Journal of Medicinal Chemistry

ARTICLE

pubs.acs.org/jmc

Structure-Based Site of Metabolism Prediction for Cytochrome P450 2D6

Samuel L. C. Moors,*[,†] Ann M. Vos,[‡] Maxwell D. Cummings,[§] Herman Van Vlijmen,[‡] and Arnout Ceulemans[†]

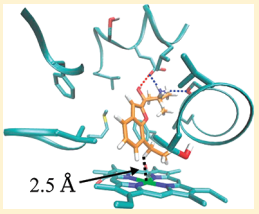
[†]Department of Chemistry, Katholieke Universiteit Leuven, Celestijnenlaan 200F, 3001 Heverlee, Belgium

[‡]Tibotec BVBA, Turnhoutseweg 30, 2340 Beerse, Belgium

[§]Johnson & Johnson Pharmaceutical Research & Development, Welsh and McKean Roads, Springhouse, Pennsylvania 19477, United States

Supporting Information

ABSTRACT: Realistic representation of protein flexibility in biomolecular simulations remains an unsolved fundamental problem and is an active area of research. The high flexibility of the cytochrome P450 2D6 (CYP2D6) active site represents a challenge for accurate prediction of the preferred binding mode and site of metabolism (SOM) for compounds metabolized by this important enzyme. To account for this flexibility, we generated a large ensemble of unbiased CYP2D6 conformations, to which small molecule substrates were docked to predict their experimentally observed SOM. SOM predictivity was investigated as a function of the number of protein structures, the scoring function, the SOM−heme cutoff distance used to distinguish metabolic sites, and intrinsic reactivity. Good SOM predictions for CYP2D6 require information from the protein. A critical parameter is the distance between the heme iron and the candidate site of metabolism. The best predictions were achieved with cutoff distances consistent with the chemistry relevant to CYP2D6 metabolism. Combination of the new ensemble-based docking method with estimated intrinsic reactivities of substrate sites considerably improved the predictivity of the model. Testing on an independent set of substrates yielded area under curve values as high as 0.93, validating our new approach.



INTRODUCTION

The human cytochromes P450 (CYP), members of the superfamily of cysteinato-heme containing CYP enzymes, are involved in the metabolism of a wide variety of endogenous and exogenous compounds.[1,2] CYP2D6 is the second most important drug metabolizing enzyme, responsible for the biotransformation of about 25% of drugs that are known to be CYP substrates.[3] Genetic polymorphism contributes to the extensive variability seen in the expression and activity of CYP2D6,[4] affecting the pharmacokinetics of 50% of the clinical drugs that are CYP2D6 substrates.[5] In drug discovery, early identification of potential CYP substrates and their site of metabolism (SOM) can guide the lead optimization process toward drug candidates that have better pharmacokinetic and toxicological profiles.[6,7] However, experimental metabolite identification remains difficult and time-consuming.[8]

Computational SOM prediction models serve as tools to help medicinal chemists decide which sites on a molecule to modify in order to avoid undesired metabolism, which in turn may help in optimizing drug half-life and reducing drug−drug interactions. To date, reported in silico models can be classified as either ligand-based or structure-based.[9]

Relatively large experimental data sets are available to facilitate the development and testing of ligand-based SOM prediction models. Another attraction of this approach is that protein structure information and related (and typically computationally expensive) calculations are not required. Reported ligand-based models include rule-based, reactivity-based, and QSAR methods. Sheridan and colleagues at Merck applied AM1 molecular orbital calculations to the prediction of dehydrogenation energies for sp^3 hybridized carbons of CYP3A4 substrates.[10] In their reactivity-based approach, prediction accuracy was improved by adding an 8 $Å^2$ minimum surface area exposure criterion for the hydrogen atoms being evaluated. More recently, in a very different approach, Sheridan et al. proposed a QSAR-based model that takes into account only the structures of the substrates to predict the SOMs.[11] The best predictivity was obtained using a relatively small set of only three substructure descriptors and two physical property descriptors. For 70% of the CYP2D6 substrates, a known SOM was accurately predicted within the top two ranked sites. Zheng et al. used support vector machines to optimize a selected set of substrate quantum chemical features for six classes of CYP-catalyzed reactions.[12] Depending on the metabolic reaction, 82−94% of the sites were correctly classified as either a SOM or not a SOM. Rydberg et al. developed Smartcyp, a model that predicts intrinsic reactivity using fragment-based rules derived from quantum chemical DFT calculations of reference small molecules.[13,14] In combination with an accessibility descriptor, the model accurately predicted a

Received: May 20, 2011
Published: July 28, 2011

dx.doi.org/10.1021/jm2006468 | J. Med. Chem. 2011, 54, 6098–6105

known SOM within the top two ranked sites for 76% of the studied CYP3A4 substrates.

Structure-based models describe the interaction between a particular CYP protein and its substrate with atomic detail and do not depend on available metabolism data for SOM prediction. With respect to metabolism of drugs, CYP3A4 is the most important CYP, responsible for the metabolism of about 50% of known drugs.[3] CYP2D6 is comparatively more selective toward substrates, with a smaller active site cavity.[15] It is therefore expected that information from the CYP2D6 protein would be important for reliable prediction of the regioselectivity of drug metabolism. de Graaf et al. tested different docking strategies using a CYP2D6 homology model to predict experimentally reported SOMs.[16] More than 80% of the substrates were docked correctly (in this case "correct" means a SOM−heme iron distance of <6 Å for the best ranked docking pose). Important advances of their work are the rescoring of the combined docking poses from different docking programs and the inclusion of explicit water molecules into the active site. Protein flexibility, however, was not considered.

The use of molecular dynamics (MD) conformational sampling for predicting metabolic regioselectivities was pioneered by Ornstein and co-workers.[17] Hritz et al. used a combination of molecular dynamics (MD) simulation and automated docking to assess the impact of protein flexibility of CYP2D6 structures on docking accuracy.[18] Conformational sampling was based on MD simulation in the presence of representative substrates docked in the CYP2D6 active site, with each substrate−protein complex simulation producing a different protein ensemble. By use of a binary decision tree to select the most appropriate protein structure for docking, 80% of the substrates were correctly docked (using the same 6 Å cutoff for "correctness" as noted above for the results of de Graaf et al.[16]). The obvious limitation of this approach is that simulation in the presence of a bound ligand yields a protein conformation ensemble that is biased toward the specific ligand, and deciding which ensemble to use for a substrate with an unknown SOM is not straightforward.

