Energetic and Kinetic Origin of CALB Interfacial Activation Revealed by PaCS-MD/MSM

*Tegar N. Wijaya[1,2] and *Akio Kitao[1]*

[1] School of Life Science and Technology, Tokyo Institute of Technology. 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8550, Japan. Tel +81-3-5734-3373

[2] Department of Chemistry, Universitas Pertamina, Jl. Teuku Nyak Arief, Simprug, Jakarta 12220, Indonesia. Tel +62 812-1149-9919

*Corresponding author: Akio Kitao, School of Life Science and Technology, Tokyo Institute of Technology, M6-13, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8550, Japan

E-mail: akitao@bio.titech.ac.jp

ABSTRACT

Conformational dynamics of *Candida antarctica* Lipase B (CALB) was investigated by molecular dynamics (MD) simulation, parallel cascade selection MD (PaCS-MD), the Markov state model (MSM), and mainly focused on the lid-opening motion closely related to substrate binding. All-atom MD simulation of CALB was conducted in water and that around the interface constructed by water and tricaprylin. CALB initially situated in water and separated by layers of water from the interface is spontaneously adsorbed onto the tricaprylin surface during MD simulation. The opening and closing motions of the lid are simulated by PaCS-MD and subsequent MSM analysis provided the free energy landscape and time scale of the conformational transitions among the closed, semi-open, and open states. The closed state is the most stable in the water system but the stable conformation in the interface system shifts to the semi-open state. In the interface system, the transition probability to the open state is higher than in the water system. These effects could explain the energetics and kinetics origin of previously reported interfacial activation of CALB. We also suggest two types of mechanisms for substrate binding in which small and hydrophilic substrates bind without interfacial activation while large and bulky substrates bind via interfacial activation. These findings could help expand the application of CALB towards a wide variety of substrates.

INTRODUCTION

Lipase is a class of enzymes with wide applications in industry[1]. Lipase catalyzes various reactions such as hydrolysis[2,3], esterification[4,5], and transesterification[6,7]. Central to lipase activity is interfacial activation[8], which enhances activity upon adsorption on the interface between water and lipids. *Candida antarctica* Lipase B (CALB) is one of the most widely used lipases in industry and academia[9] and is utilized in various applications such as biodiesel production[10,11] and chemical synthesis[12,13].

The three-dimensional structure of CALB has been resolved by X-ray crystallography[14]. The structure shows that a catalytic triad comprising Ser105, Asp187, and His224 is located behind a lid region consisting of two α helices, the small α5 (residues 142–146) helix and the large α10 (residues 268–287) helix (Figure 1A).

The interfacial activation of lipases was observed for different types of hydrophobic surface, e.g., the substrate[15,16] and solid surfaces[17,18]. An early study showed that CALB does not exhibit significant interfacial activation compared to other lipases[16]. However, a later report indicated the interfacial activation of CALB immobilized on a highly hydrophobic surface towards a bulky substrate[18]. Interfacial activation is controlled by conformational dynamics of the lid region, especially by the α5 helix, as indicated by the open and closed conformations of the crystal structures of CALB[19]. Therefore, understanding the dynamics of the lid region and its interaction with the hydrophobic surface is important to understand the mechanisms of substrate binding and catalysis.

Molecular dynamics (MD) simulation is a powerful molecular simulation method for studying complex molecular systems at the atomic level. Several MD studies have been conducted to investigate the interfacial activation of CALB. A conventional MD (CMD) simulation study

performed to construct the free energy landscape of CALB in water showed the existence of close, crystal-like, and open conformations with similar free energy values[21]. In contrast, a lower probability for the open conformation in water was reported by replica exchange MD of CALB[18]. Another study using CMD simulation indicated that interfacial activation is due to a lamellar-like nanostructure formed by triglycerides (tricaproin and tricaprylin) in a water-triglycerides system[22]. Furthermore, the role of the conformational dynamics of CALB during substrate binding was investigated by MD and the Markov state model (MSM)[20].

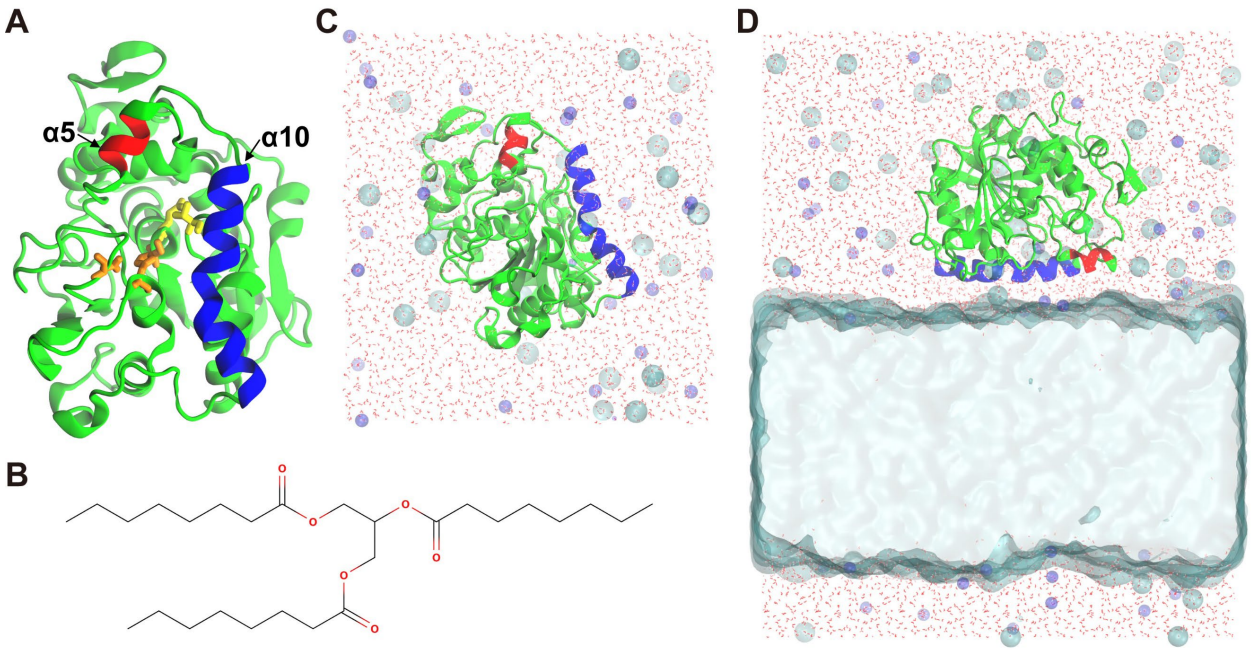

Figure 1. A. Three-dimensional structure of CALB (PDB id: 1TCA) with new cartoon representation. The α5 lid region (residues 142–146), α10 lid region (residues 268–287), the catalytic triad (Ser105, Asp187, and His224), and oxyanion hole (Thr40 and Gln106) are colored red, blue, orange and yellow, respectively. B. The chemical structure of tricaprylin. C and D are the initial configurations in the molecular dynamics simulation for CALB in water (CALB/W) and on the tricaprylin-water interface (CALB/I), respectively. Molecular images in this paper were created using Visual Molecular Dynamics(VMD)[23]

Improvements should be achieved by employing enhanced conformational sampling methods with explicit solvent models for the hydrophobic surface, followed by analysis of the free energy landscape concerning the lid open-close motion. In this regard, the combination of Parallel Cascade Selection Molecular Dynamics (PaCS-MD)[24], an enhanced sampling method, and the Markov State Model (MSM)[25–27], namely PaCS-MD/MSM, can be used. In PaCS-MD, efficient sampling of conformational changes is achieved by iteratively performing cycles of parallel short MD simulations. In each cycle, the initial structures are selected based on a collective variable describing the progression of conformational changes of interest and MD runs are restarted with re-initialized velocities, effectively enhancing the probabilities of occurrence toward the expected conformational changes. Typically, the number of MD simulations conducted in parallel (the number of replicas, n_{rep}) is 10–100, and the length of MD in each cycle is sub-nanoseconds to nanosceconds[24,28]. In the next step, PaCS-MD trajectories are used as input to build an MSM. Free energy landscape and some kinetic parameters of the conformational dynamics can be obtained from the MSM. The PaCS-MD/MSM combination has been used to study protein dynamics[29–31], protein-ligand[31–33], protein-protein[34–36], and protein-DNA interactions[37].

