

Article

Ligand Binding Path Sampling Based on Parallel Cascade Selection Molecular Dynamics: LB-PaCS-MD

Hayato Aida ¹, Yasuteru Shigeta ²  and Ryuhei Harada ^{2,*}

¹ Graduate School of Science and Technology, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8577, Japan; aida.hayato.sp@alumni.tsukuba.ac.jp

² Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8577, Japan; shigeta@ccs.tsukuba.ac.jp

* Correspondence: ryuhei@ccs.tsukuba.ac.jp

Abstract: Parallel cascade selection molecular dynamics (PaCS-MD) is a rare-event sampling method that generates transition pathways between a reactant and product. To sample the transition pathways, PaCS-MD repeats short-time MD simulations from important configurations as conformational resampling cycles. In this study, PaCS-MD was extended to sample ligand binding pathways toward a target protein, which is referred to as LB-PaCS-MD. In a ligand-concentrated environment, where multiple ligand copies are randomly arranged around the target protein, LB-PaCS-MD allows for the frequent sampling of ligand binding pathways. To select the important configurations, we specified the center of mass (COM) distance between each ligand and the relevant binding site of the target protein, where snapshots generated by the short-time MD simulations were ranked by their COM distance values. From each cycle, snapshots with smaller COM distance values were selected as the important configurations to be resampled using the short-time MD simulations. By repeating conformational resampling cycles, the COM distance values gradually decreased and converged to constants, meaning that a set of ligand binding pathways toward the target protein was sampled by LB-PaCS-MD. To demonstrate relative efficiency, LB-PaCS-MD was applied to several proteins, and their ligand binding pathways were sampled more frequently than conventional MD simulations.

Keywords: PaCS-MD; ligand binding process; ligand–protein complex formation; rare-event sampling method; molecular dynamics simulation; distributed computing



Citation: Aida, H.; Shigeta, Y.; Harada, R. Ligand Binding Path Sampling Based on Parallel Cascade Selection Molecular Dynamics: LB-PaCS-MD. *Materials* **2022**, *15*, 1490. <https://doi.org/10.3390/ma15041490>

Academic Editors: Takahiko Ban, Hideki Nabika and Takahiro Nakamura

Received: 22 December 2021

Accepted: 10 February 2022

Published: 17 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Self-assembly/organization is a fundamental phenomenon in which well-designed molecules spontaneously gather to form unique structures. These processes are common in nature and are strongly related to molecular recognition, transfer, reaction, and catalysis [1–4]. Typical examples include self-assembly of a nanocube [5], self-organization of biomolecules such as double helix DNAs and tertiary/quaternary structures of proteins, membrane formation by the self-assembly of lipid bilayers [6], and formation of amyloid fibers [7]. In biological systems, molecular self-assembly/organization plays a crucial role in cell function [8]. Similarly, ligand–protein complex formation upon ligand binding plays an important role in molecular recognition. More specifically, a target protein recognizes a ligand through their ligand–protein interactions. Therefore, from a molecular design point of view, understanding the ligand binding mechanism is essential. Structurally, a target protein complexed with a ligand is determined by various non-covalent interactions, such as hydrogen bonding [9] and van der Waals (vdW) interactions [10]. Finally, a thermodynamically stable complex is selected from possible ligand–protein configurations. However, owing to their complex interactions, the molecular mechanisms of ligand binding processes remain unclear. For this reason, theoretical and computational studies are required to address ligand binding processes at an atomic resolution. In this study,

we regard ligand binding processes as self-assembly/self-organization and elucidate their detailed mechanisms using all-atom molecular dynamics (MD) simulations.

As previously described, target proteins recognize their ligands based on ligand–protein interactions. Elucidation of ligand binding pathways is important for understanding the biological functions of complex formations. To form a ligand–protein complex, proteins use their fluctuations to bind to ligands, meaning that protein fluctuation plays a crucial role in ligand recognition. When the target proteins and ligands are rigid, rigid docking simulations predict their ligand–protein complexes. [11–14] However, it is difficult to treat flexible ligands and/or target proteins with buried binding cavities in terms of rigid docking simulations. Generally, these types of ligand binding processes tend to couple with the conformational transitions of target proteins. To treat ligand binding processes explicitly, MD simulations allow for the evaluation of the flexibility of protein–ligand systems at an atomic resolution. Therefore, MD simulations are increasingly being applied to elucidate ligand–protein binding processes by explicitly considering protein fluctuations.

The characterization of a ligand binding pathway is a challenging issue judging from the timescale gap between conventional MD (cMD) simulations and ligand–protein binding processes. In previous studies, cMD simulations allowed for the sampling of several binding/unbinding events [15,16], some utilizing an Anton (a special-purpose supercomputer) [17–20]. Notable efforts have also been made in sampling ligand binding pathways using distributed MD simulations combined with Markov state models (MSMs) [21,22] or enhanced sampling methods, such as metadynamics [23–27] and replica exchange methods [28–30]. These enhanced sampling methods allow one to sample ligand binding pathways for several proteins.

Generally, ligand binding pathways are regarded as rare events since they are stochastically induced over relatively long timescales, i.e., microseconds, and efficient techniques for sampling a set of ligand binding pathways are yet to be developed. In previous studies, we have developed several rare-event sampling methods based on distributed computing, where their overviews have been introduced in our review papers [31,32]. In distributed computing, short-time MD simulations are independently restarted from important configurations (different initial structures). To promote frequent transitions, a cycle of conformational resampling is repeated, where the cycle consists of (1) selections of the important configurations and (2) the short-time MD simulations from them. When restarting the short-time MD simulations from the different important configurations, their initial velocities are randomly regenerated based on the Maxwell–Boltzmann distribution. As our previous study reported [33], the regenerations of initial velocities promote the conformational transitions of proteins. In restarting the short-time MD, a set of different configurations/velocities works as a perturbation to promote transitions of a target protein. Thus, it is noted that LB-PaCS-MD is designated for finding ligand docking pathway and lacks time–series information and exact statistical ensemble. However, once the ligand docking pathways are obtained, the free-energy landscape (FEL) can be accurately evaluated. Here, we evaluated the rough FEL using the conformations sampled by LB-PaCS-MD for the ligand-screening purpose. To increase the transition probability, appropriate initial structures that are relevant to the rare events should be selected by defining reaction coordinates (RCs) that characterize the transitions of the target protein. By referring to the RC values, the initial structures of the short-time MD simulations are appropriately selected. Therefore, the optimal RC should be specified.

As a standard rare-event sampling method, parallel cascade selection MD (PaCS-MD) series [34–39] have been proposed. PaCS-MD is categorized into two types: (1) targeted PaCS-MD (t-PaCS-MD) [34] generates transition pathways from a given reactant to a product under the condition that a set of end-point structures (a specified reactant and product) are preliminarily known; (2) non-targeted PaCS-MD (nt-PaCS-MD) [35] generates transition pathways from reactants to neighboring metastable states without referring to any product. t-PaCS-MD and nt-PaCS-MD have been applied to several biological systems

to identify rare events and have succeeded in searching broad free-energy landscapes of several proteins [40–42].

Based on PaCS-MD, this study concerns the development of ligand binding PaCS-MD (LB-PaCS-MD): a t-PaCS-MD technique to sample ligand binding pathways of a target protein. Under a ligand-concentrated environment, where multiple ligand copies are randomly arranged around a target protein, frequent sampling of the ligand binding pathways is permitted. To select important initial structures of short-time MD simulations, LB-PaCS-MD specifies the center-of-mass (COM) distance between each ligand and the binding site of the target protein, d_{COM} . By referring to the d_{COM} value, snapshots of the short-time MD simulations are ranked; for each cycle, snapshots with smaller d_{COM} values are selected as the initial structures. These important configurations are resampled by the short-time MD simulations. With subsequent cycles of LB-PaCS-MD, the d_{COM} value gradually decreases and converges to a constant, meaning that a set of ligand binding pathways to the target protein has been successfully sampled by LB-PaCS-MD.

For demonstrations, LB-PaCS-MD was also applied to several ligand–protein complexes, and their ligand binding pathways were successfully sampled with nanosecond-order computational costs. Finally, MSMs were constructed using the sampled ligand binding trajectories to evaluate their free-energy profiles.

2. Materials and Methods

2.1. *t*-PaCS-MD

To promote transitions from a given product to a reactant, t-PaCS-MD repeats short-time MD simulations from important configurations. Here, root-mean-square deviation (RMSD) is a simple RC that allows for the selection of configurations that are structurally similar to the product. At every cycle, t-PaCS-MD selects configurations with smaller RMSD values for the product ($\text{RMSD}_{\text{product}}$) and restarts short-time MD simulations from them. By repeating conformational resampling cycles, the $\text{RMSD}_{\text{product}}$ value converges to a small constant, indicating that t-PaCS-MD successfully samples the transition pathways from the reactant to product when the following condition is satisfied: $\text{RMSD}_{\text{product}} < \text{cutoff}$ (a threshold). For RCs, any physical variable can be specified. For example, another physical variable has been proposed in our previous study [43].

