Chem-Bio Informatics Journal, Vol.17, pp.38-52 (2017)

Prediction of the site of CYP3A4 metabolism of tolterodine by molecular dynamics simulation from multiple initial structures of the CYP3A4-tolterodine complex

Atsuko Sato[1,2], Hitomi Yuki[3], Chiduru Watanabe[3], Jun-ichi Saito[4], Akihiko Konagaya[2], Teruki Honma[2,3*]

[1] *Chemical Research Laboratories, Research Function Unit, R&D Division, Kyowa Hakko Kirin Co., Ltd., 1188, Shimotogari, Nagaizumi-cho, Sunto-gun, Shizuoka, 411-8731, Japan*
[2] *School of Computing, Department of Computer Science, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama, Kanagawa, 226-8503, Japan*
[3] *Center for Life Science Technologies, RIKEN, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan*
[4] *R&D Planning Department, R&D Division, Kyowa Hakko Kirin Co., Ltd., 1-9-2 Otemachi, Chiyoda-ku, Tokyo, 100-0004, Japan*

**E-mail: honma.teruki@riken.jp*

(Received March 15; accepted April 25; published online May 31, 2017)

Abstract

CYP3A4 contributes to the metabolism of more than 30% of drugs in clinical use. Predicting the sites of metabolism (SOM) by CYP3A4, as well as the binding modes, for target compounds is important for the design of metabolically more stable drugs. Precisely predicting the structures of CYP3A4–ligand complexes is enormously challenging owing to the high number of conformational possibilities with its numerous binding substrates. We previously described a method for predicting the SOM of carbamazepine by means of docking and molecular dynamics (MD) simulations starting from multiple initial structures. To validate our method, we have now applied it to tolterodine, which is more flexible than carbamazepine. In addition, we evaluated the effectiveness of two methods for selecting the initial structures for MD. In analyzing the MD trajectories, we calculated the frequency with which carbon atoms at each of four groups of the tolterodine molecule approached to within a certain cutoff distance of the heme iron, and we also calculated binding free energy. We found that compared to the other three groups, the position to the experimentally determined SOM was close to the heme most frequently and had the lowest average $\Delta G_{\mathbf{binding}}$. For selecting the MD initial structures, clustering on the basis of protein–ligand interaction fingerprints (PLIF) was substantially more robust at predicting accessibility compared with clustering based on root-mean-square deviations. These findings demonstrate that our method is applicable for a flexible ligand and that PLIF clustering is a promising method for selecting structures for MD. We succeeded to predict the experimentally determined SOM of tolterodine together with the appropriate binding mode. The predicted binding mode is useful to design metabolically more stable compounds.

Chem-Bio Informatics Journal, Vol. 17, pp.38-52(2017)

Key Words: Cytochrome P450 (CYP), Sites of metabolism (SOM), Molecular dynamics simulation, Multiple initial structures, Tolterodine

Area of Interest: *In silico* drug discovery

1. Introduction

Cytochrome P450s (CYPs) are a group of enzymes involved in the metabolism of drugs and endogenous compounds. CYPs are expressed at appreciable levels in the small intestinal mucosa, lung, kidney, brain, olfactory mucosa, and skin but are concentrated predominantly in the liver, which is the principal organ of drug elimination [1]. Of the many CYPs, CYP3A4 is one of the most important for drug metabolism, contributing to the metabolism of more than 30% of clinically used drugs [2]. In addition to playing a crucial role in the first-pass metabolism of oral medications, CYPs, including CYP3A4, are also involved in drug–drug interactions (e.g., CYP inhibition), and the adverse effects of these interactions are an important obstacle to drug development.

The development of many drug candidates has been discontinued or delayed in the early stages owing to this pharmacokinetic problem, and improving the stability of drug candidates with respect to metabolism by CYPs is one of the highest priority issues in drug discovery. Medicinal chemists often synthesize analogs of lead molecules with the goal of reducing clearance by CYPs. Various approaches are used for this purpose, such as (a) reducing ring size, (b) adding hydroxyl groups, (c) introducing steric hindrance at a labile site, (d) cyclization at a labile site, (e) removal of labile groups, and (f) adding a F or Cl substituent at a site of metabolism (SOM) [3]. Many of these approaches, including the introduction of an additional substituent [4, 5], require accurate prediction of the SOM. Methods for SOM prediction can provide information not only about metabolically unstable groups but also about the binding mode of CYP–drug complexes and the binding mode can be used to guide the above-mentioned approaches. During drug discovery, there is often a trade-off between a compound's binding affinity to a target protein and its metabolic stability. For example, increasing the hydrophobicity of a compound tends to increase its affinity against the target protein while lowering its metabolic stability by increasing its affinity for CYPs.

To precisely predict a compound's SOM, information about the structure of the metabolic enzyme is critically important. As of 2017, 25 CYP3A4 X-ray structures had been registered in Protein Data Bank. Superimposition of the structures reveals that the conformations of the F-G loop and the side chains of the neighboring residues depend strongly on the type of substrate bound to the enzyme (Figure 1). Two important factors must be considered when predicting a substrate's SOM by CYP3A4: (1) the accessibility of each atom in the substrate to the oxygenated iron atom of the heme in CYP3A4 and (2) the oxidative reactivity of each atom in the substrate. For prediction of the former, docking simulation using the substrate and a CYP3A4 structure is widely used.