In this paper, we employ tricaprylin (Figure 1B) as a model bulky molecule to construct the hydrophobic surface. Conformational dynamics of CALB in water (hereafter CALB/W) and on the interface of water and tricaprylin (CALB/I) was investigated by MD and PaCS-MD/MSM, which allows direct observation of the effects of tricaprylin on the free energy surface and kinetic properties of CALB conformational dynamics.

METHODS

Parameterization of tricaprylin

The atom types and forcefield parameters for tricaprylin were adopted from the AMBER lipid14 forcefield[38]. All quantum calculations were performed using the Gaussian16 software package[39], while topology generation and subsequent MD simulations were conducted using the AMBER package[40] except where otherwise stated. The initial conformation of tricaprylin monomer was constructed by the molefacture plugin of Visual Molecular Dynamics (VMD) software[23], and geometry optimization and electrostatic potential (ESP) calculations were performed. To determine the atomic partial charges of tricaprylin monomer, the RESP (Restrained Electrostatic Potential) procedure[41] at the B3LYP/6-31G* level of theory was employed. Atomic partial charges were fitted using Antechamber[42]. Then, an ensemble of tricaprylin conformations was obtained from 60-ns MD simulation of 144 tricaprylin molecules in the liquid phase at 298 K and 1 atm. For each of the 144 conformers of tricaprylin taken from the last snapshot of the MD trajectory, ESPs were calculated and the partial charges were determined using the aforementioned procedure. The final atomic partial charges were obtained by averaging. To examine the obtained force field, a 50-ns MD simulation of a tricaprylin in the gas phase and that of 144 tricaprylins in the liquid phase were conducted. The obtained density and enthalpy of vaporization was 0.954 $kg.L^{-1}$ and 144.4 $kJ.mol^{-1}$, which reproduced the experimental values of 0.954 $kg.L^{-1}$ [43] and 135.4 $kJ.mol^{-1}$ [44], respectively. This parameter set for tricaprylin was used in the following simulations.

CMD simulation of CALB in water

To construct the CALB/W system, the crystal structure of CALB taken from the Protein Data Bank (PDB id: 1TCA) was protonated and solvated with 12930 OPC water molecules[45] into a cubic box using the tleap module of the AMBER package[40]. The system was neutralized and additional NaCl ions were added to reproduce a 0.15 M NaCl solution. The Amber FF19SB forcefield[46] was used to describe the protein. The constructed CALB/W system (Figure 1C) was energy minimized using the steepest descent method to remove bad contacts between atoms introduced during preparation of the model. The minimized system was equilibrated at 298.15 K using the Langevin thermostat for 500 ps while the volume of the system was kept constant. During equilibration, positional restraints were applied to the Cα atoms of the protein with a force constant of 10 $kcal.mol^{-1}Å^{-2}$. The system was then brought to correct density by isotropically performing a 500 ps NPT equilibration (wherein the Number of particles, Pressure, and Temperature are all constant) run at 298.15 K and 1 atm using the Langevin thermostat and Berendsen barostat while keeping the positional restraints. The following pre-production run was performed without the positional restraints at 298.15 K and 1 atm using the Langevin thermostat and Mote Carlo barostat for 500 ps. Finally, 10 independent production runs were performed for 500 ns using the same settings as the pre-production run. After equilibration, the length of the box edge was around 7.5 nm.

The following procedure was common for both CALB/W and /I. During the equilibration and production runs, the long-range electrostatic interactions were calculated by the Particle Mesh Ewald (PME) method and the real-space nonbonded cutoff was made at 1 nm. All of the procedures above were conducted using the GPU-capable pmemd module of AMBER20[40].

MD simulation of CALB in the tricaprylin-water interface system

The tricaprylin-water interface was constructed by a method similar to those employed in the previous reports[22,47]. First, 144 tricaprylin and 3806 water molecules were randomly inserted into a cubic box using Packmol[48], to prepare the mixture of tricaprylin and water. The number of each molecule type was chosen, so that the volumes of tricaprylin and water would be the same. The system was energy minimized, and then equilibrated using the NVT ensemble with the V-rescale method at 300 K for 15 ns. Phase separation spontaneously occurred during this step. Then, the z direction was selected as the direction of phase separation, and the system was equilibrated for 50 ns using the NPAT ensemble at 300 K and 1 atm with the V-rescale method and the Berendsen barostat, keeping the box size along the x and y directions constant while the box size along the z direction was free to change. In the next step, the equilibrated interface system was duplicated along the x and y directions. The large interface system was further equilibrated for 50 ns using the NPAT ensemble at 298 K and 1 atm. To insert CALB into the water phase of the tricaprylin-water interface system, water molecules were removed except for those situated within 0.3 nm of the tricaprylin molecules. Then, CALB and the surrounding water molecules were taken from the equilibrated CALB/W system and placed 0.7 nm above the tricaprylin surface, so that the lid faced the interface. Finally, the system was re-solvated with water and 0.15 M NaCl. The above preparation method was conducted using the GROMACS 2019 package.[49] The final CALB/I system contained 576 trycaprylin and 21391 water molecules.

The CALB/I system (Figure 1D) was energy minimized and equilibrated for 1.0 ns under isothermal-isochoric conditions at 298.15 K using the Langevin thermostat, imposing positional restraints onto the Cα atoms of the protein and the heavy atoms of tricaprylin with a force constant of 10 $kcal.mol^{-1}Å^{-2}$. The system was then equilibrated for 1.0 ns with the NPAT

ensemble at 298.15 K and 1 atm using the Langevin thermostat and Monte Carlo barostat while keeping the positional restraints. During this step, the box size along the x and y directions was kept constant while the box size along the z coordinate was free to move. A pre-production run was then performed with the NPAT ensemble without the positional restraints at 298.15 K and 1 atm using the Langevin thermostat and Mote Carlo barostat for 1.0 ns. Corresponding to CALB/W, 10 independent production runs were performed for 500 ns. After equilibration, the box size was around $8.9\times10.4\times12.6$ nm^3.

PaCS-MD Procedure

The closed structure of both systems was chosen as the input for PaCS-MD. The initial closed structure of CALB/W was selected directly from the last snapshot of the ninth CMD run. However, the initial closed structure of CALB/I could not be obtained directly because all ten CMD runs did not sample closed conformations adequately in the case of CALB/I. Therefore, a single trial of PaCS-MD which consisted of 200 cycles was performed according to the cPaCS-MD procedure, explained later. The input for this single trial was chosen from the last snapshot of the fifth CMD run. The snapshot from the last cycle was used as the initial closed structure for CALB/I. The obtained initial closed structure of both systems was used in the subsequent PaCS-MD procedure.