2.2. *LB*-PaCS-MD

LB-PaCS-MD is an extension of t-PaCS-MD to sample ligand binding pathways toward a ligand–protein complex. In LB-PaCS-MD, a target protein with multiple ligand copies is regarded as a reactant. LB-PaCS-MD samples ligand binding pathways from a ligand-unbounded state to a ligand-bound state. To sample ligand binding pathways frequently, multiple ligand copies were randomly arranged around the target protein. In this ligand-concentrated environment, LB-PaCS-MD more frequently samples ligand binding pathways than a ligand-diluted environment, e.g., the target protein with a single ligand.

Figure 1 shows the procedure of LB-PaCS-MD. In LB-PaCS-MD, d_{COM} is an RC that selects initial structures to be resampled by short-time MD simulations. At every cycle, d_{COM} is calculated for all the snapshots generated by the short-time MD simulations at the previous cycle. Here, the total snapshots per cycle is $N \times M$ when N short-time MD simulations are independently performed and each short-time MD records M snapshots. Then, all the $N \times M$ snapshots are ranked by their d_{COM} values, i.e., ascending order in d_{COM} . Finally, the highly ranked top N snapshots with smaller d_{COM} values are selected from the $N \times M$ snapshots as initial structures to of N short-time MD simulations at the next cycle. The total number of short-time MD simulations (N) is specified depending on available computational resources. The large N value allows one to sample the ligand binding pathways smoothly. As a strategy to efficiently perform LB-PaCS-MD, it might be better to prevent ligand aggregations during conformational resampling cycles. The ligand aggregations decrease the ligand binding sampling efficiency. Therefore, it is recommended to reset the number of ligands when the arranged ligand copies aggregated during LB-

PaCS-MD. By repeating the cycles, the d_{COM} value gradually decreases and converges to a constant, indicating that a set of ligand binding pathways toward the binding site of a target protein is sampled by LB-PaCS-MD. To terminate LB-PaCS-MD, one should monitor the d_{COM} profiles during the cycles. More specifically, LB-PaCS-MD is terminated when the following condition is satisfied: $d_{COM} < \text{cutoff}$ (a threshold).

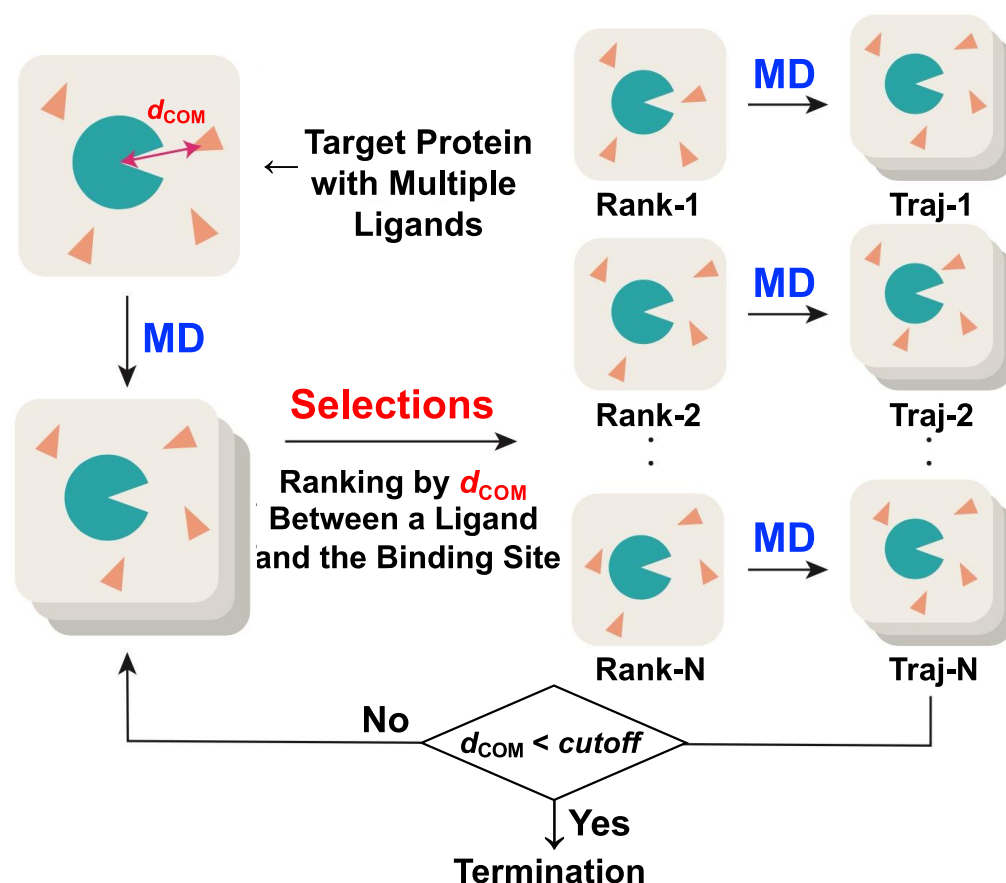


Figure 1. Flowchart of ligand binding path sampling based on LB-PaCS-MD. To frequently sample ligand binding pathways, LB-PaCS-MD repeats conformational resampling cycles consisting of: (1) selections of important configurations (initial structures) with smaller COM distance between the binding site of a target protein and each ligand (d_{COM}); (2) restarting of (short-time) MD simulations from the selected initial structures. LB-PaCS-MD is terminated when a d_{COM} value is less than cutoff (a threshold).

The present strategy for selecting initial structures is applicable for different ligand entries under the condition that more than one ligand exist. In that case, d_{COM} is defined as a COM distance between each ligand and each binding site. Therefore, LB-PaCS-MD might be applied to systems with plural ligands as an extension.

2.3. Free-Energy Profiles on Ligand-Binding Pathways

To quantitatively evaluate the performance of LB-PaCS-MD, we calculated the free-energy profiles of the sampled ligand-binding pathways using stationary distributions obtained by MSM constructions. The distributions of the ligand-binding trajectories were obtained by projection onto a conformational subspace spanned by reasonable RCs. In this evaluation, d_{COM} was selected as an RC. The construction of a reliable MSM consists of two steps: (1) defining microstates and (2) setting a reasonable lag time, τ . To define the microstates in step (1), the distributions are clustered using the k -means algorithm to yield N clusters. Each snapshot is then assigned to the closest cluster (microstate). Next, transitions among the N microstates are counted for each trajectory to obtain a maximum

likelihood ($N \times N$) transition matrix, T . Each component (T_{ij}) represents the transition rate from the i th to j th microstates, with τ averaged over all the trajectories. In step (2), the i th slowest implied time, t_i , is estimated from the i th largest eigenvalue, $\lambda_i(\tau)$, of T , defined as follows:

$$t_i(\tau) = -\frac{\tau}{\ln \lambda_i(\tau)} \quad (1)$$

where the optimal value of the implied time is determined by adjusting τ until $t_i(\tau)$ reaches an approximately constant value. Convergence of the implied time ensures that the target system satisfies the Markov assumption.

After determining the optimal τ , transitions among the microstates are again counted to estimate T under a constrained detailed balance. To calculate the free-energy profile, a stationary distribution, π ($\sum_i \pi_i = 1$), is obtained from T ($\pi = T \pi$), where π corresponds to one of the eigenvectors of T . Finally, the free-energy profile under a given temperature (T) is calculated for each microstate using the Boltzmann constant (k_B), as follows:

$$F_i = -k_B T \ln \frac{\pi_i}{\max_k \pi_k} \quad (i = 1, 2, \dots, N) \quad (2)$$

where the origin of the free-energy profile is defined by the maximum value of π . The following review papers help one to construct MSMs [44,45].

2.4. Demonstrations of LB-PaCS-MD

In this study, a set of systems was used to demonstrate LB-PaCS-MD, i.e., LB-PaCS-MD efficiently sampled a set of ligand binding pathways of the following proteins: (1) a mutant of T4 lysozyme (T4L) and (2) SARS-CoV-2 main protease (SARS-CoV2 M^{Pro}). For T4L, a holo-form was experimentally determined as a protein data bank (PDB) structure, PDBid: 3DMX [46], i.e., T4L complexed with a ligand (benzene molecule). For the SARS-CoV-2 M^{Pro}, several compounds with binding affinities were screened in previous studies [47–49], and four of these compounds were employed in this demonstration. The followings are the simulated compounds, i.e., A—*x*77; B—methyl rosmarinate; C—niclosamide; D—5,7,3',4'-tetrahydroxy-2'-(3,3-dimethylallyl). Here, all the cysteine residues of SARS-CoV2 M^{Pro} were modelled in their thiole form. For protonation states in both systems, each histidine residue was treated as a histidine with hydrogen on the epsilon nitrogen. To construct a set of parameters of all the compounds, their restrained electrostatic potential (RESP) charges were calculated based on the gas phase HF/6-31G(d) after the structural optimizations based on B3LYP/6-31G(d) as quantum mechanics calculations. Based on the RESP charges, the parameters of all the compound were generated (Supplementary Materials).

