Journal of Molecular Graphics and Modelling 54 (2014) 90–99

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Effects of protein flexibility on the site of metabolism prediction for CYP2A6 substrates

Yayun Sheng, Yingjie Chen, Lei Wang, Guixia Liu, Weihua Li*, Yun Tang*

Shanghai Key Laboratory of New Drug Design, School of Pharmacy, East China University of Science and Technology, 130 Meilong Road, Shanghai 200237, China

ARTICLE INFO

Article history:
Received 7 August 2014
Received in revised form
25 September 2014
Accepted 30 September 2014
Available online 22 October 2014

Keywords:
CYP2A6
Site of metabolism
Flexibility
Molecular dynamics simulations
Molecular docking

ABSTRACT

Structure-based prediction for the site of metabolism (SOM) of a compound metabolized by human cytochrome P450s (CYPs) is highly beneficial in drug discovery and development. However, the flexibility of the CYPs' active site remains a huge challenge for accurate SOM prediction. Compared with other CYPs, the active site of CYP2A6 is relatively small and rigid. To address the impact of the flexibility of CYP2A6 active site residues on the SOM prediction for substrates, in this work, molecular dynamics (MD) simulations and molecular docking were used to predict the SOM of 96 CYP2A6 substrates. Substrates with known SOM were docked into the snapshot structures from MD simulations and the crystal structures of CYP2A6. Compared to the crystal structures, the protein structures obtained from MD simulations showed more accurate prediction for SOM. Our results indicated that the flexibility of the active site of CYP2A6 significantly affects the SOM prediction results. Further analysis for the 40 substrates with definite K_m values showed that the prediction accuracy for the low K_m substrates is comparable to that by ligand-based methods.

1. Introduction

Cytochrome P450s (CYPs) are a superfamily of heme-containing enzymes, which play important roles in the metabolism of a wide range of endogenous and exogenous compounds [1,2]. There are 57 CYP isozymes in humans, in which the families 1, 2, and 3 are responsible for the phase I metabolism of approximately 75% of clinically used drugs [3,4]. The efficacy and safety of these drugs are greatly affected by the CYP-mediated metabolism. Thus, an understanding of drug metabolism by CYPs would help medicinal chemists to modify and design drug candidates so as to modulate their metabolism in the early stage of drug discovery.

Identification of site of metabolism (SOM) is important for improving the bioavailability of a drug and avoiding possible toxic products, and thereby attracting much attention in the past years. However, SOM identification by experiments is resource-demanding and time-consuming. In recent years, a number of computational methods have been developed for SOM prediction. These methods can be divided into two categories: ligand-based [5–8] and structure-based [9,10]. The ligand-based methods, including SMARTCyp [11], RS-Predictor [12] and XenoSite [13], mainly use the data of CYP substrates. Recently, a new ligand-based method was developed based on 2D xenobiotic structural formulas and PASS prediction algorithm [14]. These ligand-based methods are fast and relatively accurate. However, they require calculating different descriptors for substrates and the results heavily depend on the quality of training sets. In addition, they hardly consider CYP–ligand interactions, which are thought important for the stereoselectivity and regioselectivity of substrate oxidation.

The structure-based methods predict SOM of substrates mainly by means of molecular docking. Huang et al. recently developed a program called DR-Predictor for SOM prediction [15], which incorporates flexible docking-derived information and multiple-instance-learning methods. The structure-based methods can fully consider the effect of CYP enzymes on substrate binding. However, the high flexibility of the CYP active site is a huge challenge for accurate prediction of the appropriate substrate binding modes and thus SOM of the substrate. Hritz et al. employed docking combined with molecular dynamics (MD) simulations for investigating the impact of the plasticity and flexibility on SOM prediction of CYP2D6 [2]. Perić-Hassler et al. used multiple ligand and protein conformations from MD simulations for predicting the binding affinity of CYP2D6 [16]. Another similar work performed by Moors et al. used the ensemble docking by multiple CYP2D6 conformations to account for the impact of the flexibility on the SOM prediction [17]. More recently, Hayes et al. conducted an ensemble-docking experiment

* Corresponding authors. Tel.:+86 21 64251052; fax: +86 21 64253651.
E-mail addresses: whli@ecust.edu.cn (W. Li), ytang234@ecust.edu.cn (Y. Tang).

http://dx.doi.org/10.1016/j.jmgm.2014.09.005
1093-3263/

of 195 CYP3A4 substrates with induced fit docking to elucidate the substrate promiscuity of CYP3A4 [18]. These studies indicate that the choice of the protein conformation for docking is very important for accurate SOM prediction. Their results also suggest that the flexibility of the CYP active site plays a vital impact on docking performance and using multiple protein conformations can increase the accuracy of SOM prediction.

As a hepatic enzyme, CYP2A6 is a member of CYP2A subfamily and can bioactivate numerous xenobiotics, such as drugs, procarcinogens, carcinogens, and many environmental contaminants [19]. Compared with other drug-metabolizing CYPs, CYP2A6 has a small and relatively rigid active site [20,21]. In general, proteins with large active-site flexibility or malleability have much impact on the docking-based SOM prediction [2,21]. Whether the small and rigid active site of CYP2A6 has the similar effect on SOM prediction of substrates remains unclear. In the present work, to investigate the effect of the flexibility of the active site residues of CYP2A6 on the SOM prediction of its substrates, MD simulation and molecular docking were used to predict SOM of 96 CYP2A6 substrates. By comparing the results from crystal structures and from different conformations generated by MD simulations, how the flexibility of the CYP2A6 active site affects the substrate SOM prediction is explored here.

2. Materials and methods

2.1. Data collection and preparation

Ninety-six CYP2A6 substrates were collected from literature [15]. The molecular weights of these compounds range from 74.08 to 586.18 kD. Their 2D structures are presented in Fig. S1 (see Supporting Information) and the experimentally determined SOMs are indicated. All the 3D structures were downloaded from the PubChem database in SDF format [22]. Each structure was visually checked to avoid any error. After minimization with OPLS 2005 force field by MacroModel (Schrödinger, LLC, New York, NY) [23], these substrates were converted into MOL2 format.

Currently, several crystal structures of CYP2A6 complexed with different ligands are available in Protein Data Bank (PDB) [24]. From these, eight were selected for this study, including 2FDU, 2FDV, 2FDW, 2FDY [25], 1Z10, 1Z11 [20], 3T3R [26], and 4EJJ [27] (PDB code). Each structure contains four asymmetric 2A6 monomers (namely chains A, B, C, and D), and hence was separated into four files, each with one monomer for subsequent docking.

2.2. Self-docking and cross-docking based on CYP2A6 crystal structures

To select an appropriate crystal structure for docking prediction, self-docking and cross-docking were carried out. All docking simulations were performed with GOLD version 5.2 [28]. The Chemscore scoring function parameterized for heme-containing proteins was used to rank the docking poses, which was proven to outperform Goldscore scoring function in the previous studies [1,2,29]. All protein structures used in this section were converted into MOL2 format. The standard Tripos atom and bond types were assigned for the protein residues. The original ligands were extracted from the corresponding crystal structures and saved as MOL2 format. Every ligand was docked back into the original CYP2A6 protein to reproduce its crystallized binding mode and then docked into the other seven crystal structures to evaluate the performance of the crystal structure recognizing other crystallized ligands.

The Fe atom of the heme group was defined as the center of the binding site and the residues within 15.0 Å from this center point were included. Fifty solutions were output for each docking. All docking poses were ranked according to the Chemscore values. The crystallized ligand reference file was used to execute automated RMSD calculations against GOLD solutions in each docking. All results were evaluated by the RMSD values. The single CYP2A6 crystal structure with relatively lower RMSD values was considered to be more appropriate to predict SOM. Two CYP2A6 crystal structures that exhibited the best performance were selected for the SOM prediction of all 96 substrates and for further MD simulations.

