A Refined 3-Dimensional QSAR of Cytochrome P450 2C9: Computational Predictions of Drug Interactions

Sreedhara Rao,[†] Ron Aoyama, Michael Schrag, William F. Trager,[‡] Allan Rettie,[‡] and Jeffrey P. Jones*,[†]

Department of Chemistry, Washington State University, Pullman, Washington, and Department of Medicinal Chemistry, School of Pharmacy, University of Washington, Seattle, Washington

Received February 8, 2000

A ligand-based model is reported that predicts the K_i values for cytochrome P450 2C9 (CYP2C9) inhibitors. This CoMFA model was used to predict the affinity of 14 structurally diverse compounds not in the training set and appears to be robust. The mean error of the predictions is 6 μM. The experimentally measured K_i values of the 14 compounds range from 0.1 to 48 μM. Leave-one-out cross-validated partial least-squares gives a q^2 value of between 0.6 and 0.8 for the various models which indicates internal consistency. Random assignment of biological data to structure leads to negative q^2 values. These models are useful in that they establish a pharmacophore for binding to CYP2C9 that can be tested with site-directed mutagenesis. These models can also be used to screen for potential drug interactions and to design compounds that will not bind to this enzyme with high affinity.

Introduction

The cytochrome P450 superfamily of enzymes (CYP) is the most important family of enzymes involved in drug metabolism. However, only a small subset of the CYP enzymes are responsible for the majority of drug-metabolizing events.[1] Since only a handful of P450 enzymes are responsible for the metabolism of drugs, it is inevitable that during polytherapy multiple drugs will compete for the active site of a given P450. Furthermore, since some of these P450 enzymes are polymorphic, differential response within the population of drug users can lead to different therapeutic outcomes. To this end it would be extremely valuable to (1) identify the major P450 enzyme responsible for the metabolism of a given drug, (2) be able to rapidly determine a drugs affinity for this enzyme, and (3) be able to "design away" from potential drug interactions. While in vitro screens for inhibition and metabolism can provide some of this information, rapid computational methods can provide this information earlier in the drug development cycle and at significantly lower expense. This information can also be used to determine potential interactions of virtual compounds, enhancing drug design efforts. Furthermore, since a large number of P450 enzymes are responsible for homeostasis, members of the P450 superfamily are potential targets for the design of new drugs. An example of the later type of P450 is P450 aromatase, which is responsible for the conversion of androgens to estrogens. Knowledge about how these enzymes bind substrates and models that predict binding differences within the families are important in designing drugs for this type of target. Finally, these types of models can also play an important role in the design of biochemical experiments aimed at understanding the important characteristics of the enzyme active site.[2]

Given the intense interest in the P450 enzyme family surprisingly few quantitative ligand-based models have been made. Studies have been published on affinity models for CYP2B6, CYP2C9, CYP2E1, and CYP3A4. All of these models use ligand-based methods to develop a QSAR of the active site, which will predict the affinity of a given substrate for that enzyme. Other modeling efforts, in particular those for CYP2D6, have developed interesting pharmacophore models of the active site which predict the regiochemistry of metabolism.[3]

A quantiative ligand-based model for CYP2E1 was constructed for the clearance of 12 volatile organic compounds and was interpreted as a three-phase binding interaction including a recognition phase, an accessibility phase, and a reactivity phase.[4] No specific interactions are predicted in the active site which is consistent with the small nondescript nature of the compounds used in the training set. Two different models were constructed for CYP2B6 by Ekins et al.[5] based on two different modeling methods. One model predicted the important active site interactions to include a hydrogen bond and three hydrophobic regions, although the authors encouraged caution in the use of this model due to a potential lack of statistical significance. The second modeling method was not capable of predicting active site characteristics since it was an alignment independent method but was able to predict affinities for four out of five compounds not used in the training set. These two studies are unique in that they predict the kinetic properties K_M and clearance. A similar study to that reported for CYP2B6 was conducted for CYP3A4, but it used K_i or IC_{50} values as a source of biological data.[6] In this study three separate training sets were used to generate models. A total of 1−3 hydrogen-bonding sites and 1−3 hydrophobic regions were predicted depending on the training set. More recently, modeling has been conducted on CYP3A4 with 38 substrates.[7] These results indicate that the

* To whom correspondence should be addressed at: Rm. 477, Fulmer Synthesis, Dept. of Chemistry, P.O. Box 644630, Pullman, WA 99164-4630. Tel: (509) 335-5083. Fax: (509) 335-8867. E-mail: jpj@wsu.edu.
[†] Washington State University.
[‡] University of Washington.

10.1021/jm000048n CCC: $19.00 © 2000 American Chemical Society
Published on Web 06/30/2000

enzyme has two hydrogen bond acceptors, one hydrogen bond donor, and one hydrophobic region. These results by Ekins are particularly impressive since they use either an alignment-independent method (PLS MS-WHIM) or an automated alignment method (Catalyst) and thus are not dependent on user dependent alignments. All of these studies show the power of in silico simulation to help interpret complex data sets.