For both systems, apo-form PDB structures were used as reactants, i.e., 3DMV [46] for T4L and (2) 6M03 for the SARS-CoV-2 M^{Pro}, without their ligands. Figure 2 shows a set of overviews of both systems. Figure 2a shows the structure of T4L (holo-form), where a set of key residues (M102 and F114) around the binding site is highlighted. Figure 2c shows the structure of SARS-CoV2 M^{Pro} (apo form), where a set of catalytic dyad residues (H41 and C145) around the binding site is highlighted.

These structures were solvated with the TIP3P water model [50]. The solvent box sizes were initially set in Å as follows: (L_x, L_y, L_z) = (70.8, 71.0, 71.0) for T4L and (L_x, L_y, L_z) = (87.2, 87.8, 87.6) for SARS-CoV2 M^{Pro}. To neutralize the solvated systems, counter ions were arranged around each protein. Here, 8 Cl⁻ (4 Na⁺) ions were added to the T4L (SARS-CoV2 M^{Pro}) system after the solvation. From the modeled systems, 100 ps MD simulations were initiated. Specifically, 100 ps *NVT* ($T = 300$ K) and 100 ps *NPT* ($T = 300$ K and $P = 1$ bar) MD simulations were performed for each solvated configuration. The final snapshots of the 200 ps MD simulations of both systems were regarded as relaxed configurations without their ligands. Before launching each LB-PaCS-MD trial, six ligand copies were randomly arranged around each relaxed configuration and were regarded as the reactants of the LB-PaCS-MD trials. For example, Figure 2b,d show a set of overviews of the positions of

the six randomly arranged ligand copies around these proteins. Generally, it is difficult to specify an optimal number of ligand copies. Therefore, we specified the number of ligand copies ($n = 6$) so that the randomly arranged ligand copies in the initial states did not strongly interact with each other to avoid the aggregation.

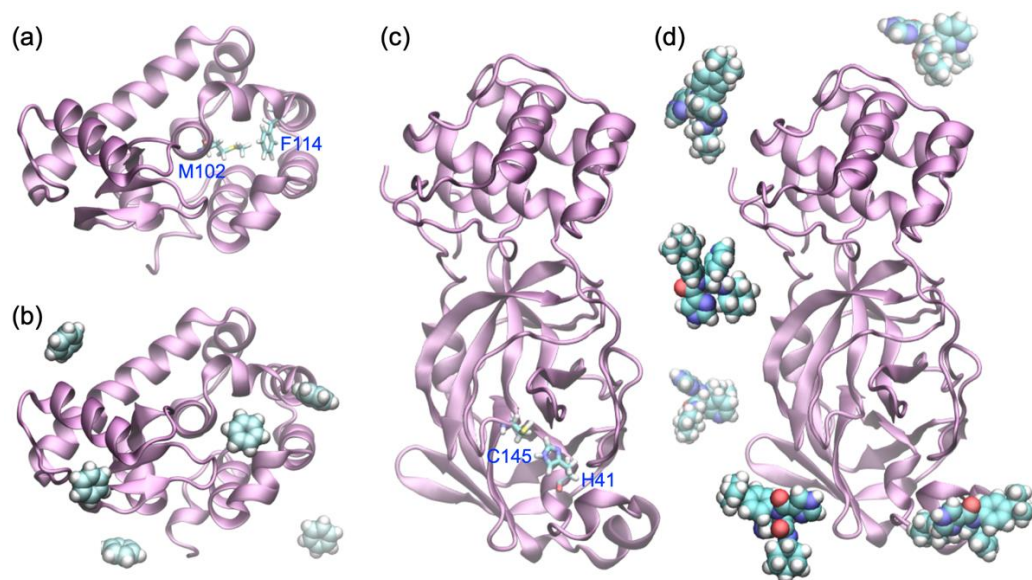


Figure 2. (a) Holo-form structure of T4L. The key residues (M102 and F114) around the binding site are depicted in licorice representation. (b) T4L with the positions of six randomly arranged ligand copies (benzene molecules in vdW representation). (c) Apo-form structure of SARS-CoV2 M^{Pro}. The dyad residues (H41 and C145) around the binding site are depicted in licorice representation. (d) SARS-CoV2 M^{Pro} with six randomly arranged ligand copies (compound A molecules in vdW representation).

The following are the MD conditions specified in each LB-PaCS-MD trial. To increase the time step of each MD simulation to 2 fs, the chemical bonds of the solutes were treated with the LINCS algorithm [51]. The chemical bonds of water molecules were treated with the SETTLE algorithm [52]. The modified Berendsen thermostat [53] and Parrinello–Rahman barostat [54,55] were used to control the temperature and pressure, respectively. The particle mesh Ewald method [56] evaluated electrostatic interactions using a real-space cutoff of 10 Å. The cutoff value for the vdW interactions was set to 10 Å. For both demonstrations, all MD simulations were performed using the GPU version of the GROMACS 2018 package [57], where the AMBER 14SBonlysc force field parameter [58] was specified. During the LB-PaCS-MD cycles, all the snapshots were ranked by their d_{COM} values. At every cycle, the highly ranked snapshots were selected as initial structures to be resampled by short-time (100 ps) MD simulations. Therefore, a computational cost per cycle was 1 ns (10 initial structures \times 100 ps MD simulations) in both systems.

We first confirmed the validity of our MD simulations. Here, to ensure that both proteins stay intact, the averages of backbone RMSD with their standard deviations were calculated using the trajectories of the final LB-PaCS-MD cycles. The averages and standard deviations for each reactant without each ligand were as follows: T4L— 1.00 ± 0.14 Å; SARS-CoV2 M^{Pro}— 1.70 ± 0.19 Å (compound A), 1.58 ± 0.27 Å (compound B), 1.56 ± 0.34 Å (compound C), 1.25 ± 0.15 Å (compound D). These backbone RMSD values ensure that both proteins were structurally stable during the LB-PaCS-MD cycles. A set of RMSD profiles versus time of all the systems is given in Figure S1 (Supplementary Materials).

3. Results

3.1. The Ligand Binding Path Sampling Efficiency of LB-PaCS-MD

The ligand binding path sampling efficiency was analyzed using T4L. As a reactant of LB-PaCS-MD, six benzene molecules were randomly arranged around the apo-form of T4L. Starting from the reactant, three LB-PaCS-MD trials independently sampled a set of ligand binding processes toward the binding site (M102) of T4L. As references, two types of ligand binding searches were considered: (1) three normal PaCS-MD trials under a dilute system (T4L with a single benzene molecule) and (2) a 100 ns cMD simulation under a dilute system (T4L with a single benzene molecule).

To monitor the ligand binding pathways, d_{COM} was defined as the COM distance between each benzene molecule and the binding site residue (M102). For T4L, a cutoff value (4.0 Å) was specified based on d_{COM} measured in the experimental ligand–protein complex (PDBid: 3DMX). By referring to the PDB structure, each COM was defined for the binding site (M102) and the benzene molecule using their carbon atoms. For the experimental structure, the COM distance between M102 and the benzene molecule was measured as about 4.0 Å. Generally, a cutoff value is specified as a COM distance between each binding site and ligand when some experimental complexes are available. Each COM can be simply defined using coordinates of carbon atoms of each binding site of a target protein and ligand. Note here that this cutoff value was not the firm choice. We have often faced with the problem of deciding when to finish conformational search [59–63]. To be more precise, one may consider the convergence of the distribution evaluated from LB-PaCS-MD. For example, a difference in the distribution of RC values averaged over n cycle is one choice since the distribution does not largely change when the conformations reach nearby a product state (a ligand binding structure).

Figure 3 shows a set of d_{COM} profiles for the three LB-PaCS-MD trials. As shown in Figure 3a, all the LB-PaCS-MD trials successfully sampled the ligand binding pathways until the 100th cycle (100 ns), i.e., all the d_{COM} profiles converged to small values ($d_{\text{COM}} \sim 3.7$ Å) after the 100th cycle. Here, 1 of the 3 normal PaCS-MD trials failed to sample the ligand-binding pathways after the 100th cycle ($d_{\text{COM}} \sim 4.7$ Å). In contrast, the 100 ns cMD simulations failed to sample pathways at the same computational cost, indicating a higher sampling efficiency of LB-PaCS-MD than both normal PaCS-MD and cMD.

Here, the ligand binding forms sampled by LB-PaCS-MD were compared with the X-ray structure. Figure S2a (Supplementary Materials) shows a sampled ligand binding snapshot, where it was superimposed for the X-ray structure. In Figure S2a, the key residue at the binding site (M102) is depicted in vdW representation, and each benzene molecule is depicted in licorice representation (blue—LB-PaCS-MD; red—the X-ray structure). The RMSD value between these benzene molecules after superimposition of T4L was 0.79 Å. To focus on the ligand binding form, a set of key residues (M102 and F114) around the binding site was depicted with each ligand (Figure S2b,c). In both binding forms, the benzene molecules maintained the common interaction with M102. For the binding form sampled by LB-PaCS-MD (Figure S2b, Supplementary Materials), the ligand had an additional contact with F114, which might be a weak stacking of their aromatic rings. As previous studies reported [64,65], the binding cavity of T4L tends to fluctuate to catch the ligand since the key residues around the binding site is deeply buried and sterically inaccessible from bulk solvent, indicating that the binding form of the ligand might be affected by the thermal fluctuation of the binding cavity and takes variations.

