

**Ligand Gaussian accelerated molecular dynamics 2 (LiGaMD2):
Improved calculations of ligand binding thermodynamics and
kinetics with closed protein pocket**

Jinan Wang and Yinglong Miao*

Center for Computational Biology and Department of Molecular Biosciences, University of
Kansas, Lawrence, KS 66047, USA

* Corresponding email: miao@ku.edu

Abstract

Ligand binding thermodynamics and kinetics are critical parameters for drug design. However, it has proven challenging to efficiently predict ligand binding thermodynamics and kinetics from molecular simulations due to limited simulation timescales. Protein dynamics especially in the ligand binding pocket often plays an important role in ligand binding. Based on our previously developed Ligand Gaussian accelerated molecular dynamics (LiGaMD), here we present LiGaMD2 in which a selective boost potential was applied to both the ligand and protein residues in the binding pocket to improve sampling of ligand binding and dissociation. To validate the performance of LiGaMD2, the T4 lysozyme (T4L) mutants with open and closed pockets bound by different ligands were chosen as model systems. LiGaMD2 could efficiently capture repetitive ligand dissociation and binding within microsecond simulations of all T4L systems. The obtained ligand binding kinetic rates and free energies agreed well with available experimental values and previous modeling results. Therefore, LiGaMD2 provides an improved approach to sample opening of closed protein pockets for ligand dissociation and binding, thereby allowing for efficient calculations of ligand binding thermodynamics and kinetics.

Introduction

ligand binding to target receptors plays a critical role in many fundamental biological processes¹, as well as in the design of more effective and selective drugs for treating human diseases². A number of experimental techniques³ have been developed to explore protein-small molecule interactions. For example, structural biology techniques^{3b} have been widely applied to determine protein-ligand complex structures. However, X-ray crystallography and cryo-electron microscopy (cryo-EM) could provide only static snapshots of protein-small molecule interactions. It is rather challenging for experimental methods to capture ligand binding and dissociation pathways and determine potential intermediate states of ligand binding to the protein target site.

Recently, ligand binding kinetics has been recognized to be critical for drug design⁴. The ligand dissociation rate (k_{off}) or residence time ($1/k_{off}$) appears to correlate with drug efficacy better than ligand binding free energy. However, ligand binding kinetic rates have proven more challenging to predict, due to slow processes of ligand dissociation and rebinding^{4b, 5}. With remarkable advances in computer hardware and method developments, conventional molecular dynamics (cMD) simulations nowadays are able to capture spontaneous ligand binding to target proteins and predict corresponding binding association rates (k_{on})⁶. However, it remains challenging to use cMD to simulate repetitive ligand binding and dissociation processes, precluding accurate prediction of ligand binding kinetic rates. For example, based on tens-of-microsecond cMD simulations, Shan et al.⁷ successfully captured spontaneous binding of the Dasatinib drug to its target Src kinase and accurately predicted the ligand association rate (k_{on}). However, no dissociation event was observed in the cMD simulations. Similar results were observed in the binding of benzene (BEN) to the L99A mutant of T4 lysozyme (T4L)⁸. The cMD simulations with lengths of 2 to 8 μ s successfully captured BEN binding to the L99A T4L⁸. Tens

to hundreds of repetitive guest binding and dissociation from the β -CD host were observed in microsecond cMD simulations⁹, which enabled accurate prediction of host-guest binding thermodynamics and kinetics. Tens-of-microsecond cMD simulations^{6d} captured repetitive binding and dissociation of six small-molecule fragments with weak millimolar binding affinities to the protein FKBP, which enabled accurate prediction of fragment binding free energies and kinetic rates. Nevertheless, cMD simulations have not captured repetitive binding and dissociation of typical small-molecule ligands to target proteins so far.

In this regard, enhanced sampling methods¹⁰ have been developed to extend accessible timescale of MD simulations, including Metadynamics¹¹, Steered MD¹², Umbrella Sampling^{12a, 13}, Replica Exchange MD¹⁴, Random Acceleration Molecular Dynamics (RAMD)¹⁵, Scaled MD¹⁶, accelerated MD (aMD)¹⁷, Gaussian accelerated MD (GaMD)¹⁸, Markov State Model (MSM)^{8, 19}, Weighted Ensemble²⁰, and so on. Using T4L as a model system, a total of $\sim 12 \mu\text{s}$ infrequent Metadynamics^{11a, 11b} simulations successfully captured 20 times of ligand binding and dissociation events and predicted the values of k_{on} and k_{off} at $3.5 \pm 2 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$ and $7 \pm 2 \text{ s}^{-1}$, being comparable to experimental values of $0.8 \pm 1 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$ and $950 \pm 20 \text{ s}^{-1}$, respectively. However, Metadynamics requires predefined collective variables (CVs) before running the simulations. Thus, *a priori* knowledge of the systems is often required. In comparison, Weighted Ensemble²¹ and MSM²² combine a large number of short cMD simulations to predict ligand binding kinetic rates. For example, Weighted Ensemble^{20b} of a total of $29 \mu\text{s}$ cMD was able to accurately predict the dissociation rate of BEN from the L99A T4L as 1000 s^{-1} , being highly consistent with the experimental value of $950 \pm 200 \text{ s}^{-1}$. One MSM⁸ built on a total of $69 \mu\text{s}$ cMD simulations predicted the values of k_{on} and k_{off} at $21 \pm 91 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$ and $311 \pm 130 \text{ s}^{-1}$, being in reasonable agreement with

experiment values of $0.8 \pm 1 \times 10^6 \text{M}^{-1} \text{s}^{-1}$ and $950 \pm 20 \text{ s}^{-1}$, respectively. However, these calculations required very expensive computational resources.

GaMD was developed to provide both unconstrained enhanced sampling and free energy calculations of large biomolecules¹⁸. It works by applying a harmonic boost potential to reduce system energy barriers. The boost potential normally exhibits a near Gaussian distribution, which enables proper reweighting of the free energy profiles through cumulant expansion to the second order¹⁸. Ligand GaMD (LiGaMD)²³ has recently been developed to allow for more efficient sampling of ligand dissociation and rebinding processes, being able to accurately predict ligand binding thermodynamics and kinetics. In LiGaMD, one selective boost was applied to the ligand non-bonded interaction potential energy to accelerate ligand dissociation. Another boost was applied on the remaining potential to facilitate ligand rebinding. Recently, an increasing number of studies suggested that the flexibility of proteins, particularly those with closed pockets, played an important role in ligand binding^{8, 24}. Protein structural flexibility allows the opening, closing, and adaptation of the binding pocket, which are critical for ligand binding to protein^{8, 25}. Additionally, enhanced sampling of the protein binding pocket has proven to significantly improve the efficiency and accuracy of simulating protein-ligand interaction^{14d, 26}. Wang et al.²⁶ proposed FEP/REST to improve the accuracy of ligand binding free energy calculation, which combined the free energy perturbation (FEP) and replica exchange with solute tempering (REST) to enhance sampling of protein residues in the binding site. FEP/REST was demonstrated to achieve much quicker simulation convergence and more accurate binding free energy calculations. Sugita et al.^{14d} developed the generalized replica exchange with solute tempering (gREST) to simulate small molecule binding to the L99A T4L. By enhanced sampling of ligand and protein residues in the binding site, the gREST simulations successfully captured ligand binding to the L99A T4L in a

total of 2.4 μ s simulations. The obtained free energy profiles of binders (BEN, ethylbenzene, and n-hexylbenzene) were distinct from those of nonbinders (phenol and benzaldehyde). Another study^{25a} from the same group combined the gREST with the replica exchange umbrella sampling (gREST/REUS) to capture the binding of compound PP1 to its target Src kinase. The gREST/REUS simulations^{25a} could capture multiple ligand binding and unbinding events.

Built on our previously developed LiGaMD, here we developed a novel approach termed LiGaMD2, in which a selective boost was added to both the ligand and protein pocket residues to accelerate ligand dissociation and binding. Various T4L mutants bound by BEN and indole (IND) with open and closed pockets were chosen as testing systems. The T4L mutants with small hydrophobic cavities that can accommodate different ligands have been widely used as model systems for benchmarking computational methods²⁷. Microsecond LiGaMD2 simulations could capture repetitive ligand binding and unbinding processes in all the simulated T4L systems. LiGaMD2 thus enabled highly efficient and accurate prediction of ligand binding thermodynamics and kinetics. Simulation predictions agreed well with experimental binding free energy and kinetic rates. Since the chosen systems included different ligands and protein mutants with distinct binding pockets, the simulations validated the robustness of LiGaMD2. Furthermore, multiple ligand binding and dissociation pathways were identified by LiGaMD2 simulations, being highly consistent with published simulation results^{8, 15a, 20b, 28}.