Structure-based and reactivity-based models are complementary and can therefore be combined. Examples of such hybrid models can be found in the commercial programs MetaSite[19] and Stardrop (Optibrium Ltd.), which consider the CYP active site geometry to some extent. In Stardrop, simple orientation and steric hindrance factors are combined with a semiempirical quantum chemical model that estimates the intrinsic vulnerability of a potential metabolic site. The model parameters for each isoform are trained using the experimentally observed metabolites for a diverse set of compounds.

Various methods for measuring the quality of a prediction model have been described, each with its own strengths and weaknesses. Sheridan et al.[11] have discussed some "measures of goodness" and the challenge of comparing the results of different studies in this field. Receiver operating characteristic (ROC) analysis is often used as a suitable way to quantify the overall prediction performance of a model. A more practical measure for use in a typical medicinal chemistry context may be the ability to rank a SOM higher than the other sites of a molecule. Another important aspect in structure-based SOM prediction concerns how to identify the "correct" ligand binding pose. Usually, a relatively large SOM−heme iron cutoff distance (6 Å) is used as a "correctness" criterion.[13,16,18] The physical interpretation of this criterion is not clear, since these long cutoffs do not preclude the closer approach of several other non-SOM atoms to the heme Fe. While the longer cutoff may be useful in method development and method comparison, it will fall short in a real world role where we seek to guide medicinal chemistry, since in this context our objective is to predict a single SOM in a molecule or chemotype of interest.

To fully assess the role of the protein conformational ensemble used in docking, we studied the CYP2D6-mediated metabolism of a set of 54 known substrates[20] using extensive ensemble-based docking experiments. The protein ensemble was generated with tCONCOORD,[21] a method that generates a set of random protein conformations satisfying a number of distance constraints derived from a single ligand-free protein structure. The structural features and docking parameters that influence prediction of the SOM were investigated using a receiver operating characteristic (ROC) analysis. It was found that predictivity is highly dependent on the SOM−heme cutoff distance applied to the automated docking poses, the scoring function used in docking, and the number of protein structures in the ensemble. To test whether prediction quality could be further improved, the model was combined with Smartcyp[14]-predicted intrinsic reactivities and with Stardrop. The conclusions drawn from initial studies with the set of 54 CYP2D6 substrates were subsequently tested and validated against an independent set of 109 substrates.

METHODS

Protein Ensemble Generation. Protein simulations started from the X-ray conformation of apo-CYP2D6s (PDB code 2F9Q),[15] which is the only CYP2D6 crystal structure currently deposited at the Protein Databank. The missing loop (residues 42−51) and some missing side chains were added, and mutations (Asp230Leu, Arg231Leu, Met374Val) were reverted to the wild-type sequence using the Swiss-PdbViewer.[22] The program tCONCOORD[21] (version 1.0) was used to generate ensembles of protein conformations. Constraints were added to preserve planarity of the heme group and to keep the Cys443 SG atom aligned with the heme Fe atom. Ten tCONCOORD runs were performed, each run generating 100 protein structures. For each member of the resulting ensemble of 1000 protein structures, protonation states were reassigned with the Schrödinger utility Protassign and the structures were subsequently minimized with the Impref utility, allowing a maximum root-mean-square deviation (rmsd) of the atom displacement of 0.5 Å. After minimization, the largest pairwise heavy-atom rmsd of the 1000 conformations was as high as 7.1 Å.

CYP2D6 Substrate Preparation. An initial set of 54 known CYP2D6 substrates (the "small set") was taken from de Groot et al.[20] A few compounds from their study were omitted. Terfenadine has been shown to be hydroxylated primarily by CYP3A4.[23,24] The *R*-(−) enantiomer of mirtazapine is not significantly metabolized by CYP2D6.[25] Some SOMs were adapted to better reflect available CYP2D6 regioselectivity data: debrisoquine is metabolized primarily at the C4-position;[26] bufuralol is hydroxylated only at the C1′-position;[27] and cinnarazine undergoes para-hydroxylation at the cinnamyl ring only.[28] The list of substrates and substrate numbering can be found in Figure S1 (Supporting Information). The 3D structures with relevant protonation states and tautomeric states were generated with the Schrödinger program LigPrep. Nonoptimal ligand states were assigned a scoring penalty using the Schrödinger program Epik.[29]

Ensemble Docking. Automated docking was performed with version 5.5 of Glide.[30,31] For grid generation, the cubic inner and outer boxes had side lengths of 10 and 32 Å, respectively. All ligand states were docked into all protein structures, treating the ligands as fully flexible while keeping the protein structure fixed. Docking poses were evaluated using the Glide SP scoring function. For each ligand state per protein conformation, up to 10 poses with a docking score of <0 kcal mol^{-1} were

retained and rescored in place with Glide XP. Cases where a substrate could not be docked to a particular conformation were considered as nonessential to the ensemble of complexes and thus discarded. The OPLS-2005 force field was used for minimization and grid generation, while OPLS-2001 was used for docking.