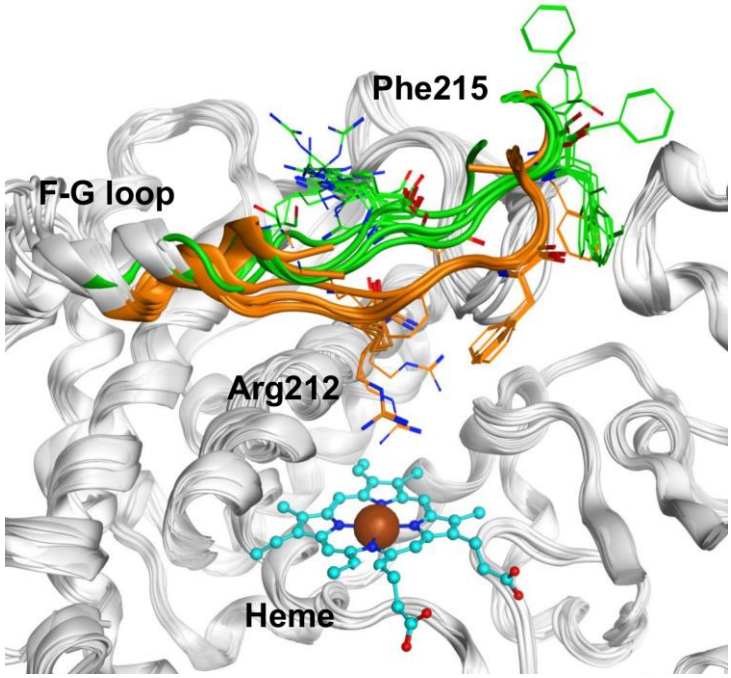

Figure 1. Superposition of crystal structures for CYP3A4
The F-G loops (Lys209-Leu216) are colored orange for structures (PDB IDs: 3UA1, 4I3Q, 4K9V, 4K9X, 1W0G, 1W0E, 1TQN, 1W0F, the side chains of Arg212 and Phe215 are located inside the heme pocket), colored green for structures (PDB IDs: 3NXU, 4I4G, 4I4H, 4K9T, the side chains of Arg212 and Phe215 are located outside the heme pocket).

There are many methods for SOM prediction [6], including SMARTCyp [7] and MetaSite [8]. The reported methods use a set of oxidative reactivities of fragment molecules (precalculated by quantum mechanical techniques such as density functional theory [DFT]) in combination with accessibility estimation methods such as the use of topological accessibility descriptors and superimposition of substrates with binding site grids. For example, MetaSite uses GRID-derived molecular interaction fields for the protein and the ligand based on protein structural information as well as molecular orbital calculations to estimate the likelihood of a metabolic reaction at a certain atom. More recently, researchers have proposed SOM prediction methods that consider side chain and ligand flexibility, as well as activation energy [9, 10]. Although these proposed methods estimate accessibility by means of protein–ligand docking, they do not adequately account for the effect of main chain flexibility.

We previously reported the first use of a combination of docking, molecular dynamics (MD) simulation [11], and quantum chemistry–derived reactivity calculations to predict the SOM of carbamazepine (CBZ) by CYP3A4. We concluded that docking simulation alone was not sufficient to accurately account for molecular flexibility of CYP3A4 and that MD simulation was necessary. Moreover, we reported that MD simulation starting from multiple initial structures of complex was effective for efficiently exploring possible binding modes. Although we successfully predicted the SOM of the rigid CBZ molecule, application of our method to flexible compounds is necessary to validate its robustness. In general, predicting binding modes is more difficult for flexible compounds than for rigid ones. CBZ has a rigid structure containing only one rotatable bond, and conformational differences have little influence on the accessibility of each atom. In this study, we selected the more flexible molecule tolterodine, an antimuscarinic drug used for treatment of

urinary incontinence symptoms [12], as a second test of our prediction method, with the goal of improving the method. CYP3A4 metabolizes tolterodine to *N*-dealkylated tolterodine by oxidizing one of the alkyl groups on the nitrogen (Figure 2) [13].

Recent increases in computing power now permit large-scale MD simulations; nevertheless, effective selection of initial structures for MD is still important for minimizing computation time, especially for practical applications such as drug design in pharmaceutical companies. Herein, we report efficient methods for minimizing the number of initial structures. Specifically, to extract and select initial structures for MD, we clustered poses generated by docking simulation of the interaction between multiple CYP3A4 structures and tolterodine. Conventionally, clustering methods based on root-mean-square deviations (RMSDs) of distance or dihedral angle have been used to partition MD trajectories into groups of structures that share similar conformational features. In recent years, structural interaction fingerprint (SIFt) [14] or protein–ligand interaction fingerprints (PLIFs) have begun to be used to cluster docking poses and MD trajectories. This method can reveal striking similarities between the patterns of protein–ligand interactions in the structures of complexes. In the case of SOM prediction for CBZ, we evaluated the accessibility of CBZ atoms to the heme iron using 5 MD simulations involving initial structures that were selected on the basis of RMSD clustering. In this work, we selected initial structures on the basis of both RMSD and PLIF clustering. We also investigated the number of initial structures that was necessary and sufficient for accurate prediction of the SOM of the flexible substrate tolterodine.

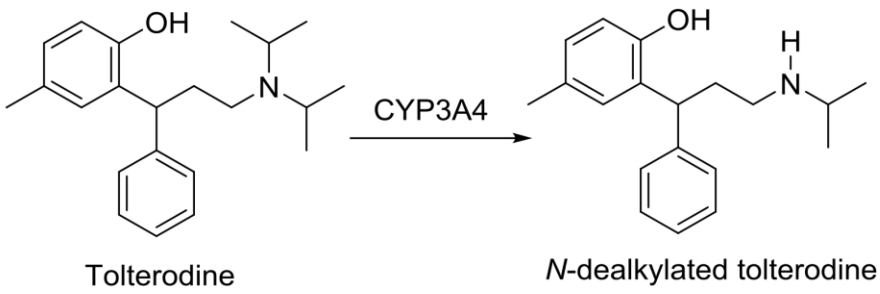

Figure 2. Main pathway for metabolism of tolterodine in human liver microsomes

2. Materials and Methods

In the crystal structure of the apo protein of CYP3A4 (PDB ID: 1TQN), the side chain of Arg212 occupies the center of the heme pocket, where the side chain presumably would inhibit the binding of tolterodine to the pocket. Therefore, we superimposed the 1TQN structure on the 1W0G structure, which is the structure of the CYP3A4–metyrapone complex and in which the Arg212 side chain is in a location where it cannot interfere with substrate docking. Subsequently, we replaced the Arg212 coordinates of the 1TQN structure with those of the 1W0G structure by using the MOE software package [15]. Before the docking procedure, the water molecules in the crystal structures were removed.

