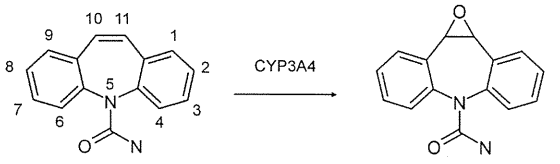

Scheme 1. Metabolic pathway of carbamazepine (CBZ) by CYP3A4.

in more detail. If accessibilities of the benzene ring atoms to the CYP3A4 heme iron are much lower than that of the 10, 11-double bond, the hydroxides cannot be produced. If the activation energy of epoxidation of the 10, 11-double bond is lower than those of hydroxylation of the benzene rings, kinetically 10, 11-epoxide is produced. About the latter case, Hata et al. also investigated the activation energies of epoxidation and hydroxylation by theoretical calculations. Their calculation suggested that the activation energy level of the first elementary step of 10, 11-epoxidation was lower than that of hydroxylation of the aromatic ring. The previous work of reactivity prediction should combine with accessibility prediction. In this study, to predict accessibility of CBZ atoms to the heme iron, we carried out MD simulations started from multiple CYP3A4-CBZ complexes and demonstrated advantage of utilizing MD simulation for the prediction of the SOM.

2. Materials and methods

Six CYP3A4 X-ray structures (PDB ID: 1TQN,[37] 1W0E, 1W0F, 1W0G,[38] 2J0D, 2V0M[39]) were available at 2009. Among them, we used four structures (1TQN, 1W0E, 1W0F, 1W0G) for docking study because 2J0D has missing residues in F–G loop close to its heme pocket and the resolution of 2V0M is over 3 Å. The structure of 1TQN is apo-protein and the side chain of Arg212 occupies the central part of its heme pocket. Since conformation of Arg212 was assumed to inhibit binding of CBZ to the heme pocket in 1TQN, the structure of 1TQN was superimposed on the structure of 1W0G. The position of the side chain of Arg212 in 1TQN is appropriate for docking, and then Arg212 coordinates of 1TQN were replaced with those of 1W0G using the program MOE.[68] Before docking, the water molecules in the crystal structures were removed.

2.1. Initial structure preparation by docking

The standard-precision (SP) mode of Glide 4.5[69] was used to generate the initial structures of CYP3A4-CBZ complex for MD simulation. The van der Waals scaling factor for receptor atoms changed to 0.5–1.0 by 0.1. CBZ docking calculations were performed with 24 docking conditions and top 10 poses by SP GlideScore for the respective conditions were saved for subsequent selection. We selected 33 complex structures under the condition that at least one hydrogen bond is formed between CBZ and CYP3A4 and the shortest distance between the heme iron and any CBZ carbon atoms are less than 7 Å. Cluster analysis was performed for the CBZ poses, and then 14 clusters were generated. Accordingly 14 representative complex structures were selected, by comparing SP GlideScore in each cluster. To optimize the CBZ poses, the extra-precision (XP) mode of Glide docking simulation was performed. After cluster analysis for the results, five complex structures with lower XP GlideScore and wide variety of CBZ poses were selected so that the selected structures had different interacting residues and CBZ directions among them. Consequently, these five complex structures were used as initial structures in MD simulations (Table 1, Fig. 3).

Table 1
Initial structures for MD simulation generated by Glide docking calculations

	Glide XP		
Model	Docking score	PDB ID	Residue to hold CBZ (kcal/mol)
1	−7.82	1TQN	Phe213
2	−7.54	1TQN	Ser119
3	−6.73	1TQN	Phe304
4	−6.70	1TQN	Phe304
5	−6.04	1W0F	Phe213

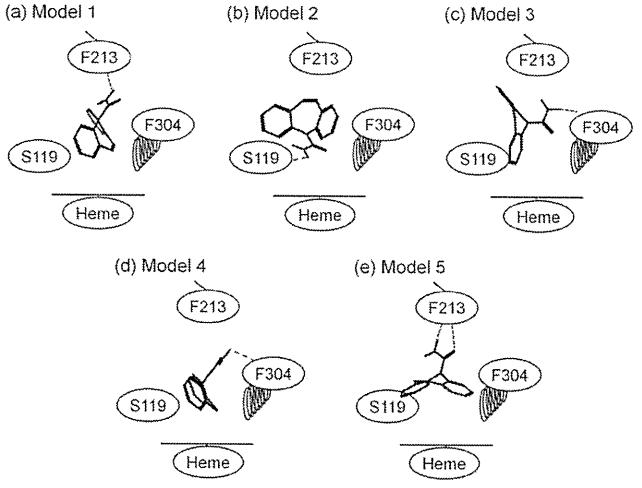

Figure 3. Binding modes in the respective models.

2.2. MD simulation protocols

The charges of the heme group and CBZ was derived by quantum mechanics calculation by Gaussian03 software[70] in the B3LYP/6-31G** level and fitted to respective atoms employing the RESP method[71] in a manner similar to that in the previous work.[72] Each model was inserted in a rectangular TIP3P water box[73] with a minimal distance of 18 Å from the box boundary. Counter ions were added to neutralize the system. The ff03 force field[74] was adopted. MD simulations were performed with the PMEMD module of AMBER 10 software package.[75] Each system was minimized (25,000 cycles of the steepest descent method, followed by 45,000 cycles of the conjugated gradient method) with applying distance restraints with 10 kcal mol^{-1} $Å^{-2}$ to the three atoms (one hydrogen atom and two hetero atoms) forming a hydrogen bond between CBZ and CYP3A4. The system was heated to 300 K for 130 ps under the NVT ensemble condition and equilibrated for 1.1 ns under NPT condition. After equilibration, additional simulation was performed under the distance restraints 5 kcal mol^{-1} $Å^{-2}$ for 100 ps and 2 kcal mol^{-1} $Å^{-2}$ for 100 ps to release restraints gradually. Then unrestrained simulation of 10 ns was performed. The snapshots in 10 ns were stored every 1 ps and used for analysis. Through the simulations, the SHAKE algorithm[76] was employed to constrain the bonds involving hydrogen atoms. Integration time step was set to 1 fs. The cut-off distance for non-bonded terms was set to 10 Å.