3.2. Elucidations of the Ligand Binding Pathways of SARS-CoV2 Main^{Pro}

For the second demonstration, LB-PaCS-MD was applied to sample a set of ligand-binding pathways of SARS-CoV-2 M^{Pro}. Based on a previous study that screened compounds with binding affinities [47–49], four compounds were used as a set of ligand candidates. To model each complex system, six ligand copies were randomly arranged around SARS-CoV-2 M^{Pro}. For SARS-CoV-2 M^{Pro}, the complex structures have not been experimentally determined, i.e., the explicit d_{COM} values are unavailable. Therefore, a

common cutoff value (6.0 Å) was used to terminate the LB-PaCS-MD cycles. Starting from each reactant, the LB-PaCS-MD trials successfully sampled the ligand binding pathways of SARS-CoV-2 M^{pro}.

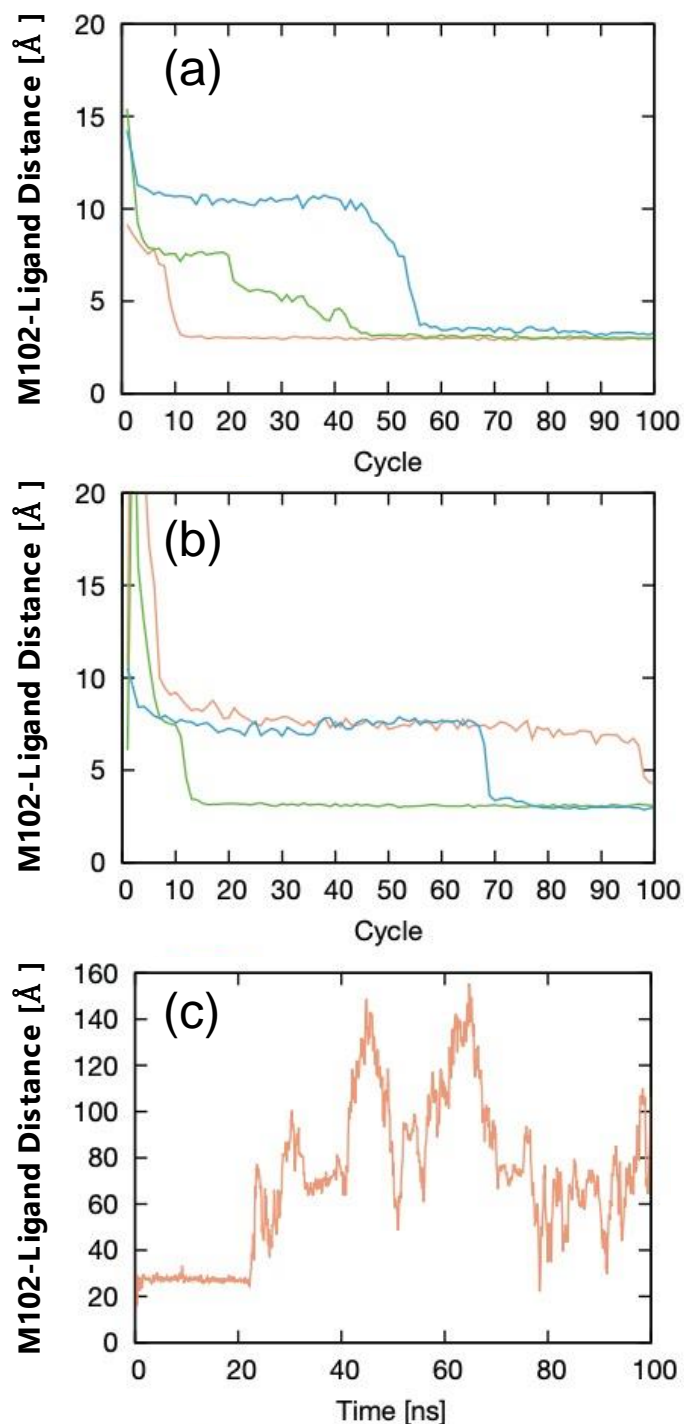


Figure 3. Profiles of the COM distance between each benzene molecule and the binding site residue (M102) of T4L for the ligand-binding trajectories sampled by three LB-PaCS-MD trials. (a) LB-PaCS-MD trials using multiple ligand copies. (b) Normal PaCS-MD trials using a single ligand. (c) The trajectory sampled by a 100 ns cMD simulation under a dilute condition (T4L with one benzene molecule).

Figure 4 shows a set of d_{COM} profiles between each compound and the binding site residue (C145). The d_{COM} values converged sufficiently until the 50th cycle. All the LB-

PaCS-MD trials successfully sampled the ligand binding pathways for compounds A, B, and C. In contrast, one of the three LB-PaCS-MD trials failed to sample the ligand-binding pathways for compound D. Based on these analyses, LB-PaCS-MD allows one to sample a set of ligand-binding pathways of a given target protein with a reasonable (nanosecond-order) computational cost.

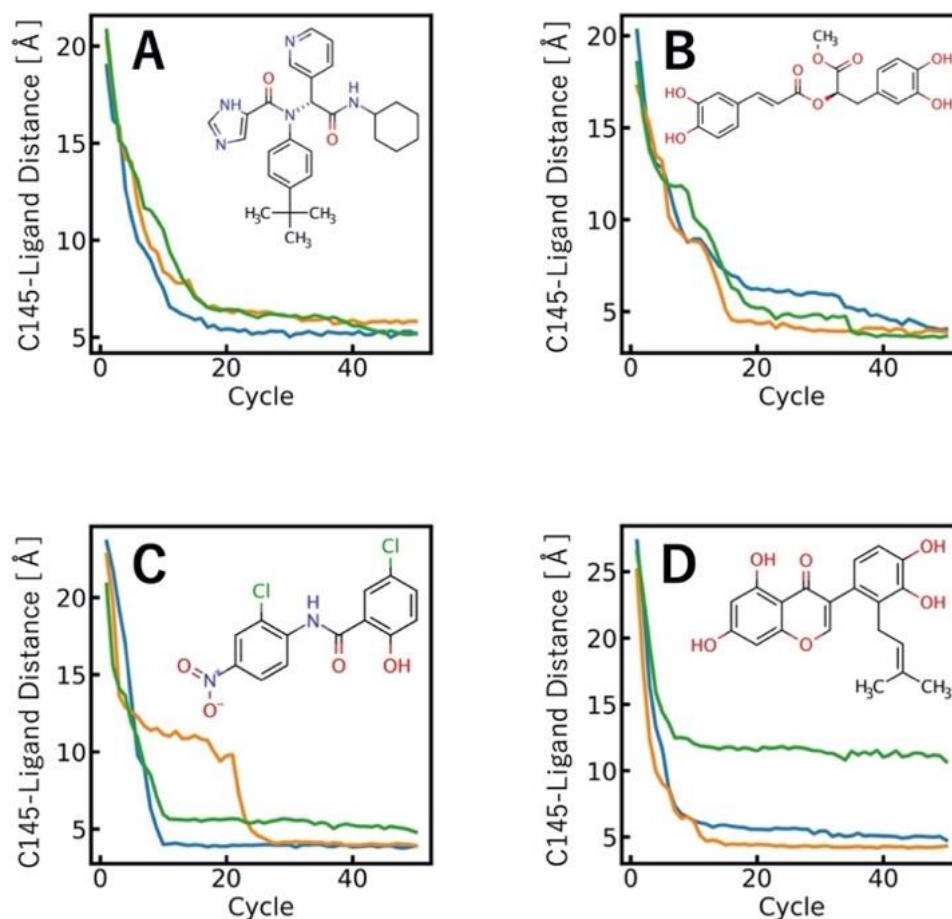


Figure 4. Profiles of the COM distances between the binding site residue (C145) of SARS-CoV-2 M^{Pro} and all the compounds for their ligand binding trajectories. (A) x77; (B) methyl rosmarinate; (C) niclosamide; (D) 5,7,3',4'-tetrahydroxy-2'-(3,3-dimethylallyl) isoflavone.

The ligand binding pathways sampled by the LB-PaCS-MD trials were evaluated quantitatively by constructing MSMs. The MSMs were constructed using the distributions projected onto the d_{COM} subspace, providing a set of stationary distributions that allowed for the calculation of free-energy profiles as a function of d_{COM} (Figure 5). The free-energy profiles showed characteristic covertures along d_{COM} . Especially, the free-energy profiles for compounds C and D had steep valleys within 10 Å of d_{COM} , while those for compounds A and B had gradual valleys around 20 Å of d_{COM} , indicating that compounds C and D had higher binding affinities than compounds A and B. As a relative comparison of the binding affinities between compounds B and D, a previous study evaluated their values [48]. More specifically, the binding affinity of compound D (−29.57 kcal/mol) was higher than that of compound B (−20.62 kcal/mol). Here, compound D showed the highest binding affinity in the considered compounds [48], which reflects the covertures of the free-energy profiles of these compounds, validating compound D with the higher binding affinity had the steep valley along d_{COM} , while compound B with the lower binding affinity had the gradual valley along d_{COM} .