2.1 Use of docking to prepare initial structures for MD simulation

The docking function in MOE [15] was used to generate the initial structure candidates of the CYP3A4–tolterodine complex for MD simulations. We specified the docking site by using the

following 3 entities: (a) dummy atoms detected by means of SiteFinder [16], (b) Phe213, and (c) the iron atom in the heme; default values were used for all other settings. The main reason for the selection of Phe213 to specify the docking site was that CBZ stably interacted with this residue during the MD simulations carried out in our previous study [11]. The heme iron was assumed to be unoxidized during the approach of the substrate [17]. The Alpha Triangle algorithm was used to place the tolterodine into the substrate-binding pocket, and the binding energy was scored according to the London ΔG function until 100 nonequivalent crude docking conformations were obtained [18]. The Amber12:EHT force field was used to minimize the structures. We selected CYP3A4–tolterodine structures for which the score (GBVI/WSA Δ*G*) was less than –5 kcal/mol and the shortest distance between the heme iron and any one of the tolterodine carbon atoms was less than 7 Å. The docked poses were clustered by 2 methods. First, PLIF analysis was carried out with MOE, and 4 bits were extracted from the interactions between Arg105, Ala370, Arg372, Glu374, and tolterodine. With these 4 bits, cluster analysis was performed by means of the k-means algorithm [19]. Second, MOE was used for cluster analysis of the tolterodine poses on the basis of the RMSDs of the compound's atomic coordinates. Then representative structures were selected by comparison of the docking scores (GBVI/WSA Δ*G*) of the poses in each cluster. The structures selected by this process were used as initial structures in the subsequent MD simulations.

2.2 MD simulation protocols

The charges on the heme group and the tolterodine molecule were derived by quantum mechanical calculations carried out with *Gaussian03* software [20] at the B3LYP/6-31G** level, and the charges were fitted to their respective atoms by means of the RESP method [21] in a manner similar to that described previously [22]. Each model of CYP3A4–tolterodine structure was inserted in a rectangular TIP3P water box [23] at a minimum distance of 18 Å from the box boundary. Counterions (Na^+, Cl^-) were added to neutralize the system. The ff03 force field [24] was used. MD simulations were performed with the PMEMD module of the Amber 12 software package [25]. The system was minimized by means of 5,000 steps each of the steepest descent and conjugate gradient methods for hydrogen optimization; 5,000 steps each of the same methods for side chains and solvent optimization; 30,000 steps each of the same methods for protein and solvent optimization; and then 30,000 steps each of the same methods for optimization of the entire system including the ligand. The system was heated to 300 K for 130 ps under the NVT ensemble condition and equilibrated for 1.1 ns under the NPT condition with application of a positional restraint of 10 kcal mol^{-1} $Å^{-2}$ to tolterodine. After equilibration, additional simulation was performed under ligand positional restraints of 5 kcal mol^{-1} $Å^{-2}$ for 100 ps and 2 kcal mol^{-1} $Å^{-2}$ for 100 ps; the goal of this additional simulation was to release the restraints gradually. Next, unrestrained simulation was performed for 10 ns as a product run. During the 10 ns simulations on a NVIDIA Tesla C2075 processor, snapshots were stored every 1 ps and used for analysis. Throughout the simulations, the SHAKE algorithm [26] was employed to constrain bonds involving hydrogen atoms. The integration time step was set to 1 fs. The cutoff distance for nonbonded terms was set to 10 Å.

2.3 Analysis of MD simulations

The ptraj module of the Amber software package was used for analyzing MD trajectories such as calculation of RMSDs and measurement of distances. To compare the stability of each binding mode while taking solvent effects into consideration, we calculated the binding free energy ($\Delta G_{binding}$) for each mode by using the MM/PBSA method in Amber. In general, $\Delta G_{binding}$ for binding between CYP3A4 and tolterodine is given by the following equation:

$$\Delta G_{binding} = \Delta G_{MM} + \Delta G_{sol} - T\Delta S$$

where ΔG_{MM} is the molecular mechanics term, ΔG_{sol} is the solvation term, and $T\Delta S$ is the entropy term. ΔG_{MM} and ΔG_{sol} were calculated by means of the MM/PBSA method implemented in Amber. In our MD simulations of the tolterodine–CYP3A4 complex, we neglected calculation of $T\Delta S$ because the ligand molecule is identical. The binding energy calculation was carried out for 10,000 snapshots acquired at 1-ps intervals throughout the 10 ns product run.

2.4 Prediction of reactivity

We evaluated the oxidative reactivity of each carbon atom of tolterodine by using SMARTCyp [7]. This application estimates the reactivity by calculation of the activation energies of oxidation, and all atoms of a substrate are assigned by matching molecular fragments with patterns in a fragment library for which the activation energies have been precomputed by means of DFT transition state calculations.

3. Results and Discussion

3.1 Preparation of initial structures for MD simulations

There have been many studies of the mechanism and catalytic cycle of oxidation by cytochrome P450s [27]. The oxidation state of the heme iron as the substrate approaches the heme is important. Whether the iron is oxidized after substrate binding or while the substrate is absent has not been conclusively established, because experimental determination of the oxidation state at the moment of substrate binding is difficult. To simplify the simulations in this study, we assumed that the heme iron was unoxidized while the substrate was approaching.

As of 2017, there were 25 available CYP3A4 X-ray structures (PDB IDs: 1TQN [28], 1W0E, 1W0F, 1W0G [29], 2J0D, 2V0M [30], 3NXU [31], 3UA1 [32], 3TJS [33], 4I4H, 4I4G, 4I3Q [34], 4K9X, 4K9W, 4K9V, 4K9U, 4K9T [35]) were available as of 2013. Of these, 1TQN has the highest resolution (2.05 Å) and is not missing any residues in the F-G loop. 1TQN is the structure of the apo protein, and the side chain of Arg212 protrudes inward the heme pocket. We assumed that in this conformation, Arg212 would inhibit binding of tolterodine to the pocket, so we superimposed the structure of 1TQN on the structure of 1W0G (Figure S1) because the position of the Arg212 side chain in 1W0G is appropriate for substrate docking and because among the structures available at the time, this structure was the only one complexed with a substrate in the heme pocket. After superimposing the structures, we replaced the Arg212 coordinates of 1TQN with those of 1W0G by using MOE [15].