For both CALB/W and /I, PaCS-MD simulations were performed in two stages. After preliminary MD to generate multiple initial structures, PaCS-MD was performed to sample conformational dynamics from the closed to open conformation, which here we call oPaCS-MD. In the second stage, open to closed conformational dynamics were sampled starting from the end snapshots from oPaCS-MD. The second stage is denoted as cPaCS-MD. A more detailed PaCS-

MD procedure in each stage is pictured in Figure 2. The inter-lid distance ***d*** used in the structural ranking of the PaCS-MD trajectories is defined as the inter-center of mass distance between the α5 helix (residues 142–146) and half of the α10 helix (residues 278–287). The length of MDs in each PaCS-MD cycle was 0.1 ns and n_{rep} was 30. First, a 1 ns preliminary MD simulation was performed, followed by selection of the top 30 structures with longer ***d*** as the initial structures for the first cycle of oPaCS-MD. After 0.1 ns MD simulations, structural ranking from the obtained trajectories and selection of the initial structures for the next cycle were performed. This cycle was repeated until ***d*** became longer than 3.0 nm. Then, the simulation was switched to cPaCS-MD by selecting the snapshots with shorter ***d*** until reaching 0.8 nm.

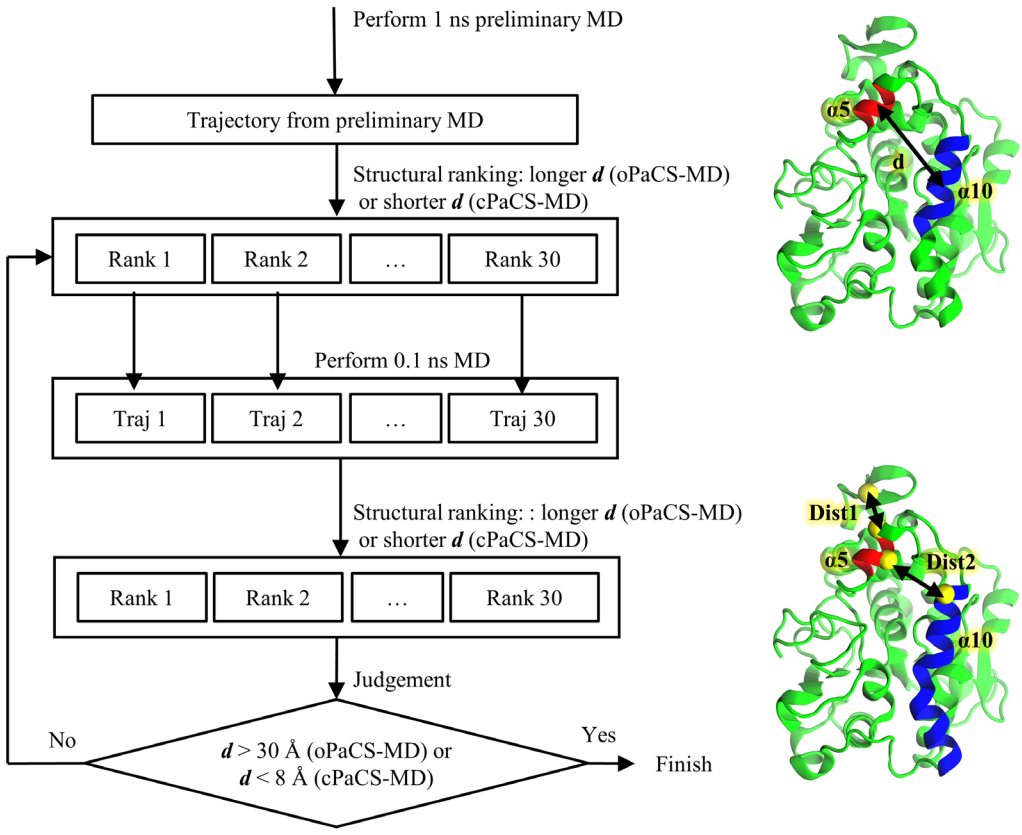

Figure 2. PaCS-MD procedures (left panel) and parameter (right top panel). MSM features (Dist1 and Dist2, right bottom panel) are also shown.

Table 1 summarizes the conducted CMDs and PaCS-MDs. The number of oPaCS-MD trials conducted for CALB/W and /I was 20 and 5, respectively. After each trial of oPaCS-MD, two trials of cPaCS-MD were continued with different initial velocities. The simulation cost of PaCS-MD per trial is defined as 0.1 ns × the number of cycles × n_{rep}. The total PaCS-MD simulation cost for CALB/W and /I was 5.60 and 8.89 μs, respectively.

Table 1. Summary of CMD and PaCS-MD.

	CMD		oPaCS-MD		cPaCS-MD	
	CALB/W	CALB/I	CALB/W	CALB/I	CALB/W	CALB/I
Number of replicas	-	-	30	30	30	30
Number of trials	10	10	20	5	40	10
Average number of cycles per trial	-	-	49.6±9.5	292.0±64.2	22.5±5.4	150.4±46.1
Average simulation cost per trial (μs)	0.5	0.5	0.149±0.028	0.876±0.193	0.067±0.016	0.451±0.138
Total simulation cost (μs)	5.0	5.0	2.98	4.38	2.70	4.51

Analysis by MSM

Analysis by MSM was conducted by PyEMMA2.5.12[50]. The inter-Cα distances between Arg309 and Leu144 (Dist1) and Ala146 and Val286 (Dist2) were chosen as the features to construct MSM (Right bottom of Figure 2). Dist2 directly measures the lid opening, and Dist1 shows the effect of the lid opening to the β-hairpin situated outside α5. As the lid becomes more open, Dist1 and Dist2 tend to be shorter and longer, respectively. Using these two features, the

snapshots sampled by PaCS-MD were clustered into 1000 microstates using k-means clustering[28] with a k-means++ initialization strategy[51].

Each element of the transition probability matrix of MSM, $T = \{T_{ij}(\tau)\}$, was calculated based on the features at time t ($x(t)$) and time $t + \tau$ ($x(t + \tau)$) according to Eq. (1), where τ is the lag time, and S_i and S_j represent the microstate before and after transition[26].

$$T_{ij}(\tau) = \mathbb{P}[x(t + \tau) \in S_j | x(t) \in S_i] \quad (1)$$

The stationary probability of microstates, $\boldsymbol{\pi} = \{\pi_i\}$, is obtained by solving the eigenfunction in Eq. (2)[26], and the free energy of each state was calculated using Eq. (3).

$$\pi = \pi T \quad (2)$$

$$G_i = -k_B T \ln \pi_i \quad (3)$$

The free energy landscape was obtained by projecting the free energy into the two-dimensional space spanned by Dist1 and Dist2.

To obtain a more coarse-grained view of the open-close motion of the lid, the macrostates analysis was conducted. The assignment of the macrostates was performed using PCCA++[52]. The flux network between the obtained macrostates was analyzed using transition path theory (TPT)[26].

RESULTS

Interactions between CALB and the tricaprylin surface

CMD simulations of the interface system (CALB/I) showed strong interaction between CALB and the tricaprylin surface as follows. While CALB and the tricaprylin surface were initially separated by layers of water around 7 Å thickness with the lid region oriented to the surface (Figure 1C), CALB was adsorbed onto the tricaprylin surface by the end of all CMD simulations

(Figure 3). This is expected because the lid region consists of mostly hydrophobic residues. Strong interaction between the lid and tricaprylin surface was also reported previously[22,53].

