counterions were added to neutralize the system. The detailed procedure for MD simulation was adopted as previously described.[33] Briefly, the system was relaxed by energy minimization with decreasing the force constant restraints to protein atoms. After the relaxation, the system was gradually heated from 0 to 300 K over 60 ps under the NVT ensemble condition and equilibrated at 300 K for 100 ps. Finally, a 10-ns MD simulation was conducted under the NPT ensemble at 300 K and 1 atm. The SHAKE algorithm[34] was applied to constrain all bonds involving hydrogen atoms. The Particle Mesh Ewald method[35] was used to calculate long-range electrostatic interactions. A time step of 2 fs and a nonbonding interaction cutoff distance of 10 Å were used. Coordinates were saved every 1 ps during the entire process.

Random acceleration molecular dynamics

The RAMD method[15,18] was applied to the CYP2A6-coumarin complex to discover the possible substrate channels in CYP2A6. In RAMD, a randomly oriented force is exerted to the center of mass of the ligand molecule. The direction of the random force is maintained for a predefined number of MD steps (N). After N steps of MD simulation, the distance that the ligand moved is computed. If the ligand moves less than the predefined distance (r_{min}), a new force direction is chosen randomly, otherwise, the same direction is kept for another N steps. By using the RAMD approach, the ligand is able to unbiasedly search for the possible exit routes in a relatively short timescale (from picoseconds to nanoseconds). The advantage of RAMD is that the enhanced ligand egress does not need to specify the search pathway or direction in advance, which is required in the SMD method. Therefore, RAMD is particularly suitable for simulating those systems with the binding pocket buried in the center of the protein fold. To date, RAMD has been successfully applied to several such systems including P450s,[15,18] G-protein coupled receptor,[36,37] nuclear receptor,[38] and insulin.[39]

It is worth noting that the size and effect of the artificial random force should be kept as low as possible in any RAMD application. After initial trials of effect of different magnitudes of accelerations on substrate egress, two random accelerations, 0.15 and 0.20 kcal/Å/g, were used in the present RAMD simulations. The values were adopted similarly as previous studies.[15] The N was set to 40 and 100 steps, respectively. The r_{min} was set to 0.001, 0.005, and 0.01 Å, separately. For validation of the RAMD simulations, three snapshot structures saved at 8-ns, 9-ns, and 10-ns of above MD simulations were used as the starting points for the RAMD simulations. For each snapshot, eight different RAMD parameter combinations were conducted using different random number generator seeds. This resulted in 40 RAMD simulations for each snapshot and a total of 120 RAMD simulations for the CYP2A6 complex.

Steered molecular dynamics

Compared with RAMD, SMD[25,26] allows a more detailed mapping of the force and energy profiles along the predetermined pathway. SMD simulations were performed under the NPT ensemble and Langevin dynamics, and external restraints were applied to pull the ligand out of the active site of CYP2A6 complex. The pulling directions were determined on the basis of the most frequently observed pathways obtained by RAMD. The pulling direction was defined along a vector formed between the initial position of coumarin and the Cα atom of Pro431 (channel 2c) as well as between the initial position and Cα of Tyr114 (channel 6).The constant-velocity SMD was used in the current simulations. The pulling velocity was set to 0.01 Å/ps, which was slower than those used in some previous SMD simulations on other systems.[40,41] A spring constant of 4 kcal/mol/Å^2 was applied to the center of mass of the substrate based on the stiff spring approximation theory.[42,43] The pulling force at t time,

$$F(t) = 2k(vt - x(t)) \quad (1)$$

where k is the spring constant of the constraint, v is the pulling velocity, $x(t)$ is the position of the ligand at time t. The work W was calculated by integrating the force over the pulled trajectory:

$$W(x(t)) = \int_0^{x(t)} F(t)\mathrm{d}x(t) \quad (2)$$

SMD simulations were done starting from the snapshot structure at 10-ns. To avoid the protein translation and rotation, two Cα atoms of Leu290 and Leu320 in the most rigid helix I and the Fe atom of heme were restrained by a harmonic potential with a force constant of 50 kcal/mol/Å^2. SMD simulations were repeatedly carried out multiple times with different random seeds for computing the force and work.

Construction of potential of mean force

Jarzynski's equality[27] was used to compute the PMF along the reaction coordinate from work distributions. Jarzynski's equality established a relation between equilibrium free energy differences and work done through nonequilibrium process. The formula of Jarzynski's equality is given by

$$\exp(-\beta\Delta F) = \langle \exp(-\beta W_{0\to\lambda}) \rangle_{ave} \quad (3)$$

Since the exponential work in the above equality is dominated by small work values, it requires sufficient trajectories to accurately estimate the free energy. In practice, however, the trajectories usually can not be sampled enough. Therefore, many approximation methods were introduced to improve the free energy calculation with finite sampling data.[42–45] On the basis of the previous

Table I
Statistics Summary of RAMD Simulations of the CYP2A6-Coumarin Complex

Channel	Accel (kcal/Å/g)	*N* (step)	r_{min} (Å)	Trajectory length (ps)	No. of successful egress		
					10-ns	9-ns	8-ns
2c	0.15, 0.20	40 100	0.001, 0.05, 0.01 0.01	124–1434	9	15	7
6	0.15, 0.20	40 100	0.001, 0.05, 0.01 0.01	64–1998	8	5	8
3	0.20	40 100	0.001, 0.05, 0.01 0.01	78–850	2	3	2
2a	0.20	40 100	0.001, 0.05, 0.01 0.01	294–1252	3	2	2
2e	0.20	40 100	0.001, 0.05, 0.01 0.01	464–794	2	0	2
Solvent	0.20	40 100	0.001, 0.05, 0.01 0.01	1080–1442	3	0	0
Other[a]	0.20	40 100	0.001, 0.05, 0.01 0.01	192–1692	1	3	4

[a]Other channels not identified by previous crystal structure analysis or MD simulations.

studies,[42,44,46] with the stiff-spring approximation, the PMF calculation with the second-order cumulant expansion method yielded satisfactory estimation for free energy change within limited sampling sizes. Therefore, the second-order cumulant expansion was used to construct the PMF in this study. The formula is given by:

$$\Delta F = \langle W \rangle - \frac{\beta}{2}(\langle W^2 \rangle - \langle W \rangle^2) \quad (4)$$

For constructing the PMF, SMD trajectories were repeated for 34 times to reduce the statistical error. The uncertainty was estimated by using the bootstrap approach according to the recent study.[45] Briefly, 34 work files were first generated from the SMD simulations. Then, the PMF was estimated via Eq. (4) using 34 work values chosen at random (with replacement) from the above date set containing 34 values. Next, the above step was repeated by a large number of times until the mean and standard deviation of the free energy estimates converged. We used 100,000 trials in this study. Finally, the uncertainty was given by the standard deviation of these PMF estimates.

RESULTS AND DISCUSSION

Channel search in the CYP2A6 crystal structures

Previous analysis on the crystal structure of the CYP2A6-coumarin complex demonstrated that no obvious channels were observed in the complex.[14] To date, nine more crystal structures of CYP2A6 were determined, which include different ligand-bound complexes and multiple mutants. Some of residues (e.g., the Phe-cluster) in the active site exhibit different conformations upon different ligands binding. To detect whether openings are present in these structures, we analyzed all the crystal structures of CYP2A6 by using MOLE,[28] a newly developed tool to explore channels, pores, and tunnels in molecular structures. The analysis results demonstrated that no channels were open enough to allow coumarin to pass in the CYP2A6 crystal structures, which was evidenced by the fact that the bottleneck radii of all the channels identified by MOLE are smaller than 1.4 Å [Supporting Information Table S1]. Compared with other P450s such as CYP3A4 and CYP2C9, which have obvious openings from the protein surface to the active site, CYP2A6 has more packed active site and adopts a completely closed conformation form with no obvious channel(s) for ligand passage. Therefore, other approaches are needed to identify which part(s) will open up to allow the ligand to enter or exit the active site of CYP2A6.

Coumarin egress channels identified by RAMD

The RAMD method, developed by Wade's group,[15] is a very powerful tool to identify the channels of those macromolecules with the active site buried inside the structural fold. Herein, RAMD was applied to CYP2A6 to identify the possible ligand channels. In order to obtain the equilibrium structures for the RAMD simulations, a 10-ns conventional MD simulation was carried out on the 2A6-coumarin complex prior to the RAMD simulations (Supporting Information Fig. S1). Details of RAMD simulations resulting in coumarin egress are summarized in Table I. The substrate was found to exit the active site of CYP2A6 via different pathways. The pathways were defined by the secondary structure elements lining them. For keeping consistency with Wade's nomenclature,[14] these pathways were denoted channel 2a, 2c, 2e, 3, 6, and the solvent channel, as shown in Figure 1(A). Channel 2a is between the F-G loop and β1 sheet. Channel 2c is located between the B' helix/B-C loop, G and I helices. Channel 2e is penetrating through the B-C loop. Channel 3 is positioned between the F and G

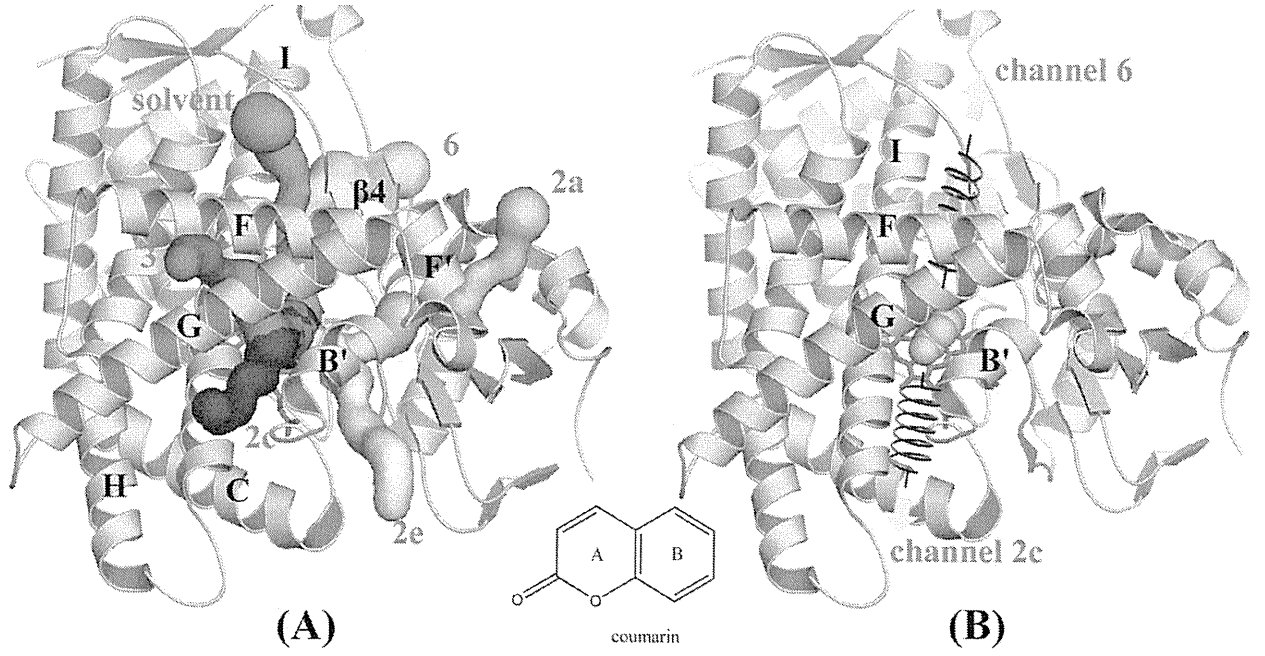

Figure 1

(**A**) Coumarin egress channels from the active site of the substrate-bound CYP2A6 complex, identified by RAMD simulations. The channels are defined according to Wade's nomenclature.[14] The major secondary elements of the complex structure are labeled. (**B**) The two most frequently observed egress channels, channel 2c and channel 6, were chosen for SMD simulations. The bound coumarin was pulled out of the active site using a harmonic potential symbolized by an artificial spring that is connected to the center of mass of coumarin. Coumarin is represented by a Corey-Pauling-Koltun model.

helices. Channel 6 is located between helices I and K and the β4 sheet. The solvent channel is between the E, F, and I helices and the β4 sheet. All the channels found in this study have been previously described except Channel 6, which is a new channel that was not been discovered by previous studies.[14] Although several different channels were found potentially for substrate egress, the egress occurrence frequency of the individual channel differed significantly. Among the successful egress trajectories, two channels, namely channel 2c and channel 6, were most frequently observed for substrate egress. The percents of substrate egress via these two channels are 37% and 26%, respectively. The other channels, channel 2a, 2e, 3, and the solvent channel, were observed rarely (4–10%). In addition, for the smaller acceleration, 0.15 kcal/Å/g, successful egress of coumarin was only observed for channel 2c and channel 6. This means that the substrate preferred exiting via channel 2c and channel 6 under a relatively smaller size of artificial random force. For any RAMD simulation, one should keep the size and effect of artificial random force exerted on the ligand as low as possible in order to reduce the perturbation to the protein structure.