2.3. Analysis of MD simulations

The ptraj module of AMBER program[75] was used for handling MD trajectories such as calculation of RMSD and measurement of distances. To compare the stability of each binding mode taking solvent effect into consideration, we obtained a calculated $\Delta G_{binding}$ value using MM/PBSA. The binding free energy between

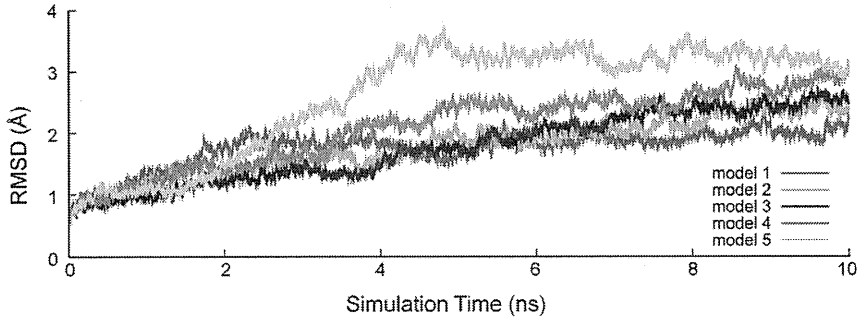

Figure 4. RMSDs of main chain atoms during the 10 ns MD simulations. Model 1 is represented by red line, model 2 is green, model 3 is blue, model 4 is magenta and model 5 is cyan.

CYP3A4 and CBZ estimated in the MM/PBSA method is obtained by the equation:

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

ΔG_{MM} and ΔG_{sol} terms were calculated by MM/PBSA method implemented in the AMBER program. In the MD simulations for CBZ-CYP3A4 complex, the entropy term was not calculated for the following two reasons. The purpose of obtaining calculated $\Delta G_{binding}$ is to suggest indices for stabilities of different binding modes of the same compound to estimate its accessibility to the heme iron. The entropy has only slight effect on $\Delta G_{binding}$ in the case of the same protein and compound, especially an inflexible compound. In this work, the number of rotatable bond of CBZ is only one and the difference of binding modes seems to hardly affect the value of $T\Delta S$. To predict stability of more flexible compounds, we need to take the effect of entropy into account. The binding energy calculation was carried out for 10,000 snapshots acquired every 1 ps through 10 ns MD simulation.

3. Results

3.1. MD simulation

Figure 4 shows the RMSD of main chain atoms between the initial structure at 0 ns and snapshots during 10 ns MD simulation. Although RMSD values were increased from 2–4 Å in every models, the values become almost constant after 5 ns. The fluctuations are also small. This suggested that calculation systems were equilibrated after simulation. To confirm the energetic stability during simulations, time course of the system total energy for each simulation was analyzed (Fig. S1). In all the simulations, the total energy steadily decreased through heating process and there was no remarkable energy shift at the moment when the distance restraint was switched off. This indicated that unusual abrupt change in structure scarcely occurred in the simulations. Therefore these results of time course analyses indicate an appropriateness of the MD simulations and the CYP-CBZ snapshots will reproduce CBZ binding in CYP3A4. The trajectories during the simulations were used for the analysis of accessibility of CBZ atoms to the heme.

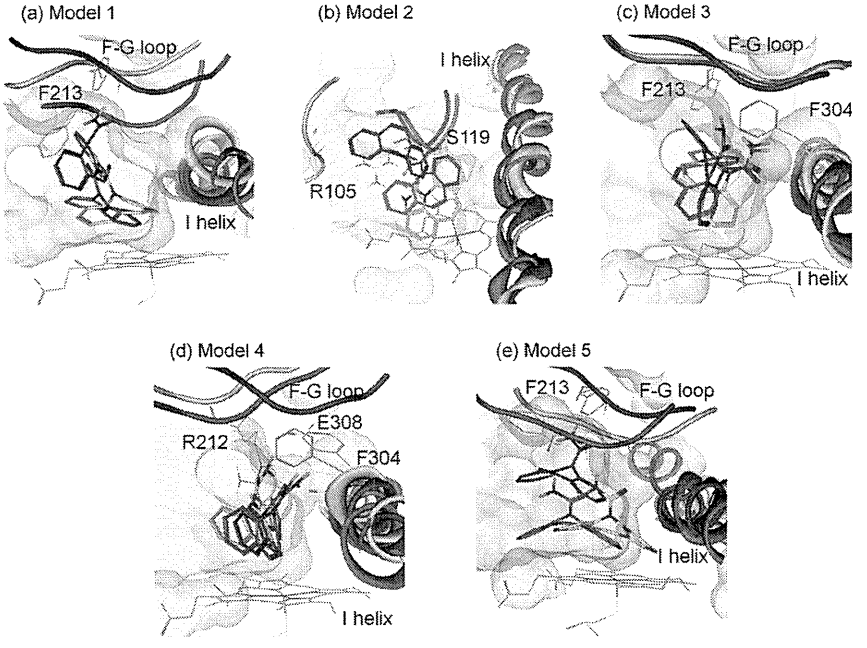

Figure 5. Representative snapshots of model 1–5 in MD simulations. (a) model 1, (b) model 2, (c) model 3, (d) model 4, (e) model 5. Red: initial complex, Green: the most frequent binding mode, Blue: the secondly-most frequent binding mode

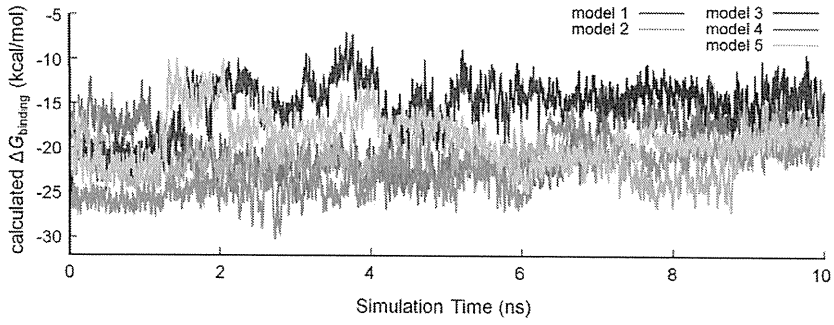

Figure 6. Calculated $\Delta G_{binding}$ value estimated by MM/PBSA method. Model 1 is represented by red line, model 2 is green, model 3 is blue, model 4 is magenta and model 5 is cyan.