In a previous paper we reported a model for CYP2C9 based on 27 inhibitors of (*S*)-warfarin metabolism in vitro.[8] This model is referred to as 2C9(27) model in this manuscript. The 2C9(27) model predicted an aromatic binding site and two electrostatic interaction sites in the active site. Combining this CoMFA-generated active site model with an earlier homology model produced for CYP2C9[9] resulted in the (unpublished) prediction that two residues, F110 and F114, may play a role in binding the aromatic residues.[10] Subsequent studies implicated F114, but not F110, as an important residue in the aromatic binding site.[2] These data indicated to us that one of the highest-affinity inhibitors, sulfaphenazole, had been aligned incorrectly. This belief stems from the fact that the CoMFA model predicted that sulfaphenazole would *not* be affected by the aromatic stacking change and the site-directed mutagenesis results indicated that interaction with the putative aromatic stacking region *did* change the affinity. We have modified our model to be consistent with this result. Furthermore, the ability of 14 new compounds to inhibit (*S*)-warfarin metabolism was measured and the data used to validate our modified model. While our initial model contained mostly coumarin containing compounds, this new data set is mostly comprised of sulfonamides with only one of these compounds being a coumarin. Furthermore, these compounds have activities which range from 0.1 to 48 μM, close to the range of the initial models biological data. Finally, the 14 new compounds are incorporated into our model to create a more robust predictive model for CYP2C9. Herein we report the modification, validation, and expansion of our initial model for CYP2C9.

Methods

Biological Methods. The methods for determining K_i values have been presented.[11] K_i values for sulfaphenazole and (*S*)-warfarin inhibition of CYP2C9 metabolism were calculated using the Cheng−Prusoff equation, which equates K_i to $IC_{50}/2$ for competitive inhibitors, if $[S] = K_m$. IC_{50} values were obtained directly from plots relating percentage residual activity versus log inhibitor concentration over the range 0−100 mM. K_m and V_{max} were estimated using the k.cat Program (Biometallics Inc.), which fits data to a nonlinear kinetics program. Spectral P450 measurements were made by the method of Estabrook et al.[12] Protein concentrations were measured using the Bradford assay.[13]

Alignment Rules. Our alignment rules remain very similar to those established for our first model which was based on the rigid warfarin analogues **3** and **4** in Table 1.[8] The binding rules align an aromatic ring, the benzylic carbon, and the hydrogen attached to the aligned benzylic carbon when possible. These rules work particularly well for coumarin-containing compounds. The sulfaphenazole analogues were aligned based on the modified alignments of sulfaphenazole (**29**). These alignments place the aniline aromatic ring in the π stacking/aromatic region, one sulfonamide oxygen pointing toward the 2-keto oxygen of the coumarin analogues, and the other sulfonamide toward the putative anion/negative dipole interaction region. Compounds **22**−**28** all have the aromatic regions aligned in the π stacking region and the carbonyl groups aligned in close proximity to those at the coumarin C-2 carbonyl. All of the compounds given in Tables 2 and 3 are aligned by a least-squares overlay of the aromatic region to structurally similar compounds in Table 1 and bonds that are free rotors adjusted to maximize overlap. The most difficult alignments, and those for which we have the least confidence, are for compounds **24**−**28** since they have multiple free rotors, are relatively small in volume, and have limited representation in the training set.

Homology Model. The construction of the homology model of CYP2C9 used Modeler 4.[14] The homology model for CYP2C9 was generated in a fashion similar to those published earlier.[9] The alignments are those of Haseman et al.[15] with P450s BM3 and Cam. P450EryF was aligned with P450BM3 and P450Cam as described by Poulos.[16] This program uses a combination of molecular dynamics and restraints based on the known structures of the homologous enzymes P450BM3, P450Cam, and P450EryF. The coordinates were obtained from the Protein Data Bank and used the pdb file 3cpp for P450Cam,[17] pdb file 2bmh for P450BM3,[18] and pdb file 1oxa for P450EryF.[16] This program will give a 3-dimensional representation of an alignment, which has a structure similar to the known structures of BM3, Cam and EryF. The degree of similarity between target and model is dependent on the degree of implied homology in the alignments. Thus, structures in areas with inserts not aligned with the known crystal structure will be determined solely by the molecular dynamics force field; areas without inserts will give a weighted average of the restraints and the force field. The resulting models were determined to be structurally reasonable based on Ramachandran plots.

To align the CoMFA field in the homology model, cyclocoumarol **3** was placed with the 7 position of the coumarin ring over the heme since this is the known region of metabolism.[11] Molecular mechanics using the Tripos force field in Sybyl was used to relax the substrate in the active-site. This was not meant to be a rigorous computational simulation but only to serve as a method to get a single low-energy binding mode for this compound. It is of course possible and likely that alternate binding modes exist for this substrate. The CoMFA field was than visualized in the active site. In this fashion, amino acids R105, F114, F110, and D293 were identified as potential residues involved in interaction with the substrate, and the structures shown in Figures 1 and 3 were generated.