2.3. MD simulations

To evaluate the impact of protein flexibility on SOM prediction, MD simulations of the ligand bound CYP2A6 complex structures were performed on the selected proteins from self-docking and cross-docking.

The protonation state of the residues in each complex was calculated using PROPKA [30]. According to the calculation results, His84, His229, His254, and His415 were fully protonated at both nitrogen atoms; His72, His328, and His477 were protonated at the ε nitrogen; and all other histidines were protonated at the δ nitrogen. In addition, Glu448 was set to its protonation state according to the calculated value (p*Ka* = 9.34).

All crystallization water molecules in the selected CYP2A6 crystal structures were kept. The initial models were pretreated using xleap in AMBER12 [31]. The treated models were solvated with water in a truncated octahedron periodic box, in which the distance between the box edges and the protein was 10.0 Å. The TIP3P model was used for water. The systems were neutralized by adding counterions.

All MD simulations were performed with the sander module of AMBER12. The AMBER ff12SB all atom force field was used for the proteins and the general AMBER force field (GAFF) was used for the substrate molecules. The detailed procedure for MD simulation was adopted as described in our previous work [32]. Briefly, energy minimization with decreasing the force constant restraints to protein atoms was carried out to relax the system. In each minimization, 1000 steps with the steepest descent method followed by 2000 steps with the conjugated gradient method were employed. After the minimizations, the systems were heated gradually from 0 to 300 K over 60 ps under the NVT ensemble condition and then equilibrated at 300 K for 50 ps. Finally, 12 ns unrestrained MD simulations at 1 atm and 300 K under the NPT ensemble condition were conducted for the CYP2A6-ligand complex system. During the MD simulations, the SHAKE algorithm [33] was used to fix all bonds including hydrogen atoms. The long-range electrostatic interactions were treated by using the Particle Mesh Ewald method [34]. Non-bonded interactions were calculated with a cutoff distance of 10.0 Å, and every time step was set to 2 fs. Interaction energies and coordinates were stored every 1.0 ps during all MD simulations.

2.4. Docking for SOM prediction

The SOM predictions of 96 substrates were first carried out based on the above selected two crystal structures and on the snapshots from MD simulations. Ten snapshots were extracted from last 2 ns of each selected protein. These snapshots were subjected to a 300-step minimization and then prepared using Protein Preparation Wizard in Maestro 9.3. Details were similar to self-docking and cross-docking. Two docking trials were done. One is using the heme iron as the center of binding site, and the other is using the point with 5.0 Å above the heme iron as the center of binding site. According to the trial results, the point above 5.0 Å from heme iron atom was chosen as the center point of binding site here and the radius was set to 20.0 Å to include the key residues of the active site. For each substrate, 50 GA runs were performed with automatic GA parameter settings. The population size was set to 100

and the number of operations was 100,000. To explore how the protein flexibility impacts on prediction, the 96 substrates were also docked into the two crystal structures for comparison.

2.5. Analysis of docking results

For SOM prediction of the 96 substrates, the choice of criterion for determining an accurate SOM prediction is important. In the previous study [2], a relatively wide criterion was applied, in which the prediction was considered to be correct as long as the experimentally observed SOM is within 6.0 Å from the heme iron atom. Yet, this might have a little risk when this criterion is applied to SOM prediction for a new substrate. In general, the first ranked pose should correspond to the binding pose with the highest binding affinity. Unfortunately, the scoring functions are not reliable enough to accurately predict the top1 ranked pose. Moreover, the top1 ranked pose with the highest score is not necessarily the only one catalytically active pose. Consequently, more top *k* ranked poses should be considered to increase accuracy, where *k* is usually 1, 2, or 3 [35].

A previous study by Hritz et al. used a distance criterion to judge if the prediction is correct or not. In their criterion, a docking pose was considered to be correct if the known SOM is located within 6.0 Å of the heme iron. This criterion is relatively wide. In fact, the intrinsic reactivity of a substrate is also important in addition to the distance between Fe and the SOM. In this work, both factors, the distance and the potential reactivity, were considered. The docking pose was considered to be accurate prediction if the known SOM points toward and locates within 6.0 Å from the heme iron. Based on these, the first 3 ranked poses, termed top1, top2 and top3, in all 50 solutions were statistically analyzed according to the above described criterion. The prediction accuracy is the percentage of correct prediction for all 96 substrates. For comparison, Hritz et al.'s criterion (termed top1 plus 6.0 Å in this work) was also used.

3. Results

3.1. Self-docking and cross-docking

Self-docking and cross-docking were conducted against 32 CYP2A6 crystal structure models (eight complexes, each with four chains). To evaluate the docking performance, the root-mean-square deviation (RMSD) values of GOLD outputs from the crystallized ligands were computed. The average and lowest RMSD values for all dockings are listed in Table 1 and Table S1 (see Supporting Information). Among eight wild type crystal structures, the resolutions of 2FDV, 2FDU, 1Z10, and 2FDY are relatively better (1.65 Å, 1.85 Å, 1.90 Å, and 1.95 Å, respectively).

For self-docking, all four chains of 4EJJ and 2FDU_B have larger average RMSD values (about 4.22–4.45 Å for 4EJJ and 3.63 Å for 2FDU_B) compared to other structural models. The average RMSD values of 1Z10_A, 2FDU_C, 3T3R_A, and 3T3R_D are 1.16, 1.68, 1.09, and 1.10 Å, respectively. The highest RMSD values by self-docking are seen for four chains of 4EJJ, which are from 4.18 to 4.41 Å (Table S1). The other RMSDs are relatively low (0.16–1.07 Å). The RMSD values of 2FDU_A (0.28 Å) and 2FDY_D (0.33 Å) indicate a better reproduction of the crystallized ligand binding modes.

Table 1
The average RMSD values for each chain in self-docking and cross-docking.

PDB code (resolution)	Chain	RMSD (Å)							
		Lig0	Lig1	Ligu	Ligv	Ligw	Ligy	Ligr	Ligj
1Z10 (1.90 Å)	A	**1.16**[a]	0.95	2.71	1.13	1.57	1.42	1.61	4.35
	B	**0.54**	0.86	3.23	1.01	1.24	1.87	1.67	4.51
	C	**0.80**	0.82	5.53	2.37	1.14	1.87	1.82	4.32
	D	**0.90**	0.67	4.25	3.29	1.38	2.04	1.90	4.44
1Z11 (2.05 Å)	A	0.98	**0.64**	4.59	1.78	1.41	1.32	1.67	4.37
	B	0.74	**0.54**	5.28	1.82	1.26	1.75	1.49	4.53
	C	0.90	**0.72**	5.67	1.15	1.07	1.84	1.77	4.29
	D	1.03	**0.77**	5.76	2.05	1.65	2.13	2.12	4.53
2FDU (1.85 Å)	A	0.75	1.41	**0.94**	0.39	0.79	1.30	1.35	4.20
	B	0.72	0.80	**3.63**	0.60	0.97	1.88	1.84	4.45
	C	0.84	1.11	**1.68**	0.58	0.94	1.89	1.84	4.39
	D	0.74	1.17	**0.93**	0.76	0.96	1.88	1.94	4.42
2FDV (1.65 Å)	A	0.96	1.41	3.54	**0.37**	0.88	1.51	1.78	4.35
	B	1.44	0.81	3.88	**0.39**	0.93	1.84	1.83	4.42
	C	0.92	1.24	5.65	**0.42**	1.00	1.85	1.86	4.39
	D	0.96	1.15	5.62	**0.45**	0.87	2.19	1.97	4.39
2FDW (2.05 Å)	A	0.86	1.32	5.78	0.61	**0.99**	1.51	1.81	4.39
	B	0.82	0.88	5.96	0.79	**0.76**	1.79	1.76	4.54
	C	0.84	0.99	5.90	0.81	**0.93**	1.59	1.68	4.28
	D	0.92	0.94	5.83	1.00	**0.48**	1.88	1.72	4.41
2FDY (1.95 Å)	A	1.22	0.96	5.53	2.74	1.32	**0.58**	0.95	4.37
	B	1.73	1.72	2.87	0.65	1.11	**0.80**	1.18	4.55
	C	0.84	1.02	5.27	1.23	1.03	**0.72**	1.55	4.26
	D	0.22	0.17	2.68	0.90	0.94	**0.60**	1.65	4.44
3T3R (2.40 Å)	A	1.01	2.08	5.47	4.02	1.38	1.24	**1.09**	4.38
	B	0.84	2.25	5.35	3.70	1.40	1.49	**0.71**	4.42
	C	0.91	1.19	5.51	4.61	1.22	1.33	**0.52**	4.38
	D	1.24	3.16	5.43	4.16	1.39	1.14	**1.10**	4.40
4EJJ (2.30 Å)	A	1.82	1.16	5.47	5.20	1.25	1.53	1.44	**4.36**
	B	0.79	1.72	5.54	4.71	1.15	1.47	1.77	**4.22**
	C	3.66	1.76	5.03	3.22	1.46	1.52	2.22	**4.26**
	D	1.49	1.12	5.50	4.92	1.32	1.33	1.33	**4.45**

[a] The average RMSD values for self-docking results are shown in **bold**.