Prediction Models. Potential SOMs are defined as carbon atoms or atom groups with at least one pendent hydrogen atom. Symmetry-related atoms or atom groups were treated as a single potential SOM. For each docking pose and for each potential SOM, the site−heme distance was calculated as the distance between the site-attached hydrogen atom closest to the heme Fe atom and the heme Fe atom. The rationale for this distance choice is that our initial studies established that the SOM prediction was slightly better than when the carbon site−heme Fe distance was used (results not shown). The predicted activity of a potential SOM was calculated as the weighted fraction of docking poses with a site−heme distance smaller than a given cutoff. In model A, all docking poses were given the same weight. For the other models, docking poses were weighted using a Boltzmann weight factor $\exp[-E/(kT)]$, where k is the Boltzmann constant and T is the temperature (300 K). In model B, E is the Glide SP docking score. In model C, E is the Glide XP score. In model D, E is the sum of the Glide XP score and a fraction of the local protein energy; 5% of the local protein energy gave the best overall prediction. The local protein energy was calculated using the OPLS-2005 force field as the energy of interaction of 32 residues that comprise the active site cavity (110, 112, 117−122, 213, 214, 216, 217, 220, 244, 247, 248, 297, 300−306, 308, 309, 370, 373, 374, 482−484) plus the heme group. Subsequently, the weighted fractions were normalized for each substrate. The predicted activities of model D were combined with the intrinsic reactivities calculated by Smartcyp, version 1.1. In model E, all sites with a Smartcyp predicted activation energy of reaction (E_a) of >77 kJ mol^{-1} were scaled by a factor 0.05 (E_a values ranged between 39.8 and 94.6 kJ mol^{-1}). In model F, in addition to model E, all sites with a predicted activity of zero were ranked according to the Smartcyp predicted E_a. This was done by assigning a very small activity value, inversely proportional to the E_a, to those sites. Finally, in model G, the predicted activities of model E were averaged with SOM predictions calculated by Stardrop, version 4.3, using a model F/Stardrop weight ratio of 10. With the small set, the only parameters that have been optimized are the local protein energy, the Smartcyp-based E_a cutoff value and scaling factor, and the Stardrop averaging weight ratio.

Prediction Quality Measurement. To assess the overall performance of the prediction models, we used ROC analysis.[32] The area under the ROC curve (AUC) was calculated as a measure of the quality of the SOM prediction model. The AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance. The AUC value ranges between 0.5 (random classifier, no predictive value) and 1.0 (ideal classifier, perfect prediction). Ensemble-based ROC curves and AUC values were calculated after ranking the predicted activities of all potential SOMs of all substrates.

Substrate-based predictions were analyzed by means of a z-score,

$$\frac{\langle A\rangle_{exp} - \langle A\rangle_{pot}}{\sigma(A)}\left(\frac{N_{exp}}{N_{pot} - N_{exp}}\right)^{1/2} \quad (1)$$

where $\langle A\rangle_{exp}$ and $\langle A\rangle_{pot}$ are the average predicted activities of the experimentally observed and potential SOMs, N_{exp} and N_{pot} are the number of experimental and potential SOMs of a substrate, respectively, and σ is the standard deviation. The z-score ranges between −1 and 1. This z-score is useful for comparing the prediction quality of individual substrates.

RESULTS

In this study, only SOMs that involve hydroxylation at an aliphatic or aromatic carbon site were considered. N- and O-dealkylation reactions occur via hydroxylation of the carbon adjacent to the heteroatom.[33] Although the reaction mechanisms differ, both aliphatic and aromatic hydroxylations involve insertion of an oxygen atom from the iron(IV) oxoporphyrin radical cation into a carbon−hydrogen bond,[33] requiring the substrate to bind with a short distance between the SOM and the heme Fe. Thus, a SOM−heme distance cutoff can be used to predict whether or not a given substrate C atom is a SOM or not.

Seven different prediction models (models A−G, see Methods) were compared to investigate the influence of docking score weighting, rescoring, internal protein energy weighting, combination with predicted intrinsic reactivity, and combination with Stardrop. In Figure 1, the performance of the different models was compared. It is noteworthy that for all models tested SOM prediction is highly dependent on the site−heme cutoff distance. We began our study by evaluating the performance of the pure ensemble-based docking models (models A−D). Weighting of the docking poses according to the Glide SP (model B) and Glide XP (model C) scores improved the prediction results. Inclusion of local protein internal energy into the weight factor offered a further small improvement in SOM predictivity (model D), with a maximum AUC (AUC_{max}) value of 0.893 at a cutoff distance of 2.8 Å (Table 1). As a reference, we evaluated atomic accessible surface areas (ASA)[34] as a measure of the probability of proximity to the heme Fe when the substrate is oriented randomly. ASA-based SOM prediction gave an AUC of 0.676. For comparison, when all atoms were given equal probability to be near the heme Fe, the AUC was 0.516, close to a fully random prediction. A theoretical perfect prediction (AUC = 1) maximally increases the AUC of equal atomic probabilities by 0.484 (100%). Thus, the ASA-based prediction accounts for 33% while the pure ensemble-based docking model D accounts for 78% of this maximal AUC increase.

Next, we investigated the impact of the intrinsic reactivity data from the Smartcyp program[14] on the prediction results (models E and F). Scaling down the predicted activity of all sites with high predicted E_a (model E) increased the AUC value substantially. Slight improvements at short SOM−heme cutoff distances were observed after ranking the sites with a predicted activity of zero (sites that were not docked closer to the heme Fe than the cutoff in any docking pose) according to their intrinsic E_a (model F). Smartcyp-only data provided essentially no predictive value for CYP2D6 (AUC $\approx$ 0.5). The predictions could be further improved by weighted averaging of the activity predictions with Stardrop predictions (model G). Performance was significantly lower when Stardrop was used alone (AUC = 0.84).

Figure 2 shows how the number of protein structures affects the AUC predictions. With an increasing number of structures, AUC values increase and become less dependent on the SOM−heme cutoff distance; in both cases the differences are relatively small. The AUC_{max} appears converged at 900 structures (Figure 2).

For practicing medicinal chemists, it is important to know which sites are most likely to be metabolized. Table 1 shows the fraction of substrates that contain a SOM within the top two ranked sites (f_{top2}). The values we obtained with our new approach compare favorably with previous reports, e.g., by Sheridan et al.,[11] who compared f_{top2} of their QSAR-based model

(0.70−0.72) with that obtained with the method of Singh et al.[10] (0.24−0.50) and with Metasite[19] (0.64−0.70). We note that in our study the substrate sets consider only sp^2 and sp^3 carbons as SOM types while the data reported for CYP2D6 by Sheridan et al.[11] take into account a small additional number of sulfur, phosphorus, and nitrogen atoms.

Single Structure Analysis. To test whether a single protein structure could be representative for the ensemble as a whole, we performed the ROC analysis on all protein structures separately. Additionally, single structure analysis provides insight into the general structural features that determine the docking pose of the substrate.