To check whether the initial structures affect the RAMD results, three snapshot structures extracted at the equilibration stage were used as the starting structures for the RAMD simulations by using the same parameter set. On the basis of the statistical results in Table I, it appeared that the observed occurrence frequency has a minor dependence on the starting structures. Coumarin egress frequently took place through channel 2c and channel 6 in the three different structures. The event that the starting structures had little effect on the egress occurrence frequency was also observed in the previous study of the ligand unbinding from the vitamin D receptor.[38]

Channel 2c is lined mainly by residues Val117, Phe111, Phe118, Phe107 in the B' helix/B-C loop, Asn297, Ile300, Leu296, Met293 in the helix I, and Glu245 in the helix F. This channel has been previously suggested to be the most likely for ligand egress in CYP2C5[15] and one of the two possible channels for testosterone egress in CYP2B1,[16] identified by molecular dynamics simulations. Channel 2c is also observed to be open in the ligand-free crystal structures of CYP2C5[47] and CYP2B4.[48] Channel 6 is located the opposite direction of channel 2c. This channel extends from the active site to the surface of the protein along the I helix. Channel 6 has the common pathway in its first terminal part with the so-called solvent channel, which is located between helices I and F and the β4 sheet. Channel 6 is lined by residues close to the C-terminus of the I helix and several residues in the

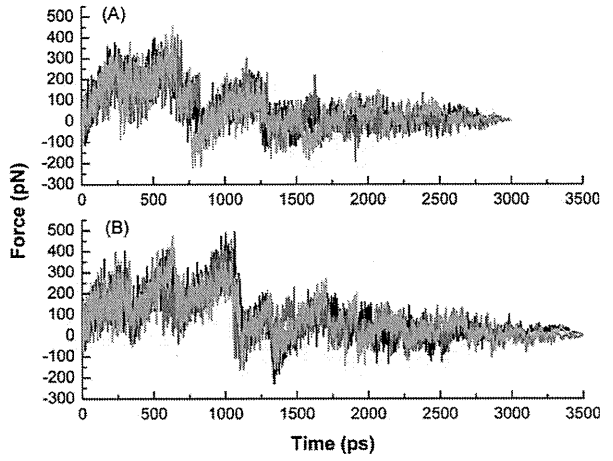

Figure 2

Force profiles in pulling coumarin out of the active site of CYP2A6 along channel 2c (**A**) and channel 6 (**B**). Three colors represent data from three representative SMD simulations along each channel.

K helix and its following loop. These residues include Glu304, Thr305, Thr308, Thr309, Tyr312, Leu316 in the I helix, and Val365, Ile366, Gly363, Phe362, and Phe411 in the K helix and its following loop. Although channel 6 was not identified by previous studies, the individual residue located in this channel has been proved to significantly alter the catalytic activity in other P450s. For instance, substitution of Ile359 with Leu in CYP2C9 showed a dramatic decrease in the catalytic activity of this enzyme toward several substrates.[49] Therefore, the possibility of channel 6 serving as a substrate access/exit channel cannot be ruled out. Nevertheless, the preference of channels as the most likely channel for coumarin egress in CYP2A6 warrants further investigation. This was accomplished by the subsequent SMD simulations and PMF calculations. The following SMD simulations and PMF calculations was based on the 10-ns snapshot structure.

SMD simulations

SMD simulations were performed on the 2A6-coumarin complex along the two most frequently observed channels [Fig. 1(B)], channel 2c and channel 6, identified by RAMD, which aimed at providing the quantitative estimates of rupture forces and characterizing the interactions between the substrate and protein residues during the substrate egress. To obtain reasonable PMF profiles of coumarin dissociation from CYP2A6, a number of external works were sampled by repeating the SMD trajectories along channel 2c and channel 6. Totally, 34 SMD trajectories were performed for each channel. Each SMD simulation was run 3.0 ns for channel 2c and 3.5 ns for channel 6 on the basis of the different channel lengths.

In the crystal structure of the CYP2A6-coumarin complex,[24] coumarin forms a direct hydrogen bond with Asn297 and is confined within the active site consisting of a "Phe-cluster" formed by residues Phe107, Phe111, Phe108, Phe209, and Phe480. During the 10-ns conventional MD simulation, the hydrogen bond between the substrate and Asn297 maintained and the substrate did not significantly deviate from the initial bound state, as shown in Supporting Information Figure S1. Therefore, coumarin first needs to break these strong interactions to exit the active site.

Egress along channel 2c

Figure 2(A) shows the typical force profiles for pulling coumarin out of the active site of CYP2A6 along channel 2c. In the first ~250 ps, a steady increase of the applied force was observed. At this stage, the substrate was tightly confined within the active site above the heme group due to the hydrogen bonding with Asn297 and the strong hydrophobic contacts with the active site residues Phe107, Phe111, Val117, Phe209, Gly301, and Leu370, as shown in Figure 3(A). After ~250 ps, the phenyl B ring of coumarin started to move away from the heme group and thus weakened the hydrophobic interactions with several active site residues, such as Val117, Leu370, and Gly301. This event was manifested by a slight decrease of the force. While the B ring of coumarin moved away, the carbonyl oxygen atom of coumarin still kept the hydrogen bonding with Asn297.

After ~450 ps, the force increased slowly again, which resulted from the additional interactions of coumarin with the residues in the middle of the I helix as well as in the B' and F helices, especially residues Phe209, Leu296, and Ile300 whose hydrophobic side-chains point to the egress pathway. The enhanced van der Waals contacts resulted in a force peak from ~550 to ~700 ps. During this period, the substrate mainly formed hydrophobic interactions with Phe209 in the F helix, Leu296 and Ile300 in the I helix and π-π interactions with the phenyl rings of Phe111 and Phe107 in the B' helix, as shown in Figure 3(B). In addition, both oxygen atoms of coumarin formed hydrogen bonds with Asn297. After ~750 ps, the force has a sharp decrease primarily due to the breakup of the interactions of the substrate with Phe209 and weakened interactions with Phe107 and Ile300. At this point, the B ring of the substrate has moved out of the binding pocket and reached the entrance of channel 2c. The following increase in the force, between ~1000 and ~1250 ps, mainly due to the interactions of the substrate with the residues at the channel entrance. The substrate was sandwiched between Met293 and Phe111. The A ring of coumarin formed the π-π interaction with the phenyl ring of Phe111. At the same time, the hydrogen bond between coumarin and Asn297 remained, as shown in Figure 3(C). After ~1300 ps, the

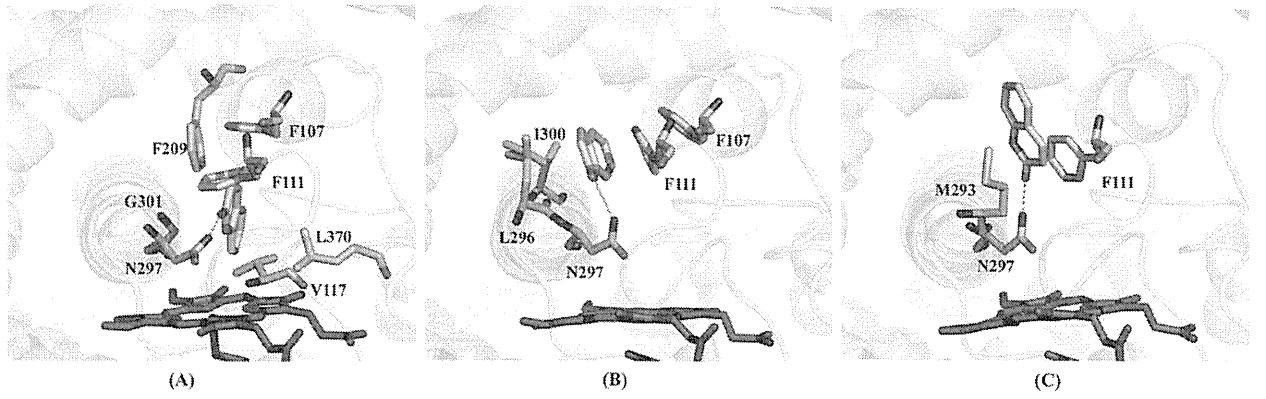

Figure 3

Snapshots of the relative positions of coumarin and the protein throughout the SMD simulation along channel 2c (**A**, 0 ps; **B**, 650 ps; and **C**, 1100 ps). Important residues interacting with the substrate are labeled and shown as sticks. The hydrogen bonds are shown in dotted lines.

hydrogen bond between the substrate and Asn297S broke and the π-π interaction disappeared, indicating the substrate has passed through the bottleneck of the channel. The fluctuation in the force after ~1300 ps was due to the interactions of coumarin with the residues at the protein surface as well as the solvent molecules.

Egress along channel 6

Channel 6 is completely unrelated to the above channel and located on the opposite side of channel 2c. The typical force profiles of coumarin egress along channel 6 are shown in Figure 2(B). In the first ~250 ps, the force has a steady increase, which is similar as that of egress along channel 2c. During this period, coumarin was stabilized by the strong hydrophobic interactions with the active site residues and remained the hydrogen bonding with Asn297. After ~250 ps, the hydrogen bond between coumarin and Asn297 was broken, which led to a decrease in the applied force. After that, the A ring of the substrate first moved away from the heme group. At ~550 ps, the A ring of coumarin reached the location between Phe209 and Phe480. At this time, the substrate formed a new hydrogen bond with the side-chain of Thr212 and the π-π interaction with the aromatic ring of Phe209. At the same time, coumarin had hydrophobic contacts with Val365, Ile366, and Ala481. These interactions were so strong that a large force was necessary to break them.

With the movement of the substrate, the hydrophobic interactions became weakened. After ~700 ps, the B ring of started to enter the channel while the A ring remained the hydrogen bonding with Thr212. After ~1150 ps, the only hydrogen bond was broken and the substrate completely entered the channel. This led to a significant decrease in the force. The force peak from ~1150 to 1300 ps was mainly caused by the hydrophobic interactions of coumarin with residues Thr308, Thr309, and Tyr312 in the I helix, Val365 and Gly363 in the K helix, as well as Thr482 and Pro484 in the β4 sheet region. At ~1600 ps, the substrate encountered two relatively large residues Tyr312 and Phe362 to block its movement. At this time point, the substrate formed the π-π interaction with the aromatic ring of Tyr312. After ~1800 ps, the substrate crossed over the blockage and entered the solvent. Accordingly, the force decreased.

Comparative analysis of coumarin egress via two channels

Figure 2 illustrated that the force profiles of coumarin egress along channel 2c and channel 6 differed significantly. The maximum force peak of coumarin egress along channel 2c occurred at ~600 ps. At this point, the primary hindrance was from the hydrophobic interactions with Phe209, Ile300, and Leu296 and the hydrogen bond with Asn297. Once the interactions were broken, the pathway was relatively smooth. In contrast, substrate egress via channel 6 had several large blockages, each of which corresponds to a peak in the force profile [Fig. 2(B)]. The peaks between ~950 and ~1300 ps indicated substrate egress encountered large hindrance along channel 6 during this time period, and thus a large force was required to break the strong interactions of the substrate with channel residues. We also analyzed the movement of the backbone atoms of those residues lining the channels. For the substrate egress via channel 2c, the maximum RMS displacement was 2.7 Å found for Val110 in the B' helix. By comparison, several residues lining channel 6 had relatively large RMS displacement, for instance, 4.0 Å for Gly363 and 3.4 Å for Asp364 and Phe362. The difference also suggested that only relatively small backbone displacement was sufficient to allow the substrate to leave the binding pocket via channel 2c when compared with channel 6.

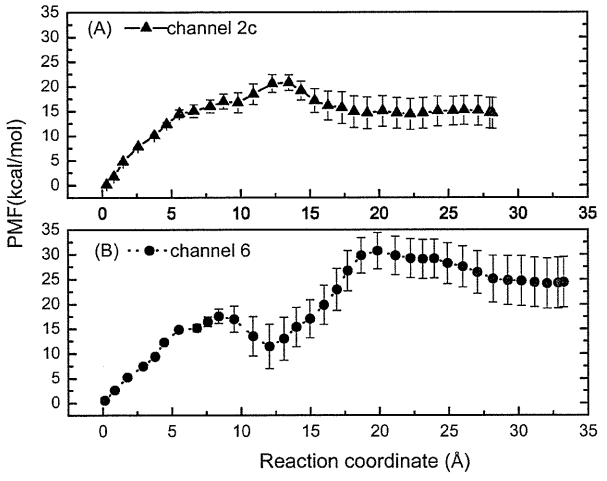

Figure 4

PMF profiles of coumarin egress along channel 2c (triangle) and channel 6 (dot). The error bars indicate the standard deviation of PMF via the bootstrap approach.

PMF construction

To gain more quantitative insights into the relative likelihood of above two channels serving as the substrate egress channel, the energetic changes along the substrate dissociation coordinate were estimated by constructing the PMF. The dissociation of coumarin from the active site by SMD is a nonequilibrium process. Jarzynski's equality,[27] which establishes a relation of the free energy with irreversible work, was used to construct the PMF. However, direct use of Jarzynski's equality needs to sample sufficient work values along the reaction coordinate. This cannot be attainable in practice. Several methods were put forward to solve the problem.[44,45] With the stiff-spring approximation, we utilized the second-order cumulant expansion method to calculate the PMF from finite sampling sizes.

Figure 4 shows the PMF profiles of coumarin dissociation along channel 2c and channel 6. The PMFs exhibit the different profiles along the two channels. The PMF along channel 6 displays a steep increase after 12 Å of pulling. This is mainly due to the different exit behaviors along the two channels. When escaping along channel 2c, the B ring of coumarin first left the heme group while keeping the hydrogen bonding with Asn297. Once breaking the shackles of the active site residues, coumarin was able to cross over the channel smoothly. In contrast, when coumarin escaping along channel 6, it required first breaking the hydrogen bond between coumarin and Asn297 so that the A ring of coumarin moved away from heme prior to the B ring. The A ring formed a new hydrogen bond with Thr212 when the substrate did not yet move out of the binding pocket. Because channel 6 is long and narrow, coumarin had strong hydrophobic contacts with the residues in channel 6. On the other hand, the clash between coumarin and the channel entrance triggered the loop following helix K to open greatly. All these facts contributed to the relatively large energy barrier when coumarin escaped along channel 6. This makes channel 6 less likely serving as a channel for coumarin escape.

The PMFs indicated free energy difference of ~24.4 ± 5.0 kcal/mol for channel 6 and ~14.6 ± 3.1 kcal/mol for channel 2c. Both overestimated the dissociation energy of coumarin from CYP2A6 derived from the experimentally determined K_d (~7.8–8.8 kcal/mol at 300 K depending on different assay methods).[4] This is likely due to the insufficient sampling and nonequilibrium effects, as interpreted in several previous studies,[39,50,51] in which estimated free energy via nonequilibrium simulations overestimated the experimental dissociation free energy by a factor one or two. However, we aimed at the difference between the two channels to suggest the most likely channel for coumarin egress rather than at the real binding free energies. The differentiated PMF profiles between channel 6 and channel 2c clearly indicated the preference of the channels for coumarin egress in CYP2A6. It is noteworthy that free energy is a state function, the two channels should give the same free energy difference upon coumarin completely unbinding from the binding pocket. However, as pointed out by the researchers in a previous study,[41] the substrate interaction with different part of the protein and incomplete relaxation of the protein following the removal of coumarin resulted in the final energy difference between the two channels.

