

Determination of Multidirectional Pathways for Ligand Release from the Receptor: A New Approach Based on Differential Evolution

Hoang Linh Nguyen, Nguyen Quoc Thai, and Mai Suan Li*



Cite This: <https://doi.org/10.1021/acs.jctc.1c01158>



Read Online

ACCESS |



Metrics & More

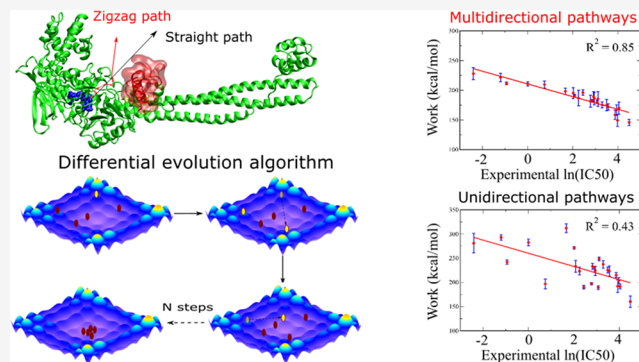


Article Recommendations



Supporting Information

ABSTRACT: Steered molecular dynamics (SMD) simulation is a powerful method in computer-aided drug design as it can be used to access the relative binding affinity with high precision but with low computational cost. The success of SMD depends on the choice of the direction along which the ligand is pulled from the receptor-binding site. In most simulations, the unidirectional pathway was used, but in some cases, this choice resulted in the ligand colliding with the complex surface of the exit tunnel. To overcome this difficulty, several variants of SMD with multidirectional pulling have been proposed, but they are not completely devoid of disadvantages. Here, we have proposed to determine the direction of pulling with a simple scoring function that minimizes the receptor–ligand interaction, and an optimization algorithm called differential evolution is used for energy minimization. The effectiveness of our protocol was demonstrated by finding expulsion pathways of Huperzine A and camphor from the binding site of Torpedo California acetylcholinesterase and P450cam proteins, respectively, and comparing them with the previous results obtained using memetic sampling and random acceleration molecular dynamics. In addition, by applying this protocol to a set of ligands bound with LSD1 (lysine specific demethylase 1), we obtained a much higher correlation between the work of pulling force and experimental data on the inhibition constant IC₅₀ compared to that obtained using the unidirectional approach based on minimal steric hindrance.



INTRODUCTION

Determining the binding affinity of a ligand to a receptor is an important problem in drug development. Various computational methods are used to solve this problem, starting with molecular docking,¹ which is fast but not accurate enough due to insufficient sampling. However, due to its high computation speed, it is widely used for the initial screening of potential compounds from large databases. More precise methods are based on molecular dynamics (MD) modeling, including free energy perturbation, thermodynamics integration (TI), molecular mechanics Poisson–Boltzmann surface area (MM-PBSA), molecular mechanics generalized Born surface area (MM-GBSA) methods,² and so forth.

Beside binding affinity, binding and unbinding kinetics and ligand pathways provide important information about drug efficacy. Many methods have been developed to investigate the association and dissociation of the ligand–protein complex.^{3–16} Using concurrent adaptive sampling, which is a weighted ensemble approach, Ahn et al. ranked ligand-binding kinetics for target β -cyclodextrin.¹⁷ Wolf et al. studied the ligand dissociation dynamics using targeted molecular dynamics (TMD) simulations with correction for dissipation.^{18,19} Metadynamics is another approach to access kinetic profiles,^{20–22} while You and Chang combined several methods to study the dissociation process.²³ In particular, funnel

metadynamics²⁴ can produce not only binding free energy but also unbinding pathways. Furthermore, machine learning emerges as a promising approach to enhance sampling related to ligand dissociation.²⁵

The steered molecular dynamics (SMD)^{26–29} was first implemented to probe the molecular mechanisms of biomolecular processes studied using single-molecule spectroscopy techniques such as atomic force microscopy (AFM) and laser optical tweezers.^{30,31} Later, SMD was employed to investigate the protein–ligand interaction by various groups.^{32–38} Today, various modifications such as hybrid SMD,³⁹ ensemble-based SMD,⁴⁰ DelPhiForce SMD,⁴¹ and so forth have been developed. The influence of the pulling velocity and the number of simulation trajectories on the prediction of protein–ligand affinity was examined.⁴²

Recently, it has been shown that SMD is as accurate as MM/PBSA⁴³ but fast enough to deal with a large number of

Received: November 16, 2021

ligands.^{34–36,39,43,44} It should be noted that if SMD is combined with Jarzynski equality⁴⁵ to calculate the absolute equilibrium free energy, then SMD takes more time than MM/PBSA. However, when SMD is used at a high pulling speed to obtain the relative binding affinity, by comparing the pulling work, SMD is observed to be faster.^{34–36,39,43,44}

SMD results depend on the pulling direction, which can be either unidirectional or multidirectional. A unidirectional pulling path can be obtained using different methods such as CAVER,⁴⁶ MOLE,⁴⁷ and MSH (minimal steric hindrance),⁴⁸ which yield the agreement between the simulation and experiment for many systems^{32,34,49–51} (note that MSH is based on SMD simulations, while CAVER and MOLE identify paths using a static structure). However, as shown below, even MSH will lead to a low correlation between the pulling work, which can be used to distinguish binders from nonbinders,⁴⁸ and the experimental IC50 values of 24 compounds bound to the anti-cancer therapy target LSD1 (lysine specific demethylase 1).⁵² One of the possible reasons for SMD failure in this case is that LSD1 has a narrow exit channel that is not suitable for pulling the ligand along a straight pathway, which prompts us to try multidirectional pulling.

There are several methods to find the best zigzag direction. In the random acceleration MD (RAMD) method,⁵³ a force of random direction is applied on the ligand, which may or may not be updated according to the ligand displacement in a certain period of time. This process continues until the ligand is released from the receptor. However, due to the random choice of the force orientation, the success rate in finding egress routes may be low compared to that in other methods.⁵⁴ Here, the success rate is defined as the percentage of trajectories leading to a ligand exit. A recent development of RAMD is τ_{RAMD} ,^{55,56} which only requires the magnitude of the random force as a parameter for ranking the residence times of diverse sets of compounds. Another method for finding the ligand binding pathway is the string method, which interpolates intermediate states between bound and unbound states.^{57,58} Combining RAMD, SMD, and umbrella sampling, it has been shown that very long simulations are needed in order to obtain an accurate ranking of the distinct access and egress routes and the free energy profiles.^{59,60}

Rydzewski et al. proposed two memetic algorithms called memory random acceleration (MERA) and the immune algorithm (IA) to find the optimal ligand egress pathway.^{54,61} In MERA, each position of the ligand leaves a trace for the next attempt. The ligand is pulled by a random force, which depends on the concentration of the traces. This algorithm is equivalent to the ant colony optimization algorithm. The MERA method requires an initial trace distribution before performing MERA and MD simulation. This distribution can be obtained by using RAMD or locally enhanced sampling.⁶¹ In the IA, the objective function, which is the interaction energy between the ligand and the protein, is obtained using the Hammett linear free energy function.⁵⁴ This algorithm searches for a path that has the minimum interaction energy at each pulling step, and the pulling direction coincides with this path. MERA and the IA have a higher success rate of pulling out the ligand than RAMD.⁵⁴ The IA has been successfully used to find egress pathways of camphor from cytochrome P450cam.⁶² Rydzewski and Valsson⁶⁴ recently proposed a method for finding pathways for ligand unbinding using convex optimization of a function that describes the protein–ligand interaction.

To solve this problem, we can find a multidirectional pathway using one of the methods mentioned above^{53,54,61,63,64} because popular RAMD is built into major software packages such as NAMD,⁶⁵ AMBER,^{66,67} and GROMACS,⁵⁶ while the code for the recent work of Rydzewski et al.⁶⁴ (maze) is available on PLUMED.⁶⁸ However, since SMD is a valuable tool in predicting the relative binding of a ligand, it would be interesting to use this method to find multidirectional pathways for ligand egress from the protein active site. Such a SMD-based approach has been developed by several groups.^{69,70} Yang et al.⁶⁹ combined on-the-fly SMD with a multi-population genetic algorithm to find minimal energy paths. Gu et al.⁷⁰ built the objective function based on the criterion that the best path provides the least rupture force. Here, the multi-population genetic algorithm was also used to solve the optimization problem. In this approach, the initial direction is chosen manually, and its population may be small, affecting the efficiency of the protocol.

Since zigzag dissociation channels have not been studied in detail,^{69,70} it is unclear if the on-the-fly SMD approach can reveal all pathways in complex systems such as the P450cam-camphor complex. In addition, the code related to the adaptive SMD^{70,71} is not available online and is not easy to implement, prompting us to develop our own protocol based on differential evolution (DE)⁷¹ to determine a multidirectional pathway.

Our tool interacts with the GROMACS package through a bash shell script. We showed that our in-house program provides pathways that have been observed using previous methods^{53,62,72} for the dissociation of Huperzine A (HupA) from Torpedo californica acetylcholinesterase (TcAChE) and camphor from P450cam. This indicates the reliability of our protocol in detecting zigzag pathways for ligand exit from the receptor-binding site. Moreover, our protocol significantly improved the correlation between non-equilibrium work and experimental IC50 data for LSD1, suggesting that it can be applied to study the binding affinity of a complex with a narrow channel.

MATERIALS AND METHODS

Set of 24 Ligands Bound to LSD1. We will study the binding affinity of 24 ligands to LSD1, the IC50 of which was experimentally measured (Table S1 in the [Supporting Information](#)). Note that the IC50 is not directly related to binding affinity but can be used to measure it since the IC50 is proportional to the dissociation constant K_i through the Cheng–Prusoff equation. The initial structure of LSD1 was obtained from the protein data bank (PDB) with PDB code 2UXN⁷³ (Figure 1). The structures of 24 ligands were retrieved from PubChem⁷⁴ and ChemSpider⁷⁵ (Table S1).