Methods

LiGaMD2: Selectively boosting both the ligand and protein pocket

LiGaMD is an enhanced sampling technique for characterizing ligand binding thermodynamics and kinetics. It works by adding a selective harmonic boost potential to the non-bonded ligand interaction potential energy. Detail of the LiGaMD method has been described in our previous study²⁹. Here, we briefly describe the algorithm for further development of the LiGaMD2 method.

We consider a system of ligand L binding to a protein P in a biological environment E . The system comprises of N atoms with their coordinates $r \equiv \{\vec{r}_1, \dots, \vec{r}_N\}$ and momenta $p \equiv \{\vec{p}_1, \dots, \vec{p}_N\}$. The system Hamiltonian can be expressed as:

$$H(r, p) = K(p) + V(r), \quad (1)$$

where $K(p)$ and $V(r)$ are the system kinetic and total potential energies, respectively. The protein P could be further divided into two parts: residues in the binding pocket (Pb) and other parts (Po).

We decompose the potential energy into the following terms:

$$\begin{aligned} V(r) = & V_{Pb,b}(r_{Pb}) + V_{Po,b}(r_{Po}) + V_{PbPo,b}(r_{PbPo}) + V_{L,b}(r_L) + V_{E,b}(r_E) \\ & + V_{PbPb,nb}(r_{Pb}) + V_{PoPo,nb}(r_{Po}) + V_{LL,nb}(r_L) + V_{EE,nb}(r_E) \\ & + V_{PbL,nb}(r_{PbL}) + V_{PbE,nb}(r_{PbE}) + V_{PbPo,nb}(r_{PbPo}) \\ & + V_{PoL,nb}(r_{PoL}) + V_{PoE,nb}(r_{PoE}) + V_{LE,nb}(r_{LE}) \end{aligned} \quad (2)$$

where $V_{Pb,b}$, $V_{Po,b}$, $V_{L,b}$ and $V_{E,b}$ are the bonded potential energies in the binding pocket of protein Pb, the remaining atoms of protein Po, ligand L and environment E , respectively. $V_{PbPo,b}(r_{PbPo})$ is the bonded potential energies involving atoms between the protein binding pocket and the other parts. $V_{PbPb,nb}$, $V_{PoPo,nb}$, $V_{LL,nb}$ and $V_{EE,nb}$ are the self non-bonded potential energies in the protein binding pocket Pb, the remaining atoms of protein Po, ligand L and environment E , respectively. $V_{PbPo,nb}(r_{PbPo})$, $V_{PbL,nb}(r_{PbL})$, $V_{PbE,nb}(r_{PbE})$, $V_{PoL,nb}(r_{PoL})$, $V_{PoE,nb}(r_{PoE})$ and

$V_{LE,nb}$ are the non-bonded interaction energies between Pb-Po, Pb-L, Pb-E, Po-L, Po-E and L-E, respectively. According to classical molecular mechanics force fields³⁰, the non-bonded potential energies are usually calculated as:

$$V_{nb} = V_{elec} + V_{vdW}, \quad (3)$$

where V_{elec} and V_{vdW} are the system electrostatic and van der Waals potential energies. The bonded potential energies are usually calculated as

$$V_b = V_{bond} + V_{angle} + V_{dihedral} \quad (4)$$

where V_{bond} , V_{angle} and $V_{dihedral}$ are the system bond, angle and dihedral potential energies. As mentioned above, flexibility of the protein pocket often plays a critical role in ligand binding. Therefore, the ligand essential potential energy in the LiGaMD2 is defined as

$$V_L(r) = V_{L,b}(r_L) + V_{Pb,b}(r_{Pb}) + V_{LL,nb}(r_L) + V_{Pb,nb}(r_{Pb}) + V_{PbL,nb}(r_{PbL}) + V_{PoL,nb}(r_{PoL}) + \\ V_{LE,nb}(r_{LE}) + V_{PbPo,nb}(r_{PbPo}) + V_{PbE,nb}(r_{PbE}) \quad (5)$$

In the Pd-GaMD, we add boost potential selectively to the ligand essential potential energy according to the GaMD algorithm:

$$\Delta V_L(r) = \begin{cases} \frac{1}{2} k_L (E_L - V_L(r))^2, & V_L(r) < E_L \\ 0, & V_L(r) \geq E_L, \end{cases} \quad (6)$$

where E_L is the threshold energy for applying boost potential and k_L is the harmonic constant. The LiGaMD2 simulation parameters are derived similarly as in the previous LiGaMD. When E is set to the lower bound as the system maximum potential energy ($E=V_{max}$), the effective harmonic force constant k_0 can be calculated as:

$$k_0 = \min(1.0, k'_0) = \min\left(1.0, \frac{\sigma_0}{\sigma_V} \frac{V_{max} - V_{min}}{V_{max} - V_{avg}}\right), \quad (7)$$

where V_{max} , V_{min} , V_{avg} and σ_V are the maximum, minimum, average and standard deviation of the boosted system potential energy, and σ_0 is the user-specified upper limit of the standard deviation of ΔV (e.g., $10k_B T$) for proper reweighting. The harmonic constant is calculated as $k = k_0 \cdot$

$\frac{1}{V_{max} - V_{min}}$ with $0 < k_0 \leq 1$. Alternatively, when the threshold energy E is set to its upper bound

$E = V_{min} + \frac{1}{k}$, k_0 is set to:

$$k_0 = k_0'' \equiv \left(1 - \frac{\sigma_0}{\sigma_V}\right) \frac{V_{max} - V_{min}}{V_{avg} - V_{min}}, \quad (8)$$

if k_0'' is found to be between 0 and 1. Otherwise, k_0 is calculated using Eqn. (7).

In addition to selectively boosting the ligand and surrounding protein residues in the binding site, another boost potential is applied on the protein and solvent to enhance conformational sampling of the protein and facilitate ligand rebinding. The second boost potential is calculated using the total system potential energy other than the ligand essential potential energy as:

$$\Delta V_D(r) = \begin{cases} \frac{1}{2} k_D (E_D - V_D(r))^2, & V_D(r) < E_D \\ 0, & V_D(r) \geq E_D \end{cases} \quad (9)$$

Where V_D is the total system potential energy other than the ligand essential potential energy, E_D is the corresponding threshold energy for applying the second boost potential and k_D is the harmonic constant. This leads to dual-boost LiGaMD2 with the total boost potential $\Delta V(r) = \Delta V_L(r) + \Delta V_D(r)$.

Energetic Reweighting of LiGaMD2

To calculate potential of mean force (PMF)³¹ from LiGaMD2 simulations, the probability distribution along a reaction coordinate is written as $p^*(A)$. Given the boost potential $\Delta V(\vec{r})$ of each frame, $p^*(A)$ can be reweighted to recover the canonical ensemble distribution, $p(A)$, as:

$$p(A_j) = p^*(A_j) \frac{\langle e^{\beta \Delta V(\vec{r})} \rangle_j}{\sum_{i=1}^M \langle p^*(A_i) e^{\beta \Delta V(\vec{r})} \rangle_i}, \quad j = 1, \dots, M, \quad (10)$$

where M is the number of bins, $\beta = k_B T$ and $\langle e^{\beta \Delta V(\vec{r})} \rangle_j$ is the ensemble-averaged Boltzmann factor of $\Delta V(\vec{r})$ for simulation frames found in the j^{th} bin. The ensemble-averaged reweighting factor can be approximated using cumulant expansion:

$$\langle e^{\beta \Delta V(\vec{r})} \rangle = \exp \left\{ \sum_{k=1}^{\infty} \frac{\beta^k}{k!} C_k \right\}, \quad (11)$$

where the first two cumulants are given by

$$\begin{aligned} C_1 &= \langle \Delta V \rangle, \\ C_2 &= \langle \Delta V^2 \rangle - \langle \Delta V \rangle^2 = \sigma_v^2. \end{aligned} \quad (12)$$

The boost potential obtained from LiGaMD2 simulations usually follows near-Gaussian distribution. Cumulant expansion to the second order thus provides a good approximation for computing the reweighting factor^{18b, 32}. The reweighted free energy $F(A) = -k_B T \ln p(A)$ is calculated as:

$$F(A) = F^*(A) - \sum_{k=1}^2 \frac{\beta^k}{k!} C_k + F_c, \quad (13)$$

where $F^*(A) = -k_B T \ln p^*(A)$ is the modified free energy obtained from LiGaMD2 simulation and F_c is a constant.