In Silico Methods. All charges were determined with the MNDO Hamiltonian[19] and structures minimized with the AM1 Hamiltonian.[20] MNDO charges were used since MNDO charges more closely reflect ab initio charges.[21] Rotations of bond were performed in SYBYL,[22] and strain was relieved with the TRIPOS force field followed by AM1 calculations. No effort was made to account for strain energy as a possible contributor to binding energy. All CoMFA methodology used the standard defaults. CoMFA is a module in the SYBYL suite of programs. QSAR models were determined by partial least-squares (PLS) analysis of the inverse natural logarithm of biological activity versus electrostatics and sterics. The electrostatic and steric fields for each molecule were determined by probing the molecules in the database with a sp^3 carbon atom with a single positive charge with the default 2 Å step size. Guided region cross-validated q^2 values were determined using the method of Cho and Tropsha[23] for each model and are as follows: 2C9-(27b) $q^2 = 0.805$, 2C9(31) $q^2 = 0.785$, 2C9(41) $q^2 = 0.667$. All other q^2 values reported use the standard method in CoMFA and are 0.724 for 2C9(27b), 0.757 for 2C9(31), and 0.629 for 2C9(41).

The nomenclature that is used to designate the models describes the enzyme, 2C9, followed by the number of compounds in the model in parentheses and a revision letter if it is a refined model. For example, 2C9(27) refers to the initial model presented first in Jones et al.,[8] which contains 27 compounds, and model 2C9(27b) refers to a revised model presented for the first time in this manuscript.

The PLS results provide a number of values that can be

Table 1. CoMFA Comparison with Experimental Activity for Various Inhibitors to Human CYP450-2C9

COMPOUND	Ki	*Ki*	Ki	Ki	Ki
1 9R	ND	ND	ND	ND	7
2 9S	ND	ND	ND	ND	4
3 9S,11S	17	13	16	15	20
4 9S,11R	9	9	10	10	12
5 9R,11R	15	12	12	10	10
6 9R,11S	12	11	12	11	7
7	2	1.3	1.2	1.2	0.6
8 9R	0.6	0.9	0.7	0.8	0.5
9 9S	0.5	0.7	0.7	0.7	0.6
10 9S,11R	1	0.3	0.4	0.3	0.1
11 9S,11S	1	1.7	1.6	1.5	1
12 9R,11S	16	22	21	20	14
13 9R,11R	30	21	18	22	37
14 9R	11	17	17	17	10
15 9S	3	6.6	6.8	7	11
16 9R	23	11	12	12	29
17 9S	9	14	14	21	21
18 9R	14	14	14	15	13
19 9S	24	22	23	18	17
20 9R	17	27	28	28	14
21 9S	2	2.2	2.0	1.7	8

COMPOUND	Ki	*Ki*	Ki	Ki	Ki
22	9	8.3	10	16	10
23	6	19	33	41	70
24	14	7.3	6.5	3.4	11
25	33	15	13	16	45
26	16	15	15	15	15
27	10	28	31	34	50
28	6	2.1	1.8	1.9	2
29	1	1.3	1.2	0.4	0.2

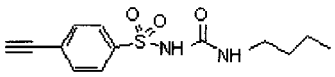
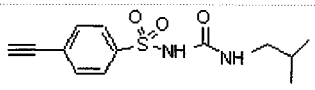
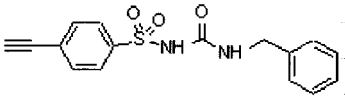
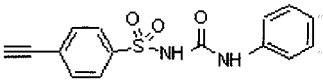
Ki Crossvalidated predictions from model 2C9(27) See Ref. 1
Ki Crossvalidated predictions from model 2C9(27b)
Ki Crossvalidated predictions from Model 2C9(31)
Ki Crossvalidated predictions from model 2C9(41)
Ki Measured values
ND Not determined

used to judge the predictive value of the models. Of these the q^2 value is probably the most useful.[24] This value is 1 − (the ratio of the sum of the squares of the difference of the predicted value and the mean value to the sum of the squares of the difference of the actual and the mean value). Thus, it is assumed that the mean value of the measured values provides the baseline model, and this model would result in a q^2 value of 0. A model that predicts values farther from the measured values than the mean will give a negative q^2 value. It has been determined that a q^2 value of around 0.3 gives a significant model at the 95% confidence limit.[24]

Results and Discussion

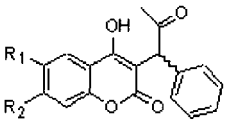
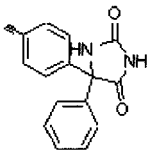
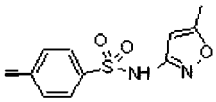
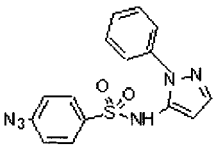
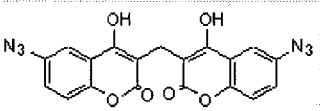
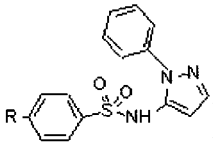
Modifications to the Existing Model. We used our original CoMFA model (designated 2C9(27) in Table 1 and throughout this manuscript) to design experiments to test the models robustness. One of our assumptions