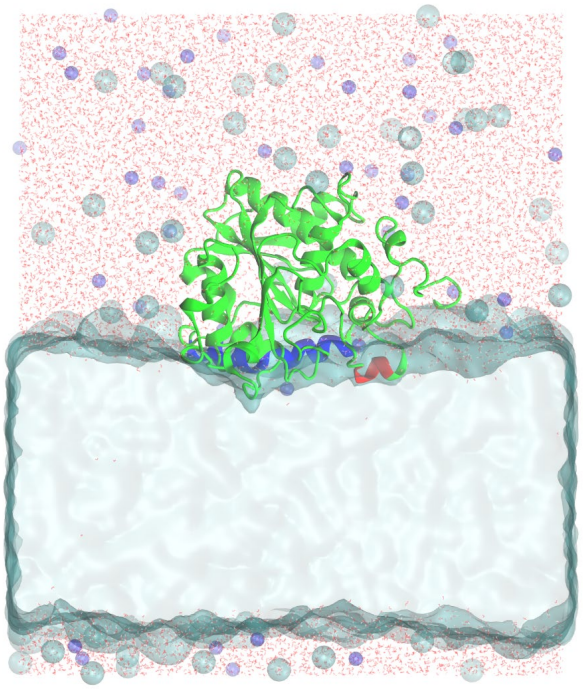

Figure 3. A representative snapshot from the CMD simulations showing CALB adsorbed onto the tricaprylin surface.

Conformational dynamics of CALB investigated by CMD and PaCS-MD

To compare the conformational space sampled by CMD and PaCS-MD, the obtained trajectories were projected onto the Dist1-Dist2 space (Figure 4). In CALB/W, PaCS-MD sampled a much wider area than CMD and the difference is more apparent for CALB/I.

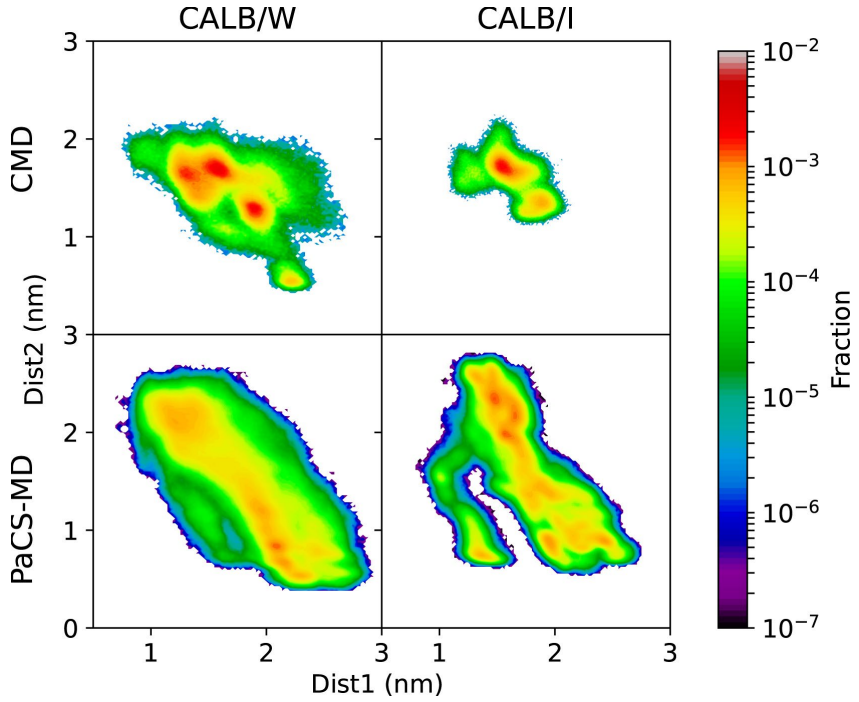

Figure 4. Projections of CMD and PaCS-MD trajectories (left and right) for both water and the interface system plotted on the Dist1–Dist2 space.

MSM analysis

MSMs for both systems were constructed by using Dist1 and Dist2 as the features, and the snapshots obtained by PaCS-MD were clustered into 1000 microstates using the k-means clustering approach (See Methods). The appropriate lag time and validity of the MSMs were determined from the implied timescales (ITS) plot (Figure S1). Each MSM was validated by plotting ITS over a series of lag times from 1 to 50 ps. The MSM is validated if the slowest timescale in the ITS plot reaches a certain constant value. The ITS plot showed that the slowest timescale reaches a relatively constant value after 30 ps for CALB/W and /I, thus validating the MSM. For a typical long MD simulation, the smallest possible lag time is chosen to prevent any information lost due to a long lag time[26]. However, a longer lag time assures better Markovianity

than a shorter lag time[26]. Since the PaCS-MD trajectories are short (100 ps each) in this work, the longest reasonable lag time is half of the MD length (50 ps). Therefore, to achieve better Markovianity for the MSM of PaCS-MD trajectories, the longest possible lag time should be chosen as long as the slowest implied timescale reaches a constant value. Therefore, 50 ps was selected. The use of 100 ps MD and 50 ps lag time was shown to reproduce experimentally determined free energy values in many cases[32–37].

The free energy landscape (FEL) of both systems (Figure 5) was calculated from the stationary distribution obtained from the transition probability matrix (Eqs. (2) and (3)). Here, we categorized the CALB conformations into three types, closed: Dist2 ≤ 1.0 nm, semi-open: 1.0 nm < Dist2 < 2.0 nm, and open: Dist2 ≥ 2.0 nm. In this classification, the crystal structure (PDB id: 1TCA) is categorized into the semi-open state. For CALB/W, the FEL has four free energy minima (*w1–4*), numbered from the lowest free energy. Three minima (*w1−3*) are located close to each other while the other (*w4*) is located far in the left area and has a much higher free energy. The global minimum *w1* and local minimum *w2* are classified into the closed conformation while local minima *w3* and *w4* correspond to the semi-open conformation. On the other hand, the FEL of CALB/I shows only two minima located close to each other (*i1* and *i2*), which are categorized into the semi-open conformation. The depths of the two minima are comparable.

The FEL also indicates differences in low free energy areas in the conformational space. The FEL of CALB/W covers larger conformational space than CALB/I, which indicates more flexibility of the lid region in water. This is expected because CALB in the interface system is adsorbed on the tricaprylin surface which restricts lid movement. The low free energy areas around *i1* and *i2* are more restricted compared to the low free energy areas around *w1–3*, which indicates more stabilization or less flexibility of the lid region in CALB/I.

