is poorly estimated by this model, limiting its application as a general, predictive model for compounds structurally even more diverse than the 50 compounds used in building the model. On the other hand, given the previous success for modeling CYP2D6 inhibition,[25] we attempt to build a QSAR model for the inhibition of the metabolism of BFC by CYP3A4 on the basis of the "standard" fuzzy pharmacophore-based descriptors implemented in the BPPM.[29,30] This "standard" BPPM synergy model, however, is shown to be poorly fitted (r^2 of 0.57 and mean-squared-error of 0.51, Figure 4) and, thus, likely to be ineffective as a filter because it underestimates the activity of compounds that strongly inhibit CYP3A4 (e.g., compounds with IC50 values lower than 1 μM, or equivalently, pIC50 values greater than 6.0, where pIC50 is −log[IC50, in M]). Such compounds would become a liability if they were to be taken further into drug development on the basis of the in silico predictions. Thus, unlike the enzyme inhibition in CYP2D6 that is more accessible to QSAR modeling,[25,41−43] CYP3A4 inhibition proves to be more difficult and does not benefit from an expanded structure−activity data set (such as the BioPrint data set) nor from using diverse descriptors (pharmacophore descriptors combined with topological and fragment descriptors). We seek some rationale for the generally poor results of QSAR modeling for CYP3A4 inhibition and, therefore, a better understanding of the enzyme−inhibitor interactions that may help improve the modeling. This motivation leads to the alternative hypothesis of multiple pharmacophores of ligand binding in the CYP3A4 catalytic pocket.

Rationale and Supporting Evidence for MPH. As described in the Introduction, the proficiency of the CYP3A4 enzyme in recognizing the wide variety of molecules and in catalyzing the broad range of metabolizing reactions suggests that the enzyme has a rather large binding pocket that is equipped to accommodate and achieve a broad range of molecular recognition.[9] In addition, enzyme kinetics studies show noncompetitive inhibition for many of its substrates, indicating that multiple substrates and inhibitors can simultaneously bind in the catalytic pocket of the CYP3A4 enzyme.[9−16] Thus, a potential shortcoming of the conventional QSAR approach (Figures 3 and 4) for modeling the activity of an enzyme such as CYP3A4 lies in the assumption

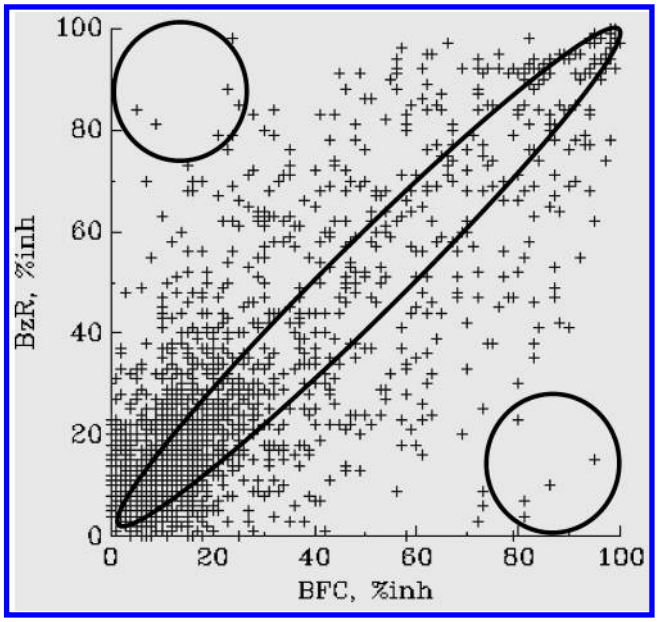

Figure 5. Relative inhibitory activity (percent inhibition at 10 μM in the primary screening assays) of 1782 BioPrint compounds for two substrates of CYP3A4, BzR and BFC. The circles indicate locations for compounds that strongly inhibit the metabolism of only one substrate. Compounds in the diagonal region (e.g., the area enclosed by the ellipse) inhibit the metabolism of both substrates equally.

that the inhibitors bind in a more or less similar binding geometry, and a single set of pharmacophoric descriptors (once found) is sufficient in modeling the structure−activity relationship. QSAR modeling without the recognition of multiple pharmacophores[26,43−45] amounts to an attempt to derive a relationship for the wide-ranging inhibitory activities (of the proximal and the distal binding subgroups, for example) from the relatively limited set of interactions probed by single substrates. Thus, a "simple pharmacophore" approach is likely to be limited in the derivation of useful structure−activity relationships for the inhibition of CYP3A4.

A further support for the consideration of a multiple pharmacophore description for the enzyme inhibition of CYP3A4 is provided by the comparison of the primary screening inhibition data for two different substrates, as shown in Figure 5 for BFC and BzR. For two substrates with non-overlapping binding sites, some inhibitors could be expected to interfere with the binding of only one substrate but not the other and, therefore, populate the off-diagonal regions in Figure 5. In the other extreme, the binding sites of two substrates might share a high degree of overlap, and the region near the diagonal in Figure 5 would be expected to be more highly occupied: a compound that inhibits the binding of one substrate because of steric hindrance is likely to inhibit the binding of the other substrate because of similar steric effects. Although subtle differences in the key interactions with the enzyme and the steric effects will ultimately combine to determine the inhibitory activities of a compound, the actual distribution shown in Figure 5 suggests that the inhibition of the metabolism of these two substrates is neither of the two extremes, namely, that the binding sites of the two substrates are partially overlapping such that most inhibitors interfere differentially with the metabolism of both substrates but not exclusively with the metabolism of only one substrate.

MPH-Directed QSAR Modeling. From the schematic diagram shown in Figure 2, a potential successful application of MPH to QSAR modeling would be the proper separation and identification of multiple pharmacophore sets that are specific for the recognition of different substrates when possible. The relationship of the inhibitory activities shown in Figure 5 for the substrate pair of BFC and BzR suggests that the data sets can be used for identifying proximal and distal binders under MPH. Thus, we make a pairwise comparison of the two inhibition data sets for BFC and BzR, as described in the MPH subsection in the Materials and Methods, and divide the IC50 data set for each substrate into two subgroups (Table 1), one for proximal binding (group p) and one for distal binding (group d), each for an independent search of linear QSAR models. For a better signal-to-noise ratio, we choose the IC50 values instead of the primary screening inhibition values for the activity comparison. The strategy of activity comparison is essentially a first-order approximation of the complex binding relationships among the two substrates and the inhibitors. It attempts to identify separate subgroups that categorize the approximate geometric relationship of the inhibitor binding relative to the substrates. As a reference, a QSAR model was also derived for the original undivided data set (group R). The results from the QSAR equation search for fitting the activity values of BFC and BzR are shown in Figure 6a and b, respectively. (For the purpose of developing and validating MPH, only multivariate linear modeling in the BPPM is employed.)

