

# Synergistic Use of Compound Properties and Docking Scores in Neural Network Modeling of CYP2D6 Binding: Predicting Affinity and Conformational Sampling

Peter S. Bazeley,<sup>†</sup> Sridevi Prithivi,<sup>†</sup> Craig A. Struble,<sup>†</sup> Richard J. Povinelli,<sup>‡</sup> and Daniel S. Sem\*<sup>§</sup>

Department of Mathematics, Statistics, and Computer Science, Department of Electrical and Computer Engineering, and Chemical Proteomics Facility at Marquette, Department of Chemistry, Marquette University, P.O. Box 1881, Milwaukee, Wisconsin 53233-1881

Received June 26, 2006

Cytochrome P450 2D6 (CYP2D6) is used to develop an approach for predicting affinity and relevant binding conformation(s) for highly flexible binding sites. The approach combines the use of docking scores and compound properties as attributes in building a neural network (NN) model. It begins by identifying segments of CYP2D6 that are important for binding specificity, based on structural variability among diverse CYP enzymes. A family of distinct, low-energy conformations of CYP2D6 are generated using simulated annealing (SA) and a collection of 82 compounds with known CYP2D6 affinities are docked. Interestingly, docking poses are observed on the backside of the heme as well as in the known *active site*. Docking scores for the *active site* binders, along with compound-specific attributes, are used to train a neural network model to properly bin compounds as strong binders, moderate binders, or nonbinders. Attribute selection is used to preselect the most important scores and compound-specific attributes for the model. A prediction accuracy of  $85 \pm 6\%$  is achieved. Dominant attributes include docking scores for three of the 20 conformations in the ensemble as well as the compound's formal charge, number of aromatic rings, and AlogP. Although compound properties were highly predictive attributes (12% improvement over baseline) in the NN-based prediction of CYP2D6 binders, their combined use with docking score attributes is synergistic (net increase of 23% above baseline). Beyond prediction of affinity, attribute selection provides a way to identify the most relevant protein conformation(s), in terms of binding competence. In the case of CYP2D6, three out of the ensemble of 20 SA-generated structures are found to be the most predictive for binding.

## INTRODUCTION

Drug lead screening has been an active area of research for many years. Due to the tedious and expensive nature of experimental screening procedures, computational compound screening has been pursued extensively in recent years.<sup>1</sup> Approaches often incorporate molecular docking simulations into the virtual screening process.<sup>2,3</sup> Traditionally, a major limitation of docking algorithms has been a lack of protein flexibility, which would otherwise require excessive computation time to process. This limitation is known as the "protein docking problem". To compensate, some docking applications permit limited binding site side-chain movement in order to introduce some flexibility within the protein target;<sup>4</sup> nevertheless, there still exists a need for backbone movement in order to fully capture the physiologically relevant binding conformation(s). Furthermore, even under optimal conditions, docking alone is usually not very predictive of binding affinity for flexible proteins and may therefore benefit by supplementing with 1D QSAR models to enhance predictability.

While an exhaustive simulation of 3-dimensional protein search space is currently infeasible because of computational

limitations, knowledge-based inferences can be used to reduce the complexity of the protein docking problem. For example, consideration of movement only around the binding site may sufficiently preclude unnecessary search space. Since binding sites tend to comprise a small percentage of a protein's volume, a simulation of the binding site alone will significantly reduce the search complexity. Additionally, certain segments of the protein are likely to be less important than others for the binding process. By excluding such segments from consideration when evaluating binding site flexibility, the search space is further restricted. Such an approach to docking is best developed on a model enzyme system with a highly flexible binding site.

A well-known enzyme family responsible for the metabolism of a multitude of drug compounds and other xenobiotics is the cytochrome P450 (CYP) family. The CYP family has the broadest substrate specificity of any enzyme family. These monooxygenases, or mixed-function oxidase enzymes, metabolize the majority of clinically utilized drugs. The CYPs are heme proteins, whose central iron is the site of oxidation in the *active site*. Among the prominent drug-metabolizing CYPs in humans is CYP2D6, a liver enzyme that metabolizes 20–25% of all drugs.<sup>5</sup> CYP2D6 is of particular interest in the pharmaceutical industry because many of its substrates cannot be metabolized by other enzymes, and overly rapid metabolism leads to short serum half-life for a drug. Furthermore, inhibition of CYP2D6-based metabolism of one drug by another drug can lead to

\* Corresponding author phone: (414)288-7859; e-mail: daniel.sem@marquette.edu. Corresponding author address: 535 North 14th Street, Milwaukee, WI 53233.

<sup>†</sup> Department of Mathematics, Statistics, and Computer Science.

<sup>‡</sup> Department of Electrical and Computer Engineering.

<sup>§</sup> Chemical Proteomics Facility at Marquette, Department of Chemistry.

dangerous drug–drug interactions. CYP2D6 is an especially good model system for our studies, because of its broad substrate specificity. Over 400 ligands have recently been reported<sup>6</sup> for CYP2D6, of which ~60% are substrates and ~40% are inhibitors. To accommodate such a wide range of ligands must require significant binding site flexibility and malleability, which is not easily modeled using most current docking approaches. The docking process described herein is specifically designed to address binding site flexibility. The only other CYP isoforms that bind more substrates and inhibitors than CYP2D6 are CYP3A4 and CYP2C, but our studies begin with CYP2D6 because the class of ligands bound by it is more easily identified—it generally prefers to bind amines. Future studies will be directed to the somewhat more challenging CYP3A4 and CYP2C isoforms.

In this paper, we present an approach to circumventing the protein docking problem and improving the predictability of binding affinities, using CYP2D6 as the target enzyme. There are 4 major steps in the process: (1) identification of structurally invariant segments of a protein, across the gene family to which it belongs; (2) conformational sampling, while retaining low-energy structures; (3) molecular docking of inhibitors into these distinct, low-energy protein conformations; and (4) attribute selection and construction of neural network (NN) models based on docking scores and compound-specific attributes.

The first step identifies portions of the protein that are not expected to be important for binding specificity. This is done by performing 3D alignment with other human CYPs and labeling sections of the backbone that have a high degree of 3D structural homology across that gene family. Given that these other CYPs bind different substrates, it can be inferred that these “highly homologous” (structurally conserved) segments are not important for binding specificity and can therefore be kept fixed during molecular dynamics conformational sampling and subsequent docking studies.

The second step utilizes molecular dynamics (MD) to generate multiple conformations of CYP2D6. The highly homologous segments identified in step 1 are held fixed during dynamics, thereby precluding their contribution to conformational fluctuations. Twenty CYP2D6 conformations are selected, which have distinct binding sites and yet still retain low-energy conformations.

The third step involves automated docking of a set of 82 inhibitors with known binding affinities for CYP2D6. These compounds are docked into all 20 protein conformations, and the binding affinities are calculated using two different scoring functions. Ultimately, we seek to identify structures that allow for more accurate docking simulations, i.e., calculated binding affinities that are closer to the true affinities. Since docking scores alone are often not highly predictive of affinity, these are supplemented with other data in step 4.

The fourth step uses the docking scores as attributes, along with compound-specific attributes (molecular weight, AlogP, etc.), to generate neural network models that predict binding affinity. In addition to generating models, this process also identifies protein conformations that may be most relevant for inhibitor binding. This information is obtained from the attribute selection process, in which protein conformations that are most predictive of binding affinities are selected.