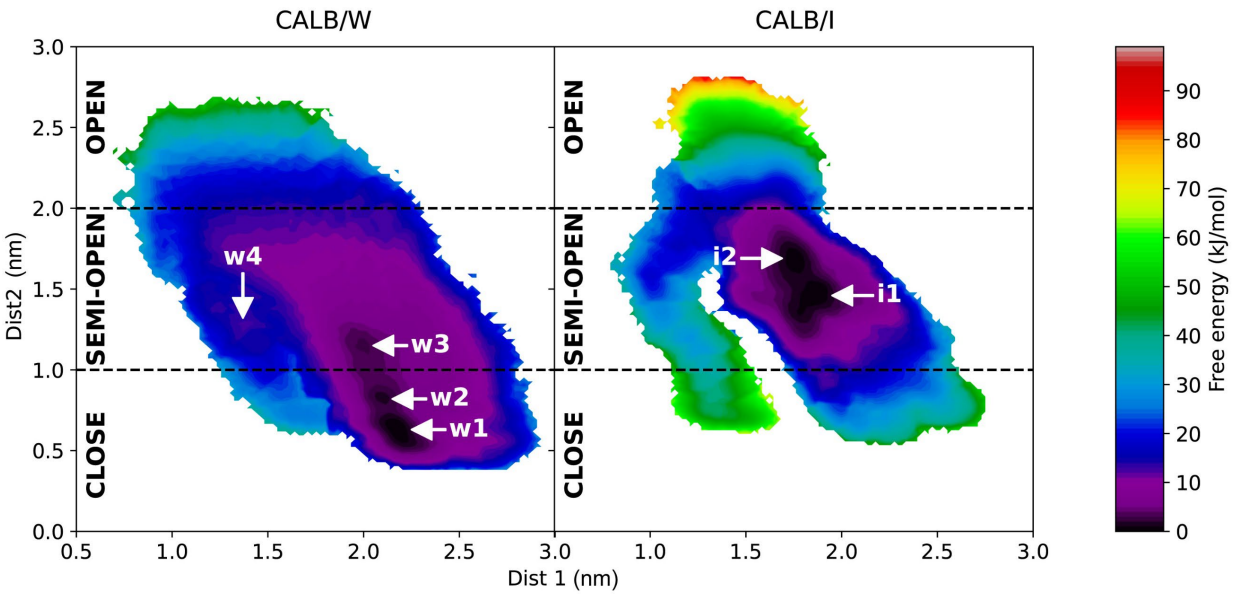

Figure 5. Free energy landscape (FEL) of CALB/W and /I obtained by PaCS-MD/MSM.

To obtain a more macroscopic view of FEL, macrostate assignment was performed using PCCA++ (Figure 6). The number of macrostates for both systems was chosen so that the conformational space of the global minima is well defined. This requirement was satisfied with six macrostates for both cases. In CALB/W, the conformational space was divided into three closed states (WC1, WC2, and WC3), one semi-open state (WS), and two open states (WO1 and WO2). The global free energy minimum *w1* was assigned to WC2, while *w2* and *w3* were assigned to WS. Although WC2 contains the global minimum, its free energy in the macroscopic view is slightly higher (0.9 kJ/mol) than WS. Transition times to WS from the other macrostates tend to be faster (≤ 1 ns) than transitions from WS to the other states (> 2.5 ns) but the transitions between WC2 and WS are comparable. Transition times from the semi-open state (WS) to the open states (WO1 and WO) are 47 and 46 ns, respectively. From WS, transition to WC2 and

WC3 is 2.5 and 3.7 ns, respectively, which shows that the transition from the semi-open to the closed state is one-order of magnitude faster than that to the open state.

For CALB/I, the conformational space was divided into three closed states (IC1, IC2, and IC3), two semi-open states (IS1 and IS2), and one open state (IO). Both *i1* and *i2* were assigned to IS2, which shows the lowest free energy, and the other macrostates are situated in higher energy areas of the FEL. The transition time from the semi-open state (IS2) to the open state (IO) is 493 ps, which is much faster than the similar transition in CALB/W. The transition time from IO to IS2 is also fast (373 ps), indicating that the transitions between the semi-open and open states occur frequently in less than the nanosecond time scale in CALB/I.

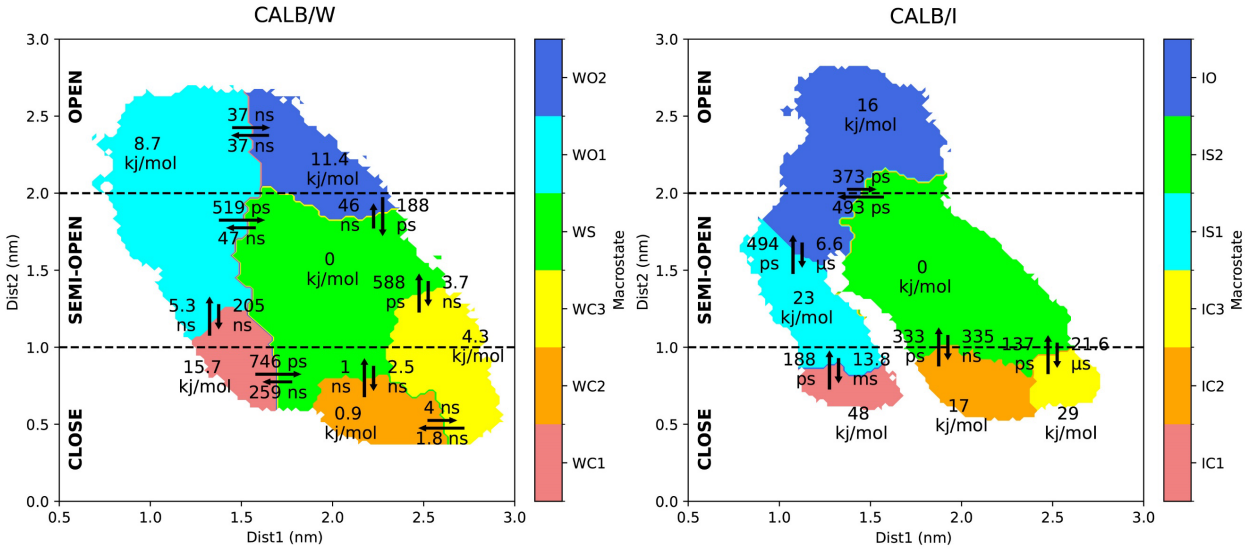

Figure 6. Macrostate assignment using PCCA++ for both CALB/W and /I. PCCA++ results and MFPT between macrostates for CALB/W (left) and CALB/I (right). Transition time and free energy difference are also indicated.

Analysis by transition path theory

To investigate the main transition pathway from the closed to open state, the flux network was calculated by the coarse-grained MSM obtained by PCCA++ and TPT (Figure 7). For CALB/W, WC2 and WO2 were selected as the starting and end points, respectively, and IC2 and IO were selected as the starting and end points for CALB/I, respectively.

The flux network of CALB/W showed several pathways from the closed (WC2) to open (WO2) state, and the major pathway is WC2→WS→WO2. Interestingly, the intermediate WS is the most populated macrostate. In CALB/I, the closed (IC2) to open (IO) state pathways are more restricted, the major pathway being IC2→IS2→IO. Similar to CALB/I, the closed to open transition goes through the most populated macrostate, IS2. IC1 and IS1 located in the left arm of the free energy landscape (Figure 6) are disconnected from the network, indicating that the closed to open transition does not go through these macrostates. Interestingly, the most populated macrostate of each system has a different committor probability. The committor probability of WS in CALB/W was 0.12 while that of IS2 in CALB/I was 0.7. In this particular case, a larger committor probability means that the probability of transition to the open state is higher than the transition to the closed state and vice versa. Therefore, CALB tends to make a transition to the closed state in water but moves to the open state in the interface. This result is consistent with the aforementioned transition time results.