The statistical parameters of the QSAR models derived from model fitting for MPH groups shown in Figure 6a and b are summarized in Table 2. Also included in Table 2 for a direct comparison are the corresponding parameters for the linear portion of the "standard BPPM model" (Figure 4). For each model, MSE and r^2 are listed; see ref 46, for example, for a discussion of the use of these parameters in regression analysis. Also listed is the compound parameter MSE/r^2, which can be considered as an "effective" MSE: given two models of equal MSE values, the one with a higher value of r^2 (better correlation) will be considered as the better model and indicated as such by the smaller value for the effective MSE. Thus, the ratio MSE/r^2 serves as a single parameter for comparing the fitness of QSAR models. The MPH models for groups p and d in Table 2 are shown to have smaller mean-squared errors and higher correlation coefficients (thus, smaller MSE/r^2 ratios) than the reference model of group R, indicating the better fit of the respective QSAR equations to the experimental data for the proximal and the distal groups individually. The fitting of the activity of highly active compounds (outliers in the reference model of group R) is significantly improved in MPH models (Figure 6a,b). These comparisons show that the MPH and the division of the compound set into the proximal and the distal groups are rational and useful approaches for improving the modeling of CYP3A4 inhibition.

As a negative control, pairs of groups are constructed by random selections from group R such that the group sizes of each pair correspond to groups p and d (Table 2). Here, the individual control models may have better statistics; for example, the C1p model for BFC shows a lower MSE and a higher r^2 than the MPH model of group p, or the C2p model shows a smaller MSE. But, the MSE/r^2 ratios for each pair of control groups are higher than those of the MPH models

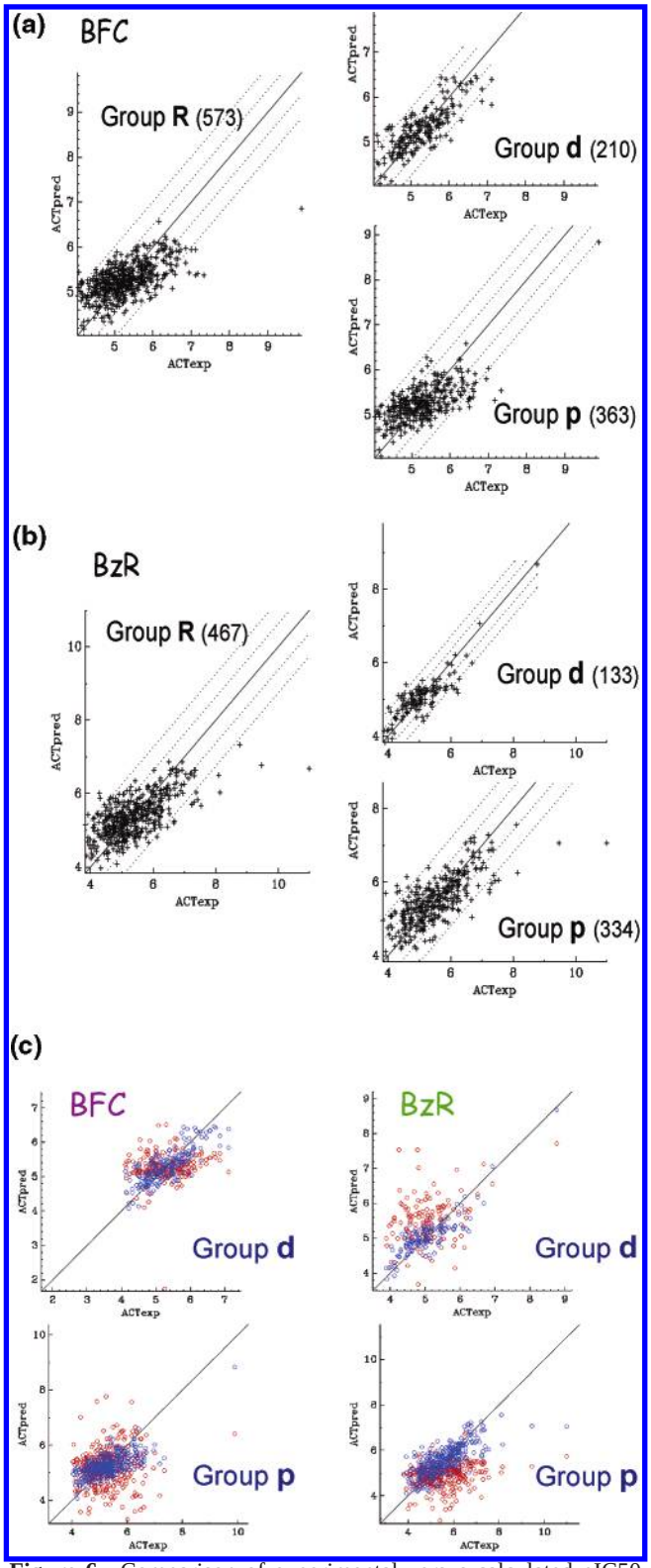

Figure 6. Comparison of experimental versus calculated pIC50 values for various linear MPH models for the BFC substrate (a) and for the BzR substrate (b). The number of compounds for each model is enclosed in parentheses. The two sets of lines parallel to the diagonal mark are, respectively, one- and two-standard-errors away from the perfect linear relationship. (c) Cross-group predictions from the subgroup models for each substrate. The native data for each model (in blue) are reproduced from Figure 6a and b. For each substrate, the calculated values for compounds in group d using the group p model and vice versa are shown in red. The poor distribution of the data points obtained from cross-group model calculations (marked in red) validates the specificity of each subgroup model.