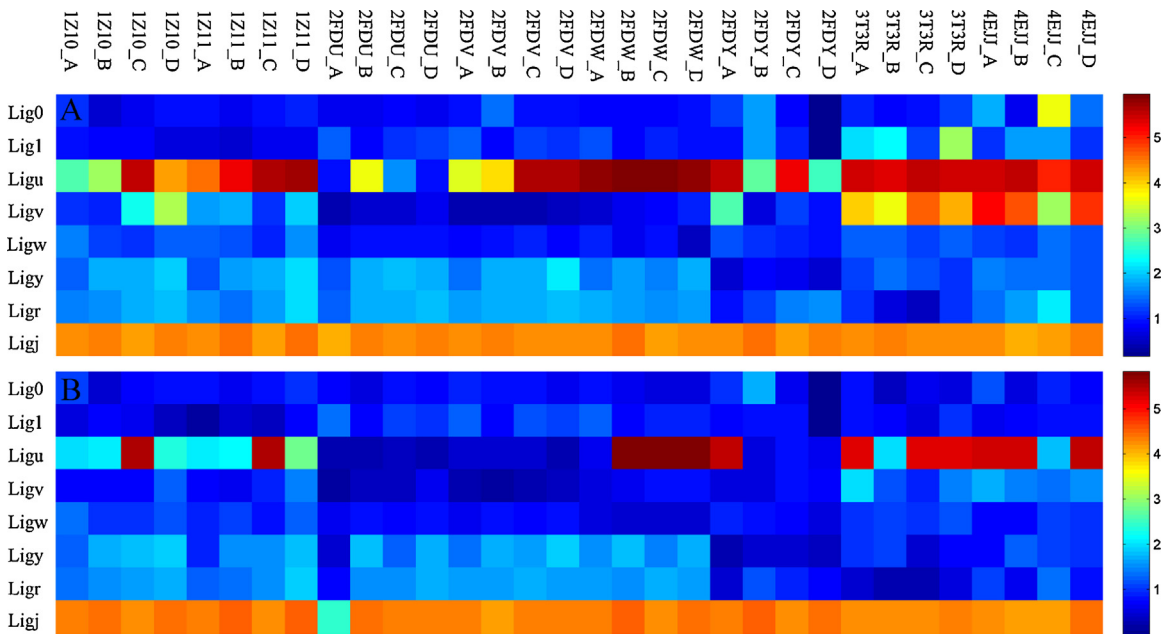

Fig. 1. The heat map for the average RMSD values (A) and the lowest RMSD values (B) in self-docking and cross-docking.

For cross-docking, both the average and lowest RMSD values have large range from 0.17 to 5.96 Å and 0.03 to 5.83 Å, respectively. The average RMSDs of 2FDU_A vary from 0.39 to 4.20 Å, while the lowest RMSDs vary from 0.19 to 2.40 Å. Especially when nicotine (Ligj) was docked into 2FDU_A, the lowest RMSD value is 2.40 Å, which is much lower than those when Ligj was docked into other molecules (4.18–4.54 Å). For 2FDY_D, the seven average RMSDs range from 0.17 to 4.44 Å, while the lowest RMSDs vary from 0.03 to 4.41 Å. When the original ligand of 2FDU (Ligu) was docked into this protein, the lowest and average RMSD values are 0.58 and 2.68 Å, respectively. These values are lower than those when Ligu was docked into other structures.

To evaluate the docking performance of self-dockings and cross-dockings more visually, the average and lowest RMSD values in Table 1 and Table S1 were used to produce two heat maps, as shown in Fig. 1. The heat maps reflect that 2FDU_A and 2FDY_D have better docking results than other models. Consequently, 2FDY_D and 2FDU_A were chosen for further MD simulations and the SOM prediction of 96 substrates.

3.2. MD simulations

Twelve nanosecond MD simulations were carried out for the two CYP2A6 complex systems (2FDU_A and 2FDY_D) in order to look into the effects of the flexibility of active site residues on the SOM prediction. The RMSD values relative to the initial structure were monitored during the MD simulations, which are shown in Fig. 2. For the 2FDU_A complex system, the RMSD values of the backbone atoms maintained ~1.0 Å after 6.0 ns. In contrast, the RMSD values of the backbone atoms of 2FDY_D complex system maintained ~1.2 Å after 9.0 ns. The slight fluctuation of RMSDs after 10.0 ns suggests that the protein structure is equilibrated sufficiently. 10 snapshots were extracted every 0.2 ns from the last 2 ns of MD trajectories for each complex system.

After MD simulations, the active site volumes were calculated by the program Voidoo [36] with a 1.4 Å probe and a grid mesh of 0.4 Å as described [27]. The active site volumes of the 10 snapshot structures for 2FDY_D range from 320.7 to 400.9 $Å^3$, and 306.0 to 346.1 $Å^3$ for the 10 snapshot structures of 2FDU_A (see Table S2 in Supporting Information). In addition, the difference between the snapshot structures from MD simulations and the crystal structures was analyzed. A superimposition of the crystal structure 2FDY_D and its 11.0 ns snapshot structure reveals that the active site residues have an obvious positional change (Fig. 3A). Residues F107, F209 and F480 undergo a remarkable shift. The side chains of F209 and F480 move out of the active site, which results in a more open active site relative to the crystal structure. The side chain of F107 has a shift close to the active site. In addition, F111, T305, and I366 also have a slight shift away from the active site. The same phenomena were seen for 2FDU_A. Fig. 3B shows the superimposition of 10.2 ns snapshot structure and the crystal structure of 2FDU_A. F209 and F107 move away from the active site and F111 has a slight shift away from the pocket. In contrast, F480 and F118 are close to the ligand and stabilize the binding of the ligand. N297 also forms a hydrogen bond with the ligand.

3.3. SOM prediction of 96 substrates

For each system, 96 substrate molecules were docked into the 10 protein snapshot structures, which are taken from MD simulations at 10.2, 10.4, 10.6, 10.8, 11.0, 11.2, 11.4, 11.6, 11.8, and 12.0 ns, respectively. The corresponding docking scores were calculated in each docking. All ranked top3 poses were analyzed according to the criterion described in Section 2. Table 2 shows the prediction results for all protein structures. When only

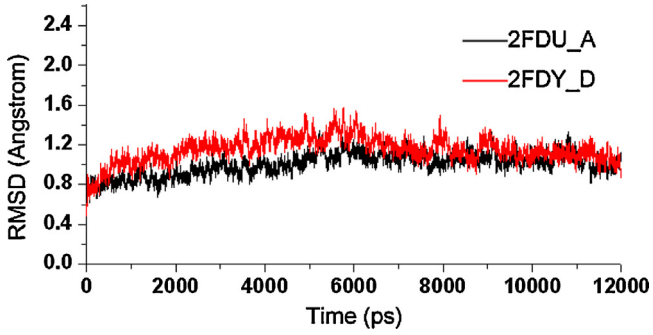

Fig. 2. RMSD with respect to simulation time for 12 ns MD simulations for backbone atoms of 2FDU_A and 2FDY_D.