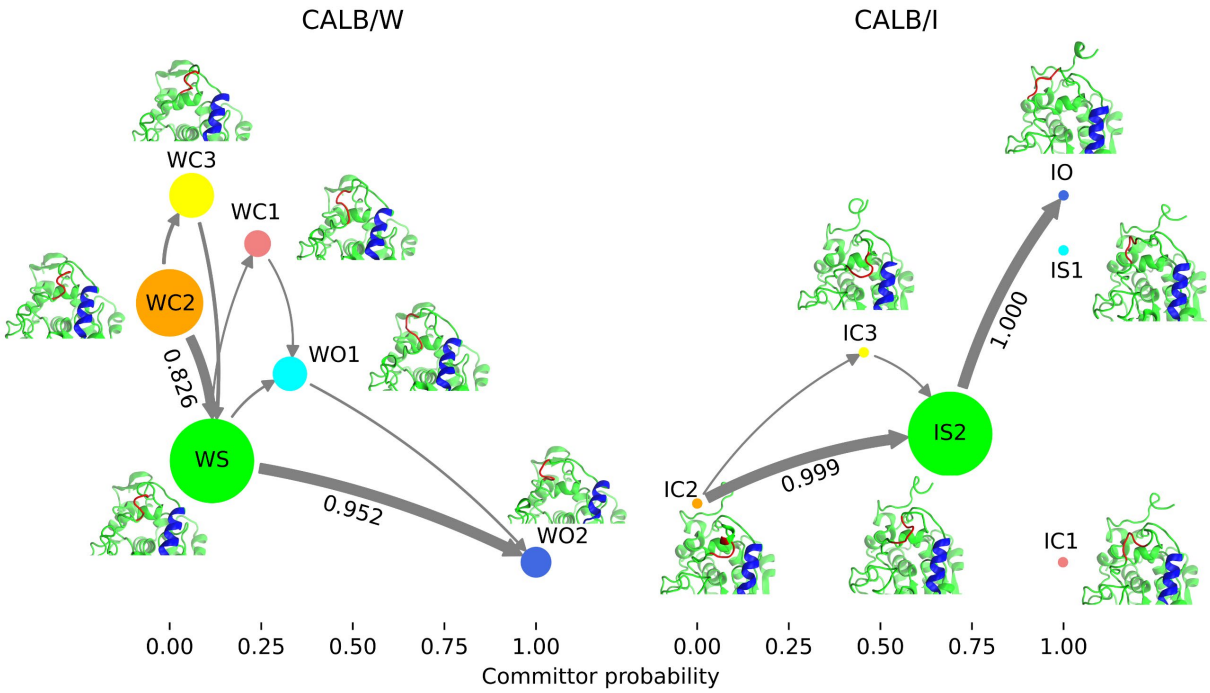

Figure 7. Flux network of from the closed to open conformation for CALB/W and /I. The macrostates were arranged by their committor probabilities along the abscissa. The values near the arrows are the flux fraction of each transition between macrostates. For CALB/W, the circle area is proportional to $-1/(\ln \pi)$ of the corresponding macrostate. For CALB/I, the ratio of circle area between IS2 and other macrostates is fixed at 100:1 for clarity due to stationary probability of IS2 which is over 0.99.

DISCUSSION

The results of this research indicate that the tricaprylin surface affects conformational dynamics of the CALB lid in two ways. First, it shifts the global minimum from the closed conformation in water to the semi-open conformation in the interface system. Second, the semi-open conformation in CALB/I has higher probability to transition to the open conformation than to the closed conformation. On the other hand, the semi-open conformation in water is more

likely to move to the closed conformation. Adding these two effects together, the tricaprylin surface has significant effects of promoting a more open conformation, leading to easier substrate binding. These results are consistent with experimental evidence for CALB interfacial activation when CALB is immobilized on highly hydrophobic silanized beads[18].

It was reported that CALB apparently does not undergo interfacial activation for small and hydrophilic substrates[16,18] while interfacial activation occurs for bulky substrates at the hydrophobic surface[18]. A self-activation mechanism has been proposed to explain this observation based on the similar free energy between the open and closed conformations obtained from the CMD simulation of CALB in water[21]. However, our findings are in disagreement with this mechanism because we found that the free energy of the closed conformation is much lower than that of the open conformation in water. Our results are consistent with other MD simulations[18,22]. Therefore, an additional explanation is necessary to account for lack of interfacial activation for small and hydrophilic substrates.

We examined the crystal structure of CALB complexed with Tween 80 (substrate) (PDB id: 1LBT)[54] to understand the relation between lid conformation and substrate binding. In contrast to tricaprylin having three ester bonds and three tails, Tween 80 has only one ester bond and two tails. Therefore, Tween 80 might occupy less space around the catalytic site. In this complex, Tween 80 is bound to CALB in an appropriate position for catalysis to occur, and CALB shows Dist1 and Dist2 values of 16.9 and 15.5 Å, respectively, which is assigned as the semi-open conformation in our definition. This conformation also belongs to the most stable macrostate (WS) in water, which is easily accessible in water. From this analysis, interfacial activation might not be necessary for Tween 80 binding because the semi-open conformation is enough for the binding event. Smaller and hydrophilic substrates might require even less space for binding

than Tween 80. On the other hand, bulky substrates require larger opening in the CALB lid region and therefore interfacial activation was observed. Triglycerides, including tricaprylin, might also require larger opening because its catalysis by CALB is slower than that of mono- and diglycerides[55].

CONCLUSIONS

In this paper, PaCS-MD was employed to sample the conformational dynamics of CALB both in water and on the water-tricaprylin interface. The analysis of PaCS-MD trajectories using MSM indicated that the closed state is the most stable in the water system but the stable conformation in the interface system shifts to the semi-open state. In the interface system, the transition probability to the open state is higher than in the water system. These effects could explain the energetics and kinetics origin of previously reported interfacial activation of CALB. We also suggest two types of mechanisms of substrate binding in which small and hydrophilic substrates bind without interfacial activation, while large and bulky substrates bind via interfacial activation. These findings could help expand CALB applications towards a wide variety of substrates.

ACKNOWLEDGMENT

This research was supported by MEXT/JSPS KAKENHI Nos. JP19H03191, JP20H05439, JP21H05510, and JP22H04745 to A.K., and by MEXT as a “Program for Promoting Researches on the Supercomputer Fugaku” (Application of Molecular Dynamics Simulation to Precision Medicine Using Big Data Integration System for Drug Discovery, JPMXP1020200201 and Biomolecular Dynamics in a Living Cell, JPMXP1020200101) to A.K. This work used

computational resources of the supercomputer TSUBAME provided by Tokyo Institute of Technology, FUGAKU through the HPCI System Research Project (Project IDs: hp210029, hp210172, hp210177, hp220164, and hp220170), Research Center for Computational Science, The National Institute of Natural Science, and The Institute for Solid State Physics, The University of Tokyo.

REFERENCES

(1) Sarmah, N.; Revathi, D.; Sheelu, G.; Yamuna Rani, K.; Sridhar, S.; Mehtab, V.; Sumana, C. Recent Advances on Sources and Industrial Applications of Lipases. *Biotechnology Progress*. John Wiley and Sons Inc. January 1, 2018, pp 5–28. https://doi.org/10.1002/btpr.2581.

(2) Brígida, A. I. S.; Amaral, P. F. F.; Coelho, M. A. Z.; Gonçalves, L. R. B. Lipase from Yarrowia Lipolytica: Production, Characterization and Application as an Industrial Biocatalyst. *Journal of Molecular Catalysis B: Enzymatic*. Elsevier B.V. 2014, pp 148–158. https://doi.org/10.1016/j.molcatb.2013.11.016.

(3) Abe, K.; Gomi, K.; Hasegawa, F.; Machida, M. Impact of Aspergillus Oryzae Genomics on Industrial Production of Metabolites. *Mycopathologia*. September 2006, pp 143–153. https://doi.org/10.1007/s11046-006-0049-2.