Table 2. Statistical Parameters of QSAR Models[a]

"Standard" BPPM Model (Full BioPrint Compound Set)

CYP 3A4−BFC			
	MSE	r^2	MSE/r^2
CYP 3A4 (1590)	0.505	0.567	0.892

MPH (BioPrint Compounds with Experimental IC50)

CYP 3A4−BFC				CYP 3A4−BzR			
data group	MSE	r^2	MSE/r^2	data group	MSE	r^2	MSE/r^2
R (573)	0.262	0.369	0.711	R (467)	0.404	0.459	0.880
p (363)	0.245	0.446	0.550	p (334)	0.363	0.545	0.666
d (210)	0.140	0.622	0.224	d (133)	0.125	0.728	0.172
C1p (363)	0.207	0.487	0.425	C1p (334)	0.388	0.490	0.793
C1d (210)	0.220	0.491	0.449	C1d (133)	0.178	0.753	0.237
C2p (363)	0.201	0.416	0.484	C2p (334)	0.375	0.505	0.742
C2d (210)	0.233	0.565	0.412	C2d (133)	0.173	0.751	0.230
C3p (363)	0.270	0.374	0.723	C3p (334)	0.344	0.561	0.613
C3d (210)	0.161	0.582	0.276	C3d (133)	0.235	0.629	0.374
C4p (363)	0.279	0.398	0.701	C4p (334)	0.395	0.506	0.780
C4d (210)	0.150	0.539	0.278	C4d (133)	0.207	0.667	0.311
C5p (363)	0.243	0.410	0.594	C5p (334)	0.411	0.491	0.836
C5d (210)	0.171	0.593	0.288	C5d (133)	0.171	0.722	0.236
				C6p (334)	0.324	0.542	0.597
				C6d (133)	0.213	0.747	0.285

[a] Parameters for the linear portion of the "standard" BPPM model are included for a direct comparison with MPH models. Each model is identified by the label shown in column 1, with the number of compounds enclosed in parentheses. *F* values (calculated according to ref 50) are well above 10.0 ($P \ll 0.05$) for all models. The ratio of mean-squared error (MSE, column 2) to r^2 (column 3) is an "effective MSE" as described in the text. It is used as a compound indicator for the comparison of the quality of models (column 4). QSAR models for group p and group d are referred to as MPH models, whereas the model for group R (combined from compounds in both groups p and d) is the reference model. For negative control, group R is randomly divided into two subgroups of appropriate sizes; these are identified by C1, C2, etc. The reference model from group R shows improvement over the linear BPPM model, a result likely due to modeling on only active compounds.

as a pair; namely, the pair of control group models do not have MSE/r^2 ratios that are both better than the corresponding group p and group d models. This indicates the specificity of the constitution of the MPH groups above the random, nonspecific grouping from the data set. Figure 6c shows the cross-group predictions of the group p and the group d models for each substrate, that is, substituting descriptor values of members in group d into the QSAR equation of group p and vice versa. The poor predictions from cross-group models further validate the specificity of MPH models derived for the subgroups.

For implementing the MPH strategy of comparing the relative activity of a compound against two different substrates, we chose the IC50 data set given that IC50 values have a higher signal-to-noise ratio and that they are representative of the dose dependency of the inhibitory effect. As a result, however, weak inhibitors (i.e., compounds with a lower than 30% inhibition at a 10 μM concentration, cf. the in vitro assay subsection in the Materials and Methods) are excluded from MPH modeling (Table 1). Lowering the threshold for IC50 determination from 30% would somewhat remedy this drawback, by increasing the number of IC50 values in the data set, to the extent that high IC50 values can be reliably obtained experimentally in such cases.

Table 3. (a) Comparison of MPH Models for BFC,[a] (b) Descriptor Comparison between MPH Models for BFC and BzR,[b] and (c) Descriptor Comparison with Models from Random Subgroups[c]

(a)

BFC proximal		BFC distal		BFC reference	
fr2_thiazole	0.8875	fr2_imidazole	0.4095	zexp(MonH+)	-0.3687
S_aaS	0.1788	MonH+	0.5357	zexp2(IINC)	0.8868
SArom^2	0.2149	zsig3(HD-HD5)	0.8021	fr2_imidazole	0.1623
zsig3(HA-HD11)	-0.9072	DistH+	-0.0358	zexp(Hp-HA6)	-0.2462
HA-NC4	-0.071	zsig3(#H)	-0.2994	Ar-HA13	0.0566
zexp(#A)	-0.3013	Ar-HA6	0.062	Ar-HA6	0.0336
zsig3(Hp-NC10)	0.5442	HA-PC7	-0.2588	DipH-	-0.1451
zexp2(HD-HD5)	-0.4428	Hp-Ar7	-0.021	fr2_sulfide	0.1704
S_ssO	0.0121	zsig3(HD-PC12)	1.3738	fr2_intrahbonds	-0.1337
Ar-HD7	-0.0286	Ar-PC11	0.3305	zsig3(LD~RA)	0.5658
chi3v_ch	-1.616	Ar-PC10	-0.4407	zsig3(HA-HD11)	-0.7284
fr2_nitrile	-0.2819	zexp2(LH~RD)	0.4974	LRA~RD	-0.3989
u:5(Ar-HD5)	-0.3113	zexp2(Ar-NC12)	0.8406	fr2_nitrile	-0.3187
zsig3(LD~P)	-0.4541	LP~RD	1.266	zexp2(LR~AD)	-0.2304
LRA~RD	-0.3044	Hp-HD7	0.0659	fr2_sulfonamd	0.2020
zsig3(LD~RA)	0.7611	Hp-Hp14	-0.1637	HA-HD12	-0.2143
fr2_sulfonamd	0.2658	zexp(HA-HA7)	-0.5382	zexp2(HD-PC4)	0.2176
HA-HA14	-0.1998	LA~RA	-0.1698	fr2_quatN	-0.4341
HA-HD12	-0.2115	Ar-PC9	0.2166	fr2_C=O_noCOO	-0.0564

(b)