Table 2. Addition of Four New Sulfonamide Analogues

COMPOUND		*Ki*	Ki	**Ki**	**Ki**
	**30**	36	19	23	18
	**31**	34	43	49	48
	**32**	39	19	20	25
	**33**	28	19	21	7

Ki Predictions from model 2C9(27b)
Ki Crossvalidated predictions from Model 2C9(31)
Ki Crossvalidated predictions from model 2C9(41)
Ki Measured values

Table 3. Substituted Anlogues of Inhibitors of CYP450-2C9: CoMFA Prediction and Comparison with Experimental Values

COMPOUND		*Ki*	Ki	**Ki**	**Ki**
 R1 R2					
—≡ H	**34**	2.2	1.7	2.5	1.5
H —≡	**35**	2.0	1.7	1.8	1.1
	**36**	13	13	20	18
	**37**	3.4	2.8	2.7	3
	**38**	2.2	2.0	2.4	0.4
	**39**	0.8	0.8	0.7	3
 R=					
CH3-CO-	**40**	3.6	2.7	1.5	2.4
H-	**41**	1.5	1.2	0.5	2
OH-	**42**	2.4	1.9	0.4	6
CH3-	**43**	2.8	2.3	3.9	0.1

Ki Predictions from model 2C9(27b)
Ki Predictions from Model 2C9(31)
Ki Crossvalidated predictions from model 2C9(41)
Ki Measured values
All Values are in μM.

in the construction of the original CoMFA model for P4502C9 was that a π stacking interaction was important in the orientation and binding affinity of substrates. A homology model of CYP2C9 in combination with the CoMFA model suggested both F110 and F114 as likely residues involved in this interaction.[10] Mutation of phenylalanine 110 to leucine has very little effect on binding. However, mutation of phenylalanine 114 to leucine decreases the affinity of warfarin and diclofenac. These results are consistent with our original CoMFA model which assumed both these compounds had aromatic stacking interactions. In addition, the affinity of the inhibitor sulfaphenazole is also significantly decreased (over 100-fold) by this mutation. In contrast, in model 2C9(27), we placed sulfaphenazole parallel to the heme and below the aromatic stacking region. This is illustrated Figure 1, in which sulfaphenazole (old) is docked in the homology model using the orientation used in the CoMFA model 2C9(27). It should be noted that the conclusion about this interaction reflects the mutagensis data and the homology model was used to illustrate this point, not to arrive at this conclusion. The alignment in our initial 2C9(27) model predicted that the anilino nitrogen of sulfaphenazole would interact with a positively charged electrostatic site some distance removed from the aromatic stacking region. To make the model consistent with the mutagenesis data, sulfaphenazole was moved to an alignment with the anilino group in the π stacking region (see Figure 1 new) which would remove the possibility of interaction at the electrostatic site. (Data presented in this paper support the hypothesis that an electrostatic interaction with this nitrogen is not important, vide infra.) The resulting model has a q^2 value of 0.72, as compared to our initial model that had a q^2 value of 0.70. The new model results in a nearly identical pharmacophore for the inhibition of CYP2C9. Finally, we made a number of minor adjustments to the alignments, in particular for compounds **22** (phenytoin) and **23,** and did a more thorough job of aligning the aromatic rings in the stacking region to arrive at model 2C9(27b), which has a q^2 value of 0.81. The resulting predictions are given in Table 1. Interestingly, while the cross-validated q^2 value increased, the predictive capacity for related compounds is no better and at times worse. This decrease in predictive power is most apparent among the 9*R*-coumarins, compounds **5**, **6**, **8**, **12**−**14**, **16**, **18**, and **20**. In this series only compounds **5** and **6** are predicted more accurately by the new model than by the older model. However, this is counterbalanced by an increase in the predictive power for most other compounds.

A second alignment was done with the anilino group close to the hypothetical position of the heme, with the potential for coordination with the iron, and the phenyl group in the π stacking region. For the 27 compounds in Table 1, this orientation gave a q^2 value of 0.59 for a four-component model. It should be noted that this alignment is consistent with spectral data reported by others.[25] These authors hypothesize that coordination of the anilino nitrogen with the iron is a major binding determinant for sulfaphenazole. However, on the basis of our model and the mutagensis data, we hypothesize that this orientation represents a minor component of the total bound sulfaphenazole. This hypothesis is

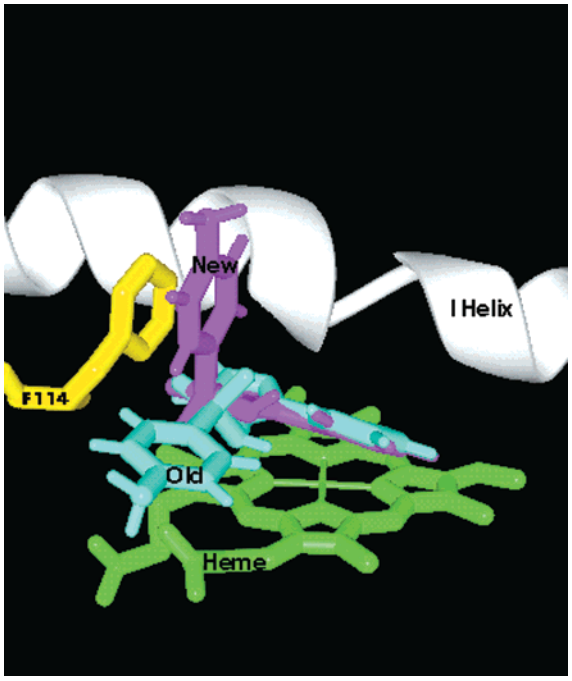


Figure 1. Sulfaphenazole bound in the original (old) orientation of model 2C9(27). No π stacking interactions are present is shown in light blue. Sulfaphenazole bound in the new orientation with π stacking interactions is shown in magenta.