3.2. The movement of CBZ during simulations

To compare representative CBZ poses at each simulation, we superimposed the initial CBZ poses, the most and secondly-most frequent poses in Figure 5.

3.2.1. Model 1

The orientation of CBZ on CYP3A4 was not significantly changed although its position was shifted from the initial position to the place close to heme (Fig. 5a). The hydrogen bond was often observed between the amino group of CBZ and the main chain carbonyl group of Phe213. Through the MD simulation, F–G loop bounded to CBZ moved toward heme with the motion of CBZ. This movement enabled close access of CBZ to heme for oxidation.

3.2.2. Model 2

Unlike other models, CBZ was dramatically moved from the initial position during the MD simulation (Fig. 5b). CBZ molecule turned more than 90 degree from the initial orientation and the 10, 11-double bond gradually approached to the iron. A hydrogen bond with Ser119 was seen in the docking model and other one with Arg105 was also observed through the simulation. The larger movements were observed at the regions around hydrogen bonded Arg105 and Ser119 than those of F–G loop.

3.2.3. Model 3

In the initial structure, CBZ formed a hydrogen bond with Phe304 as well as in the case of the initial structure of model 4, although its azepine ring was located in the opposite direction (Fig. 3c and Fig. 5c). During the MD simulation, the direction of CBZ plane was slightly changed. Because of this direction change, in the secondly most frequent pose, the interaction between CBZ and Phe213 became similar to those in model 1 and model 5.

3.2.4. Model 4

There was almost no change in CBZ position during the simulation. However, several hydrogen-bonding patterns appeared (Fig. 5d). Most frequently observed pattern was the interaction with the carboxyl group of Glu308 (Fig. 5d green). Secondly most-frequently observed one was the same as the initial binding mode, that is, hydrogen bonds with Phe304 and with the guanidine group of Arg212 (Fig. 5d blue) were formed. When MD simulation started, CBZ was already within a distance for enabling the access to heme and this CBZ position had been kept during simulation.

3.2.5. Model 5

The motion of CBZ at model 5 was similar to model 1 (Fig. 5e). F–G loop moved to heme pocket as CBZ moved. Keeping its orientation, CBZ became close to the heme iron. Models 1 and 5 demonstrated that the shape of binding pocket of CYP3A4 is adjusted to substrate due to the swing movement of F–G loop. Phe213 in F–G loop formed a hydrogen bond with CBZ. The aromatic rings of CBZ were almost parallel to the heme and π–π interaction was formed.

3.3. Comparison of binding modes determined through MM/PBSA method

Figure 6 shows the time course of calculated $\Delta G_{binding}$ of each model during the whole MD simulation time. It was found that changes of binding mode of CBZ caused the fluctuation in calculated $\Delta G_{binding}$ and the fluctuation ranges were different among models. Any model hardly kept a single stable state through the simulation. The main factor which caused a large fluctuation in calculated $\Delta G_{binding}$ value is the presence or absence of a hydrogen bond(s) between CBZ and CYP3A4. For example, calculated $\Delta G_{binding}$ value of CBZ drastically rose from −21 to −14 kcal/mol at about 1 ns in model 3. This rise was caused by the disruption of two hydrogen bonds between amino group at CBZ and the oxygen at Phe213 main chain and between carbonyl group at CBZ and the nitrogen at Phe213 main chain. Although π–π interaction is not explicitly calculated by MM/PBSA method, the large hydrophobic surface contact between CBZ aromatic rings and heme affected the CBZ position and its binding energy. The hydrophobic interaction was observed in model 5. Calculated $\Delta G_{binding}$ value increased at around 1 to 2 ns and decreased at 4 ns. At first, an aromatic ring of CBZ had a parallel position with respect to heme by the hydrophobic contact. After 1 ns, CBZ moved away from the heme and the hydrophobic contact disappeared and CBZ returned to the position possible for interaction at 4 ns.

4. Discussion

Carbon atoms on CBZ rings were labeled as shown in Fig. 7. Because CBZ has a symmetric structure, carbon atoms were categorized into five groups. As a guide of accessibility of a compound to the heme iron, we calculated the frequency of the carbon position within the distance for oxidization and $\Delta G_{binding}$ value of each binding mode. To determine the proper cut-off distance of a compound atom and the heme iron for oxidation, we surveyed the crystal structures of 12 CYP-substrate complexes reported in 2009. The distances of the metabolized carbon atoms and the heme iron were within 6 Å and the average was 4.14 Å in nine complexes (Table S2). The binding sites of the other three compounds were reported as the primary recognition site[38,77] and the position to inhibit other compounds metabolism.[78] Furthermore, the structures of the intermediate dioxygen complex of cytochrome P450cam with camphor, which were very informative for the cut-off distance, were reported.[79,80] The distance average of three cytochrome P450cam complexes was 4.37 Å. Therefore, the criteria for the oxidation of compound atoms was set within 6 Å from the heme iron

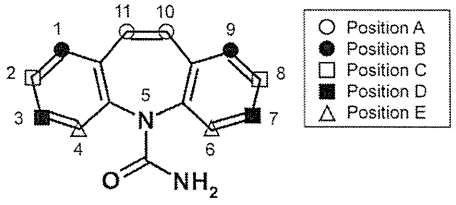

Figure 7. Position labels assigned for CBZ carbon atoms.

by adding 1–2 Å margin. Figure 8 shows frequencies of the carbon position within the distance for oxidization through MD simulations. If a carbon atom is the closest position to the heme iron and the distance is within 6 Å, the black line is drawn at the time point. Among all the positions, position A, which is an epoxidized site, marked the highest frequency when CBZ comes close to the heme iron. The number of counts that position A meets the above criteria was 21,938 through the simulation and it is about 44 percent of trajectories, averaged over all five models. In models 2 and 4, position A was constantly near the heme. The percentage of meeting the criteria of position A in model 4 is over 90 percent and model 2 is about 75 percent. Through the first half of simulation in model 1, carbons at position A were frequently observed near the heme. The secondly most accessed position was position B which accounted for 28 percent of all trajectories. Position B appeared through the last half of simulation of model 1 and 5. The rest of carbon atoms, positions C, D, and E were less accessed to the heme iron than positions A and B. In the first half of simulation of model 5, the carbon atoms of position C were the nearest to the heme iron and the carbon atoms of positions D and E were the nearest in the simulation of model 3.