Before substrate docking, the water molecules in the crystal structures were removed. This same modified 1TQN was also used in the previous study [11]. The docking program implemented in MOE was used to generate the initial structure candidates of the CYP3A4–tolterodine complex for MD simulations. From 236 docking poses, we selected 154 poses for which the docking score (GBVI/WSA ΔG) was less than –5 kcal/mol and for which the shortest distance between the heme iron and at least one of the tolterodine carbon atoms was <7 Å. For the tolterodine poses, cluster analysis was performed on the basis of RMSDs of the compound's atomic coordinates and by PLIF, and then 23 clusters were generated by tuning the cutoff values. Because several of the clusters included multiple docking poses docking patterns, we further divided the clusters by visual

inspection. This procedure resulted in the selection of 27 representative structures that had the lowest docking score in each cluster (Figure 3). These 27 structures were used as the initial structures for the MD simulations (Figure 4 shows the workflow).

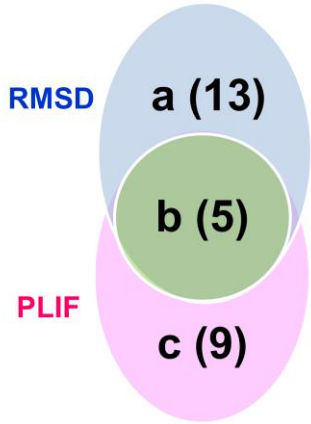

Figure 3. Venn diagram of 27 initial structures generated by RMSD and PLIF clustering
Numbers in parentheses are numbers of structures.

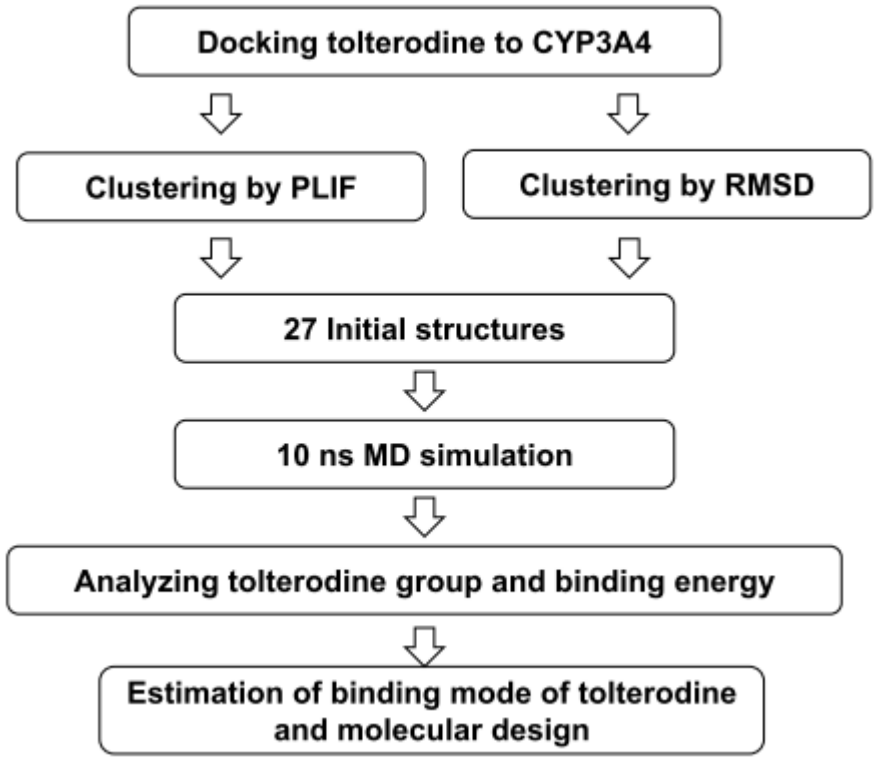

Figure 4. Workflow of for MD simulations using with multiple initial structures

3.2 Analysis of MD simulations

MD simulations were performed with the PMEMD module of the Amber 12 software package in much the same way as described previously [22]. For CBZ, in an additional simulation after equilibration, we applied distance restraints to reflect the presence of a hydrogen bond between CBZ and CYP3A4. In contrast, in this study, we restrained only the initial coordinates of tolterodine because no key hydrogen bonds between this compound and the enzyme were predicted. This method can be adapted for future work because it can be applied to a wide range of compounds. After the process for equilibration, unrestrained 10 ns simulations were performed as product runs.

During the 10 ns simulations, snapshots were stored every 1 ps and used for analysis. The ptraj module of Amber was used for analyzing MD trajectories as described in section 2.3. To compare the stability of each binding mode while taking solvent effects into consideration, we calculated a $\Delta G_{\text{binding}}$ value by using MM/PBSA.

In the MD simulations of the tolterodine–CYP3A4 complex, we did not calculate the entropy term, for the following reason. The purpose of calculating $\Delta G_{\text{binding}}$ is to represent indices for the stabilities of different binding modes of the same compound to predict the compound's accessibility to the heme iron. In the case of the same protein and same compound, the entropy has little influence on $\Delta G_{\text{binding}}$.

Evaluation of RMSD values for the main chain atoms of the initial poses obtained from docking and the corresponding atoms in snapshots during the 10 ns MD product runs revealed no major structural changes in the trajectories during the runs (Figure S2). To confirm that tolterodine did not break loose from CYP3A4 during the simulations, we analyzed the time course of $\Delta G_{\text{binding}}$ for each simulation (Figure S3). In all the simulations, there were no remarkable energy shifts at the moment when the distance restraint was switched off. This result indicates that no unusual abrupt changes in structure occurred during the simulations. The results of these time-course analyses suggest that the MD simulations were appropriately carried out and that the CYP3A4–tolterodine snapshots could be expected to reproduce tolterodine binding to CYP3A4 from the selected initial structures. The trajectories during the simulations were used for analysis of the accessibility of the tolterodine carbon atoms to the heme.