## RESULTS AND DISCUSSION

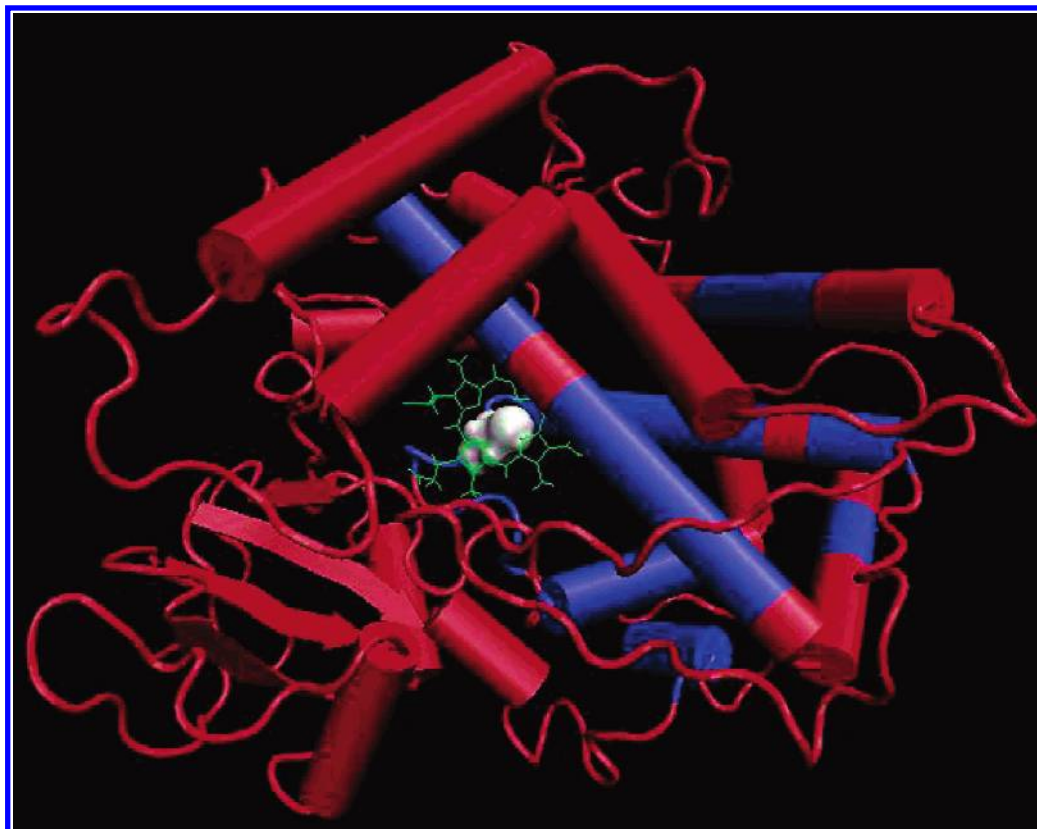
**3D Alignment.** Altogether, 9 segments comprising 82 residues, are found to be highly homologous (structurally conserved) between the 9 CYP structures aligned in 3D space. These are shown in blue in Figure 1, which shows a “cartoon” representation of CYP2D6 secondary structure, viewed from above the binding site. The white surface behind the green heme is the proximal cysteine ligand, common to all CYPs. Of special interest is the structural conservation of the long I helix, which sits on top of the heme and comprises a wall of the *active site* pocket. This structural conservation is disrupted in the region of helix I that comprises the *active site* wall and which has a well-known distortion/kink, that may have functional relevance. This region is apparent as the red segment above the heme.

While 115 total residues are found to be structurally conserved across the CYPs, a requirement that there be 4 contiguous residues that are structurally conserved across all 9 protein structures reduced this to 82 residues (shown in blue in Figure 1). A cutoff of 4 sequential residues is chosen to avoid random correlations between structures, such as  $\beta$ -turns and crossings of the backbone, which would not constitute a true structural alignment of the CYP backbone residues. Since a secondary structural element usually consists of at least 4 residues, this seemed to be an appropriate threshold to ensure true backbone structural alignments across the 9 structures.

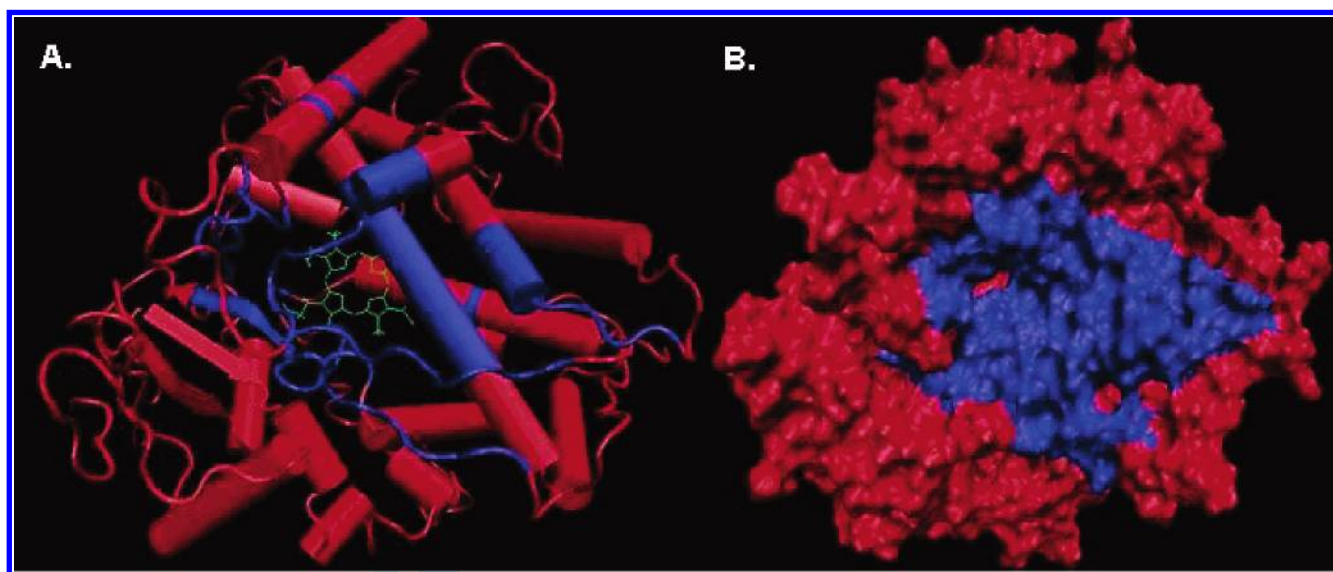
Given that these segments are extremely similar in 3D positioning between multiple CYP enzymes (2C8, 2C9, and 3A4) and given that these enzymes also have different substrates bound (3 with 2C9 and 4 with 3A4), it can be inferred that these structurally homologous segments are less significant for binding specificity than other regions of the protein. Consequently, these portions of CYP2D6 are immobilized during the MD simulations, to simplify the calculations and to prevent conformational changes associated with their movement.

**Molecular Dynamics.** The MD (molecular dynamics) calculations generated 20 protein structures that have overall low potential energies and whose binding site atoms are distinct in 3D space. Twenty conformations are sought so as to capture a range of binding site conformations, while 3-dimensional distinction is necessary to produce nonredundant structures for docking calculations. It is important to maintain conformations with low potential energy, which indicates stability and therefore greater concentration in solution. The structurally homologous segments of CYP2D6 are kept rigid during the MD simulations. Specifically, the C $_{\alpha}$  atoms for these residues are held fixed, such that large structural changes cannot take place in close proximity to these segments. Significant conformational changes in the protein are limited to other regions, including the majority of the binding site residues. A first attempt at performing MD without significant heating generated little structural variation, with a 50 ps run at 400 K generating 50 000 structures that all differed by <0.8 Å RMSD from each other (C $_{\alpha}$  atoms). Hence, a simulated annealing (SA) approach is required to better sample conformational space.

The 20 protein structures are generated by performing a SA heating protocol 20 times, using the same initial conformation. SA involves heating the system to an ex-



**Figure 1.** Structural invariance in CYP2D6. Highly homologous (structurally conserved) segments of CYP2D6, compared to other CYPs, are colored in blue.