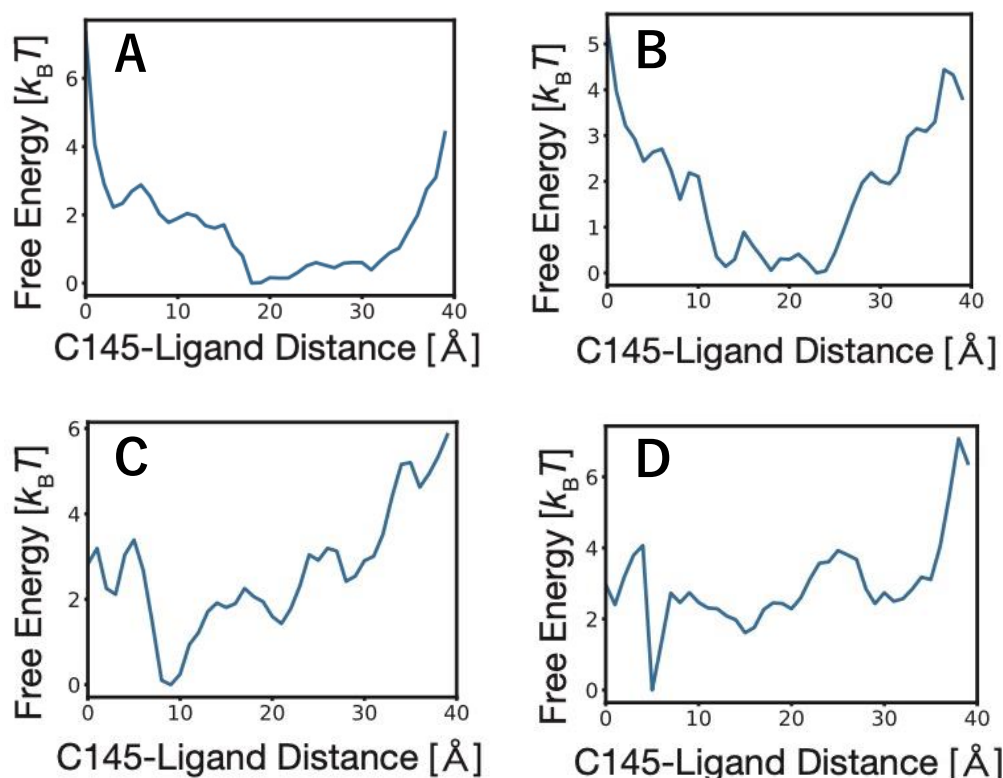


Figure 5. Free-energy profiles projected onto d_{COM} of the ligand binding pathways sampled by three LB-PaCS-MD trials for compounds A, B, C, and D. d_{COM} is defined as the COM distance between the binding site residue (C145) and each compound. (A) $\times 77$; (B) methyl rosmarinate; (C) niclosamide; (D) 5,7,3',4'-tetrahydroxy-2'-(3,3-dimethylallyl) isoflavone.

To address the interaction between the binding site and each compound, the characteristic snapshots sampled at the final (50th) cycle are shown in Figure 6. To highlight the ligand–protein interactions, the catalytic dyad residues (C145 and H41) around the binding site were depicted. The importance of these residues has been already reported [48]. Judging from Figure 6, all the compounds interacted with the thiole and imidazole groups of C145 and H41. As shown in Figure 6, compounds A and B interacted with these groups via a single site. In contrast, compounds C and D interacted with these groups via multiple sites. The difference in the interaction patterns might reflect the difference in the binding affinities of these compounds. More specifically, compound A and B tend to fluctuate due to the single site interaction, which might be an origin of the gradual valley in the free-energy profiles. Similarly, compounds C and D tend to tightly interact with the catalytic dyad residues via the multiple sites, which might be an origin of the steep valley in the free-energy profiles.

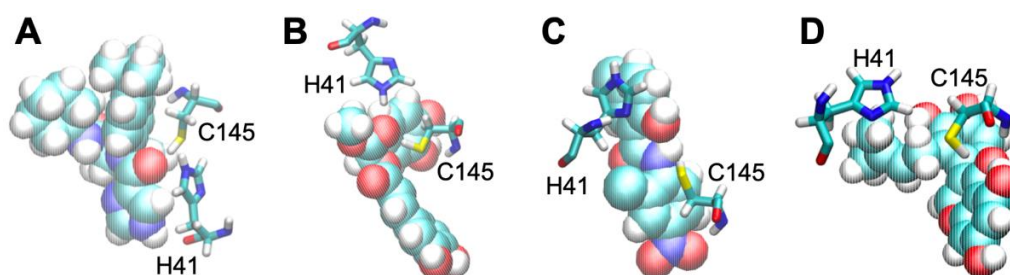


Figure 6. Ligand–protein interactions for compounds (A–D). Each compound is depicted in vdW representation. The dyad residues (H41 and C145) around the binding site are depicted in licorice representation.

In the present study, one chain of SARS-CoV2 M^{Pro} was used to perform the short-time MD simulations. As a previous study reported [66], the protein dynamics and binding pocket geometry might be different between monomer and dimer. However, we used the monomer to sample the ligand binding paths to the monomer as a simple demonstration since the dimer simulation becomes complicated and takes longer time. When simulating the dimer, LB-PaCS-MD must consider the double binding sites. Since one of us has already performed the dimer simulation with a ligand in the A chain [67,68], the dimer simulation will be a next application of LB-PaCS-MD for the purpose of *in silico* screening of compounds.

In addition, the simple protonation for the HIS residues might not reflect the physiological state rigorously. Owing to the present system setup, the SARS-CoV2 M^{Pro} monomer with the simple HIS protonation was only used as a demonstration. Therefore, it might be difficult to derive biological/pharmaceutical implications from the preset results.

4. Discussion

The ligand binding path sampling efficiency of LB-PaCS-MD might depend on the number of ligand copies to be arranged around a target protein. Generally, it is difficult to optimally arrange multiple ligand copies, meaning that the optimal number is unknown. Another concern was that the arranged ligand copies might aggregate at high concentrations, which would result in insufficient ligand binding path sampling. For this issue, weakly repulsive intermolecular interactions among the arranged ligand copies might prevent aggregation. These highly ligand-concentrated environments with the repulsive intermolecular interactions enhance ligand binding site searches on protein surfaces [69].

For an RC of LB-PaCS-MD, the COM distance between a ligand and the binding site of a target protein was specified as an RC. However, future studies should consider other RCs to characterize more complicated binding poses. In addition to d_{COM} , some RCs might be more appropriate for characterizing ligand orientations upon binding to the active site. To treat complicated ligand binding pathways, a scoring function could help to select important configurations and can be defined by multiple RCs. Ligand binding path sampling in a higher-dimensional RC subspace based on the scoring function could theoretically increase the ligand path-sampling efficiency of LB-PaCS-MD.

Here, we discuss the validity of d_{COM} as a measure when COM of a ligand is flexible. To treat the flexibility, the shape of a target ligand is characterized by its radius of gyration (R_g) or accessible surface area (SASA). These measures allow one to treat a flexible ligand rather than a rigid one. To select initial structures during LB-PaCS-MD cycles, a scoring function can be defined using R_g and SASA in addition to d_{COM} . To treat this delicate ligand binding process, both d_{COM} and R_g might be appropriate measures to rank snapshots by considering the ligand shape. The definition of a reasonable scoring function with multiple measures might be important to sample more complicated ligand-binding pathways. Figure S3 (Supplementary Materials) shows a comparison to explain a difference in the ligand-binding path sampling strategy using a single RC and multiple RCs in LB-PaCS-MD.

To quantitatively calculate the binding and dissociation constants of a ligand for a target protein, intensive sampling of both binding and unbinding pathways is required. For the dissociation path sampling, we have proposed several methods based on distributing computing [31,32]. Therefore, combinations of these dissociation sampling methods with LB-PaCS-MD allow one to intensively sample both binding and unbinding trajectories, enabling one to estimate the binding and dissociation constants quantitatively.

Instead of d_{COM} , an alternative RC might be considered to characterize intermolecular interactions between a target protein and a ligand. For example, the intermolecular interactions might be described by a distance matrix, whose matrix elements correspond to atom-to-atom distances between the ligand and protein. To select important configurations in LB-PaCS-MD, the difference between a pair of distance matrices of an instantaneous snapshot (A_{ij} , the distance between the i th and j th atoms), and the target complex (B_{ij} , the distance between the i th and j th atoms) is regarded as a reasonable measure. As an

option, these distance matrices can be locally defined for atoms around the binding site not all the atoms of the target protein complex. For example, a sum of all the elements of the difference matrix ($\Delta = \sum_{ij} A_{ij} - B_{ij}$) can be defined as a measure. In terms of Δ , LB-PaCS-MD selects snapshots with smaller Δ values as important configurations by considering the ligand–protein intermolecular interactions. According to cycles, the Δ value gradually decreases as the ligand binding process proceeds.