Structures of P450cam-Camphor and TcAChE–HupA Complexes. To test our new protocol, we studied two cases: HupA binds to Acetylcholinesterase (AChE) and camphor binds to P450cam. In the first case, the structure of Torpedo California AChE (TcAChE) was obtained from the protein data bank (PDB) with PDB id 1AES,⁷⁶ while for the heme–camphor complex, we used the PDB structure with id 2CPP.⁷⁷ The parameters for heme were collected from Shahrokh et al.⁷⁸

Molecular Docking. AutoDock Tools 1.5.4⁷⁹ was used to prepare the input for docking simulation with the PDBQT format. The docking simulation is carried out using AutoDock Vina 1.5.4,⁸⁰ where the ligand structure was flexible. The docking simulation box was centered at the binding site, and its

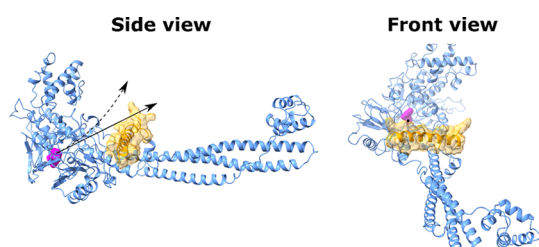


Figure 1. Schematic for the initial complex structure of LSD1 and the ligand (magenta). The pulling direction is shown as a black arrow in the side view and a black dot in the front view, which represents the direction points to the reader. The obstacle regions are shown in yellow with a transparent surface. The dotted black arrow suggests a pathway that ligands can avoid clash with red regions.

grid size and grid center for all targets are shown in Table S2. To access good quality of the docking, the completeness parameter for global search was set to 600. The docking structure obtained in the lowest binding energy mode of the receptor–ligand complex was used as the initial configuration for SMD simulation.

MD Simulations. MD simulations were carried out using the GROMACS 2018 package.⁸¹ The complex structure was solvated using the TIP3P water model⁸² in a rectangular box with a minimum distance of 2 nm between the protein and the box edges. Counter ions were added to neutralize the charge of the system. We used the AMBER99SB-ILDN⁸³ force field to parameterize the protein.

The parameters of the ligand that binds to LSD1 were computed using Antechamber⁸⁴ and AcPype,⁸⁵ which are based on the general AMBER force field.⁸⁶ Atomic point charges were determined using AM1-BCC.⁸⁷ The system was relaxed using the steepest descent algorithm⁸⁸ and was then equilibrated for 500 ps in an NVT ensemble at 300 K reserved using the ν -rescale algorithm.⁸⁹ It was then held at 1 bar using the Parrinello-Rahman algorithm for 5 ns in an NPT ensemble,⁹⁰ and to avoid an abrupt change in the structure, the heavy atoms of the protein and ligand were restrained by a harmonic potential with a spring constant of 1000 kJ/mol/nm.² Then, to achieve a properly equilibrated structure of the ligand, an additional 100 ns NPT simulation was performed without restriction. The cutoff for nonbonded interactions was 1.0 nm.

SMD Simulation. In SMD, an external force is applied to a dummy atom that is linked to the ligand atom closest to the CoM of the ligand using a spring with a stiffness k . Then, the force experienced by the ligand is $F = k(\Delta x - vt)$, where v is the pulling speed and Δx is the pulled atom displacement from the initial position. We chose $k = 600$ kJ/mol·nm², which is the typical value used in the AFM experiment.⁹¹

From the force–time/position profile, we collect the rupture force F_{\max} which is necessary for the dissociation of the protein–ligand complex and can be used to rank binding affinity. The pulling work W_{pull} , which is another metric for comparing binding affinity, was calculated using the following formula

$$W_{\text{pull}} = \int F(x)dx = \frac{1}{2} \sum_{i=1}^{n-1} (F_{i+1} + F_i)(x_{i+1} - x_i) \quad (1)$$

where F_i and x_i are the force and displacement at step i , respectively, and n is the number of simulation steps. Since the work⁴⁸ and rupture force^{72,92} depend on the pulling speed but

the correlation between the simulation data, obtained in a highly non-equilibrium regime, and the experimental data does not depend on it,⁴⁸ we used $v = 5$ nm/ns. For each protein–ligand complex, we performed 20 independent SMD trajectories in both unidirectional and multidirectional simulations, and W_{pull} was averaged over all trajectories. The number of trajectories was chosen so that the error bar, presented as the standard deviation, was less than or equal to 10% of the mean.

Choice of the Pulling Pathway. We performed SMD simulations with unidirectional and multidirectional pulling pathways. In the unidirectional case, the ligand was pulled in the direction determined using the MSH algorithm.⁴⁸ Such a pathway is shown in Figure 1 for LSD1 (black arrow). For each receptor–ligand complex, 20 SMD trajectories of 2 ns each were conducted.

The multidirectional pathway was found using our new protocol, which is described in the Results and Discussion section. In this case, the pulling direction was changed every 20 ps, and the rationale for this choice is also shown in the Results and Discussion section.

Pearson Correlation Coefficient. The Pearson correlation coefficient between the pulling work W and experimental IC50 values was calculated as follows

$$R = \frac{\langle W - \langle W \rangle \rangle \langle \text{IC50} - \langle \text{IC50} \rangle \rangle}{\sigma_W \sigma_{\text{IC50}}} \quad (2)$$

where X and σ_X represent the average and standard deviation of quantity X , respectively.

RESULTS AND DISCUSSION

Unidirectional SMD Simulation Provides a Poor Correlation with Experimental IC50 of LSD1–Ligand Complexes. Our docking simulation showed that all 24 ligands have the same binding site in LSD1 (Figure S1). Consequently, the pulling directions obtained using the MSH method are almost the same since the exit tunnel is narrow.

Previous studies^{49,93} have shown that pulling work W_{pull} obtained from SMD simulations correlates better with experimental IC50 values than the rupture force (maximum force in the force–extension/time profile). This is due to the fact that work is defined for the whole process, while the maximum force is achieved in one state. Therefore, in this study, we calculate work profiles for a set of 24 ligands bound with LSD1 (Figure S2), and W_{pull} is defined as work at the last point. As mentioned above, IC50 is not directly related to the binding free energy ΔG , but it is reasonable to assume that $\Delta G \sim \ln(\text{IC50})$. Thus, we showed the correlation between W_{pull} and $\ln(\text{IC50})$ (Figure 2) and obtained a moderate correlation level⁹⁴ with the coefficient $R^2 = 0.43$. To understand why the correlation between the SMD simulation and experiment is not high, we examined force–time profiles. For trajectories 2, 5, 9, and 10 of ligand 1, a second peak occurs after the major maximum (Figure S3), indicating a possible collision between the ligand and the receptor during its egress. A similar situation also occurred with other ligands (data not shown), which implies that a straight direction obtained by using the MSH protocol does not work well for the narrow exit tunnel, and this may result in a low correlation between the simulation and experiment. Furthermore, since MSH has been shown to have better performance⁴⁸ than other similar methods,^{46,47} the existing unidirectional approaches cannot improve the

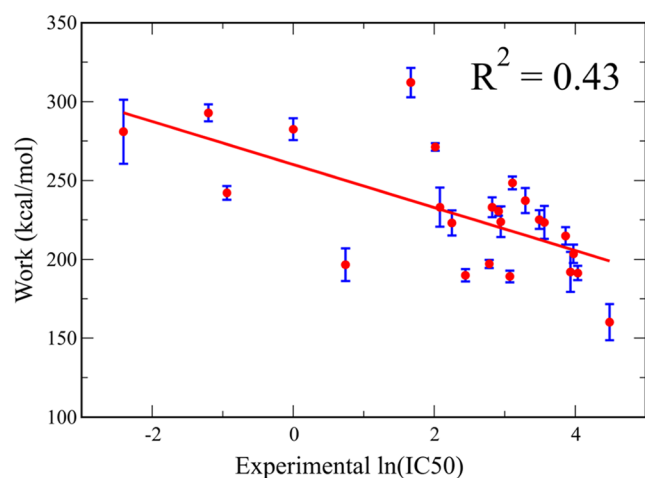


Figure 2. Pulling work obtained from straight pulling direction SMD simulations as a function of experimental $\ln(\text{IC}_{50})$ with IC_{50} measured in moles per liter. The red line represents the fitting function of data. R^2 indicates the correlation coefficient. The results were averaged over 20 SMD trajectories, and error bars represent standard deviations.

correlation with the experiment predicted using MSH ($R^2 = 0.43$).

The pulling directions obtained using the MSH algorithm are vectors that point from regions E308–A309 of chain A to regions G314–R316, A809–V811 of chain A and are finally directed to a location that is close to regions R308–G313 of chain B and K371–F396 of chain A (Figure S4).

Although the external force does not direct the ligand to pass through the R308–G313 and K371–F396 regions of LSD1, the side chains of these regions can interfere with the ligand movement due to their proximity to the pulling path. This can be seen by the strong interaction between the ligand and these areas (Figure S3), the minima of which are close to the second peak of the force experienced by the ligand [see Figure S3 where the distance between the center of mass (CoM) of the ligand and the CoM of regions R308–G313 and K371–F396 of LSD1 is shown]. The average of minimum distance of all ligands is about $13.9 \pm 1.0 \text{ \AA}$, which indicates that the ligands approach closer to these regions during SMD simulation.

Since the collision problem in the R308–G313 and K371–F396 regions arises when a unidirectional pathway is used, we expect this problem to be avoided if the ligand moves along the orange path (Figure 1). However, in unidirectional pulling, if these regions can be avoided, the ligand may interfere with other protein residues on the surface of the binding pocket. Therefore, a different approach is required to find a pathway that retains the ligand both from the binding pocket surface and from the R308–G313 and K371–F396 regions.

Adaptively Changing Direction in SMD Simulations.