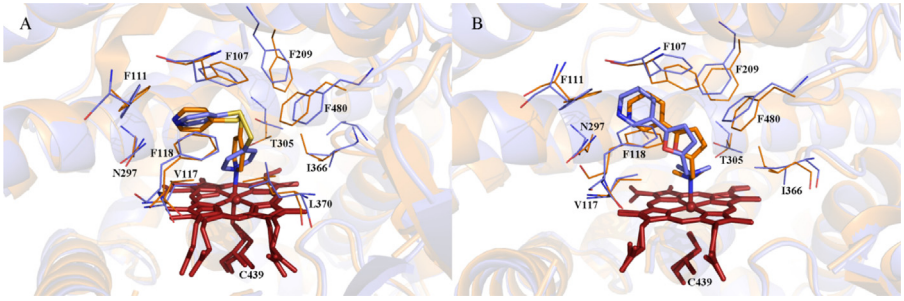

Fig. 3. Superposition of the crystal structures (orange) and 11.0 ns (A) and 10.2 ns (B) protein conformation (blue) of 2FDY_D and 2FDU_A. The ligands and heme are shown with sticks. The residues in active site are shown with lines.

considering the ranked top1 docking pose, the prediction accuracies range from 40.6% to 57.3% for the 10 snapshots of 2FDY_D (Table 2). They have 2.1–10.4% increase when the top2 docking poses were considered. The highest value is 63.5%. When considering the ranked top3 docking poses, the highest and lowest prediction accuracies are 65.6% and 49.0%, respectively. Among the 10 different protein snapshot structures, the structure at 11.0 ns achieves the best SOM prediction accuracies for all top1, top2, and top3 docking poses, which are 57.3%, 63.5%, and 65.6%, respectively.

In contrast, the prediction accuracies for 2FDU_A are lower than those for 2FDY_D (Table 2). The prediction accuracies for the 10 2FDU_A snapshots are 36.5–51.0%, 39.6–53.1%, and 44.8–56.3% for the top1, top2, and top3 ranked poses. The highest prediction accuracies within top1, top2, and top3 ranked docking poses are 51.0%, 53.1%, and 56.3%, respectively, when the protein structure at 10.2 ns was used. The prediction accuracies for all protein structures of this system are 54.2–68.8% when the known SOM within 6.0 Å from the heme iron atom within top1 ranked docking pose was considered. When considering this previous criterion, the highest prediction accuracy for 2FDY_D is slightly higher (~1.0%) than that of 2FDU_A. Overall, the prediction accuracies for 2FDY_D are higher than that for 2FDU_A at the same criterion.

When the original crystal structure was used for docking prediction, the prediction accuracies of 2FDY_D in top1, top2, top3 ranked docking poses are 32.3%, 34.4%, and 37.5%, respectively (Table 2). All of them are lower than those of MD simulation snapshot structures. In addition, the previous criterion (top1 plus 6.0 Å) was also applied here. The snapshot structures accurately predicted SOM within the top1 ranked docking pose for 59.4–69.8% of the substrates. When this wide criterion was used, the prediction accuracies are relative higher both for 10 MD protein conformations and crystal structures. For the crystal structure 2FDU_A, the top1, top2, and top3 prediction accuracies are about 24.0%, 25.0%, and 26.0%, respectively, which are lower than that for 2FDY_D within corresponding ranked poses.

In summary, the prediction accuracies for the refined MD simulation protein structures are higher than those of crystal structures for both systems.

4. Discussion

4.1. Selection of 2FDU_A and 2FDY_D

There are eight crystal structures of CYP2A6 bound with different ligands available in PDB. The active site volumes of these crystal structures range from 238.9 $Å^3$ to 287.8 $Å^3$ and the RMSD values of the backbone atoms between these structures are small (see Table S3 in Supporting Information). To select the appropriate CYP2A6 structure(s) for predicting SOM of CYP2A6 substrates, self-docking and cross-docking were first performed using the crystallized ligands. The crystal structures that are able to accommodate all the crystallized ligands and reproduce the binding mode in the original crystal structures were chosen for further docking prediction. This idea has been used in a previous study [15], in which the redocking experiments based on crystal structures were performed to ascertain whether Autodock Vina was suitable for evaluating CYP binding modes.

In general, the lower RMSD values suggest better reproduction of crystal binding modes for original ligands. Two CYP2A6 structures, 2FDU_A and 2FDY_D, exhibited the lowest RMSD values in self-docking, which are 0.28 and 0.33 Å, respectively. Fig. 4 shows the superimposition of the docking pose with the lowest RMSD value with the crystal ligand in 2FDU_A and 2FDY_D. The lower average RMSD values for the two structures were also observed in the cross-dockings (see Table 1). Overall, the self-docking and crossing docking results indicate that 2FDY_D and 2FDU_A are more suitable for the SOM prediction when compared to other CYP2A6 crystal structures.

4.2. Impact of protein flexibility on SOM prediction

The prediction accuracies of the 2FDY_D crystal structure for the top1, top2, and top3 ranked docking poses range from 32.3% to 37.5%, which are lower than most of the results of the refined protein structures from MD simulations. Among the 10 snapshots, the 11.0 ns protein structure achieves the highest prediction accuracy (65.6% for top3). As analyzed in Section 3, the obvious difference was seen for several residues in the active site. A volume analysis

Table 2
Prediction accuracies of SOM for the 10 snapshot structures from MD simulation and the crystal structure of 2FDY_D and 2FDU_A.

Chain	Prediction Accuracy (%)	10.2	10.4	10.6	10.8	11.0	11.2	11.4	11.6	11.8	12.0	Crystal structure
2FDY_D	Top1	54.2	40.6	54.2	57.3	57.3	55.2	47.9	44.8	53.1	45.8	32.3
	Top2	56.3	45.8	58.3	62.5	63.5	58.3	51.0	55.2	56.3	47.9	34.4
	Top3	57.3	50.0	59.4	62.5	65.6	60.4	51.0	59.4	58.3	49.0	37.5
	Top1 + 6.0 Å	68.8	65.6	65.6	68.8	65.6	69.8	66.7	59.4	67.7	60.4	63.5
2FDU_A	Top1	51.0	46.9	36.5	46.9	50.0	47.9	47.9	44.8	39.6	43.8	24.0
	Top2	53.1	51.0	39.6	50.0	52.1	49.0	50.0	51.0	45.8	51.0	25.0
	Top3	56.3	54.2	44.8	54.2	53.1	50.0	50.0	52.1	46.9	52.1	26.0
	Top1 + 6.0 Å	65.6	65.6	64.6	68.8	63.5	62.5	59.4	54.2	56.3	62.5	62.5