Ligand binding free energy calculations from 3D potential of mean force

We calculate ligand binding free energy from 3D potential of mean force (PMF) of ligand displacements from the target protein as the following³³:

$$\Delta G^0 = -\Delta W_{3D} - RT \ln \frac{V_b}{V_0}, \quad (14)$$

where V_0 is the standard volume, $V_b = \int_b e^{-\beta W(r)} dr$ is the average sampled bound volume of the ligand with $\beta = 1/k_B T$, k_B is the Boltzmann constant, T is the temperature, and ΔW_{3D} is the depth of the 3D PMF. ΔW_{3D} can be calculated by integrating Boltzmann distribution of the 3D PMF $W(r)$ over all system coordinates except the x, y, z of the ligand:

$$\Delta W_{3D} = -RT \ln \frac{\int_u e^{-\beta W(r)} dr}{\int_u dr}, \quad (15)$$

where $V_u = \int_u dr$ is the sampled unbound volume of the ligand. The exact definitions of the bound and unbound volumes V_b and V_u are not important as the exponential average cut off contributions far away from the PMF minima^{33b}. A python script “PyReweighting-3D.py” in the *PyReweighting* tool kit (<http://miao.compbio.ku.edu/PyReweighting/>)³⁴ was applied for reweighting LiGaMD2 simulations to calculate the 3D reweighted PMF and associated ligand binding free energies.

Ligand binding kinetics obtained from reweighting of LiGaMD2 Simulations

Reweighting of ligand binding kinetics from LiGaMD2 simulations followed a similar protocol using Kramers’ rate theory that has been recently implemented in kinetics reweighting of the GaMD³⁵, LiGaMD^{34b}, Pep-GaMD²⁹ and PPI-GaMD³⁶ simulations. Provided sufficient sampling of repetitive ligand dissociation and binding in the simulations, we record the time periods and

calculate their averages for the ligand found in the bound (τ_B) and unbound (τ_U) states from the simulation trajectories. The τ_B corresponds to residence time in drug design³⁷. Then the ligand dissociation and binding rate constants (k_{off} and k_{on}) were calculated as:

$$k_{off} = \frac{1}{\tau_B}. \quad (16)$$

$$k_{on} = \frac{1}{\tau_U [L]}, \quad (17)$$

where $[L]$ is the ligand concentration in the simulation system.

According to Kramers' rate theory, the rate of a chemical reaction in the large viscosity limit is calculated as³⁵:

$$k_R \cong \frac{w_m w_b}{2\pi\xi} e^{-\Delta F/k_B T}, \quad (18)$$

where w_m and w_b are frequencies of the approximated harmonic oscillators (also referred to as curvatures of free energy surface³⁸) near the energy minimum and barrier, respectively, ξ is the frictional rate constant and ΔF is the free energy barrier of transition. The friction constant ξ is related to the diffusion coefficient D with $\xi = k_B T/D$. The apparent diffusion coefficient D can be obtained by dividing the kinetic rate calculated directly using the transition time series collected directly from simulations by that using the probability density solution of the Smoluchowski equation³⁹. In order to reweight ligand kinetics from the LiGaMD2 simulations using the Kramers' rate theory, the free energy barriers of ligand binding and dissociation are calculated from the original (reweighted, ΔF) and modified (no reweighting, ΔF^*) PMF profiles, similarly for curvatures of the reweighed (w) and modified (w^* , no reweighting) PMF profiles near the ligand bound ("B") and unbound ("U") low-energy wells and the energy barrier ("Br"), and the ratio of apparent diffusion coefficients from simulations without reweighting (modified, D^*) and with

reweighting (D). The resulting numbers are then plugged into Eq. (17) to estimate accelerations of the ligand binding and dissociation rates during LiGaMD2 simulations³⁵, which allows us to recover the original kinetic rate constants.

System Setup

The complex structures of benzene (BEN) bound to the L99A T4L, M102A T4L and F104A T4L were obtained from the 3HH4⁴⁰, 220L⁴¹ and 227L⁴¹ PDB files, respectively. The crystal structure of indole (IND) bound to the L99A T4L was obtained from the 185L⁴² PDB file. The AMBER ff14SB force field⁴³ was used for the protein and the GAFF force field⁴⁴ for the ligand with AM1-BCC charges. Each system was solvated in a periodic box of TIP3P water molecules with a distance of 18 Å from the solute to the box edge using *tleap*. Therefore, the ligand concentration was 0.0034 M in the simulation system. Appropriate number of Cl⁻ ions were added to achieve system neutrality.

Simulation Protocol

Each system was energy minimized and gradually heated to 300 K in 1 ns with the Langevin thermostat and harmonic restraints of 50 kcal/mol/Å² on all non-hydrogen atoms of the protein and the ligand using the AMBER22 software⁴⁵. The simulation system was firstly energy minimized with 1.0 kcal/mol/Å² constraints on the heavy atoms of the proteins, including the steepest descent minimization for 50,000 steps and conjugate gradient minimization for 50,000 steps. The system was then heated from 0 K to 300 K for 200 ps. It was further equilibrated using the NVT ensemble at 300 for 200 ps and the NPT ensemble at 300 K and 1 bar for 1 ns with 1 kcal/mol/Å² constraints on the heavy atoms of the protein, followed by 2 ns short cMD without any constraint. The LiGaMD2 simulations proceeded with 14 ns short cMD to collect the potential

statistics, 49.2 ns LiGaMD2 equilibration after adding the boost potential and then three independent 1,000 ns production runs. It provided more powerful sampling to set the threshold energy for applying the boost potential to the upper bound (i.e., $E = V_{\min} + 1/k$) in our previous study ligand dissociation and binding using LiGaMD²⁹. Therefore, the threshold energy for applying the ligand essential potential was also set to the upper bound in the LiGaMD2 simulations. The selective boost potential was applied to both the ligand and protein pocket residues within 5 Å of ligand in the LiGaMD2 simulations. For the second boost potential applied on the system total potential energy other than the ligand essential potential energy, sufficient acceleration was obtained by setting the threshold energy to the lower bound. In order to observe ligand dissociation during LiGaMD2 equilibration while keeping the boost potential as low as possible for accurate energetic reweighting, the (σ_{OP} , σ_{OD}) parameters were finally set to (9.0 kcal/mol, 6.0 kcal/mol), (7.0 kcal/mol, 6.0 kcal/mol), (9.0 kcal/mol, 6.0 kcal/mol) and (9.0 kcal/mol, 6.0 kcal/mol) for the LiGaMD2 simulations of the BEN bound to the L99A T4L (T4L:L99A-BEN), F104A T4L (T4L:F104A-BEN), M102A T4L (T4L:M102A-BEN) and IND bound to the L99A T4L (T4L:L99A-IND). LiGaMD2 production simulation frames were saved every 0.4 ps for analysis.

Simulation Analysis

The VMD⁴⁶ and CPPTRAJ⁴⁷ tools were used for simulation analysis. The number of ligand dissociation and binding events (N_D and N_B) and the ligand binding and unbinding time periods (τ_B and τ_U) were recorded from individual simulations (**Tables 1 & S1**). With high fluctuations, τ_B and τ_U were recorded for only the time periods longer than 1 ns. The 1D, 2D and 3D PMF profiles, as well as the ligand binding free energy, were calculated through energetic reweighting of the LiGaMD2 simulations. The center-of-mass distance between the ligand and the protein pocket (defined by protein residues within 5 Å of ligand) and ligand heavy atom RMSDs relative to X-

ray structures with the protein aligned were chosen as the reaction coordinate for calculating the 1D PMF profiles. The bin size was set to 1.0 Å. The software *trj_cavity*⁴⁸ implemented in GROMACS was used to calculate the pocket volume. 2D PMF profiles of the ligand RMSD and pocket volume were calculated to analyze conformational changes of the protein upon ligand binding. The bin size was set to 50 Å³ for pocket volume. The cutoff for the number of simulation frames in one bin was set to 500 for reweighting 1D and 2D PMF profiles. The 3D PMF profiles of ligand displacements from the T4L mutants in the X, Y and Z directions were further calculated from the LiGaMD2 simulations. The bin sizes were set to 1 Å in the X, Y and Z directions. The cutoff of simulation frames in one bin for 3D PMF reweighting (ranging from 100-400 for three individual LiGaMD2 simulations) was set to the minimum number below which the calculated 3D PMF minimum will be shifted. The ligand binding free energies (ΔG) were calculated using the reweighted 3D PMF profiles and binding kinetic rates by $\Delta G = -RT\ln(k_{off}/k_{on})$, respectively. The ligand dissociation and binding rate constants (k_{on} and k_{off}) were calculated from the LiGaMD2 simulations with their accelerations analyzed using the Kramers' rate theory (**Table S2**).