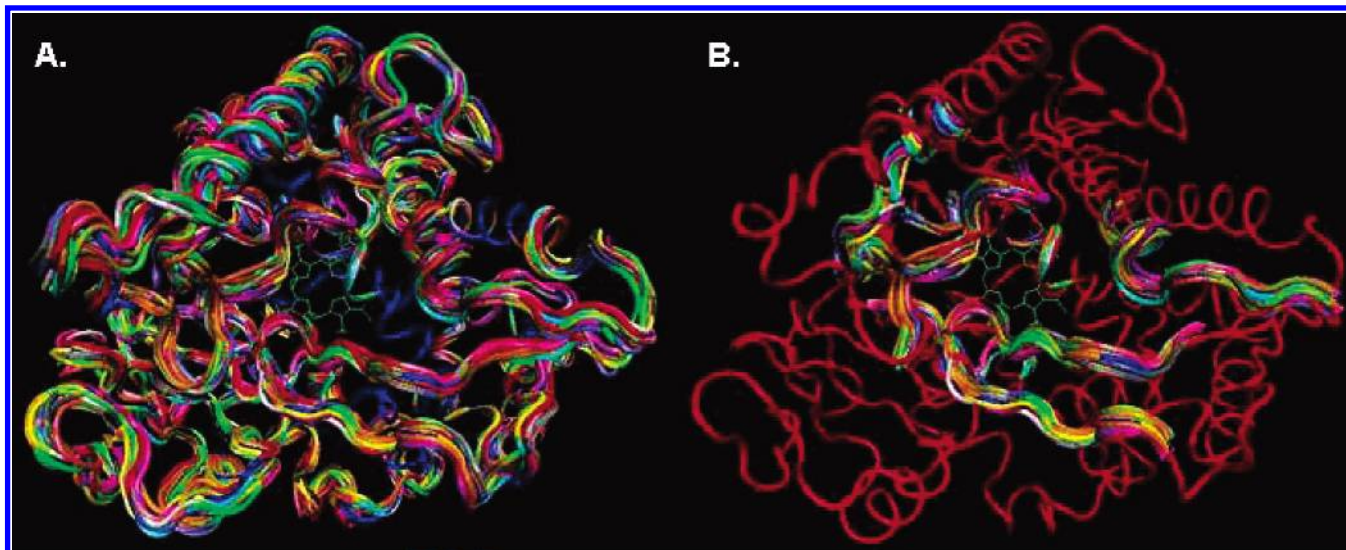
**Figure 2.** Binding site of CYP2D6. Binding site residues for CYP2D6, as defined in this study, are colored in blue in either a cartoon (A) or surface representation (B).

tremely high temperature and then cooling it slowly until it reaches a low-energy conformation. The SA protocol is often used to generate multiple low-energy conformers,<sup>7</sup> since the extreme temperature allows the protein structure to escape local energy minima that would otherwise prevent significant conformational changes. The potential energy of the protein is assessed after each SA run and compared to the energy of the initial structure. Distinction between structures is assessed by measuring RMSD differences between atoms of binding site residues. These residues are selected by visual inspection of the *active site* pocket, using the program VMD.<sup>8</sup>

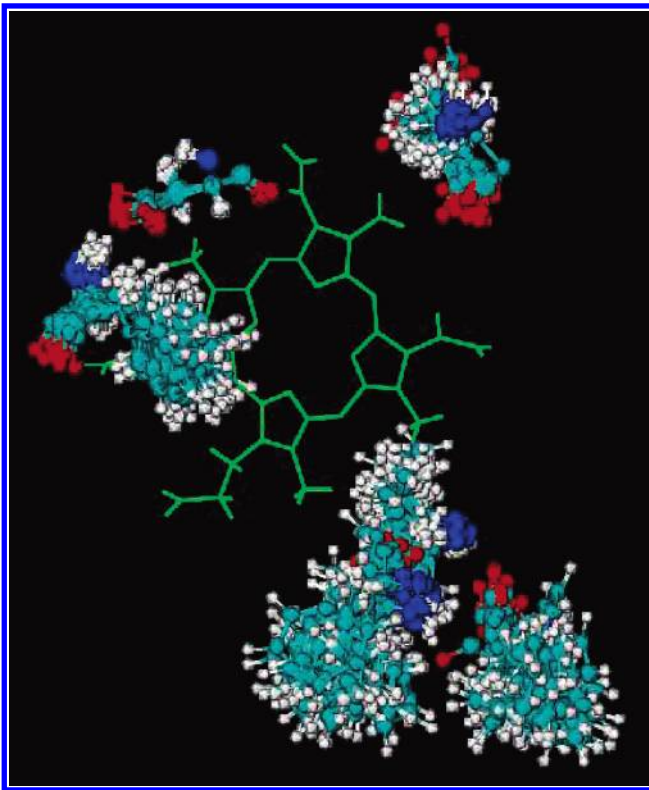
Figure 2 shows the binding site face of CYP2D6, as defined for the calculations performed herein, with the designated binding site residues specified in blue.

The SA calculations generate some significant fluctuations in side-chain and backbone conformations, although large shifts of helices and/or loops do not occur. Figure 3A shows backbone variation for the 20 SA runs. In this figure, the 20 structures are overlaid so as to illustrate the structural variation across the whole protein, although the structural conservation of helix I is apparent (the blue helix located above the heme). As expected, the F–G loop (left side of





**Figure 3.** SA results for CYP2D6. Twenty overlaid conformations of CYP2D6, which resulted from the 20 separate SA runs, are shown either for the entire backbone (A) or for highly variable regions in and near the *active site* (B).



**Figure 4.** Side-chain variability for *active site* and related residues. For the 20 structures generated in the SA process, side chains are shown for residues that are thought to be important for substrate and inhibitor binding: Asp301, Glu216, Phe120, Phe481, Phe483, and Val374.

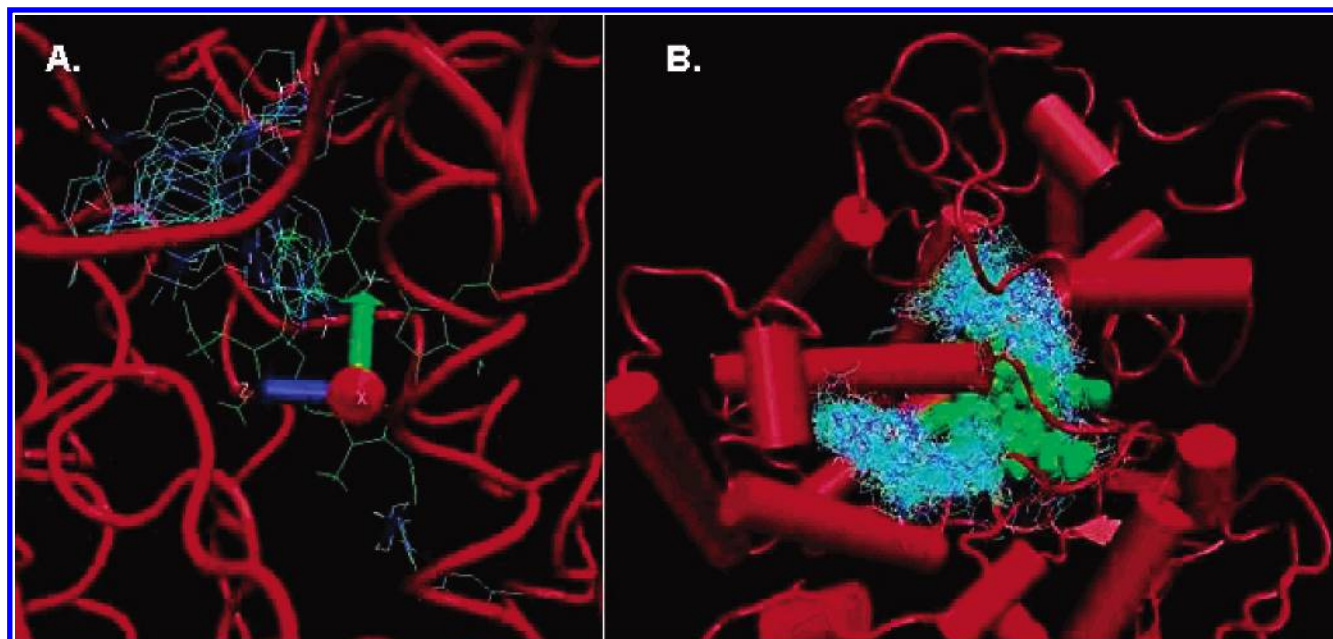
figure) shows significant variability. This variability is more apparent in Figure 3B, where only the variation of backbone residues around the *active site* cavity is shown. This figure illustrates the residues that are considered during RMSD calculations. Especially relevant for docking calculations is the variability that is observed for side chains of residues that have previously been reported to have a role in substrate and/or inhibitor binding (Figure 4).