(4) Xiao, Z.; Hou, X.; Lyu, X.; Zhao, J. yi; Xi, L.; Li, J.; Lu, J. R. Enzymatic Synthesis of Aroma Acetoin Fatty Acid Esters by Immobilized Candida Antarctica Lipase B. *Biotechnol Lett* **2015**, *37* (8), 1671–1677. https://doi.org/10.1007/s10529-015-1834-0.

(5) Sharma, S.; Kanwar, S. S.; Dogra, P.; Chauhan, G. S. Gallic Acid-Based Alkyl Esters Synthesis in a Water-Free System by Celite-Bound Lipase of Bacillus Licheniformis SCD11501. *Biotechnol Prog* **2015**, *31* (3), 715–723. https://doi.org/10.1002/btpr.2072.

(6) Garlapati, V. K.; Kant, R.; Kumari, A.; Mahapatra, P.; Das, P.; Banerjee, R. Lipase Mediated Transesterification of Simarouba Glauca Oil: A New Feedstock for Biodiesel Production. *Sustainable Chemical Processes* **2013**, *1* (1). https://doi.org/10.1186/2043-7129-1-11.

(7) Clementz, A. L.; del Peso, G.; Canet, A.; Yori, J. C.; Valero, F. Utilization of Discard Bovine Bone as a Support for Immobilization of Recombinant Rhizopus Oryzae Lipase Expressed in Pichia Pastoris. *Biotechnol Prog* **2016**, *32* (5), 1246–1253. https://doi.org/10.1002/btpr.2321.

(8) Verger, R. "Interfacial Activation" of Lipases: Facts and Artifacts. *Trends Biotechnol* **1997**, *15* (1), 32–38. https://doi.org/10.1016/S0167-7799(96)10064-0.

(9) Ortiz, C.; Ferreira, M. L.; Barbosa, O.; dos Santos, J. C. S.; Rodrigues, R. C.; Berenguer-Murcia, Á.; Briand, L. E.; Fernandez-Lafuente, R. Novozym 435: The "Perfect" Lipase Immobilized Biocatalyst? *Catal Sci Technol* **2019**, *9* (10), 2380–2420. https://doi.org/10.1039/c9cy00415g.

(10) Zhang, H.; Liu, T.; Zhu, Y.; Hong, L.; Li, T.; Wang, X.; Fu, Y. Lipases Immobilized on the Modified Polyporous Magnetic Cellulose Support as an Efficient and Recyclable Catalyst for Biodiesel Production from Yellow Horn Seed Oil. *Renew Energy* **2020**, *145*, 1246–1254. https://doi.org/10.1016/j.renene.2019.06.031.

(11) Du, W.; Xu, Y.; Liu, D.; Zeng, J. Comparative Study on Lipase-Catalyzed Transformation of Soybean Oil for Biodiesel Production with Different Acyl Acceptors. *J Mol Catal B Enzym* **2004**, *30* (3–4), 125–129. https://doi.org/10.1016/j.molcatb.2004.04.004.

(12) Cha, H. J.; Park, J. B.; Park, S. Esterification of Secondary Alcohols and Multi-Hydroxyl Compounds by Candida Antarctica Lipase B and Subtilisin. *Biotechnology and Bioprocess Engineering* **2019**, *24* (1), 41–47. https://doi.org/10.1007/s12257-018-0379-1.

(13) Santos, L. D. F.; Coutinho, J. A. P.; Ventura, S. P. M. From Water-in-Oil to Oil-in-Water Emulsions to Optimize the Production of Fatty Acids Using Ionic Liquids in Micellar Systems. *Biotechnol Prog* **2015**, *31* (6), 1473–1480. https://doi.org/10.1002/btpr.2156.

(14) Uppenberg, J.; Hansen, M. T.; Patkar, S.; Jones, T. A. The Sequence, Crystal Structure Determination and Refinement of Two Crystal Forms of Lipase B from Candida Antarctica. *Structure* **1994**, *2* (4), 293–308. https://doi.org/10.1016/S0969-2126(00)00031-9.

(15) Wang, S.; Xu, Y.; Yu, X. W. A Phenylalanine Dynamic Switch Controls the Interfacial Activation of Rhizopus Chinensis Lipase. *Int J Biol Macromol* **2021**, *173*, 1–12. https://doi.org/10.1016/j.ijbiomac.2021.01.086.

(16) Martinelle, M.; Holmquist, M.; Hult, K. On the Interfacial Activation of Candida Antarctica Lipase A and B as Compared with Humicola Lanuginosa Lipase. *Biochimica et Biophysica Acta (BBA)/Lipids and Lipid Metabolism* **1995**, *1258* (3), 272–276. https://doi.org/10.1016/0005-2760(95)00131-U.

(17) Zhang, J.; Wang, Z.; Zhuang, W.; Rabiee, H.; Zhu, C.; Deng, J.; Ge, L.; Ying, H. Amphiphilic Nanointerface: Inducing the Interfacial Activation for Lipase. *ACS Appl Mater Interfaces* **2022**, *14* (34), 39622–39636. https://doi.org/10.1021/acsami.2c11500.

(18) Zisis, T.; Freddolino, P. L.; Turunen, P.; van Teeseling, M. C. F.; Rowan, A. E.; Blank, K. G. Interfacial Activation of Candida Antarctica Lipase B: Combined Evidence from Experiment and Simulation. *Biochemistry* **2015**, *54* (38), 5969–5979. https://doi.org/10.1021/acs.biochem.5b00586.

(19) Stauch, B.; Fisher, S. J.; Cianci, M. Open and Closed States of Candida Antarctica Lipase B: Protonation and the Mechanism of Interfacial Activation. *J Lipid Res* **2015**, *56* (12), 2348–2358. https://doi.org/10.1194/jlr.M063388.

(20) Lu, C.; Peng, X.; Lu, D.; Liu, Z. Global and Kinetic Profiles of Substrate Diffusion in Candida Antarctica Lipase B: Molecular Dynamics with the Markov-State Model. *ACS Omega* **2020**, *5* (17), 9806–9812. https://doi.org/10.1021/acsomega.9b04432.

(21) Luan, B.; Zhou, R. A Novel Self-Activation Mechanism of: Candida Antarctica Lipase B. *Physical Chemistry Chemical Physics* **2017**, *19* (24), 15709–15714. https://doi.org/10.1039/c7cp02198d.

(22) Benson, S. P.; Pleiss, J. Self-Assembly Nanostructures of Triglyceride-Water Interfaces Determine Functional Conformations of Candida Antarctica Lipase B. *Langmuir* **2017**, *33* (12), 3151–3159. https://doi.org/10.1021/acs.langmuir.6b04570.

(23) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J Mol Graph* **1996**, *14* (1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

(24) Harada, R.; Kitao, A. Parallel Cascade Selection Molecular Dynamics (PaCS-MD) to Generate Conformational Transition Pathway. *Journal of Chemical Physics* **2013**, *139* (3). https://doi.org/10.1063/1.4813023.

(25) Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science. *J Am Chem Soc* **2018**, *140* (7), 2386–2396. https://doi.org/10.1021/jacs.7b12191.

(26) *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*; Bowman, G. R., Pande, V. S., Noé, F., Eds.; Advances in Experimental Medicine and Biology; Springer Netherlands: Dordrecht, 2014; Vol. 797. https://doi.org/10.1007/978-94-007-7606-7.

(27) Noé, F.; Rosta, E. Markov Models of Molecular Kinetics. *J Chem Phys* **2019**, *151* (19), 190401. https://doi.org/10.1063/1.5134029.