By use of the full protein ensemble, the optimal cutoff value (model F) was 2.7 Å. The AUC decreased slowly with the cutoff distance up to ∼6 Å, while a sharp decline was seen at greater distances (Figure 3). For most single structures, however, the optimal cutoff distance was greater than 5 Å and the observed AUC_{max} values were much lower. The average AUC value was highest at a cutoff distance of 5.0 Å (Figure 3), while at 2.7 Å cutoff distance, AUC values ranged between 0.52 and 0.80. Steric effects and/or strong protein−substrate interactions often keep the substrate from closely approaching the reactive center, even if the SOM is closest to heme Fe, as illustrated for (*S*)-propafenone in Figure 4a. Remarkably, the performance of the best single protein structure (AUC_{max} = 0.904; f_{top2} = 0.78) was close to that of the optimal ensemble-based approach (AUC_{max} = 0.920; f_{top2} = 0.84) but at a much larger cutoff distance (7.0 Å). The X-ray structure on its own, with an AUC_{max} of 0.83 at a 6.5 Å, is clearly not well suited for SOM prediction (Figure 3).

We applied the functional mode analysis[35] algorithm to determine the motion of the active site cavity that is maximally correlated with the single-structure AUC (Figure 5). In structures that correspond to increased AUC the active site cavity opens up, with increasing distances between helix I, the loop between helix K and sheet β_{1-4}, and the Phe120 side chain; concomitantly, relative to the cavity, the heme plane tilts/moves away from helix I and toward the K−β_{1-4} loop.

Single Substrate Analysis. To estimate the prediction quality of individual substrates, we calculated *z*-scores (see Methods) for single substrates using a distance cutoff of 2.7 Å. For a given

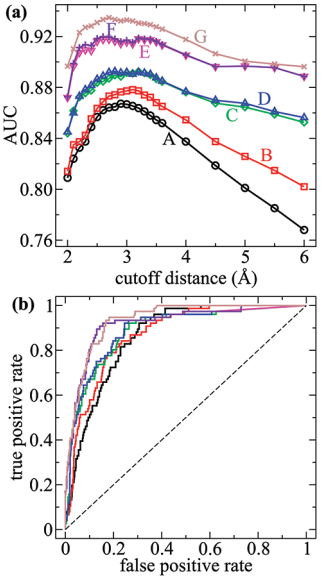


Figure 1. AUC as a function of cutoff distance for the tested prediction models (a) and corresponding ROC curves for a 2.7 Å cutoff distance (b).

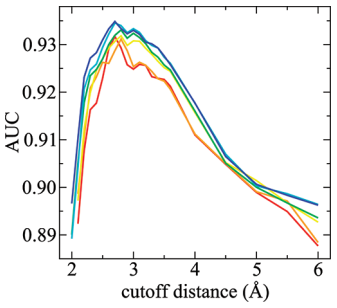


Figure 2. AUC as a function of cutoff distance (model G). Increasing number of protein structures are shown: 100 (red), 200 (orange), 400 (yellow), 700 (green), 900 (cyan), 1000 (blue).

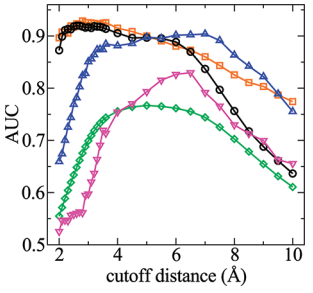


Figure 3. AUC as a function of cutoff distance (model F) using the full ensemble of the small set (black) and large set (orange). Also shown are the following: the small set averaged over all single structure values (green); highest single-structure AUC values (blue); AUC of the X-ray structure (magenta).

Table 1. Full Ensemble-Based SOM Prediction Results[a]

	model D	model F	
substrate set	AUC_{max} (d_{max})	AUC_{max} (d_{max})	f_{top2} (d_{max})
small set	0.89 ± 0.04 (2.8)	0.92 ± 0.03 (2.6)	0.85 (2.6)
large set	0.88 ± 0.03 (3.6)	0.93 ± 0.02 (2.8)	0.88 (2.7)
large set minus α-N-hydroxylation	0.90 ± 0.03 (3.6)	0.93 ± 0.03 (2.8)	0.90 (2.7)

[a] AUC_{max}, maximum AUC; f_{top2}, fraction of substrates that contain an experimentally observed SOM within the top two ranked sites; d_{max}, SOM−heme cutoff distance (Å) corresponding to AUC_{max} or f_{top2}. Margins of error are given at the 95% confidence level. We note that models D and F are highly correlated, as are the large sets with and without the α-N-hydroxylation SOMs. Therefore, the statistical significance of their different AUC_{max} values is not limited by the given margins of error.

dx.doi.org/10.1021/jm2006468 |*J. Med. Chem.* 2011, 54, 6098–6105

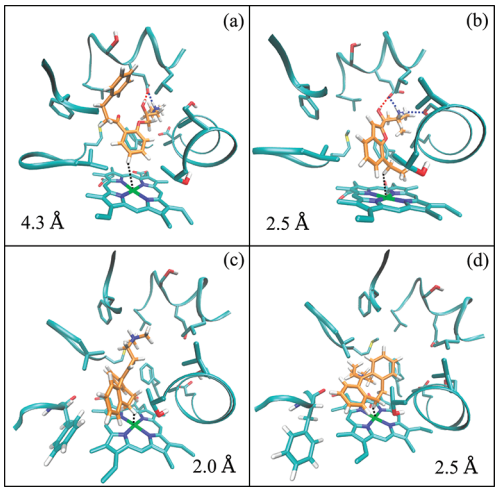


Figure 4. Example substrates docked into the CYP2D6 active site cavity. The noted SOM—heme distances are shown as black dotted lines, and H-bonds are shown as blue or red dotted lines. (a) (*S*)-Propafenone. Although in this binding mode the (*S*)-propafenone SOM is closest to the heme Fe, H-bond interactions with Glu216 and steric interactions with the cavity residues prevent docking of the SOM closer to the heme iron. (b) High *z*-score binding mode of (*S*)-bufuralol. H-Bonds with Glu216 and Ser304 are indicated. (c, d) Two alternative high *z*-score binding modes of amitriptyline.