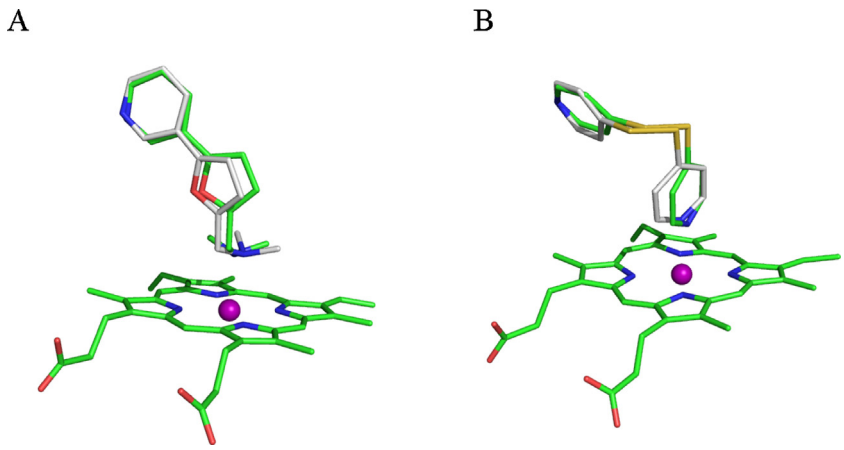

Fig. 4. Self-docking results for the two best crystal structures: 2FDU_A (A); 2FDY_D (B). The ligand and heme in the crystal structures are shown in the green stick, and the first pose from self-docking is shown in the gray stick. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

between the 11.0 ns snapshot structure and the crystal structure reveals that the volume of active site becomes larger after MD simulation (400.9 Å^3 vs 267.4 Å^3). Thus, the larger active site can accommodate more substrates with various sizes. This led to a better docking performance for this snapshot structure. Similar to the phenomena seen for 2FDY_D, the prediction accuracies of the 10 snapshots of 2FDU_A are also higher than those of the crystal structure, in which 10.2 ns snapshot structure achieved the best prediction accuracies. The volume analysis was also performed between the 10.2 ns snapshot structure of 2FDU_A and crystal structure. The volume of the active site for 10.2 ns snapshot structure (306.0 Å^3) is also larger than that of the crystal structure (262.8 Å^3).

Comparing the prediction accuracies of 2FDY_D and 2FDU_A, we found that the performance of 2FDY_D is better than that of 2FDU_A for both the MD snapshot structures and the crystal structure. For the crystal structure, the prediction accuracies of 2FDY_D for the top1, top2, and top3 are about 8.3%, 9.4%, and 11.5% higher than those of 2FDU_A, respectively. For the MD snapshot structures, the highest prediction accuracies of 2FDY_D are about 6.3%, 10.4%, and 9.3% higher than those of 2FDU_A for the top1, top2, and top3 ranked docking poses, respectively. Fig. 5A and B shows the superimpositions of the two crystal structures as well as the two snapshots with the highest prediction accuracies. As seen in Fig. 5A, both F209 and F118 in 2FDU_A have their side chains to point to the active site, whereas the two same residues in 2FDY_D are away from the active site. These result in a larger active site in 2FDY_D than 2FDU_A. The benzene rings of F107 and F111 in 2FDY_D have a slight rotation away from the active site. These make a larger binding pocket for 2FDY_D over 2FDU_A. A similar phenomenon is also seen in the comparison of the 11.0 ns snapshot structure of 2FDY_D and the 10.2 ns snapshot structure of 2FDU_A (Fig. 5B). Similarly, the side chains of F209 have two different orientations in the two systems. F209 in 2FDU_A points to the binding pocket while it turns outward in 2FDY_D. In addition, F107, F111, F480, T305, and I300 have obvious shifts away from the active site of 2FDY_D. Compared to 2FDU_A, these differences lead a larger binding pocket to accommodate various substrates.

Our docking results showed that (a) the SOM prediction accuracies of protein structures from MD simulations are higher than those from crystal structures; (b) the SOM prediction of 2FDY_D is better than that of 2FDU_A. These suggested that the protein flexibility of CYP2A6 plays a determining role in the SOM prediction. Several studies have proved that protein flexibility has significant impacts on substrate binding and conformation sampling from MD simulations can improve the SOM prediction for CYPs. The use of MD conformational sampling for predicting metabolic

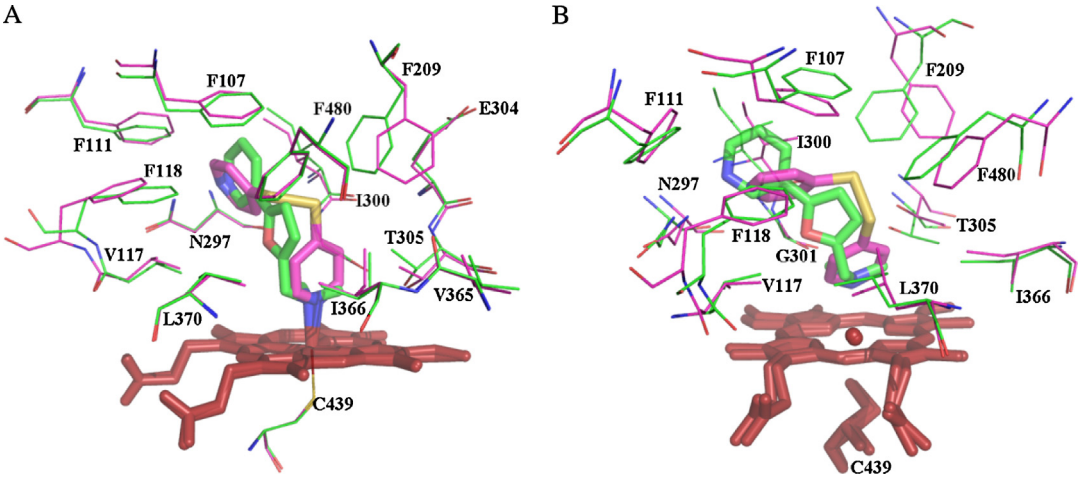

Fig. 5. (A) Superposition of the active sites of the two crystal structures. Purple: 2FDY_D chain; Green: 2FDU_A chain. (B) Superposition of the active sites of the refined MD simulation protein structures. Purple: 11.0 ns of 2FDY_D chain; Green: 10.2 ns of 2FDU_A chain. The ligands and heme are shown with sticks. The residues in active site are shown with lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Classifying the substrates according to K_m values. Prediction accuracies for low K_m (1–20 μM), medium K_m (20–100 μM), medium-high K_m (100–500 μM), high K_m (>500 μM) substrates within top1, top2, top3, and top1 plus 6.0 Å.

K_m Class	Number	Prediction accuracy			
		Top1	Top2	Top3	Top1 + 6.0 Å
Low	19	84.2	94.7	94.7	84.2
Medium	8	75.0	75.0	87.5	75.0
Medium-high	10	60.0	60.0	60.0	70.0
High	3	66.7	66.7	66.7	66.7
Cumulative stastics[a]	40	75.0	80.0	82.5	77.5
Cumulative stastics[b]	81	61.7	69.1	70.4	69.1

[a] The whole 40 substrates with available K_m values produce the modified cumulative statistics.
[b] The 15 substrates with a little role in metabolism were removed to produce modified cumulative statistics.

regioselectivities was pioneered by Loida et al. [37]. Hritz et al. used a combination of MD simulation and molecular docking to explore the impact of protein flexibility of the CYP2D6 structures on docking accuracy [2]. In another study by Danielson et al. [38], MD simulations were used to generate an ensemble of diverse protein structures used for SOM prediction. Recently, Perić-Hassler et al. used MD simulations to create multiple ligand and protein conformations to predict CYP2D6 binding affinity [16]. Stjernschantz et al. found that the ligand–protein binding affinity predictions for CYP2C9 can be improved by using multiple binding modes from MD simulations [39]. Vasanthanathan et al. used the LIE models to predict the binding free energy for CYP1A2 ligands [40]. These findings together with our results prove that MD simulation conformational sampling incorporated molecular docking can achieve better performance in the SOM prediction of CYPs.

4.3. Comparison with ligand-based methods

Several ligand-based methods have been developed and applied to the SOM predictions on the important CYP enzymes. DR-Predictor can correctly predict the SOMs with an accuracy of 83% for 100 CYP2A6 substrates if considering the top2 ranked positions [15]. Another ligand-based program, XenoSite, can correctly predict 85.7% of 105 CYP2A6 substrates within top2 ranked positions [13]. Our results showed that the highest prediction accuracies for 96 substrates within the top1, top2, and top3 ranked docking poses are 57.3%, 63.5%, and 65.6%, respectively. Judging from these values, we have to admit that the prediction rates of our results are not particularly high when compared with the aforementioned ligand-based methods. However, a deep analysis may lead to a different perspective.

