

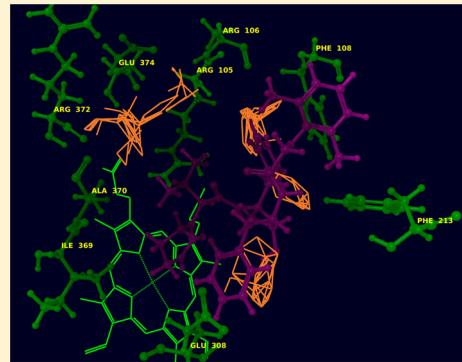
# Elucidating Substrate Promiscuity in the Human Cytochrome 3A4

Christina Hayes, Daniel Ansbro, and Maria Kontoyianni\*

Department of Pharmaceutical Sciences, School of Pharmacy, Southern Illinois University Edwardsville, Edwardsville, Illinois 62034, United States

## Supporting Information

**ABSTRACT:** The human cytochrome P450 enzymes (CYPs) are heme-protein monooxygenases, which catalyze oxidative reactions of a broad spectrum of substrates. Consequently, they play a critical role in the metabolism of xenobiotics, such as drugs and carcinogens, and the catabolism of endogenous lipophilic factors. Bioavailability and toxicity, both of which can be related to CYPs, continue to pose problems in the development of new drugs. The isoform which metabolizes over one-third of drugs, CYP 3A4, was investigated employing ensemble-docking experiments of a 195-substrate library with induced fit and GOLD docking algorithms and a number of scoring functions. Enzyme conformations included three currently available CYP 3A4 crystal structures. All docking experiments were performed in duplicates with and without inclusion of crystallographic waters. Resultant poses were assessed based on accuracy of site of metabolism prediction. Analyses of the docked solutions pertaining to ranking efficacy, ligand molecular properties, stabilizing residues in the ligand–enzyme complexes, and metabolic reactions are discussed. Our analyses suggest that certain residues make favorable interactions with the bound substrates. Employing multiple receptor conformations enhances the accuracy of catalytic prediction, while ligand size and flexibility impact docking performance. The presence of waters observed in crystal complexes does not necessarily lead to improved performance.



## INTRODUCTION

The cytochrome P450 (CYPs) monooxygenases carry out oxidations and are thus responsible for the biotransformation of most drugs and carcinogens. Recent reports on failures in phases II and III for the periods 2008–2010 and 2007–2010, respectively, indicated that safety and efficacy continue to be challenging, while only 1% of failures were attributed to pharmacokinetics and bioavailability.<sup>1,2</sup> Even though metabolism is seemingly linked to just pharmacokinetics and bioavailability at first, safety and efficacy are also impacted by metabolic liability. For example, bioactivation of a drug may result in intermediates, which could covalently bind to proteins and in turn lead to toxicity and reduced safety.<sup>3–5</sup> Also, individuals deficient in a particular P450 will not be able to metabolize a given dose of a drug, thus causing an exacerbated pharmacological response and subsequent unwanted effects.<sup>4</sup> When expression and activity levels of CYPs are elevated by induction because of the activation of key transcription factors, biotransformation of a dose of a drug will be accelerated leading to lower pharmacological efficacy.<sup>6</sup> Similarly, concurrent administration of certain drugs, with one being a CYP inhibitor, may result in slower metabolism of the other, provided that they bind to the same site of the isozyme. In turn, slower metabolism of an administered drug can lead to toxicity. Nevertheless, inhibition is desirable when drug biotransformation leads to toxic or carcinogenic metabolites.<sup>7</sup>

Fifty-seven P450 isozymes exist in humans,<sup>8,9</sup> while eight of these enzymes are responsible for the metabolism of most clinical drugs to date.<sup>10,11</sup> The most abundant CYP in the

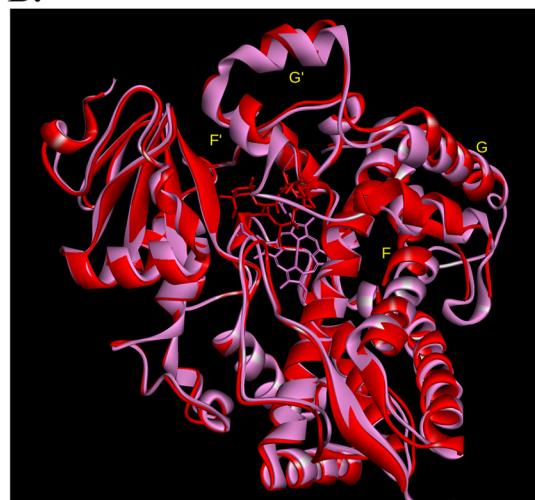
human liver is 3A4, which is responsible for the metabolism of more than one-third of the currently clinically administered drugs and many endogenous compounds.<sup>12–16</sup> Consequently, given the breadth of its substrates, the potential for drug–drug interactions is high. Therefore, understanding the factors driving substrate selectivity is critical, especially to the pharmaceutical industry where screening for CYP 3A4 inhibition has become a filter in early drug discovery.

Several crystal structures of CYP 3A4 have been solved; recent ones are complexed with erythromycin (PDB accession code 2J0D),<sup>17</sup> ketoconazole (PDB accession code 2V0M),<sup>17</sup> ritonavir (PDB 3NXU),<sup>18</sup> bromoergocryptine (PDB 3UA1),<sup>19</sup> pyridine-, imidazole-, and oxazole-substituted deoxyritonavir analogues (PDB 4I3Q, 4I4G, 4I4H),<sup>20</sup> alkyl-substituted deoxyritonavir derivatives (PDB 4K9T, 4K9U, 4K9V, 4K9X, 4K9W),<sup>21</sup> and desthiazolylmethylcarbonyl ritonavir (PDB 3TJS).<sup>22</sup> An examination of the unliganded structure (PDB 1TQN)<sup>23</sup> shows short F and G helices, and a large highly ordered hydrophobic core of phenylalanines (213, 215, 219, 220, 241, 304) above the active site. The ketoconazole–3A4 complex was resolved with two molecules in the binding pocket (see Figure 1A for an overlay of the unliganded 3A4 crystal structure with the ketoconazole–3A4 complex).<sup>17</sup> Obvious changes in the F and G helices and intervening loops were observed. The hydrophobic cluster comprised of the phenylalanine residues was broken up with some chains exposed to

Received: November 19, 2013

Published: February 27, 2014



**A.****B.**

**Figure 1.** (A) Overlay of the unliganded CYP 3A4 structure (pink, PDB entry 1TQN) onto the ketoconazole-bound CYP 3A4 (cyan, PDB entry 2V0M), with one heme (apo structure) shown. The position of Arg 212 is depicted (ball-and-stick, yellow) because of its switch toward the surface in the CYP 3A4–ketoconazole complex. (B) Overlay of the unliganded CYP 3A4 (pink) with the erythromycin-bound CYP 3A4 structure (red, PDB entry 2J0D).