The basic idea of our method is that to get the pulling direction on the fly in SMD simulations, we split the simulations into N short intervals (Figure 3A). Below, we show that the time interval $\tau = 20 \text{ ps}$ is optimal in our method, when SMD was carried out at a pulling speed of 5 nm/ns . For each interval, the so-called DE⁷¹ was used to identify the optimal unidirectional path. Once such a path was found, we used SMD to transfer the ligand from point i to point $i + 1$. This process is repeated N times until the ligand reaches an unbound state in solution (Figure 3A). N is equal to the simulation time τ_{sim} divided by the time interval $\tau = 20 \text{ ps}$ ($N =$

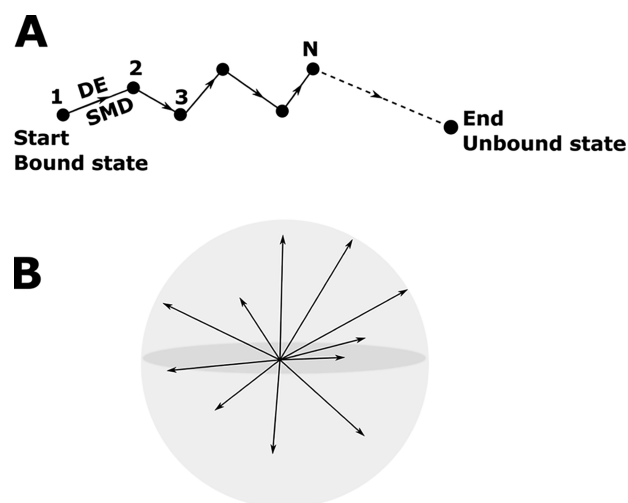


Figure 3. (A) Schematic description of our method, which involves N intervals to transition from a bound state in a binding site to an unbound state. One interval corresponds to the time interval $\tau = 20 \text{ ps}$ of SMD simulation at a pulling speed of 5 nm/ns . For each step, we first applied DE to identify the short straight path, and once that path was found, SMD was performed to pull the ligand from point i to point $i + 1$. This procedure was repeated until the ligand reached the unbound state. (B) Representation of 3000 vectors randomly generated in DE. The maximum vector length is 7 \AA .

τ_{sim}/τ) and depends on the system (Table S3) in the range from 50 (LSD1) to 150 (TcAChE). To find the exit paths, we will execute 20 trajectories for LSD1 and TcAChE and 60 trajectories for P450cam (Table S3). The duration of each trajectory is usually shorter than the cutoff time τ_{sim} , resulting in practical steps less than the N indicated in Table S3 (see below).

To find an optimal path by using DE, we define the scoring E_{score} as follows

$$E_{\text{score}} = E_{\text{elec}}^{\text{current}} + E_{\text{vdw}}^{\text{current}} - E_{\text{elec}}^{\text{previous}} - E_{\text{vdw}}^{\text{previous}} + \alpha |E_{\text{elec}}^{\text{current}} + E_{\text{vdw}}^{\text{current}}| \quad (3)$$

where E_{elec} and E_{vdw} are the electrostatic and van der Waals interaction energies between the ligand and receptor, respectively. The upper “current” refers to the probe configuration that is generated from the random translation of the ligand, while the “previous” refers to the configuration obtained at the end of a short step in the SMD simulation. At each short step, we calculate the minimum of E_{score} , which is the local energy.

Why do we need an $\alpha |E_{\text{elec}}^{\text{current}} + E_{\text{vdw}}^{\text{current}}|$ term if we can assume that $E_{\text{score}} = E_{\text{elec}}^{\text{current}} + E_{\text{vdw}}^{\text{current}} - E_{\text{elec}}^{\text{previous}} - E_{\text{vdw}}^{\text{previous}}$? Imagine that we get into a global minimum or some state with very low energy, then the searching procedure would stop, but adding this term helps avoid the trap. Term $\alpha |E_{\text{elec}}^{\text{current}} + E_{\text{vdw}}^{\text{current}}|$ can also prevent the ligand and receptor overlap.

To settle the coefficient α , we carried out, for example, 20 SMD trajectories for the TcAChE–HupA complex with $\alpha = 0, 1, 2, 3,$ and 4 . The numbers of the obtained paths are 10, 4, 6, 4, and 2 for $\alpha = 0, 1, 2, 3,$ and 4 , respectively, which shows that α values of 0 and 2 give more paths than other values of α . However, the camphor ligand gets stuck at several positions along the SMD trajectory for the P450cam–camphor complex with $\alpha = 0$. Thus, we chose $\alpha = 2$ to balance the ability to

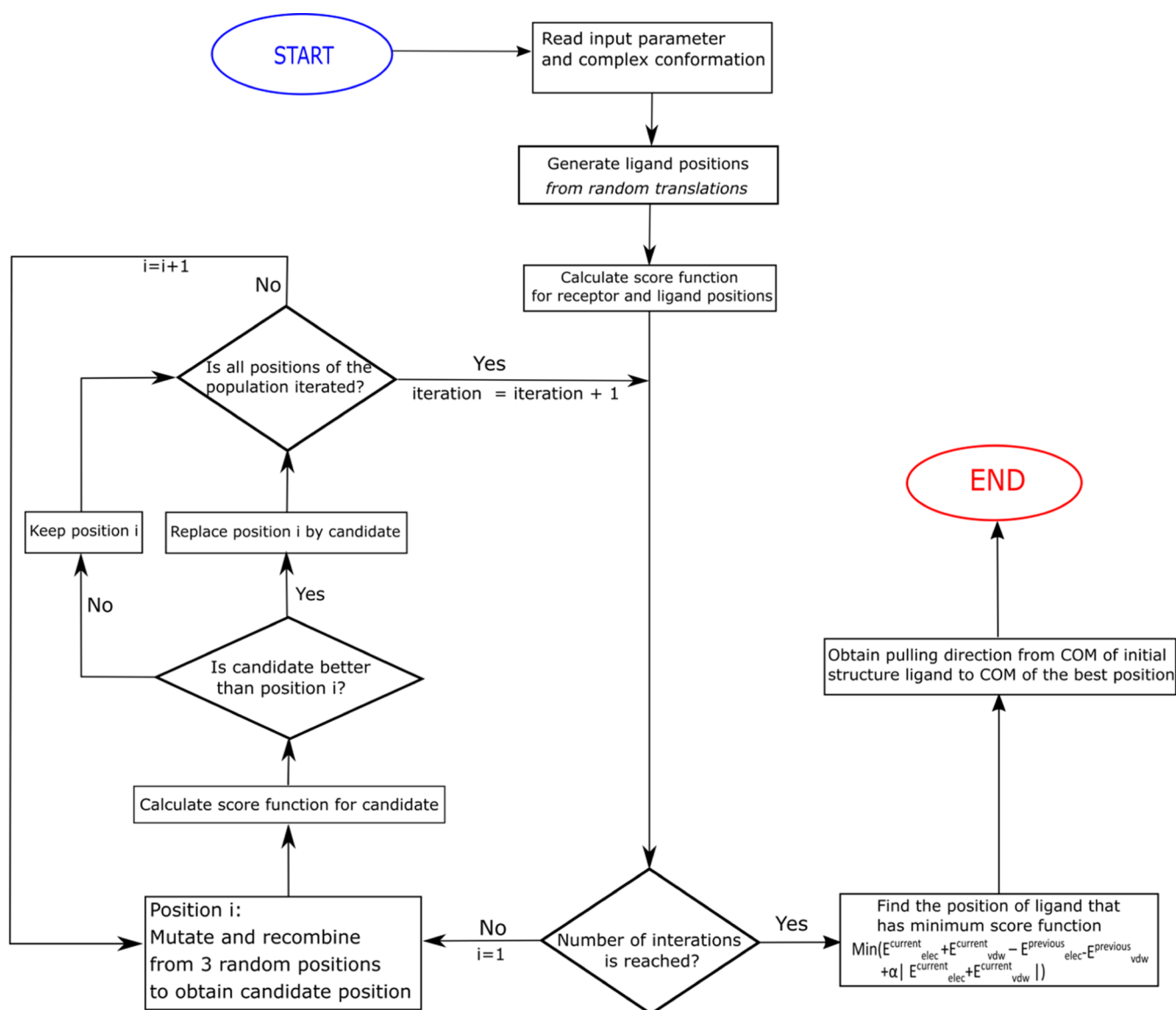


Figure 4. Main steps in the novel protocol for obtaining the optimal zigzag pulling pathway. COM refers to the center of mass of the ligand.

capture rare paths and prevent ligand stuck. Force field parameters of atoms are the same as those in conventional MD simulations.

In our method, the direction in each step is determined corresponding to the minimum E_{score} . From this point of view, our method differs from RAMD,⁵³ where the direction is chosen at random and changes if the ligand encounters the cavity surface of the protein. In addition, the success rate of the RAMD depends on the constant force and threshold velocity, which are not easy to select,⁶¹ but this disadvantage was addressed in the τ_{RAMD} protocol⁵⁵ by providing a rather straightforward procedure for determining the parameters of the random force for protein–ligand dissociation.

Our score function (eq 3) is also different from the other studies that use optimization algorithms. For instance, Gu et al. used information entropy to model a multi-objective problem,⁷⁰ and the initial pulling direction is not selected by these algorithms but manually to make sure that the ligand is pulled toward the solvent. In our algorithm, the objective function is simple to reduce the computational effort, and the program can automatically determine the initial direction.

Let us describe the DE⁷¹ algorithm in more detail. This algorithm is an optimization method used to find the minimum of an objective function that is a D-dimensional real-valued function. It directly searches the optimal solution from the objective function, which does not require gradient information. The main steps of DE are described in Figure 4. First, a set of ligand positions are generated by random translations (Figure 3B). For the first step shown in Figure 3A, the ligand conformations obtained at the end of the NPT simulation were used for the DE algorithm. However, for the remaining short intervals, we used the ligand conformations generated in the SMD simulation to perform the optimization using DE.