In light of statistical results of RAMD, the force profiles, structural changes of SMD simulations, as well as the PMF comparison, we conclude that channel 2c is the favorable pathway for coumarin egress in CYP2A6.

Mechanism of channel 2c opening

Coumarin egress along channel 2c involves a rotation of the side-chain of Phe111 and a relatively slight backbone movement of the B' helix. The crystal structure of the CYP2A6-coumarin complex shows that Phe111 covers above the substrate and constitutes the roof of the active site with other phenylalanine residues, such as Phe107 and Phe209. No obvious channels are found in the crystal structure of CYP2A6. Analysis of the substrate egress trajectories of both RAMD and SMD simulations along channel 2c revealed that the side-chain of Phe111 needed to rotate in order to make enough space for coumarin to pass. Phe111 is located at the entrance of channel 2c and appears to act as a gatekeeper to regulate substrate egress. To quantify the ring rotation of Phe111, the change of the side-chain torsion χ_2 (CA, CB, CG, and CD1) of the residue was monitored during a representative SMD simulation, as shown in Figure 5. In the first ~700 ps, the torsion angle of Phe111 remains around 125°. In the following ~600 ps, the torsion angle of Phe111 changes from ~150°

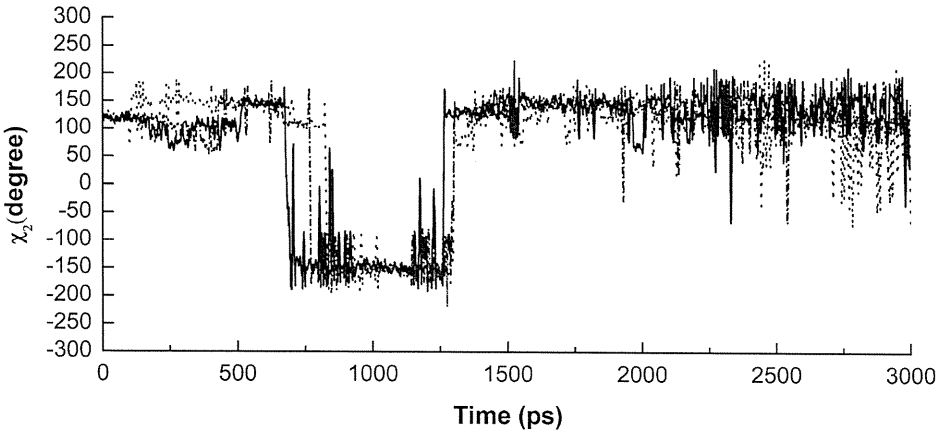

Figure 5

Variations of the side-chain torsion χ_2 (CA, CB, CG, and CD1) of Phe111 with respect to simulation time during coumarin egress along channel 2c. Three different line types represent data from three representative SMD simulations.

to $\sim -150°$. After the substrate passes through the bottleneck, the torsion angle of Phe111 rotated back and fluctuated around $\sim 125°$. The same event has been observed in other P450s including CYP101,[17] 2B1,[16] and 3A4[12,13] during their ligands egress. A similar phenomenon also occurred in other systems.[39]

In addition to the rotation of Phe111, the helix B'-C region underwent a slight swing to expand the channel so as to allow coumarin to pass, as shown in Figure 6. The helix B'-C region exhibits considerable flexibility in several mammalian P450s upon different ligands binding, such as CYP2C9,[52] 2C5,[53] and 2B4.[48,54] It is worth noting that there are two conserved GlyXGly motifs flanking the region between helix B' and the B'-C loop in most family 2 P450s including CYP2A6. The peptide backbone of glycine can readily adopt a wide range of conformations due to the lack of a side chain. This increases the flexibility of the helix B'-C region and thereby facilitates the opening of channel 2c.

Noticeably, the present result found in CYP2A6 in conjunction with previous findings in 2B1 and 2C5, three of which are family 2 P450s with two conserved GlyXGly motifs in helix B'-C region, points to channel 2c being the substrate dissociation channel. This makes us speculate that channel 2c might be the common substrate channel in the family 2 P450s with the two conserved GylXGly motifs in helix B'-C region. However, this idea warrants further investigation on other P450s.

Further evidence of channel 2c as a ligand channel

Channel 2c serving as a ligand channel has been reported in several previous studies of mammalian P450s. They include our own work about CYP2B1[16] and the study of CYP2C5 from Wade's group.[15] Channel 2c was shown to be the predominant egress channel for several ligands in CYP2C5. Crystal structures analysis also demonstrated that channel 2c was open in multiple mammalian P450s including CYP2B4,[48] 2C5,[47] 2C8,[55] and 2C9.[56] Furthermore, site-directed mutagenesis data for CYP2B1 also provide evidence for a substrate access/exit channel in the channel 2c region.[9] Replacement of several residues in the N-terminal part of helix I by other residues was shown to significantly affect the catalytic activity of CYP2B1 toward some substrates. For example, the mutants I290F and L293A abolished the catalytic activity of CYP2B1 toward testosterone 16β-hydroxylation.[9] These two residues correspond to Met293 and

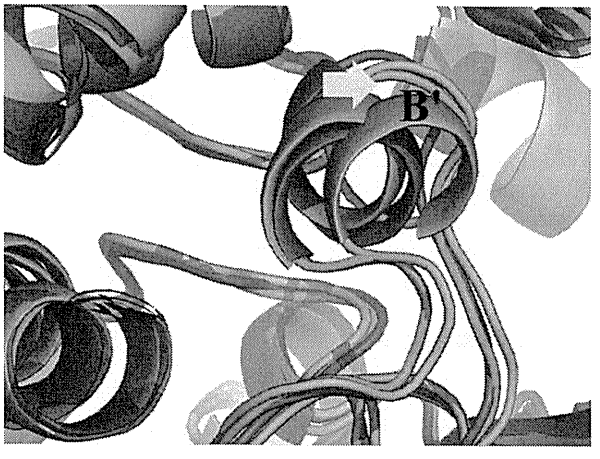

Figure 6

Movement of B' helix during coumarin egress along channel 2c. Green, starting structure; Cyan, at 750 ps; and Purple, at 1400 ps. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Leu296 of CYP2A6 in their alignment positions, both of which were identified to be important for coumarin egress. Phe108 in the B' helix, corresponding to Phe107 in 2A6, had dramatic effect on the stereoselectivity of 2B1 toward testosterone when substituted by alanine.[8]

CONCLUSIONS

An interesting and important issue about CYP2A6 is how its substrate coumarin dissociates from the highly packed active site, which is deeply buried at the center of structural fold. Geometric analysis of the available crystal structures of CYP2A6 demonstrated that no obvious channels were open enough to allow substrates to pass through. Therefore, RAMD, an improved MD approach, was utilized to explore the possible pathways for coumarin egress. Several plausible channels were found by RAMD. Two different channels, namely channel 2c and channel 6, were most frequently observed. The former is through helices I and G and the helix B'-C region and the latter is located between helices I and K and β4. Multiple SMD simulations were further exploited to estimate the rupture forces and characterize the intermediate states of coumarin egress. Both SMD results and PMF calculations indicated channel 2c was the most likely channel for coumarin egress in CYP2A6. The opening of channel 2c involved rotations of Phe111 and displacement of helix B'. Since two conserved GlyXGly motifs flank the helix B'/B'-C loop, the increased flexibility of this region renders the opening of channel 2c easier. This may be related to the kinetic results that a rapid binding and release dynamic behavior of coumarin occurred in its oxidation by CYP2A6. The identified channel 2c is common in other family 2 P450s, such as 2C5 and 2B1. These data will assist in understanding of the catalytic mechanism of CYP2A6 and gain further insight into the channel selectivity of P450s.

ACKNOWLEDGMENTS

The authors thank Dr. Rebecca Wade for her valuable discussion about the usage of RAMD and her kindness in sharing the RAMD code (http://projects.eml.org/mcm/software/amber). They are grateful to Dr. F. Marty Ytreberg for his kindness in offering us the script for calculating the uncertainty of PMF. Usage of AMBER is gratefully acknowledged.

REFERENCES

1. Nebert DW, Russell DW. Clinical importance of the cytochromes P450. Lancet 2002;360:1155–1162.
2. Denisov IG, Makris TM, Sligar SG, Schlichting I. Structure and chemistry of cytochrome P450. Chem Rev 2005;105:2253–2277.
3. Isin EM, Guengerich FP. Kinetics and thermodynamics of ligand binding by cytochrome P450 3A4. J Biol Chem 2006;281:9127–9136.
4. Yun CH, Kim KH, Calcutt MW, Guengerich FP. Kinetic analysis of oxidation of coumarins by human cytochrome P450 2A6. J Biol Chem 2005;280:12279–12291.
5. Wade RC, Motiejunas D, Schleinkofer K, Sudarko, Winn PJ, Banerjee A, Kariakin A, Jung C. Multiple molecular recognition mechanisms. Cytochrome P450–a case study Biochim Biophy Acta 2005;1754:239–244.
6. Wade RC, Winn PJ, Schlichting I, Sudarko . A survey of active site access channels in cytochromes P450. J Inorg Biochem 2004;98:1175–1182.
7. Winn PJ, Ludemann SK, Gauges R, Lounnas V, Wade RC. Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine. Proc Natl Acad Sci USA 2002;99:5361–5366.
8. Honma W, Li W, Liu H, Scott EE, Halpert JR. Functional role of residues in the helix B' region of cytochrome P450 2B1. Arch Biochem Biophys 2005;435:157–165.
9. Scott EE, Liu H, Qun He Y, Li W, Halpert JR. Mutagenesis and molecular dynamics suggest structural and functional roles for residues in the N-terminal portion of the cytochrome P450 2B1 I helix. Arch Biochem Biophys 2004;423:266–276.
10. Domanski TL, Halpert JR. Analysis of mammalian cytochrome P450 structure and function by site-directed mutagenesis. Curr Drug Metab 2001;2:117–137.
11. Fishelovitch D, Shaik S, Wolfson HJ, Nussinov R. How does the reductase help to regulate the catalytic cycle of cytochrome P450 3A4 using the conserved water channel? J Phys Chem B 2010;114:5964–5970.
12. Fishelovitch D, Shaik S, Wolfson HJ, Nussinov R. Theoretical characterization of substrate access/exit channels in the human cytochrome P450 3A4 enzyme: involvement of phenylalanine residues in the gating mechanism. J Phys Chem B 2009;113:13018–13025.
13. Li W, Liu H, Luo X, Zhu W, Tang Y, Halpert JR, Jiang H. Possible pathway(s) of metyrapone egress from the active site of cytochrome P450 3A4: a molecular dynamics simulation. Drug Metab Dispos 2007;35:689–696.
14. Cojocaru V, Winn PJ, Wade RC. The ins and outs of cytochrome P450s. Biochim Biophy Acta 2007;1770:390–401.
15. Schleinkofer K, Sudarko, Winn PJ, Ludemann SK, Wade RC. Do mammalian cytochrome P450s show multiple ligand access pathways and ligand channelling? EMBO Rep 2005;6:584–589.
16. Li W, Liu H, Scott EE, Grater F, Halpert JR, Luo X, Shen J, Jiang H. Possible pathway(s) of testosterone egress from the active site of cytochrome P450 2B1: a steered molecular dynamics simulation. Drug Metab Dispos 2005;33:910–919.
17. Ludemann SK, Lounnas V, Wade RC. How do substrates enter and products exit the buried active site of cytochrome P450cam? II. Steered molecular dynamics and adiabatic mapping of substrate pathways. J Mol Biol 2000;303:813–830.
18. Ludemann SK, Lounnas V, Wade RC. How do substrates enter and products exit the buried active site of cytochrome P450cam? I. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms. J Mol Biol 2000;303:797–811.
19. Podust LM, Poulos TL, Waterman MR. Crystal structure of cytochrome P450 14alpha -sterol demethylase (CYP51) from *Mycobacterium tuberculosis* in complex with azole inhibitors. Proc Natl Acad Sci USA 2001;98:3068–3073.
20. Yun CH, Shimada T, Guengerich FP. Purification and characterization of human liver microsomal cytochrome P-450 2A6. Mol Pharmacol 1991;40:679–685.
21. Yano JK, Denton TT, Cerny MA, Zhang XD, Johnson EF, Cashman JR. Synthetic inhibitors of cytochrome P-450 2A6: inhibitory activity, difference spectra, mechanism of inhibition, and protein cocrystallization. J Med Chem 2006;49:6987–7001.
22. Pianezza ML, Sellers EM, Tyndale RF. Nicotine metabolism defect reduces smoking. Nature 1998;393:750.
23. Guengerich FP, Sorrells JL, Schmitt S, Krauser JA, Aryal P, Meijer L. Generation of new protein kinase inhibitors utilizing cytochrome p450 mutant enzymes for indigoid synthesis. J Med Chem 2004;47:3236–3241.