3.3 Prediction of accessibility and reactivity

The carbon atoms of tolterodine were numbered and categorized into 4 groups (groups A–D, Figure 5). As indicators of the accessibility of the substrate to the heme iron, we calculated the frequency with which at least one atom at each of the 4 carbon groups was within 6 Å of the heme iron and we calculated the corresponding $\Delta G_{\text{binding}}$ values, using all snapshots (10,000 snapshots for each simulation). We used 6 Å as the cutoff distance at which an oxidative reaction could occur as in the previous study [11] which confirmed that this cutoff distance was appropriate. 179,019 snapshots in which at least one carbon atom was ≤6 Å of the heme iron were selected, and the snapshots were grouped on the basis of the position of the carbon atom nearest to the iron.

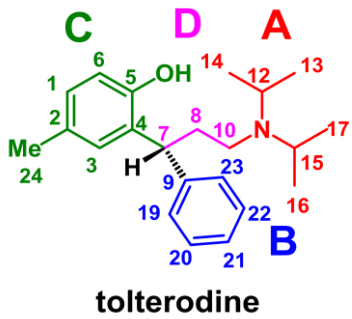

Figure 5. Position labels assigned to tolterodine carbon atoms

Table 1. Accessibility and reactivity

[a] Frequencies at which the carbon atom positions were located ≤6 Å from the heme iron, along with reactivities of selected carbon atoms. The distances were monitored every 1 ps throughout the 10 ns MD simulations. ΔGbinding values were calculated with MM/PBSA, and activation energies were calculated with SMARTCyp.

	Distance $_{Fe-C}$ ≤ 6 Å				Reactivity	
	Frequecy		$\Delta G_{binding}$ (kcal/mol)			
Group	No. of snapshots	%	Mean	Standard deviation	Atom	Activation Energy (kcal/mol)
A	72,980	40.8	–24.10	5.89	C12, C15	9.82
B	37,455	20.9	–16.84	5.68	C21	20.63
C	67,887	37.9	–20.98	4.99	C24	15.87
D	697	0.4	–22.01	2.92	C10	9.82

To predict the SOM, we calculated the frequency of appearance within the cutoff distance and the $\Delta G_{binding}$ of the CYP3A4–tolterodine complex for groups A–D, and we calculated the reactivity of each carbon atom of tolterodine using SMARTCyp [7] (Table 1). Analysis of the snapshots for all 27 MD simulations indicated that the proportion of snapshots in which group A was closest to the heme was the highest (40.8%), and the corresponding proportions for groups B, C, and D were 20.9%, 37.9%, and 0.4%, respectively. Table 1 and S4 show the mean $\Delta G_{binding}$ values calculated by the MM/PBSA method for each group. These Figures indicate that the tolterodine poses in which an atom at group A was closest to the heme iron had lower mean $\Delta G_{binding}$ values than the poses in which an atom at one of the other 3 groups was closest to the iron. The mean $\Delta G_{binding}$ values for the 4 groups increased in the order A < D < C < B. Therefore, we concluded that the tolterodine carbon atoms at group A were the most accessible to the heme.

In terms of oxidative reactivity, SMARTCyp indicated that C12 (group A), C15 (group A), and C10 (group D) all had a predicted activation energy of 9.82 kcal/mol, which was lower than the predicted activation energies of the other atoms (which were >15 kcal/mol, Table 1). According to these results, both group A and group D are highly reactive. However, group D (frequency, 0.4%; $\Delta G_{binding}$, –22.01 kcal/mol) was less accessible than group A (40.8%, –24.10 kcal/mol) because the group D atoms are located at the center of the molecule. Taken together, the accessibility results

derived from the MD simulations and the reactivity results calculated by means of SMARTCyp predict that group A is the SOM of tolterodine by CYP3A4, which is consistent with experimental results [13].

3.4 Prediction of binding mode

In addition to SOM prediction, determination of the binding mode at the time of the oxidative reaction is also informative for the design of more metabolically stable drugs. To determine a reasonable binding mode for tolterodine, we sampled 2,700 MD snapshots (each 100 ps) and extracted 734 snapshots in which carbon atoms at group A were close to the heme. PLIF [15] bits were generated (Figure S5), and cluster analysis was performed with the Pipeline Pilot [36]. The threshold level was adjusted by visual inspection in such a way that members of each cluster had a similar binding mode; as a result, 16 clusters were generated. Table S1 shows the frequency of each cluster, along with the average calculated $\Delta G_{\text{binding}}$ values and average minimum Fe–C distances. Cluster 1 had the most members, the second lowest average $\Delta G_{\text{binding}}$, and the smallest Fe–C distance. From this result, we inferred that cluster 1 contained a complex that could reasonably be assumed to represent the binding mode of tolterodine. In contrast, complexes in cluster 7 had the lowest average $\Delta G_{\text{binding}}$ values among the clusters that accounted for >1% of the snapshots (Table S1).

Figure S6 shows the most stable complexes in cluster 1 (blue) and cluster 7 (gray). The complex extracted from cluster 1 had the lowest $\Delta G_{\text{binding}}$ among all the complexes (–38.56 kcal/mol); in contrast, $\Delta G_{\text{binding}}$ for the most stable complex in cluster 7 was –33.05 kcal/mol. By taking into account the frequency and the binding energy, we deduced that the complex in cluster 1 was likely to be the binding mode of tolterodine at the time of the oxidative reaction. This structure suggests several design modifications that might interfere with the binding between CYP3A4 and tolterodine, as long as the modifications did not affect the resulting compounds' activity against the target protein. For instance, breaking the hydrogen bond between Arg345 and the hydroxyl group of tolterodine might reduce the binding affinity to CYP3A4. In addition, a bulky substituent in the meta or para position on the benzene ring might effectively reduce the binding affinity because there appears to be no room in the substrate-binding site for any additional substituents.