The number of positions and the magnitude of the translation vector can be set by the user, but we can show that 3000 ligand positions and the vector length of 7 Å are a reasonable choice. Figure S5 shows the effect of the number of conformations on the pulling direction of ligand 10 in LSD1. Our result indicates that the pulling direction is converged at 2000 conformations. Therefore, our choice of 3000 is safe. Figure S6 shows the dependence of the pulling direction obtained by using DE with 3000 ligand conformations versus

the vector length for ligand 10 interacting with LSD1. It is obvious that the pulling direction is converged at a length of 7 Å. Hence, this value guarantees locality and is not too small to allow the ligand to wobble around its original structure. If the vector length is too long, the direction may point through the protein into the outer space because the potential value at that position is small, and this is what we want to avoid.

The DE algorithm generates new candidate solutions by adding a weighted difference between the two candidates to the third candidate, and the new candidate is called a mutation. This mutation is mixed with the parameters (crossover) of another candidate (target candidate), resulting in a trial candidate. If the trial candidate has a lower value of objective function than the target candidate, the trial candidate replaces the target candidate in the population. This process is iterated until the termination criterion is met (Figure 4). To find out how many iterations are sufficient to obtain reasonable results, we examined the dependence of the pulling direction on the number of iteration steps for ligand 10 leaving LSD1 (Figure S7). After 25 and 50 iterations, the pulling direction does not converge. However, when the number of iterations is equal or greater than 100, the result converges. Therefore, we chose 100 as the number of iterations in this work to balance performance and reliability.

In short, based on the DE algorithm,⁷¹ the ligand positions are refined for the best score function value at each iteration (see eq 3 for the score function). In this work, the cross probability (CR) and differential weight (F) of the DE algorithm were chosen as 0.9 and 0.5, respectively, that is, we used the same set of parameters as those in the original paper of Storn and Price.⁷² After 100 iterations, the positions with the smallest score function are selected to find the next step of the pulling direction (Figure 3A). As soon as the ligand leaves the protein, the pulling direction is kept unchanged.

As mentioned above, for each small step (Figure 3A), after DE, we performed a 20 ps SMD simulation at a pulling speed $v = 5$ nm/ns, where the ligand was allowed to be flexible. To show that the choice of the time interval $\tau = 20$ ps is reasonable, we tested the time intervals $\tau = 10, 20,$ and 40 ps for SMD simulation for the HupA–TcAChE complex. A typical force–time profile for one trajectory is shown in Figure S8. We excluded $\tau = 40$ ps because there is a sharp change after passing the peak (see the black box at the bottom). Since the rupture forces are similar for $\tau = 10$ and 20 ps, we chose 20 ps to save simulation time. The user can use shorter intervals for smoother interval switching change between intervals.

Finally, for the convenience of the user, we summarize the parameters used in our method in Table 1. We recommend this set of parameters, but the user can try other sets as well.

Table 1. Parameters Used in Our Method

name	value
number of populations (random vectors) (see Figure 3B)	3000
maximum length of random vectors	7 Å
number of iterations in DE for each step (see Figure 3A)	100
pulling speed v	5 nm/ns
time interval τ	20 ps
α (see eq 3)	2
CR	0.9
differential weight (F)	0.5

Our Protocol Can Reproduce Two Previously Known Exit Paths for HupA from TcAChE. Our protocol was aimed at solving problems with pulling of ligands from LSD1, but we must first check its reliability for systems with known unbinding pathways. Previous computational studies^{72,95,96} have identified two dissociation pathways for HupA from the TcAChE binding site (Figure 5). The first pathway is called the front exit pwf, and the second is the transiently opened side exit pws.

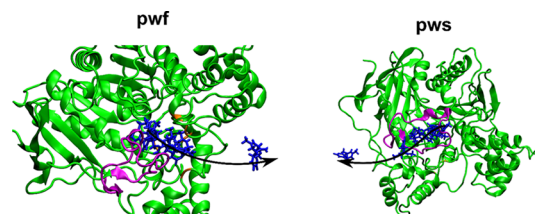


Figure 5. Two pathways of HupA to exit from the TcAChE binding site. The result obtained in this work coincides with the paths previously reported by other groups.^{72,95,96} Snapshots of the ligand are shown in blue licorice, while the Ω -loop is shown in purple.

Pwf is a tunnel from the CAS (catalytic anionic site) of TcAChE that involves residues Y70, V71, D72, Y121, W279, F290, F331, and Y334. There is no rearrangement of TcAChE residues, which is required for the HupA escape along this pathway. The second pathway pws from the CAS to the solvent is formed by a transient channel via the Ω -loop, which comprises residues 67–94 (Figure 5). The research by Tara et al. suggested that pws can play as a channel for water and ion molecules in the HupA-bound structure.⁹⁶ In the simulations carried out by Rydzewski et al.,⁷² HupA can leave through pws, but the Ω -loop opening is needed.⁷² Due to the presence of the transient channel that includes the Ω -loop opening, the ability to capture the two dissociation pathways of HupA can be considered a benchmark for our protocol.

Using our protocol, we have also obtained two pathways, which coincide with previously identified pwf and pws⁷² (Figure 5), which means that the protocol based on DE is reliable in determining dissociation pathways of the TcAChE–HupA complex.

In our simulations, HupA exits from the binding pocket via pwf and pws pathways in 14 and 6 SMD trajectories, respectively. This observation is understandable, and since there is a structure restructuring to open the Ω loop, it takes a while for pws to execute whereas pwf is already open, which forces HupA to favor pwf over pws. To further support this result, we show the time dependence of the root mean square deviation (RMSD) of the Ω -loop in two pathways (Figure S9). For pwf, we have a monotonic dependence, while for pws, a sudden change at about 1650 ps indicates the opening of the Ω -loop.

The average simulation times to determine the exit channels are about 1636 and 1989 ps for channels 1 and 2, respectively (Table S4), which is less than the cutoff value of 3000 ps (Table S3). This is because, as mentioned above, the search process usually ends before the maximum time is reached.

We calculated the nonbonded interaction energy between HupA and TcAChE for two pathways (Figure S9). Since the simulation is terminated when HupA is pulled out of TcAChE, but they are not far away, interactions are still present at the end of the simulation. In both cases, the interaction energy

changes sharply when HupA leaves TcAChE, but the change in the pwf case is softer, which is consistent with the Ω -loop opening in pws.

Since the difference in the interaction energies between the initial and final states in pws is smaller than in pwf (Figure S9), we anticipate that although pws requires the Ω -loop opening, the potential barrier is lower along this pathway. Therefore, opening of the Ω -loop lowers the potential barrier, indicating that both these pathways are feasible. In addition, we did not observe any evidence that HupA leaves TcAChE via the so-called backdoor pathway (pwb) that was reported for other ligands.⁹⁷ This result shows that pwb is not an HupA dissociation pathway, which is consistent with Rydzewski et al.⁷²

Our Protocol Can Identify Five Previously Known Exit Paths for Camphor from P450cam. It is already known that camphor can escape from the P450cam binding site via five major pathways,^{53,62} making the problem more difficult than in the HupA case.⁹⁸ Because the interior of the P450cam molecule is complex, we performed 60 independent SMD trajectories to determine the exit paths for camphor using our protocol. For comparison, we used the same structural nomenclature of P450cam as Poulos et al.⁷⁷ and Rydzewski and Nowak.⁶² P450cam comprises 13 helices labeled A, B, B', and C-L and five β sheets labeled 1–5 (Figure S10).

Camphor passes between helices I and C in the first path (PW1) (Figure 6), which depends on the flexibility of spiral C.

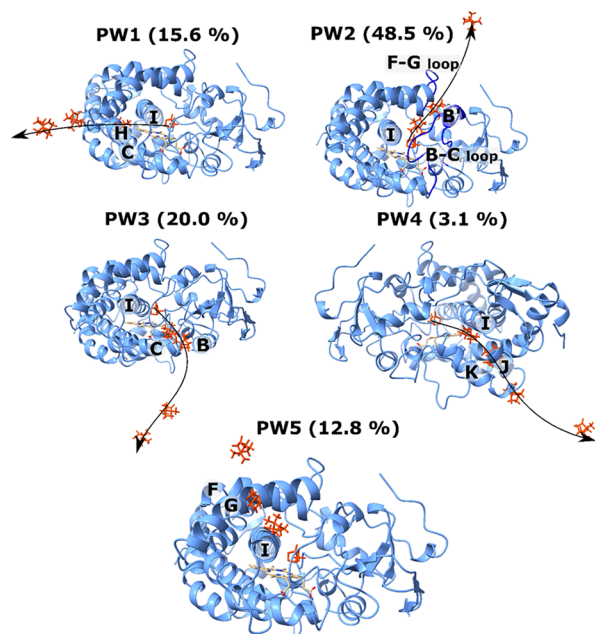


Figure 6. Schematics for dissociation pathways of camphor from P450cam. The arrows represent pathways obtained from this work. Several snapshots of ligands are shown in orange licorice. HEM group is shown in licorice. The number in brackets represents the population of the pathway obtained from bootstrap analysis.

This path has been discovered from previous simulations using RAMD and thermal motion pathway methods,^{53,99,100} as well as using the machine learning-based dimensionality reduction method.⁶²

The second pathway (PW2) starts from the L244-G248 residues of helix I and passes between the B' helix (residues E91–A92) and the F–G loop (R186–F193) (Figure

6).^{62,99,101} It has been suggested to subclass this path, as in the RAMD simulations,⁵³ but Rydzewski and Nowak⁶² found only one class that is the most likely.

The third pathway (PW3) includes helix I and B–C loop and passes between helices B and C including residues Y75–H80 and D104–P106, which is believed to be the channel for 5-hydroxyl-camphor.^{77,102} This pathway was also found by Rydzewski and Nowak.⁶² Previous research studies suggest that this pathway may play a role in the electron transport or water network for proton transport.^{99,103} The fourth pathway (PW4) was identified in a study by Rydzewski and Nowak,⁶³ which is adjacent to helix I and passes helices J and K (E273–R280). This pathway is not observed in previous RAMD studies. The fifth pathway (PW5) was found in the RAMD simulation,⁵³ which is called pathway class 3. However, this pathway was not observed by Rydzewski and Nowak.⁶² The ligand traverses helix I at residues L244–G248 and then crosses the space between helices F (I174–D182) and G (A199–L204) (Figure 6).