Since all the 96 substrates used in this study were taken from the previous study [12], a detailed examination of the original literature reveals that CYP2A6 has a significantly differential role in the oxidation for these substrates. Among 96 substrates, 40 substrates were found to have definite K_m values by CYP2A6 oxidation and 15 substrates are relevant to CYP2A6 to a small extent. On the basis of this fact, the prediction results were statistically reanalyzed. The 40 substrates with definite K_m values (indicated by '&' in Supporting Information) were separated into four classes: low (1–20 μM), medium (20–100 μM), medium-high (100–500 μM), and high (>500 μM) based on the K_m values reported in the literature [12]. Table 3 presents the prediction accuracies based on the 11.0 ns snapshot structure of 2FDY_D (the best protein template) for each class, the whole 40 substrates with available K_m values, and the library removing the 15 substrates. The 40 substrates with available K_m values are composed of 19 low K_m substrates, eight medium K_m substrates, 10 medium-high K_m substrates, and three high K_m substrates. Although the prediction accuracies for the whole 96 substrates are worse than those predicted by the ligand-based methods, the prediction for the 40 available K_m values substrates (80.0%) is comparable to that by the ligand-based methods. The prediction accuracies for the 19 low K_m substrates within top1, top2, and top3 ranked poses are 84.2%, 94.7%, and 94.7%, respectively. When the previous criterion (top1 plus 6.0 Å) was used in the statistics of the library, the prediction accuracy for the low K_m substrates (84.2%) is also higher than that for medium K_m substrates (75.0%), medium-high K_m substrates (70.0%) and high K_m substrates (66.7%). When the 15 substrates with less relevant to CYP2A6 were removed from the whole 96 substrates, the prediction accuracies within top1, top2, top3 and top1 plus 6.0 Å have a slight increase. In addition, we found that the SOM prediction is more accurate for low K_m substrates than for high K_m substrates. In general, a substrate with a low K_m value has a higher affinity with CYPs. Currently it remains a challenge for the structure-based method to accurately predict the SOM for the substrates with high K_m. To improve the prediction for the high K_m substrates, other strategies may be required, such as using a combination of ligand-based and structure-based methods [41] or more advanced methods such as umbrella sampling [42].

For comparison, the 96 substrates were also predicted by the online web server of SMARTCyp (version 2.4.2) [11], a reactivity-based model for human cytochrome P450-based metabolism [43]. For all 96 substrates, SMARTCyp can achieve a prediction accuracy of 85.4% (82/96 = 85.4%) within top2, which is better than that of our method (50–60%). A total of 14 substrates (indicated by

Table 4
The number of substrates predicted correctly among the 14 substrates that cannot be accurately predicted by SMARTCyp in the 10 protein conformations of 2FDY_D and 2FDU_A.

Chain	Number of correct prediction	10.2	10.4	10.6	10.8	11.0	11.2	11.4	11.6	11.8	12.0
2FDY_D	Top1	5	4	3	5	5	6	4	4	6	3
	Top2	5	4	5	6	8	8	5	5	8	4
	Top3	6	4	5	6	8	8	5	6	8	4
	Top1 + 6.0 Å	9	6	7	8	9	8	9	5	9	6
2FDU_A	Top1	5	6	2	5	3	6	5	6	2	5
	Top2	6	6	2	5	3	6	5	6	3	5
	Top3	6	6	2	7	3	6	5	6	3	5
	Top1 + 6.0 Å	6	6	4	6	5	6	6	6	3	6

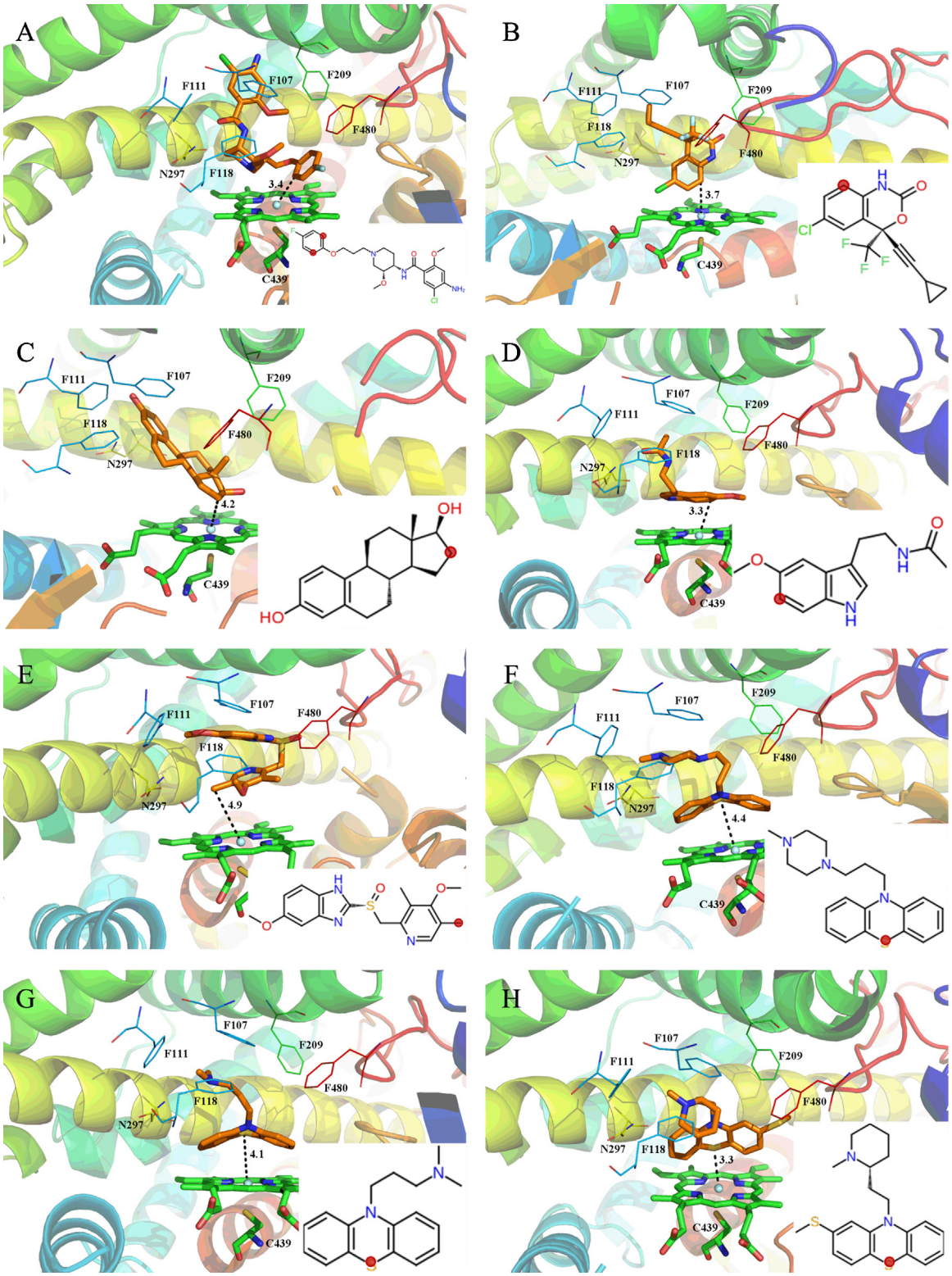

Fig. 6. Docking results for the eight substrates: (A) cisapride; (B) efavirenz; (C) estradiol; (D) melatonin; (E) omeprazole; (F) perazine; (G) promazine; and (H) thioridazine. Residues in active site are shown as lines. Substrate and Heme are shown as sticks. Distance between SOM and heme iron atom is also indicated.