supported by the binding data for sulfaphenazole analogues which bind tightly and do not have a basic nitrogen in this position. These include compounds **40**−**43** (Table 3), all of which bind with high affinity. In particular compound **43**, in which the NH_2 group is replaced by a methyl group. This compound actually binds tighter than the parent compound and cannot possibly have a strong electrostatic interaction at this site. Thus, while we do not believe that the small differences in q^2 values allow us to assign a given binding mode, the partial least-squares analysis shows that this is not required to explain the biological data. However, it is possible that compounds **40**−**43** do not share a common binding orientation with compound **29**.

Realignment of the compounds to give the 2C9(27b) model resulted in a change in the predicted pharmacophore. In our initial model two electrostatic interaction sites that interact with partially negatively charged portions of the substrate are predicted to be important (Figure 2a). Two possible sites of interaction with positive charge in the substrate were also identified in this model. The refined model (2C9(27b)) predicts a single strong negative charge interaction site and 1−2 positive charge interaction sites (Figure 2b). The possible interpretations of the pharmacophore prediction will be presented when we discuss model 2C9(41) which includes all of the compounds for which binding data is available.

Validation of the Model. While cross-validation as represented by the q^2 value, provides information about the internal consistency and thus the predictive capacity of a model, the ultimate test is to challenge the model to predict values of a set of compounds not included in the training set. Thus, the modified 2C9(27b) model was used to predict the K_i values for 14 compounds not included in the initial training set. Each prediction uses the equation from the full partial least-squares results for the 27 compounds and predicts the affinity of the unknown test compound for a given alignment and charge distribution. The results for each of the 14 compounds are given in Tables 2 and 3. Overall, the predictions for the 14 new compounds are reasonable. This is surprising in that the training set only has two sulfonamides, while in the validation set 10 of the 14 compounds are sulfonamides.

To test if increased representation in the training set could increase the predictive capacity, 4 of the sulfonamides were arbitrarily chosen to enhance the representation of this class of compounds in the training set. After including these compounds, the model with 31 training compounds, model 2C9(31), was used to predict the affinity of the remaining 10 compounds. The predictions of this enhanced model are for the most part better. Now most of the sulfonamide binding affinities are well-predicted (Table 2), with the largest absolute error occurring for prediction of the sulfaphenazoles analogues **38** and **43**.

Finally, all 41 molecules were included in the training set (see Tables 2 and 3). This model is designated 2C9-(41). The predicted values are better. The q^2 value for this model is 0.647. The reason for the decrease in the internal predictive capacity appears to be the inability of the model to reconcile the differences in binding affinities between compounds **29, 42,** and **43**. While compounds **42** and **29** both contain electronegative atoms and similar electrostatics, they vary in affinity by 30-fold. Compounds **29, 42,** and **43** all have similar steric bulk but different electrostatics. For example, compound **43** has much less partial negative charge. From these three compounds it would appear that another descriptor may be required to reconcile the differences in activity. One possible explanation is that compound **42** exists to some extent as an anion. Since the pK_a of the proton on the phenol oxygen of **42** is around 7.9, a significant fraction of the compound exists as an anion. Since 2C9 is predicted by others to bind anions it may appear contradictory that an anion would bind to 2C9 with a lower affinity than the neutral compound.[1,25−27] However our results are not in conflict with this hypothesis since either an anion localized on the phenol oxygen could be in a place not stabilized by a countercharge in the protein or the molecule may be forced to bind in a different, sterically compromised orientation, to increase electrostatic interactions. Removing compound **42** from the training set changes the q^2 value from 0.65 to 0.76.

Evaluation of the Model. Given the nature of partial least-squares analysis in general, it is natural to wonder about the possibility of a chance correlation and possibility of data sets strongly dependent on a subset of the molecules in the data set. To test these possibilities we used two approaches that, while not entirely thorough, increase our confidence that the models are predictive and significant.