**Docking.** The third phase of this work utilizes the 20 protein structures for molecular docking simulations. To perform the docking, the program AutoDock 3.0<sup>9</sup> is used.

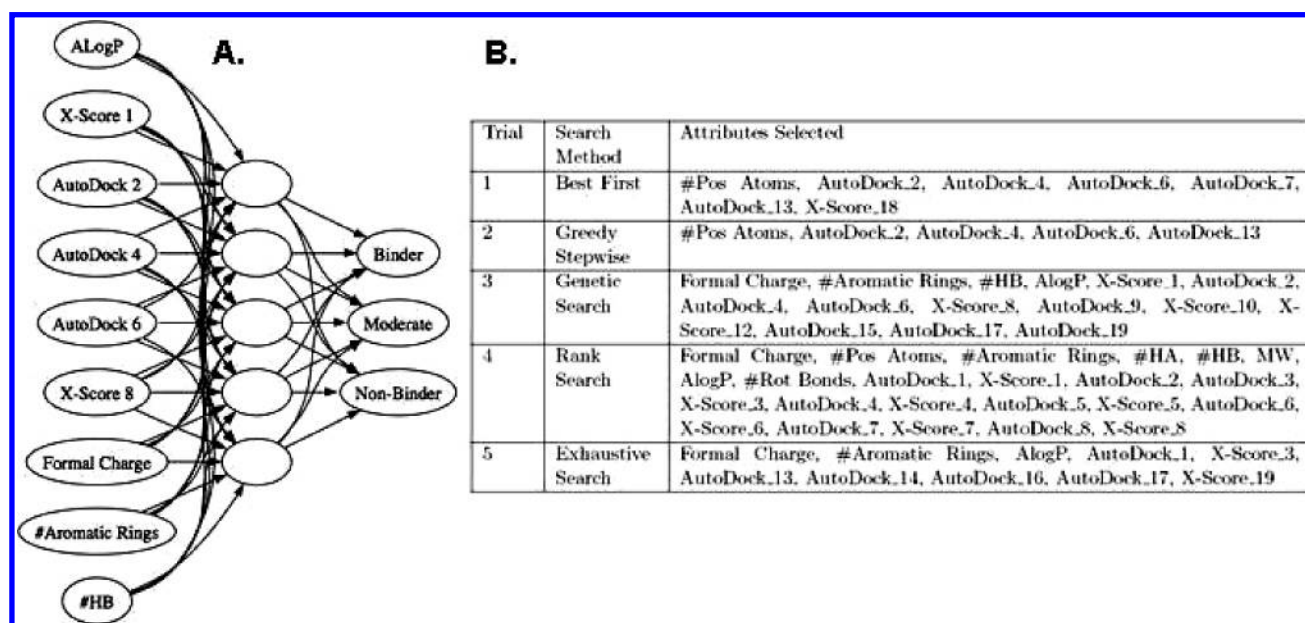
Additionally, the graphical interface, AutoDockTools,<sup>10</sup> is used to prepare the protein structures for the docking runs.

A data set of 82 compounds, selected from an article by Kemp et al.,<sup>11</sup> is used for automated docking. This data set consists of three smaller data sets of 21, 30, and 31 compounds, which are found in refs 12 and 13 and from NCI,<sup>14</sup> respectively. After docking the 82 compounds 10 times each into each of the 20 structures, there are 16 400 total docking results. In this case, a docking result consists of a calculated free energy of interaction (i.e. “score”) and a docked conformation of the compound within the protein (i.e. “pose”). The majority of these poses are located inside the binding site and close to the heme—where catalysis occurs (i.e. the *active site*). But, some poses are on the opposite side of the heme, outside of this *active site*. Figure 5A shows the 10 poses for one compound docked into one of the CYP2D6 structures. In this figure, nine of the 10 are inside the binding site near the heme, while the 10th, on the lower right, is on the other side of the heme. On this remote face of the heme there are two pockets that were found to be occupied with high frequency in the docking calculations, and they are summarized in Figure 5B for all of the dockings. For purposes of initial scoring, these distal site binders are ignored. But it is possible that this may actually be another inhibitor binding site, since it is at the protein–protein interaction interface where cytochrome P450 oxidoreductase (CYPOR) is thought to bind, so inhibitors could have an effect on reaction rate by interfering with electron transfer by disrupting the CYP/CYPOR complex.

The AutoDock algorithm utilizes an internal scoring function to assess docking accuracy throughout the simulation. The resulting free energy score is essentially the final calculation of this function. In addition to AutoDock’s scoring function, we rescore each set of 10 docked poses for each compound with X-Score,<sup>15</sup> a stand-alone scoring function application. In previous studies comparing the ability of different scoring functions to accurately predict binding affinities for known protein–ligand complexes, X-Score was found to perform better than a number of other scoring functions.<sup>16,17</sup>



**Figure 5.** Docking poses. (A) Docking for one compound into CYP2D6 generated 10 poses: 9 in the *active site* and one on the backside of the heme. View is from above the *active site* side of the heme. (B) In analyzing all of the docking results, many poses are observed on the backside of the heme—as summarized here. These are shown in CPK colors for the two clusters of poses relative to the heme prosthetic group, which is rendered as a green space filling model. View is from the back of the heme, opposite the *active site*.



**Figure 6.** Neural network (NN) model. (A) Neural network graph of a typical model generated in this study. (B) Results of attribute selection using five different search methods.

Comparison of calculated free energy scores with the known affinities gives relatively poor correlations with both AutoDock and X-Score, with correlation coefficients no higher than 0.3 for any of the 20 structures. This is true whether correlations are made for the 82 compounds with each of the 20 protein structures, or whether the best score of the 20 structures is used in the correlation. Hence, a neural network (NN) approach is employed, whereby the docking scores are used as attributes along with compound-specific attributes to train a NN model.

**Neural Network Model.** The 40 docking scores (20 AutoDock, 20 X-Score) are used as attributes for training neural network models, along with compound-specific attributes that include the following: HA (*number of hydro-*

*gen bond acceptors*), HB (*number of hydrogen bond donors*), MW (*molecular weight*), *number of rotatable bonds*, ALogP, *formal charge*, *number of aromatic rings*, and *number of positive atoms* (Figure 6). A baseline NN model with all attributes and without scores for the second binding site is only slightly better than labeling all compounds as “binder” ( $63 \pm 6\%$  vs  $62\%$ ). Including scores for the second binding site on the back of the heme (Figure 5) improves accuracy ( $73 \pm 6\%$ ), suggesting it might be a relevant site. Attribute selection was done using five different search methods (Table 1). Rank search and exhaustive search improve accuracy slightly ( $74 \pm 6\%$ ), while other attribute selection algorithms do not show improvements over using all attributes, with the second binding site scores. Analysis of the confusion



**Table 1.** Attribute Selection Process

trial	search method	attributes selected	accuracy
1	best first	#pos atoms, AutoDock_2, AutoDock_4, AutoDock_6, AutoDock_7, AutoDock_13, X-Score_18	70.7 $\pm$ 6.5
2	greedy stepwise	#pos atoms, AutoDock_2, AutoDock_4, AutoDock_6, AutoDock_13	68.3 $\pm$ 6.5
3	genetic search	formal charge, #aromatic rings, #HB, AlogP, X-Score_1, AutoDock_2, AutoDock_4, AutoDock_6, X-Score_8, AutoDock_9, X-Score_10, X-Score_12, AutoDock_15, AutoDock_17, AutoDock_19	70.7 $\pm$ 6.5
4	rank search	formal charge, #pos atoms, #aromatic rings, #HA, #HB, MW, AlogP, #rot bonds, AutoDock_1, X-Score_1, AutoDock_2, AutoDock_3, X-Score_3, AutoDock_4, X-Score_4, AutoDock_5, X-Score_5, AutoDock_6, X-Score_6, AutoDock_7, X-Score_7, AutoDock_8, X-Score_8	74.4 $\pm$ 6.5
5	exhaustive search	formal charge, #aromatic rings, AlogP, AutoDock_1, X-Score_3, AutoDock_13, AutoDock_14, AutoDock_16, AutoDock_17, X-Score_19	74.4 $\pm$ 6.5

Binder	Moderate	Non-binder	Classified As
44	8	6	Binder
7	14	3	Moderate
0	0	0	Non-binder

Trial 1.