'*' in Fig. S1) cannot be accurately predicted the known SOMs by SMARTCyp within top2 rank-positions. However, docking based on the MD simulation snapshot structures can accurately predict some of 14 substrates. The numbers of substrates predicted correctly among these 14 substrates in the two systems are present in Table 4. 2FDY_D is more appropriate for SOM prediction of the 14 substrates. The highest prediction accuracy within top2 is 57.1% (8/14 = 57.1%). Among all snapshot structures, 11.0 ns, 11.2 ns, and 11.8 ns structures of 2FDY_D can achieve this performance. Taking 11.0 ns protein structure as an example, eight of the whole 14 substrates can be accurately predicted. They are cisapride, efavirenz, estradiol, melatonin, omeprazole, perazine, promazine,

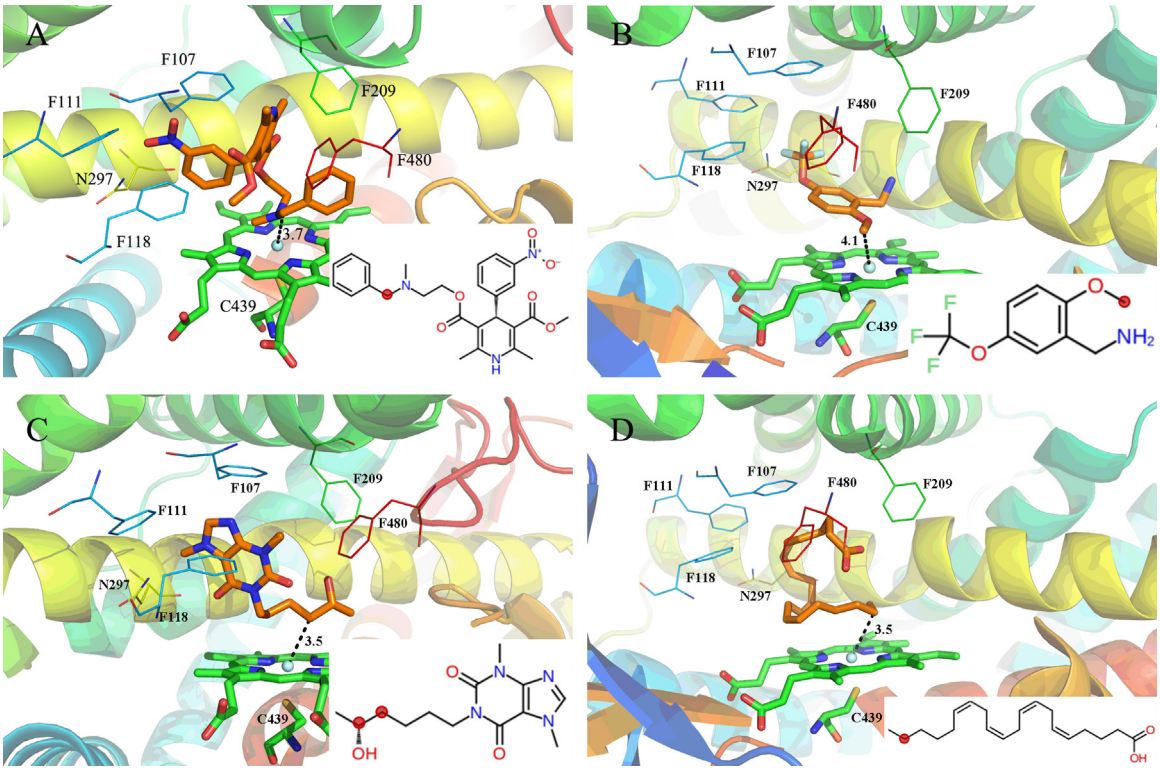

Fig. 7. Docking results for the four substrates: (A) nicardipine; (B) cp_677_623; (C) lisofylline; and (D) arachidonic_acid. Residues in active site are shown as lines. Substrate and heme are shown as sticks. Distance between SOM and heme iron atom is also indicated.

and thioridazine, which have average molecular weight of about 328.27 kD. As seen from Fig. 6, the known SOMs of these substrates are close to the heme iron and the distances between the SOMs and the iron are within 6.0 Å. Cisapride, estradiol, and melatonin can form a hydrogen bond with N297 in the active site. In addition, the other four of 14 substrates, nicardipine, cp_677_623, lisofylline, and arachidonic_acid, can be accurately predicted by the snapshot structure at 11.8 ns (see Fig. 7). However, progesterone and cp_125_611 cannot be correctly predicted for all protein structures.

Our results indicate that the structure-based method by combining docking and MD simulations can achieve the accuracy comparable to the ligand-based methods when carefully considering the substrate data sources. In contrast to the ligand-based methods, the structure-based methods do not require calculating any descriptors and, in principle, can be applied to prediction of any new molecules. In addition, the structure-based methods can consider the effect of CYP active site residues on the SOM prediction. This is especially important for the stereoselectivity of substrate oxidation.

5. Conclusions

Accurate prediction of SOM of compounds can help medicinal chemists to improve the efficacy of the drugs and reduce the toxicity of the drugs. In this work, to address the effect of active site residues of CYP2A6 on the SOM prediction of its substrates, molecular docking combined MD simulation was conducted for 96 CYP2A6 substrates.

Self-docking and cross-docking results indicated that 2FDY_D and 2FDU_A exhibited better performance on recognizing the crystallized ligands when compared to other crystal structures. A comparison of the prediction accuracies reveals that the prediction based on refine MD snapshot structures is more accurate than that based on the crystal structures. The difference in the prediction accuracy results from the different conformations of the active site residues of CYP2A6. These results suggest that the flexibility of the active site residues of CYP2A6 has a significant impact on SOM prediction of its substrates, regardless of the fact that the CYP2A6 active site is relative small and rigid. Although our structure-based prediction is less accurate than that by the ligand-based methods when using a comprehensive set of 96 substrates, some substrates that are difficult to be accurately predicted by SMARTCyp can be predicted well using our method. Further analysis shows that the prediction for the low K_m substrates is more accurate than for high K_m substrates. In addition, if the 40 substrates with available K_m values were predicted within top2 ranked positions, the prediction accuracy is comparable with that with the ligand-based methods. In future, the structure-based SOM prediction methods combined with other approaches such as ligand-based methods are expected to improve the accuracy for SOM prediction.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grants 81373328 and 81373329), the 863 Project (Grant 2012AA020308), and the Fundamental Research Funds for the Central Universities (Grant WY1113007).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2014.09.005.

References

[1] P. Vasanthanathan, J. Hritz, O. Taboureau, L. Olsen, F.S. Jørgensen, N.P. Vermeulen, C. Oostenbrink, Virtual screening and prediction of site of metabolism for cytochrome P450 1A2 ligands, J. Chem. Inf. Model. 49 (2009) 43–52.