Furthermore, to compare stability of binding modes among all sites of metabolism, we performed statistical analysis of calculated $\Delta G_{binding}$ for CYP3A4-CBZ complexes with each position group in addition to the frequency analysis. Trajectories whose carbon atom was within 6 Å from the heme iron were selected and separated based on the position of the carbon atom nearest to the iron. Figure 9 shows the averages and standard deviations of the $\Delta G_{binding}$ values by MM/PBSA method by each position group. This figure indicates that the CBZ poses in the position A group have the lowest $\Delta G_{binding}$ value compared to those of other positions. The average calculated $\Delta G_{binding}$ values increased in the order of positions B, C, D, and E. This is almost the same order as the frequency of each position. One of the reasons for the difference in calculated $\Delta G_{binding}$ among five position groups will be the presence of hydrophobic contacts such as π–π interaction between the heme and the aromatic rings of CBZ. Because positions D and E (model 3 in Fig. 5c) are located at the edge of CBZ ring, the aromatic rings of CBZ cannot closely contact to the heme and at most only one aromatic ring of CBZ can form π–π interaction. In contrast, positions A and B are located at the center of CBZ ring and one or two aromatic ring can form parallel or vertical (T-shaped) π–π interactions more easily. The loss of hydrophobic contact will result in the reduction in the binding affinity at positions D and E. The facts that position A was most frequently observed and showed the best averaged

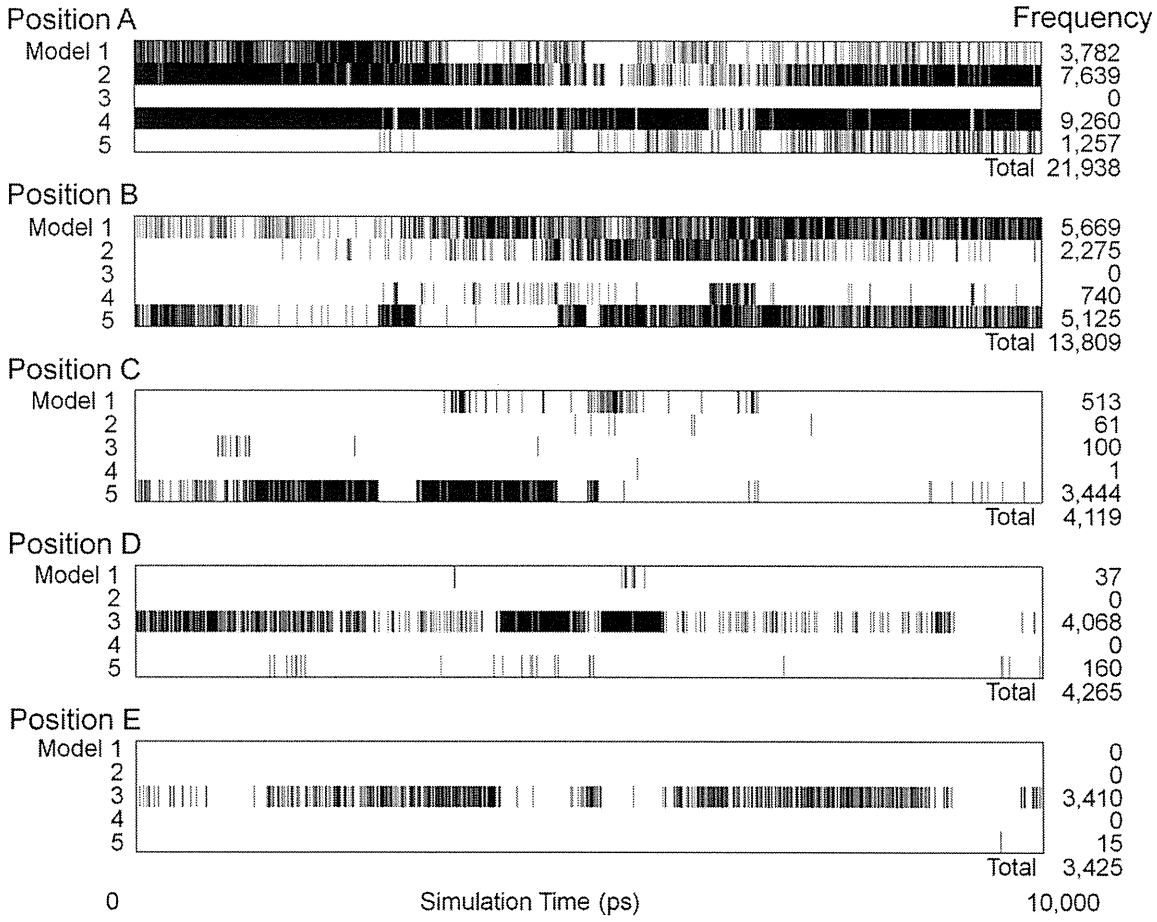

Figure 8. Heat map of the frequencies of the appearance of carbon atoms close to the heme. Black line means the carbons in each position are located within 6 Å from heme iron. The distances were monitored every 1 ps through 10 ns MD simulations.