24. Yano JK, Hsu MH, Griffin KJ, Stout CD, Johnson EF. Structures of human microsomal cytochrome P450 2A6 complexed with coumarin and methoxsalen. Nat Struct Mol Biol 2005;12:822–823.
25. Grubmuller H, Heymann B, Tavan P. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. Science 1996;271:997–999.
26. Isralewitz B, Gao M, Schulten K. Steered molecular dynamics and mechanical functions of proteins. Curr Opin Struct Biol 2001;11:224–230.
27. Jarzynski C. Nonequilibrium equality for free energy differences. Phys Rev Lett 1997;78:2690–2693.
28. Petrek M, Kosinova P, Koca J, Otyepka M. MOLE: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels. Structure 2007;15:1357–1363.
29. Hui Li, Robertson AD, Jensen JH. Very fast empirical prediction and interpretation of protein pKa values. Proteins 2005;61:704–721.
30. Gilson MK, Sharp KA, Honig BH. Calculating the electrostatic potential of molecules in solution. J Comput Chem 1988;9:327–335.
31. Case DA, Cheatham TE, III, Darden T, Gohlke H, Luo R, Merz KM, Jr, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. J Comput Chem 2005;26:1668–1688.
32. Jorgensen WL, Chandrasekhar J, Madura J, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
33. Li W, Ode H, Hoshino T, Liu H, Tang Y, Jiang H. Reduced catalytic activity of P450 2A6 mutants with coumarin: a computational investigation. J Chem Theory Comput 2009;5:1411–1420.
34. Rychaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. J Comput Phys 1977;23:327–341.
35. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh ewald potential. J Chem Phys 1995;103:8577–8592.
36. Wang T, Duan Y. Chromophore channeling in the G-protein coupled receptor rhodopsin. J Am Chem Soc 2007;129:6970–6971.
37. Wang T, Duan Y. Ligand entry and exit pathways in the beta2-adrenergic receptor. J Mol Biol 2009;392:1102–1115.
38. Perakyla M. Ligand unbinding pathways from the vitamin D receptor studied by molecular dynamics simulations. Eur Biophys J 2009;38:185–198.
39. Vashisth H, Abrams CF. Ligand escape pathways and (un)binding free energy calculations for the hexameric insulin-phenol complex. Biophys J 2008;95:4193–4204.
40. Shen J, Li W, Liu G, Tang Y, Jiang H. Computational insights into the mechanism of ligand unbinding and selectivity of estrogen receptors. J Phys Chem B 2009;113:10436–10444.
41. Zhang D, Gullingsrud J, McCammon JA. Potentials of mean force for acetylcholine unbinding from the alpha7 nicotinic acetylcholine receptor ligand-binding domain. J Am Chem Soc 2006;128:3019–3026.
42. Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K. Free energy calculation from steered molecular dynamics simulations using Jarzynsk's equality. J Chem Phys 2003;19:3559–3566.
43. Park S, Schulten K. Calculating potentials of mean force from steered molecular dynamics simulations. J Chem Phys 2004;120:5946–5961.
44. Jensen MO, Park S, Tajkhorshid E, Schulten K. Energetics of glycerol conduction through aquaglyceroporin GlpF. Proc Natl Acad Sci USA 2002;99:6731–6736.
45. Ytreberg FM, Zuckerman DM. Efficient use of nonequilibrium measurement to estimate free energy differences for molecular systems. J Comput Chem 2004;25:1749–1759.
46. Ytreberg FM. Absolute FKBP binding affinities obtained via nonequilibrium unbinding simulations. J Chem Phys 2009;130:164906.
47. Williams PA, Cosme J, Sridhar V, Johnson EF, McRee DE. Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. Mol Cell 2000;5:121–131.
48. Scott EE, He YA, Wester MR, White MA, Chin CC, Halpert JR, Johnson EF, Stout CD. An open conformation of mammalian cytochrome P450 2B4 at 1.6-Å resolution. Proc Natl Acad Sci USA 2003;100:13196–13201.
49. Takanashi K, Tainaka H, Kobayashi K, Yasumori T, Hosakawa M, Chiba K. CYP2C9 Ile359 and Leu359 variants: enzyme kinetic study with seven substrates. Pharmacogenetics 2000;10:95–104.
50. Grater F, de Groot BL, Jiang H, Grubmuller H. Ligand-release pathways in the pheromone-binding protein of *Bombyx mori*. Structure 2006;14:1567–1576.
51. Cuendet MA, Michielin O. Protein-protein interaction investigated by steered molecular dynamics: the TCR-pMHC complex. Biophys J 2008;95:3575–3590.
52. Wester MR, Yano JK, Schoch GA, Yang C, Griffin KJ, Stout CD, Johnson EF. The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-Å resolution. J Biol Chem 2004;279:35630–35637.
53. Wester MR, Johnson EF, Marques-Soares C, Dijols S, Dansette PM, Mansuy D, Stout CD. Structure of mammalian cytochrome P450 2C5 complexed with diclofenac at 2.1 Å resolution: evidence for an induced fit model of substrate binding. Biochemistry 2003;42:9335–9345.
54. Scott EE, White MA, He YA, Johnson EF, Stout CD, Halpert JR. Structure of mammalian cytochrome P450 2B4 complexed with 4-(4-chlorophenyl)imidazole at 1.9-A resolution: insight into the range of P450 conformations and the coordination of redox partner binding. J Biol Chem 2004;279:27294–27301.
55. Schoch GA, Yano JK, Wester MR, Griffin KJ, Stout CD, Johnson EF. Structure of human microsomal cytochrome P450 2C8. Evidence for a peripheral fatty acid binding site J Biol Chem 2004;279:9497–9503.
56. Williams PA, Cosme J, Ward A, Angove HC, Matak Vinkovic D, Jhoti H. Crystal structure of human cytochrome P450 2C9 with bound warfarin. Nature 2003;424:464–468.

Computational Analysis on the Binding of Epitope Peptide to Human Leukocyte Antigen Class I Molecule A*2402 Subtype

MD Iqbal Mahmood, Yuri Matsuo, Saburo Neya, and Tyuji Hoshino*

Graduate School of Pharmaceutical Sciences, Chiba University; 1–33 Yayoi-cho, Inage-ku, Chiba 263–8522, Japan.
Received June 15, 2011; accepted July 25, 2011; published online July 26, 2011

Immunological response induced by small amino peptide has attracted much recent attention in the field of immunotherapy. Wilms' tumor (WT1) protein is one of the potent tumor antigens inducing immunological response in mouse and human, because WT1 is over expressed in many types of leukemia and various kinds of solid tumors. A 9-mer peptide encoded in WT1 protein (CMTWNQMNL; amino acid 235—243) is known to serve as antigenic peptide for human leukocyte antigen (HLA)-A*2402 molecule. It was reported that the replacement of the second amino residue, which is deeply responsible for the peptide binding to HLA, induced strong immunological response compared to the natural peptide. In this study, 19 kinds of single amino substitutions were introduced at position 2 of this 9-mer WT1 peptide. We performed molecular dynamics simulation on the complex of each of WT1 epitope peptides and HLA-β2 micro globulin (β2m) molecule, and subsequently estimated the binding affinity using molecular mechanics/generalized-Born surface area method combined with normal mode analysis. Our computation indicated that the peptide containing M2Y or M2W mutation showed high binding affinity to the HLA-β2m molecule as well as the natural peptide. We have also examined the role of the residue at position 2 in peptide binding to HLA-β2m. The calculation showed that van der Waals interaction between the side chain of the residue at position 2 and hydrophobic residues inside B-pocket of HLA are important. These findings will be helpful to search other potent peptides that will enhance strong immunological response specific to HLA-A*2402 molecule.

Key words human leukocyte antigen; molecular dynamics simulation; binding affinity; epitope peptide; anchor residue

Major histocompatibility complex (MHC) is a transmembrane glycoprotein that plays an important role in immunological system. Human MHC molecule is usually called as human leukocyte antigen (HLA) and HLA molecules are grouped into class I and II. MHC class I molecule is a heterodimer of heavy chain called α chain, whose mass-weight is 45 kDa, and light chain called β2 micro globulin (β2m) with a mass-weight of 12 kDa. A complex of an HLA and a peptide derived from antigen is displayed on the surface of nucleated cell or platelet. HLA heavy chain has a quite large diversity and thousands of genes have been detected so far, and the number of known genes for HLA heavy chain is still increasing. Each individual has two or three kinds of HLA out of several thousands of HLA genes.

Viral protein or cancer-related antigens are detected as foreign molecules in a cell and dissociated by proteasome into peptide fragments consisting of 6—15 amino residues.[1] This peptide fragment is bound to HLA class I molecule and the complex is carried to the cell surface in antigen display cells. Cytotoxic T lymphocytes (CTL) expressing T cell receptor recognize the complex, which results in inducing the immunological response to attack the virus-infected or oncogenic cells to exclude virus or tumor.[2]

Immunological response has attracted much recent attention because of its potential for immunotherapy. At present, several peptides are assumed to be effective for medical treatment of cancer, leukemia, hepatitis C, and applied to patients in a clinical trial as vaccine through the controlled vaccination.[3—5] This peptide-vaccine therapy is expected to lead the regression of disease-related cells without damaging normal tissues. Since the combination of HLA genes has a large diversity, the selection of peptide adequate for each individual is one of the important issues to enhance the performance of immunotherapy.

The discovery of a peptide inducing strong immunological response is a key factor in immunotherapy. Since the peptide firmly bound to HLA molecule will be effective as medical material, the search of potent peptide, usually called as epitope, is critically important. In this study, we focus on Wilms' tumor (WT1) protein.[6] WT1 protein is encoded in β2 Wilms' tumor gene and overexpressed in leukemic and solid tumor cell. A 9-mer amino acid peptide encoded in WT1 protein was already known to work as an antigenic peptide for HLA-A*2402 molecule. Oka and his co-workers reported the experimental findings that the replacement of the second amino acid residue, which is considered to be deeply responsible for the peptide binding to HLA-A*2402, induced strong immunological response of WT1 specific CTLs compared to the natural peptide. A peptide containing a single amino residue mutation is currently applies for the clinical trial for a vaccination against solid tumor or leukemia.[7]

Computer simulation enables us to visualize the interaction of proteins or the interaction of a chemical compound and its target protein. Molecular dynamics (MD) is one of the major computational approaches for analyzing the interaction between epitope peptide and HLA molecules to predict a suitable epitope inducing immunological response. The feasibility of MD simulation was examined through a comparison with experimental methods.[8,9] In this work, we employed MD simulation to analyze the interaction between HLA molecule and epitope peptide. Atoms in the molecules are described as charged particles, and motions of the respective particles can be predicted in a time-evolution manner with calculating the forces among charged particles. Hence MD simulation produces a virtual system in which physiological conditions such as temperature and pressure are set equal to the designated experimental setup.

We also performed computational analysis for estimating

* To whom correspondence should be addressed. e-mail: hoshino@chiba-u.jp

the binding affinity between epitope peptide and HLA-β2m heterodimer. Several computer programs to predict the binding affinity of peptides to HLA were already released so far. MHCpred[10] and PepDict[11] are well known programs for HLA-peptide prediction. Those programs were developed on the knowledge-based approach or the neural network method making use of *in vitro* experimental data. These programs effectively and accurately predict the peptide binding with respect to the known kinds of HLAs. However, the accuracy of the prediction is insufficient for variety kinds of HLAs, especially for the HLA of which little experimental data are available. In this study, we carried out molecular dynamics simulation, followed by molecular mechanics generalized Born/surface area (MM-GB/SA) method and normal mode analysis.[12] MD simulation provides a stable molecular structure of a complex of HLA-β2m heterodimer and epitope peptide. MM-GB/SA method and normal mode analysis gives the binding free energy of the peptide to HLA, which directly corresponds to the binding affinity. The present computation is not knowledge-based approach and then requires no experimental data. Hence, the calculation method employed in this study will be applied to not only to major HLAs but also rare HLA molecules.

In the first step of this study, we built calculation models for a complex of HLA-A*2402 and WT1-derived epitope peptide; CMTWNQMNL, and also a complex of HLA-A*2402 and every kind of single amino residue-mutated peptides. MD simulation was applied to these complex models to analyze the binding structure and to estimate the binding free energy. Simulation showed a difference in binding free energy between the natural peptide and the mutated ones. The discussion was developed in terms of compatibility with the experimentally measured immunological response. Since the second residue at N-terminus of epitope (p2) is important for the peptide binding, the interactions of p2 residue with amino residues in HLA molecule were closely investigated. This investigation will be helpful for designing the epitope peptide enhancing the binding to HLA molecules.

Experimental

Build of Simulation Model A structure of HLA-A*2402 was already resolved by an X-ray crystallographic analysis and registered in protein data bank (PDB)[13] in a complex form with an antigen peptide; PDB code 2BCK, in a resolution of 2.80 Å.[14] This crystal structure contains two pairs of HLA-β2m complexed with epitope peptides, sulfate ions, glycerols, and waters, while there are 8 missing residues. The epitope peptide in 2BCK structure is a 9-mer fragment of telomerase reverse transcriptase and the sequence is different from WT1 protein. Then, computational model was built by homology modeling. One of the HLA-β2m-epitope complexes in 2BCK crystal structure was selected as a reference. Modeller9.2 was employed for homology modeling,[15—17] in which the original epitope peptide in the crystal structure was replaced by WT1 peptide. Further, p2 residue was converting from methionine of the natural peptide into other kinds of amino acids. The complex was solvated with about 5000 TIP3P waters,[18] generated in a rectangular box using leap module of AMBER9[19,20] in a manner shown in Fig. 1a.

Computational Procedure MD simulation was carried out using sander module of AMBER9 program package.[20] AMBER ff03 force field was employed. MD simulation was executed in three steps of minimization, heating and equilibration. Atom geometry was energetically minimized in the first step. The minimization was performed with the steepest descent method[21] for the earlier 3000 cycles and with the conjugated gradient method for the later 10000 cycles, with only water molecules allowed to move freely. The minimization was executed in a similar manner, with water molecules and epitope peptides allowed to move. Again the minimization was executed without any constraint for waters, peptide and HLA-β2m. The heating calculation was executed in the NVT ensemble condition. The temperature of the model system was elevated from 0 to 310 K for 200 ps. The equilibration calculation was performed in the NPT ensemble condition with a temperature of 310 K and a pressure of 1 atm. Langevin dynamics method was applied for controlling temperature and pressure with a collision frequency of 0.1 ps^{-1}.[22—24] The long-distance electrostatic and van der Waals interaction was calculated under the periodic boundary condition, setting the cutoff distance to be 12 Å. The expansion and shrinkage of all covalent bonds connecting to hydrogen atoms were constraint with SHAKE algorithm.[25] The integration time step was 2 fs. Six nanoseconds equilibration calculation was carried out for all models.

Estimation of Binding Affinity The binding affinity of HLA-A*2402-β2m and epitope peptide was estimated from the binding free energy ΔG_{bind},[26]

$$\Delta G_{bind} = \Delta H_{bind} - T\Delta S_{bind}$$

in which the enthalpic and entropic terms were computed by MM-GB/SA method[27] and normal mode analysis,[12] respectively.