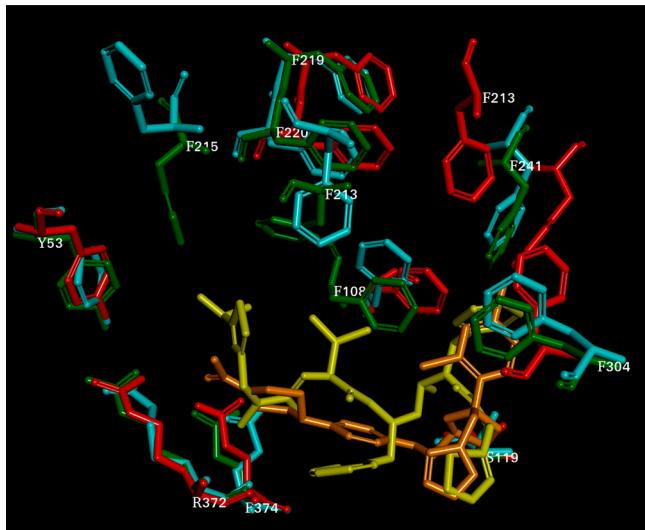
the surrounding medium. Residues 210–213 adopted a helix–helix shape that prolongs the F helix by one extra turn, with Arg 212 being on the surface, contrary to the apo-structure in which the arginine is pointing toward the active site. In contrast, Leu 211 switches from the surface (apo structure) to the interior of the active site.<sup>17</sup> The erythromycin–3A4 complex does not show a productive binding mode. Nevertheless, the complex still displays distinct changes from the uncomplexed structure, and in turn it offers a look into alternative 3A4 conformations with F and G helices dominating the changes (Figure 1B). Helices F and G are shifted, however the direction of the shift is opposite to that of ketoconazole–3A4; similar shifts are observed in helices F' and G' as well. It should be pointed out that the shift in helix F results in a large opening in the binding pocket.<sup>17</sup> Both complexes reinforce that dramatic conformational transitions are taking place upon ligand binding, thus providing an explanation for the wide array of drugs binding to CYP 3A4 as substrates, inhibitors, or inducers. Similar to the erythromycin–3A4 complex, the ritonavir–3A4 structure (Figure 2) does not show a productive binding mode either. However, changes in the F–G region are mirroring what was observed in the complex with ketoconazole. Furthermore, four waters were crystallized within 4 Å of the inhibitor.<sup>18</sup> Differences among the binding pockets of the three crystal ligand–CYP 3A4 complexes are depicted in Figure 3. It can be seen that the phenylalanine cluster, including phenylalanines 213, 215, 241 and to a lesser degree phenylalanines 219, 220, and 304, shows the most deviations among the three complexes. Furthermore, some of the side chains have shifted from outward to inward orientations.

Because the breadth of 3A4 substrates ranges from polar to nonpolar and varies substantially in size, we set out to investigate computationally plausible substrate binding modes in an attempt to understand mechanistic details and the role of waters in metabolism. Specifically, our goals were to address orientation of substrates in the binding pocket of 3A4, explore the role of active-site waters in pose prediction, and assess whether the waters found in 3A4 crystal complexes with ligands should be explicitly included in docking. Of the computational



**Figure 2.** Overlay of the unliganded CYP 3A4 (pink) structure onto the ritonavir–CYP 3A4 crystal complex (green, PDB entry 3NXU), with the bound ligand shown in yellow sticks.

studies regarding P450s, few are structure-based, with three being the closest to the study reported herein. Santos et al. carried out molecular dynamics (MD) simulations of CYP 2D6 complexed with R-3,4-methylenedioxy-N-ethylamphetamine (MDEA) to generate protein conformations and determine hydration sites.<sup>24</sup> Subsequently, eight 2D6 structures were selected, followed by docking of 11 similar and 53 dissimilar to MDEA substrates. Findings could not unequivocally confirm the affect of waters regarding improvements in orientations since reduced docking performance was seen in several cases when water was included. Thus, it was suggested that water presence and its impact are dependent upon the protein conformation and size of the substrates. Similarly, docking of 20 CYP 1A2 substrates was undertaken using GOLD and various scenarios of water inclusion/exclusion in the binding pocket of



**Figure 3.** Active site residues of the crystal CYP 3A4 complexes with ketoconazole (cyan), erythromycin (red), and ritonavir (green). Two of the bound ligands are shown in orange and yellow, while heme is not depicted.

the enzyme.<sup>25</sup> The contribution of water toward accuracy of pose predictions proved difficult to decipher. In some cases, the presence of water led to a substrate orientation accurately ranked among the top five, while in others present waters resulted in inaccurately predicted binding modes. Furthermore, the investigators carried out a virtual screening (VS) experiment of a compound collection consisting of 20 substrates, 20 inhibitors, and seeded with 3331 decoys; varying numbers of hits were identified when several fractions of the database were considered. For example, at 5% of the collection, eight substrates were recovered when water was not included and three when crystallographic waters were present. Of the inhibitors and for the same fraction of the database, just four were returned when water was not present. Finally, the most relevant to CYP 3A4 work has been reported by Teixeira et al., who utilized an earlier crystal structure of 3A4 in complex with metyrapone (inhibitor) and progesterone (substrate).<sup>26</sup> Sixteen known ligands were docked with Autodock into various MD-generated conformational transitions of the enzyme. Interestingly, these investigators showed that the ability of each ligand to be successfully docked into one of the 125 MD-generated receptor conformations ranged from 6% to 45%. Reasons for the low docking performance were thought to be insufficient sampling of the rings in some of the ligands and the flexibility of 3A4.

Because our objectives were to explore and identify productive binding modes of 3A4 substrates, we first validated observed poses of all ligands reported in complexes with CYPs. For this part of the work, we employed LigandFit,<sup>27</sup> GOLD,<sup>28,29</sup> and Glide,<sup>30–32</sup> along with nine scoring functions. The best performing docking programs were GOLD and Glide. Subsequently, docking of 195 known 3A4 substrates (Supporting Information Figure S1) into the three crystal structures 2J0D, 3NXU, and 2V0M was undertaken with some enhancements to the best performers identified in the previous step. Specifically, to address the widely reported receptor flexibility, we employed a modified induced fit docking (IFD)<sup>33–35</sup> methodology with Glide, that is every substrate was docked into a more “open” CYP 3A4 structure with forty poses of each

substrate returned, thus 40 ligand–3A4 complexes per substrate were generated. Subsequently, protein refinement of the complexes was performed with the active site restored, followed by a second stage docking of each substrate into its respective 40 receptor conformations that were previously minimized in the second step of the protocol. These second-stage docking simulations generated 40 solutions per substrate. In summary, the adapted IFD protocol enables us to perform extensive conformational sampling of the CYP 3A4 structures, with the ultimate goal of exploring substrate orientations within the binding pocket. Regarding docking with GOLD, heme parameters were included<sup>36</sup> and ten poses per substrate were generated. All sites of metabolism (SOM) were derived from experimental data; predicted poses from both GOLD and IFD Glide runs were evaluated based on the vicinity of SOMs to the heme iron. Docking simulations were performed in duplicates in water-free proteins and in binding pockets with crystallographic waters as an integral part of the protein.

## ■ COMPUTATIONAL METHODS

**Preparation of Proteins and Substrate Library.** All computations were carried out on quad core Intel 3.0 GHz Xeon processors X5472. The atomic coordinates of the human cytochrome P450 3A4 in complex with erythromycin (PDB ID 2J0D), ritonavir (PDB ID 3NXU), and ketoconazole (PDB ID 2V0M) were obtained from the Protein Data Bank and hydrogen atoms were added within the graphical interface of the respective docking program. Specifically, within the Maestro interface (Schrodinger, LLC, New York, NY), the receptor was prepared by assigning bond orders, adding hydrogens and finding overlaps, followed by hydrogen bond optimization and minimization using the Protein Preparation Wizard. Minimization employed the “Impref” utility, which runs a series of constrained impact minimizations with gradually decreasing strength of the heavy-atom restraining potential. Two minimizations were initially performed with the heavy-atom restraint potential force constant at 10. In the first minimization, the torsional potential was turned off to improve hydrogen optimization, whereas the second minimization restored the torsional potential. The restraining potential force constant was subsequently reduced to 3, 1, 0.3, and 0.1. If the output structure from a minimization exceeds the specified RMSD threshold, relative to the starting structure, the program stops and returns the structure from the previous minimization. Thus, the RMSD is checked at the end of each round of minimization. Receptor grid generation was subsequently employed with the van der Waals (vdw) radius scaling factor set to 1.0, and partial charge cutoff at 0.25. Similarly, within the Hermes interface (Cambridge Crystallographic Data Centre, Cambridge, United Kingdom), the protein preparation wizard was employed. Validation docking experiments were performed on the above crystal 3A4 complexes and on a number of other crystal CYPs to add confidence in the final selection of appropriate docking algorithms for the study. Docking simulations employed GOLD5.1,<sup>37</sup> LigandFit within Discovery Studio,<sup>38</sup> and Glide in Maestro.<sup>39</sup> Specifically, crystal complexes of CYP 2B6 with 4-(4-chlorophenyl)imidazole (PDB ID 3IBD),<sup>40</sup> CYP 2C9 with flurbiprofen (PDB ID 1R9O),<sup>41</sup> and CYP 2E1 with 4-methylpyrazole (PDB ID 3E4E)<sup>42</sup> were also included in the original validation experiments. Reference ligands were extracted from their cognate proteins, their atom and bond types were corrected, hydrogens were added and partial charges

were calculated using the Momany–Rone method.<sup>43</sup> Following validation, duplicate docking runs were performed with and without crystallographic waters, and results were evaluated based on the ability of each docking program to reproduce experimental binding modes of the ligands.