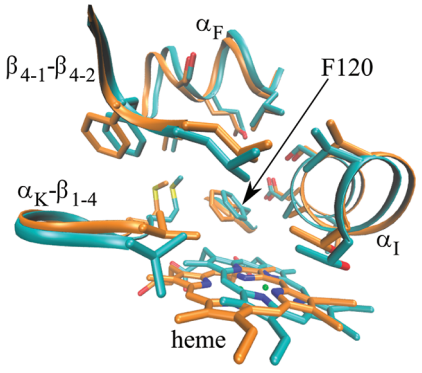


Figure 5. Functional mode analysis. Superposition of two extreme conformations along the cavity motion that maximally correlates with the single-structure AUC values (model F) for a 2.7 Å cutoff distance. Some important residues that form the active site are shown. Low AUC is in cyan, and high AUC is in orange. Rc values for model building (800 structures) and cross-validation (200 structures) were 0.75 and 0.71, respectively.

substrate, a positive *z*-score means that, on average, SOMs are ranked higher than non-SOMs. Most substrates (43 out of 54) have a positive ensemble-based *z*-score (Figure 6).

Substrates with a high ensemble-based *z*-score and a low fraction of positive single-structure *z*-scores are indicative of strong protein conformational selection. The high *z*-score complexes formed by a given substrate and a small fraction of protein structures are energetically more stable than the low *z*-score complexes formed by the majority of structures. For example, (*S*)-bufuralol (substrate 8) has a positive *z*-score for only 64 out of 1000 structures while its ensemble-based *z*-score (0.57) is quite high (Figure 4b). Thus, (*S*)-bufuralol selects a small subset of protein conformations to bind preferentially in a pose that is

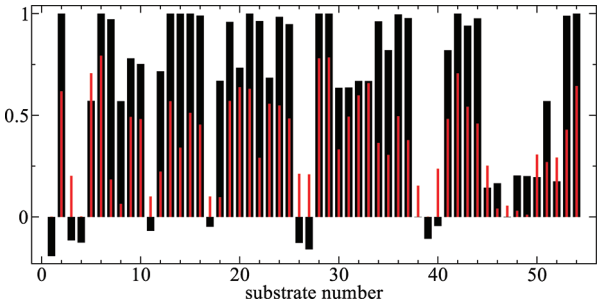


Figure 6. Prediction quality comparison of the individual substrates (model F): *z*-score based on the full ensemble (black); fraction of protein structures with positive *z*-score (red).

suitable for oxidation at its SOM and that greatly stabilizes the complex.

Substrates with both a low ensemble-based *z*-score and a low fraction of positive single-structure *z*-scores are indicative of conformational sampling deficiencies, with either tCONCOORD or the docking approach or with both used in this study. For example, only two protein conformations had a positive *z*-score for amitriptyline (substrate 4). Although the amitriptyline binding mode is very different for both protein conformations (Figure 4c, d), the two protein conformations are very similar to each other while at the same time both are quite exceptional with respect to the complete conformational ensemble. Both conformations have very large active site cavities, which is necessary for the amitriptyline SOM to closely approach the active site heme Fe. These two structures are also unique in that the K″—L loop is positioned on the distal side of the heme plane rather than the much more typical proximal side. Similar results were observed for the analogue nortryptiline (substrate 39).

Validation against an External Set. To further validate our approach, an independent set of 109 substrates was collected from the literature data, which we will refer to as the "large set". The large set contains 1129 potential SOMs, 122 of which are experimentally observed SOMs. Of these observed SOMs, 26 are at sp^2 carbons and 94 are at sp^3 carbons (Figure S2). In contrast to the small set, which contains only SOMs that undergo hydroxylation adjacent to an oxygen atom, the large set also contains 31 SOMs that involve hydroxylation in a position α to a basic N atom. The structural diversity of this set of substrates was established using ChemmineR.[36] By application of a similarity cutoff of 0.7, the 109 small molecules formed 87 clusters, 70 of which have a single member (singletons), with the largest cluster having only 4 members. For the purpose of estimating the structural similarity between the small and large sets, the two sets were combined and analyzed. When the same method was applied, the combined set formed 112 clusters, only 6 of which had members from both sets. Thus, the small and large sets of substrates contain distinct and nonoverlapping chemotypes.

With the large set, the dependence of AUC_{max} on the cutoff distance is similar to that of the small set, albeit with a more gradual performance decline with increasing cutoff distance (Figure 3). Compared to the small set, the performance of the large set is better with model F at essentially the same optimal cutoff distance but not as good with model D (Table 1). Discarding the SOMs that undergo α-N-hydroxylation allows for a more direct comparison with the small set and allows evaluation of the performance with respect to α-N-hydroxylation SOMs. With the α-N-hydroxylation SOMs discarded from the

large set, the AUC_{max} fully recovers to that observed for the small set (Table 1). The Smartcyp-predicted E_a of sites next to amine nitrogens is relatively low,[14] compensating for the reduced probability of finding these SOMs close to the heme Fe (in comparison to other SOMs).

DISCUSSION

The results of the present work are consistent with the hypothesis that the protein environment is generally the most important determinant of which site on a substrate molecule is preferentially metabolized by CYP2D6.

With strictly structure-based models A−D, based on ensembles of CYP2D6 conformations that are not biased by the presence of a bound ligand during conformer generation, our ability to predict known SOMs is quite good. The difference between the results obtained with the Glide SP (model B) and Glide XP (model C) scoring functions is fully consistent with the deep and hydrophobic CYP2D6 active site cavity and the extra "hydrophobic enclosure" term in the Glide XP function.[30]

SOM prediction was improved substantially by simply scaling down the activity of sites with a high Smartcyp-predicted E_a. Thus, for highly and moderately labile substrate sites, the likelihood of being metabolized depends largely on access to the reactive Fe, whereas for relatively stable sites the intrinsic reactivity is more important. In contrast, sites next to a basic nitrogen seem to be much less dependent on the nonbonded substrate−protein interaction, even with low predicted E_a. The basic nitrogen strongly interacts with the acidic residues Glu216 and Asp301,[37] disfavoring orientation of the nitrogen toward the catalytic center. Our results suggest that reaction at a position α to a basic nitrogen occurs by CYP2D6 only when more accessible sites are too stable to react. This is consistent with the results of Bonn et al., who aligned a set of compounds to reference substrates and concluded that N-dealkylation occurs when the preferred site (the site that overlaps with the SOM of the reference) is blocked.[38]

In the present study we evaluated the ensemble-based SOM−heme Fe cutoff distance and determined that the optimal cutoff distance was very short (~2.7 Å for H to Fe). This is contrary to the 6 Å cutoff most typically encountered in studies of this type.[13,16,18] The consensus catalytic cycle for hydroxylation of C−H groups involves a two-step hydrogen abstraction/oxygen rebound mechanism.[39] Considering the ferryl oxoporphyrin cation radical (also called[40] compound **I**) as the reactive heme species, 2.7 Å is within reacting distance and thus is highly chemically relevant, as it is fully consistent with the atom transfer mechanism (the Fe−O bond length[41] for compound **I** is 1.65 Å). The strong dependence of AUC on the cutoff distance suggests that binding pose selectivity (and hence regioselectivity) occurs predominantly at a mechanistically relevant distance.