Five hundred snapshot structures were extracted from the trajectory of the last 1 ns run of the equilibrating MD simulation. For every snapshot structure, binding enthalpy was calculated by MM-GB/SA method and the average value of the binding enthalpy was obtained. Ten snapshot structures were extracted from the last 1 ns trajectory of MD simulation to calculate entropic term with the normal mode analysis and the averaged value was obtained for the entropic term. In the calculation of enthalpic term, the respective model structures for HLA-β2m, epitope peptide, and HLA-β2m-peptide complex were built from each of 500 snapshot structures, and ΔH_{bind} was computed with the following equation.

$$\Delta H_{bind} = H_{complex} - (H_{HLA\text{-}\beta 2m} + H_{peptide})$$

where $\Delta H_{HLA\text{-}\beta 2m}$, $\Delta H_{peptide}$, and $\Delta H_{complex}$ represent the free energies of HLA-β2m, epitope peptide and their complex, respectively. In the normal mode analysis, the entropy for rotation and vibration was computed using the models for HLA-β2m, epitope peptide, and their complex and the change in entropy due to peptide binding was obtained from the following equation.

$$T\Delta S_{bind} = TS_{complex} - (TS_{HLA\text{-}\beta 2m} + TS_{peptide})$$

In this study, we introduced the deformation effect of epitope peptides due to binding to HLA. The ligand molecule in this work is a 9-mer peptide chain. The peptide chain has a large conformational diversity due to the rotation around the covalent bond. That is, 9-mer peptide shows a considerably large difference in structure between the HLA-bound and unbound states. This is a marked difference from low mass-weight compounds like medical drugs. In the usual MM-GB/SA method, the structure of ligand is assumed to be identical between the receptor-bound and unbound states. However the estimation of enthalpic and entropic component is not accurate if the diversity in configuration of the ligand molecule is large. The deformation effect should be taken into account to improve the energy estimation. In order to incorporate the deformation effect, additional MD simulations were performed for the model representing epitope peptide only, followed by evaluation of the binding free energy with MM-GB/SA method and normal mode analysis. The epitope peptide was placed in a rectangular box and solvated with about 5000 water molecules. The cutoff distance for electrostatic and van der Waals forces was set to 12 Å. Totally 6 ns MD simulation was performed for equilibration and the trajectory for the last 1 ns was used for estimating the binding energy. Snapshot structures were extracted every 10 ps from the last 1 ns MD simulation and 100 structures were collected. MM-GB/SA method and normal mode analysis were executed to compute the enthalpic and entropic contributions. The modified binding free energy was obtained from the following equations,

$$\Delta G_{mod} = \Delta H_{mod} - T\Delta S_{mod}$$
$$\Delta H_{mod} = H_{complex} - (H_{HLA} + H'_{peptide})$$
$$T\Delta S_{mod} = TS_{complex} - (TS_{HLA} + TS'_{peptide})$$

where $H'_{peptide}$ and $TS'_{peptide}$ were calculated from MD simulation for epitope peptide only.

In order to examine the interaction between the side chain of p2 residue of epitope peptide and all the residues of HLA molecule, alanine-scanning calculation[28,29] was performed, in which alanine was substituted for the amino acid at the residue of which the interaction energy with p2 residue was estimated. Comparison in energy between substituted and non-substituted mod-

els indicated the contribution of the residue to the total binding energy. The binding free energies between p2 residue and the residues of HLA were calculated by MM-GB/SA method.[27]

Results and Discussion

Binding Structure of Epitope Peptide to HLA Class I Molecule HLA class I molecule has six binding pockets labeled A—F as shown in Fig. 1b. The binding of epitope peptide to the pockets will stabilize the complex structure.[14,30] The amino acid residue of the peptide playing an important role in binding to HLA-β2m is called "anchor residue." The position of anchor residue and the location of significantly important pocket vary depending on the type of HLA. The second residue from N-terminus; p2, and the ninth residue; p9, are the anchor residues for HLA-A*2402. The favorable amino acid for anchor residue is also changed, depending on the type of HLA. For example Tyr or Phe at p2 residue makes the peptide binding strong in HLA-A*2402,[14] while Leu or Met generates a firm binding to HLA-A*2402.[31]. According to the experiment by Tsuboi *et al.*,[7] the substitution of Tyr for Met at p2 anchor residue makes the binding of WT1 peptide stronger. The residues responsible for the peptide binding to HLA are supposed to have little influence on the association with CTL. Therefore, the replacement of anchor residue with other kind of amino acids will be one of the effective approaches to enhance the binding of epitope peptide to HLA. In this study, the binding affinity was examined with changing the amino acid at p2 anchor residue of the natural WT1 peptide.

Snapshot structures for analyzing binding mode and binding affinity were extracted from the last 1 ns trajectory of MD simulation. The model system should be sufficiently equilibrated for the sake of reliable analysis. One of the standard techniques to confirm the equilibration is to measure the root mean square deviation (RMSD) value between the snapshot structure at the time point concerned and the structure at the starting point. If the time evolution of RMSD shows constant, the system fluctuates in a single conformation and is confirmed to be equilibrated. RMSD value during MD simulation was monitored on the HLA-β2m peptide complex, with respect to the Cα, N, and C main chain atoms, setting the snapshot structure after heating calculation as a reference at the starting point. The RMSD plot indicates that RMSD value fluctuates in a range of 2.8±0.5 Å for the last 1 ns (data not shown). This result means that the model system was in equilibration.

Principal component analysis[32,33] is another method to confirm the equilibration of the calculation system. Since protein usually fluctuates and has several probable confirmations, there exist several minimum points in potential energy surface. If the energy barrier separating potential minima is low enough, the structure fluctuates among some energetically minimized points during MD simulation. When the model system is in equilibration, the structure fluctuates only around a stable minimum point. Principle component analysis was executed for the calculations on the natural peptide and 19 mutants (data not shown). In every case, there appear several clusters, and the conformation converges to a single cluster with the progress of MD simulation. This result suggests that 6 ns MD simulation leads structural equilibration for the HLA-β2m-peptide complex.

Binding Free Energy Table 1 shows the comparison of the calculated binding free energy, and the binding constant obtained from experiment for HLA-A*2402 bound with the natural peptide and M2Y variant. The binding constant K_d in Table 1 was measured in the *in vitro* experiment using flow cytometry by Oka *et al.*[7] According to their *in vitro* experiment, K_d value for the natural peptide was measured to be 1.82×10^{-5} M and that for M2Y variant was 6.40×10^{-7} M. Hence, M2Y showed a higher binding affinity to HLA-

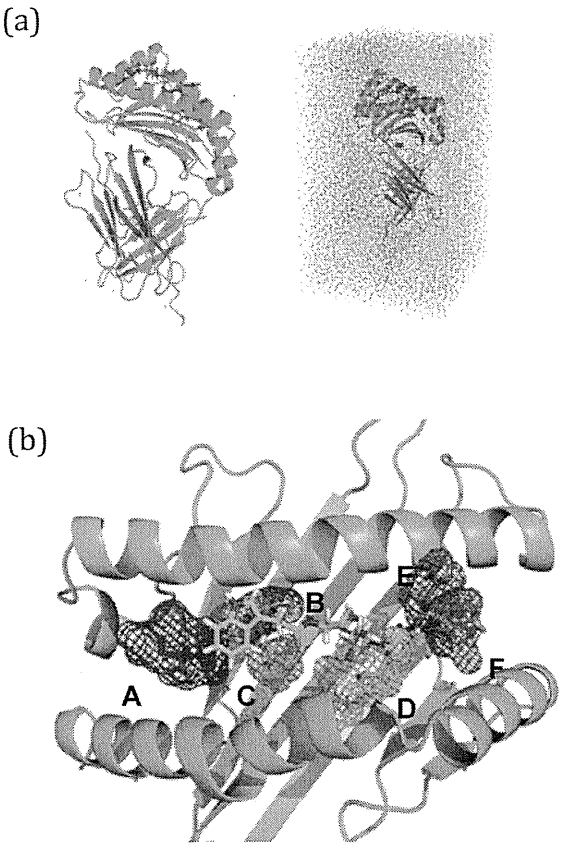

Fig. 1. (a) Structure of HLA Class I Molecule Combined with 9-mer Peptide

9-mer peptide is shown in the stick representation. 9-mer peptide and HLA complex was solvated with about 5000 of TIP3P waters, generated in a rectangular box for molecular dynamics simulation.

(b) Structure of Binding Pockets of HLA Class I Molecule

This molecule has six binding pockets labeled A—F. The binding of an epitope peptide into the pockets will stabilize the complex.

Table 1. Comparison of the Calculated Binding Free Energy and the Experimentally Measured Binding Constant

Peptide	A sequence	K_d (M)	ΔG_{exp} (kcal/mol)	ΔG_{bind} (kcal/mol)	ΔG_{mod} (kcal/mol)
NP	C**M**TWNQMNL	1.82×10^{-5}		−59.6	−44.8
M2Y	C**Y**TWNQMNL	6.40×10^{-7}	−2.0	−64.4	−45.7

NP, natural peptide. Underlined capital letter indicates anchor position; bold capital letter indicates a substituted amino acid.

Table 2. Calculated Binding Free Energy for Every Kind of Peptides Combined with HLA

Peptide	ΔH_{ele} (kcal/mol)	ΔH_{vwd} (kcal/mol)	ΔH_{int} (kcal/mol)	ΔH_{sol} (kcal/mol)	ΔH_{mod} (kcal/mol)	$T\Delta S_{mod}$ (kcal/mol)	ΔG_{mod} (kcal/mol)
NP	−139.8	−89.0	6.4	144.2	−78.2	−33.4	−44.8
M2Y	−159.5	−93.4	10.0	157.4	−85.5	−39.8	−45.7
M2W	−190.8	−93.7	10.6	188.2	−85.7	−37.7	−48.2
M2V	−153.0	−71.1	5.5	152.4	−66.2	−37.7	−28.5
M2I	−155.0	−75.0	5.1	152.4	−72.5	−32.2	−40.3
M2L	−148.6	−77.3	0.4	150.5	−75.0	−46.1	−28.9
M2A	−142.1	−71.2	2.4	147.0	−63.9	−31.6	−32.3
M2F	−101.3	−71.1	1.8	106.4	64.2	−39.2	−25.0
M2G	−152.9	−80.3	7.0	153.9	−72.3	−38.0	−34.3
M2S	−173.7	−74.9	2.2	167.4	−79.0	−37.5	−41.5
M2T	−177.0	−73.5	4.6	176.4	−69.5	−30.6	−38.9
M2P	−182.2	−84.7	8.1	180.1	−78.7	−37.4	−41.3
M2C	−185.0	−89.4	6.7	183.5	−84.2	−40.8	−43.5
M2N	−163.4	−84.6	6.1	165.2	−76.7	−43.4	−33.3
M2Q	−187.2	−83.7	6.0	182.7	−82.2	−38.7	−43.5
M2H	−157.8	−83.5	5.4	161.3	−74.6	−33.9	−40.7
M2K	−234.3	−65.9	3.5	239.1	−57.6	−39.0	−18.6
M2R	−257.7	−83.6	5.6	261.3	−74.3	−45.0	−29.3
M2D	−102.9	−79.9	−0.8	110.0	−73.6	−44.9	−28.7
M2E	−63.3	−77.6	5.0	71.5	−64.1	−32.4	−31.7

NP, natural peptide; ΔH_{ele}, electrostatic energy; ΔH_{vdw}, van der Walls interaction; ΔH_{int}, internal energy from bond, angle and dihedral; ΔH_{sol}, solvation energy; ΔH_{mod}, modified averaging enthalpic energy; $T\Delta S_{mod}$, modified averaging entopic energy; ΔG_{mod}, modified binding free energy.

A*2402, compared to the natural peptide. The calculation of binding free energy without deformation effect, ΔG_{bind}, indicated that M2Y mutant is more stable than the natural peptide. Accordingly, the usual technique for calculating binding free energy could predict the binding affinity measured in experiments. ΔG_{mod} of Table 1 shows a comparison in energy between the natural peptide and M2Y variant obtained with the deformation effect of epitope peptide took into account. The energy comparison shows that the binding of M2Y peptide is more stable than the natural peptide, which is also compatible with the experimental findings. The energy difference estimated from the binding constant is −2.0 kcal/mol between the natural peptide and M2Y. The calculated energy difference is −0.9 kcal/mol. Accordingly the combination of MM-GB/SA method and normal mode analysis with considering the deformation effect will correctly predict the binding affinity of epitope peptide bound to HLA-β2m. The result suggests that the deformation effect has a certain degree of importance in the prediction of the binding free energy.

It is informative to clarify what chemical structure increases the binding affinity between HLA-β2m and epitope peptide. Hence, the p2 amino residue of the peptide was replaced with another kind of amino acids. The complex models for 19 kinds of variants bound to HLA-β2m were built and MD simulations were performed for all models, followed by free energy calculation with MM-GB/SA method and normal mode analysis. Our concern is how we can improve the binding affinity replacing an amino acid with another one to increase the electrostatic or van der Waals interaction. Tyr and Trp have aromatic ring and contain polar atoms possible for hydrogen bonding. Non-polar amino acid residues, Val, Leu, Ile, Phe, Pro, were substituted at p2 anchor position, because they have large hydrophobic side chain to increase van der Waals interaction. Ser and Thr have hydroxy group in their side chains to make it possible to form hydrogen bonds. Cys, Asn and Gln can also make hydrogen bonds because of the presence of polar hydrogen atoms in their side chains. Basic polar amino acid residues, His, Lys, Arg, and acidic polar amino acid ones, Asp, Glu, can generate ionic interaction while hydrophobic interaction is little. Non-polar amino acid residues, Gly, Ala, have small side chains, but may contribute to keep peptide flexibility.