The 3D ligand library included 195 substrates previously reported as CYP 3A4 substrates.<sup>11,44</sup> Molecules were CHARMm energy minimized within PipelinePilot. Special attention was given to the protonation states of the ligands, with ionizable groups (amines, carboxylic acids, phosphates, amidines) assumed to be ionized at physiological pH. The library was saved in maestro and sdf formats, using Maestro's sdconvert utility.

**Validation Docking.** GOLD 5.1. Docking was performed using GOLD 5.1 and both ChemScore and GoldScore were employed as fitness functions. Standard default settings were used in all calculations, with search efficiency set to 50%, and the GoldScore annealing parameters depending on the calculated number of GA operations. All single bonds were treated as rotatable. At the beginning of the runs, external vdw energies range from 9 to 4, thus allowing a few bumps to be tolerated at first. Similarly, the hydrogen bonding parameters are set via the max\_distance, which is the distance between donor hydrogen and fitting point and must be less than max\_distance for the bond to count toward the fitness score. This in turn allows poor hydrogen bonds to occur at the beginning of a GA run. A 10.0 Å radius from each bound ligand was used to define the active site. Early termination was not allowed, while the option for diverse solutions was selected. The heme scoring function parameter file was employed for both GoldScore and ChemScore to take into account different hydrogen bond acceptor types.<sup>36</sup>

Glide 5.7. Glide, version 5.7,<sup>30–32,39</sup> was used in this study. Glide grids were generated with ligand scaling of 0.8 for the vdw radii. Ligands were docked and ranked using standard precision (SP), followed by postdocking minimization. All docking experiments were performed with default settings and 30 poses per ligand were saved for further consideration.

LigandFit. LigandFit<sup>27,37,38</sup> implemented in Discovery Studio Client version 3.1 was used. The binding sites were defined with a 10 Å radius from the bound ligand. Subsequently, conformational sampling of the ligands was performed employing Monte Carlo simulations (10 000 trials). Both vdw and electrostatic energy terms were considered in the calculations. Each docked pose was further fitted into the binding pocket through a number of rigid-body minimizations. A maximum of 30 poses were saved for each molecule to be subsequently scored using LigScore1, LigScore2, PLP1, PLP2, and PMF. The LigScores 1 and 2 were calculated using the Dreiding force field.

**Docking for SOM Predicted Modes.** Docking experiments were performed with IFD and GOLD. All simulations were run in duplicates, thus a total of 10 docking computations were undertaken.

**Induced Fit.** In the first stage docking, certain residues were trimmed from the crystal structures. Specifically, amino acids Phe 108, Met 114, and Gly 481 were trimmed from the ritonavir– and ketoconazole–CYP 3A4 complexes, whereas residues Phe 57, Arg 106, and Phe 241 were trimmed from the erythromycin-bound 3A4 crystal structure. The above side-chains were initially removed in order to provide more room in the cavity and allow multiple binding modes of the substrates to be identified. Of all these residues, only phenylalanines 108 and

241 have been reported to be part of the hydrophobic cluster consisting of six residues, thus involved in ligand interactions. It should be noted that the original side-chains were later restored. Substrates were subsequently docked with SP, and forty complexes were generated per substrate. The general vdw radius scaling partial charge cutoff for the receptor was set to 0.25, while the partial charge cutoff for vdw radii scaling, the Coulomb-vdw cutoff, and the hydrogen bond filter cutoff for the ligand were set to 0.15, 100.0, and 0.0, respectively. The *x*, *y*, and *z* dimensions of the ligand centroid bounding box and of the outer (grid) box were 10.0 and 30.0 Å, respectively. The general vdw radius scaling factors for the receptor and the ligand were set at 0.50. Once the first-stage docking was completed, deleted residues were restored and all residues and ligand within 4 Å from each pose were energy minimized using Prime.<sup>45,46</sup> Next, poses with Prime energies within 30 kcal/mol of the lowest pose were sorted, and the top 10 were considered for further investigations. Finally, a second-stage docking was employed for each of the substrates docked into its respective forty receptor conformations with the same values as in first stage docking above, with no positive Coulomb-vdw energies allowed, and ligand scaling factor for vdw radius scaling set to 0.8. Forty poses per substrate were generated in this final stage. Solutions were visually inspected and processed based on the IFD score and the SOM's vicinity to the heme iron. The IFD score is computed based on GlideScore and a small fraction of the Prime energy as follows: IFDScore = 1.0 × GlideScore + 0.05 × Prime\_Energy.

**GOLD 5.1.** All calculations were performed as described in the section under Validation Docking with ten solutions, and the ChemScore and GoldScore heme parameters.

**Identification of Binding Pockets.** FTMap. We employed the FTMap algorithm at the server (<http://ftmap.bu.edu/>)<sup>47</sup> following the typical procedure, that is fragments were used as probes for the binding surface of each starting crystal structure after it was prepared with the protein preparation wizard within maestro. Subsequently, the best poses were retained and minimized with the protein atoms held fixed and only the probe atoms allowed to move. Minimized fragments were clustered and ranked, and consensus sites were assigned, followed by overlapping the docked poses corresponding to productive binding modes of the CYP 3A4 substrates onto the consensus sites from FTMap.

## ■ RESULTS AND DISCUSSION

CYPs have been challenging structurally due to their plasticity and ability to accommodate a plethora of diverse compounds. Mammalian P450s are connected to the membrane via an N-terminal  $\alpha$ -helix. This helix is removed or replaced by hydrophilic residues in order to facilitate crystallization. CYPs have a buried, mostly hydrophobic binding pocket, which in 3A4 consists mainly of a phenylalanine cluster right on top of the active site cavity.

Although their association with the membrane has complicated structural and functional investigations, various studies have proposed a biophysical model for the CYPs in an approximate *in vivo* environment. The first crystal structure of a mammalian microsomal CYP presented an assembly in which the CYP was not fully integrated into the membrane, but instead interacted with the membrane through a hydrophobic cluster on one face of the protein.<sup>48</sup> Subsequent atomic force microscopy experiments reinforced the above structural template.<sup>49</sup> Specifically the current model for CYP 3A4

Table 1. Summary of Docking Successes

	2V0M <sup>a</sup>		2J0D <sup>a</sup>		2J0D HOH		3NXU <sup>a</sup>		3NXU HOH	
	total <sup>b</sup>	accurate <sup>c</sup>	total	accurate	total	accurate	total	accurate	total	accurate
IFD	181	152 (84)	190	170 (89)	186	117 (63)	183	137 (75)	181	125 (69)
GOLD	194	147 (76)	195	133 (68)	195	146 (75)	195	113 (58)	195	145 (74)

<sup>a</sup>Protein databank accession codes correspond to CYP 3A4 complexes with ketoconazole (2V0M), erythromycin (2J0D), and ritonavir (3NXU).