BFC proximal		BzR distal		BzR proximal		BFC distal	
fr2_thiazole	0.8875	zexp(H~AD)	-1.3274	zsig3(SH)	-1.4414	fr2_imidazole	0.4095
S_aaS	0.1788	Ar-HA13	0.2318	fr2_imidazole	0.4874	MonH+	0.5357
SArom^2	0.2149	zexp2(HD-HD5)	-0.5173	SArom^2	0.4641	zsig3(HD-HD5)	0.8021
zsig3(HA-HD11)	-0.9072	zsig3(Ar-NC8)	1.4029	LH~A	0.0132	DistH+	-0.0358
HA-NC4	-0.071	chi3v_c	0.2548	Ar-HA6	0.0621	zsig3(#H)	-0.2994
zexp(#A)	-0.3013	zexp(CosA3)	0.2426	zexp(Hp-Ar14)	-0.9876	Ar-HA6	0.062
zsig3(Hp-NC10)	0.5442	zexp(H~R)	-0.6466	Hp-Ar13	-0.111	HA-PC7	-0.2588
zexp2(HD-HD5)	-0.4428	H~A	0.1562	zsig3(Hp-NC6)	1.0653	Hp-Ar7	-0.021
S_ssO	0.0121	zsig3(HD-HD11)	-0.6387	zsig3(SHBD^2)	-1.4584	zsig3(HD-PC12)	1.3738
Ar-HD7	-0.0286	zexp(HA-HA14)	1.8895	zsig3(HD-PC8)	0.744	Ar-PC11	0.3305
chi3v_ch	-1.616	chi3_ch	-1.757	LRA~RD	-0.716	Ar-PC10	-0.4407
fr2_nitrile	-0.2819	zexp(LPI)	-0.8039	fr2_quatN	-1.109	zexp2(LH~RD)	0.4974
u:5(Ar-HD5)	-0.3113	chi4_c	-0.1869	zsig3(Hp-HA14)	0.9658	zexp2(Ar-NC12)	0.8406
zsig3(LD~P)	-0.4541	Hp-HA11	-0.1091	zexp(LPI)	-0.498	LP~RD	1.266
LRA~RD	-0.3044	zexp2(HA-PC5)	0.3581	zexp(HD-PC13)	-0.4933	Hp-HD7	0.0659
zsig3(LD~RA)	0.7611	zexp2(HA-PC14)	-2.1732	zexp(HA-HA12)	-0.5333	Hp-Hp14	-0.1637
fr2_sulfonamd	0.2658	Ar-PC12	-0.3688	HA-HD8	-0.1897	zexp(HA-HA7)	-0.5382
HA-HA14	-0.1998	Hp-PC10	-0.1886	HA-HD14	-0.7159	LA~RA	-0.1698
HA-HD12	-0.2115	zexp2(D~AD)	-0.5152	zexp(Ar-HD10)	0.3792	Ar-PC9	0.2166

(c)

BFC proximal(random)		BzR distal		BzR proximal		BFC distal(random)	
zexp2(LH~R)	-0.3645	zexp(H~AD)	-1.3274	zsig3(SH)	-1.4414	S_aaS	0.8018
zexp(IINC)	0.7533	Ar-HA13	0.2318	fr2_imidazole	0.4874	LR~A	0.0208
zsig3(Ar-HD15)	-0.8268	zexp2(HD-HD5)	-0.5173	SArom^2	0.4641	zexp2(LA~AD)	-1.0422
Ar-Ar8	0.0215	zsig3(Ar-NC8)	1.4029	LH~A	0.0132	zexp2(Ar-Ar10)	-0.4023
Hp-PC5	0.1025	chi3v_c	0.2548	Ar-HA6	0.0621	S_aasN	0.1521
fr2_imidazole	0.1953	zexp(CosA3)	0.2426	zexp(Hp-Ar14)	-0.9876	zexp(#A)	-0.6186
HD-HD6	-0.2912	zexp(H~R)	-0.6466	Hp-Ar13	-0.111	chi1v	0.0422
LRA~RD	-0.4202	H~A	0.1562	zsig3(Hp-NC6)	1.0653	zsig3(Ar-NC10)	1.3751
fr2_Imine	-0.3153	zsig3(HD-HD11)	-0.6387	zsig3(SHBD^2)	-1.4584	Ar-Ar4	0.1268
fr2_imide	-0.4758	zexp(HA-HA14)	1.8895	zsig3(HD-PC8)	0.744	Hp-Ar5	-0.0945
fr2_nitrile	-0.3465	chi3_ch	-1.757	LRA~RD	-0.716	zsig3(P~RD)	1.9955
zsig3(Ar-HA11)	-0.5903	zexp(LPI)	-0.8039	fr2_quatN	-1.109	HA-HD12	-0.3015
DistH-	-0.0858	chi4_c	-0.1869	zsig3(Hp-HA14)	0.9658	HA-HA4	-0.0945
zexp(LH~D)	-0.2713	Hp-HA11	-0.1091	zexp(LPI)	-0.498	zexp2(Ar-PC6)	-0.4649
fr2_pyridine	-0.1859	zexp2(HA-PC5)	0.3581	zexp(HD-PC13)	-0.4933	zexp2(LA~P)	-0.3566
Ar-PC6	-0.0644	zexp2(HA-PC14)	-2.1732	zexp(HA-HA12)	-0.5333	LR~P	-0.0998
zexp2(Ar-HA10)	0.2015	Ar-PC12	-0.3688	HA-HD8	-0.1897	S_dssC	-0.0785
zexp(Ar-HA4)	0.2797	Hp-PC10	-0.1886	HA-HD14	-0.7159	fr2_thiophene	-0.7871
Ar-HA12	-0.0767	zexp2(D~AD)	-0.5152	zexp(Ar-HD10)	0.3792	PC-PC4	0.3575

[a] For each MPH model, the descriptors and their corresponding coefficients in the QSAR equation is listed. Names of fragment descriptors for functional groups are preceded with the substring "fr2_"; names of topological descriptors (e.g., chi3v_ch and S_aaS) follow the convention of Kier and Hall.[31−34] Other descriptors, such as "HA-NC4", belong to the class of pharmacophoric descriptors in the BPPM; as an example, "HA-NC4" specifies the pharmacophore feature pair of hydrogen-bond acceptor (HA) and negatively charged (NC) separated at a distance of 4 Å. Other pharmacophore features and additional transformation functional forms (e.g., zexp for exponential or zsig for sigmoidal) are summarized in the text, and the full details can be found in the references for the BPPM.[25,29,30] Common descriptors between subgroup models and that of group R are shown color-coded and are described in detail in the text. The MPH models for groups p and d show additional group-specific and contrasting features (descriptors enclosed in boxes), for example, the thiazole group in the proximal binding vs imidazole in distal binding or the bipolar pharmacophore feature pair of aromatic-hydrogen-bond donor vs the aromatic-hydrogen-bond acceptor pair. [b] Common descriptors (in light blue) and similar descriptors (in dark blue) are found for the pair of MPH models (BFC group p and BzR group d) and similarly for the pair of MPH models for BzR group p and BFC group d. Common descriptors are also found for the two proximal models (in orange). These observations indicate that the models characterize two substrate-binding sites in close proximity with (and possibly overlapping) each other. [c] As a negative control, descriptors in MPH models for a "random subgroup" pair for BFC (listed in Table 2 and described in the text) are compared with MPH models for BzR (i.e., replacing the BFC models in Table 3b). Descriptors in the BzR models do not appear in random subgroup models, even those descriptors that are common or similar to MPH models for BFC (in light blue and dark blue, respectively, inherited from Table 3b).