[2] J. Hritz, A. de Ruiter, C. Oostenbrink, Impact of plasticity and flexibility on docking results for cytochrome P450 2D6: a combined approach of molecular dynamics and ligand docking, J. Med. Chem. 51 (2008) 7469–7477.
[3] F.P. Guengerich, Z.L. Wu, C.J. Bartleson, Function of human cytochrome P450s: characterization of the orphans, Biochem. Biophys. Res. Commun. 338 (2005) 465–469.
[4] F.P. Guengerich, Cytochrome, P450s and other enzymes in drug metabolism and toxicity, AAPS J. 8 (2006) E101–E111.
[5] Y.W. Hsiao, C. Petersson, M.A. Svensson, U. Norinder, A pragmatic approach using first-principle methods to address site ofmetabolism with implications for reactive metabolite formation, J. Chem. Inf. Model. 52 (2012) 686–695.
[6] K. Sato, Y. Yamazoe, Unimolecular and bimolecular binding system for the prediction of CYP2D6-mediated metabolism, Drug Metab. Dispos. 40 (2012) 486–496.
[7] M. Zheng, X. Luo, Q. Shen, Y. Wang, Y. Du, W. Zhu, H. Jiang, Site of metabolism prediction for six biotransformations mediated by cytochromes P450, Bioinformatics 25 (2009) 1251–1258.
[8] M.J. Sykes, R.A. McKinnon, J.O. Miners, Prediction of metabolism by cytochrome P450 2C9: alignment and docking studies of a validated database of substrates, J. Med. Chem. 51 (2008) 780–791.
[9] A. Tarcsay, R. Kiss, G.M. Keseru, Site of metabolism prediction on cytochrome P450 2C9: a knowledge-based docking approach, J. Comput. Aided Mol. Des. 24 (2010) 399–408.
[10] R. Santos, J. Hritz, C. Oostenbrink, Role of water in molecular docking simulations of cytochrome P450 2D6, J. Chem. Inf. Model. 50 (2010) 146–154.
[11] P. Rydberg, D.E. Gloriam, J. Zaretzki, C. Breneman, L. Olsen, SMARTCyp: a 2D method for prediction of cytochrome P450-mediated drug metabolism, ACS Med. Chem. Lett. 1 (2010) 96–100.
[12] J. Zaretzki, P. Rydberg, C. Bergeron, K.P. Bennett, L. Olsen, C.M. Breneman, RS-Predictor models augmented with SMARTCyp reactivities: robust metabolic regioselectivity predictions for nine CYP isozymes, J. Chem. Inf. Model. 52 (2012) 1637–1659.
[13] J. Zaretzki, M. Matlock, S.J. Swamidass, XenoSite: accurately predicting CYP-mediated sites of metabolism with neural networks, J. Chem. Inf. Model. 53 (2013) 3373–3383.
[14] A.V. Rudik, A.V. Dmitriev, A.A. Lagunin, D.A. Filimonov, V.V. Poroikov, Metabolism site prediction based on xenobiotic structural formulas and PASS prediction algorithm, J. Chem. Inf. Model. 54 (2014) 498–507.
[15] T.W. Huang, J. Zaretzki, C. Bergeron, K.P. Bennett, C.M. Breneman, DR-Predictor: incorporating flexible docking with specialized electronic reactivity and machine learning techniques to predict CYP-mediated sites of metabolism, J. Chem. Inf. Model. 53 (2013) 3352–3366.
[16] L. Perić-Hassler, E. Stjernschantz, C. Oostenbrink, D.P. Geerke, CYP 2D6 binding affinity predictions using multiple ligand and protein conformations, Int. J. Mol. Sci. 14 (2013) 24514–24530.
[17] S.L. Moors, A.M. Vos, M.D. Cummings, H. Van Vlijmen, A. Ceulemans, Structure-based site of metabolism prediction for cytochrome P450 2D6, J. Med. Chem. 54 (2011) 6098–6105.
[18] C. Hayes, D. Ansbro, M. Kontoyianni, Elucidating substrate promiscuity in the human cytochrome 3A4, J. Chem. Inf. Model. 54 (2014) 857–869.
[19] S. Yamano, J. Tatsuno, F.J. Gonzalez, The CYP2A3 gene product catalyzes coumarin 7-hydroxylation in human liver microsomes, Biochemistry 29 (1990) 1322–1329.
[20] J.K. Yano, M.H. Hsu, K.J. Griffin, C.D. Stout, E.F. Johnson, Structures of human microsomal cytochrome P450 2A6 complexed with coumarin and methoxsalen, Nat. Struct. Mol. Biol. 12 (2005) 822–823.
[21] J. Skopalík, P. Anzenbacher, M. Otyepka, 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 112 (2008) 8165–8173.
[22] E.E. Bolton, Y. Wang, P.A. Thiessen, S.H. Bryant, PubChem: integrated platform of small molecules and biological activities, Annu. Rep. Comput. Chem. 4 (2008) 217–241.
[23] Maestro v9.3, Macromodel, Protein Preparation Wizard: Schrödinger, LLC, New York, NY.
[24] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.
[25] J.K. Yano, T.T. Denton, M.A. Cerny, X. Zhang, E.F. Johnson, J.R. Cashman, Synthetic inhibitors of cytochrome P-450 2A6: inhibitory activity, difference spectra, mechanism of inhibition, and protein cocrystallization, J. Med. Chem. 49 (2006) 6987–7001.
[26] N.M. DeVore, K.M. Meneely, A.G. Bart, E.S. Stephens, K.P. Battaile, E.E. Scott, Structural comparison of cytochromes P450 2A6, 2A13, and 2E1 with pilocarpine, FEBS J. 279 (2012) 1621–1631.
[27] N.M. DeVore, E.E. Scott, Nicotine and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone binding and access channel in human cytochrome P450 2A6 and 2A13 enzymes, J. Biol. Chem. 287 (2012) 26576–26585.
[28] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, J. Mol. Biol. 267 (1997) 727–748.
[29] P. Rydberg, P. Vasanthanathan, C. Oostenbrink, L. Olsen, Fast prediction of cytochrome P450 mediated drug metabolism, Chem. Med. Chem. 4 (2009) 2070–2079.
[30] H. Li, A.D. Robertson, J.H. Jensen, Very fast empirical prediction and rationalization of protein pKa values, Proteins 61 (2005) 704–721.
[31] D. Case, T. Darden, T. Cheatham III, C. Simmerling, J. Wang, R. Duke, R. Luo, R. Walker, W. Zhang, K. Merz, AMBER 12, University of California, San Francisco, 2012.
[32] W. Li, H. Ode, T. Hoshino, H. Liu, Y. Tang, H. Jiang, Reduced catalytic activity of P450 2A6 mutants with coumarin: a computational investigation, J. Chem. Theory Comput. 5 (2009) 1411–1420.
[33] J.P. Rychaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, J. Comput. Phys. 23 (1977) 327–341.
[34] U. Essmann, L. Perera, W.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh ewald potential, J. Chem. Phys. 103 (1995) 8577–8592.
[35] G. Cruciani, E. Carosati, B. De Boeck, K. Ethirajulu, C. Mackie, T. Howe, R. Vianello, MetaSite: understanding metabolism in human cytochromes from the perspective of the chemist, J. Med. Chem. 48 (2005) 6970–6979.
[36] G.J. Kleywegt, J.Y. Zou, M. Kjeldgaard, T.A. Jones, International Tables for Crystallography, Volume F, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001 (353–356 and 366–367).
[37] P.J. Loida, S.G. Sligar, M.D. Paulsen, G.E. Arnold, R.L. Ornstein, Stereoselective hydroxylation of norcamphor by cytochrome P450cam. Experimental verification of molecular dynamics simulations, J. Biol. Chem. 27 (1995) 5326–5330.
[38] M.L. Danielson, P.V. Desai, M.A. Mohutsky, S.A. Wrighton, M.A. Lill, Potentially increasing the metabolic stability of drug candidates via computational site of metabolism prediction by CYP2C9: the utility of incorporating protein flexibility via an ensemble of structures, Eur. J. Med. Chem. 46 (2011) 3953–3963.
[39] E. Stjernschantz, C. Oostenbrink, Improved ligand-protein binding affinity predictions using multiple binding modes, Biophys. J. 98 (2010) 2682–2691.
[40] P. Vasanthanathan, L. Olsen, F.S. Jørgensen, N.P. Vermeulen, C. Oostenbrink, Computational prediction of binding affinity for CYP1A2-ligand complexes using empirical free energy calculations, Drug Metab. Dispos. 38 (2010) 1347–1354.
[41] L.J. Kingsley, G.L. Wilson, M.E. Essex, M.A. Lill, Combining structure- and ligand-based approaches to improve site of metabolism prediction in CYP2C9 substrates, Pharm. Res. (2014), http://dx.doi.org/10.1007/s11095-014-1511-3 (online).
[42] Y. Yang, S.E. Wong, F.C. Lightstone, Understanding a substrate's product regioselectivity in a family of enzymes: a case study of acetaminophen binding in cytochrome P450s, PLoS One 9 (2014) e87058.
[43] J. Kirchmair, M.J. Williamson, A.M. Afzal, J.D. Tyzack, A.P. Choy, A. Howlett, P. Rydberg, R.C. Glen, FAst MEtabolizer (FAME): a rapid and accurate predictor of sites of metabolism in multiple species by endogenous enzymes, J. Chem. Inf. Model 53 (2013) 2896–2907.