<sup>b</sup>Total number of substrates docked. <sup>c</sup>Numbers in parentheses correspond to percent accuracy.

suggests that it is tethered to the endoplasmic reticulum membrane via its N-terminus, while other parts of the enzyme, including the F'G' loop and the substrate access/exit channels are embedded into the bilayer.<sup>48,50</sup> The remainder of the CYP, which is predominantly hydrophilic, is exposed to the aqueous cytoplasmic environment and dynamically interacts with reductase and other cofactors.

Because neither the membrane nor the N-terminal tether are required for the CYP catalytic mechanism *in vitro*,<sup>51–53</sup> we were not concerned they were not included in the present investigation. Our primary aim was to find optimal docking strategies of ligands into representative structures of CYP 3A4, and to also consider the effect that the presence or absence of active site waters can have on docking performance. It should be pointed out that our methodology inherently supports either the conformational selection theory<sup>54</sup> or the induced fit paradigm<sup>55</sup> as accountable for the observed plasticity of this P450 isoform. Furthermore, we were able to ascertain protein–ligand interactions in molecular detail, and indirectly evaluate the docking methodology in its ability to predict substrate binding to 3A4.

As alluded to earlier in the Introduction, available CYP 3A4 crystal structures do not suffice to address the enzyme's flexibility. Earlier reports showed a positive correlation between flexibility and substrate promiscuity in 3A4, which was also noted to be more flexible in the active site relative to other P450s.<sup>26,56</sup> In an effort to address the conformational versatility of CYP3A4, we employed two methodologies: (i) A complex and tailored to the system IFD methodology with Glide<sup>30–32</sup> and (ii) GOLD<sup>28,29</sup> incorporating heme parameters<sup>36</sup> for a more tailored binding mode sampling. Because the metabolic products of the CYP3A4 substrates have been reported, we decided to consider both primary and secondary SOMs when kinetics data were not available. In turn, the ability to predict poses observed experimentally provides further confidence in our approach (substrates docked in this study are depicted in Supporting Information Figure S1).

As mentioned above, evaluation of LigandFit,<sup>27</sup> Glide,<sup>30–32</sup> and GOLD<sup>28,29</sup> to reproduce experimental binding modes observed in various CYP–ligand complexes preceded any subsequent studies on 3A4. Thus, several crystal structures of CYP 3A4 bound complexes, along with the human CYPs 1A2, 2B6, 2C9, and 2E1 were considered in this step. LigandFit reproduced the 3A4–ritonavir, 3A4–erythromycin, and 3A4–ketoconazole complexes. Glide was successful in reproducing the observed binding modes of  $\alpha$ -naphthoflavone with CYP 1A2, 4-(4-chlorophenyl)imidazole with 2B6, and flurbiprofen, 4-methylpyrazole, and ketoconazole into 2C9, 2E1, and 3A4, respectively (results not shown). However, not all ligands in the above crystal structures are substrates, and some of the binding modes do not correspond to productive ones. Furthermore, preliminary calculations involving a VS experiment of a small substrate library employing LigandFit failed to reproduce reported SOMs (results not shown). Consequently, we elected

to use GOLD and IFD with Glide for the work reported herein. The IFD workflow included three crystal CYP 3A4 complexes, with three side-chain residues in the binding pockets being trimmed during the first docking stage. By scaling back the vdw radii and increasing the Coulomb-vdw cutoff, more “breathing space” is given to the binding pocket, thus generating poses that might not have otherwise been attainable due to potential steric clashes. Forty poses per substrate were generated in this stage, which in turn resulted in forty receptor conformations depending on the bound ligand. Subsequently, the deleted amino acid side chains were rebuilt, followed by Prime<sup>45,46</sup> reorientation of the amino acid rotamers surrounding the docked ligand and final optimization of each complex. Every substrate was then redocked for the second time into each of the forty minimized receptor conformations, and forty poses per receptor were returned from each docking experiment. The above calculations were run in duplicates with and without crystallographic waters. By employing this workflow, we were able to replicate the induced fit theory of binding in the first stage docking and the conformational selection theory in the second stage docking. Final poses were evaluated based on the vicinity of SOMs to the heme iron. Substrate binding modes were considered acceptable if within 5 Å from the heme, and provided that no other moieties were positioned in such a way that the reactive molecular center would be hindered from coordinating with the iron. In an independent evaluation, we also considered the IFD scoring function which incorporates the Prime energy and GlideScore.

It can be seen in Table 1 that with few exceptions, substrates are accurately docked in most of the scenarios we explored here. In Supporting Information Table S.1, failures and successes on a per substrate basis are being tabulated. The most challenging appear to be alfentanil, aprepitant, cisapride, dextromethorphan, fentanyl, itraconazole, and pimozide, in that only one or two scenarios, out of the total 10 runs, succeeded in docking them accurately. It should be pointed out that quite a few of these drugs are also inhibitors; however, our evaluation of docked solutions is based on the distance of respective SOMs to the heme iron to exclude the possibility of inhibitor binding orientations which would simply occupy the binding pocket. As seen in Table 1, best performance ranges from 74% to 89%, depending on the docking algorithm or scenario employed. Of the three original crystal structures included in this study, it appears that 2J0D (CYP 3A4–erythromycin) would be the best to employ if accuracy of predicted binding modes is the objective (see Table 1). It is also telling that better performance is observed when waters are not included in IFD. This may not be surprising given that the waters in the two crystal structures (2J0D and 2V0M, CYP 3A4–erythromycin and CYP 3A4–ketoconazole, respectively) under investigation are not highly conserved. Therefore, removing the waters will give more space in the binding pocket and allow ligands to dock in orientations that could lead to feasible receptor–ligand complexes. However, when waters are kept, they become part

of the grid and will not move out of the way in order to enable a more optimal fit for the substrates. In contrast, GOLD is seemingly performing better when waters are included (Table 1). GOLD assigns fitting points to solvent accessible binding site atoms and water oxygens. Water molecules can form four hydrogen bonding interactions, thus waters along with binding site fitting points will be preferentially mapped to corresponding ligand atoms with acceptor/donor properties when sterically feasible in GOLD.

In an effort to delineate the types of metabolic reactions that are accurately predicted the most, we present a summary of the breakdown of docking performance by CYP 3A4-mediated reactions (Table 2) for the two crystal structures we had the

**Table 2. Successful Predictions of Metabolic Pathways**

type	total <sup>a</sup>		success			
	primary	secondary	IFD		GOLD	
			2J0D <sup>b</sup>	2V0M <sup>b</sup>	2J0D	2V0M
aliphatic oxidation	73	12	68	62	70	53
N-dealkylation	68	20	59	57	64	52
aromatic hydroxylation	20	5	20	19	20	17
O-dealkylation	18	4	17	14	18	12
aromatization	8	1	6	6	8	3
sulfoxidation	5	6	5	5	5	2
N-oxidation	3	7	2	1	3	4
other	4	0	3	3	4	4

<sup>a</sup>Total refers to reported metabolic 3A4-mediated reactions of the substrates. <sup>b</sup>Complexes of CYP 3A4 with erythromycin (2J0D) and ketoconazole (2V0M).

most success overall. It can be seen that GOLD outperforms IFD when the erythromycin–CYP complex is considered for all primary metabolic reactions. In contrast, IFD is more predictive in regards to the ketoconazole–3A4 crystal complex in reference to the primary metabolic reactions. When primary and secondary 3A4-mediated reactions are both taken into consideration, it appears that N-dealkylations, sulfoxidations, and N-oxidations present the most challenges. In contrast, aromatic hydroxylations, O-dealkylations, aliphatic oxidations, and aromatizations are the most consistent in regards to accuracy of predictions. This is not surprising since N-demethylations are quite common, yet the proximity of the reactive center to the heme iron can be hindered by the aliphatic substituent on the heavy atom.