Using our protocol and 60 SMD trajectories, we obtained five major egress pathways of camphor, of which the first four pathways are the same as those indicated by Rydzewski and Nowak⁶³ (Figure 6), while PW5 is similar to pathway class 3 in RAMD simulations.⁵³ PW1–PW5 occurred in 11, 31, 9, 2, and 7 trajectories, respectively. By performing a bootstrap analysis with the size of the resampled data set of 20,000, we can show that the average populations are 15.6 ± 3.4 , 48.5 ± 5.1 , 20.0 ± 4.1 , 3.1 ± 0.8 , and 12.8 ± 3.7 for paths 1, 2, 3, 4, and 5, respectively. Thus, in line with the previous work, we obtained the most probable PW2 with a population of 48.5% but not the subclasses pathways identified in RAMD simulations.⁵³

PW4 has the lowest probability ($2/60 \approx 3\%$), which is consistent with Rydzewski and Nowak. PW5 ($\approx 12\%$) is more likely than PW4, and this pathway was observed in RAMD simulations⁵³ but not in Rydzewski and Nowak.⁶² In RAMD simulations, PW5 has the least population, which agrees with our result.

From the time dependence of the nonbonded interaction energy (Figure 7), it is clear that the ligand takes the shortest time to escape from P450cam along PW2 (see also Table S5), which is consistent with the fact that PW2 is the most probable route. In addition, W_{pull} is the smallest in PW2 (Table 2), which also confirms this observation. However, W_{pull} in PW1 is

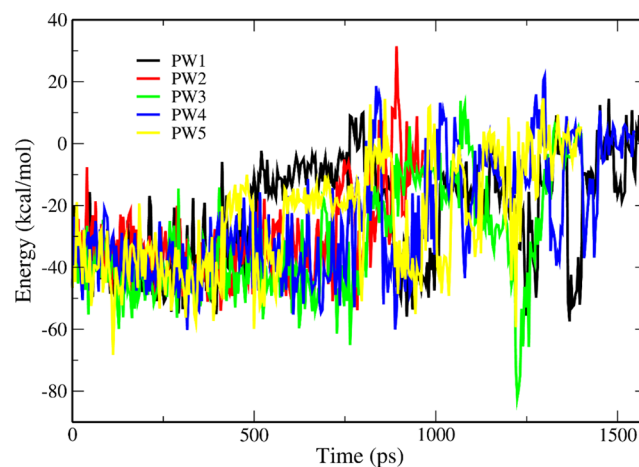


Figure 7. Time dependence of the energy of interaction between P450cam and camphor along five pathways.

Table 2. Pulling Work (kcal/mol) of Camphor Obtained in Five Unbinding Pathways^a

PW1	PW2	PW3	PW4	PW5
254.47 ± 26.41	128.97 ± 5.71	169.06 ± 14.24	389.42 ± 46.77	152.51 ± 13.99

^aThe error represents the standard deviation.

larger than in PW3 and PW5, which contradicts the fact that the population of PW1 is higher than that of PW3 and PW5. This may be due to the fact that the number of SMD runs is not enough to get a reliable result. Overall, our protocol is robust because it can predict the same pathways as those in previous studies for both P450cam-camphor and TcAChE–HupA complexes.

Zigzag Pulling Improves the Performance of SMD Simulation for LSD1. As mentioned above, in the case of LSD1, unidirectional pulling led to a second peak in the force–time/extension profile. This artifact arises from the collision of the ligand with the surface of the narrow exit tunnel, which affects the rupture force and pulling work, resulting in a low correlation between the theory and experiment. To solve this problem, instead of pulling with a constant direction, we took a multidirectional approach. Since our DE-based protocol works well for the HupA–TcAChE and P450cam-camphor systems, we used it to study the binding affinity of 24 ligands for LSD1.

In SMD simulations, the pulling direction is determined on the fly, and the initial conformations as well as other setup parameters are the same as those in the straight pulling. The multidirectional pathway avoids the region where the collision between the ligand and the receptor occurs in unidirectional stretching (Figure 8). Figure S11 shows typical force–time

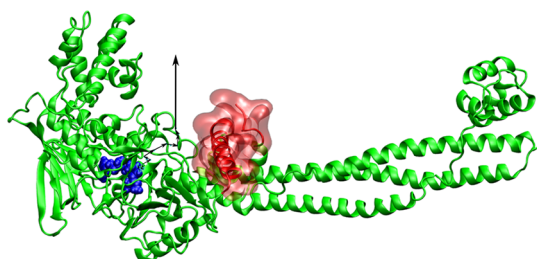


Figure 8. Schematic for the zigzag pulling direction (black arrow) of the ligand (blue) from LSD1. The obstacle region is shown in red with a transparent surface.

profiles, obtained with straight and zigzag pulling. Clearly, the second peak is much more pronounced in the unidirectional case, which implies that the multidirectional path avoids collision between the ligand and the receptor. In accordance with this observation, W_{pull} performed by the ligand during the exit along the zigzag path is lower than in the forward case (Figure 9). Correlation between work values obtained using two pulling methods is $R^2 = 0.45$.

The correlation between W_{pull} and experimental $\ln(\text{IC}_{50})$ is $R^2 = 0.85$ (Figure 10), which is better than $R^2 = 0.43$, obtained in SMD simulations using a constant pulling direction. The improved correlation between the theory and experiment proves the superiority of the new protocol.

It should be noted that the correlation level $R^2 = 0.85$ (Figure 10) was obtained at $v = 5$ nm/ns. In order to show that the result is independent of the pulling speed, we ran simulations at $v = 1$ nm/ns and obtained $R^2 = 0.86$ (Figure S12), which is very close to 0.85. Thus, as in previous

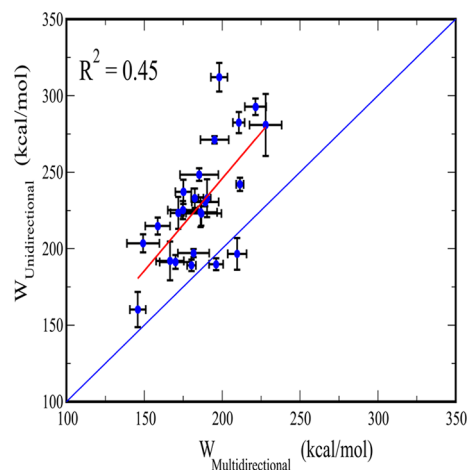


Figure 9. Pulling work obtained using SMD simulations with unidirectional and multidirectional pulling for 24 ligands bound to LSD1. The red line is a linear fit of the two data sets with a correlation level $R^2 = 0.45$. The results were obtained at $v = 5$ nm/ns and averaged over 20 SMD trajectories, and error bars represent standard deviations.

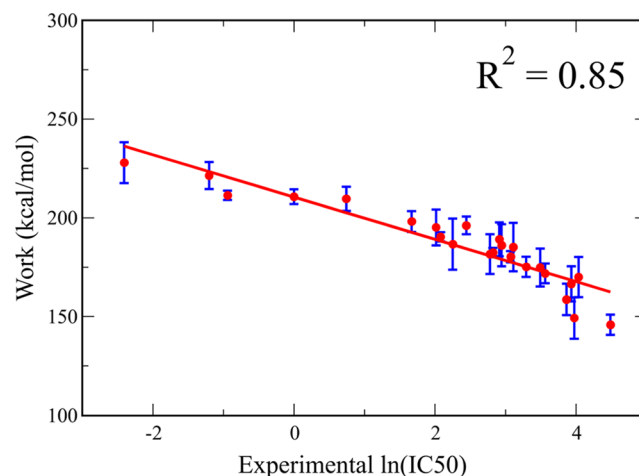


Figure 10. Pulling work as a function of the experimental $\ln(\text{IC}_{50})$ with IC_{50} measured in moles per liter. Results were obtained using SMD simulations, in which the DE-based protocol was employed to identify the zigzag pulling direction. The red line represents the data fit function. R^2 is the correlation coefficient. The results were obtained at $v = 5$ nm/ns and averaged over 20 SMD trajectories, and error bars represent standard deviations.

studies,⁴⁸ the correlation between W_{pull} and experimental $\ln(\text{IC}_{50})$ is robust against v .

Finding Zigzag Paths Using the New Protocol Does Not Take Much Time. When simulating the TcAChE–HupA and P450cam-camphor complexes, there is no trajectory along which ligands get stuck inside the receptor, which shows that our protocol has 100% success. From this point of view, our method is better than RAMD, which has a success rate below 100%.^{53,104}

The simulation time to determine the five exit pathways of camphor out of P450cam is about 1000 ps (Table S5), which is much higher than the RAMD time spent on pathways 1, 2, and 5.⁵³ Since both methods used almost the same number of trajectories, our method is more than an order of magnitude slower than RAMD. Overall, however, we can improve performance by increasing the pulling speed. Although our method is slower than RAMD, the computation time required to find the egress channels is much less than the total time. Using a node of two CPUs, Intel(R) Xeon(R) Gold 6136 CPU @ 3.00GHz with 12 physical cores per CPU, and four GPUs, NVIDIA Tesla V100-SXM2-32GB, our protocol consumes less than 3% of the total computational time for LSD1, TcAChE, and P450cam targets, indicating that the DE-based algorithm is not time-consuming and is effective in finding release paths in complex systems. For example, the simulation time is about 8 h for 10 trajectories of LSD1–ligand complexes, and the wall clock time is about 8 h 12 min.