MPH Models for BFC and BzR Inhibition. The characteristics and the internal consistency of the MPH models for BFC and BzR are illustrated by a detailed examination of descriptor variables in the QSAR equations. The descriptor variables in a QSAR model equation represent the structural features of substrates and inhibitors that the enzyme recognizes for binding specificity. For example, the sulfonamide and the nitrile functional groups are shared by proximal binders for BFC, while the imidazole group is shared by distal binders, and similarly, the field descriptor LRA∼RD (for aromatic-hydrogen-bond donor groups in the aromatic-hydrogen-bond acceptor field) and the bipolar descriptor Ar-HA6 (for the aromatic-hydrogen-bond-acceptor feature pair at a 6 Å distance) are specific to proximal and distal binders, respectively (Table 3a). Color-coded to either the proximal or the distal model, these group-specific structural features appear concomitantly in the group R model, for which the whole IC50 data set is used and no attempt has been made in the distinction between the proximal or the distal binding. The separation of the data set into the proximal and distal groups thus allows the structure features of the binding sites for the two substrates to be selected specifically for each model. On the other hand, the proximity (and the possible overlap) of the binding sites of BFC and BzR is indicated by the common descriptor terms (descriptor type and descriptor coefficient) between the two proximal models (color-coded in orange in Table 3b) and the common descriptor terms of the proximal model of one substrate and the distal model of the other (color-coded in light and dark blue in Table 3b). Last, models from control groups (random subgroups for negative control, Table 2) do not share common structural features with MPH subgroup models (Table 3c), indicating that the subgroups established from the multiple pharmacophore hypothesis uniquely specify features for the proximal and the distal binding sites.

MPH Models for TES and MID. Two other inhibition data sets are available in BioPrint, using TES or MID as the substrate for the CYP3A4 inhibition assay. Similar to the data sets for BFC and BzR, these data sets are also candidates for the MPH modeling approach. For the TES−BFC pair, the comparative binding is shown in Figure 7a, and the statistics of MPH models are shown in Table 4a. Similar to Table 2, the MSE/r^2 ratios here are smaller for the MPH models for both substrates than those of the corresponding reference models (and the control models generally). The pairing of TES with BzR under MPH is shown in Table 4b, although the MPH models from this pairing are less favorable than those from the TES−BFC pair because the distal group for BzR is much smaller in size in comparison to the proximal group and, thus, is more susceptible to an overfitting of the experimental data in the QSAR equation search. Unlike other substrate pairs, the comparative activity for MID with other substrates is highly asymmetric (cf. Figure 7b for the MID−BFC pair). This asymmetry is due to the generally lower inhibitory activity of BioPrint compounds against MID, which is consistent with the higher apparent K_M for midazolam relative to those of the other CYP3A4 substrates.[7,27,47−49] Thus, for midazolam, the relationship of the activity relative to those of other substrates is different from the relationship of other substrate pairs (perhaps because of unique binding characteristics of the midazolam binding site). The MPH approximation for subgroup identification would place most compounds in the distal group for MID, clearly an oversimplification that requires refinement of the MPH strategy. In practice, because of the asymmetric distribution, and therefore the highly unblanaced group sizes (as shown in Figure 7b), the smaller of the two subgroups is at a risk of model overfitting.

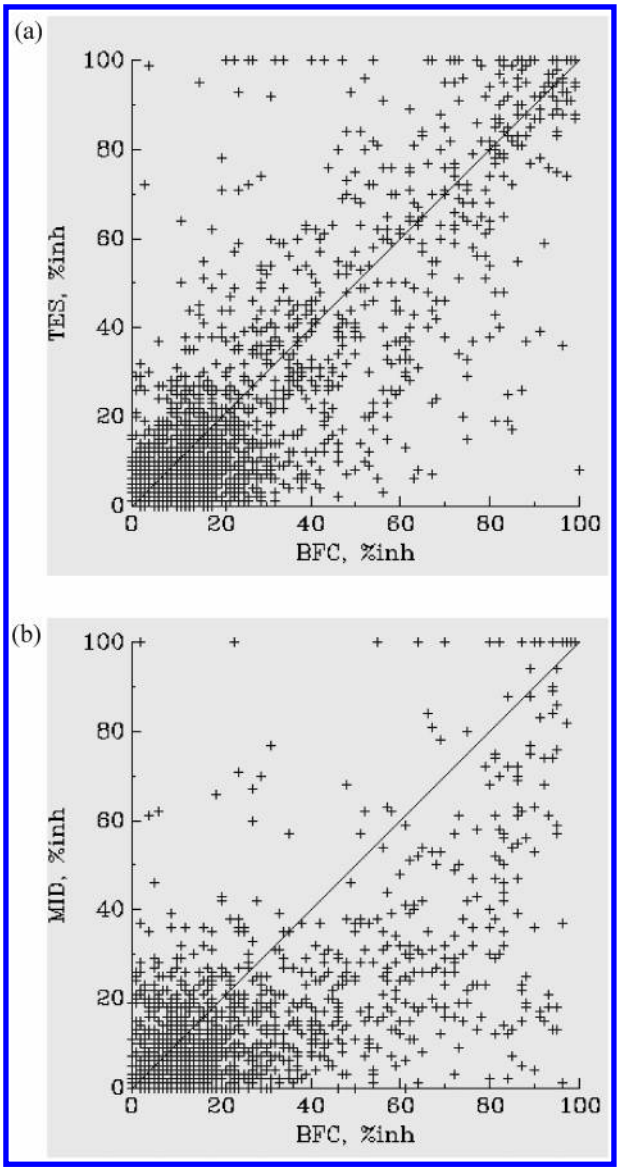

Figure 7. Comparison of primary inhibition data sets for testosterone and BFC (a) and for midazolam and BFC (b).

Application of MPH Models for Prediction ("When and How" To Use MPH Models). For the MPH modeling of in vitro CYP3A4 inhibition, the full IC50 data sets in BioPrint at the time had been used in the derivation of QSAR models, with no data left for a test data set. However, since the MPH models have been developed, additional IC50 values for BFC and BzR (13 and 15 compounds, respectively) have been obtained experimentally and entered into the BioPrint database. We use these new data to test the MPH models and to develop protocols for virtual screening applications. For testing and virtual screening purposes, the SVM classifier (described in Materials and Methods) is used for determining whether the group p model or the group d model should be used for predicting the inhibitory activity of a compound. For the BFC substrate, one of the 13 compounds is classified

Table 4. Statistics for MPH Models for Testosterone[a]

(a) MPH (BioPrint Compounds with Experimental IC50)