We thought it might be worthwhile to examine whether we see more failures within certain therapeutic categories over others. Table 3 depicts the number of compounds that failed to be docked in each scenario per disease category. Even though no clear distinctions can be made, it is noticeable that certain schemes were more effective than others, thus suggesting that some categories displayed higher rates of failures than others; for example, almost all calculations resulted in mostly productive poses for miscellaneous and analgesics contrary to cardiovascular drugs. Consequently, we felt it might be informative to look at failures from a molecular structural perspective. Toward that goal, we explored the plausible correlation between ligand molecular properties and SOM predictive performance, even though our intent was not a pharmacophore-based approach. Common variables such as molecular weight, hydrogen-bond acceptors, hydrogen-bond donors, number of rotatable bonds, and molecular volume were calculated for our substrate library (results not shown) in an effort to assess whether certain properties are the most critical in terms of performance. We noted that structures with an average molecular weight (around 400) and a high number of rotatable bonds (7–14) were challenging because of their conformational flexibility, which in turn would deter an SOM from being predicted near the heme iron. Similarly, ligands with molecular weights in the same range but with a smaller number of rotatable bonds are rather inflexible; consequently, the docking algorithms may succeed in sampling their conformational space, however, due to their low(er) molecular weight, several options for placements within the large and rather flexible sites exist. In conclusion, size and flexibility play a determining role in docking predictions.

A critical question in docking is in regards to binding patterns and residues involved in receptor–ligand complex formation (Table 4). Results for the most often observed amino acids are shown for poses in the erythromycin–3A4 complex. It can be seen that a majority of the interactions are with hydrophobic amino acids, which is not surprising given the nature of the active site of CYP 3A4. Specifically, the phenylalanine cluster is maintained; bonding interactions are observed with phenylalanines 108, 213, 304, and to a lesser degree with Phe 220. Additional interactions are seen with the side chains of Arg 105, Ile 301, Ile 369, and the side and main chains of Ala 370 and Leu 482. Hydrogen-bonding interactions are formed with the side-chain of Ser 119. Finally, electrostatic interactions with the backbone and side chains of Glu 374 are observed. Figure 4 depicts a diagram of the main binding

**Table 3. Failed Ligand Disease Classes/Categories**

	2V0M <sup>a</sup>		2J0D <sup>a</sup>		2J0D HOH		3NXU <sup>a</sup>		3NXU HOH	
	IFD	GOLD	IFD	GOLD	IFD	GOLD	IFD	GOLD	IFD	GOLD
anti-infectives (25)	3	2	1	5	5	5	4	8	2	6
cardiovascular (32)	7	9	4	7	14	3	9	16	11	8
oncology (18)	2	3	2	8	5	7	5	6	5	5
psychotropic (39)	3	9	3	11	18	12	6	17	11	9
men's/women's (17)	2	2	1	8	4	3	6	4	5	4
analgesia/sedation (13)	4	5	3	3	8	3	4	7	7	3
diabetes (4)	2	1	1	2	1	2	2	2	2	1
immunomodulators (3)	1	3	0	2	2	2	1	3	0	3
miscellaneous (44)	4	13	5	16	12	12	9	19	12	11
total failures	28	47	20	62	69	49	46	82	55	50

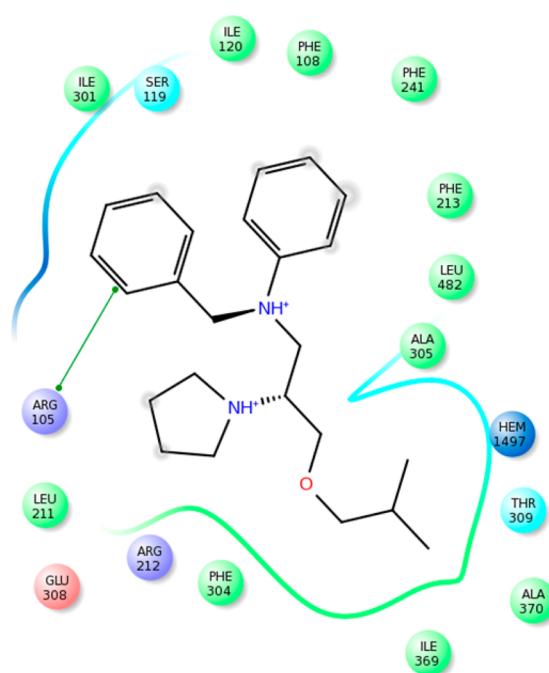
<sup>a</sup>Complexes of CYP 3A4 with erythromycin (2J0D), ketoconazole (2V0M), and ritonavir (3NXU).

Table 4. Ligand–Receptor Interactions ( $2\text{J}0\text{D}$  with No Water Included)<sup>a</sup>

Table 4. continued

Table 4. continued

"aaSC" and "BB" denote side-chain and backbone interactions, while "X" marks predicted interactions with respective residues.

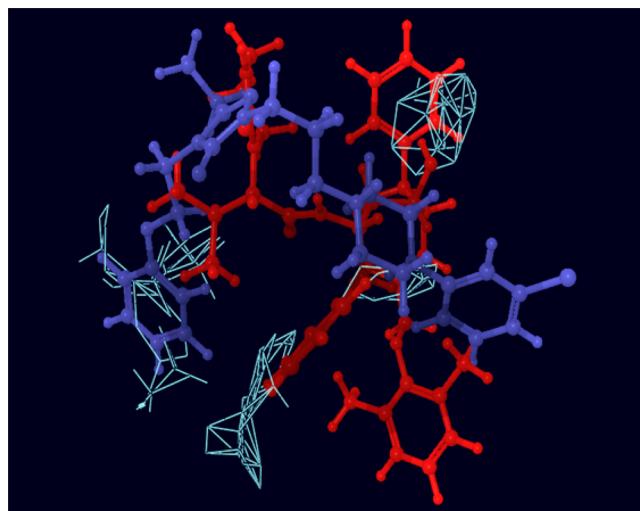


**Figure 4.** Schematic depiction of a representative substrate with the most populated amino acids forming bonding interactions.

interactions observed with a representative substrate. Our findings are in agreement with Davydov et al., who identified Arg 105, Leu 373, and Ser 119 as the residues involved in the substrate-recognition region, along with Ile 369, Phe 213, and Phe 304.<sup>57</sup> It should be noted that Arg 212, which is critical in the final stage of bromoergocryptine's positioning closer to the heme,<sup>19</sup> was not predicted to significantly contribute to CYP 3A4-ligand complex stabilization in the study here. The rationale might be that in 2J0D the Arg 212 side chain is on the surface of the protein, contrary to the apo 1TQN (see Figure 2A) where it points toward the active site.

To assess scoring accuracy, we calculated the IFD scores and examined the ability of the scoring functions IFD, GoldScore, and ChemScore to rank the correct substrate poses (Table 5). Several studies have already shown that scoring functions are not able to distinguish experimentally observed binding modes from the rest of the docked poses.<sup>58–60</sup> We report the top-ranked poses, as well as those in the top five and top seven ranks in Table 5. GoldScore and ChemScore are efficient in differentiating accurate from inaccurate poses almost entirely at a 5% fraction of the substrate library. The plausible reason for the discrepant results between the scoring functions employed in GOLD versus IFD scoring may be due to the heme parameters employed by GOLD.