Recent CYP docking studies used a much longer cutoff distance of 6 Å between the substrate site (C atoms) and the heme Fe atom (corresponds to H-to-Fe distance of ~5 Å) to discriminate between correctly and incorrectly docked complexes.[13,16,18] However, the physical basis for this long cutoff distance is unclear. For example, suppose that for a given docking pose a given SOM is within a 5 Å cutoff, positioned 4.9 Å from the heme Fe atom. At this distance, it is quite possible that a number of other relevant sites of the substrate (atoms that are theoretical SOMs but are not known SOMs) are even closer to the heme Fe than the experimentally known SOM. To better understand how the SOM prediction is related to the proximity of the substrate to the heme Fe, we have applied an alternative distance criterion (Figure 7). Rather than the commonly employed single-sided maximum cutoff distance, whereby a docking pose contributes to the predicted activity of all potential SOMs *(known or theoretically possible SOMs)* that are positioned below the cutoff distance, the alternative criterion uses a *cutoff distance window* of 1 Å, whereby a docking pose contributes to the predicted activity of a potential SOM only if this potential SOM is the atom closest to the heme Fe and within the cutoff window (Figure 7). At distances below 3.2 Å (e.g., between 2.2 and 3.2 Å), the alternative criterion performs as well as the single-sided criterion, reflecting the low probability of finding more than one site (as defined above) within the short cutoff distance. At longer distances, however, the AUC of the alternative criterion degrades more sharply with increasing cutoff distance, in comparison to the standard cutoff procedure (Figure 7). This alternative analysis shows that the 1 Å cutoff windows involving longer distances contain an increased number of "incorrect" docking poses (poses where the experimentally known SOM meets the commonly applied cutoff criterion but some other atom in the molecule is actually closer to the Fe atom than the known SOM). Thus, the dependence of prediction quality on proximity to the heme Fe is much stronger than the commonly employed single-sided distance criterion might suggest. As has been observed in virtual screening studies, highly ranked binding poses of known actives often do not resemble known crystal structures, even if they are well-prioritized by the docking algorithm.[42] In the case of CYP2D6, the use of a short cutoff distance in combination with ensemble-based docking helps to reduce the fraction of binding modes for which the orientation of the substrate is inconsistent with the chemical mechanism of CYP2D6 metabolism.

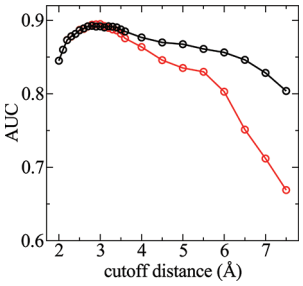


Figure 7. Influence of the distance criterion on SOM prediction (model D). (Black) Commonly employed single-sided maximum cutoff distance, whereby the distance between the potential SOM and the heme Fe atom must be less than the indicated cutoff distance. (Red) Cutoff distance window, whereby a potential SOM must be the atom closest to the heme Fe and located within the distance range "cutoff to (cutoff − 1)".

Can a single protein structure replace the full protein ensemble? In view of the AUC_{max} score of the best performing single protein conformation with the small set (0.904; cutoff distance of 7.0 Å), it is tempting to assume that this is the case. However, our work here is purely retrospective in that the small molecules studied here have been fully characterized with experimental SOM studies, and we have used that information to evaluate the performance of our computational approach. In a real-world prospective setting, one will not know a priori which protein structure and cutoff distance to use simply because the experimental SOM data will not be available (to allow evaluation of SOM prediction performance). For example, using the "best" performing single CYP2D6 conformer and cutoff distance noted

dx.doi.org/10.1021/jm2006468 |J. Med. Chem. 2011, 54, 6098–6105

above, the AUC score of the large set was 0.815, lower than the 0.904 observed with the small set. This does not mean that a well chosen and balanced reduced set of structures might not accurately represent the larger ensemble of protein states if some degree of protein flexibility is included in the docking. However, at this juncture the appropriate methods for both protein conformation selection and flexible docking remain to be determined. Clearly, these are both important topics for future research in this field.

In automated docking studies based on a single CYP2D6 homology model, de Graaf et al. showed that the inclusion of active-site water molecules strongly improves SOM predictions for this enzyme.[16] In a subsequent study, however, the explicit water molecules were observed to be quite mobile, their role in the binding mode being highly substrate-dependent and protein-conformation-dependent.[43] Given the importance of protein flexibility, as shown in the present work and also in that of Hritz et al.,[18] it appears likely that active-site water molecules partially mimic protein dynamics and/or vice versa in single-structure docking. The inclusion of water molecules in our protocol might further reduce the number of distinct protein structures required for optimal SOM prediction.