CYP 3A4−BFC				CYP 3A4−TES			
data group	MSE	r^2	MSE/r^2	data group	MSE	r^2	MSE/r^2
R (573)	0.268	0.354	0.758	R (433)	0.221	0.374	0.591
p (442)	0.258	0.399	0.647	p (213)	0.158	0.607	0.261
d (131)	0.130	0.639	0.203	d (220)	0.176	0.422	0.416
C1p (442)	0.263	0.393	0.670	C1p (213)	0.165	0.514	0.321
C1d (131)	0.085	0.761	0.112	C1d (220)	0.213	0.418	0.511
C2p (442)	0.249	0.381	0.654	C2p (213)	0.129	0.546	0.236
C2d (131)	0.141	0.695	0.202	C2d (220)	0.211	0.505	0.417
C3p (442)	0.251	0.409	0.614	C3p (213)	0.152	0.500	0.304
C3d (131)	0.092	0.756	0.122	C3d (220)	0.217	0.456	0.476
C4p (442)	0.265	0.367	0.723				
C4d (131)	0.112	0.722	0.155				
C5p (442)	0.270	0.364	0.743				
C5d (131)	0.119	0.692	0.172				

(b) MPH (BioPrint Compounds with Experimental IC50)

CYP 3A4−BzR				CYP 3A4−TES			
data group	MSE	r^2	MSE/r^2	data group	MSE	r^2	MSE/r^2
R (467)	0.412	0.447	0.922	R (433)	0.220	0.379	0.580
p (379)	0.376	0.523	0.718	p (209)	0.163	0.458	0.356
d (88)	0.092	0.659	0.140	d (224)	0.171	0.563	0.303
C1p (379)	0.382	0.480	0.796	C1p (209)	0.174	0.497	0.350
C1d (88)	0.156	0.807	0.193	C1d (224)	0.166	0.537	0.308
C2p (379)	0.340	0.461	0.737	C2p (209)	0.172	0.471	0.366
C2d (88)	0.142	0.886	0.160	C2d (224)	0.162	0.573	0.283
C3p (379)	0.412	0.432	0.954	C3p (209)	0.144	0.555	0.260
C3d (88)	0.094	0.887	0.106	C3d (224)	0.163	0.570	0.286
				C4p (209)	0.147	0.520	0.282
				C4d (224)	0.206	0.478	0.431

[a] Statistical parameters for the testosterone−BFC pair and the testosterone−BzR pair are shown in a and b, respectively.

by the SVM classifier to be a distal binder, and the predicted pIC50 value has an error of only 0.06; for the remaining 12 compounds, the predicted pIC50's from the proximal model have an average error of 0.62. For BzR, all 15 compounds are classified in the proximal group, and the average error in pIC50's is 0.60. These results are graphically summarized in Figure 8, showing that the prediction errors are within the range of fitting errors. These models will be validated in this manner continually as new data become available. Other classifiers in addition to the SVM classifier could also be developed and used in the virtual screening application.

CONCLUSIONS

Given that CYP3A4 metabolizes a large variety of drugs, the inhibition of the enzyme cannot be effectively detected by assays using any single substrate. Not inconsistently, even for a single substrate, the inhibition data could not be adequately summarized and quantitatively represented by a single global QSAR model. The MPH approach substantiates the notion that the structure−activity relationship for this enzyme would be better summarized with local models. For each substrate, a proximal and a distal binding group are identified in our MPH modeling. These are local models, each being specific to a general physical locality of the binding pocket. Implicitly, they are also "local models" in the descriptor space because each subgroup leads to a QSAR model with a specific set of descriptor variables.

As a simple, first-order approximation, MPH conceptually provides the basis for the strategy of comparative binding

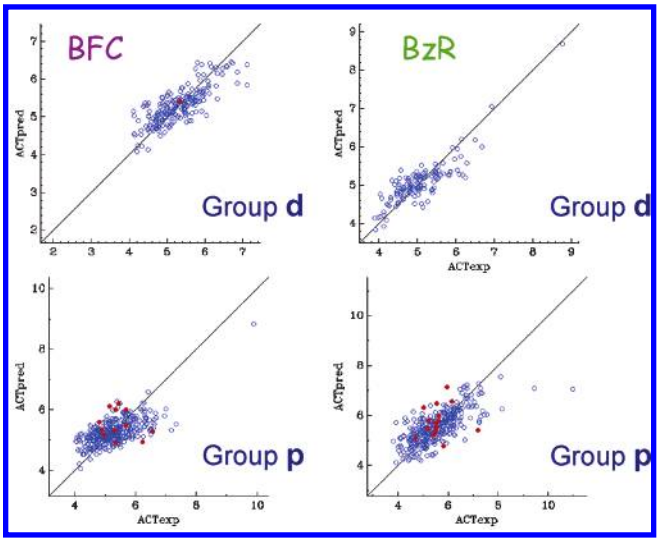

Figure 8. MPH model prediction of new IC50's for BFC and BzR. The data for building the MPH models (in blue) are reproduced from Figure 6a and b. These data provided a reference for evaluating the prediction errors for the new compounds (in red). See text for details.

that makes use of a pairwise comparison of IC50 activity values for different substrates. By this pairwise comparison, it becomes possible to categorize the proximal and the distal binding relative to each substrate and to gain structural insight into how multiple substrates of CYP3A4 may interact with the enzyme (e.g., the degree to which their binding sites may lie in close proximity or even overlap). This can provide a basis for the elaboration of other strategies and for the further understanding of the binding and the inhibition of the enzyme by individual inhibitors. From the strategic perspective of MPH, the requirement of IC50 activity values in the pairwise comparison strategy removes the inactive compounds and their confounding effects from the QSAR modeling. This also points to the need for an in silico method for the differentiation of active and inactive compounds, emphasizing further the concept of local modeling for the enzyme inhibition of CYP3A4.

ACKNOWLEDGMENT

We thank Mark Crawford and Frédéric Revah for support and for critical reading of the manuscript. We also acknowledge the comments and suggestions from the reviewers.