While studies have been published that indicate one can have confidence in CoMFA models with a q^2 value of greater than 0.3, the possibility remains that a chance correlation can be obtained.[24] This is particularly true of alignments that do not follow strict alignment rules. While we believe our alignment rules are strict we decided to test for this possibility. To test for the

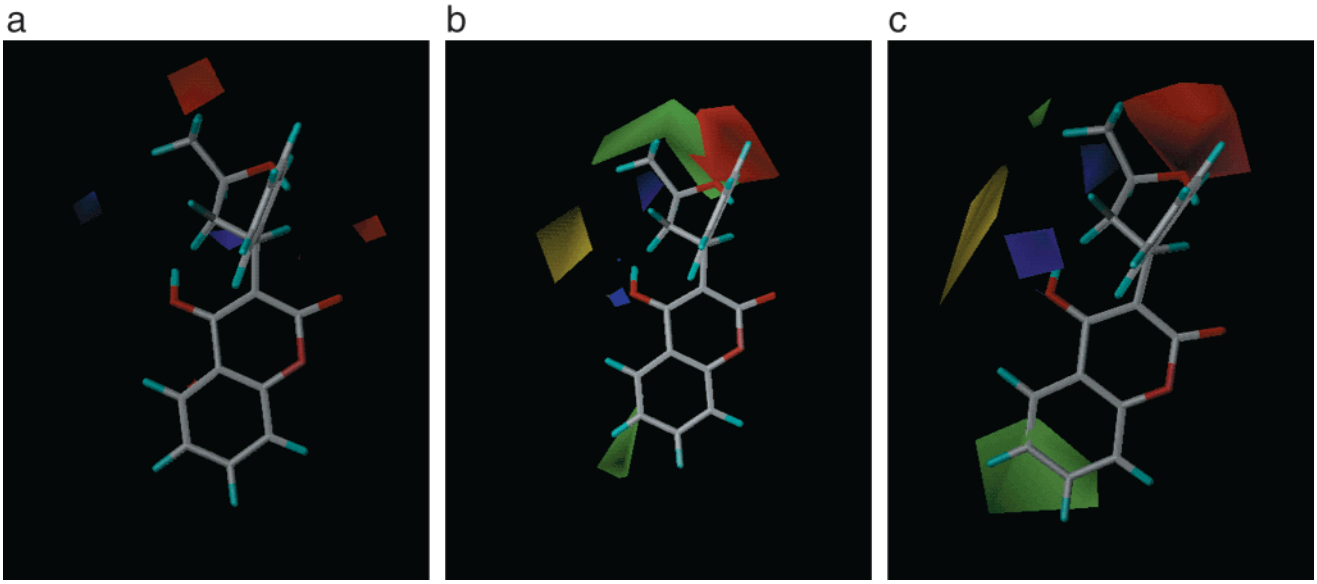


Figure 2. (a) Original 2C9 CoMFA model electrostatic interactions. Red indicates where more negative charge will enhance binding, blue where positive charge will enhance binding. The steric fields are not shown, since they obscure the electrostatic fields. (b) Model 2C9(27b) CoMFA fields. Red areas indicate where more negative charge will enhance binding, blue areas where positive charge will enhance binding, yellow areas where steric interactions hinder binding, green areas where steric interactions enhance binding. (c) Model 2C9(41) CoMFA fields. Red areas indicate where more negative charge will enhance binding, blue areas where positive charge will enhance binding, yellow areas where steric interactions hinder binding, green areas where steric interactions enhance binding.

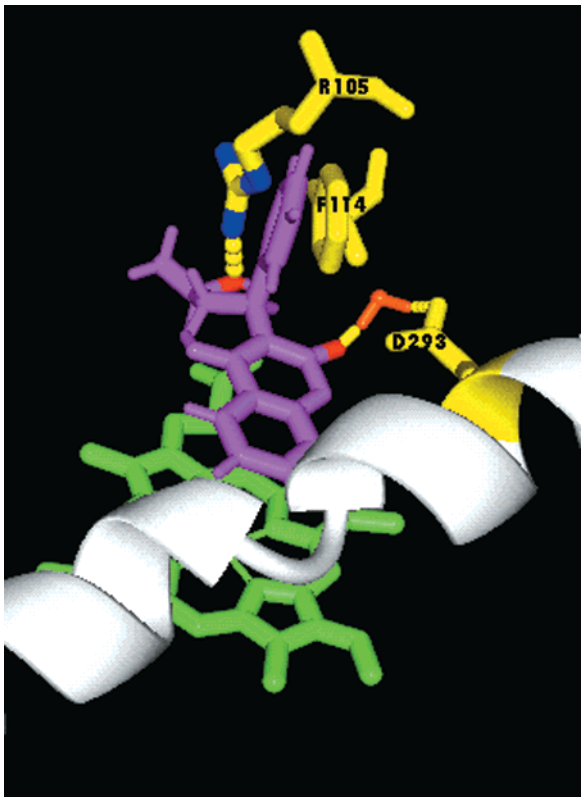


Figure 3. Docking of compound **10** in the active site of a homology model of CYP2C9. The putative interacting amino acids are in yellow, compound **10** in magenta, the bridging water in orange.

possibility of a chance correlation, we randomly assigned biological activities to structures and determined the q^2 values. In two separate randomization trials we obtained q^2 values of −0.15 and −0.24 for the 41-compound model (2C9(41)) with randomly assigned biological activities.