Table 2 shows the calculated binding free energy for the natural and other 19 kinds of peptide complexed with HLA-A*2402 molecules. The energy difference in ΔG_{mod} indicates that all of substitutions except for M2Y and M2W are less stable than the natural peptide. ΔH_{mod} in the table also indicates that HLA-β2m complexed with M2Y or M2W mutant show the lowest energy among all the complexes. Namely these epitope peptides will give the highest binding affinity. It is already known that a modified 9-mer WT1 (M2Y) peptide showed much higher HLA-A*2402-binding affinity than the natural 9-mer WT1 peptide.[7] In this study, we found that the binding free energy of M2W to HLA-β2m is slightly lower than M2Y (−2.5 kcal/mol). This means that M2W peptide substitution is the most favorable in terms of binding affinity. The van der Waals energy (ΔH_{vdw}) shows a favorable contribution to the peptide binding. ΔH_{vdw} value of the HLA-β2m complexed with M2Y or M2W is the lowest among the complexes. The van der Waals energy of M2W is slightly lower then M2Y (−0.03 kcal/mol). Therefore M2W peptide substitution is more favorable because of the presence of large hydrophobic side chain. Among the other peptides M2K shows the highest ΔH_{vdw} value and highest binding free energy ΔG_{mod}, which suggests that the presence of Lys causes a strong instability for the binding of HLA-β2m and peptide. The large change in ΔG_{mod} in Table 2 clearly demonstrates that the side chain of the anchor residue, p2, plays a quite important role for peptide binding. M2C, M2Q, M2S and M2P show the binding energies near to the natural peptide, which suggests that side chain hydrogen-bond doner or acceptor can serve as an anchor for peptide binding. Judging from the binding energy, M2H, M2I, M2T, M2G, M2N, M2A and M2E are not so stable. Further, M2V, M2F, M2L,

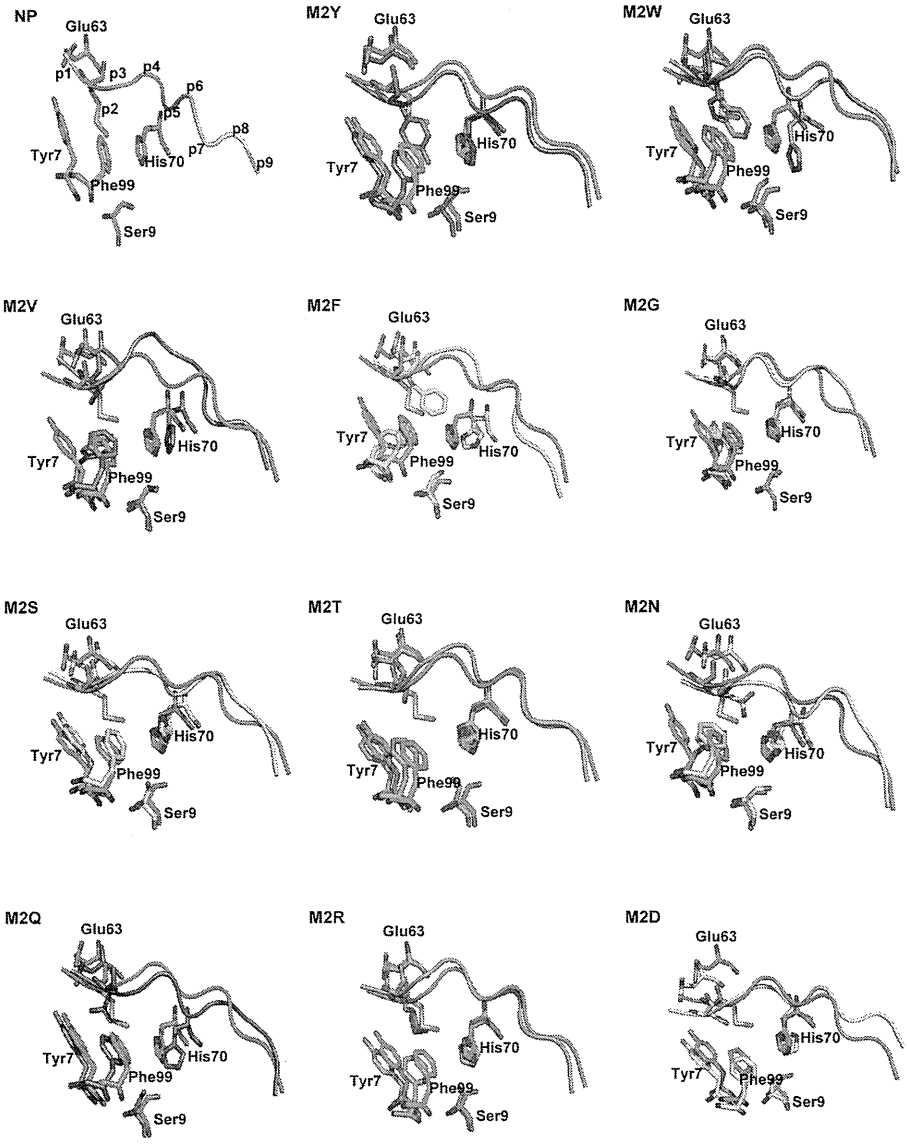

Fig. 2. Superimposition of the Average Structure of 19 Kinds of Peptide Variants on That of the Natural Peptide Bound to HLA-β2m

The structure of the natural peptide complexed with HLA-β2m is colored dark in superimposition. The position of each residue on the epitope peptide is shown in the figure at the right bottom for natural peptide (NP).

M2D and M2R hardly maintain the stability of peptide binding.

Binding Structure Figure 2 shows the superimposition of the averaged structures of 19 kinds of peptide to the natural peptide bound to HLA-β2m. The averaged structure for the last 1 ns MD simulation was obtained for the respective models, and pymol software[35] was employed to superimpose the structure of the HLA-β2m-natural peptide complex on that of the p2-mutated one, with respect to main chain atoms of HLA. The N-terminal sides of the peptide main chain for M2F and M2D models are largely deviated from that of the natural peptide. These peptides showed the instability in binding to HLA-β2m molecule in Table 2. The C-terminal side of the epitope peptide is largely deviated in M2D, M2N and M2G models from the natural peptide. This indicated that the C-terminal region of M2D, M2N and M2G peptides is loosely bound to HLA. Several X-ray crystal structures obtained for other types of HLA indicated that no significant deviation was observed for the three residues at the both terminal regions.[36] This suggests that central residues of the epitope peptide are mainly responsible for the activation of CTL and the terminal residues primarily contribute to the binding of epitope peptide to HLA. Both terminal regions of M2Y, M2W, M2Q and M2S are compatible with that of the natural peptide. All these peptide show the low binding energy in Table 2 and the complexes with HLA are stable.

Contribution of Amino Residues of HLA to Peptide Binding The interaction energy between p2 anchor residue and the respective residues of HLA were calculated by alanine scanning technique.[28,29] In the alanine scanning calculation, one of the residues of HLA is replaced by Ala and then the binding energy is calculated with the residue-replaced model. A comparison in binding free energy between the models with and without replacement indicates the contribution of the residue to the peptide binding. The energy contribution of p2 residue is computed by MM-GB/SA method. Figure 3 shows the results of alanine scanning calculation, in which low energy value means positive contribution of the

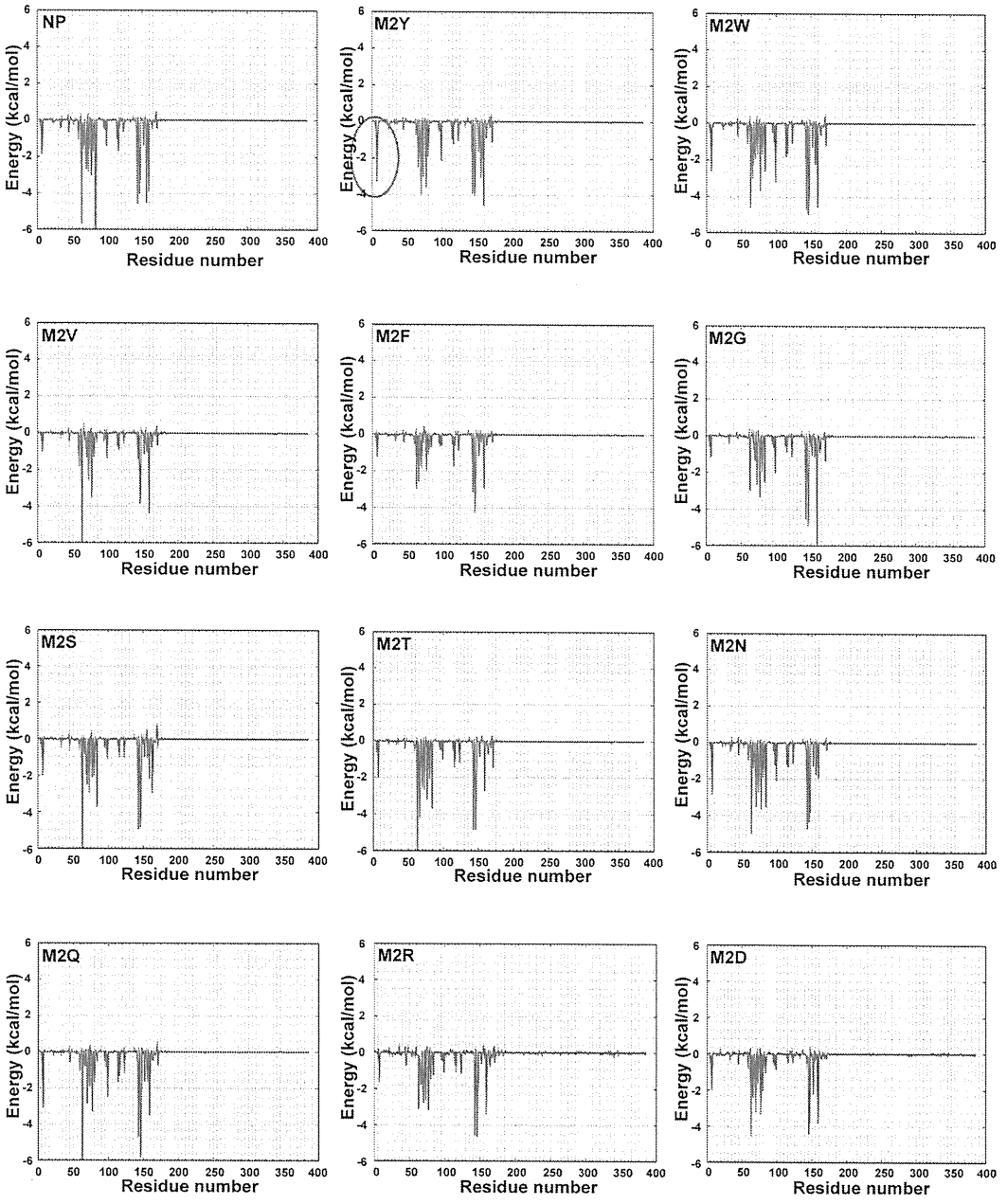

Fig. 3. Interaction Energy of p2 Residue of Epitope Peptide with the Respective Residues of HLA Class I Molecule

A comparison of the energy diagrams suggests that several residues concerning the interaction are common among the mutants but the interaction with Ser9 residue of HLA is only enhanced in M2Y.

residue to the interaction with p2 anchor residue. In all HLA-β2m-peptide models, Tyr7, Glu63, His70, Phe99, Ser9 are related to the interaction. The locations of these residues are displayed in Fig. 3 to clarify the relationship with p2 anchor residue. These residues were reported to be located at the deep bottom of B-pocket.[30] A comparison in Fig. 3 suggests that the interaction with Ser9 of HLA residue is markedly enhanced in M2Y (gray circle in Fig. 3). M2Y generates a hydrogen bond with HLA, which makes the peptide binding stable. The interaction with Ser9 of HLA residue is also enhanced in M2W. The time ratio for keeping hydrogen bond within 3.5 Å was calculated from the last 1 ns MD simulation. Ser9 keeps a hydrogen bond to M2Y for 74% of MD simulation time, while other models scarcely keep the hydrogen bond with Ser9. The interaction with Phe99 is the highest in M2W. This large interaction will be attributed to the π–π interaction between the indole group of Trp at p2 anchor residue and the aromatic ring of Phe. As for M2Y, M2Q and M2N, the interaction with Phe99 is also higher compared with the natural peptide. In M2Q, M2S, M2T and M2V, the interaction with Glu63 was increased while other peptides scarcely showed the favorable energy gain in the interaction with Glu63. In M2Q, M2S, M2T and M2V, several residues of protein also show small degree of interaction with p2 residue.

Table 3 shows the respective components of the binding free energy between p2 residue and HLA-β2m, where $\Delta H_{\text{ele-p2}}$, $\Delta H_{\text{vdw-p2}}$, and $\Delta H_{\text{sol-p2}}$ represent the electrostatic, van der Waals, and solvation energies, respectively. There are many hydrophobic residues at the inside of B binding pocket

Table 3. Components of the Binding Free Energy between p2 Residue and HLA

Peptide	$\Delta H_{ele\text{-}p2}$ (kcal/mol)	$\Delta H_{vwd\text{-}p2}$ (kcal/mol)	$\Delta H_{int\text{-}p2}$ (kcal/mol)	$\Delta H_{sol\text{-}p2}$ (kcal/mol)	ΔH_{p2}* (kcal/mol)
NP	−20.7	−10.1	8.1	5.3	−17.4
M2Y	−17.6	−13.4	7.9	4.6	−18.5
M2W	−17.8	−15.4	8.6	7.0	−17.6
M2V	−27.4	−6.8	8.2	3.3	−22.7
M2I	−23.8	−8.7	8.1	4.9	−19.9
M2L	−19.5	−7.9	7.8	2.8	−16.8
M2A	−20.3	−4.9	8.1	3.5	−13.6
M2F	−15.9	−9.5	8.5	1.0	−15.9
M2G	−12.6	−3.6	9.2	2.0	−5.0
M2S	−39.5	−3.0	7.7	14.7	−20.1
M2T	−19.6	−6.9	7.7	6.7	−12.1
M2P	−9.9	−7.7	9.3	1.9	−6.4
M2C	−20.7	−6.0	7.8	4.5	−14.4
M2N	−23.8	−8.2	8.6	9.6	−13.8
M2Q	−30.7	−8.8	8.2	14.1	−17.2
M2H	−21.3	−10.3	8.0	7.5	−16.1
M2K	−115.4	−6.9	8.1	104.0	−10.2
M2R	−103.7	−10.5	8.5	93.8	−11.9
M2D	25.5	−6.8	7.6	−36.6	−10.3
M2E	31.7	−7.3	8.7	−44.0	−10.9

NP, natural peptide; $\Delta H_{ele\text{-}p2}$, electrostatic energy; $\Delta H_{vdw\text{-}p2}$, van der Walls interaction; $\Delta H_{int\text{-}p2}$, internal energy from bond, angle and dihedral; $\Delta H_{sol\text{-}p2}$, solvation energy; ΔH_{p2}*, total binding free energy between p2 residue and HLA-β2m.