Is the Dissociation Pathway Reversible? In the previous section, we identified the ligand dissociation pathways for several complexes. An interesting question emerges: is the dissociation pathway reversible? Since the E_{score} function only has a term that prevents a ligand from being energetically trapped, but no assumptions have been made about the unbinding or binding process, our protocol must be applied to determine the association pathways. To test the reversibility of the pathway, we tried to find a pathway for HupA to associate with TcAChE, which begins with the last snapshot of the dissociation pathway. Out of five attempts, only in one case (20%) does the ligand come back to the binding site following the same unbinding pathway. Although the success rate is not high, this demonstrates that, as a proof of concept, our protocol works for ligand–protein association. However, the solution of the problem of reversibility of the dissociation path using our approach with different sets of parameters requires further research because verifying whether this method can be used to compute the ligand association, as we have done here, and determining whether the dissociation paths are reversible are generally different problems.¹⁰⁵

CONCLUSIONS

Using SMD simulations with a straight pulling direction, we cannot properly describe the ligand release from LSD1 as the correlation between the simulated pulling work and experimental data on IC50 is quite low. This is due to the narrow tunnel, which leads to the collision of the ligand with some residues of the receptor. To overcome this difficulty, we proposed a new approach for defining a multidirectional pulling path and used it for SMD simulations instead of a unidirectional path. In our protocol, the release pathway is derived from a minimum of receptor–ligand interaction using a DE algorithm. We have shown that the protocol is successful in identifying pathways previously obtained using other methods for the TcAChE–HupA and P450cam–camphor complexes.

By applying the new method to study the binding affinity of 24 ligands to LSD1, we obtained a much better agreement with the experiment compared to the direct exit pathway SMD simulation. It would be useful to check the reliability of our protocol for other systems.

Recently, it was shown that similar to SMD,⁴⁸ TMD can be used to discern binding affinities of small compounds based on the non-equilibrium work.¹⁹ In TMD, the ligand position can

be changed as long as the distance (reaction coordinate) is maintained by the constraint. This differs from SMD, where the ligand movement is navigated by the direction of the pulling force. Hence, in SMD, the pulling direction is important, while in TMD, the constraint is crucial. In the case of the unidirectional pathway, TMD and SMD are very similar because the straight pulling direction can also be a distance constraint. However, in complex systems such as LSD1, in SMD, the zigzag pulling direction must be determined in order to steer the ligand out from the binding site, while in TMD, a reasonable constraint is needed to obtain the ejection pathways. Since the pulling direction and constraint can be different, it would be interesting to know if it is possible to use TMD instead of SMD in our protocol to find multidirectional paths. Work in this direction is in progress.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.1c01158>.

CID, name, and IC50 of the ligands studied in this work; grid size and grid center used in docking simulation for all targets; simulation time τ_{sim} for each SMD trajectory; simulation time for determining egress pathways of HupA from TcAChE; simulation time (ps) for determining exit pathways of camphor from P450cam; initial structure for SMD simulations of LSD1 and ligands obtained from docking simulation; work profiles obtained for 24 ligands bound to LSD1; dependence of the distances between the COM of ligand 1 and regions R308–G313 and K371–F396 of LSD1, force experienced by ligand 1, and the interaction energy between ligand 1 and regions R308–G313 and K371–F396 of LSD1 on the displacement; the pulling directions obtained using the MSH algorithm for LSD1; Figure S5: dependence of the pulling direction on the number of positions for ligand 10 interacting with LSD1; dependence of the pulling direction on the vector length for ligand 10 interacting with LSD1; dependence of the pulling direction on the number of iterations; force–time profiles obtained using SMD for the HupA–TcAChE complex for $\tau = 10, 20,$ and 40 ps; time dependence of nonbonded interaction energies of HupA and TcAChE along pathways pwf and pws and of RMSD of the Ω -loop; structure of cytochrome P450cam; force–time profiles obtained with straight and zigzag pulling for LSD1; and pulling work obtained at $v = 1$ nm/ns as a function of the experimental $\ln(\text{IC}_{50})$ (PDF)

Main code and scripts used to run simulations (ZIP)

AUTHOR INFORMATION

Corresponding Author

Mai Suan Li – Institute of Physics, Polish Academy of Sciences, Warsaw 02-668, Poland; orcid.org/0000-0001-7021-7916; Phone: +48 886813018; Email: masli@ifpan.edu.pl

Authors

Hoang Linh Nguyen – Life Science Lab, Institute for Computational Science and Technology, Ho Chi Minh City 729110, Vietnam; Ho Chi Minh City University of

Technology (HCMUT), Ho Chi Minh City 740500, Vietnam; Vietnam National University, Ho Chi Minh City 71300, Vietnam; orcid.org/0000-0003-4141-1642

Nguyen Quoc Thai – Life Science Lab, Institute for Computational Science and Technology, Ho Chi Minh City 729110, Vietnam; Dong Thap University, Cao Lanh City, Dong Thap 81100, Vietnam; orcid.org/0000-0001-9096-1611

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jctc.1c01158>

Notes

The authors declare no competing financial interest.
Data and software availability: the initial and representative structures of the two exit pathways of HupA from TcAChE are presented in the PW_TcAChE.zip file. The PW_P450cam.zip file contains the initial and representative structures of the five pathways to expel camphor from P450cam. The initial and representative structures of ligand 1 in complex with LSD1, obtained in a multidirectional trajectory, are stored in the LSD1_lig1.zip file. Our main code in C++ and bash script files for running simulations to determine unbinding paths are available in the “Code-Scripts.zip” file. The AutoDockTools and AutoDock Vina, used for docking simulations are available free of charge at <http://autodock.scripps.edu/resources/adt> and <http://vina.scripps.edu>, respectively. The AmberTools, which consists of Antechamber, can be downloaded for free at <https://ambermd.org/AmberTools.php>. The Acypype tool was downloaded from <https://github.com/alanwilter/acypype>. The GROMACS 2018 package was obtained from <https://manual.gromacs.org>. 3D visualization was performed by using the Pymol package and VMD software, which can be obtained from <https://pymol.org/2/> and <https://www.ks.uiuc.edu/Research/vmd/>, respectively. Qtgrace, used for plotting, can be downloaded from <https://sourceforge.net/projects/qtgrace/>. Instructions for using our code are available in the README.txt file.

ACKNOWLEDGMENTS

This work was supported by the Department of Science and Technology at Ho Chi Minh city (grant 07/2019/HĐ-KHCNTT), Narodowe Centrum Nauki in Poland (grant 2019/35/B/ST4/02086), the supercomputer centre TASK in Gdansk, PLGrid infrastructure, Poland, and the computer cluster at ICST, Vietnam. NHL was funded by Vingroup JSC and supported by the Master, PhD Scholarship Programme of Vingroup Innovation Foundation (VINIF), Institute of Big Data, code VINIF.2021.TS.029.

REFERENCES

- (1) Pagadala, N. S.; Syed, K.; Tuszynski, J. Software for molecular docking: a review. *Biophys. Rev.* **2017**, *9*, 91–102.
- (2) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **2015**, *10*, 449–461.
- (3) Jagger, B. R.; Kochanek, S. E.; Haldar, S.; Amaro, R. E.; Mulholland, A. J. Multiscale simulation approaches to modeling drug–protein binding. *Curr. Opin. Struct. Biol.* **2020**, *61*, 213–221.
- (4) Nunes-Alves, A.; Kokh, D. B.; Wade, R. C. Recent progress in molecular simulation methods for drug binding kinetics. *Curr. Opin. Struct. Biol.* **2020**, *64*, 126–133.
- (5) Ahalawat, N.; Mondal, J. An Appraisal of Computer Simulation Approaches in Elucidating Biomolecular Recognition Pathways. *J. Phys. Chem. Lett.* **2021**, *12*, 633–641.
- (6) Decherchi, S.; Cavalli, A. Thermodynamics and Kinetics of Drug-Target Binding by Molecular Simulation. *Chem. Rev.* **2020**, *120*, 12788–12833.
- (7) Ribeiro, J. M. L.; Tsai, S.-T.; Pramanik, D.; Wang, Y.; Tiwary, P. Kinetics of Ligand–Protein Dissociation from All-Atom Simulations: Are We There Yet? *Biochemistry* **2019**, *58*, 156–165.
- (8) Bernetti, M.; Masetti, M.; Rocchia, W.; Cavalli, A. Kinetics of Drug Binding and Residence Time. *Annu. Rev. Phys. Chem.* **2019**, *70*, 143–171.
- (9) Limongelli, V. Ligand binding free energy and kinetics calculation in 2020. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020**, *10*, No. e1455.
- (10) De Benedetti, P. G.; Fanelli, F. Computational modeling approaches to quantitative structure–binding kinetics relationships in drug discovery. *Drug Discovery Today* **2018**, *23*, 1396–1406.
- (11) Bruce, N. J.; Ganotra, G. K.; Kokh, D. B.; Sadiq, S. K.; Wade, R. C. New approaches for computing ligand–receptor binding kinetics. *Curr. Opin. Struct. Biol.* **2018**, *49*, 1–10.
- (12) Lu, H.; Iuliano, J. N.; Tonge, P. J. Structure–kinetic relationships that control the residence time of drug–target complexes: insights from molecular structure and dynamics. *Curr. Opin. Chem. Biol.* **2018**, *44*, 101–109.
- (13) Schuetz, D. A.; de Witte, W. E. A.; Wong, Y. C.; Knasmueller, B.; Richter, L.; Kokh, D. B.; Sadiq, S. K.; Bosma, R.; Nederpelt, I.; Heitman, L. H.; Segala, E.; Amaral, M.; Guo, D.; Andres, D.; Georgi, V.; Stoddart, L. A.; Hill, S.; Cooke, R. M.; De Graaf, C.; Leurs, R.; Frech, M.; Wade, R. C.; de Lange, E. C. M.; Ijzerman, A. P.; Müller-Fahrnow, A.; Ecker, G. F. Kinetics for Drug Discovery: an industry-driven effort to target drug residence time. *Drug Discovery Today* **2017**, *22*, 896–911.
- (14) Votapka, L. W.; Jagger, B. R.; Heyneman, A. L.; Amaro, R. E. SEEKR: Simulation Enabled Estimation of Kinetic Rates, A Computational Tool to Estimate Molecular Kinetics and Its Application to Trypsin–Benzamidine Binding. *J. Phys. Chem. B* **2017**, *121*, 3597–3606.
- (15) Mollica, L.; Decherchi, S.; Zia, S. R.; Gaspari, R.; Cavalli, A.; Rocchia, W. Kinetics of protein-ligand unbinding via smoothed potential molecular dynamics simulations. *Sci. Rep.* **2015**, *5*, 11539.
- (16) Gobbo, D.; Piretti, V.; Di Martino, R. M. C.; Tripathi, S. K.; Giabbai, B.; Storici, P.; Demitri, N.; Giroto, S.; Decherchi, S.; Cavalli, A. Investigating Drug–Target Residence Time in Kinases through Enhanced Sampling Simulations. *J. Chem. Theory Comput.* **2019**, *15*, 4646–4659.
- (17) Ahn, S.-H.; Jagger, B. R.; Amaro, R. E. Ranking of Ligand Binding Kinetics Using a Weighted Ensemble Approach and Comparison with a Multiscale Milestoning Approach. *J. Chem. Inf. Model.* **2020**, *60*, 5340–5352.
- (18) Wolf, S.; Lickert, B.; Bray, S.; Stock, G. Multisecond ligand dissociation dynamics from atomistic simulations. *Nat. Commun.* **2020**, *11*, 2918.
- (19) Wolf, S.; Amaral, M.; Lowinski, M.; Vallée, F.; Musil, D.; Güldenhaupt, J.; Dreyer, M. K.; Bomke, J.; Frech, M.; Schlitter, J.; Gerwert, K. Estimation of Protein–Ligand Unbinding Kinetics Using Non-Equilibrium Targeted Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2019**, *59*, 5135–5147.
- (20) Marques, S. M.; Bednar, D.; Damborsky, J. Computational Study of Protein-Ligand Unbinding for Enzyme Engineering. *Front. Chem.* **2019**, *6*, 650.
- (21) Capelli, R.; Lyu, W.; Bolnykh, V.; Meloni, S.; Olsen, J. M. H.; Rothlisberger, U.; Parrinello, M.; Carloni, P. Accuracy of Molecular Simulation-Based Predictions of koff Values: A Metadynamics Study. *J. Phys. Chem. Lett.* **2020**, *11*, 6373–6381.
- (22) Casasnovas, R.; Limongelli, V.; Tiwary, P.; Carloni, P.; Parrinello, M. Unbinding Kinetics of a p38 MAP Kinase Type II Inhibitor from Metadynamics Simulations. *J. Am. Chem. Soc.* **2017**, *139*, 4780–4788.