It is also possible that a high q^2 value can reflect only the information in a small subset of molecules. To test if the correlations we obtained are dependent on a small subset of the data, cross-validation was be carried out with only 80% of the data as opposed to the more standard leave-one-out method. In this case models 2C9-(27b), 2C9(31), and 2C9(41) have q^2 values of 0.593, 0.607, and 0.712, respectively. While these numbers will differ for different leave-20%-out groups, our results showed similar q^2 values to those obtained for the leave-one-out method.

Thus, the models are internally consistent as judged by the leave-one-out q^2 values. The models can predict data not in the training sets as judged by our 2C9(27b) models ability to predict the K_i values of 14 compounds not in the training set. Finally, the models do not appear to be the result of a chance correlation as judged by our leave-20%-out experiments and the negative q^2 values obtained for the randomly assigned biological activities.

Predicting Drug Interactions. The models can be interpreted in terms of their potential use to predict drug interactions. The majority of drug interactions involving CYP2C9 appear to be competitive. Thus, for a drug interaction to occur, two drugs must have a K_i similar to their therapeutic concentration. For example, a drug that has a therapeutic concentration of around 1 nM and a K_i value for CYP2C9 of 1 μM is unlikely to show any drug interaction even with the drugs that display relatively high affinity for 2C9. Since a large number of drugs have low micromolar to submicromolar therapeutic concentrations, an arbitrary definition of a potential problem compound with respect to drug interactions can be assumed to be around 10 μM. Thus, those compounds with affinities <10 μM for CYP2C9 have a strong potential for drug interactions, while the compounds with >10 μM K_i values have a more limited chance for a drug interactions. Out of 41 compounds

(compounds **3**−**43**), 19 have measured K_i values less than 10 μM. In all cases but 2 (compounds **6** and **33**) the results from any of the three models presented herein place the compound in the correct class. Conversely, 19 compounds have K_i values greater than 10 μM, and all but 3 (compounds **4, 15**, and **24**) are predicted to be in the correct class with all of the models presented in this paper. Thus, given this admittedly arbitrary definition and for compounds related to those in the training set, the CoMFA model can be used to predict potential drug interactions. Of course knowledge of the potential therapeutic concentration can make these classes better defined, but this exercise points out both the utility and the method that could be used to predict potential problems.

Active Site Characteristics. It has been known for some time that CYP2C9 binds compounds with large dipoles or negative charges.[1,25−27] Thus, oxygen-rich compounds such as carboxylic acids, sulfonamides, and alcohols are substrates for 2C9. More recently it has been determined that an aromatic ring or lipophilic interaction is important in binding substrates based on site-directed mutagenesis.[2] In our initial model for CYP2C9 we hypothesized two positive electrostatic sites in the enzyme and an aromatic stacking site (Figure 2a). Our present model is shown in Figure 2c. This model can be interpreted in terms of the CoMFA fields to have a single site where negative charge is important on the substrate (red in Figure 2c), one bad steric interaction (yellow in Figure 2c), and two positions where increasing steric bulk increases affinity (green in Figure 2c). Two small areas where positive charge increases affinity are also present (blue in Figure 2c). The position of these good positive sites indicates to us that they reflect the opposite end of the dipole that positions the negative atom in the active site to interact with a positive or partial positive charge on the enzyme.

In an attempt to map the steric and electrostatic interactions onto the active site of 2C9 we placed the CoMFA model into the active site of a homology model constructed for 2C9 based on the known bacterial crystal structures. The construction of this model is presented in Methods, and the model is very similar to our previously published model of CYP2C9.[28] A number of compounds were aligned by placing the aromatic ring in close proximity to F114 and the 7 position of the coumarin ring close to the heme iron. The results indicate that the area of increased steric bulk (green area in Figure 2c) fills the area over the heme near the I helix and the β-5 sheet region (structural features relative to P450Cam structure). The anionic interaction site is predicted to be R105 from this admittedly low-resolution homology model. Finally, the carbonyl oxygen at the C-2 position of coumarin may interact with D293 in the I helix, possibly through a bridging water molecule. The warfarin alcohol **10** docked in the homology model shown in Figure 3 illustrates these potential interactions.

In conclusion, ligand-based models have been developed for one of the major P450 enzymes involved in drug metabolism, CYP2C9. These models can explain the observed biological data for compounds not included in the model. This information may prove to be important for predicting drug interactions and to our understanding of the features in the active site important in binding of substrates or inhibitors.

Acknowledgment. We thank the National Institutes of Health for financial support through Grants GM32165 (W.F.T., A.R.) and ES09122 (J.P.J.).