Binder	Moderate	Non-binder	Classified As
44	10	4	Binder
7	12	5	Moderate
0	0	0	Non-binder

Trial 2.

Binder	Moderate	Non-binder	Classified As
44	9	0	Binder
7	8	3	Moderate
0	5	6	Non-binder

Trial 3.

Binder	Moderate	Non-binder	Classified As
46	7	1	Binder
4	8	1	Moderate
1	7	7	Non-binder

Trial 4.

Binder	Moderate	Non-binder	Classified As
46	5	2	Binder
5	11	3	Moderate
0	6	4	Non-binder

Trial 5.

**Figure 7.** Confusion matrices for the five NN models presented in Figure 6B and Table 1.

matrices (Figure 7) reveals that although errors in classifying ligands as binders, nonbinders, or moderate binders do occur, the most serious misclassification of binder as nonbinder (and nonbinder as binder) is rare. Indeed, not a single such misclassification occurred in trial 3, where attribute selection was done using the genetic search. Only two such misclassifications occurred in trials 4 and 5 (rank search and exhaustive search), which yielded slightly more predictive models overall.

Several interesting conclusions about protein conformations result from the classification and attribute selection. Protein conformation #6 (63  $\pm$  6%) is found to provide the most predictive docking scores, followed by #5 (62  $\pm$  6%) and #14 (62  $\pm$  6%), suggesting they may represent relevant conformations of the CYP2D6 protein, in terms of binding competence. The only scoring attributes selected by at least four of the search methods are the AutoDock scores for docking into CYP2D6 conformations #2, #4, and #6 (see Table 2), suggesting the relevance of these protein conformations, relative to the other 17. Conformation #6 is selected only once with #5 and not at all with #14 (Table 1), suggesting these conformations have related binding site

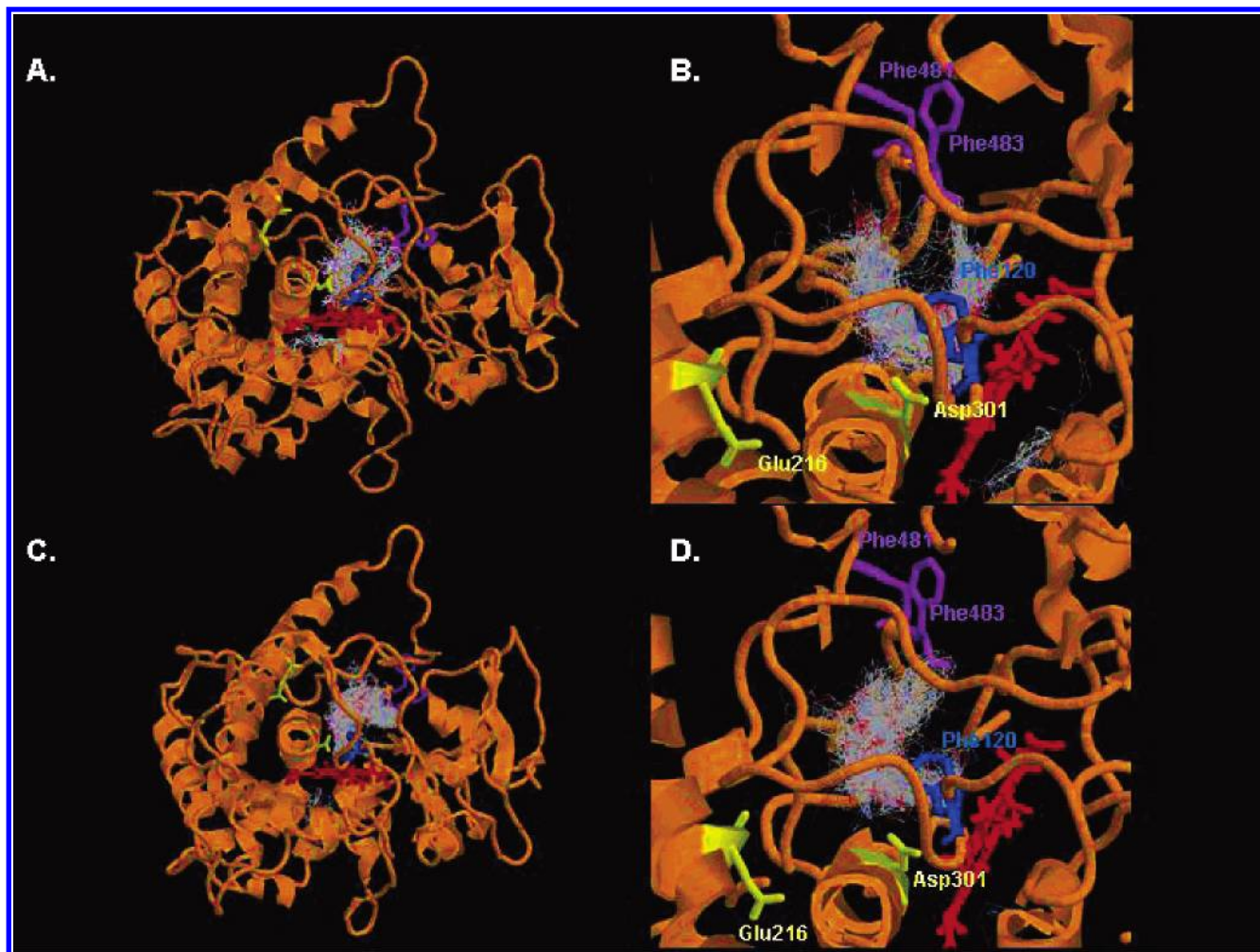
**Table 2.** Attribute Selection Frequency in Trials 1–5

attribute	count	attribute	count
formal charge	3	X-Score_3	2
#pos atoms	3	AutoDock_4	4
#aromatic rings	3	AutoDock_6	4
#HB	2	AutoDock_7	2
AlogP	3	X-Score_8	2
AutoDock_1	2	AutoDock_13	3
X-Score_1	2	AutoDock_17	2
AutoDock_2	4		

shapes that confer similar binding affinities. In contrast, conformations #2 and #4 complement #6 in terms of predicting binding affinity, so may represent different *active site* conformations that are sampled in solution in order to accommodate different ligands. It is especially noteworthy that all attribute selection trials resulted in the selection of multiple docking scores (Table 1), suggesting that an ensemble of protein conformations is involved in ligand recognition.