of HLA. This suggests that hydrophobic residues of epitope peptide play an important role in peptide binding. For example, $\Delta H_{vdw\text{-}p2}$ of M2I, M2L and M2F are slightly larger than those of other mutations into a similar size of residue and further $\Delta H_{sol\text{-}p2}$ values of these residues are markedly low. Then, M2I, M2L and M2F indicate the favorable total binding energy in ΔH_{p2} compared with the polar residues of M2K, M2R, M2D, M2E as seen in Table 3. $\Delta H_{vdw\text{-}p2}$ of M2S and M2G are −3.0 and −3.6 kcal/mol, and they are the highest among 20 peptides. This result is reasonably explained from the smallness of the side chain of Gly and the hydrophilicity of Ser. In contrast, $\Delta H_{vdw\text{-}p2}$ of M2W is −15.4 kcal/mol, which is the lowest among the peptides. $\Delta H_{vdw\text{-}p2}$ of M2Y is also lower than other kinds of peptide. The indole group of M2W and the aromatic rings of M2Y peptide have a large contribution to hydrophobic interaction in B-binding pocket. Accordingly, the energetic gain in van der Waals term seems some degree of correlation with the total binding affinity. The p2 anchor residue should have a large hydrophobicity to enhance the binding affinity to HLA-β2m. The indole group of Trp is important to enhance the binding with the B-pocket of HLA. The hydroxyl group of Tyr is also important for binding with the formation of hydrogen bond. The importance of hydrogen bond by Tyr was also suggested in the recent report that human immunodeficiency virus (HIV) changes its residues from Tyr to Phe to escape from the immunotherapy using a peptide vaccine.[37]

Energy Calculation with Another Type of HLA As an example for predicting binding affinity between epitope and HLA, similar calculation were carried out for HLA-A*0201 and 4 kinds of peptide.[36] HLA-A*0201, as well as HLA-A*2402, is one of the most popular HLA types in Japan. Table 4 shows the calculated binding free energy, in which computational prediction was compatible with the experimentally measured affinity of the peptide binding to HLA-A*0201. These results suggest that our present approach is one of the effective techniques to select the favorable peptides as HLA epitope. We have also examined the role of p2 residue of peptide in binding to HLA-β2m in this example. In the calculation, a large van der Waals interaction between the side chain of p2 residue and hydrophobic residues inside B-pocket of HLA was observed.

Table 4. Calculated Binding Free Energy for Several Peptides Binding to HLA-A*0201

PDB ID	Peptide sequence	K_d (nM)	ΔG_{calc} (kcal/mol)
1HHG	TLTSVNTSV	294.0	6.0
1HHJ	ILKEPVHGV	242.0	9.7
1HHK	LLFGYPVYV	11.0	−1.8
1HHI	GILGFVFTL	6.0	−30.1

The application of MD simulation for selecting a peptide strongly bound to HLA was challenged in many research groups.[38—41] Although there seems to be a room for improvement in practical use,[38] the recent high-performance computers offer us a hope of accurate prediction of antigenic peptides.[40] Some theoretical studies suggested the importance of the flexibility for peptide recognition.[42,43] MD simulation is useful to obtain an insight into the dynamic behavior of molecules. For example, an increase in the peptide flexibility was observed in the binding groove.[43] The conformational change of HLA molecule due to the peptide binding has been analyzed.[44—46] Such a conformational change was shown to be closely related to the recognition of the peptide-HLA complex by CTLs,[45] and the conformational difference between the peptide-bound and -unbound HLAs was clearly indicated by the removal of the bound peptide from a complex model.[46] Nojima *et al.* investigated the structural changes of the peptide-binding groove due to the removal of the peptide using the normal mode analysis.[47,48] They suggested that two membrane-proximal domains had significant influence on the activity and stability of HLA.[49,50]

The computational study is also advantageous to analyze the effect of amino acid substitution of epitope peptides.[51—54] The reason for the change in peptide recognition among different types of HLA molecules has been clearly explained.[55—57] In the computational studies, the accurate prediction of the complex structure of HLA and peptide is essentially important and some trials were attempted for the construction of appropriate complex structures.[58,59] Further, reliable estimation of the binding energy is indispensable for selecting a promising peptide, and MM-GB/SA and MM-Poisson–Boltzmann (PB)/SA methods are the most standard approaches for estimating the binding energy.[60,61]

All the above previous studies suggest the usefulness of molecular simulation for predicting the affinity and specificity of antigenic peptide binding to HLA. Because of the importance of peptide flexibility, the evaluation of entropic terms is indispensable for estimating peptide binding affinity. The prediction of the difference in binding affinity due to the amino acid substitution is possible through the theoretical calculations with high-performance computers of today. Accordingly, our present work will be one of the examples demonstrating an effectiveness of computational analysis, performed by taking full advantage of MD simulation.

Conclusion

We investigated the binding affinity between HLA-A*2402 and epitope peptide derived from Wilms' tumor (WT1) protein, performing molecular dynamics simulations, followed by MM-GB/SA method and normal mode analysis. The binding affinity was also contrasted between the 9-mer natural WT1 peptide and the modified peptide in which 19 kinds of amino acids were substituted for Met located at the second residue of the N-terminal side. WT1 protein is highly expressed in solid tumor or blood in leukemia patient, and a vaccine targeting WT1 is expected to be one of the promising approaches in immunotherapy.[5] In this work, we demonstrated that computational prediction of the peptide binding affinity to HLA is possible. The contribution of p2 residue of the epitope peptide to the binding free energy was due to the formation of hydrogen bond at the bottom of B-pocket of HLA and van der Waals interaction inside B-pocket. For the stable binding of epitope peptide to HLA-A*2402, p2 residue should have the following property. First, p2 residue should bear alkyl chain or aromatic ring to make van der Waals interaction strong enough inside B-pocket. Second, the residue containing polar atom is advantageous to make a hydrogen bond at the bottom of B-pocket. Third, the residue should have a large volume comparative to Tyr or Phe. These findings will be helpful to design the peptide vaccine inducing high immunological response or the peptide-mimic chemical for medical drugs.

Acknowledgments Calculations were performed at Research Center for Computational Science, Okazaki, Japan, and Information Technology Center of the University of Tokyo, and also by the high-performance computer system at Institute for Media Information Technology of Chiba University. A part of this work was supported by Grant-in-Aid for Scientific Research (C) from Japan Society for the Promotion of Science (JSPS).

References

1) Rock K. L., Goldberg A. L., *Annu. Rev. Immunol.*, **17**, 739—779 (1999).
2) Townsend A., Bodmer H., *Annu. Rev. Immunol.*, **7**, 601—624 (1989).
3) Kageyama S., Nagata Y., Miyahara Y., Hiasa A., Naota H., Okumura S., Imai H., Shiraishi T., Masuya M., Nishikawa M., Sunamoto J., Akiyoshi K., Kanematsu T., Scott A. M., Murphy R., Hoffman E. W., Old L. J., Shiku H., *Clin. Cancer Res.*, **12**, 7397—7405 (2006).
4) Oka Y., Udaka K., Tsuboi A., Elisseeva O. A., Ogawa H., Aozasa K., Kishimoto T., Sugiyama H., *J. Immunol.*, **164**, 1873—1880 (2000).
5) Oka Y., Tsuboi A., Kawakami M., Elisseeva O. A., Nakajima H., Udaka K., Kawase I., Oji Y., Sugiyama H., *Curr. Med. Chem.*, **13**, 2345—2352 (2006).
6) Oka Y., Tsuboi A., Taguchi T., Osaki T., Kyo T., Nakajima H., Elisseeva O. A., Oji Y., Kawakami M., Ikegame K., Hosen N., Yoshihara S., Wu F., Fujiki F., Murakami M., Masuda T., Nishida S., Shirakata T., Nakatsuka S., Sasaki A., Udaka K., Dohy H., Aozasa K., Noguchi S., Kawase I., Sugiyama H., *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 13885—13890 (2004).
7) Tsuboi A., Oka Y., Udaka K., Murakami M., Masuda T., Nakano A., Nakajima H., Yasukawa M., Hiraki A., Oji Y., Kawakami M., Hosen N., Fujioka T., Wu F., Taniguchi Y., Nishida S., Asada M., Ogawa H., Kawase I., Sugiyama H., *Cancer Immunol. Immunother.*, **51**, 614—620 (2002).
8) Zhang H., Wang P., Papangelopoulos N., Xu Y., Sette A., Bourne P. E., Lund O., Ponomarenko J., Nielsen M., Peters B., *PLoS ONE*, **5**, e9272 (2010).
9) Morikis D., Lambris J. D., *Trends Immunol.*, **25**, 700—707 (2004).
10) Guan P., Doytchinova I., Zygouri C., Flower D., *Nucleic Acids Res.*, **31**, 3621—3624 (2003).
11) Tomer Hertz C. Y., *BMC Bioinformatics*, **7**, S3 (2006).
12) Brooks B., Karplus M., *Proc. Natl. Acad. Sci. U.S.A.*, **80**, 6571—6575 (1983).
13) Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H., Shindyalov I. N., Bourne P. E., *Nucleic Acids Res.*, **28**, 235—242 (2000).
14) Cole D. K., Rizkallah P. J., Gao F., Watson N. I., Boulter J. M., Bell J. I., Sami M., Gao G. F., Jakobsen B. K., *Eur. J. Immunol.*, **36**, 170—179 (2006).
15) Sali A., Blundell T. L., *J. Mol. Biol.*, **234**, 779—815 (1993).
16) Martí-Renom M. A., Stuart A. C., Fiser A., Sánchez R., Melo F., Sali A., *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 291—325 (2000).
17) Fiser A., Do R. K., Sali A., *Protein Sci.*, **9**, 1753—1773 (2000).
18) Jorgensen W. L., Chandrasekhar J., Madura J. D., Impey R. W., Klein M. L., *J. Chem. Phys.*, **79**, 926—935 (1983).
19) Case D. A., Cheatham T. E. 3rd, Darden T., Gohlke H., Luo R., Merz K. M. Jr., Onufriev A., Simmerling C., Wang B., Woods R. J., *J. Comput. Chem.*, **26**, 1668—1688 (2005).
20) Pearlman D. A., Case D. A., Caldwell J. W., Ross W. S., Cheatham T. E., Debolt S., Ferguson D., Seible G., Kollman P., *Comput. Phys. Commun.*, **91**, 1—41 (1995).
21) Brooks B. R., Brooks C. L. III, Mackerell A. D., Nilsson L., Petrella R. J., Roux B., Won Y., Archontis G., Bartels C., Boresch S., Caflisch A., Caves L., Cui Q., Dinner A. R., Feig M., Fischer S., Gao J., Hodoscek M., Im W., Kuczera K., Lazaridis T., Ma J., Ovchinnikov V., Paci E., Pastor R. W., Post C. B., Pu J. Z., Schaefer M., Tidor B., Venable R. M., Woodcock H. L., Wu X., Yang W., York D. M., Karplus M., *J. Comput. Chem.*, **4**, 187—217 (1983).
22) Pastor R. W., Brooks B. R., Szabo A., *Mol. Phys.*, **65**, 1409—1419 (1988).
23) Loncharich R. J., Brooks B. R., Pastor R. W., *Biopolymers*, **32**, 523—535 (1992).
24) Izaguirre J. A., Catarello D. P., Wozniak J. M., Skeel R. D., *J. Chem. Phys.*, **114**, 2090—2098 (2001).
25) Ryckaert J. P., Ciccotti G., Berendsen H. J. C., *J. Comput. Phys.*, **23**, 327—341 (1977).
26) Kollman P., *Chem. Rev.*, **93**, 2395—2417 (1993).
27) Golhke H., Case D. A., *J. Comput. Chem.*, **25**, 238—250 (2004).
28) Manning T. C., Schlueter C. J., Brodnicki T. C., Parke E. A., Speir J. A., Garcia K. C., Teyton L., Wilson I. A., Kranz D. M., *Immunity*, **8**, 413—425 (1998).
29) Zoete V., Michielin O., *Proteins*, **67**, 1026—1047 (2007).
30) Chelvanayagam G., *Immunogenetics*, **45**, 15—26 (1996).
31) Rammensee H., Bachmann J., Emmerich N. P., Bachor O. A., Stevanovic S., *Immunogenetics*, **50**, 213—219 (1999).
32) Teeter M. M., Case D. A., *J. Phys. Chem.*, **94**, 8091—8097 (1990).
33) Kitao A., Go N., *Curr. Opin. Struct. Biol.*, **9**, 164—169 (1999).
34) Zeh H. J. III, Leder G. H., Lotze M. T., Salter R. D., Tector M., Stuber G., Modrow S., Storkus W. J., *Hum. Immunol.*, **39**, 79—86 (1994).
35) DeLano W. L., The DeLano Scientific, San Carlos, CA, U.S.A., 2002.
36) Madden D., Garboczi D., Wiley D., *Cell*, **75**, 693—708 (1993).
37) Fujiwara M., Tanuma J., Koizumi M., Kawashima Y., Honda K., Matsuoka-Aizawa S., Dohki S., Oka S., Takiguchi M., *J. Virol.*, **82**, 138—147 (2008).
38) Tsurui H., Takahashi T., *J. Pharmacol. Sci.*, **105**, 299—316 (2007).
39) Antes I., Siu S. W., Lengauer T., *Bioinformatics*, **22**, e16—e24 (2006).
40) Flower D. R., Phadwal K., Macdonald I. K., Coveney P. V., Davies M. N., Wan S., *Immunome Res.*, **6** (Suppl. 2), S4 (2010).
41) Lim J. S., Kim S., Lee H. G., Lee K. Y., Kwon T. J., Kim K., *Mol. Immunol.*, **33**, 221—230 (1996).
42) Pöhlmann T., Böckmann R. A., Grubmüller H., Uchanska-Ziegler B., Ziegler A., Alexiev U., *J. Biol. Chem.*, **279**, 28197—28201 (2004).
43) Starikov E. B., Nilsson L., Hülsmeyer M., *Eur. Biophys. J.*, **33**, 651—655 (2004).
44) Stavrakoudis A., Tsoulos I. G., Uray K., Hudecz F., Apostolopoulos V., *J. Mol. Model.*, **17**, 1817—1829 (2011).
45) Knapp B., Omasits U., Schreiner W., Epstein M. M., *PLoS ONE*, **5**, e11653 (2010).
46) Painter C. A., Cruz A., López G. E., Stern L. J., Zavala-Ruiz Z., *PLoS ONE*, **3**, e2403 (2008).
47) Nojima H., Takeda-Shitaka M., Kurihara Y., Adachi M., Yoneda S., Kamiya K., Umeyama H., *Chem. Pharm. Bull.*, **50**, 1209—1214 (2002).
48) Nojima H., Takeda-Shitaka M., Kurihara Y., Kamiya K., Umeyama H., *Chem. Pharm. Bull.*, **51**, 923—928 (2003).
49) Nojima H., Takeda-Shitaka M., Kanou K., Kamiya K., Umeyama H., *Chem. Pharm. Bull.*, **56**, 635—641 (2008).