3.5 Effective selection of MD initial structures for accessibility prediction

In this study, 27 clusters were generated by means of cluster analysis of tolterodine poses on the basis of both RMSD and PLIF calculations. As the initial structures for the MD simulation, a representative CYP3A4–tolterodine structure with the lowest MOE docking score was selected from each cluster, for a total of 27 structures (Figure 3). There were 5 initial structures that appeared in both the RMSD cluster and the PLIF cluster. Because representative structures were selected by comparing the docking scores, these 5 structures were selected from different but partially overlapping clusters.

Figure 6 shows the frequencies of the occurrences of the carbon atoms at groups A–D at a distance of ≤6 Å from the heme iron (that is, within the distance required for oxidation) during the 27 MD simulation runs. At least one of the carbon atoms assigned to a particular group (A–D) was closest to the heme iron and ≤6 Å from it at a given time point during a simulation run are indicated by black lines at the given time point. Overall, carbons in group A showed the highest frequency of occurrence within the specified distance. Interestingly, the high frequency was particularly noticeable for the initial structures that were found only in the PLIF clusters (see c in the top panel of Figure 6). The frequency for group B was approximately half that for group A. The second highest frequency was observed for snapshots in which group C came near to the heme. In the case

of group C, there was not much difference between structures selected from the PLIF clusters and those selected from the RMSD clusters (c and a, respectively, in Figure 6). There were few snapshots in which group D was close to the heme because the carbon atoms at this group are located at the center of the tolterodine molecule and are buried by surrounded substituent groups.

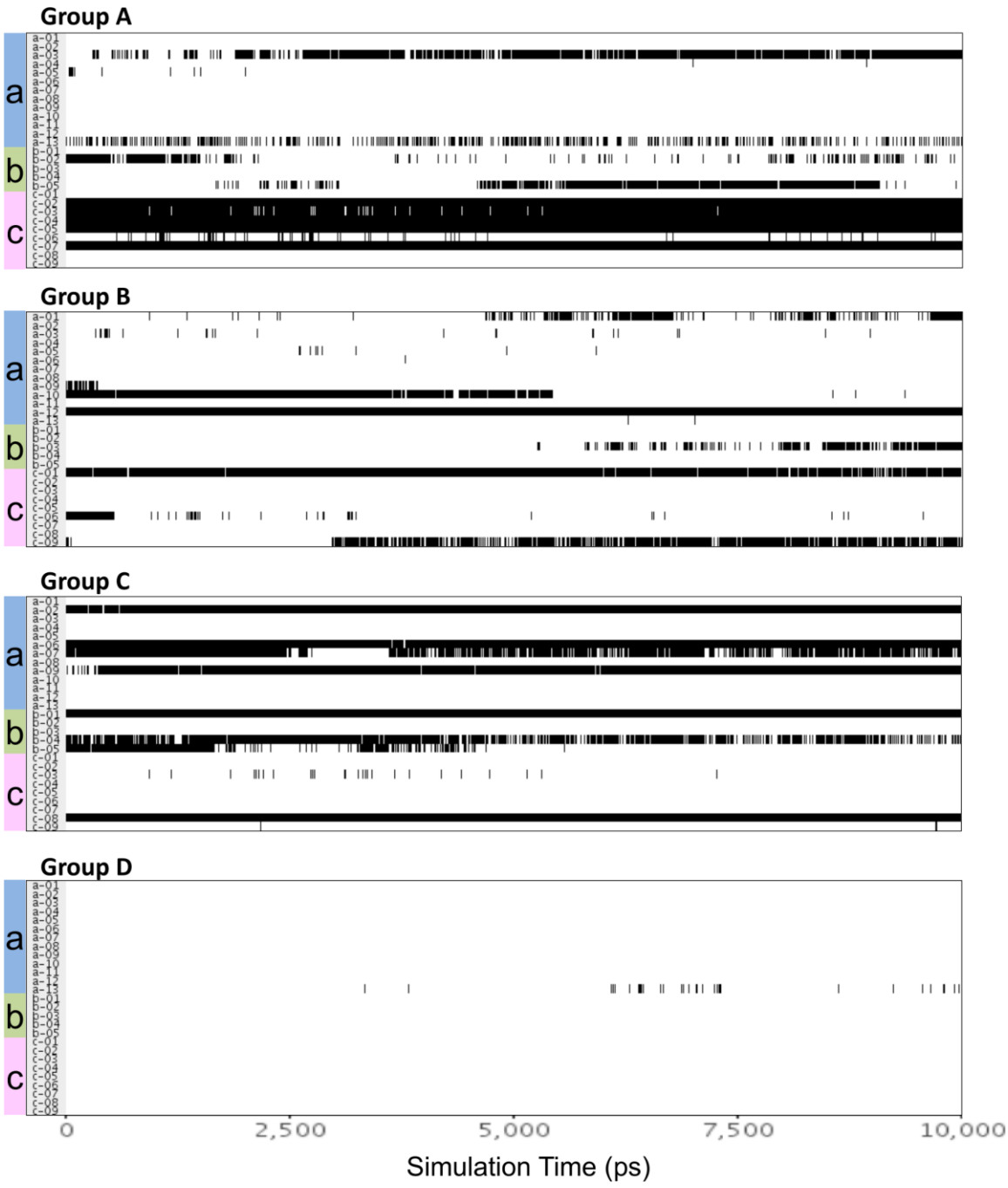

Figure 6. Heat map of the frequencies at which carbon atom positions were located close to the heme

Each black line indicates a snapshot in which at least one carbon atom at a given position (A–D) was ≤6 Å from the heme and was closer to the heme than any other carbon atom in the molecule. Distances were monitored every 1 ps throughout the 10 ns MD simulations. The clustering methods used to select the 27 initial structures are indicated on the left of each heat map (see Figure 3 for a key to the colors and letters).

On the basis of our results, we compared the effectiveness of the RMSD and PLIF clustering methods at minimizing the number of initial structures for MD (Table 2). The rankings in order of frequency and $\Delta G_{binding}$ obtained with the 14 structures selected from the PLIF cluster were the same as the rankings obtained with all 27 initial structures, whereas the rankings obtained with the 18 structures from the RMSD cluster were incorrect.