Judging from ligand dynamics, its diffusion might play an important role in binding to a target protein. The diffusion of the ligand is described by the mean-square-deviation (MSD) of COM of a target ligand. To consider the mobility, MSD of the ligand might be a measure in addition to d_{COM} to define a reasonable scoring function for the selections of important configurations in LB-PaCS-MD.

Next, we discuss an extension of LB-PaCS-MD for sampling other self-assembly processes. In this study, we only treated ligand–protein complex formation using T4L and SARS-CoV-2 M^{Pro}. In addition to these biological targets, LB-PaCS-MD might be applicable to the self-assembly of materials such as nanocubes of organic compounds. In a previous study, a set of dissociation processes of a carbon nanocube (hexamer) was sampled using our proposed rare-event sampling method [5,70,71]. In the sampled dissociation processes, the hexamer was decomposed into monomers. As a post-processing treatment after sampling the dissociation processes based on other rare-event sampling methods [72,73], LB-PaCS-MD might inversely sample a set of association processes from the dissociated monomers toward the self-assembled hexamer. Therefore, LB-PaCS-MD might allow one to sample a set of self-assembly processes of organic molecules toward a higher-order structure. In summary, LB-PaCS-MD can be used to elucidate several self-assembly processes, not only biological molecules, but also organic molecules.

Finally, we compare LB-PaCS-MD with other enhanced ligand sampling methods for their ligand-binding path sampling strategy. In a review paper [74], several enhanced ligand binding path sampling methods have been introduced. About the sampling of ligand binding pathways, umbrella sampling [75–77] or metadynamics and their variants [78–81] adopted external biases along a specified RC such as d_{COM} . In these biased sampling methods, RCs are specified prior to applications. In this sense, LB-PaCS-MD have a common issue of RC specifications with these biased sampling methods. For implementation, LB-PaCS-MD can be performed using a simple script, allowing one to control the conformational resampling cycles without modifying source codes and imposing external biases. For a disadvantage, LB-PaCS-MD is difficult to be applied when the ligand binding site of a target protein is unknown. Therefore, alternative methods are required to identify the binding sites prior to applications. For alternative methods, the cosolvent or mixed-solvent MD simulations are regarded as computational tools in identifying the ligand binding sites [82,83]. These methods use solvent mixtures that usually consist of organic molecules or fragments that possessed key pharmacophore features crucial for ligand binding. Once the cosolvent or mixed solvent MD simulations identified the ligand binding sites, LB-PaCS-MD intensively samples the ligand binding paths toward the hotspots.

5. Conclusions

In this study, LB-PaCS-MD, an extension of PaCS-MD, was used to frequently sample ligand binding pathways toward a target protein. Multiple ligand copies were randomly arranged around a given target protein to increase the chance of sampling ligand-binding pathways. In a ligand-concentrated environment, LB-PaCS-MD allowed for frequent ligand binding path sampling. LB-PaCS-MD specified the COM distances between each ligand and the binding site residue of the target protein, d_{COM} . The snapshots generated by LB-PaCS-MD were ranked by their d_{COM} values. In each cycle, snapshots with smaller d_{COM} values were selected as important configurations (initial structures), and they were conformationally resampled by short-time MD simulations. By repeating conformational resampling cycles, the d_{COM} value gradually decreased and converged to a constant, indicating that a set of ligand binding pathways was successfully sampled by LB-PaCS-

MD. As demonstrations, LB-PaCS-MD was applied to two proteins (T4L and SARS-CoV-2 M^{Pro}), and their ligand binding pathways were sampled more frequently than both normal PaCS-MD and cMD in the dilute environments, validating the higher performance of ligand-binding path sampling.

In distributed computing, such as LB-PaCS-MD, it is essential to select important configurations with high potential to make transitions toward a target state. The important configurations are selected based on a criterion for defining the high potential of snapshots. Generally, the criterion is non-trivial and empirically determined depending on the target system. Therefore, several measures (physical variables) are considered to define the criterion. For a future perspective, these measures might be automatically determined by a machine learning algorithm. For example, our previous study [84] adopted a machine learning like algorithm to automatically search a set of optimal measures for selecting the important configurations. When selecting the important configurations in LB-PaCS-MD, some machine learning algorithms might increase the ligand-binding sampling efficiency.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ma15041490/s1>, 1. Supplementary Materials for the T4L and SARS-CoV2 M^{Pro} Systems: Figure S1: Time series of the backbone RMSD of each protein. For each system, the trajectories generated by each final cycle (the 100th cycle for T4L and the 50th cycle for SARS-CoV2 M^{Pro}) were used to calculate the backbone RMSD values. For the RMSD calculations, each relaxed configuration (the snapshots after the 200-ps MD simulations from the experimental structures for both systems) was used as each reference. (a) T4L. (b) SARS-CoV2 M^{Pro} with compound A. (c) SARS-CoV2 M^{Pro} compound B. (d) SARS-CoV2 M^{Pro} compound C. (e) SARS-CoV2 M^{Pro} compound D.; Figure S2: (a) Ligand-binding form sampled by LB-PaCS-MD. The key residue (M102) for the ligand-binding process is depicted in vdW representation with each benzene molecule in licorice representation (blue: LB-PaCS-MD, red: the X-ray structure). (b,c) The ligand-binding forms are highlighted around the binding site with a set of key residues (M102 and F114). (top) The benzene molecule sampled by LB-PaCS-MD. (bottom) The benzene molecule in the X-ray structure.; Figure S3: Flowchart of selections of initial configurations in LB-PaCS-MD. (a) Selections of initial configurations using a single reaction coordinate (RC) and (b) those using multiple RCs in terms of a pre-defined scoring function. 2. Parameter Files for All the Organic Compounds (Benzene, Compound A, B, C, and D).

Author Contributions: Conceptualization, H.A. and R.H.; methodology, R.H.; software, H.A.; validation, H.A., Y.S. and R.H.; formal analysis, H.A.; investigation, R.H.; resources, R.H.; data curation, H.A.; writing—original draft preparation, H.A., Y.S. and R.H.; writing—review and editing, R.H.; visualization, R.H.; supervision, R.H.; project administration, R.H.; funding acquisition, Y.S. and R.H. All authors have read and agreed to the published version of the manuscript.