- (23) You, W.; Chang, C.-e. A. Role of Molecular Interactions and Protein Rearrangement in the Dissociation Kinetics of p38 α MAP Kinase Type-I/II/III Inhibitors. *J. Chem. Inf. Model.* **2018**, *58*, 968–981.
- (24) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel metadynamics as accurate binding free-energy method. *Proc. Natl. Acad. Sci.* **2013**, *110*, 6358–6363.
- (25) Wang, Y.; Lamim Ribeiro, J. M.; Tiwary, P. Machine learning approaches for analyzing and enhancing molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2020**, *61*, 139–145.
- (26) Grubmüller, H.; Heymann, B.; Tavan, P. Ligand binding: Molecular mechanics calculation of the streptavidin biotin rupture force. *Science* **1996**, *271*, 997–999.
- (27) Isralewitz, B.; Izrailev, S.; Schulten, K. Binding pathway of retinal to bacterio-opsin: a prediction by molecular dynamics simulations. *Biophys. J.* **1997**, *73*, 2972–2979.
- (28) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.* **1997**, *72*, 1568–1581.
- (29) Izrailev, S.; Stepaniants, S.; Isralewitz, B.; Kosztin, D.; Lu, H.; Molnar, F.; Wrigger, W.; Schulten, K. Steered molecular dynamics. *Computational Molecular Dynamics: Challenges, Methods, Ideas*; Springer, 1999; pp 39–65.
- (30) Evans, E.; Ritchie, K.; Merkel, R. Sensitive force technique to probe molecular adhesion and structural linkages at biological interfaces. *Biophys. J.* **1995**, *68*, 2580.
- (31) Florin, E.-L.; Moy, V. T.; Gaub, H. E. Adhesion forces between individual ligand-receptor pairs. *Science* **1994**, *264*, 415–417.
- (32) Colizzi, F.; Perozzo, R.; Scapozza, L.; Recanatini, M.; Cavalli, A. Single-molecule pulling simulations can discern active from inactive enzyme inhibitors. *J. Am. Chem. Soc.* **2010**, *132*, 7361–7371.
- (33) Patel, J. S.; Berteotti, A.; Ronsisvalle, S.; Rocchia, W.; Cavalli, A. Steered Molecular Dynamics Simulations for Studying Protein–Ligand Interaction in Cyclin-Dependent Kinase 5. *J. Chem. Inf. Model.* **2014**, *54*, 470–480.
- (34) Mai, B. K.; Viet, M. H.; Li, M. S. Top leads for swine influenza A/H1N1 virus revealed by steered molecular dynamics approach. *J. Chem. Inf. Model.* **2010**, *50*, 2236–2247.
- (35) Villarreal, O. D.; Yu, L.; Rodriguez, R. A.; Chen, L. Y. Computing the binding affinity of a ligand buried deep inside a protein with the hybrid steered molecular dynamics. *Biochem. Biophys. Res. Commun.* **2017**, *483*, 203–208.
- (36) Ytreberg, F. M. Absolute FKBP binding affinities obtained via nonequilibrium unbinding simulations. *J. Chem. Phys.* **2009**, *130*, 164906.
- (37) Capelli, A. M.; Bruno, A.; Entrena Guadix, A.; Costantino, G. Unbinding Pathways from the Glucocorticoid Receptor Shed Light on the Reduced Sensitivity of Glucocorticoid Ligands to a Naturally Occurring, Clinically Relevant Mutant Receptor. *J. Med. Chem.* **2013**, *56*, 7003–7014.
- (38) Capelli, A. M.; Costantino, G. Unbinding Pathways of VEGFR2 Inhibitors Revealed by Steered Molecular Dynamics. *J. Chem. Inf. Model.* **2014**, *54*, 3124–3136.
- (39) Chen, L. Y. Hybrid Steered Molecular Dynamics Approach to Computing Absolute Binding Free Energy of Ligand-Protein Complexes: A Brute Force Approach That Is Fast and Accurate. *J. Chem. Theory Comput.* **2015**, *11*, 1928–1938.
- (40) Potterton, A.; Hussein, F. S.; Southey, M. W. Y.; Bodkin, M. J.; Heifetz, A.; Coveney, P. V.; Townsend-Nicholson, A. Ensemble-Based Steered Molecular Dynamics Predicts Relative Residence Time of A2A Receptor Binders. *J. Chem. Theory Comput.* **2019**, *15*, 3316–3330.
- (41) Peng, Y.; Yang, Y.; Li, L.; Jia, Z.; Cao, W.; Alexov, E. DFMD: Fast and Effective DelPhiForce Steered Molecular Dynamics Approach to Model Ligand Approach Toward a Receptor: Application to Spermine Synthase Enzyme. *Front. Mol. Biosci.* **2019**, *6*, 74.
- (42) Okimoto, N.; Suenaga, A.; Taiji, M. Evaluation of protein–ligand affinity prediction using steered molecular dynamics simulations. *J. Biomol. Struct. Dyn.* **2017**, *35*, 3221–3231.
- (43) Li, M. S.; Mai, B. K. Steered molecular dynamics - A promising tool for drug design. *Curr. Bioinf.* **2012**, *7*, 342–351.
- (44) Do, P.-C.; Lee, E. H.; Le, L. Steered Molecular Dynamics Simulation in Rational Drug Design. *J. Chem. Inf. Model.* **2018**, *58*, 1473–1482.
- (45) Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.
- (46) Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput. Biol.* **2012**, *8*, No. e1002708.
- (47) Sehnal, D.; Svobodová Vařeková, R.; Berka, K.; Pravda, L.; Navrátilová, V.; Banáš, P.; Ionescu, C.-M.; Otyepka, M.; Koča, J. MOLE 2.0: advanced approach for analysis of biomacromolecular channels. *J. Cheminf.* **2013**, *5*, 39.
- (48) Vuong, Q. V.; Nguyen, T. T.; Li, M. S. A new method for navigating optimal direction for pulling ligand from binding pocket: application to ranking binding affinity by steered molecular dynamics. *J. Chem. Inf. Model.* **2015**, *55*, 2731–2738.
- (49) Truong, D. T.; Li, M. S. Probing the Binding Affinity by Jarzynski’s Nonequilibrium Binding Free Energy and Rupture Time. *J. Phys. Chem. B* **2018**, *122*, 4693–4699.
- (50) Mai, B. K.; Li, M. S. Neuraminidase inhibitor R-125489—a promising drug for treating influenza virus: steered molecular dynamics approach. *Biochem. Biophys. Res. Commun.* **2011**, *410*, 688–691.
- (51) Truong, D. T.; Nguyen, M. T.; Vu, V. V.; Ngo, S. T. Fast pulling of ligand approach for the design of β -secretase 1 inhibitors. *Chem. Phys. Lett.* **2017**, *671*, 142–146.
- (52) Mould, D. P.; McGonagle, A. E.; Wiseman, D. H.; Williams, E. L.; Jordan, A. M. Reversible Inhibitors of LSD1 as Therapeutic Agents in Acute Myeloid Leukemia: Clinical Significance and Progress to Date. *Med. Res. Rev.* **2015**, *35*, 586–618.
- (53) Lüdemann, S. K.; Lounnas, V.; Wade, R. C. How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms. *J. Mol. Biol.* **2000**, *303*, 797–811.
- (54) Rydzewski, J.; Nowak, W. Memetic algorithms for ligand expulsion from protein cavities. *J. Chem. Phys.* **2015**, *143*, 124101.
- (55) Kokh, D. B.; Amaral, M.; Bomke, J.; Grädler, U.; Musil, D.; Buchstaller, H.-P.; Dreyer, M. K.; Frech, M.; Lowinski, M.; Vallee, F.; Bianciotto, M.; Rak, A.; Wade, R. C. Estimation of Drug-Target Residence Times by τ -Random Acceleration Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2018**, *14*, 3859–3869.
- (56) Kokh, D. B.; Doser, B.; Richter, S.; Ormersbach, F.; Cheng, X.; Wade, R. C. A workflow for exploring ligand dissociation from a macromolecule: Efficient random acceleration molecular dynamics simulation and interaction fingerprint analysis of ligand trajectories. *J. Chem. Phys.* **2020**, *153*, 125102.
- (57) Yu, A.; Lau, A. Y. Energetics of Glutamate Binding to an Ionotropic Glutamate Receptor. *J. Phys. Chem. B* **2017**, *121*, 10436–10442.
- (58) Suh, D.; Jo, S.; Jiang, W.; Chipot, C.; Roux, B. String Method for Protein–Protein Binding Free-Energy Calculations. *J. Chem. Theory Comput.* **2019**, *15*, 5829–5844.
- (59) Sgrignani, J.; Magistrato, A. Influence of the Membrane Lipophilic Environment on the Structure and on the Substrate Access/Egress Routes of the Human Aromatase Enzyme. A Computational Study. *J. Chem. Inf. Model.* **2012**, *52*, 1595–1606.
- (60) Magistrato, A.; Sgrignani, J.; Krause, R.; Cavalli, A. Single or Multiple Access Channels to the CYP450s Active Site? An Answer from Free Energy Simulations of the Human Aromatase Enzyme. *J. Phys. Chem. Lett.* **2017**, *8*, 2036–2042.