Of the compound-specific attributes, *AlogP* is the most predictive (71  $\pm$  6%) followed by *number of positive atoms* (70  $\pm$  6%). The most frequently selected compound-specific attributes are as follows: *formal charge*, *number of positive atoms*, *number of aromatic rings*, and *AlogP*. However, since *formal charge* and *number of positive atoms* are common with only two of the five methods, they appear to correlate, as expected. So, consistent with what is already known for CYP2D6 substrates and inhibitors, CYP2D6 tends to bind hydrophobic compounds with a positive charge and an aromatic ring. Such attributes are highly predictive, since an accuracy of 74  $\pm$  6% is achieved using *formal charge*, *number of aromatic rings*, and *AlogP* alone. Conversely, use of only docking scores as attributes provides a similarly high level of accuracy. Therefore, the next question we sought to address is: *does combined usage of docking scores and compound properties provide synergistic increases in accuracy?*

The answer to this question is yes including docking scores as NN attributes together with compound properties increases accuracy significantly. The effectiveness of docking scores as attributes is shown through the evaluation of several different NN structures in combination with various subsets of the attributes, all of which include at least one docking score. Three optimized NNs have accuracies of 78  $\pm$  6%, and their architectures are detailed in the Experimental Section. But, this model was built using docking scores for a given protein conformation that is most predictive across the whole set of ligands in the training set, which may not



**Figure 8.** Overlay of the 82 ligands bound in lowest energy poses, with energies calculated using the AutoDock (A, B) or X-Score (C, D) scoring function. Protein is CYP2D6 in conformation #6.

accurately model the actual binding process. Since a flexible protein binding site might sample any of our 20 conformations, different structures might be relevant for different ligands. That is, protein structure #6 might give the lowest energy complex for ligand A, while protein structure #10 might give the lowest energy complex for ligand B. To model this induced fit type of binding, we used as attributes the “best X-Score” or “best AutoDock score”, where *best* means the lowest energy pose for a given ligand, among the 20 possible CYP2D6 conformations. The most accurate NN model is obtained by using these docking score features. The resulting model (including also HA, MW, ALogP, formal charge, and number of aromatic rings) obtains an accuracy of  $85 \pm 6\%$ , a significant improvement over the model constructed with docking scores using the same CYP2D6 conformation for all ligands ( $78 \pm 6\%$ ) or without docking scores at all ( $74 \pm 6\%$ ). Choosing the *best* docking score (and corresponding CYP2D6 conformation) is therefore much more predictive than forcing a single conformation across all ligands and is a physically more realistic model for enzymes with flexible binding sites, since it simulates “induced fit” binding.

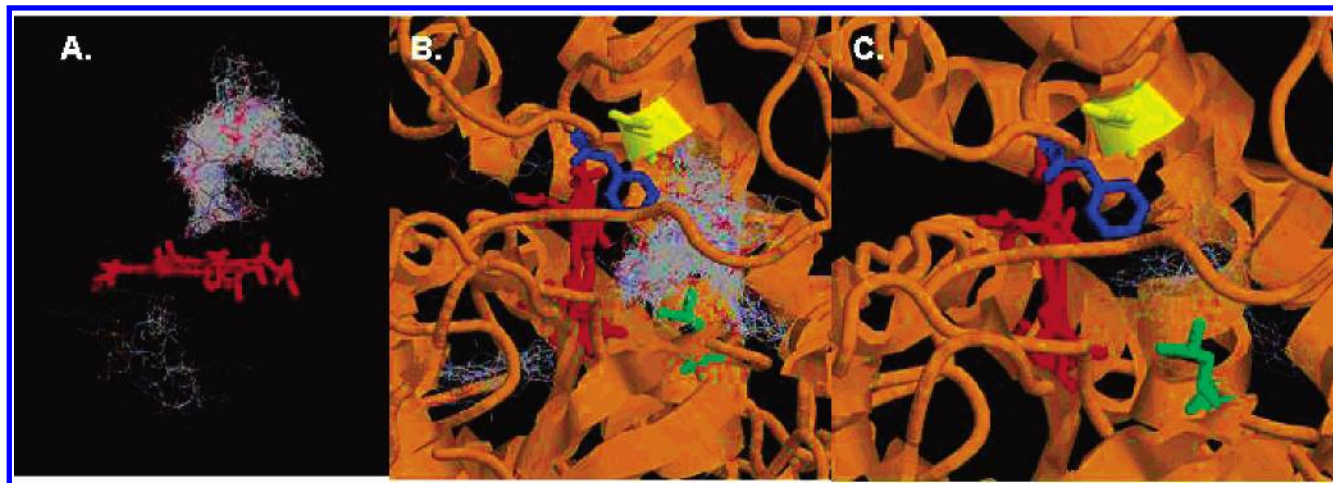
Analysis of the docking poses for all 82 ligands (Figure 8) provides some insight into structure–function relationships for binding site residues. It can be seen that there are clusters of ligands close to the heme as well as somewhat removed and extending to the F–G loop region. The latter ligands

will be more challenging to dock and score accurately, since this region of the CYP structure is known to have the highest flexibility. Positively charged amines on ligands (colored blue) can be seen at various regions in the binding site, although proximity to Asp301 and Phe120 is seen for some of the higher affinity ligands (Figure 9), as expected.<sup>18,19</sup> Interestingly, the cluster of 82 ligand poses assumes a U-shape, with the curvature apparently caused by steric interference from Leu484. This residue, along with Phe120, creates a 6 Å constriction in the binding site (Figure 9), defining a hydrophobic wall for ligands. It can be seen that X-Score and AutoDock scoring functions deal slightly differently with these interactions (Figure 8), although both show this type of U-shaped clustering due to Leu484. It should also be noted that although previous literature reports have suggested a role for Phe481/Phe483<sup>20,21</sup> in binding<sup>16c,d</sup> our docking results suggest that these residues do not participate directly in binding. Rather, their location is such that they interact with the F–G loop, possibly locking it into place to permit proper formation of that surface of the binding site. It is also possible that Phe481/Phe483 interactions permit proper placement of Leu484 which is itself centrally located in the binding site (Figure 9B,C).

#### SUMMARY

This study presents an approach to predicting binding affinity that implicitly considers binding site flexibility. It





**Figure 9.** Overlay as in Figure 8C,D, highlighting the effect of the hydrophobic wall comprised of Leu484 and Phe120 on cluster shape. (A) Protein has been removed for clarity to show how docking poses cluster in a U-shaped manner. (B) Protein is rendered as a cartoon with heme (red), Asp301 (yellow), Leu484 (green), and Phe120 (blue) shown, demonstrating the proximity of Leu484 to the U-curve in the clustered poses (6 Å). (C) Rendering as in B, but only the 13 highest affinity ligands are shown ( $IC_{50} < 1 \mu M$ ).

starts by identifying structurally conserved regions across a gene family, to determine regions to be kept fixed in subsequent simulated annealing runs, to generate multiple conformations for docking studies. Applied to CYP2D6, this results in 20 different conformations in which to dock potential ligands or inhibitors. The conformations that are most relevant for binding are selected by treating docking scores as attributes for a NN model and using attribute selection to identify those attributes (i.e. structures) that are most predictive of binding affinity. This NN/attribute selection strategy shows that only 3 out of 20 conformations of CYP2D6 are most relevant for binding. Binding affinity predictability is significantly enhanced by the inclusion of compound specific attributes as inputs to the NN model. In the case of CYP2D6, the most important compound specific attributes are the *number of positive charges*, ALogP (hydrophobicity), and the *number of aromatic rings*. This strategy therefore combines the benefits of 1D QSAR modeling with 3D docking and is able to achieve a prediction accuracy of  $85 \pm 6\%$ . This approach implicitly includes a consideration of side-chain flexibility and the possibility of multiple binding modes, which is especially important for proteins with large and malleable/dynamic binding sites, like the CYPs.

## EXPERIMENTAL SECTION

**3D Alignment.** The 3D alignment phase involves a homology model of CYP2D6 (created as described previously<sup>22</sup>) and Protein Data Bank (PDB)<sup>23</sup> crystal structures for CYP2C8, CYP2C9, and CYP3A4. CYPs 2C9 and 3A4 have three and four available structures, respectively, due to crystallization with different substrates or inhibitors. These additional structures are included so as to expand the conformational search space. Altogether there are 9 PDB structures that are compared. CYPs 2C8 and 2C9 are dimers, with A and B chains: only the A chains are used in the alignment. Furthermore, only  $C_{\alpha}$  atoms are aligned.

Segments of the CYP2D6 binding site are designated as less significant than others in determining binding specificity if they are found to be “highly homologous” (structurally conserved) among the other human CYPs. To assess

structural homology, rigorous 3D alignment is performed. While sequence alignment is a useful approach for deciphering protein homology, a 3D alignment is more appropriate for this problem, given the 3D nature of the search space.

There are 2 overall steps to the 3D alignment phase: (1) a coarse-grained, global alignment and (2) a fine-tuned, local alignment. Both of these steps involve RMSD calculations using  $C_{\alpha}$  atoms. The global alignment involves pairwise combinations of CYP2D6 with one of the other 8 structures, using the Combinatorial Extension<sup>24</sup> (CE) 3D alignment tool. This application performs iterative alignment of two protein chains using distance-based heuristics. The globally aligned pairs are then subjected to local alignment with an in-house Perl script, which uses a 1.5 Å RMSD window to designate individual residues as “highly homologous”. Segments of the backbone that have four contiguous highly homologous (structurally overlaid/proximal) residues in all nine structures are designated as highly homologous segments.