To further validate the accuracy of the docking simulations, we carried out calculations using computational solvent mapping (FTMAP server),<sup>61,62</sup> which docks small organic molecules, followed by identification of favorable positions on the protein surface, clustering of all conformations, and finally ranking the clusters according to their average free energy. Subsequently, the low energy clusters are grouped into consensus sites with the largest being considered as active sites. Figure 5 shows mapping results of two representative



**Figure 5.** FTMap of lopinavir (red) and nefazodone (blue).

ligands, nefazodone and lopinavir. It can be seen that both substrate conformations coincide with hot spots identified by the program. Specifically, consensus regions detected by FTMap correspond to the ionic pocket of Arg 372, Arg 105, and Glu 374, and the hydrophobic cluster comprised of Phe 108 and Phe 213, residues which can also be seen in Table 4 of residues involved in stabilizing CYP 3A4-substrate complexes.

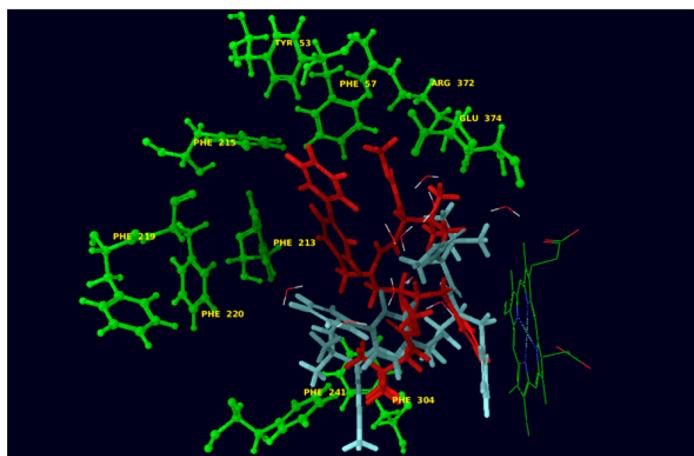
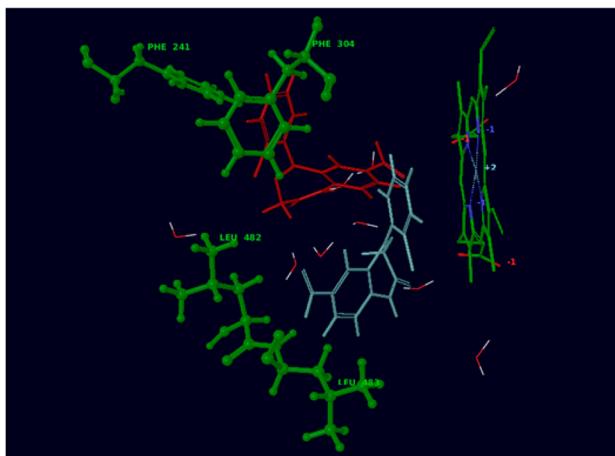
As discussed in the Introduction, another objective of this work was the role of crystallographic waters and their contribution to docking performance. Going back to Table 1, we noted that water inclusion is favorable for some docking algorithms over others. To reiterate, GOLD's performance is improved when waters are included in the docking simulations. In contrast, IFD with Glide performs better when the crystallographic waters are not included in the computations. It would appear reasonable to anticipate Glide's poor performance with hydration because waters are part of the grid and could possibly interfere or preclude a particular substrate from binding. On the other hand, GOLD may not place a ligand next to a water molecule, it will nevertheless favor hydrogen bonding interactions of ligands being positioned in close proximity to waters. However, the effect of water

**Table 5. Cumulative Rankings for Docked Substrates**

docking	scoring	2V0M <sup>a,b,c</sup>			2J0D <sup>a,b,c</sup>			2J0D HOH <sup>a,b</sup>			3NXU <sup>a,b,c</sup>			3NXU HOH <sup>a,b</sup>		
		1	5	7	1	5	7	1	5	7	1	5	7	1	5	7
IFD (GLIDE)	IFD	26	68	92	22	42	50	12	21	28	18	30	32	35	43	48
GOLD	GoldScore	52	111	135	45	102	115	31	93	122	47	89	103	48	118	132
	ChemScore	39	114	137	33	99	116	35	101	123	41	94	108	36	108	124

<sup>a</sup>Top-ranked poses or in the top five or top 7 in respective docking and scoring schemes. <sup>b</sup>Numbers are cumulative with the previous column(s).

<sup>c</sup>Complexes of CYP 3A4 with erythromycin (2J0D), ketoconazole (2V0M), and ritonavir (3NXU).

**A.****B.**

**Figure 6.** (A) Atazanavir orientation in the presence (blue) or absence (red) of waters in the CYP3A4–ritonavir complex. The presence of waters had a worsening effect on identifying the correct pose. (B) Clonazepam orientation in the presence (red) or absence (blue) of waters CYP3A4–ritonavir complex. Contrary to that shown in panel A where the effect of waters is seen, the correct pose was identified in the presence of waters.

inclusion is dependent upon the conformation and the size of the substrate. The latter agrees with previously reported findings on other P450s. Specifically, Santos et al.<sup>24</sup> investigated the role of water in molecular docking simulations of CYP 2D6, and concluded that hydration was not always necessary for better performance in predicting SOMs. Similarly, Vasanthanathan et al.<sup>25</sup> explored VS of a library of substrates and inactives of CYP 1A2. They reported that the presence of individual waters was at times a deterrent to identifying the correct poses of some substrates. In contrast, de Graaf et al.<sup>63</sup> in a much earlier study, explored scenarios for binding mode predictions in P450cam, P450 2C5, and P450EryF. These investigators showed that inclusion of either crystallographic or computationally predicted active site waters resulted in better docking performance than when the waters were omitted. What we have seen is not different from the aforementioned papers. Figure 6A and B depicts atazanavir and clonazepam, respectively, as representative examples of docking predictions where the presence or absence of waters enhanced or diminished docking performance. Specifically, in the absence of water, atazanavir was predicted to be close to the heme iron, while waters in the active site precluded identification of the correct pose by favoring a binding mode in an unfavorable angle to the heme. In contrast, clonazepam is accurately predicted to bind when waters are present because they are pushing the ligand toward the phenylalanine cluster, while elimination of waters results in a pose that is shifted toward the bottom of the binding pocket. However, it should be pointed out that contrary to de Graaf et al.,<sup>63</sup> our investigation does not include predicted waters, neither do we see that docking is better at predicting SOM than reproducing observed binding conformations. Toward that end, it might be interesting in the future to allow certain waters to spin in an attempt to optimize hydrogen-atom positions by enabling them to translate within a predefined distance. Nonetheless, our findings are in agreement with previous investigations on other CYPs, where improved catalytic site prediction was not unambiguous but rather depended upon different waters on a per substrate basis. We are indeed able to use our methodology for predictions of catalytic sites and X-ray ligand-target complexes.

## CONCLUSION

In this study, we employed two methodologies to investigate the conformational requirements of each ligand binding to CYP 3A4. We are able to show that an ensemble of protein conformations stemming from an adapted IFD protocol can sample the target binding space and predict catalytic sites for the majority of substrates, with percentages varying depending upon the starting 3A4 crystal structure. Even though docking performance was somewhat compromised, inclusion of waters and heme parameters also led to accurate catalytic site predictions when GOLD was employed. It should be pointed out that the erythromycin–3A4 crystal structure used as a starting point in our study does not include a productive binding conformation of the drug. However, as has been suggested by several experimentalists in the past, the plasticity observed in 3A4 is still evident when substrates are simply occupying the binding site, and suffices to provide an open enough pocket to accommodate diverse ligands.<sup>64–70</sup> Our analyses suggest that in the ligand–CYP complexes, certain residues are repeatedly predicted to make favorable and stabilizing interactions with the bound substrates. Docking in multiple receptor conformations is seemingly increasing the chances of accurate catalytic prediction. With the exception of 3.5% of the total substrate library, the rest of the ligands are docked in productive binding modes. Furthermore, the size and degree of flexibility of the ligands seem to influence docking accuracy, while the presence of crystallographic waters does not necessarily yield improved performance. In turn, this suggests that extensive solvation needs to be undertaken to fully evaluate the contributing effect of waters to substrate binding.