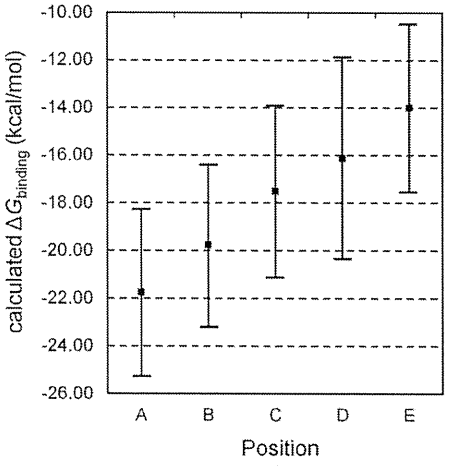

Figure 9. Calculated $\Delta G_{binding}$ value with respect to the carbons in each position, calculated by MM-PBSA method. Average of calculated $\Delta G_{binding}$ value is represented by black square and standard deviation is shown by error bar.

calculated $\Delta G_{binding}$ indicate that the 10, 11-double bond is the most accessible and metabolically hot spot in CBZ by CYP3A4.

From the frequency analysis of the nearest carbon position, the carbon atoms at positions A and B can access to the heme iron more frequently than those of other positions. Furthermore the calculated $\Delta G_{binding}$ of binding mode at position A was lower than position B by 2 kcal/mol. This site of metabolism of CBZ can be predicted by both the accessibility prediction in this study and the reactivity prediction reported by Hata et al.[67] The activation energy for oxidation of positions A and B were 21.3 and 26.9 kcal/mol, respectively. This lower activation energy and lower calculated $\Delta G_{binding}$ value of position A suggest that the oxidation mainly occurred at position A. It is consistent with the experimental result that the position A oxidized metabolite (10, 11-epoxide) is the major product by CYP3A4 and that of position B are scarcely detected 64. Thus, both accessibility and reactivity for the atoms of a given compound can predict SOM and the ratio of each metabolite.

To evaluate the effectiveness of MD simulation using multiple initial structures for flexible CYP3A4, we compared the result from our approach with those from docking and subsequent MD using only one initial structure. Docking scores, the nearest carbons to the heme, and distance to the heme iron of the representative docking poses by Glide docking are summarized in Table 2. The nearest carbon positions suggested by docking simulation were A (models 1 and 4) and D (models 2, 3, and 5). Those are at the central double bond and the edge of side benzene rings. Although model 1 was the docking pose with the best docking score, the distance between the nearest carbon atom and the heme iron was 6.65 Å, which is rather long for metabolizing the carbon atom. In terms of distance, model 4 showed the smallest distance (2.86 Å). However, GlideScore was −6.70 kcal/mol, which is 1.12 kcal/mol larger than that of model 1 (−7.82 kcal/mol). Considering only small distance in models 3 (4.59 Å) and 4 (2.86 Å), docking scores of the both models are almost same (models 3: −6.73 kcal/mol and 4: −6.70 kcal/mol). Glide docking can provide multiple possible sites (A and D in this case) for metabolism, which contains a correct answer. However, it seems difficult to assess accurate accessibility such as distance, frequency, energetic stability and distinguish which position is the major site of metabolism only by docking calculation.

We also docked CBZ using GOLD version 4.0.1.[81,82] We used GoldScore or ChemScore function. Cluster analysis was performed for the CBZ poses and 10 representative complex structures with

Table 2
Nearest carbon position to the heme in each initial structure and the most frequently observed one during MD simulation

Model	Docking (Glide)			Most frequent carbon position during MD simulation
	Glide XP docking score (kcal/mol)	Nearest carbon position	Distance to heme iron (Å)	
1	−7.82	A (>6 Å)	6.65	B
2	−7.54	D (>6 Å)	7.59	A
3	−6.73	D	4.59	D
4	−6.70	A	2.86	A
5	−6.04	D (6 Å)	6.03	B

the highest GoldScore in the respective cluster were selected. Same cluster analysis was also performed using ChemScore. There was no pose in GoldScore cluster representatives whose carbon located within 6 Å from the heme iron. In ChemScore cluster representatives there were six poses near the heme. Two were position B and four were position C (Table S3). These Glide and GOLD results showed a difficulty of the prediction by docking methods only.

Next we assessed the performance of MD calculations using the docking poses as initial structures. Table 3 shows binding free energies and appearances of positions A to E for each MD simulation and their total. For this analysis, we used snapshots in which one carbon atom is accessible to the heme iron (≤6 Å). The positions with the best calculated $\Delta G_{binding}$ or the largest occupancy were in italics. As shown in Table 3, energetically stable and frequently appearing positions were very different among five MD simulations. This suggests that MD simulation with one initial model is not sufficient to explore conformational space and total MD with multi-initial models is more favorable. Even in one MD simulation, multiple accessible positions appeared. For example, in docking models 1 and 2, any carbon atoms are not accessible to the heme iron. However, the MD simulations from the models 1 and 2 produced positions A and B as accessible carbons. It is likely that MD simulations include the effect of the flexibility of CYP3A4 while protein-rigid docking calculations do not.

Supposed that we performed only one MD simulation using one initial structure, it is very difficult to choose an appropriate initial structure. The poses of CBZ in models 1 and 2 will be a little far away from the heme compared to the initial structures. As for model 3 which showed good GlideScore and good CBZ-heme distance, the predicted position were C (calculated $\Delta G_{binding}$ mean) and D (occupancy), which are incorrect answers. Although CBZ was distant from the heme in models 1 and 2, MD simulation resulted in the correct prediction for position A (calculated $\Delta G_{binding}$ mean) or B (occupancy) in model 1 and position A (both) in model 2. This indicates that if a substrate is approximately placed at the right position in CYP3A4, MD simulation can change the CYP conformations to fit the substrate and let the substrate move the appropriate position for metabolization. In contrast, an initial structure with an inappropriate position (model 3) has led us to the incorrect answer even if the initial structure showed strong interaction energy with CYP3A4.