We have developed an automated docking-based method of SOM prediction for CYP2D6 that advances the field in the areas of protein structure ensemble generation and distance cutoff while delivering excellent SOM prediction results. With our new approach, reliable SOM predictions are possible for a diverse array of CYP2D6 substrates. The use of a physically and chemically relevant cutoff distance criterion to identify the "correct" binding mode represents an advance in terms of overall prediction quality, mechanistic interpretation, and utility as a medicinal chemistry tool. Omission of a bound ligand during generation of the protein structure ensemble should lead to relatively unbiased sets of protein structures for use in docking. The tCONCOORD method is a revised implementation of the original CONCOORD method,[44] with tCONCOORD designed to extend conformational sampling beyond the vicinity of the reference structure. Structure generation with tCONCOORD is relatively simple and fast. Previous work has established that, while CONCOORD and normal-mode analysis techniques yielded relatively limited conformational sampling, tCONCOORD generates much more diverse ensembles, comparable to those observed with conventional MD simulation.[45] Other work has shown that tCONCOORD reliably reproduces the large conformational transitions seen in sets of X-ray crystal structures and NMR ensembles.[21] Be that as it may, the sampling method in our new SOM prediction procedure could be further improved by using more rigorous advanced structure generating methods, and we are exploring replica exchange MD[46,47] in this context. The ensemble of structures needs to be generated only once, which justifies performing long time scale simulations. Advanced ensemble docking algorithms, which simultaneously dock the ligand into the full ensemble,[48] could further increase the computational efficiency of our approach. A variety of performance criteria persist in the field of SOM prediction, posing an additional challenge to method development and advancement, and a diversity of perspectives in this regard could also be beneficial to the further refinement of our approach. Given that very good predictions were possible with the constraints-based tCONCOORD method, the combined ensemble docking−intrinsic reactivity strategy appears quite promising, and we expect that ongoing work will further advance the approach.

■ ASSOCIATED CONTENT

Supporting Information. Figures S1 and S2 containing the small set (with numbering) and the large set with indicated experimental SOMs and first and second ranked predicted potential SOMs. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +32 16 32 73 84. Fax: +32 16 79 92. E-mail: sam.moors@chem.kuleuven.be.

■ ACKNOWLEDGMENT

We thank Mark Seierstad (Johnson & Johnson Pharmaceutical Research & Development, La Jolla, CA, U.S.) for running the Stardrop calculations, and Robert Sheridan (Merck Research Laboratories, Rahway, NJ, U.S.) for insightful comments on the manuscript. This work was financially supported by the Flemish agency for Innovation by Science and Technology (IWT).

■ ABBREVIATIONS USED

AUC, area under curve; CYP, cytochrome P450; CYP2D6, cytochrome P450 2D6; MD, molecular dynamics; ROC, receiver operating characteristic; SOM, site of metabolism

■ REFERENCES

(1) Brown, C. M.; Reisfeld, B.; Mayeno, A. N. Cytochromes p450: a structure-based summary of biotransformations using representative substrates. *Drug Metab. Rev.* **2008**, *40*, 1–100.

(2) Rendic, S.; Guengerich, F. P. Update information on drug metabolism systems—2009, part II. Summary of information on the effects of diseases and environmental factors on human cytochrome P450 (CYP) enzymes and transporters. *Curr. Drug Metab.* **2010**, *11*, 4–84.

(3) *Cytochrome P450 Structure, Mechanism, and Biochemistry*; Kluwer Academic/Plenum Publishers: New York, 2005.

(4) Neafsey, P.; Ginsberg, G.; Hattis, D.; Sonawane, B. Genetic polymorphism in cytochrome P450 2D6 (CYP2D6): population distribution of CYP2D6 activity. *J. Toxicol. Environ. Health, Part B* **2009**, *12*, 334–361.

(5) De Gregori, M.; Allegri, M.; De Gregori, S.; Garbin, G.; Tinelli, C.; Regazzi, M.; Govoni, S.; Ranzani, G. N. How and why to screen for CYP2D6 interindividual variability in patients under pharmacological treatments. *Curr. Drug Metab.* **2010**, *11*, 276–282.

(6) Baranczewski, P.; Stanczak, A.; Kautiainen, A.; Sandin, P.; Edlund, P. Introduction to early in vitro identification of metabolites of new chemical entities in drug discovery and development. *Pharmacol. Rep.* **2006**, *58*, 341–352.

(7) Brown, D. Unfinished business: target-based drug discovery. *Drug Discovery Today* **2007**, *12*, 1007–1012.

(8) Nedderman, A. N. R. Metabolites in safety testing: metabolite identification strategies in discovery and development. *Biopharm. Drug Dispos.* **2009**, *30*, 153–162.

(9) Stjernschantz, E.; Vermeulen, N. P. E.; Oostenbrink, C. Computational prediction of drug binding and rationalisation of selectivity towards cytochromes P450. *Expert Opin. Drug Metab. Toxicol.* **2008**, *4*, 513–527.

(10) Singh, S. B.; Shen, L. Q.; Walker, M. J.; Sheridan, R. P. A model for predicting likely sites of CYP3A4-mediated metabolism on drug-like molecules. *J. Med. Chem.* **2003**, *46*, 1330–1336.

(11) Sheridan, R. P.; Korzekwa, K. R.; Torres, R. A.; Walker, M. J. Empirical regioselectivity models for human cytochromes p450 3A4, 2D6, and 2C9. *J. Med. Chem.* **2007**, *50*, 3173–3184.

(12) Zheng, M. Y.; Luo, X. M.; Shen, Q. C.; Wang, Y.; Du, Y.; Zhu, W. L.; Jiang, H. L. Site of metabolism prediction for six biotransformations mediated by cytochromes P450. *Bioinformatics* **2009**, *25*, 1251–1258.

(13) Rydberg, P.; Vasanthanathan, P.; Oostenbrink, C.; Olsen, L. Fast prediction of cytochrome P450 mediated drug metabolism. *ChemMedChem* **2009**, *4*, 2070–2079.

(14) Rydberg, P.; Gloriam, D. E.; Zaretzki, J.; Breneman, C.; Olsen, L. SMARTCyp: a 2D method for prediction of cytochrome P450-mediated drug metabolism. *ACS Med. Chem. Lett.* **2010**, *1*, 96–100.

(15) 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.

(16) de Graaf, C.; Oostenbrink, C.; Keizers, P. H. J.; van der Wijst, T.; Jongejan, A.; Vermeulen, N. P. E. Catalytic site prediction and virtual screening of cytochrome P450 2D6 substrates by consideration of water and rescoring in automated docking. *J. Med. Chem.* **2006**, *49*, 2417–2430.

(17) Loida, P. J.; Sligar, S. G.; Paulsen, M. D.; Arnold, G. E.; Ornstein, R. L. Stereoselective hydroxylation of norcamphor by cytochrome P450cam. Experimental verification of molecular dynamics simulations. *J. Biol. Chem.* **1995**, *270*, 5326–5330.

(18) Hritz, J.; de Ruiter, A.; Oostenbrink, C. Impact of plasticity and flexibility on docking results for cytochrome P450 2D6: a combined approach of molecular dynamics and ligand docking. *J. Med. Chem.* **2008**, *51*, 7469–7477.