(28) Harada, R.; Kitao, A. Nontargeted Parallel Cascade Selection Molecular Dynamics for Enhancing the Conformational Sampling of Proteins. *J Chem Theory Comput* **2015**, *11* (11), 5493–5502. https://doi.org/10.1021/acs.jctc.5b00723.

(29) Inoue, Y.; Ogawa, Y.; Kinoshita, M.; Terahara, N.; Shimada, M.; Kodera, N.; Ando, T.; Namba, K.; Kitao, A.; Imada, K.; Minamino, T. Structural Insights into the Substrate Specificity Switch Mechanism of the Type III Protein Export Apparatus. *Structure* **2019**, *27* (6), 965-976.e6. https://doi.org/10.1016/j.str.2019.03.017.

(30) Takaba, K.; Tran, D. P.; Kitao, A. Edge Expansion Parallel Cascade Selection Molecular Dynamics Simulation for Investigating Large-Amplitude Collective Motions of Proteins. *J Chem Phys* **2020**, *152* (22), 225101. https://doi.org/10.1063/5.0004654.

(31) Tran, D. P.; Taira, Y.; Ogawa, T.; Misu, R.; Miyazawa, Y.; Kitao, A. Inhibition of the Hexamerization of SARS-CoV-2 Endoribonuclease and Modeling of RNA Structures Bound to the Hexamer. *Sci Rep* **2022**, *12* (1). https://doi.org/10.1038/s41598-022-07792-2.

(32) Tran, D. P.; Takemura, K.; Kuwata, K.; Kitao, A. Protein-Ligand Dissociation Simulated by Parallel Cascade Selection Molecular Dynamics. *J Chem Theory Comput* **2018**, *14* (1), 404–417. https://doi.org/10.1021/acs.jctc.7b00504.

(33) Hata, H.; Tran, D. P.; Sobeh, M. M.; Kitao, A.; Sobeh, M. M. Binding Free Energy of Protein/Ligand Complexes Calculated Using Dissociation Parallel Cascade Selection Molecular Dynamics and Markov State Model. *Biophys Physicobiol* **2021**, *18*, 305–316. https://doi.org/10.2142/biophysico.bppb-v18.037.

(34) Tran, D. P.; Kitao, A. Dissociation Process of a MDM2/P53 Complex Investigated by Parallel Cascade Selection Molecular Dynamics and the Markov State Model. *Journal of Physical Chemistry B* **2019**, *123* (11), 2469–2478. https://doi.org/10.1021/acs.jpcb.8b10309.

(35) Hata, H.; Nishihara, Y.; Nishiyama, M.; Sowa, Y.; Kawagishi, I.; Kitao, A. High Pressure Inhibits Signaling Protein Binding to the Flagellar Motor and Bacterial Chemotaxis through Enhanced Hydration. *Sci Rep* **2020**, *10* (1). https://doi.org/10.1038/s41598-020-59172-3.

(36) Tran, D. P.; Kitao, A. Kinetic Selection and Relaxation of the Intrinsically Disordered Region of a Protein upon Binding. *J Chem Theory Comput* **2020**, *16* (4), 2835–2845. https://doi.org/10.1021/acs.jctc.9b01203.

(37) Sobeh, M. M.; Kitao, A. Dissociation Pathways of the P53 DNA Binding Domain from DNA and Critical Roles of Key Residues Elucidated by DPaCS-MD/MSM. *J Chem Inf Model* **2022**, *62* (5), 1294–1307. https://doi.org/10.1021/acs.jcim.1c01508.

(38) Dickson, C. J.; Madej, B. D.; Skjevik, Å. A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C. Lipid14: The Amber Lipid Force Field. *J Chem Theory Comput* **2014**, *10* (2), 865–879. https://doi.org/10.1021/ct4010307.

(39) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratch, D. J. Gaussian 16, Revision B.01. Gaussian, Inc., Wallingford CT 2016.

(40) Case, D. A.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; III; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Kasavajhala, K.; Kovalenko, A.; Krasny, R.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Man, V.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wilson, L.; Wolf, R. M.; Wu, X.; Xiong, Y.; Xue, Y.; York, D. M.; Kollman, P. A. Amber 2020. University of California: San Francisco 2020.

(41) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *Journal of Physical Chemistry* **1993**, *97* (40), 10269–10280. https://doi.org/10.1021/j100142a004.

(42) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J Mol Graph Model* **2006**, *25* (2), 247–260. https://doi.org/10.1016/j.jmgm.2005.12.005.

(43) CRC Handbook. *CRC Handbook of Chemistry and Physics, 88th Edition*, 88th ed.; CRC Press, 2007.

(44) Kishore, K.; Shobha, H. K.; Mattamal, G. J. Structural Effects on the Vaporization of High Molecular Weight Esters. *J Phys Chem* **1990**, *94* (4), 1642–1648. https://doi.org/10.1021/j100367a077.

(45) Izadi, S.; Anandakrishnan, R.; Onufriev, A. v. Building Water Models: A Different Approach. *Journal of Physical Chemistry Letters* **2014**, *5* (21), 3863–3871. https://doi.org/10.1021/jz501780a.

(46) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. Ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J Chem Theory Comput* **2020**, *16* (1), 528–552. https://doi.org/10.1021/acs.jctc.9b00591.

(47) Gruber, C. C.; Pleiss, J. Lipase B from Candida Antarctica Binds to Hydrophobic Substrate-Water Interfaces via Hydrophobic Anchors Surrounding the Active Site Entrance. *J Mol Catal B Enzym* **2012**, *84*, 48–54. https://doi.org/10.1016/j.molcatb.2012.05.012.

(48) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J Comput Chem* **2009**, *30* (13), 2157–2164. https://doi.org/10.1002/jcc.21224.

(49) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

(50) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J Chem Theory Comput* **2015**, *11* (11), 5525–5542. https://doi.org/10.1021/acs.jctc.5b00743.

(51) Arthur, D.; Vassilvitskii, S. K-Means++: The Advantages of Careful Seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*; SODA '07; Society for Industrial and Applied Mathematics: USA, 2007; pp 1027–1035.

(52) Röblitz, S.; Weber, M. Fuzzy Spectral Clustering by PCCA+: Application to Markov State Models and Data Classification. *Adv Data Anal Classif* **2013**, *7* (2), 147–179. https://doi.org/10.1007/s11634-013-0134-6.

(53) Filipe, H. A. L.; Almeida, M. C. F.; Teixeira, R. R.; Esteves, M. I. M.; Henriques, C. A.; Antunes, F. E. Dancing with Oils - The Interaction of Lipases with Different Oil/Water Interfaces. *Soft Matter* **2021**, *17* (30), 7086–7098. https://doi.org/10.1039/d1sm00590a.

(54) Uppenberg, J.; Oehrner, N.; Norin, M.; Hult, K.; Kleywegt, G. J.; Patkar, S.; Waagen, V.; Anthonsen, T.; Jones, T. A. Crystallographic and Molecular-Modeling Studies of Lipase B from Candida Antarctica Reveal a Stereospecificity Pocket for Secondary Alcohols. *Biochemistry* **1995**, *34* (51), 16838–16851. https://doi.org/10.1021/bi00051a035.

(55) Fedosov, S. N.; Brask, J.; Pedersen, A. K.; Nordblad, M.; Woodley, J. M.; Xu, X. Kinetic Model of Biodiesel Production Using Immobilized Lipase Candida Antarctica Lipase B. *J Mol Catal B Enzym* **2013**, *85–86*, 156–168. https://doi.org/10.1016/j.molcatb.2012.09.011.