50) Nojima H., Kanou K., Kamiya K., Atsuda K., Umeyama H., Takeda-Shitaka M., *Chem. Pharm. Bull.*, **57**, 1193—1199 (2009).
51) Tang Y., Lin Z., Ni B., Wei J., Han J., Wang H., Wu Y., *Cancer Immunol. Immunother.*, **56**, 319—329 (2007).
52) Joseph M. A., Mitchell M. L., Evanseck J. D., Kovacs J. R., Jia L., Shen H., Meng W. S., *Mol. Immunol.*, **44**, 322—331 (2007).
53) Toh H., Savoie C. J., Kamikawaji N., Muta S., Sasazuki T., Kuhara S., *Biopolymers*, **54**, 318—327 (2000).
54) Toh H., Kamikawaji N., Tana T., Sasazuki T., Kuhara S., *Protein Eng.*, **11**, 1027—1032 (1998).
55) Fabian H., Huser H., Narzi D., Misselwitz R., Loll B. Zieglerv, Böckmann RA., Uchanska-Ziegler B., Naumann D., *J. Mol. Biol.*, **376**, 798—810 (2008).
56) Sieker F., Straatsma T. P., Springer S., Zacharias M., *Mol. Immunol.*, **45**, 3714—3722 (2008).
57) Sieker F., Springer S., Zacharias M., *Protein Sci.*, **16**, 299—308 (2007).
58) Todman S. J., Halling-Brown M. D., Davies M. N., Flower D. R., Kayikci M., Moss D. S., *J. Mol. Graph. Model.*, **26**, 957—961 (2008).
59) Fagerberg T., Cerottini J. C., Michielin O., *J. Mol. Biol.*, **356**, 521—546 (2006).
60) Cárdenas C., Bidon-Chanal A., Conejeros P., Arenas G., Marshall S., Luque F. J., *J. Comput. Aided Mol. Des.*, **24**, 1035—1051 (2010).
61) Wan S., Coveney P. V., Flower D. R., *J. Immunol.*, **175**, 1715—1723 (2005).

Bioorganic & Medicinal Chemistry 20 (2012) 775–783

Contents lists available at SciVerse ScienceDirect

Bioorganic & Medicinal Chemistry

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

Prediction of sites of metabolism in a substrate molecule, instanced by carbamazepine oxidation by CYP3A4

Hitomi Yuki [a,b], Teruki Honma [b], Masayuki Hata [c,*], Tyuji Hoshino [a,*]

[a] Graduate School of Pharmaceutical Sciences, Chiba University, 1-8-1 Inohana, Chuo-ku, Chiba 260-8675, Japan
[b] RIKEN Systems and Structural Biology Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan
[c] College of Pharmaceutical Sciences, Matsuyama University, 4-2 Bunkyo-cho, Matsuyama 790-8578, Japan

ARTICLE INFO

Article history:
Received 12 October 2011
Revised 30 November 2011
Accepted 1 December 2011
Available online 8 December 2011

Keywords:
Cytochrome P450 (CYP)
Carbamazepine (CBZ)
Sites of metabolism (SOM)
Molecular dynamics simulation
Multiple initial structures

ABSTRACT

In drug discovery process, improvement of ADME/Tox properties of lead compounds including metabolic stability is critically important. Cytochrome P450 (CYP) is one of the major metabolizing enzymes and the prediction of sites of metabolism (SOM) on the given lead compounds is key information to modify the compounds to be more stable against metabolism. There are two factors essentially important in SOM prediction. First is accessibility of each substrate atom to the oxygenated Fe atom of heme in a CYP protein, and the other is the oxidative reactivity of each substrate atom. To predict accessibility of substrate atoms to the heme iron, conventional protein-rigid docking simulations have been applied. However, the docking simulations without consideration of protein flexibility often lead to incorrect answers in the case of very flexible proteins such as CYP3A4. In this study, we demonstrated an approach utilizing molecular dynamics (MD) simulation for SOM prediction in which multiple MD runs were executed using different initial structures. We applied this strategy to CYP3A4 and carbamazepine (CBZ) complex. Through 10 ns MD simulations started from five different CYP3A4-CBZ complex models, our approach correctly predicted SOM observed in experiments. The experimentally known epoxidized sites of CBZ by CYP3A4 were successfully predicted as the most accessible sites to the heme iron that was judged from a numerical analysis of calculated $\Delta G_{binding}$ and the frequency of appearance. In contrast, the predictions using protein-rigid docking methods hardly provided the correct SOM due to protein flexibility or inaccuracy of the scoring functions. Our strategy using MD simulation with multiple initial structures will be one of the reliable methods for SOM prediction.

1. Introduction

Cytochrome P450 (CYP) is a typical monooxygenase including heme inside and plays important roles for metabolism of small molecules in animals, plants, and fungi.[1–4] In particular, human CYP family expressed in liver has been intensively studied because CYPs are involved in metabolizing not only endogenous small molecules but also exogenous bioactive compounds such as drugs and pollutants.[5–8] CYPs can be classified into some families and subfamilies based on their sequence similarities and about thirty family members are known for human.[9] Among many CYP members, CYP3A4 is one of the most important enzymes due to its higher expression level and treatability of wide range substrates.[10–13] Therefore if one substrate inhibits the metabolic activity of CYP3A4, it can widely affect the pharmacokinetics of other substrate. This inhibition sometimes causes severe side effects and is called as drug–drug interaction.[14–17] For this reason, it is essentially important to pay attention to interaction of drug candidates with CYP3A4 in drug discovery. At the early stage of the drug discovery process, the above mentioned CYP inhibition, binding affinity for CYP and metabolic stability of the compounds in liver microsome are examined or predicted in silico and the results together with other ADME/Tox profile are used for assessment of quality of lead compounds.[18–27] To reduce the CYP inhibition and enhance the metabolic stability of compounds, predictions of sites of metabolism (SOM) or binding modes with CYPs are quite helpful in addition to assays of the inhibition and stability in experiments. Using the predicted SOM, in other words, metabolic hot spots, a substituent around the sites can be converted into a metabolically more stable functional group.[28–30] The docking models between compounds and CYPs are also informative to design new derivatives without CYP inhibition.

There are two main factors for the SOM prediction. One is accessibility of each atom in a substrate to the oxygenated Fe atom of heme in a CYP protein. The other is the oxidative reactivity of each

* Corresponding authors. Tel.: +81 89 926 7265; fax: +81 89 926 7162 (M.H.); tel./fax: +81 43 226 2936 (T.H.).
E-mail addresses: mhata@cc.matsuyama-u.ac.jp (M. Hata), hoshi@p.chiba-u.ac.jp (T. Hoshino).

0968-0896/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2011.12.004

atom in a substrate. To predict accessibility of substrate atoms to the Fe atom, docking simulation using the substrate and a CYP protein is one of the potent approaches. In the docking simulation, structural information on protein is needed. However, until the early 2000s, it had been difficult to crystallize human CYP structures because they are membrane-associated proteins. Therefore homology models based on X-ray structures of CYPs of other species had been used for in silico study.[31–36] In 2004, two groups published the first human CYP3A4 crystal structures and then complex structures with several substrates have been reported.[37,38] At 2009, six structures were registered in Protein Data Bank (PDB). These structures deepened our understanding on protein behaviors such as induced fit of CYP3A4 by interacting with various small compounds. From the superimposition of the six CYP3A4 structures, large conformational change of the F–G loop and side chains of residues around the loop were observed (Fig. 1). The F–G loop was disordered in 2J0D because of its flexibility and could not be modeled with sufficient reliability.[39] In the heme pocket, progesterone in 1W0F and erythromycin in 2J0D occupied the position of the F–G loop in other CYP3A4 structures (1W0E, 1W0G, and 2V0M) and this means that both substrates pushed the loop away from the original position.[39] If the substrate-specific binding region is flexible such as CYP3A7,[40] conventional protein-rigid docking calculation would not be effective and lead to incorrect answers. To assess the performance of the conventional docking software for CYP3A4, we extracted the ligands in six CYP3A4 X-ray structures and docked them to other five CYP3A4 structures (cross-docking simulation) using Glide. Root-mean-square deviation (RMSD) between the docked ligand poses and their native poses were calculated. Surprisingly, none of the top five poses by Glide score was satisfactorily close to the correct answer (RMSD <2 Å). To reproduce induced fit, docking protocols considering protein flexibility or facilitating molecular dynamics (MD) simulations have been examined.[41–44] One of the docking approaches considering protein flexibility is the combination of protein conformation ensemble and protein-rigid docking.[45,46] Currently, this approach seems to be a standard technique to predict side chain conformations or relatively small change of main chain conformations. Since the main chain of F–G loop of CYP3A4 is largely changed according to the X-ray structures,[47–50] MD simulation is suitable for more accurate prediction of the binding mode.

In this study, we utilized MD simulation for SOM prediction in which multiple MD runs were executed using different initial structures. In general, the results of MD simulations with nano second (ns) time scale are largely dependent on the initial structures.[51] Movements of atoms are not satisfactorily reproduced only with a single MD run using one initial structure. In particular in the case of a flexible protein, conformations appearing in a single simulation are limited and calculation time more than micro second would be needed.[52] Multiple MD simulations studies were reported to efficiently explore conformation space of protein–substrate complex structures and improve ligand–protein binding affinity predictions.[53–55] To improve the independency on the initial structures in the nano-second scale simulation of the flexible CYP3A4, we prepared multiple CYP3A4-compound complexes by docking simulation (Fig. 2). The docking simulation generated diverse substrate-binding structures. A series of MD simulations with the respective complexes are expected to more comprehensively explore possible binding modes of a given compound. We selected carbamazepine (CBZ) as a compound to evaluate the above protocol for the purpose of SOM prediction. Since CBZ is widely used as an antiepileptic drug, its metabolites have been studied by many groups.[56–63] Another reason is its rigid framework. If a compound has too many rotatable bonds, the compound conformations have a large influence on the accessibilities of each atom rather than protein conformations. Therefore, it is safe to

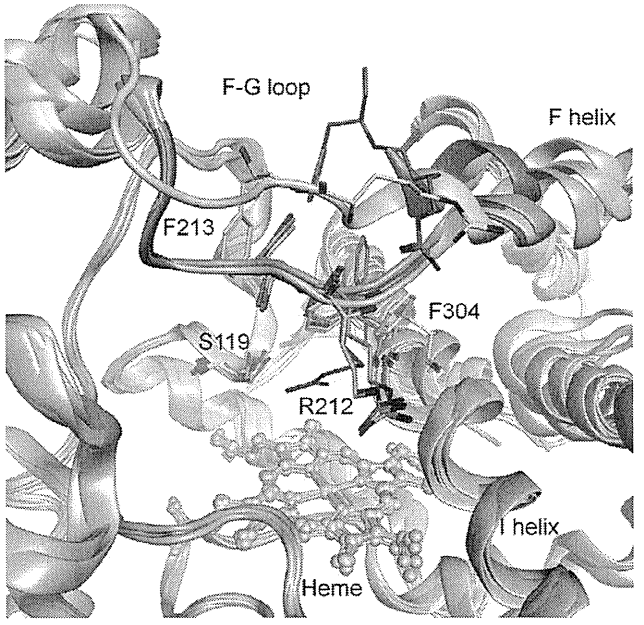

Figure 1. Superimposition of 6 crystal structures for CYP3A4. The heme is depicted in a ball-and-stick representation. Structures in red, magenta, yellow, orange, blue and green are 1TQN, 1W0E, 1W0F, 1W0G, 2J0D and 2V0M, respectively.

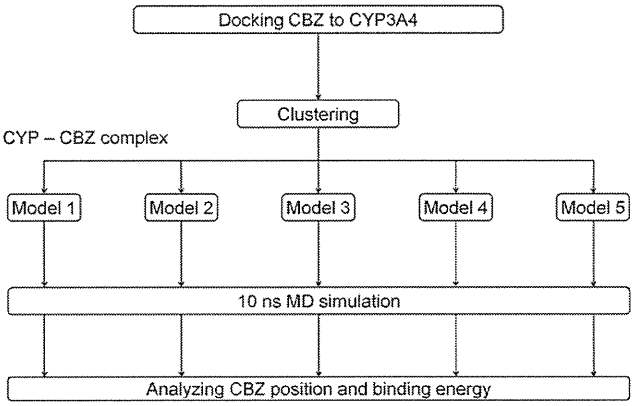

Figure 2. Workflow of MD simulations using multiple initial structures.

exclude compound flexibility in the evaluation of our present approach. CBZ has the tricyclic framework and consists of the central azepine ring and two benzene rings on both sides. CYP3A4 is a primary enzyme responsible for the metabolic clearance of CBZ and CYP2C8 is a minor one.[64] The main metabolite form is carbamazepine 10, 11-epoxide, whose double bond at the azepine ring is epoxidized (Scheme 1). In other metabolites, hydroxides of the benzene rings, *N*-glucuronides at the carbamoyl side chain were reported.[65,66] Kerr et al. showed the epoxidation was mainly caused by CYP3A4 and CYP3A4 hardly involved in production of hydroxides of the benzene rings.[64]

To analyze the reactivity of atoms on CBZ, Hata et al. examined density functional theory calculations of metabolites of CBZ (10, 11-epoxide and hydroxides of the benzene ring).[67] According to their analysis, 10, 11-epoxide showed about 20 kcal/mol higher potential energy than those of hydroxides of the benzene ring. This suggests that hydroxylation of the benzene ring will be thermodynamically major products. However, as described above, only 10, 11-epoxide form was produced by CYP3A4. Hence, accessibility and activation energy of oxidative reaction should be examined