**Molecular Dynamics.** The second phase of this work involves performing molecular dynamics (MD) simulations in order to generate 20 distinct yet low-energy conformations. The CYP2D6 structure used in this work is minimized with the AMBER94<sup>25</sup> molecular mechanics force field. Prior to molecular dynamics, the CYP2D6 structure is converted to the CHARMM22<sup>26</sup> atom types, and energy minimization is performed. Both of these steps are carried out with the HyperChem version 7.5 (Hypercube, Inc.; Gainesville, FL) molecular modeling package. Steepest descent minimization is used to remove the large initial forces: 4600 cycles are performed, which leaves the energy well below 0 kcal/mol. This is followed by conjugate gradient minimization, which continues until the structure has a root-mean-square gradient of less than  $0.1 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$ .

Version 2.5 of NAMD<sup>27</sup> is used to perform the MD simulations. NAMD requires 4 input files—protein coordinates file, protein structure file (PSF, which contains information about atoms and geometries), force field parameter file, and NAMD configuration file. The coordinates and PSF files are generated with psfgen, a NAMD utility program, while the parameter file is obtained from the force field, CHARMM22 in this case. The dynamics are broken



into 2 main phases, and thus 2 configuration files are used. The first phase involves energy minimization for 1 ps, followed by heating from 0 K to 300 K over 3 ps, and, finally, 200 ps of equilibration. The second phase, simulated annealing, heats the system from 300 K to 1000 K over 3 ps, keeps a constant 1000 K for 10 ps, and cools the system to 300 K over 10 ps. The simulated annealing protocol is performed 20 times, each using the same system configuration resulting from the first phase of dynamics. The highly homologous segments are held fixed during both dynamics phases. Scaled 1–4 interactions are used, with a 1–4 scaling of 1.0. Additional parameters include a time step of 1 fs, cutoff distance of 12.0 Å, switch distance of 10.0 Å, pairlist distance of 14 Å, nonbonded frequency of 1, steps per cycle of 1, and dielectric of 1.0.

An in-house Perl script is used to calculate the RMSD between individual corresponding atoms and then to calculate an average RMSD for the binding site. The script compares the average RMSD for each structure to every other structure (including the original structure), noting whether the difference is greater than 1.0 Å, the predetermined threshold.

**Molecular Docking.** Molecular docking is performed with version 3.0 of AutoDock. To prepare the protein structures, AutoDockTools is used to assign partial charges and solvation parameters. AutoDockTools does not recognize non-standard protein residues, such as the heme prosthetic group; however, the iron and pyrrole nitrogens are fully included in the final structures. This is accomplished by adding the heme partial charges to the protein files prior to running AutoGrid3 and changing the VDW and epsilon values for the M and X atoms in the protein GPF files prior to running AutoDock3.

Preparation of the compounds for docking involves 5 or 6 steps, depending on the initial data. Three data sets, gleaned from Kemp et al.,<sup>11</sup> are used in this work. These are designated the Ekins,<sup>12</sup> Strobl,<sup>13</sup> and NCI<sup>14</sup> sets. The Ekins and Strobl sets are only available as drawings in the respective papers and therefore need to be drawn into 2 dimensions. This is done with ChemDraw, a component of ChemOffice Ultra (CambridgeSoft Corp.; Cambridge, MA). The NCI set is available for download as 2D structures. The three data sets are converted into 3D with CORINA version 3.2 (Molecular Networks GmbH; Erlangen, Germany) and then minimized using Gaussian AM1 calculations (revision C.02, Gaussian, Inc.; Wallingford, CT). The initial NCI set has 33 compounds, but two do not pass Gaussian minimization and are therefore discarded. Subsequently, the compounds are assigned atom types and converted to MOL2 format with Mol2Mol (version 5.02, Interware). Partial charges (Gasteiger) are then added with version 4.1.5 of Pipeline Pilot (SciTegic, Inc.; San Diego, CA). Finally, the program AutoTors, which is part of the AutoDock suite, is used to assign rotatable bonds and format the compounds for use with AutoDock.

AutoDock's Lamarckian Genetic Algorithm (LGA) is used with all default values, and 10 runs are performed for each compound/CYP2D6-conformation pair. The GA component defaults include 50 initial individuals, 25 000 maximum energy evaluations, or 27 000 maximum generations, an elitism of 1 individual, a mutation rate of 0.02, and a crossover rate of 0.80. Additionally, a search space of 60 Å × 60 Å × 60 Å is used, which is more than sufficient to

surround the entire protein. Each compound is docked into each CYP2D6-conformation 10 times, for a total of 200 docking runs per compound.

**Neural Network Model Construction.** Multilayer perceptron neural network (NN) models are used to predict binding affinities for compounds. Each compound is represented by attributes, which include the 40 docking scores (20 AutoDock, 20 X-Score) and the following compound-specific attributes calculated with Pipeline Pilot: HA (*number of hydrogen bond acceptors*), HB (*number of hydrogen bond donors*), MW (*molecular weight*), *number of rotatable bonds*, ALogP, *formal charge*, *number of aromatic rings*, and *number of positive atoms* (Figure 6A). Training and testing a NN requires data with known affinities. The Kemp data for the 82 compounds was used for training and testing with their affinities binned as “binder” (IC<sub>50</sub> < 20 μM), “moderate binder” (20 μM ≤ IC<sub>50</sub> < 500 μM), and “nonbinder” (IC<sub>50</sub> > 500 μM). Since the data set contains 51 known binders out of 82 total compounds, a classifier that uses only the prior data distribution has an accuracy of 62% (51/82). Hence, accuracies greater than 62% are considered to be above chance. All data values were normalized to a [−1,1] scale using a linear transformation.

The process of constructing the NNs involves two steps. The first step uses attribute selection algorithms to select highly predictive subsets of attributes from the initial 40 docking scores and 8 compound-specific attributes. Five search methods are used for attribute selection: best first, greedy stepwise, genetic search, rank search (with information gain), and exhaustive search (Figure 6B). A C4.5 decision tree model, due to its efficient implementation, is used to evaluate attribute subsets. Use of multiple attribute search methods permits the construction of different NN models. This strategy of exploring model space using different feature sets was chosen over a GA-based construction of multiple models, because it avoids the need to combine models using techniques such as voting or regression modeling.

The second step in constructing NNs involves selecting a NN architecture and a training duration. NNs are highly parametrized models, and the following paragraphs elaborate on model architectures investigated. A 10-fold cross-validation is used to evaluate NN accuracy. In 10-fold cross-validation,<sup>28</sup> the data are partitioned into 10 nonoverlapping and approximately equal sized sets. A neural network model is trained on 9 of the 10 partitions, and its accuracy is measured on the 10th. This process is repeated so that each partition is used as a test set. The accuracy estimate is therefore the average accuracy obtained across the 10 folds. The error estimates are obtained from the accuracy variance across the 10-folds. Because the test set is not used when training the model, accuracy estimates are not subject to overfitting effects. Indeed, when our best neural network model is trained and tested on the same set of 82 compounds, the accuracy is 92%, which would be an overestimate of accuracy and indicative of overfitting. It should also be noted that since we are not predicting numeric affinities, but rather are predicting categories of affinities, cross-validation is the best approach for measuring accuracy (as opposed to  $R^2$  and  $Q^2$ ). Weka version 3.4.4<sup>29</sup> is used to perform attribute selection and NN design, training, and testing.

A systematic NN architecture is used to evaluate the results of attribute selection, resulting in a total of 6 NNs (one using all compound attributes, and one for each of the five attribute selection algorithms). The NN architecture consists of an input layer with one node per selected attribute, an output layer with three nodes (one for each affinity bin), and a single hidden layer with the number of nodes given by  $(\#attributes + \#bins)/2$ . Each node uses the logistic activation function. Training uses back-propagation for 500 epochs, where 1 epoch involves propagating each of the 82 instances through the NN.