Table 2. Comparison of the effectiveness of selection methods for minimizing the number of initial structures for MD

[a] See Figure 3 for an explanation of the lowercase letters in this column.

Selection method	Group	Frequency			$\Delta G_{binding}$ (kcal/mol)		No. of initial structures
		No. of snapshots	%	Rank	Mean	Rank	
PLIF +	A	72,980	40.8	1	–24.10	1	
RMSD	B	37,455	20.9	3	–16.84	4	
(all) [a]	C	67,887	37.9	2	–20.98	3	27
	D	697	0.4	4	–22.01	2	
	Total	179,019					
RMSD	A	22,493	22.1	2	–21.28	2	
(a + b) [a]	B	20,855	20.5	3	–16.09	4	
	C	57,598	56.7	1	–20.94	3	18
	D	697	0.7	4	–22.01	1	
	Total	101,643					
PLIF	A	57,796	53.8	1	–24.54	1	
(b + c) [a]	B	18,929	17.6	3	–18.32	3	
	C	30,716	28.6	2	–21.50	2	14
	D	0	—	—	—	—	
	Total	107,441					

This comparative analysis for tolterodine, which is a relatively flexible compound, indicates that selection of the initial structures from the PLIF cluster was more effective and robust than selection using conventional RMSD clustering. As already mentioned, PLIF describes patterns of forming or not forming protein–ligand interactions, and the clusters generated by using PLIF can effectively identify the binding modes. In contrast, RMSD clustering detects simple positional changes, and sometimes the changes are not meaningful for every binding mode. Our comparative analysis for tolterodine shows that initial structures should be selected by focusing on interaction patterns for MD simulations carried out for the purpose of analyzing protein–ligand accessibility. In terms of the minimum number of initial structures for that purpose, we were able to reduce the number to 14 in the case of tolterodine. However, the appropriate number of initial structures can be expected to depend on the flexibility and the pharmacophore profile of the substrate as well as the length of the product MD runs.

4. Conclusion

In this study, we came to a couple of conclusions. Firstly, we successfully predicted the experimentally determined SOM of tolterodine, which is a more flexible ligand than CBZ, on the basis of accessibility to the heme iron in CYP3A4 by analyzing MD simulation results obtained with multiple initial structures. In our analysis, simulation snapshots in which the experimentally metabolized carbons approached the heme iron showed the highest frequency and the most stable binding. The carbons that showed high accessibility also showed high calculated reactivities, as indicated by SMARTCyp. It indicated that our SOM prediction method, which combines MD simulation with quantum mechanics-based reactivity prediction by means of SMARTCyp can be applied to flexible ligands.

Secondly, we found the effective means in the selection of initial structures for MD simulations. The MD simulations using initial structures selected by PLIF clustering achieved significantly more accurate prediction of accessibility than those using structures selected from RMSD clusters. To establish a more robust protocol for selection of the minimum number of appropriate initial structures, we must further validate our method; however, PLIF clustering appears to be promising.

The advantage of our method is that it can be used not only to predict the SOM but also to select the proper binding mode using MD trajectories, and the binding mode can in turn be used to design metabolically more stable compounds. At this point, the method remains time-consuming: 62 h on a NVIDIA Tesla C2075 processor was required to predict the SOM of one compound. Additional work will be required before the method becomes mature enough for practical use in drug discovery.

Supplementary Information

Figures S1-S6, Table S1 are in the supplementary materials (1) available at: http://doi.org/10.1273/cbij.17.38

Acknowledgements

For this research, we used the computational resources of the RICC (RIKEN Integrated Cluster of Clusters). This research funding was supported by Kyowa Hakko Kirin Co. We are grateful to all those who supported us at RIKEN and at Kyowa Hakko Kirin Co. This research is (partially) supported by the Platform Project for Supporting in Drug Discovery and Life Science Research （Platform for Drug Discovery, Informatics, and Structural Life Science） from Japan Agency for Medical Research and Development (AMED).

References

[1] Lin, Y.; Lu, P.; Tang, C.; Mei, Q.; Sandig, G.; *et al*. Substrate Inhibition Kinetics for Cytochrome P450-Catalyzed Reactions. *Drug Metab. Dispos.* **2001**, *29* (4), 368-374.

[2] Zanger, U. M.; Schwab, M. Cytochrome P450 Enzymes in Drug Metabolism: Regulation of Gene Expression, Enzyme Activities, and Impact of Genetic Variation. *Pharmacol. Ther.* **2013**,