To investigate the effect of the number of initial models on accessibility prediction, the fluctuation of calculated $\Delta G_{binding}$ and occupancy using 1 to 5 initial models were plotted (Table S4, Fig. S5). As the initial models, the models with lower Glide docking score were added one by one. When the number of models is one and two, there is little difference in accessibility of each position in terms of the calculated $\Delta G_{binding}$ fluctuation. The difference appeared in the cases of more than three initial models. On the other hand, occupancy of positions A and B using one and two models were higher than those of other positions. As the number of initial models increases, the difference of positions A and B becomes prominent. Therefore, at least three initial models were required

Table 3
Calculated $\Delta G_{binding}$ value and occupancy of carbon atoms in each position in five MD simulation results and their sum total

Carbon position	Experimental[a]	Glide XP mode	MD from Model 1		MD from Model 2		MD from Model 3		MD from Model 4		MD from Model 5		MD multi-initial models	
		XP docking score (kcal/mol)	Calculated ΔG mean (kcal/mol)	Occupancy (%)	Calculated ΔG mean (kcal/mol)	Occupancy (%)	Calculated ΔG mean (kcal/mol)	Occupancy (%)	Calculated ΔG mean (kcal/mol)	Occupancy (%)	Calculated ΔG mean (kcal/mol)	Occupancy (%)	Calculated ΔG mean (kcal/mol)	Occupancy (%)
A	Major	−6.70 (model 4)	*−20.38*	37.8	*−21.28*	*76.6*	–[c]	0.0	*−22.97*	92.6	−19.82	12.6	*−21.75*	*46.1*
B	N.D.[c]		−19.26	56.7	−20.65	22.8	–[c]	0.0	−19.44	7.4	*−19.94*	*51.2*	−19.75	29.0
C	N.D.[b]	–[c]	−20.19	5.1	−19.35	0.6	*−20.45*	1.3	−22.89	0.0	−16.99	34.4	−17.51	8.7
D	N.D.[b]	*−6.73* (model 3)	−19.65	0.4	–[c]	0.0	−16.13	*53.8*	–[c]	0.0	−15.50	1.6	−16.13	9.0
E	N.D.[c]	–[c]	–[c]	0.0	–[c]	0.0	−13.96	44.9	–[c]	0.0	−19.56	0.2	−13.99	7.2

Occupancy means the percentage of snapshots every 1 ps when the heme iron and nearest CBZ carbon located within 6 Å through 10 ns MD simulation. Italized cells mean the lowest calculated $\Delta G_{binding}$ value or the most frequently observed carbon position in the MD runs.

[a] Data obtained from Ref. 64.
[b] Not detected from cDNA expressed CYP3A4.
[c] The heme iron–CBZ carbon distances are always longer than 6 Å during simulation.

and four or five models were desirable for robust prediction in the case of CBZ. Next we examined the effects of simulation time on calculated $\Delta G_{binding}$ and occupancy using 2–10 ns MD simulation time (Table S6, Fig. S7). All through the simulation time, the position A showed lowest calculated $\Delta G_{binding}$ and highest occupancy. Especially the calculated $\Delta G_{binding}$ and the occupancy almost became constant after 8 ns. Thus, the accessibility prediction appeared to be robust if more than 8 ns MD simulation was executed, although the tendency was observed at 2–6 ns simulation time.

The initial models by docking calculations seem to be local minimum and there would be large energy barriers between the minimums. In particular, if ligand orientations are inverted, it needs about 100 ns MD simulation time to change other binding modes of the compound.[83] According to the studies by Shan et al.[83], Src kinase inhibitors required totaling 35 and 115 μs simulation time and several MD runs to reproduce the sound binding modes of inhibitors. With the increase of the flexibility, it will take more time to simulate all the behavior of compounds in CYP3A4. To apply SOM predictions in drug discovery stage, both the prediction accuracy and speed are required. For this purpose, our protocol using multiple initial structures and 10 ns MD calculations seems practical. We performed five 10 ns MD simulations from different initial structures and the most stable complex structures during the simulations were also different (Fig. 6). Although it is ideal that the correct and most stable pose appears in any simulations from any initial structures, it seems to be difficult to explore substrate binding poses only with one 10 ns simulation in particular for flexible proteins such as CYP3A4. In the case of CBZ, five initial structures and 10 ns MD simulation gave the reasonable prediction. If new compounds for prediction have high flexibility, the required number of initial structures and length of simulation time should be increased. The proper settings of initial structures and simulation time for diverse new compounds will be investigated in the future work.

5. Conclusion

We performed the prediction of substrate accessibility to the heme iron in CYP3A4 by the analysis of MD simulation results using multiple initial structures. In this study, we exemplified well studied carbamazepine (CBZ) as a substrate. The main metabolite of CBZ by CYP3A4 is carbamazepine 10, 11-epoxide. Our approach correctly predicted the carbon atoms experimentally epoxidized by CYP3A4. The predicted carbon atoms are the most accessible ones to the heme iron in terms of calculated $\Delta G_{binding}$ and frequency of appearance. The prediction using only molecular docking methods such as Glide and GOLD could not provide the correct position due to the lack of flexibility of the enzyme. Although performing MD simulations can be a solution to this induced fit problem, the simulation result largely depends on its initial structures. In spite of time consuming calculations, we tried MD simulations using not only one substrate–enzyme complex structure but multiple complexes whose substrates have different binding poses. While each MD simulation from one initial structure generated various accessible sites to the heme, total analysis of five MD simulation results have led us to improve the accuracy for predicting SOM from the most stable and frequent position. Although some MD simulations from one initial structure could reach the same answer, it is difficult to select one appropriate structure. For the SOM prediction of a given compound, firstly the accessible positions to the heme iron with high frequency are identified. The ratio of metabolites from the identified positions can be evaluated from both calculated $\Delta G_{binding}$ and reactivity for the positions. Our strategy using diverse multiple initial structures will be potent method to generate the complex between a flexible protein such as CYP3A4 and a

substrate. In future, we will apply this method to the combinations of other CYP families and drug candidates for enhancing efficient drug discovery.

Acknowledgment

This work was supported by Grant-in-Aid for Scientific Research (C) (KAKENHI (19590137, 22590050)).

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2011.12.004.