Results

Flexibility of the protein pocket plays an important role in dissociation of buried ligands

The ligand binding pockets are different in the L99A and F104A mutants of T4 lysozyme (T4L). The binding pocket in L99A T4L is deeply buried (**Fig. 1A**), while the pocket is exposed to the solvent in the F104A mutant (**Fig. 1D**). LiGaMD and LiGaMD2 testing simulations with σ_{OP} increasing from 6.0 to 10.0 kcal/mol and σ_{OD} at 6.0 kcal/mol were performed on these two systems. Ligand dissociation was captured in both LiGaMD and LiGaMD2 simulations of benzene (BEN)

binding in the F104A mutant with an exposed binding pocket (**Fig. 1E& 1F**). With LiGaMD, the ligand dissociated from the F104A T4L at 5.45 ns, 7.20 ns, 18.22 ns and 7.80 ns with the σ_{0P} values of 6.0, 7.0, 8.0 and 9.0 kcal/mol, respectively. Interestingly, LiGaMD with the σ_{0P} value at 9.0 could capture both the ligand dissociation and rebinding within the 49.2 ns equilibration simulation. LiGaMD2 captured the ligand dissociation at 3.60 ns, 11.20 ns, 4.50 ns and 29.60 ns with the σ_{0P} value of 6.0, 7.0, 8.0 and 9.0 kcal/mol, respectively. The LiGaMD2 with the σ_{0P} values of 7.0 and 8.0 kcal/mol could capture the ligand dissociation and rebinding within the 49.2 ns equilibration simulation.

For the L99A T4L with a buried protein pocket, the LiGaMD could not capture ligand dissociation even with the value of σ_{0P} increased to 10.0 kcal/mol (**Fig. 1B**). In comparison, the LiGaMD2 simulations could capture the ligand dissociation and rebinding with the σ_{0P} values of 9.0 and 10.0 kcal/mol (**Fig. 1B&1C**). The ligand dissociated at 12.52 ns and rebound at ~29.50 ns in the LiGaMD2 simulation with the σ_{0P} value of 9.0 kcal/mol. In the LiGaMD2 simulation with the σ_{0P} value of 10.0 kcal/mol, the ligand dissociation and rebinding occurred at ~23.60 ns and ~26.80 ns, respectively.

Next, we identified correlation of the ligand dissociation and conformational changes of the protein pocket in the L99A T4L (**Fig. S1**). The binding pocket exhibited low RMSD of ~1-2 Å in the LiGaMD simulations with all the tested parameters as no selective boost potential was applied to the protein pocket (**Fig. S1A**). In comparison, RMSD of the binding pocket significantly increased in the LiGaMD2 simulations with the σ_{0P} values of 9.0 and 10.0 kcal/mol, respectively (**Fig. S1B**). As the σ_{0P} value of 9.0 kcal/mol was the lowest value that enabled the ligand dissociation and rebinding, we further focused on the binding pocket in this simulation. During ligand dissociation around 12 ns, RMSD of the binding pocket increased to ~1.7 Å. RMSD of the

binding pocket decreased to mostly <1.0 Å when the ligand completely dissociated to the solvent. The binding pocket RMSD increased to ~ 2 Å when the ligand rebound to the pocket at ~ 30 ns (**Fig S1C**). After the ligand bound completely to the protein pocket, RMSD of the binding pocket decreased to ~ 1 Å again. We further calculated volume of the protein pocket during the LiGaMD2 simulation (**Fig S1D**). The volume of the protein pocket was overall smaller in the ligand-free (*apo*) state than in the ligand-bound (*holo*) state (**Fig S1D**). Similar opening of the binding pocket was observed in previous aMD simulation of BEN dissociation from the L99A T4L⁴⁹. In summary, LiGaMD2 showed improved sampling of ligand binding to the buried protein pockets, where dynamics of the binding pocket played an important role in the ligand dissociation and rebinding. For the system with an open pocket, both LiGaMD and LiGaMD2 worked well.

Microsecond LiGaMD2 simulations captured repetitive ligand dissociation and rebinding to the T4L mutants

As the σ_{op} values of 9.0 and 7.0 kcal/mol were the lowest to capture the ligand dissociation and rebinding in LiGaMD2 equilibration simulations of the L99A and F104A T4L systems (**Figs. 1C & 1F**), they were used for further three independent 1,000 ns production simulations. Furthermore, two more systems with buried binding pockets were added to demonstrate the robustness of LiGaMD2, including the M102A T4L bound by BEN and the L99A T4L bound by a different ligand indole (IND). In summary, LiGaMD2 simulations were performed on three complexes of BEN bound to the L99A T4L (T4L:L99A-BEN), M102A T4L (T4L:M102A-BEN) and F104A T4L (T4L:F104A-BEN), and another complex of IND bound to the L99A T4L (T4L:L99A-IND) (**Figs. 2A-D**). Three independent 1,000 ns LiGaMD2 production simulations were performed on each of the four systems (**Table 1**). The LiGaMD2 simulations of the T4L:L99A-BEN system recorded an average boost potential of 107.77-109.84 kcal/mol with 9.58-9.67 kcal/mol standard

deviation. The LiGaMD2 simulations of the T4L:M102A-BEN system recorded an average boost potential of 108.04-109.21 kcal/mol with 9.49-10.04 kcal/mol standard deviation. In comparison, the average boost potential was 70.45-72.53 kcal/mol with 7.81-8.11 kcal/mol standard deviation in the three simulations of the T4L:F104A-BEN. The boost potential applied in simulations of the T4L:L99A-IND system was slightly larger than that of the T4L:L99A-BEN system, with average of 115.80-118.16 kcal/mol and 9.68-9.79 kcal/mol standard deviation (**Table 1**).

RMSDs of the ligand relative to the X-ray structures with the T4L aligned were calculated (**Figs. 2A-2D**) to record the number of ligand dissociation and binding events (N_D and N_B) in each of the 1,000 ns LiGaMD2 simulations. With close examination of the ligand binding trajectories, RMSD cutoffs of the ligand unbound and bound states were set to $>15 \text{ \AA}$ and $<5.0 \text{ \AA}$, respectively. Because of ligand fluctuations, we recorded only the corresponding binding and dissociation events that lasted for more than 1.0 ns. In each simulation of the T4L:L99A-BEN system, about 5-6 binding and 5-6 dissociation events were observed (**Fig. 2A & Table 1**). In each simulation of the T4L:F104A-BEN system, about 3-4 dissociation and 3-4 binding events were observed (**Fig. 2B & Table 1**). In each simulation of the T4L:M102A-BEN system, about 3-7 dissociation and 3-7 binding events were observed (**Fig. 2C & Table 1**). A similar number of ligand dissociation (3-6) and binding (3-7) events were observed in simulations of the T4L:L99A-IND system (**Fig. 2D & Table 1**). In summary, repetitive ligand dissociation and rebinding were successfully captured in each of the 1,000 ns LiGaMD2 simulations of various T4L mutants bound by different ligands (**Figs. 2**).

Next, we explored the correlation between conformational changes of the binding pocket and ligand binding in the LiGaMD2 production simulations. The ligand RMSD and volume of the binding pocket were used as reaction coordinates to calculate 2D PMF (**Fig. 3**). Five low-energy

states were identified in the 2D PMF profile of the T4L:L99A-BEN system including the Bound (“B”), Intermediate (“I”), Unbound U1 (“U1”), Unbound U2 (“U2”) and Unbound U3 (“U3”) states (**Fig. 3A**). The ligand RMSD and pocket volume in the B, I, U1, U2 and U3 states centered around (2.5 Å, 300 Å³), (9.3 Å, 425.7 Å³), (24.5 Å, 259.9 Å³), (34.8 Å, 360.3 Å³), and (41.6 Å, 251.9 Å³), respectively (**Fig 3A**). In the T4L:F104A-BEN system, three low-energy states were identified, including the Bound (“B”), Intermediate (“I”) and Unbound (“U”) states (**Fig. 3B**). The ligand RMSD and the pocket volume in the B, I and U centered around (3.8 Å, 176.80 Å³), (11.6 Å, 59.6 Å³) and (35.0 Å, 54.7 Å³), respectively (**Fig. 3B**). In the T4L:M102A-BEN system, five low-energy states were identified including the Bound (“B”), Intermediate (“I”), Intermediate (“I2”), Unbound (“U1”) and Unbound (“U2”) states (**Fig. 3C**). The ligand RMSD and pocket volume in the B, I, I2, U1 and U2 states centered around (2.7 Å, 219.7 Å³), (10.7 Å, 356.5 Å³), (11.0 Å, 113.4 Å³), (32.9 Å, 258.8 Å³), and (37.8 Å, 132.9 Å³), respectively (**Fig 3C**). In the T4L:L99A-IND system, four low-energy states were identified including the Bound (“B”), Intermediate (“I”), Intermediate (“I2”) and Unbound (“U”) states (**Fig. 3D**). The ligand RMSD and pocket volume in the B, I, I2, and U states centered around (1.9 Å, 363.2 Å³), (5.0 Å, 468.8 Å³), (11.5 Å, 357.4 Å³), and (28.8 Å, 155.4 Å³), respectively (**Fig 3D**). In compared with the bound state, a larger binding pocket volume was identified in the intermediate I state in the systems with an buried protein pocket including the T4L:L99A-BEN, T4L:M102A-BEN and T4L:L99A-IND (**Figs. 3A, 3B & 3D**). In contrast, the binding pocket volume was significantly smaller in the intermediate I state in the T4L:F104A-BEN with an exposed binding pocket (**Fig. 3B**). The representative intermediate conformational states from the four systems were shown in **Fig. 3E-3H**. Compared to the bound X-ray structures, helices C and G in the intermediate “I” state moved outward in the T4L:L99A-BEN system, leading to opening of the binding pocket (**Fig. 3E**). In the