Additional architectures are constructed using insights from the authors to improve on systematically constructed NNs. Three optimized NNs are described. The first NN uses the first nine attributes (*formal charge*, *#aromatic rings*, *AlogP*, *AutoDock 1*, *X-Score 3*, *AutoDock 13*, *AutoDock 14*, *AutoDock 16*, *AutoDock 17*) selected using Weka's exhaustive search, one hidden layer with six neurons, and 500 epochs for training. The second NN architecture uses the first nine attributes (*formal charge*, *#pos atoms*, *#aromatic rings*, *#HA*, *#HB*, *MW*, *AlogP*, *#rot bonds*, *AutoDock 1*) selected using Weka's rank search and has two hidden layers with a "5–2" scheme and 550 epochs for training. The final NN architecture uses the first nine attributes (*formal charge*, *#aromatic rings*, *#HB*, *AlogP*, *X-Score 1*, *AutoDock 2*, *AutoDock 4*, *AutoDock 6*, *X-Score 8*) selected using Weka's genetic search and has two hidden layers with a "5–2" scheme and 500 epochs for training. Back-propagation and logistic activation functions are used as described in the previous paragraph.

The best NN uses the two most predictive docking score features, "best" AutoDock score and "best" X-Score score, along with five compound features: HA, MW, *AlogP*, *formal charge*, and *number of aromatic rings*. The NN contains a single hidden layer with five nodes (determined by  $(\#attributes + \#bins)/2$ ) and uses 500 epochs for training. Back-propagation and logistic activation functions are used as described above.

**Note Added in Proof.** Since the completion of this study, a 3.0 Å resolution crystal structure of human CYP2D6 (pdb code 2F9Q) was published.<sup>30</sup> This structure validates the CYP2D6 model used in our study (1.66 Å backbone RMSD difference, for 288 residues), with the highest structural match being in the binding site. The difference between our model and the crystal structure is less than the resolution of the crystal structure and is in any case sampled using our MD analysis. Greatest variation is in the F–G loop region, which is known to be dynamic, and was doubly mutated in the crystal structure to increase solubility.

#### ACKNOWLEDGMENT

We thank Aurora D. Costache for helpful advice. Remote connectivity for doing calculations was with Abilene/Internet2, funded by NSF (ANI-0333677).

**Supporting Information Available:** Scripts for processing structures. This material is available free of charge via the Internet at <http://pubs.acs.org>.

#### REFERENCES AND NOTES

- Wu, G.; Vieth, M. SDOCKER: A method utilizing existing X-ray structures to improve docking accuracy. *J. Med. Chem.* **2004**, *47*, 3142–3148.
- Ferrari, A. M.; Wei, B. Q.; Costantino, L.; Shoichet, B. K. Soft docking and multiple receptor conformations in virtual screening. *J. Med. Chem.* **2004**, *47*, 5076–5084.
- Bajorath, J. Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discovery* **2002**, *1*, 882–894.
- Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE: Efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **2001**, *308*, 377–395.
- Ingelman-Sundberg, M. Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): Clinical consequences, evolutionary aspects and functional diversity. *Pharmacogenomics J.* **2005**, *5*, 6–13.
- Rendic, S. Summary of information on human CYP enzymes: human P450 metabolism data. *Drug Metab. Rev.* **2002**, *34*, 83–448.
- Höltje, H.; Sippl, W.; Rognan, D.; Folkers, G. In *Molecular Modeling: Basic Principles and Applications*, 2nd ed.; Wiley-VCH: Weinheim, 2003; pp 18–19.
- Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics. *J. Mol. Graphics Modell.* **1996**, *14*, 33–38.
- Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
- Sanner, M. F. Python: A programming language for software integration and development. *J. Mol. Graphics Modell.* **1999**, *17*, 57–61.
- Kemp, C. A.; Flanagan, J. U.; van Eldik, A. J.; Maréchal, J.; Wolf, C. R.; Roberts, G. C. K.; Paine, M. J. I.; Sutcliffe, M. J. Validation of model of cytochrome P450 2D6: An in silico tool for predicting metabolism and inhibition. *J. Med. Chem.* **2004**, *47*, 5340–5346.
- Ekins, S.; Bravi, G.; Binkley, S.; Gillespie, J. S.; Ring, B. J.; Wikel, J. H.; Wrighton, S. A. Three and four dimensional-quantitative structure activity relationship (3D/4D-QSAR) analyses of CYP2D6 inhibitors. *Pharmacogenetics* **1999**, *9*, 477–489.
- Strobl, G. R.; von Krüedener, 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.
- Developmental Therapeutics Program, National Cancer Institute. <http://dtp.nci.nih.gov/docs/ldrugs/chemdata.html>. For database searching, use the URL <http://dtp.nci.nih.gov/dtpstandard/dwindex/index.jsp>. (Accessed January 17, 2005).
- Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 11–26.
- Wang, R.; Lu, Y.; Fang, X.; Wang, S. An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein–ligand complexes. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2114–2125.
- Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287–2303.
- Flanagan, J. U.; Marechal, J. D.; Ward, R.; Kemp, C. A.; McLaughlin, L. A.; Sutcliffe, M. J.; Roberts, G. C.; Paine, M. J.; Wolf, C. R. Phe120 contributes to the regiospecificity of cytochrome P450 2D6: mutation leads to the formation of a novel dextromethorphan metabolite. *Biochem. J.* **2004**, *380*, 353–360.
- Ellis, S. W.; Hayhurst, G. P.; Smith, G.; Lightfoot, T.; Wong, M. M.; Simula, A. P.; Ackland, M. J.; Sternberg, M. J.; Lennard, M. S.; Tucker, G. T.; Wolf, C. R. Evidence that aspartic acid 301 is a critical substrate-contact residue in the active site of cytochrome P450 2D6. *J. Biol. Chem.* **1995**, *270*, 29055–29058.
- Hayhurst, G. P.; Harlow, J.; Chowdry, J.; Gross, E.; Hilton, E.; Lennard, M. S.; Tucker, G. T.; Ellis, S. W. Influence of phenylalanine-481 substitutions on the catalytic activity of cytochrome P450 2D6. *Biochem. J.* **2001**, *355*, 373–379.
- Lussenburg, B. M.; Keizers, P. H.; de Graaf, C.; Hidestrand, M.; Ingelman-Sundberg, M.; Vermeulen, N. P.; Commandeur, J. N. The role of phenylalanine 483 in cytochrome P450 2D6 is strongly substrate dependent. *Biochem. Pharmacol.* **2005**, *70*, 1253–1261.
- Yao, H.; Costache, A. D.; Sem, D. S. Chemical proteomic tool for ligand mapping of CYP anti-targets: An NMR-compatible 3D QSAR descriptor in the heme-based coordinate system. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1456–1465.
- Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535–542.
- Shindyalov, I. N.; Bourne, P. E. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* **1998**, *11*, 739–747.
- Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, J. P.



- P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (26) MacKerell, A. D., Jr.; Bashford, D.; Bellot, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (27) Kalé, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. NAMD2: Greater scalability for parallel molecular dynamics. *J. Comput. Phys.* **1999**, *151*, 283–312.
- (28) Kohavi, R. A Study of cross-validation and bootstrap for accuracy estimation and model selection. *Int. Joint Conf. Artif. Intell.* **1995**, *14*, 1137–1145.
- (29) Witten, E. F.; Frank, E. *Data Mining: Practical machine learning tools and techniques*, 2nd ed.; Morgan Kaufmann: San Francisco, CA, 2005; pp 265–267.
- (30) Rowland, P.; Blaney, F. E.; Smyth, M. G.; Jones, J. J.; Leydon, V. R.; Oxbrow, A. K.; Lewis, C. J.; Tennant, M. G.; Modi, S.; Eggleston, D. S.; Chenery, R. J.; Bridges, A. M. Crystal structure of human cytochrome P450 2D6. *J. Biol. Chem.* **2006**, *281*, 7614–7622.

CI600267K