References and notes

1. Wrighton, S. A.; Stevens, J. C. *Crit. Rev. Toxicol.* **1992**, *22*, 1.
2. Williams, P. A.; Cosme, J.; Sridhar, V.; Johnson, E. F.; McRee, D. E. *Mol. Cell* **2000**, *5*, 121.
3. Ortiz de Montellano, P. R. *Cytochrome P450: Structure, Mechanism, and Biochemistry*; Springer: New York, 2005.
4. Nelson, D. R. *Hum. Genomics* **2009**, *4*, 59.
5. Bernhardt, R. *Rev. Physiol. Biochem. Pharmacol.* **1996**, *127*, 137.
6. Nebert, D. W.; Russell, D. W. *Lancet* **2002**, *360*, 1155.
7. Ding, X.; Kaminsky, L. S. *Annu. Rev. Pharmacol. Toxicol.* **2003**, *43*, 149.
8. Lewis, D. F. *Pharmacogenomics* **2004**, *5*, 305.
9. Guengerich, F. P. *FASEB J.* **1992**, *6*, 745.
10. Shimada, T.; Yamazaki, H.; Mimura, M.; Inui, Y.; Guengerich, F. P. *J. Pharmacol. Exp. Ther.* **1994**, *270*, 414.
11. Li, A. P.; Kaminski, D. L.; Rasmussen, A. *Toxicology* **1995**, *104*, 1.
12. Guengerich, F. P. *Annu. Rev. Pharmacol. Toxicol.* **1999**, *39*, 1.
13. Evans, W. E.; Relling, M. V. *Science* **1999**, *286*, 487.
14. Li, A. P.; Maurel, P.; Gomez-Lechon, M. J.; Cheng, L. C.; Jurima-Romet, M. *Chem. Biol. Interact.* **1997**, *107*, 5.
15. Li, A. P.; Lu, C.; Brent, J. A.; Pham, C.; Fackett, A.; Ruegg, C. E.; Silber, P. M. *Chem. Biol. Interact.* **1999**, *121*, 17.
16. Wang, R. W.; Newton, D. J.; Liu, N.; Atkins, W. M.; Lu, A. Y. H. *Drug Metab. Dispos.* **2000**, *28*, 360.
17. Lau, W. C.; Waskell, L. A.; Watkins, P. B.; Neer, C. J.; Horowitz, K.; Hopp, A. S.; Tait, A. R.; Carville, D. G. M.; Guyer, K. E.; Bates, E. R. *Circulation* **2003**, *107*, 32.
18. Gombar, V. K.; Silver, I. S.; Zhao, Z. *Curr. Top. Med. Chem.* **2003**, *3*, 1205.
19. Crivori, P.; Zamora, I.; Speed, B.; Orrenius, C.; Poggesi, I. *J. Comput. Aided Mol. Des.* **2004**, *18*, 155.
20. Pelkonen, O.; Turpeinen, M.; Uusitalo, J.; Rautio, A.; Raunio, H. *Basic Clin. Pharmacol. Toxicol.* **2005**, *96*, 167.
21. Wishart, D. S. *Drugs R D* **2007**, *8*, 349.
22. Lee, P. H.; Cucurull-Sanchez, L.; Lu, J.; Du, Y. J. *J. Comput. Aided Mol. Des.* **2007**, *21*, 665.
23. Sakiyama, Y.; Yuki, H.; Moriya, T.; Hattori, K.; Suzuki, M.; Shimada, K.; Honma, T. *J. Mol. Graphics Modell.* **2008**, *26*, 907.
24. Schwaighofer, A.; Schroeter, T.; Mika, S.; Hansen, K.; ter Laak, A.; Lienau, P.; Reichel, A.; Heinrich, N.; Müller, K.-R. *J. Chem. Inf. Model.* **2008**, *48*, 785.
25. Li, H.; Sun, J.; Fan, X.; Sui, X.; Zhang, L.; Wang, Y.; He, Z. *J. Comput. Aided Mol. Des.* **2008**, *22*, 843.
26. Itokawa, D.; Yamauchi, A.; Chuman, H. *QSAR Comb. Sci.* **2009**, *28*, 629.
27. Li, W.; Ode, H.; Hoshino, T.; Liu, H.; Tang, Y.; Jiang, H. *J. Chem. Theory Comput.* **2011**, *5*, 1411.
28. Thompson, T. N. *Curr. Drug Metab.* **2000**, *1*, 215.
29. Thompson, T. N. *Med. Res. Rev.* **2001**, *21*, 412.
30. Nassar, A.-E. F.; Kamel, A. M.; Clarimont, C. *Drug Discovery Today* **2004**, *9*, 1020.
31. Modi, S.; Paine, M. J.; Sutcliffe, M. J.; Lian, L.-Y.; Primrose, W. U.; Wolf, C. R.; Roberts, G. C. K. *Biochemistry* **1996**, *35*, 4540.
32. Szklarz, G. D.; Halpert, J. R. *J. Comput. Aided Mol. Des.* **1997**, *11*, 265.
33. Afzelius, L.; Zamora, I.; Ridderström, M.; Andersson, T. B.; Karlén, A.; Masimirembwa, C. M. *Mol. Pharmacol.* **2001**, *59*, 909.
34. Ekins, S.; de Groot, M. J.; Jones, J. P. *Drug Metab. Dispos.* **2001**, *29*, 936.
35. Ridderström, M.; Zamora, I.; Fjellström, O.; Andersson, T. B. *J. Med. Chem.* **2001**, *44*, 4072.
36. Venhorst, J.; ter Laak, A. M.; Commandeur, J. N. M.; Funae, Y.; Hiroi, T.; Vermeulen, N. P. E. *J. Med. Chem.* **2003**, *46*, 74.
37. Yano, J. K.; Wester, M. R.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. *J. Biol. Chem.* **2004**, *279*, 38091.
38. Williams, P. A.; Cosme, J.; Vinković, D. M.; Ward, A.; Angove, H. C.; Day, P. J.; Vonrhein, C.; Tickle, I. J.; Jhoti, H. *Science* **2004**, *305*, 683.
39. Ekroos, M.; Sjögren, T. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13682.
40. Torimoto, N.; Ishii, I.; Toyama, K.; Hata, M.; Tanaka, K.; Shimomura, H.; Nakamura, H.; Ariyoshi, N.; Ohmori, S.; Kitada, M. *Drug Metab. Dispos.* **2007**, *35*, 484.
41. Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. *J. Med. Chem.* **2005**, *48*, 4040.
42. Ferrara, P.; Curioni, A.; Vangrevelinghe, E.; Meyer, T.; Mordasini, T.; Andreoni, W.; Acklin, P.; Jacoby, E. *J. Chem. Inf. Model.* **2006**, *46*, 254.