Funding: RH acknowledges support from Japan Society for the Promotion of Science (JSPS) KAKENHI (grant No. JP21K06094). Y.S. acknowledges support from the CREST program (grant No. JPMJCR20B3) by the Japan Agency for Science and Technology (JST) and Grant-in-aid for innovative research “Biometal Sciences” (grant No. 20H05497) by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: This research was partially supported by the Initiative on Promotion of Supercomputing for Young or Women Researchers, Information Technology Center, University of Tokyo. This research used computational resources of Cygnus provided by Multidisciplinary Cooperative Research Program in Center for Computational Sciences, University of Tsukuba.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Israelachvili, J.N.; Mitchell, D.J.; Ninham, B.W. Theory of Self-Assembly of Hydrocarbon Amphiphiles into Micelles and Bilayers. *J. Chem. Soc. Faraday Trans.* **1976**, *72*, 1525–1568. [[CrossRef](#)]
2. Whitesides, G.M.; Mathias, J.P.; Seto, C.T. Molecular Self-Assembly and Nanochemistry—A Chemical Strategy for the Synthesis of Nanostructures. *Science* **1991**, *254*, 1312–1319. [[CrossRef](#)] [[PubMed](#)]
3. Winfree, E.; Liu, F.R.; Wenzler, L.A.; Seeman, N.C. Design and self-assembly of two-dimensional DNA crystals. *Nature* **1998**, *394*, 539–544. [[CrossRef](#)] [[PubMed](#)]
4. Zhang, S.G. Fabrication of novel biomaterials through molecular self-assembly. *Nat. Biotechnol.* **2003**, *21*, 1171–1178. [[CrossRef](#)] [[PubMed](#)]
5. Harada, R.; Mashiko, T.; Tachikawa, M.; Hiraoka, S.; Shigeta, Y. Programed dynamical ordering in self-organization processes of a nanocube: A molecular dynamics study. *Phys. Chem. Chem. Phys.* **2018**, *20*, 9115–9122. [[CrossRef](#)]
6. Israelachvili, J.N.; Mitchell, D.J.; Ninham, B.W. Theory of Self-Assembly of Lipid Bilayers and Vesicles. *Biochim. Biophys. Acta* **1977**, *470*, 185–201. [[CrossRef](#)]
7. Rambaran, R.N.; Serpell, L.C. Amyloid fibrils Abnormal protein assembly. *Prion* **2008**, *2*, 112–117. [[CrossRef](#)]
8. Kushner, D.J. Self-Assembly of Biological Structures. *Bacteriol. Rev.* **1969**, *33*, 302–345. [[CrossRef](#)]
9. Ariga, K.; Kunitake, T. Molecular recognition at air-water and related interfaces: Complementary hydrogen bonding and multisite interaction. *Acc. Chem. Res.* **1998**, *31*, 371–378. [[CrossRef](#)]
10. Meyer, E.E.; Rosenberg, K.J.; Israelachvili, J. Recent progress in understanding hydrophobic interactions. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 15739–15746. [[CrossRef](#)]
11. Ewing, T.J.A.; Kuntz, I.D. Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.* **1997**, *18*, 1175–1189. [[CrossRef](#)]
12. Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748. [[CrossRef](#)] [[PubMed](#)]
13. Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; et al. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749. [[CrossRef](#)] [[PubMed](#)]
14. Trott, O.; Olson, A.J. Software News and Update 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.
15. Shan, Y.B.; Kim, E.T.; Eastwood, M.P.; Dror, R.O.; Seeliger, M.A.; Shaw, D.E. How Does a Drug Molecule Find Its Target Binding Site? *J. Am. Chem. Soc.* **2011**, *133*, 9181–9183. [[CrossRef](#)]
16. Pan, A.C.; Xu, H.F.; Palpant, T.; Shaw, D.E. Quantitative Characterization of the Binding and Unbinding of Millimolar Drug Fragments with Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2017**, *13*, 3372–3377. [[CrossRef](#)]
17. Shaw, D.E.; Deneroff, M.M.; Dror, R.O.; Kuskin, J.S.; Larson, R.H.; Salmon, J.K.; Young, C.; Batson, B.; Bowers, K.J.; Chao, J.C.; et al. Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM* **2008**, *51*, 91–97. [[CrossRef](#)]
18. Klepeis, J.L.; Lindorff-Larsen, K.; Dror, R.O.; Shaw, D.E. Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.* **2009**, *19*, 120–127. [[CrossRef](#)]
19. Shaw, D.E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R.O.; Eastwood, M.P.; Bank, J.A.; Jumper, J.M.; Salmon, J.K.; Shan, Y.B.; et al. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330*, 341–346. [[CrossRef](#)]
20. Lindorff-Larsen, K.; Piana, S.; Dror, R.O.; Shaw, D.E. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517–520. [[CrossRef](#)]
21. Senne, M.; Trendelkamp-Schroer, B.; Mey, A.S.J.S.; Schutte, C.; Noe, F. EMMA: A Software Package for Markov Model Building and Analysis. *J. Chem. Theory Comput.* **2012**, *8*, 2223–2238. [[CrossRef](#)] [[PubMed](#)]
22. Scherer, M.K.; Trendelkamp-Schroer, B.; Paul, F.; Perez-Hernandez, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.H.; Noe, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015**, *11*, 5525–5542. [[CrossRef](#)] [[PubMed](#)]
23. Gervasio, F.L.; Laio, A.; Parrinello, M. Flexible docking in solution using metadynamics. *J. Am. Chem. Soc.* **2005**, *127*, 2600–2607. [[CrossRef](#)]
24. Barducci, A.; Chelli, R.; Procacci, P.; Schettino, V.; Gervasio, F.L.; Parrinello, M. Metadynamics simulation of prion protein: Beta-structure stability and the early stages of misfolding. *J. Am. Chem. Soc.* **2006**, *128*, 2705–2710. [[CrossRef](#)] [[PubMed](#)]
25. Raiteri, P.; Laio, A.; Gervasio, F.L.; Micheletti, C.; Parrinello, M. Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *J. Phys. Chem. B* **2006**, *110*, 3533–3539. [[CrossRef](#)] [[PubMed](#)]
26. Yang, Y.I.; Zhang, J.; Che, X.; Yang, L.J.; Gao, Y.Q. Efficient sampling over rough energy landscapes with high barriers: A combination of metadynamics with integrated tempering sampling. *J. Chem. Phys.* **2016**, *144*, 094105. [[CrossRef](#)]
27. Bernetti, M.; Masetti, M.; Recanatini, M.; Amaro, R.E.; Cavalli, A. An Integrated Markov State Model and Path Metadynamics Approach to Characterize Drug Binding Processes. *J. Chem. Theory Comput.* **2019**, *15*, 5689–5702. [[CrossRef](#)]
28. Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151. [[CrossRef](#)]
29. Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **2000**, *113*, 6042–6051. [[CrossRef](#)]

30. Okamoto, Y. Generalized-ensemble algorithms: Enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. *J. Mol. Graph. Model.* **2004**, *22*, 425–439. [[CrossRef](#)]
31. Harada, R.; Takano, Y.; Baba, T.; Shigeta, Y. Simple, yet powerful methodologies for conformational sampling of proteins. *Phys. Chem. Chem. Phys.* **2015**, *17*, 6155–6173. [[CrossRef](#)] [[PubMed](#)]
32. Harada, R. Simple, yet Efficient Conformational Sampling Methods for Reproducing/Predicting Biologically Rare Events of Proteins. *Bull. Chem. Soc. Jpn.* **2018**, *91*, 1436–1450. [[CrossRef](#)]
33. Aida, H.; Shigeta, Y.; Harada, R. Regenerations of Initial Velocities in Parallel Cascade Selection Molecular Dynamics (PaCS-MD) Enhance the Conformational Transitions of Proteins. *Chem. Lett.* **2020**, *49*, 798–801. [[CrossRef](#)]
34. Harada, R.; Kitao, A. Parallel cascade selection molecular dynamics (PaCS-MD) to generate conformational transition pathway. *J. Chem. Phys.* **2013**, *139*, 035103. [[CrossRef](#)]
35. Harada, R.; Kitao, A. Nontargeted Parallel Cascade Selection Molecular Dynamics for Enhancing the Conformational Sampling of Proteins. *J. Chem. Theory Comput.* **2015**, *11*, 5493–5502. [[CrossRef](#)] [[PubMed](#)]
36. Harada, R.; Shigeta, Y. Temperature-Shuffled Parallel Cascade Selection Molecular Dynamics Accelerates the Structural Transitions of Proteins. *J. Comput. Chem.* **2017**, *38*, 2671–2674. [[CrossRef](#)]
37. Harada, R.; Shigeta, Y. Dynamic Specification of Initial Structures in Parallel Cascade Selection Molecular Dynamics (PaCS-MD) Efficiently Promotes Biologically Relevant Rare Events. *Bull. Chem. Soc. Jpn.* **2017**, *90*, 1236–1243. [[CrossRef](#)]
38. Harada, R.; Shigeta, A. On-the-Fly Specifications of Reaction Coordinates in Parallel Cascade Selection Molecular Dynamics Accelerate Conformational Transitions of Proteins. *J. Chem. Theory Comput.* **2018**, *14*, 3332–3341. [[CrossRef](#)]
39. Harada, R.; Sladek, V.; Shigeta, Y. Nontargeted Parallel Cascade Selection Molecular Dynamics Using Time-Localized Prediction of Conformational Transitions in Protein Dynamics. *J. Chem. Theory Comput.* **2019**, *15*, 5144–5153. [[CrossRef](#)]
40. Baba, T.; Harada, R.; Nakano, M.; Shigeta, Y. On the Induced-Fit Mechanism of Substrate-Enzyme Binding Structures of Nylon-Oligomer Hydrolase. *J. Comput. Chem.* **2014**, *35*, 1240–1247. [[CrossRef](#)]
41. Kitao, A.; Harada, R.; Nishihara, Y.; Tran, D.P. Parallel Cascade Selection Molecular Dynamics for Efficient Conformational Sampling and Free Energy Calculation of Proteins. *AIP Conf. Proc.* **2016**, *1790*, 020013.
42. Fujita, J.; Harada, R.; Maeda, Y.; Saito, Y.; Mizohata, E.; Inoue, T.; Shigeta, Y.; Matsumura, H. Identification of the key interactions in structural transition pathway of FtsZ from *Staphylococcus aureus*. *J. Struct. Biol.* **2017**, *198*, 65–73. [[CrossRef](#)] [[PubMed](#)]
43. Yasuda, T.; Shigeta, Y.; Harada, R. Efficient Conformational Sampling of Collective Motions of Proteins with Principal Component Analysis-Based Parallel Cascade Selection Molecular Dynamics. *J. Chem. Inf. Model.* **2020**, *60*, 4021–4029. [[CrossRef](#)] [[PubMed](#)]
44. Chodera, J.D.; Noe, F. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **2014**, *25*, 135–144. [[CrossRef](#)] [[PubMed](#)]
45. Husic, B.E.; Pande, V.S. Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* **2018**, *140*, 2386–2396. [[CrossRef](#)]
46. Liu, L.J.; Baase, W.A.; Matthews, B.W. Halogenated Benzenes Bound within a Non-polar Cavity in T4 Lysozyme Provide Examples of I center dot center dot center dot S and I center dot center dot center dot Se Halogen-bonding. *J. Mol. Biol.* **2009**, *385*, 595–605. [[CrossRef](#)]
47. Sahu, S.N.; Mishra, B.; Sahu, R.; Pattanayak, S.K. Molecular dynamics simulation perception study of the binding affinity performance for main protease of SARS-CoV-2. *J. Biomol. Struct. Dyn.* **2020**, *23*, 1–16. [[CrossRef](#)]
48. ul Qamar, M.T.; Alqahtani, S.M.; Alamri, M.A.; Chen, L.L. Structural basis of SARS-CoV-2 3CL(pro) and anti-COVID-19 drug discovery from medicinal plants. *J. Pharm. Anal.* **2020**, *10*, 313–319. [[CrossRef](#)]
49. Andrianov, A.M.; Kornoushenko, Y.V.; Karpenko, A.D.; Bosko, I.P.; Tuzikov, A.V. Computational discovery of small drug-like compounds as potential inhibitors of SARS-CoV-2 main protease. *J. Biomol. Struct. Dyn.* **2021**, *39*, 5779–5791. [[CrossRef](#)]
50. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935. [[CrossRef](#)]
51. Hess, B.; Bekker, H.; Berendsen, H.J.C.; Fraaije, J.G.E.M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472. [[CrossRef](#)]
52. Miyamoto, S.; Kollman, P.A. Settle—An Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952–962. [[CrossRef](#)]
53. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101. [[CrossRef](#)] [[PubMed](#)]
54. Parrinello, M.; Rahman, A. Crystal-Structure and Pair Potentials—A Molecular-Dynamics Study. *Phys. Rev. Lett.* **1980**, *45*, 1196–1199. [[CrossRef](#)]
55. Parrinello, M.; Rahman, A. Polymorphic Transitions in Single-Crystals—A New Molecular-Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190. [[CrossRef](#)]
56. Essmann, U.; Perera, L.; Berkowitz, M.L.; Darden, T.; Lee, H.; Pedersen, L.G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593. [[CrossRef](#)]
57. Abraham, M.J.; van der Spoel, D.; Lindahl, E.; Hess, B.; The GROMACS Development Team. GROMACS User Manual Version 2018. 2018. Available online: <https://manual.gromacs.org/documentation/2018/download.html> (accessed on 10 January 2018).
58. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. [[CrossRef](#)]