T4L:F104A-BEN system, major conformational changes upon ligand binding involved helices C and B, which moved inwards and reduced volume of the binding pocket (**Fig. 3F**). In the T4L:M102A-BEN system, helices D, F and G moved outwards in the intermediate “I” state with opening of the ligand binding pocket, accompanied by inward movement of helix C (**Fig. 3G**). In the T4L:L99A-IND system, outward movements were observed in helices C, D and G in the intermediate I state (**Fig. 3H**). Therefore, conformational changes of the protein pocket greatly facilitated the ligand dissociation and binding in the L99A and M104A systems with a buried binding pocket.

Ligand binding kinetic rates and free energies calculated from LiGaMD2 agreed well with experimental data

LiGaMD2 simulations that successfully captured repetitive ligand binding and dissociation allowed us to calculate the ligand binding kinetic rate constants. We recorded the time periods for the ligand found in the bound (τ_B) and unbound (τ_U) states throughout the LiGaMD2 simulations. Without reweighting, the ligand binding rate constants (k_{on}^*) were calculated directly from the LiGaMD2 trajectories as $8.22 \pm 5.48 \times 10^7 \text{ M}^{-1} \cdot \text{s}^{-1}$, $7.79 \pm 1.36 \times 10^6 \text{ M}^{-1} \cdot \text{s}^{-1}$, $6.81 \pm 1.27 \times 10^7 \text{ M}^{-1} \cdot \text{s}^{-1}$ and $1.67 \pm 0.67 \times 10^9 \text{ M}^{-1} \cdot \text{s}^{-1}$ in the system of T4L:L99A-BEN, T4L:M102A-BEN, T4L:F104A-BEN and T4L:L99A-IND systems, respectively (**Table 2**). The accelerated dissociation rate constants (k_{off}^*) were calculated as $3.47 \pm 2.31 \times 10^5 \text{ s}^{-1}$, $3.46 \pm 1.81 \times 10^9 \text{ s}^{-1}$, $1.72 \pm 1.44 \times 10^9 \text{ s}^{-1}$ and $2.49 \pm 1.25 \times 10^7 \text{ s}^{-1}$ in the T4L:L99A-BEN, T4L:M102A-BEN, T4L:F104A-BEN and T4L:L99A-IND systems, respectively (**Table 2**).

Next, we reweighted the LiGaMD2 simulations of ligand-T4L mutants to calculate acceleration factors of ligand binding and dissociation processes (**Table S1**) and recovered the original kinetic

rate constants using the Kramers' rate theory (**Table 2**). The dissociation free energy barrier (ΔF_{off}) significantly decreased from 10.5 ± 0.86 , 9.00 ± 1.07 , 8.55 ± 0.54 , and 7.11 ± 0.34 kcal/mol in the reweighted PMF profiles to 2.55 ± 0.43 , 2.35 ± 0.34 , 2.36 ± 0.35 and 2.23 ± 0.26 kcal/mol in the modified PMF profiles for the system of T4L:L99A-BEN, T4L:M102A-BEN, T4L:F104A-BEN and T4L:L99A-IND, respectively, respectively (**Table S1 and Fig. S2**). The free energy barrier for ligand binding (ΔF_{on}) slightly decreased from 4.93 ± 1.38 , 6.03 ± 0.11 , 5.40 ± 1.07 , 4.37 ± 0.33 kcal/mol in the reweighted profiles to 0.57 ± 0.14 , 1.15 ± 0.50 , 0.54 ± 0.053 , 0.53 ± 0.10 kcal/mol in the modified PMF profiles for the system of T4L:L99A-BEN, T4L:M102A-BEN, T4L:F104A-BEN and T4L:L99A-IND, respectively (**Table S1 and Fig. S2**). Curvatures of the reweighed (w) and modified (w^* , no reweighting) free energy profiles were calculated near the ligand Bound ("B") and Unbound ("U") low-energy wells and the energy barrier ("Br"), as well as the ratio of apparent diffusion coefficients calculated from LiGaMD2 simulations with reweighting (D) and without reweighting (modified, D^*) (**Table S1**). According to the Kramers' rate theory, the ligand association was accelerated by 11.07, 0.81, 0.02 and 558.5 times for the systems of T4L:L99A-BEN, T4L:M102A-BEN, T4L:F104A-BEN and T4L:L99A-IND, respectively. The ligand dissociation was significantly accelerated by 2.40×10^2 , 1.72×10^6 , 1.24×10^3 and 7.13×10^3 times in the T4L:L99A-BEN, T4L:M102A-BEN, T4L:F104A-BEN and T4L:L99A-IND systems, respectively. Therefore, the reweighted k_{on} in the T4L:L99A-BEN, T4L:M102A-BEN, T4L:F104A-BEN and T4L:L99A-IND systems were calculated as $7.42 \pm 4.81 \times 10^6$, $9.57 \pm 6.29 \times 10^6$, $3.16 \pm 2.29 \times 10^9$ and $2.99 \pm 2.87 \times 10^6$ M⁻¹·s⁻¹, being highly consistent with the corresponding experimental values⁵⁰ of $0.8\text{-}1.0 \times 10^6$, $3.0\text{-}5.0 \times 10^6$, $>10 \times 10^6$ and $0.7\text{-}1.0 \times 10^6$ M⁻¹·s⁻¹, respectively (**Table 2**). The reweighted k_{off} in the T4L:L99A-BEN, T4L:M102A-BEN, T4L:F104A-BEN and T4L:L99A-IND systems were calculated as 1441 ± 883 , 2011 ± 1606 , $1.38 \pm 0.67 \times 10^6$ and 3494 ± 559

s⁻¹, in good agreement with the corresponding experimental values⁵⁰ of 950, 3000, >10000 and 325 s⁻¹, respectively.

Based on the ligand binding kinetic rates (k_{on} and k_{off}), we calculated the ligand binding free energies as $\Delta G = -RT\ln(k_{off}/k_{on})$. The resulting binding free energies in the T4L:L99A-BEN, T4L:M102A-BEN, T4L:F104A-BEN and T4L:L99A-IND systems (**Table 1**) were -5.17 ± 0.72 kcal/mol, -5.01 ± 0.73 kcal/mol, -3.42 ± 0.72 kcal/mol and -4.87 ± 1.06 kcal/mol, being highly consistent with the corresponding experimental values of -4.12 kcal/mol, -4.18 kcal/mol, -4.02 kcal/mol and -4.82 kcal/mol, respectively. The root-mean square error (RMSE) of binding free energy was only 0.73 kcal/mol. Alternatively, we could calculate the ligand binding free energy using the 3D reweighted PMF profiles of the ligand displacement from the T4L binding pocket in the X, Y and Z directions (**Table 1**). The ligand binding free energies in the T4L:L99A-BEN, T4L:M102A-BEN, T4L:F104A-BEN and T4L:L99A-IND systems (**Table 1**) were estimated as -5.88 ± 0.61 kcal/mol, -6.43 ± 0.70 kcal/mol, -4.57 ± 0.12 kcal/mol and -7.40 ± 0.39 kcal/mol, respectively. The RMSE of binding free energies predicted from the 3D PMF was greater as 1.94 kcal/mol, but still within the acceptable range of binding free energy predictions (2.0 kcal/mol)⁵¹. Therefore, both efficient conformational sampling and accurate ligand binding thermodynamic and kinetic calculations were achieved through the LiGaMD2 simulations.