- (61) Rydzewski, J.; Nowak, W. Ligand diffusion in proteins via enhanced sampling in molecular dynamics. *Phys. Life Rev.* **2017**, *22*–23, 58.
- (62) Rydzewski, J.; Nowak, W. Machine learning based dimensionality reduction facilitates ligand diffusion paths assessment: a case of cytochrome P450cam. *J. Chem. Theor. Comput.* **2016**, *12*, 2110–2120.
- (63) Rydzewski, J.; Valsson, O. Finding multiple reaction pathways of ligand unbinding. *J. Chem. Phys.* **2019**, *150*, 221101.
- (64) Rydzewski, J. maze: Heterogeneous ligand unbinding along transient protein tunnels. *Comput. Phys. Commun.* **2020**, *247*, 106865.
- (65) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **2020**, *153*, 044130.
- (66) Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I.; Brozell, S. R.; Cerutti, D.; Cheatham, T.; Cruzeiro, V. W. D.; Darden, T.; Duke, R. E. *Amber 2021: Reference Manual*, 2021.
- (67) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 198–210.
- (68) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Brogna, R. A.; Parrinello, M. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.
- (69) Yang, K.; Liu, X.; Wang, X.; Jiang, H. A steered molecular dynamics method with adaptive direction adjustments. *Biochem. Biophys. Res. Commun.* **2009**, *379*, 494–498.
- (70) Gu, J.; Li, H.; Wang, X. A self-adaptive steered molecular dynamics method based on minimization of stretching force reveals the binding affinity of protein–ligand complexes. *Molecules* **2015**, *20*, 19236–19251.
- (71) Storn, R.; Price, K. Differential evolution - A simple and efficient heuristic for global optimization over continuous spaces. *J. Global Optim.* **1997**, *11*, 341–359.
- (72) Rydzewski, J.; Jakubowski, R.; Nowak, W.; Grubmüller, H. Kinetics of Huperzine A Dissociation from Acetylcholinesterase via Multiple Unbinding Pathways. *J. Chem. Theory Comput.* **2018**, *14*, 2843–2851.
- (73) Yang, M.; Culhane, J. C.; Szewczuk, L. M.; Gocke, C. B.; Brautigam, C. A.; Tomchick, D. R.; Machius, M.; Cole, P. A.; Yu, H. Structural basis of histone demethylation by LSD1 revealed by suicide inactivation. *Nat. Struct. Mol. Biol.* **2007**, *14*, 535.
- (74) Bolton, E. E.; Wang, Y.; Thiessen, P. A.; Bryant, S. H. PubChem: integrated platform of small molecules and biological activities. *Annu. Rep. Comput. Chem.* **2008**, *4*, 217–241.
- (75) Search ChemSpider, ChemSpiderSearch and share chemistry. <http://www.chemspider.com/> Accessed on May 19, 2018.
- (76) Dvir, H.; Jiang, H. L.; Wong, D. M.; Harel, M.; Chetrit, M.; He, X. C.; Jin, G. Y.; Yu, G. L.; Tang, X. C.; Silman, I.; Bai, D. L.; Sussman, J. L. X-ray Structures of Torpedo californica Acetylcholinesterase Complexed with (+)-Huperzine A and (–)-Huperzine B: Structural Evidence for an Active Site Rearrangement. *Biochemistry* **2002**, *41*, 10810–10818.
- (77) Poulos, T. L.; Finzel, B. C.; Howard, A. J. High-resolution crystal structure of cytochrome P450cam. *J. Mol. Biol.* **1987**, *195*, 687–700.
- (78) Shahrokh, K.; Orendt, A.; Yost, G. S.; Cheatham, T. E., 3rd. Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle. *J. Comput. Chem.* **2012**, *33*, 119–133.
- (79) Sanner, M. F. Python: a programming language for software integration and development. *J. Mol. Graph. Model.* **1999**, *17*, 57–61.
- (80) Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461.
- (81) 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.
- (82) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- (83) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950–1958.
- (84) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Antechamber: an accessory software package for molecular mechanical calculations. *J. Am. Chem. Soc.* **2001**, *222*, U403.
- (85) da Silva, A. W. S.; Vranken, W. F. ACPYPE-Antechamber python parser interface. *BMC Res. Notes* **2012**, *5*, 367.
- (86) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (87) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic Charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.
- (88) Bartholomew-Biggs, M. The Steepest Descent Method. *Nonlinear Optimization with Engineering Applications; Springer Optimization and Its Applications*; Springer, 2005; pp 1–8.
- (89) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (90) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (91) Binnig, G.; Quate, C. F.; Gerber, C. Atomic Force Microscope. *Phys. Rev. Lett.* **1986**, *56*, 930–933.
- (92) Rico, F.; Russek, A.; González, L.; Grubmüller, H.; Scheuring, S. Heterogeneous and rate-dependent streptavidin–biotin unbinding revealed by high-speed force spectroscopy and atomistic simulations. *Proc. Natl. Acad. Sci.* **2019**, *116*, 6594–6601.
- (93) Vuong, Q. V.; Nguyen, T. T.; Li, M. S. A New Method for Navigating Optimal Direction for Pulling Ligand from Binding Pocket: Application to Ranking Binding Affinity by Steered Molecular Dynamics. *J. Chem. Inf. Model.* **2015**, *55*, 2731–2738.
- (94) Schober, P.; Boer, C.; Schwarte, L. A. Correlation Coefficients: Appropriate Use and Interpretation. *Anesth. Analg.* **2018**, *126*, 1763–1768.
- (95) Bai, F.; Xu, Y.; Chen, J.; Liu, Q.; Gu, J.; Wang, X.; Ma, J.; Li, H.; Onuchic, J. N.; Jiang, H. Free energy landscape for the binding process of Huperzine A to acetylcholinesterase. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 4273–4278.
- (96) Tara, S.; Helms, V.; Straatsma, T. P.; McCammon, J. A. Molecular dynamics of mouse acetylcholinesterase complexed with huperzine A. *Biopolymers* **1999**, *50*, 347–359.
- (97) Sanson, B.; Colletier, J.-P.; Xu, Y.; Lang, P. T.; Jiang, H.; Silman, I.; Sussman, J. L.; Weik, M. Backdoor opening mechanism in acetylcholinesterase based on X-ray crystallography and molecular dynamics simulations. *Protein Sci.* **2011**, *20*, 1114–1118.
- (98) Cojocar, V.; Winn, P. J.; Wade, R. C. The ins and outs of cytochrome P450s. *Biochim. Biophys. Acta Gen. Subj.* **2007**, *1770*, 390–401.
- (99) Wade, R. C.; Winn, P. J.; Schlichting, I. A survey of active site access channels in cytochromes P450. *J. Inorg. Biochem.* **2004**, *98*, 1175–1182.
- (100) Lüdemann, S. K.; Carugo, O.; Wade, R. C. Substrate Access to Cytochrome P450cam: A Comparison of a Thermal Motion Pathway Analysis with Molecular Dynamics Simulation Data. *J. Mol. Model.* **1997**, *3*, 369–374.

(101) Winn, P. J.; Lüdemann, S. K.; Gauges, R.; Lounnas, V.; Wade, R. C. 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.* **2002**, *99*, 5361–5366.

(102) Oprea, T. I.; Hummer, G.; García, A. E. Identification of a functional water channel in cytochrome P450 enzymes. *Proc. Natl. Acad. Sci.* **1997**, *94*, 2133–2138.

(103) Vohra, S.; Musgaard, M.; Bell, S. G.; Wong, L.-L.; Zhou, W.; Biggin, P. C. The dynamics of camphor in the cytochrome P450 CYP101D2. *Protein Sci.* **2013**, *22*, 1218–1229.

(104) Du, J.; Liu, L.; Guo, L. Z.; Yao, X. J.; Yang, J. M. Molecular basis of P450 OleTJE: an investigation of substrate binding mechanism and major pathways. *J. Comput.-Aided Mol. Des.* **2017**, *31*, 483–495.

(105) Bhatt, D.; Zuckerman, D. M. Beyond Microscopic Reversibility: Are Observable Nonequilibrium Processes Precisely Reversible? *J. Chem. Theory Comput.* **2011**, *7*, 2520–2527.