43. Okimoto, N.; Futatsugi, N.; Fuji, H.; Suenaga, A.; Morimoto, G.; Yanai, R.; Ohno, Y.; Narumi, T.; Taiji, M. *PLoS Comput. Biol.* **2009**, *5*, e1000528.
44. Li, W.; Shen, J.; Liu, G.; Tang, Y.; Hoshino, T. *Proteins* **2011**, *79*, 271.
45. Hritz, J.; de Ruiter, A.; Oostenbrink, C. *J. Med. Chem.* **2008**, *51*, 7469.
46. Totrov, M.; Abagyan, R. *Curr. Opin. Struct. Biol.* **2008**, *18*, 178.
47. Park, H.; Lee, S.; Suh, J. *J. Am. Chem. Soc.* **2005**, *127*, 13634.
48. Otyepka, M.; Skopalík, J.; Anzenbacherová, E.; Anzenbacher, P. *Biochim. Biophys. Acta.* **2007**, *1770*, 376.
49. Skopalík, J.; Anzenbacher, P.; Otyepka, M. *J. Phys. Chem. B* **2008**, *112*, 8165.
50. Sykes, M. J.; McKinnon, R. A.; Miners, J. O. *J. Med. Chem.* **2008**, *51*, 780.
51. Genheden, S.; Ryde, U. *J. Comput. Chem.* **2011**, *32*, 187.
52. Zwier, M. C.; Chong, L. T. *Curr. Opin. Pharmacol.* **2010**, *10*, 745.
53. Seifert, A.; Tatzel, S.; Schmid, R. D.; Pleiss, J. *Proteins* **2006**, *64*, 147.
54. Sadiq, S. K.; Wright, D. W.; Kenway, O. A.; Coveney, P. V. *J. Chem. Inf. Model.* **2010**, *50*, 890.
55. Stjernschantz, E.; Oostenbrink, C. *Biophys. J.* **2010**, *98*, 2682.
56. Tybring, G.; von Bahr, C.; Bertilsson, L.; Collste, H.; Glaumann, H.; Solbrand, M. *Drug Metab. Dispos.* **1981**, *9*, 561.
57. Eichelbaum, M.; Tomson, T.; Tybring, G.; Bertilsson, L. *Clin. Pharmacokinet.* **1985**, *10*, 80.
58. Bertilsson, L.; Tomson, T. *Clin. Pharmacokinet.* **1986**, *11*, 177.
59. Wolkenstein, P.; Tan, C.; Lecoeur, S.; Wechsler, J.; Garcia-Martin, N.; Charue, D.; Bagot, M.; Beaune, P. *Chem. Biol. Interact.* **1998**, *113*, 39.
60. Pelkonen, O.; Myllynen, P.; Taavitsainen, P.; Boobis, A. R.; Watts, P.; Lake, B. G.; Price, R. J.; Renwick, A. B.; Gómez-Lechón, M. J.; Castell, J. V.; Ingelman-Sundberg, M.; Hidestrand, M.; Guillouzo, A.; Corcos, L.; Goldfarb, P. S.; Lewis, D. F. *Xenobiotica* **2001**, *31*, 321.
61. Fourie, L.; Breytenbach, J. C.; Du Plessis, J.; Goosen, C.; Swart, H.; Hadgraft, J. *Int. J. Pharm.* **2004**, *279*, 59.
62. Bu, H.-Z.; Kang, P.; Deese, A. J.; Zhao, P.; Pool, W. F. *Drug Metab. Dispos.* **2005**, *33*, 1920.
63. Kang, P.; Liao, M.; Wester, M. R.; Leeder, J. S.; Pearce, R. E.; Correia, M. A. *Drug Metab. Dispos.* **2008**, *36*, 490.
64. Kerr, B. M.; Thummel, K. E.; Wurden, C. J.; Klein, S. M.; Kroetz, D. L.; Gonzalez, F. J.; Levy, R. H. *Biochem. Pharmacol.* **1994**, *47*, 1969.
65. Lertratanangkoon, K.; Horning, M. G. *Drug Metab. Dispos.* **1982**, *10*, 1.
66. Maggs, J. L.; Pirmohamed, M.; Kitteringham, N. R.; Park, B. K. *Drug Metab. Dispos.* **1997**, *25*, 275.
67. Hata, M.; Tanaka, Y.; Kyoda, N.; Osakabe, T.; Yuki, H.; Ishii, I.; Kitada, M.; Neya, S.; Hoshino, T. *Bioorg. Med. Chem.* **2008**, *16*, 5134.
68. *Molecular Operating Environment (MOE)*; Chemical Computing Group Inc.: 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, **2010.**
69. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004**, *47*, 1739.
70. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*; Gaussian, Inc., Wallingford, CT, 2004.
71. Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1357.
72. Sano, E.; Li, W.; Yuki, H.; Liu, X.; Furihata, T.; Kobayashi, K.; Chiba, K.; Neya, S.; Hoshino, T. *J. Comput. Chem.* **2010**, *31*, 2746.
73. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
74. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999.
75. Case, D. A.; Darden, T. A.; Cheatham, III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *AMBER 10*; University of California, San Francisco, 2008.
76. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
77. Williams, P. A.; Cosme, J.; Ward, A.; Angove, H. C.; Matak Vinkovic, D.; Jhoti, H. *Nature* **2003**, *424*, 464.
78. Schoch, G. A.; Yano, J. K.; Sansen, S.; Dansette, P. M.; Stout, C. D.; Johnson, E. F. *J. Biol. Chem.* **2008**, *283*, 17227.
79. Schlichting, I.; Berendzen, J.; Chu, K.; Stock, A. M.; Maves, S. A.; Benson, D. E.; Sweet, R. M.; Ringe, D.; Petsko, G. A.; Sligar, S. G. *Science* **2000**, *287*, 1615.
80. Nagano, S.; Poulos, T. L. *J. Biol. Chem.* **2005**, *280*, 31659.
81. Jones, G.; Willett, P.; Glen, R. C. *J. Mol. Biol.* **1995**, *245*, 43.
82. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727.
83. Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. *J. Am. Chem. Soc.* **2011**, *133*, 9181.