138 (1), 103-141.
[3] Di, L.; Kerns, E. H. *Drug-like Properties: Concepts, Structure Design and Methods: From ADME to Toxicity Optimization*; Academic Press, 2010.
[4] Fisher, M. B.; Henne, K. R.; Boer, J. The Complexities Inherent in Attempts to Decrease Drug Clearance by Blocking Sites of CYP-Mediated Metabolism. *Curr. Opin. Drug Discov. Devel.* **2006**, *9* (1), 101-109.
[5] Böhm, H. -J.; Banner, D.; Bendels, S.; Kansy, M.; Kuhn, B.; *et al*. Fluorine in Medicinal Chemistry. *ChemBioChem* **2004**, *5* (5), 637-643.
[6] Kirchmair, J.; Williamson, M. J.; Tyzack, J. D.; Tan, L.; Bond, P. J.; *et al*. Computational Prediction of Metabolism: Sites, Products, SAR, P450 Enzyme Dynamics, and Mechanisms. *J. Chem. Inf. Model.* **2012**, *52* (3), 617-648.
[7] Rydberg, P.; Gloriam, D. E.; Zaretzki, J.; Breneman, C.; Olsen, L. SMARTCyp: A 2D Method for Prediction of Cytochrome P450-Mediated Drug Metabolism. *ACS Med. Chem. Lett.* **2010**, *1* (3), 96-100.
[8] Cruciani, G.; Carosati, E.; De Boeck, B.; Ethirajulu, K.; Mackie, C.; *et al*. MetaSite: Understanding Metabolism in Human Cytochromes from the Perspective of the Chemist. *J. Med. Chem.* **2005**, *48* (22), 6970-6979.
[9] Huang, T. W.; Zaretzki, J.; Bergeron, C.; Bennett, K. P.; Breneman, C. M. DR-Predictor: Incorporating Flexible Docking with Specialized Electronic Reactivity and Machine Learning Techniques to Predict CYP-Mediated Sites of Metabolism. *J. Chem. Inf. Model.* **2013**, *53* (12), 3352-3366.
[10] Tyzack, J. D.; Williamson, M. J.; Torella, R.; Glen, R. C. Prediction of Cytochrome P450 Xenobiotic Metabolism: Tethered Docking and Reactivity Derived from Ligand Molecular Orbital Analysis. *J. Chem. Inf. Model.* **2013**, *53* (6), 1294-1305.
[11] Yuki, H.; Honma, T.; Hata, M.; Hoshino, T. Prediction of Sites of Metabolism in a Substrate Molecule, Instanced by Carbamazepine Oxidation by CYP3A4. *Bioorg. Med. Chem.* **2012**, *20* (2), 775-783.
[12] Van Kerrebroeck, P.; Kreder, K.; Jonas, U.; Zinner, N.; Wein, A. Tolterodine Once-Daily: Superior Efficacy and Tolerability in the Treatment of the Overactive bladder. *Urology* **2001**, *57* (3), 414-421.
[13] Brynne, N.; Stahl, M. M.; Hallén, B.; Edlund, P. O.; Palmér, L.; *et al*. Pharmacokinetics and Pharmacodynamics of Tolterodine in Man: A New Drug for the Treatment of Urinary Bladder Overactivity. *Int. J. Clin. Pharmacol. Ther.* **1997**, *35* (7), 287-295.
[14] Deng, Z.; Chuaqui, C.; Singh, J. Structural Interaction Fingerprint (SIFt): A Novel Method for Analyzing Three-Dimensional Protein−Ligand Binding Interactions. *J. Med. Chem.* **2004**, *47* (2), 337-344.
[15] *Molecular Operating Environment (MOE)*, 2013.08; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2013.
[16] Soga, S.; Shirai, H.; Kobori, M.; Hirayama, N. Use of Amino Acid Composition to Predict Ligand-Binding Sites. *J. Chem. Inf. Model.* **2007**, *47* (2), 400-406.
[17] Shaik, S.; Hirao, H.; Kumar, D. Reactivity Patterns of Cytochrome P450 Enzymes: Multifunctionality of the Active Species, and the Two States–two Oxidants Conundrum. *Nat. Prod. Rep.* **2007**, *24* (3), 533-552.
[18] Corbeil, C. R.; Williams, C. I.; Labute, P. Variability in Docking Success Rates due to Dataset Preparation. *J. Comput. Aided Mol. Des.* **2012**, *26* (6), 775-786.
[19] MacQueen, J. B, Berkeley, University of California Press. 1967, 1, 281-297. http://www-m9.ma.tum.de/foswiki/pub/WS2010/CombOptSem/kMeans.pdf (accessed Jun 5, 2016).

[20] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; *et al*. *Gaussian 03*; Gaussian, Inc., Wallingford, CT, 2004.
[21] Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the Multimolecule and Multiconformational RESP Methodology to Biopolymers: Charge Derivation for DNA, RNA, and Proteins. *J. Comput. Chem.* **1995**, *16* (11), 1357-1377.
[22] Sano, E.; Li, W.; Yuki, H.; Liu, X.; Furihata, T.; *et al*. Mechanism of the Decrease in Catalytic Activity of Human Cytochrome P450 2C9 Polymorphic Variants Investigated by Computational Analysis. *J. Comput. Chem.* **2010**, *31* (15), 2746-2758.
[23] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926.
[24] Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; *et al*. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24* (16), 1999-2012.
[25] Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; *et al*. *Amber 12*, University of California, San Francisco.
[26] Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327-341.
[27] Hrycay, E. G.; Bandiera, S. M. Monooxygenase, Peroxidase and Peroxygenase Properties and Reaction Mechanisms of Cytochrome P450 Enzymes. In *Monooxygenase, Peroxidase and Peroxygenase Properties and Mechanisms of Cytochrome P450*; Hrycay, E. G., Bandiera, S. M., Eds.; Springer International Publishing: Cham, 2015; Vol. 851, pp 1-61.
[28] 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-A Resolution. *J. Biol. Chem.* **2004**, *279* (37), 38091-38094.
[29] Willams, P. A.; Cosme, J.; Vinkovic, D.M.; Ward, A.; Angove, H. C.; *et al*. Crystal Structures of Human Cytochrome P450 3A4 Bound to Metyrapone and Progesterone. *Science* **2004**, *305* (5684), 683-686.
[30] Ekroos, M.; Sjögren, T. Structural Basis for Ligand Promiscuity in Cytochrome P450 3A4. *Proc. Natl. Acad. Sci. USA* **2006**, *103* (37), 13682-13687.
[31] Sevrioukova, I. F.; Poulos, T. L. Structure and Mechanism of the Complex between Cytochrome P4503A4 and Ritonavir. *Proc. Natl. Acad. Sci. USA* **2010**, *107* (43), 18422-18427.
[32] 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* (5), 3510-3517.
[33] Sevrioukova, I. F.; Poulos, T. L. Interaction of Human Cytochrome P4503A4 with Ritonavir Analogs. *Arch. Biochem. Biophys.* **2012**, *520* (2), 108-116.
[34] 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* (9), 3733-3741.
[35] Sevrioukova, I. F.; Poulos, T. L. Dissecting Cytochrome P450 3A4–Ligand Interactions Using Ritonavir Analogues. *Biochemistry (Mosc.)* **2013**, *52* (26), 4474-4481.
[36] Dassault Systèmes BIOVIA, *PipelinePilot*, Release 9.2, San Diego: Dassault Systèmes, 2014.