REFERENCES AND NOTES

(1) Coon, M. J. Cytochrome P450: Nature's Most Versatile Biological Catalyst. *Annu. Rev. Pharmacol. Toxicol.* **2005**, *45*, 1−25.
(2) Hardman, J. G.; Limbird, L. E.; Gilman, A. G. *Goodman and Gilman's The Pharmacological Basis of Therapeutics*, 10th ed.; McGraw Hill: New York, 2001; p 15.
(3) Rendic, S. Summary of Information on Human CYP Enzymes: Human P450 Metabolism Data. *Drug Metab. Rev.* **2002**, *34*, 83−448.
(4) Shimada, T.; Yamazaki, H.; Mimura, M.; Inui, Y.; Guengerich, F. P. Inter-individual Variations in Human Liver Cytochrome P-450 Enzymes Involved in the Oxidation of Drugs, Carcinogens and Toxic Chemicals: Studies with Liver Microsomes of 30 Japanese and 30 Caucasians. *J. Pharmacol. Exp. Ther.* **1994**, *270*, 414−423.
(5) Furukawa, M.; Nishimura, M.; Ogino, D.; Chiba, R.; Ikai, I.; Ueda, N.; Naito, S.; Kuribayashi, S.; Moustafa, M. A.; Uchida, T.; Sawada, H.; Kamataki, T.; Funae, Y.; Fukumoto, M. Cytochrome P450 Gene Expression Levels in Peripheral Blood Mononuclear Cells in Comparison with the Liver. *Cancer Sci.* **2004**, *95*, 520−529.
(6) Shou, M.; Grogan, J.; Mancewicz, J. A.; Krausz, K. W.; Gonzalez, F. J.; Gelboin, H. V.; Korzekwa, K. R. Activation of CYP3A4: Evidence