Multiple ligand binding and dissociation pathways were identified from LiGaMD2 simulations

We closely examined the LiGaMD2 trajectories to explore ligand binding and dissociation pathways. For dissociation of BEN from the T4L:L99A, four pathways between the CD, CF, DG and FGH helices were identified (**Fig. 4**). All these pathways were involved in the BEN rebinding to the T4L:L99A. One extra binding pathway of HG was found in the BEN binding to the

T4L:L99A. LiGaMD2 captured 5, 3, 6 and 3 times of BEN dissociation events through pathways of CD, CF, DG and FGH, respectively (**Fig 4**). The BEN rebinding events through pathways of CD, CF, DG, FGH and HG were 3, 2, 3, 6 and 2, respectively. The same ligand dissociation pathways of CD, CF, DG and FGH were identified in the LiGaMD2 simulations of T4L:M102A-BEN and T4L:L99A-IND systems (**Fig. 4**). Two pathways of DG and FGH were observed in the simulations of BEN binding to the T4L:M102A. The ligand dissociation events in T4L:M102A system along pathways of CD, CF, DG and FGH were 1, 2, 1 and 12, respectively (**Fig 4**). The ligand rebinding events in the T4L:M102A-BEN through pathways of DG and FGH were 2 and 11, respectively. For dissociation of BEN from the T4L:F104A, two pathways near the A and C helices were identified. The dissociation events in the T4L:F104A-BEN were 8 and 1 via pathways A and C, respectively (**Fig. 4**). Pathway C was observed in BEN binding to the F104A T4L mutant (**Fig. 4**). For the dissociation of IND from the T4L:L99A, four pathways between the CD, CF, DG and FGH helices were identified. While only pathways CF and FGH were observed in the rebinding of IND to the L99A T4L mutant. The dissociating event via pathways CD, CF, DG and FGH in the T4L:L99A-IND system were 3, 3, 5, and 3, respectively. The IND binding events via pathways CF and FGH were 3 and 9, respectively (**Fig. 4**). The binding and dissociating pathways were consistent with earlier simulation findings using Metadynamics^{28, 52}, Weighted Ensemble^{20b}, Machine Learning⁵³, tRAMD^{15a} and aMD⁴⁹ simulations. In summary, multiple ligand binding and dissociation pathways were observed in the LiGaMD2 simulations of the T4L mutants. The ligand binding and dissociation mostly followed the same pathways.

Discussions

We have presented LiGaMD2 that improved enhanced sampling and accurate prediction of protein-ligand binding thermodynamics and kinetics for especially proteins with closed binding pockets. LiGaMD2 works by selectively boosting the potential of both ligand and protein residues in the binding pocket. LiGaMD2 shows significantly improved sampling of systems with buried binding pockets, where flexibility of the binding pocket plays a critical role in ligand binding. Microsecond LiGaMD2 simulations have allowed us to capture repetitive ligand dissociation and rebinding processes as demonstrated on four T4L mutant model systems. These simulations then enabled accurate predictions of ligand binding free energies and kinetic rate constants.

LiGaMD2 simulations revealed the critical role of protein flexibility for ligand binding, especially in the case of solvent-inaccessible buried pockets, in good agreement with previous experimental^{52a, 54} and computational studies^{14d, 20b, 49}. Protein flexibility has been recognized as one of the main factors that regulates protein-ligand binding kinetics^{25c-e}. The influence of protein flexibility on ligand binding site can vary from small changes like opening or closing of an existing pocket to the formation of a new pocket^{25e}. For example, the MSM⁸ built with 60 μ s cMD simulations revealed that the movement of helix D/G/H/J could transiently open a channel for ligand binding to the target site of the L99A T4L. Such movement of helix D/G/H were also observed in the intermediate states in the LiGaMD2 simulations (**Fig. 3**). Additionally, multiple ligand binding and dissociation pathways were identified from LiGaMD2 simulations (**Fig. 4**), being highly consistent with previous enhanced sampling simulations, including the RAMD^{15a}, aMD⁴⁹, Metadynamics²⁸, MSM⁸ and Weighted Ensemble^{20b}. For example, the dissociation pathway FGH with highest probability observed in LiGaMD2 was also captured in the simulations of RAMD^{15a}, Metadynamics²⁸, aMD⁴⁹, MSM⁸ and Weighted Ensemble^{20b}.

Compared with the cMD⁵⁵, Metadynamics⁵⁶, Weighted Ensemble,⁵⁷ MSM⁸ and Replica Exchange MD simulations^{25a}, LiGaMD2 provides an efficient and/or easier-to-use approach to simulation of ligand binding and dissociation and calculations of ligand binding thermodynamics and kinetics. It is advantageous over previous LiGaMD for proteins with buried binding pockets. Microsecond cMD simulations were able to capture benzene binding to the L99A T4L⁸. However, slower ligand dissociation was still beyond the accessibility of cMD. Weighted Ensemble^{20b} and MSM were able to accurately predict ligand binding kinetics⁸. However, tens of microsecond cMD simulations were needed for the Weighted Ensemble^{20b} and MSM⁸. For the Replica Exchange method⁵⁸, a large number of replica simulations were often needed to model protein-ligand binding. In the case of gREST simulations^{14d}, eight replicas were needed to capture ligand binding to the L99A T4L. With carefully designed CV, Metadynamics could capture both ligand binding and unbinding with high efficiency. However, the predefined CVs could potentially lead to certain constraints on the ligand binding pathways and conformational space. Such simulations could also suffer from the “hidden energy barrier” problem and slow convergence if important CVs were missing.⁵⁹ Overall, the previous methods appeared computationally expensive, requiring mostly tens-of-microsecond simulations to characterize ligand binding thermodynamics and kinetics. In this context, LiGaMD2 that has allowed us to capture repetitive ligand binding and unbinding within microsecond simulations. It provides an improved approach to characterization of ligand binding thermodynamics and kinetics, especially for proteins with buried binding pockets.

Acknowledgements

This work used supercomputing resources with allocation award TG-MCB180049 and BIO210039 through the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National

Science Foundation grant number ACI-1548562, and project M2874 through the National Energy Research Scientific Computing Center (NERSC), which is a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231, and the Research Computing Cluster and BigJay Cluster resource funded through NSF Grant MRI-2117449 at the University of Kansas. This work was supported in part by the National Institutes of Health (R01GM132572) and National Science Foundation (2121063).

Supporting Information

Tables S1, S2 and **Figures S1-S2** are provided in the Supporting Information. This information is available free of charge via the Internet at <http://pubs.acs.org>.

Reference

1. Nooren, I. M.; Thornton, J. M., Diversity of protein-protein interactions. *EMBO J* **2003**, *22* (14), 3486-92.
2. (a) Hajduk, P. J.; Greer, J., A decade of fragment-based drug design: strategic advances and lessons learned. *Nature Reviews Drug Discovery* **2007**, *6* (3), 211-219; (b) Klebe, G., Applying thermodynamic profiling in lead finding and optimization. *Nature Reviews Drug Discovery* **2015**, *14* (2), 95-110.
3. (a) Hauser, A. S.; Attwood, M. M.; Rask-Andersen, M.; Schiöth, H. B.; Gloriam, D. E., Trends in GPCR drug discovery: new agents, targets and indications. *Nature Reviews Drug Discovery* **2017**, *16* (12), 829-842; (b) Miura, K., An overview of current methods to confirm protein-protein interactions. *Protein and peptide letters* **2018**, *25* (8), 728-733; (c) Renaud, J.-P.; Chari, A.; Ciferri, C.; Liu, W.-t.; Rémigy, H.-W.; Stark, H.; Wiesmann, C., Cryo-EM in drug discovery: achievements, limitations and prospects. *Nature Reviews Drug Discovery* **2018**, *17* (7), 471-492.
4. (a) 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.; Muller-Fahrnow, A.; Ecker, G. F., Kinetics for Drug Discovery: an industry-driven effort to target drug residence time. *Drug Discov Today* **2017**, *22* (6), 896-911; (b) Tonge, P. J., Drug-Target Kinetics in Drug Discovery. *ACS chemical neuroscience* **2018**, *9* (1), 29-39; (c) Ahmad, K.; Rizzi, A.; Capelli, R.; Mandelli, D.; Lyu, W.; Carloni, P., Enhanced-Sampling Simulations for the Estimation of Ligand Binding Kinetics: Current Status and Perspective. *Frontiers in Molecular Biosciences* **2022**, *9*; (d) IJzerman, A. P.; Guo, D., Drug - Target Association Kinetics in Drug Discovery. *Trends in Biochemical Sciences* **2019**, *44* (10), 861-871; (e) Tonge, P. J., Drug-Target Kinetics in Drug Discovery. *ACS Chemical Neuroscience* **2018**, *9* (1), 29-39; (f) Ferruz, N.; De Fabritiis, G., Binding Kinetics in Drug Discovery. *Molecular Informatics* **2016**, *35* (6-7), 216-226; (g) Zhou, Y.; Fu, Y.; Yin, W.; Li, J.; Wang, W.; Bai, F.; Xu, S.; Gong, Q.; Peng, T.; Hong, Y.; Zhang, D.; Zhang, D.; Liu, Q.; Xu, Y.; Xu, H. E.; Zhang, H.; Jiang, H.; Liu, H., Kinetics-Driven Drug Design Strategy for Next-Generation Acetylcholinesterase Inhibitors to Clinical Candidate. *Journal of Medicinal Chemistry* **2021**, *64* (4), 1844-1855.
5. Pawnikar S, B. A., Wang J, Miao Y. , Binding Analysis Using Accelerated Molecular Dynamics Simulations and Future Perspectives. *Adv Appl Bioinform Chem* **2022**, *15*, 1-19.
6. (a) Shan, Y.; 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* (24), 9181-3; (b) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E., Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc Natl Acad Sci* **2011**, *108* (32), 13118-23; (c) Shaw, D. E.; Adams, P. J.; Azaria, A.; Bank, J. A.; Batson, B.; Bell, A.; Bergdorf, M.; Bhatt, J.; Butts, J. A.; Correia, T. In *Anton 3: twenty microseconds of molecular dynamics simulation before lunch*, Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, 2021; pp 1-11; (d) Pan, A. C.; Xu, H.; Palpant, T.; Shaw, D. E., Quantitative Characterization of the Binding and Unbinding of Millimolar Drug Fragments with Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation* **2017**, *13* (7), 3372-3377.
7. Shan, Y.; 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* (24), 9181-9183.
8. 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 Computational Biology* **2018**, *14* (5), e1006180.
9. Tang, Z.; Chang, C.-e. A., Binding Thermodynamics and Kinetics Calculations Using Chemical Host and Guest: A Comprehensive Picture of Molecular Recognition. *Journal of Chemical Theory and Computation* **2018**, *14* (1), 303-318.