## ASSOCIATED CONTENT

### Supporting Information

Failed ligands (docking performance on a per substrate basis for all scenarios) and list of all known CYP 3A4 substrates used in the present study. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*Telephone: 618-650-5166. Fax: 618-650-5145. E-mail: mkontoy@siue.edu.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The authors would like to thank Professor Tom Poulos, Departments of Molecular Biology and Biochemistry, Chemistry, and Pharmaceutical Sciences, University of California, Irvine, for his invaluable comments and suggestions throughout the course of this work, and for reviewing the manuscript. We are indebted to Irina Sevrioukova for carefully reviewing this paper and for her useful comments. We are also thankful to Brittany Schuchman, Department of Biological Sciences, SIUE, for some of the IFD scoring calculations.

## REFERENCES

- (1) Arrowsmith, J. Trial watch: Phase II failures: 2008–2010. *Nat. Rev. Drug Discovery* **2011**, *10*, 328–9.
- (2) Arrowsmith, J. Trial watch: Phase III and submission failures: 2007–2010. *Nat. Rev. Drug Discovery* **2011**, *10*, 87.
- (3) Liebler, D. C.; Guengerich, F. P. Elucidating mechanisms of drug-induced toxicity. *Nat. Rev. Drug Discovery* **2005**, *4*, 410–20.
- (4) Guengerich, F. P. Cytochrome P450s and other enzymes in drug metabolism and toxicity. *AAPS J.* **2006**, *8*, E101–11.
- (5) Kirchmair, J.; Williamson, M. J.; Tyzack, J. D.; Tan, L.; Bond, P. J.; Bender, A.; Glen, R. C. Computational prediction of metabolism: Sites, products, SAR, P450 enzyme dynamics, and mechanisms. *J. Chem. Inf. Model.* **2012**, *52*, 617–48.
- (6) Tompkins, L. M.; Wallace, A. D. Mechanisms of cytochrome P450 induction. *J. Biomed. Mol. Toxicol.* **2007**, *21*, 176–81.
- (7) Verras, A.; Kuntz, I. D.; Ortiz de Montellano, P. R. Computer-assisted design of selective imidazole inhibitors for cytochrome p450 enzymes. *J. Med. Chem.* **2004**, *47*, 3572–9.
- (8) Nelson, D. R. Cytochrome P450 and the individuality of species. *Arch. Biochem. Biophys.* **1999**, *369*, 1–10.
- (9) Guengerich, F. P.; Wu, Z. L.; Bartleson, C. J. Function of human cytochrome P450s: characterization of the orphans. *Biochem. Biophys. Res. Commun.* **2005**, *338*, 465–9.
- (10) Makris, T. M.; Denisov, I.; Schlichting, I.; Sligar, S. G. Activation of molecular oxygen by cytochrome P450. In *Cytochrome P450: Structure, Mechanism and Biochemistry*; Ortiz de Montellano, P. R., Ed.; Plenum Press: New York, 2005; pp 149–82.
- (11) Williams, D. A. Drug Metabolism. In *Foye's Principles of Medicinal Chemistry*, 7th ed.; Lemke, T., Williams, D. A., Roche, V. F., Zito, S. W., Eds.; Lippincott Williams & Wilkins: Baltimore, MD, 2013; pp 106–90.
- (12) Guengerich, F. P. In *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 3rd ed.; Ortiz de Montellano, P. R., Ed.; Kluwer Academic/Plenum Publishers: New York, 2005; pp 377–530.
- (13) Williams, J. A.; Hyland, R.; Jones, B. C.; Smith, D. A.; Hurst, S.; Goosen, T. C.; Peterkin, V.; Koup, J. R.; Ball, S. E. Drug–drug interactions for UDP-glucuronosyltransferase substrates: A pharmacokinetic explanation for typically observed low exposure (AUCi/AUC) ratios. *Drug Metab. Dispos.* **2004**, *32*, 1201–8.
- (14) Guengerich, F. P. Enzymatic oxidation of xenobiotic chemicals. *Crit. Rev. Biochem. Mol. Biol.* **1990**, *25*, 97–153.
- (15) Denisov, I. G.; Baas, B. J.; Grinkova, Y. V.; Sligar, S. G. Cooperativity in cytochrome P450 3A4: Linkages in substrate binding, spin state, uncoupling, and product formation. *J. Biol. Chem.* **2007**, *282*, 7066–76.
- (16) Guengerich, F. P. Cytochrome P-450 3A4: Regulation and role in drug metabolism. *Annu. Rev. Pharmacol. Toxicol.* **1999**, *39*, 1–17.
- (17) Ekroos, M.; Sjogren, T. Structural basis for ligand promiscuity in cytochrome P450 3A4. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13682–7.
- (18) Sevrioukova, I. F.; Poulos, T. L. Structure and mechanism of the complex between cytochrome P4503A4 and ritonavir. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 18422–7.
- (19) Sevrioukova, I. F.; Poulos, T. L. Structural and mechanistic insights into the interaction of cytochrome P4503A4 with bromoergocryptine, A type I ligand. *J. Biol. Chem.* **2012**, *287*, 3510–7.
- (20) Sevrioukova, I. F.; Poulos, T. L. Pyridine-substituted desoxyritonavir is a more potent inhibitor of cytochrome P450 3A4 than ritonavir. *J. Med. Chem.* **2013**, *56*, 3733–41.
- (21) Sevrioukova, I. F.; Poulos, T. L. Dissecting cytochrome P450 3A4–ligand interactions using ritonavir analogues. *Biochemistry* **2013**, *52*, 4474–81.
- (22) Sevrioukova, I. F.; Poulos, T. L. Interaction of human cytochrome P4503A4 with ritonavir analogs. *Arch. Biochem. Biophys.* **2012**, *520*, 108–16.
- (23) Yano, J. K.; Wester, M. R.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-Å resolution. *J. Biol. Chem.* **2004**, *279*, 38091–4.
- (24) Santos, R.; Hritz, J.; Oostenbrink, C. Role of water in molecular docking simulations of cytochrome P450 2D6. *J. Chem. Inf. Model.* **2010**, *50*, 146–54.
- (25) Vasanthanathan, P.; Hritz, J.; Tabouret, O.; Olsen, L.; Jorgensen, F. S.; Vermeulen, N. P.; Oostenbrink, C. Virtual screening and prediction of site of metabolism for cytochrome P450 1A2 ligands. *J. Chem. Inf. Model.* **2009**, *49*, 43–52.
- (26) Teixeira, V. H.; Ribeiro, V.; Martel, P. J. Analysis of binding modes of ligands to multiple conformations of CYP3A4. *Biochim. Biophys. Acta* **2010**, *1804*, 2036–45.
- (27) Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. LigandFit: A novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graphics Modell.* **2003**, *21*, 289–307.
- (28) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–48.
- (29) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein–ligand docking using GOLD. *Proteins* **2003**, *52*, 609–23.
- (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*, 1739–49.
- (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) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750–9.
- (33) Sherman, W.; Beard, H. S.; Farid, R. Use of an induced fit receptor structure in virtual screening. *Chem. Biol. Drug Des.* **2006**, *67*, 83–4.
- (34) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **2006**, *49*, 534–53.
- (35) Farid, R.; Day, T.; Friesner, R. A.; Pearlstein, R. A. New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies. *Bioorg. Med. Chem.* **2006**, *14*, 3160–73.
- (36) Kirton, S. B.; Murray, C. W.; Verdonk, M. L.; Taylor, R. D. Prediction of binding modes for ligands in the cytochromes P450 and other heme-containing proteins. *Proteins* **2005**, *58*, 836–44.