(19) Cruciani, G.; Carosati, E.; De Boeck, B.; Ethirajulu, K.; Mackie, C.; Howe, T.; Vianello, R. MetaSite: understanding metabolism in human cytochromes from the perspective of the chemist. *J. Med. Chem.* **2005**, *48*, 6970–6979.

(20) de Groot, M. J.; Ackland, M. J.; Horne, V. A.; Alex, A. A.; Jones, B. C. Novel approach to predicting P450-mediated drug metabolism: development of a combined protein and pharmacophore model for CYP2D6. *J. Med. Chem.* **1999**, *42*, 1515–1524.

(21) Seeliger, D.; Haas, J.; de Groot, B. L. Geometry-based sampling of conformational transitions in proteins. *Structure* **2007**, *15*, 1482–1492.

(22) Guex, N.; Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **1997**, *18*, 2714–2723.

(23) Jones, B. C.; Hyland, R.; Ackland, M.; Tyman, C. A.; Smith, D. A. Interaction of terfenadine and its primary metabolites with cytochrome P450 2D6. *Drug Metab. Dispos.* **1998**, *26*, 875–882.

(24) Ling, K. H. J.; Leeson, G. A.; Burmaster, S. D.; Hook, R. H.; Reith, M. K.; Cheng, L. K. Metabolism of terfenadine associated with CYP3A(4) activity in human hepatic microsomes. *Drug Metab. Dispos.* **1995**, *23*, 631–636.

(25) Dodd, S.; Boulton, D.; Burrows, G.; De, V. C.; Norman, T. In vitro metabolism of mirtazapine enantiomers by human cytochrome P450 enzymes. *Hum. Psychopharmacol.* **2001**, *16*, 541–544.

(26) Lightfoot, T.; Ellis, S. W.; Mahling, J.; Ackland, M. J.; Blaney, F. E.; Bijloo, G. J.; De Groot, M. J.; Vermeulen, N. P. E.; Blackburn, G. M.; Lennard, M. S.; Tucker, G. T. Regioselective hydroxylation of debrisoquine by cytochrome P4502D6: implications for active site modelling. *Xenobiotica* **2000**, *30*, 219–233.

(27) Yamazaki, H.; Guo, Z. Y.; Persmark, M.; Mimura, M.; Inoue, K.; Guengerich, F. P.; Shimada, T. Bufuralol hydroxylation by cytochrome P450 2D6 and 1A2 enzymes in human liver microsomes. *Mol. Pharmacol.* **1994**, *46*, 568–577.

(28) Kariya, S.; Isozaki, S.; Uchino, K.; Suzuki, T.; Narimatsu, S. Oxidative metabolism of flunarizine and cinnarizine by microsomes from B-lymphoblastoid cell lines expressing human cytochrome P450 enzymes. *Biol. Pharm. Bull.* **1996**, *19*, 1511–1514.

(29) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. Epik: a software program for pKa prediction and protonation state generation for drug-like molecules. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 681–691.

(30) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.

(31) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.

(32) Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874.

(33) Meunier, B.; de Visser, S. P.; Shaik, S. Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes. *Chem. Rev.* **2004**, *104*, 3947–3980.

(34) Shrake, A.; Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* **1973**, *79*, 351–371.

(35) Hub, J. S.; de Groot, B. L. Detection of functional modes in protein dynamics. *PLoS Comput. Biol.* **2009**, *5*, No. e1000480.

(36) Cao, Y.; Charisi, A.; Cheng, L. C.; Jiang, T.; Girke, T. ChemmineR: a compound mining framework for R. *Bioinformatics* **2008**, *24*, 1733–1734.

(37) Paine, M. J. I.; McLaughlin, L. A.; Flanagan, J. U.; Kemp, C. A.; Sutcliffe, M. J.; Roberts, G. C. K.; Wolf, C. R. Residues glutamate 216 and aspartate 301 are key determinants of substrate specificity and product regioselectivity in cytochrome P450 2D6. *J. Biol. Chem.* **2003**, *278* 4021–4027.

(38) Bonn, B.; Masimirembwa, C. M.; Aristei, Y.; Zamora, I. Molecular basis of CYP2D6-mediated N-dealkylation: balance between metabolic clearance routes and enzyme inhibition. *Drug Metab. Dispos.* **2008**, *36*, 2199–2210.

(39) Guengerich, F. P. Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem. Res. Toxicol.* **2001**, *14*, 611–650.

(40) Munro, A. W.; Girvan, H. M.; McLean, K. J. Variations on a (t)heme—novel mechanisms, redox partners and catalytic functions in the cytochrome P450 superfamily. *Nat. Prod. Rep.* **2007**, *24*, 585–609.

(41) Bathelt, C. M.; Zurek, J.; Mulholland, A. J.; Harvey, J. N. Electronic structure of compound I in human isoforms of cytochrome P450 from QM/MM modeling. *J. Am. Chem. Soc.* **2005**, *127*, 12900–12908.

(42) Cummings, M. D.; DesJarlais, R. L.; Gibbs, A. C.; Mohan, V.; Jaeger, E. P. Comparison of automated docking programs as virtual screening tools. *J. Med. Chem.* **2005**, *48*, 962–976.

(43) Santos, R.; Hritz, J.; Oostenbrink, C. Role of water in molecular docking simulations of cytochrome P450 2D6. *J. Chem. Inf. Model.* **2010**, *50*, 146–154.

(44) de Groot, B. L.; vanAalten, D. M. F.; Scheek, R. M.; Amadei, A.; Vriend, G.; Berendsen, H. J. C. Prediction of protein conformational freedom from distance constraints. *Proteins* **1997**, *29*, 240–251.

(45) Eyrisch, S.; Helms, V. What induces pocket openings on protein surface patches involved in protein−protein interactions? *J. Comput.-Aided Mol. Des.* **2009**, *23*, 73–86.

(46) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(47) Moors, S. L. C.; Michielssens, S.; Ceulemans, A. Improved replica exchange method for native-state protein sampling. *J. Chem. Theory Comput.* **2011**, *7*, 231–237.

(48) Huang, S. Y.; Zou, X. Q. Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. *Proteins* **2007**, *66*, 399–421.