10. (a) Kappel, K.; Miao, Y.; McCammon, J. A., Accelerated Molecular Dynamics Simulations of Ligand Binding to a Muscarinic G-protein Coupled Receptor. *Quarterly Reviews of Biophysics* **2015**, *48* (04), 479-487; (b) Saglam, A. S.; Chong, L. T., Protein-protein binding pathways and calculations of rate constants using fully-continuous, explicit-solvent simulations. *Chemical Science* **2019**, *10* (8), 2360-2372.
11. (a) Wang, Y.; Valsson, O.; Tiwary, P.; Parrinello, M.; Lindorff-Larsen, K., Frequency adaptive metadynamics for the calculation of rare-event kinetics. *The Journal of Chemical Physics* **2018**, *149* (7), 072309; (b) Wang, Y.; Martins, J. M.; Lindorff-Larsen, K., Biomolecular conformational changes and ligand binding: from kinetics to thermodynamics. *Chemical Science* **2017**, *8* (9), 6466-6473; (c) Banerjee, P.; Bagchi, B., Dynamical control by water at a molecular level in protein dimer association and dissociation. *Proceedings of the National Academy of Sciences* **2020**, *117* (5), 2302-2308; (d) Antoszewski, A.; Feng, C.-J.; Vani, B. P.; Thiede, E. H.; Hong, L.; Weare, J.; Tokmakoff, A.; Dinner, A. R., Insulin dissociates by diverse mechanisms of coupled unfolding and unbinding. *The Journal of Physical Chemistry B* **2020**, *124* (27), 5571-5587; (e) Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M., Kinetics of protein-ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proc Natl Acad Sci* **2015**, *112* (5), E386-E391; (f) Casanovas, 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* (13), 4780-4788.
12. (a) Kingsley, L. J.; Esquivel-Rodríguez, J.; Yang, Y.; Kihara, D.; Lill, M. A., Ranking protein-protein docking results using steered molecular dynamics and potential of mean force calculations. *Journal of computational chemistry* **2016**, *37* (20), 1861-1865; (b) Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K., Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J Chem Phys* **2003**, *119* (6), 3559-3566; (c) Gonzalez, A.; Perez-Acle, T.; Pardo, L.; Deupi, X., Molecular basis of ligand dissociation in beta-adrenergic receptors. *PLoS One* **2011**, *6* (9), e23815.
13. (a) Gumbart, J. C.; Roux, B.; Chipot, C., Efficient determination of protein-protein standard binding free energies from first principles. *Journal of chemical theory and computation* **2013**, *9* (8), 3789-3798; (b) Joshi, D. C.; Lin, J. H., Delineating Protein-Protein Curvilinear Dissociation Pathways and Energetics with Naïve Multiple-Walker Umbrella Sampling Simulations. *Journal of Computational Chemistry* **2019**, *40* (17), 1652-1663; (c) Reif, M. M.; Zacharias, M., Improving the Potential of Mean Force and Nonequilibrium Pulling Simulations by Simultaneous Alchemical Modifications. *Journal of Chemical Theory and Computation* **2022**, *18* (6), 3873-3893.
14. (a) Siebenmorgen, T.; Zacharias, M., Efficient Refinement and Free Energy Scoring of Predicted Protein-Protein Complexes Using Replica Exchange with Repulsive Scaling. *Journal of Chemical Information and Modeling* **2020**, *60* (11), 5552-5562; (b) Sugita, Y.; Okamoto, Y., Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters* **1999**, *314* (1-2), 141-151; (c) Sugita, Y.; Kamiya, M.; Oshima, H.; Re, S., Replica-exchange methods for biomolecular simulations. In *Biomolecular Simulations*, Springer: 2019; pp 155-177; (d) Niitsu, A.; Re, S.; Oshima, H.; Kamiya, M.; Sugita, Y., De novo prediction of binders and non-binders for T4 Lysozyme by gREST simulations. *Journal of Chemical Information and Modeling* **2019**, *59* (9), 3879-3888.
15. (a) Nunes-Alves, A.; Kokh, D. B.; Wade, R. C., Ligand unbinding mechanisms and kinetics for T4 lysozyme mutants from τ RAMD simulations. *Current Research in Structural Biology* **2021**, *3*, 106-111; (b) Kokh, D. B.; Amaral, M.; Bomke, J.; Gradler, 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 tau-Random Acceleration Molecular Dynamics Simulations. *J Chem Theory Comput* **2018**, *14* (7), 3859-3869; (c) Wang, T.; Duan, Y., Chromophore channeling in the G-protein coupled receptor rhodopsin. *J. Am. Chem. Soc.* **2007**, *129* (22), 6970-6971.
16. (a) Deb, I.; Frank, A. T., Accelerating Rare Dissociative Processes in Biomolecules Using Selectively Scaled MD Simulations. *Journal of Chemical Theory and Computation* **2019**, *15* (11), 5817-5828; (b) Schuetz, D. A.; Bernetti, M.; Bertazzo, M.; Musil, D.; Eggenweiler, H. M.; Recanatini, M.; Masetti, M.; Ecker, G. F.; Cavalli, A., Predicting Residence Time And Drug Unbinding Pathway Through Scaled Molecular Dynamics. *J Chem Inf Model* **2018**.