References

(1) Smith, D. A.; Jones, B. C. Speculations on The Substrate Structure−Activity Relationship (SSAR) of Cytochrome P450 Enzymes. *Biochem. Pharmacol.* **1992**, *44*, 2089−2098.
(2) Haining, R. L.; Jones, J. P.; Henne, K. R.; Fisher, M. B.; Koop, D. R. et al. Enzymatic determinants of the substrate specificity of CYP2C9: role of B′−C loop residues in providing the pi-stacking anchor site for warfarin binding. *Biochemistry* **1999**, *38*, 3285−3292.
(3) de Groot, M. J.; Ackland, M. J.; Horne, V. A.; Alex, A. A.; Jones, B. C. A novel approach to predicting P450 mediated drug metabolism. CYP2D6 catalyzed N-dealkylation reactions and qualitative metabolite predictions using a combined protein and pharmacophore model for CYP2D6. *J. Med. Chem.* **1999**, *42*, 4062−4070.
(4) Waller, C. L.; Evans, M. V.; Mckinney, J. D. Modeling the Cytochrome P450-Mediated Metabolism Of Chlorinated Volatile Organic Compounds. *Drug Metab. Dispos.* **1996**, *24*, 203−210.
(5) Ekins, S.; Bravi, G.; Ring, B. J.; Gillespie, T. A.; Gillespie, J. S. et al. Three-dimensional quantitative structure activity relationship analyses of substrates for CYP2B6. *J. Pharmacol. Exp. Ther.* **1999**, *288*, 21−29.
(6) Ekins, S.; Bravi, G.; Wikel, J. H.; Wrighton, S. A. Three-dimensional-quantitative structure activity relationship analysis of cytochrome P-450 3A4 substrates. *J. Pharmacol. Exp. Ther.* **1999**, *291*, 424−433.
(7) Ekins, S.; Bravi, G.; Binkley, S.; Gillespie, J. S.; Ring, B. J. et al. Three- and four-dimensional quantitative structure activity relationship analyses of cytochrome P-450 3A4 inhibitors. *J. Pharmacol. Exp. Ther.* **1999**, *290*, 429−438.
(8) Jones, J. P.; He, M. X.; Trager, W. F.; Rettie, A. E. Three-Dimensional Quantitative Structure−Activity Relationship For Inhibitors Of Cytochrome P4502c9. *Drug Metab. Dispos.* **1996**, *24*, 1−6.
(9) Korzekwa, K. R.; Jones, J. P. Predicting the cytochrome P450 mediated metabolism of xenobiotics. [Review]. *Pharmacogenetics* **1993**, *3*, 1−18.
(10) Jones, J. P. Unpublished results.
(11) He, M.; Korzekwa, K. R.; Jones, J. P.; Rettie, A. E.; Trager, W. F. Structural forms of phenprocoumon and warfarin that are metabolized at the active site of CYP2C9. *Arch. Biochem. Biophys.* **1999**, *372*, 16−28.
(12) Estabrook, R. W.; Peterson, J.; Baron, J.; Hildebrandt, A. *Methods Pharmacol.* **1972**, *2*, 303−350.
(13) Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **1976**, *72*, 248−254.
(14) Sali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779−815.
(15) Hasemann, C. A.; Kurumbail, R. G.; Boddupalli, S. S.; Peterson, J. A.; Deisenhofer, J. Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure* **1995**, *3*, 41−62.
(16) Cuppvickery, J. R.; Poulos, T. L. Structure Of Cytochrome P450eryf Involved In Erythromycin Biosynthesis. *Nat. Struct. Biol.* **1995**, *2*, 144−153.
(17) Raag, R.; Poulos, T. L. Crystal structure of the carbon monoxide-substrate-cytochrome P-450CAM ternary complex. *Biochemistry* **1989**, *28*, 7586−7592.
(18) Li, H.; Poulos, T. L. Modeling Protein-Substrate Interactions in the Heme Domain of Cytochrome P450Bm-3. *Acta Crystallogr., Sect. D* **1995**, *51*, 21.
(19) Dewar, M. J.; Theil, W. Ground States of Molecules, 38. The MNDO Method. *J. Am. Chem. Soc.* **1977**, *99*, 4899−4907.
(20) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(21) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11*, 431−439.
(22) *SYBYL*, 6.1 ed.; Tripos Associates: St. Louis.

(23) Cho, S. J.; Tropsha, A. Cross-validated R2-guided region selection for comparative molecular field analysis: a simple method to achieve consistent results. *J. Med. Chem.* **1995**, *38*, 1060−1066.
(24) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.
(25) Mancy, A.; Dijols, S.; Poli, S.; Guengerich, P.; Mansuy, D. Interaction of sulfaphenazole derivatives with human liver cytochromes P450 2C: molecular origin of the specific inhibitory effects of sulfaphenazole on CYP 2C9 and consequences for the substrate binding site topology of CYP 2C9. *Biochemistry* **1996**, *35*, 16205−16212.
(26) Mancy, A.; Broto, P.; Dijols, S.; Dansette, P. M.; Mansuy, D. The Substrate Binding Site Of Human Liver Cytochrome P450 2c9 − an Approach Using Designed Tienilic Acid Derivatives and Molecular Modeling. *Biochemistry* **1995**, *34*, 10365−10375.
(27) Jones, B. C.; Hawksworth, G.; Horne, V.; Newlands, A.; Tute, M. et al. Putative action site model for CYP2C9. *Acta Crystallogr., Sect. D* **1993**, *34*, 143p−144p.
(28) Korzekwa, K. R.; Jones, J. P. Predicting the cytochrome P450 mediated metabolism of xenobiotics. *Pharmacogenetics* **1993**, *3*, 1−18.

JM000048N