59. Harada, R.; Takano, Y.; Shigeta, Y. Fluctuation Flooding Method (FFM) for accelerating conformational transitions of proteins. *J. Chem. Phys.* **2014**, *140*, 125103. [[CrossRef](#)]
60. Harada, R.; Takano, Y.; Shigeta, Y. Efficient conformational sampling of proteins based on a multi-dimensional TaBoo SeArch algorithm: An application to folding of chignolin in explicit solvent. *Chem. Phys. Lett.* **2015**, *630*, 68–75. [[CrossRef](#)]
61. Harada, R.; Takano, Y.; Shigeta, Y. TaBoo SeArch Algorithm with a Modified Inverse Histogram for Reproducing Biologically Relevant Rare Events of Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 2436–2445. [[CrossRef](#)]
62. Harada, R.; Shigeta, Y. Efficient Conformational Search Based on Structural Dissimilarity Sampling: Applications for Reproducing Structural Transitions of Proteins. *J. Chem. Theory Comput.* **2017**, *13*, 1411–1423. [[CrossRef](#)] [[PubMed](#)]
63. Harada, R.; Shigeta, Y. Self-Avoiding Conformational Sampling Based on Histories of Past Conformational Searches. *J. Chem. Inf. Model.* **2017**, *57*, 3070–3078. [[CrossRef](#)] [[PubMed](#)]
64. Mulder, F.A.A.; Mittermaier, A.; Hon, B.; Dahlquist, F.W.; Kay, L.E. Studying excited states of proteins by NMR spectroscopy. *Nat. Struct. Biol.* **2001**, *8*, 932–935. [[CrossRef](#)] [[PubMed](#)]
65. Mondal, J.; Ahalawat, N.; Pandit, S.; Kay, L.E.; Vallurupalli, P. Atomic resolution mechanism of ligand binding to a solvent inaccessible cavity in T4 lysozyme. *PLoS Comput. Biol.* **2018**, *14*, e1006180. [[CrossRef](#)]
66. Tekpinar, M.; Yildirim, A. Impact of dimerization and N3 binding on molecular dynamics of SARS-CoV and SARS-CoV-2 main proteases. *J. Biomol. Struct. Dyn.* **2021**, 1–12. [[CrossRef](#)]
67. Nutho, B.; Mahalapbutr, P.; Hengphasatporn, K.; Pattarangoon, N.C.; Simanon, N.; Shigeta, Y.; Hannongbua, S.; Rungrotmongkol, T. Why Are Lopinavir and Ritonavir Effective against the Newly Emerged Coronavirus 2019? Atomistic Insights into the Inhibitory Mechanisms. *Biochemistry* **2020**, *59*, 1769–1779. [[CrossRef](#)] [[PubMed](#)]
68. Deetanya, P.; Hengphasatporn, K.; Wilasluck, P.; Shigeta, Y.; Rungrotmongkol, T.; Wangkanont, K. Interaction of 8-anilino-naphthalene-1-sulfonate with SARS-CoV-2 main protease and its application as a fluorescent probe for inhibitor identification. *Comput. Struct. Biotechnol.* **2021**, *19*, 3364–3371. [[CrossRef](#)]
69. Takemura, K.; Sato, C.; Kitao, A. ColDock: Concentrated Ligand Docking with All-Atom Molecular Dynamics Simulation. *J. Phys. Chem. B* **2018**, *122*, 7191–7200. [[CrossRef](#)] [[PubMed](#)]
70. Harada, R.; Nakamura, T.; Takano, Y.; Shigeta, Y. Protein Folding Pathways Extracted by OFLOOD: Outlier FLOODing Method. *J. Comput. Chem.* **2015**, *36*, 97–102. [[CrossRef](#)] [[PubMed](#)]
71. Harada, R.; Nakamura, T.; Shigeta, Y. Sparsity-weighted Outlier FLOODing (OFLOOD) method: Efficient rare event sampling method using sparsity of distribution. *J. Comput. Chem.* **2016**, *37*, 724–738. [[CrossRef](#)]
72. Harada, R.; Nakamura, T.; Shigeta, Y. A Fast Convergent Simulated Annealing Algorithm for Protein-Folding: Simulated Annealing Outlier FLOODing (SA-OFLOOD) Method. *Bull. Chem. Soc. Jpn.* **2016**, *89*, 1361–1367. [[CrossRef](#)]
73. Harada, R.; Yoshino, R.; Nishizawa, H.; Shigeta, Y. Temperature-pressure shuffling outlier flooding method enhances the conformational sampling of proteins. *J. Comput. Chem.* **2019**, *40*, 1530–1537. [[CrossRef](#)] [[PubMed](#)]
74. Lazim, R.; Suh, D.; Choi, S. Advances in Molecular Dynamics Simulations and Enhanced Sampling Methods for the Study of Protein Systems. *Int. J. Mol. Sci.* **2020**, *21*, 6339. [[CrossRef](#)] [[PubMed](#)]
75. Kokubo, H.; Tanaka, T.; Okamoto, Y. Prediction of Protein-Ligand Binding Structures by Replica-Exchange Umbrella Sampling Simulations: Application to Kinase Systems. *J. Chem. Theory Comput.* **2013**, *9*, 4660–4671. [[CrossRef](#)]
76. Ngo, S.T.; Vu, K.B.; Bui, L.M.; Vu, V.V. Effective Estimation of Ligand-Binding Affinity Using Biased Sampling Method. *ACS Omega* **2019**, *4*, 3887–3893. [[CrossRef](#)]
77. Ngo, S.T. Estimating the ligand-binding affinity via lambda-dependent umbrella sampling simulations. *J. Comput. Chem.* **2021**, *42*, 117–123. [[CrossRef](#)]
78. Capelli, R.; Carloni, P.; Parrinello, M. Exhaustive Search of Ligand Binding Pathways via Volume-Based Metadynamics. *J. Phys. Chem. Lett.* **2019**, *10*, 3495–3499. [[CrossRef](#)]
79. Ibrahim, P.; Clark, T. Metadynamics simulations of ligand binding to GPCRs. *Curr. Opin. Struct. Biol.* **2019**, *55*, 129–137. [[CrossRef](#)]
80. Provasi, D. Ligand-Binding Calculations with Metadynamics. *Biomol. Simul. Methods Protoc.* **2019**, 2022, 233–253.
81. Raniolo, S.; Limongelli, V. Ligand binding free-energy calculations with funnel metadynamics. *Nat. Protoc.* **2020**, *15*, 2837–2866. [[CrossRef](#)]
82. Kimura, S.R.; Hu, H.P.; Ruvinsky, A.M.; Sherman, W.; Favia, A.D. Deciphering Cryptic Binding Sites on Proteins by Mixed-Solvent Molecular Dynamics. *J. Chem. Inf. Model.* **2017**, *57*, 1388–1401. [[CrossRef](#)] [[PubMed](#)]
83. Zariquiey, F.S.; de Souza, J.V.; Bronowska, A.K. Cosolvent Analysis Toolkit (CAT): A robust hotspot identification platform for cosolvent simulations of proteins to expand the druggable proteome. *Sci. Rep.* **2019**, *9*, 19118. [[CrossRef](#)] [[PubMed](#)]
84. Harada, R.; Nakamura, T.; Shigeta, Y. Automatic detection of hidden dimensions to obtain appropriate reaction coordinates in the Outlier FLOODing (OFLOOD) method. *Chem. Phys. Lett.* **2015**, *639*, 269–274. [[CrossRef](#)]