for the Simultaneous Binding of Two Substrates in a Cytochrome P450 Active Site. *Biochemistry* **1994**, *33*, 6450−6455.
(7) Wang, R. W.; Newton, D. J.; Liu, N.; Atkins, W. M.; Lu, A. Y. H. Human Cytochrome P-450 3A4: In Vitro Drug-Drug Interaction Patterns are Substrate-Dependent. *Drug Metab. Dispos.* **2000**, *28*, 360−366.
(8) Atkins, W. M.; Wang, R. W.; Lu, A. Y. H. Allosteric Behavior in Cytochrome P450-Dependent in Vitro Drug−Drug Interactions: A Prospective Based on Conformational Dynamics. *Chem. Res. Toxicol.* **2002**, *14*, 338−347.
(9) Ekins, S.; Stresser, D. M.; Williams, J. A. In vitro and Pharmacophore Insights into CYP3A Enzymes. *Trends Pharmacol. Sci.* **2003**, *24*, 161−166.
(10) Koley, A. P.; Robinson, R. C.; Markowitz, A.; Friedman, F. K. Drug−Drug Interactions: Effect of Quinidine on Nifedipine Binding to Human Cytochrome P450 3A4. *Biochem. Pharmacol.* **1997**, *53*, 455−460.
(11) Domanski, T. L.; He, Y.-A.; Harlow, G. R.; Halpert, J. R. Dual Role of Human Cytochrome P450 3A4 Residue Phe-304 in Substrate Specificity and Cooperativity. *J. Pharmacol. Exp. Ther.* **2000**, *293*, 585−591.
(12) Schrag, M. L.; Wienkers, L. C. Covalent Alteration of the CYP3A4 Active Site: Evidence for Multiple Substrate Binding Domains. *Arch. Biochem. Biophys.* **2001**, *391*, 49−55.
(13) Lu, P.; Lin, Y.; Rodrigues, A. D.; Rushmore, T. H.; Baillie, T. A.; Shou, M. Testosterone, 7-Benzyloxyquinoline, and 7-Benzyloxy-4-trifluoromethyl-coumarin Bind to Different Domains within the Active Site of Cytochrome P450 3A4. *Drug Metab. Dispos.* **2001**, *29*, 1473−1479.
(14) Kenworthy, K. E.; Clarke, S. E.; Andrews, J.; Houston, J. B. Multisite Kinetic Models for CYP3A4: Simultaneous Activation and Inhibition of Diazepam and Testosterone Metabolism. *Drug Metab. Dispos.* **2001**, *29*, 1644−1651.
(15) Shou, M.; Dai, R.; Cui, D.; Korzekwa, K. R.; Baillie, T. A.; Rushmore, T. H. A Kinetic Model for the Metabolic Interaction of Two Substrates at the Active Site of Cytochrome P450 3A4. *J. Biol. Chem.* **2001**, *276*, 2256−2262.
(16) Hosea, N. A.; Miller, G. P.; Guengerich, F. P. Elucidation of Distinct Ligand Binding Sites for Cytochrome P450 3A4. *Biochemistry* **2000**, *39*, 5929−5939.
(17) Yano, J. K.; Wester, M. R.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. The Structure of Human Microsomal Cytochrome P450 3A4 Determined by X-ray Crystallography to 2.05-Å Resolution. *J. Biol. Chem.* **2004**, *279*, 36091−36094.
(18) Williams, P. A.; Cosme, J.; Vinkovic, D. M.; Ward, A.; Angove, H. C.; Day, P. J.; Vonrhein, C.; Tickle, I. J.; Jhoti, H. Crystal Structures of Human Cytochrome P450 3A4 Bound to Metyrapone and Progesterone. *Science* **2004**, *305*, 683−686.
(19) Krejsa, C. M.; Horvath, D.; Rogalski, S. L.; Penzotti, J. E.; Mao, B.; Barbosa, F.; Migeon, J. C. Predicting ADME Properties and Side Effects: The BioPrint Approach. (BioPrint is a registered trademark of Cerep). *Curr. Opin. Drug Discovery Dev.* **2003**, *6*, 470−480.
(20) Froloff, N.; Hamon, V.; Dupuis, P.; Otto-Bruc, A.; Mao, B.; Merrick, S.; Migeon, J. Construction of a Homogeneous and Informative in Vitro Profiling Database for Anticipating the Clinical Effects of Drugs. In *Chemogenomics Knowledge-based Approaches to Drug Discovery*; Jacoby, E., Ed.; Imperial College Press: London, 2006; Chapter 8, pp 175−206.
(21) de Groot, M. J.; Ekins, S. Pharmacophore Modeling of Cytochromes P450. *Adv. Drug Delivery Rev.* **2002**, *54*, 367−383.
(22) Liu, J.; Pan, D.; Tseng, Y.; Hopfinger, A. J. 4D-QSAR Analysis of a Series of Antifungal P450 Inhibitors and 3D-Pharmacophore Comparisons as a Function of Alignment. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 2170−2179.
(23) de Groot, M. J.; Kirton, S. B.; Sutcliffe, M. J. In Silico Methods for Predicting Ligand Binding Determinants of Cytochromes P450. *Curr. Top. Med. Chem.* **2004**, *4*, 1803−1824.
(24) Afzelius, L.; Zamora, I.; Masimirembwa, C. M.; Karlen, A.; Andersson, T. B.; Mecucci, S.; Baroni, M.; Cruciani, G. Conformer- and Alignment-Independent Model for Predicting Structurally Diverse Competitive CYP2C9 Inhibitors. *J. Med. Chem.* **2004**, *47*, 907−914.
(25) Gozalbes, R.; Barbosa, F.; Froloff, N.; Horvath, D. The BioPrint Approach for the Evaluation of ADMET Properties: Application to the Prediction of Cytochrome P450 2D6 Inhibition. In *Pharmacokinetic Profiling in Drug Research: Biological, Physicochemical, and Computational Strategies*; Testa, B., Krämer, S. D., Wunderli-Allenspach, H., Folkers, G., Eds.; VHCA: Zürich, Switzerland, 2006; pp 395−415.
(26) Wanchana, S.; Yamashita, F.; Hashida, M. QSAR Analysis of the Inhibition of Recombinant CYP 3A4 Activity by Structurally Diverse Compounds Using a Genetic Algorithm-Combined Partial Least Squares Method. *Pharm. Res.* **2003**, *20*, 1401−1408.
(27) Stresser, D. M.; Blanchard, A. P.; Turner, S. D.; Erve, J. L.; Dandeneau, A. A.; Miller, V. P.; Crespi, C. L. Substrate-dependent Modulation of CYP3A4 Catalytic Activity: Analysis of 27 Test Compounds with Four Fluorometric Substrates. *Drug Metab. Dispos.* **2000**, *28*, 1440−1448.
(28) Ono, S.; Hatanaka, T.; Hotta, H.; Satoh, T.; Gonzalez, F. J.; Tsutsui, M. Specificity of Substrate and Inhibitor Probes for Cytochrome P450s; Evaluation of in Vitro Metabolism using cDNA-Expressed Human P450s and Human Liver Microsomes. *Xenobiotica* **1996**, *26*, 681−693.
(29) Horvath, D. High Throughput Conformational Sampling and Fuzzy Similarity Metrics: A Novel Approach to Similarity Searching and Focused Combinatorial Library Design and Its Role in the Drug Discovery Laboratory. In *Combinatorial Library Design and Evaluation*; Ghose, A. K., Viswanadhan, V. N., Eds.; Marcel Dekker Inc.: New York, 2001; pp 429−471.
(30) Gozalbes, R.; Rolland, C.; Nicolai, E.; Paugam, M.-F.; Coussy, L.; Horvath, D.; Barbosa, F.; Mao, B.; Revah, F.; Froloff, N. QSAR Strategy and Experimental Validation for the Development of a GPCR Focused Library. *QSAR Comb. Sci.* **2005**, *24*, 508−516.
(31) Kier, L. H.; Hall, L. B. The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure−Property Modeling. In *Reviews in Computational Chemistry II*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1991; pp 367−422.
(32) Kier, L. H.; Hall, L. B. *The Molecular Description*; Wiley: New York, 1999.
(33) Kier, L. H.; Hall, L. B. Molecular Connectivity in Chemistry and Drug Research; Academic Press: New York, 1976.
(34) Tordeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, Germany, 2000.
(35) Yoshida, F.; Topliss, J. G. QSAR Model for Drug Human Oral Bioavailability. *J. Med. Chem.* **2000**, *43*, 2575−2585.
(36) Horvath, D.; Jeandenans, C. Neighborhood Behavior of in Silico Structural Spaces with Respect to in Vitro Activity Spaces−A Novel Understanding of the Molecular Similarity Principle in the Context of Multiple Receptor Binding Profiles. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 680−690.
(37) Horvath, D.; Jeandenans, C. Neighborhood Behavior of in Silico Structural Spaces with Respect to in Vitro Activity Spaces−A Benchmark for Neighborhood Behavior Assessment of Different in Silico Similarity Metrics. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 691−698.
(38) Chang, C.-C.; Lin, C.-J. *LIBSVM: A Library for Support Vector Machines*; 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm (accessed Apr 2005).
(39) Byvatov, E.; Schneider, G. SVM-Based Feature Selection for Characterization of Focused Compound Collections. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 993−999.
(40) Yap, C. W.; Chen, Y. Z. Prediction of Cytochrome P450 3A4, 2D6, and 2C9 Inhibitors and Substrates by Using Support Vector Machines. *J. Chem. Inf. Model.* **2005**, *45*, 982−992.
(41) Strobl, G. R.; von Kruedener, S.; Stockigt, J.; Guengerich, F. P.; Wolff, T. Development of a Pharmacophore for Inhibition of Human Liver Cytochrome P-450 2D6: Molecular Modeling and Inhibition Studies. *J. Med. Chem.* **1993**, *36*, 1136−1145.
(42) Susnow, R. G.; Dixon, S. L. Use of Robust Classification Techniques for the Prediction of Human Cytochrome P450 2D6 Inhibition. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1308−1315.
(43) Ekins, S.; Berbaum, J.; Harrison, R. K. Generation and Validation of Rapid Computational Filters for CYP2D6 and CYP3A4. *Drug Metab. Dispos.* **2003**, *31*, 1077−1080.
(44) Kriegl, J. M.; Eriksson, L.; Arnhold, T.; Beck, B.; Johansson, E.; Fox, T. Multivariate Modeling of Cytochrome P450 3A4 Inhibition. *Eur. J. Pharm. Sci.* **2005**, *24*, 451−463.
(45) Arimoto, R.; Prasad, M. A.; Gifford, E. Development of CYP3A4 Inhibition Model: Comparison of Machine-Learning Techniques and Molecular Descriptors. *J. Biomol. Screen.* **2005**, *10*, 197−205.
(46) Kvålseth, T. O. Cautionary Note About R^2. *Am. Stat.* **1985**, *39*, 279−285.
(47) Ekins, S.; Bravi, G.; Wikel, J. H.; Wrighton, S. A. Three-Dimensional-Quantitative Structure Activity Relationship Analysis of Cytochrome P-450 3A4 Substrates. *Drug Metab. Dispos.* **1999**, *291*, 424−433.
(48) Stresser, D. M.; Turner, S. D.; Blanchard, A. P.; Miller, V. P.; Crespi, C. L. Cytochrome P450 Fluorometric Substrates: Identification of Isoform-Selective Probes for Rat CYP2D2 and Human CYP3A4. *Drug Metab. Dispos.* **2002**, *30*, 845−852.
(49) Fowler, S. M.; Taylor, J. M.; Friedberg, T.; Wolf, C. R.; Riley, R. J. CYP3A4 Active Site Volume Modification by Mutagenesis of Leucine 211. *Drug Metab. Dispos.* **2002**, *30*, 452−456.
(50) Zar, J. H. *Biostatistical Analysis*, 3rd ed.; Prentice-Hall: Upper Saddle River, NJ, 1996; pp 415−418.

CI0600915