- (37) GOLD 5.1; Cambridge Crystallographic Data Centre: Cambridge, United Kingdom, 2005–2011.
- (38) Discovery Studio Client 3.1.0; Accelrys, Inc: San Diego, CA, 2012.
- (39) Maestro, version 9.2.112; Schrodinger, LLC: New York, NY, 2011.
- (40) Gay, S. C.; Shah, M. B.; Talakad, J. C.; Maekawa, K.; Roberts, A. G.; Wilderman, P. R.; Sun, L.; Yang, J. Y.; Huelga, S. C.; Hong, W. X.; Zhang, Q.; Stout, C. D.; Halpert, J. R. Crystal structure of a cytochrome P450 2B6 genetic variant in complex with the inhibitor 4-(4-chlorophenyl)imidazole at 2.0-Å resolution. *Mol. Pharmacol.* **2010**, *77*, S29–38.
- (41) Wester, M. R.; Yano, J. K.; Schoch, G. A.; Yang, C.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-Å resolution. *J. Biol. Chem.* **2004**, *279*, 35630–7.
- (42) Porubsky, P. R.; Meneely, K. M.; Scott, E. E. Structures of human cytochrome P-450 2E1. Insights into the binding of inhibitors and both small molecular weight and fatty acid substrates. *J. Biol. Chem.* **2008**, *283*, 33698–707.
- (43) Momany, F. A.; Rone, R. Validation of the general purpose QUANTA 3.2/CHARMm force field. *J. Comput. Chem.* **1992**, *13*, 888–900.
- (44) Cytochrome P450 Drug Interactions. In *Pharmacist's Letter/Prescriber's Letter*; Therapeutic Research Center: Stockton, CA, 2011; Vol. 22, p 220233.
- (45) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J.; Honig, B.; Shaw, D. E.; Friesner, R. A. A hierarchical approach to all-atom protein loop prediction. *Proteins* **2004**, *55*, 351–67.
- (46) Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. On the role of the crystal environment in determining protein side-chain conformations. *J. Mol. Biol.* **2002**, *320*, 597–608.
- (47) Ngan, C. H.; Bohnuud, T.; Mottarella, S. E.; Beglov, D.; Villar, E. A.; Hall, D. R.; Kozakov, D.; Vajda, S. FTMAP: Extended protein mapping with user-selected probe molecules. *Nucleic Acids Res.* **2012**, *40*, W271–5.
- (48) Williams, P. A.; Cosme, J.; Sridhar, V.; Johnson, E. F.; McRee, D. E. Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Mol. Cell* **2000**, *5*, 121–31.
- (49) Bayburt, T. H.; Sligar, S. G. Single-molecule height measurements on microsomal cytochrome P450 in nanometer-scale phospholipid bilayer disks. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6725–30.
- (50) Fishelovitch, D.; Shaik, S.; Wolfson, H. J.; Nussinov, R. Theoretical characterization of substrate access/exit channels in the human cytochrome P450 3A4 enzyme: involvement of phenylalanine residues in the gating mechanism. *J. Phys. Chem. B* **2009**, *113*, 13018–25.
- (51) Pernecke, S. J.; Larson, J. R.; Philpot, R. M.; Coon, M. J. Expression of truncated forms of liver microsomal P450 cytochromes 2B4 and 2E1 in *Escherichia coli*: influence of NH<sub>2</sub>-terminal region on localization in cytosol and membranes. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2651–5.
- (52) Li, Y. C.; Chiang, J. Y. The expression of a catalytically active cholesterol 7 alpha-hydroxylase cytochrome P450 in *Escherichia coli*. *J. Biol. Chem.* **1991**, *266*, 19186–91.
- (53) Sagara, Y.; Barnes, H. J.; Waterman, M. R. Expression in *Escherichia coli* of functional cytochrome P450c17 lacking its hydrophobic amino-terminal signal anchor. *Arch. Biochem. Biophys.* **1993**, *304*, 272–8.
- (54) Changeux, J. P.; Edelstein, S. Conformational selection or induced fit? 50 years of debate resolved. *F1000 Biol. Rep.* **2011**, *3*, 19.
- (55) Koshland, D. E. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 98–104.
- (56) Skopalik, J.; Anzenbacher, P.; Otyepka, M. Flexibility of human cytochromes P450: Molecular dynamics reveals differences between CYPs 3A4, 2C9, and 2A6, which correlate with their substrate preferences. *J. Phys. Chem. B* **2008**, *112*, 8165–73.
- (57) Davydov, D. R.; Rumfeldt, J. A.; Sineva, E. V.; Fernando, H.; Davydova, N. Y.; Halpert, J. R. Peripheral ligand-binding site in cytochrome P450 3A4 located with fluorescence resonance energy transfer (FRET). *J. Biol. Chem.* **2012**, *287*, 6797–809.
- (58) Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. Evaluation of docking performance: comparative data on docking algorithms. *J. Med. Chem.* **2004**, *47*, 558–65.
- (59) Kontoyianni, M.; Sokol, G. S.; McClellan, L. M. Evaluation of library ranking efficacy in virtual screening. *J. Comput. Chem.* **2005**, *26*, 11–22.
- (60) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912–31.
- (61) Brenke, R.; Kozakov, D.; Chuang, G. Y.; Beglov, D.; Hall, D.; Landon, M. R.; Mattos, C.; Vajda, S. Fragment-based identification of druggable “hot spots” of proteins using Fourier domain correlation techniques. *Bioinformatics* **2009**, *25*, 621–7.
- (62) Landon, M. R.; Lieberman, R. L.; Hoang, Q. Q.; Ju, S.; Caaveiro, J. M.; Orwig, S. D.; Kozakov, D.; Brenke, R.; Chuang, G. Y.; Beglov, D.; Vajda, S.; Petsko, G. A.; Ringe, D. Detection of ligand binding hot spots on protein surfaces via fragment-based methods: Application to DJ-1 and glucocerebrosidase. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 491–500.
- (63) de Graaf, C.; Pospisil, P.; Pos, W.; Folkers, G.; Vermeulen, N. P. Binding mode prediction of cytochrome p450 and thymidine kinase protein–ligand complexes by consideration of water and rescoring in automated docking. *J. Med. Chem.* **2005**, *48*, 2308–18.
- (64) Cameron, M. D.; Wen, B.; Allen, K. E.; Roberts, A. G.; Schuman, J. T.; Campbell, A. P.; Kunze, K. L.; Nelson, S. D. Cooperative binding of midazolam with testosterone and alpha-naphthoflavone within the CYP3A4 active site: A NMR T1 paramagnetic relaxation study. *Biochemistry* **2005**, *44*, 14143–51.
- (65) Sligar, S. G.; Denisov, I. G. Understanding cooperativity in human p450 mediated drug-drug interactions. *Drug Metab. Rev.* **2007**, *39*, 567–79.
- (66) Kapelyukh, Y.; Paine, M. J.; Marechal, J. D.; Sutcliffe, M. J.; Wolf, C. R.; Roberts, G. C. Multiple substrate binding by cytochrome P450 3A4: Estimation of the number of bound substrate molecules. *Drug Metab. Dispos.* **2008**, *36*, 2136–44.
- (67) Denisov, I. G.; Frank, D. J.; Sligar, S. G. Cooperative properties of cytochromes P450. *Pharmacol. Ther.* **2009**, *124*, 151–67.
- (68) Denisov, I. G.; Shih, A. Y.; Sligar, S. G. Structural differences between soluble and membrane bound cytochrome P450s. *J. Inorg. Biochem.* **2012**, *108*, 150–8.
- (69) Sevrioukova, I. F.; Poulos, T. L. Understanding the mechanism of cytochrome P450 3A4: recent advances and remaining problems. *Dalton Trans.* **2013**, *42* (9), 3116–26.
- (70) Denisov, I. G.; Sligar, S. G. A novel type of allosteric regulation: functional cooperativity in monomeric proteins. *Arch. Biochem. Biophys.* **2012**, *519*, 91–102.