17. Hamelberg, D.; Mongan, J.; McCammon, J. A., Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J Chem Phys* **2004**, *120* (24), 11919-11929.
18. (a) Wang, J.; Arantes, P. R.; Bhattacharai, A.; Hsu, R. V.; Pawnikar, S.; Huang, Y. M.; Palermo, G.; Miao, Y., Gaussian accelerated molecular dynamics (GaMD): principles and applications. *Wiley Interdiscip Rev Comput Mol Sci* **2021**, *11* (5), e1521; (b) Miao, Y.; Feher, V. A.; McCammon, J. A., Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *Journal of Chemical Theory and Computation* **2015**, *11* (8), 3584-3595.
19. (a) Buch, I.; Giorgino, T.; De Fabritiis, G., Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (25), 10184-10189; (b) Plattner, N.; Noe, F., Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat Commun* **2015**, *6*, 7653.
20. (a) Lotz, S. D.; Dickson, A., Wepy: A Flexible Software Framework for Simulating Rare Events with Weighted Ensemble Resampling. *ACS Omega* **2020**, *5* (49), 31608-31623; (b) Nunes-Alves, A.; Zuckerman, D. M.; Arantes, G. M., Escape of a Small Molecule from Inside T4 Lysozyme by Multiple Pathways. *Biophysical Journal* **2018**, *114* (5), 1058-1066.
21. Donyapour, N.; Roussey, N. M.; Dickson, A., REVO: Resampling of ensembles by variation optimization. *The Journal of Chemical Physics* **2019**, *150* (24), 244112.
22. Buch, I.; Giorgino, T.; De Fabritiis, G., Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proceedings of the National Academy of Sciences* **2011**, *108* (25), 10184-10189.
23. Miao, Y.; Bhattacharai, A.; Wang, J., Ligand Gaussian Accelerated Molecular Dynamics (LiGaMD): Characterization of Ligand Binding Thermodynamics and Kinetics. *J Chem Theory Comput* **2020**, *16* (9), 5526-5547.
24. (a) Baron, R.; McCammon, J. A., Dynamics, Hydration, and Motional Averaging of a Loop-Gated Artificial Protein Cavity: The W191G Mutant of Cytochrome c Peroxidase in Water as Revealed by Molecular Dynamics Simulations. *Biochemistry* **2007**, *46* (37), 10629-10642; (b) Karplus, M.; Gelin, B. R.; McCammon, J. A., Internal dynamics of proteins. Short time and long time motions of aromatic sidechains in PTI. *Biophysical Journal* **1980**, *32* (1), 603-618; (c) Amaral, M.; Kokh, D. B.; Bomke, J.; Wegener, A.; Buchstaller, H. P.; Eggenweiler, H. M.; Matias, P.; Sirrenberg, C.; Wade, R. C.; Frech, M., Protein conformational flexibility modulates kinetics and thermodynamics of drug binding. *Nature Communications* **2017**, *8* (1), 2276.
25. (a) Re, S.; Oshima, H.; Kasahara, K.; Kamiya, M.; Sugita, Y., Encounter complexes and hidden poses of kinase-inhibitor binding on the free-energy landscape. *Proceedings of the National Academy of Sciences* **2019**, *116* (37), 18404-18409; (b) Wilson, C.; Agafonov, R.; Hoemberger, M.; Kutter, S.; Zorba, A.; Halpin, J.; Buosi, V.; Otten, R.; Waterman, D.; Theobald, D., Using ancient protein kinases to unravel a modern cancer drug's mechanism. *Science* **2015**, *347* (6224), 882-886; (c) Pan, A. C.; Borhani, D. W.; Dror, R. O.; Shaw, D. E., Molecular determinants of drug-receptor binding kinetics. *Drug discovery today* **2013**, *18* (13-14), 667-673; (d) Klebe, G., The use of thermodynamic and kinetic data in drug discovery: decisive insight or increasing the puzzlement? *ChemMedChem* **2015**, *10* (2), 229-231; (e) Stank, A.; Kokh, D. B.; Fuller, J. C.; Wade, R. C., Protein Binding Pocket Dynamics. *Accounts of Chemical Research* **2016**, *49* (5), 809-815; (f) Teague, S. J., Implications of protein flexibility for drug discovery. *Nature Reviews Drug Discovery* **2003**, *2* (7), 527-541.
26. Wang, L.; Berne, B. J.; Friesner, R. A., On achieving high accuracy and reliability in the calculation of relative protein-ligand binding affinities. *Proceedings of the National Academy of Sciences* **2012**, *109* (6), 1937-1942.
27. (a) Nunes-Alves, A.; Kokh, D. B.; Wade, R. C., Recent progress in molecular simulation methods for drug binding kinetics. *Current Opinion in Structural Biology* **2020**, *64*, 126-133; (b) Eriksson, A. E.; Baase, W. A.; Wozniak, J. A.; Matthews, B. W., A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature* **1992**, *355* (6358), 371-373; (c) Eriksson, A. E.; Baase, W. A.; Zhang, X. J.; Heinz, D. W.; Blaber, M.; Baldwin, E. P.; Matthews, B. W., Response of a Protein Structure to Cavity-Creating Mutations and Its Relation to the Hydrophobic Effect. *Science* **1992**, *255* (5041), 178-183.

28. Capelli, R.; Carloni, P.; Parrinello, M., Exhaustive Search of Ligand Binding Pathways via Volume-Based Metadynamics. *The Journal of Physical Chemistry Letters* **2019**, *10* (12), 3495-3499.
29. Wang, J.; Miao, Y., Peptide Gaussian accelerated molecular dynamics (Pep-GaMD): Enhanced sampling and free energy and kinetics calculations of peptide binding. *J Chem Phys* **2020**, *153* (15), 154109.
30. (a) Vanommeslaeghe, K.; MacKerell, A. D., Jr., CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochim. Biophys. Acta* **2015**, *1850* (5), 861-871; (b) 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* (16), 1999-2012.
31. Roux, B., The Calculation of the Potential of Mean Force Using Computer-Simulations. *Comput Phys Commun* **1995**, *91* (1-3), 275-282.
32. Miao, Y.; Sinko, W.; Pierce, L.; Bucher, D.; McCammon, J. A., Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *J Chem Theory Comput* **2014**, *10* (7), 2677-2689.
33. (a) Doudou, S.; Burton, N. A.; Henchman, R. H., Standard Free Energy of Binding from a One-Dimensional Potential of Mean Force. *J. Chem. Theory Comput.* **2009**, *5* (4), 909-18; (b) Buch, I.; Giorgino, T.; De Fabritiis, G., Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc Natl Acad Sci U S A* **2011**, *108* (25), 10184-9.
34. (a) Miao, Y.; Sinko, W.; Pierce, L.; Bucher, D.; Walker, R. C.; McCammon, J. A., Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. *J. Chem. Theory Comput.* **2014**, *10* (7), 2677-2689; (b) Miao, Y.; Bhattarai, A.; Wang, J., Ligand Gaussian Accelerated Molecular Dynamics (LiGaMD): Characterization of Ligand Binding Thermodynamics and Kinetics. *J. Chem. Theory Comput.* **2020**.
35. Miao, Y., Acceleration of biomolecular kinetics in Gaussian accelerated molecular dynamics. *J. Chem. Phys.* **2018**, *149* (7), 072308.
36. Wang, J.; Miao, Y., Protein-Protein Interaction-Gaussian Accelerated Molecular Dynamics (PPI-GaMD): Characterization of Protein Binding Thermodynamics and Kinetics. *Journal of Chemical Theory and Computation* **2022**, *18* (3), 1275-1285.
37. Ferruz, N.; De Fabritiis, G., Binding kinetics in drug discovery. *Mol. Inform.* **2016**, *35* (6-7), 216-226.
38. (a) Doshi, U.; Hamelberg, D., Extracting Realistic Kinetics of Rare Activated Processes from Accelerated Molecular Dynamics Using Kramers' Theory. *J. Chem. Theory Comput.* **2011**, *7* (3), 575-581; (b) Frank, A. T.; Andricioaei, I., Reaction Coordinate-Free Approach to Recovering Kinetics from Potential-Scaled Simulations: Application of Kramers' Rate Theory. *J. Phys. Chem. B* **2016**, *120* (33), 8600-8605.
39. Hamelberg, D.; Shen, T.; Andrew McCammon, J., Relating kinetic rates and local energetic roughness by accelerated molecular-dynamics simulations. *J. Chem. Phys.* **2005**, *122* (24), 241103.
40. Liu, L.; Marwitz, A. J.; Matthews, B. W.; Liu, S. Y., Boron Mimetics: 1, 2-Dihydro-1, 2-azaborines Bind inside a Nonpolar Cavity of T4 Lysozyme. *Angewandte Chemie International Edition* **2009**, *48* (37), 6817-6819.
41. Baldwin, E.; Baase, W. A.; Zhang, X.-j.; Feher, V.; Matthews, B. W., Generation of ligand binding sites in T4 lysozyme by deficiency-creating substitutions. *Journal of molecular biology* **1998**, *277* (2), 467-485.
42. Morton, A.; Matthews, B. W., Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: linkage of dynamics and structural plasticity. *Biochemistry* **1995**, *34* (27), 8576-8588.
43. 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. *Journal of chemical theory and computation* **2015**, *11* (8), 3696-3713.

44. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *Journal of computational chemistry* **2004**, *25* (9), 1157-1174.
45. Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I.; Brozell, S. R.; Cerutti, D. S.; Cheatham III, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E., *Amber 2022*. University of California, San Francisco: 2022.
46. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *Journal of Molecular Graphics & Modelling* **1996**, *14* (1), 33-38.
47. Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* **2013**, *9* (7), 3084-3095.
48. Paramo, T.; East, A.; Garzón, D.; Ulmschneider, M. B.; Bond, P. J., Efficient Characterization of Protein Cavities within Molecular Simulation Trajectories: trj_cavity. *Journal of Chemical Theory and Computation* **2014**, *10* (5), 2151-2164.
49. Feher, V. A.; Schiffer, J. M.; Mermelstein, D. J.; Mih, N.; Pierce, L. C. T.; McCammon, J. A.; Amaro, R. E., Mechanisms for Benzene Dissociation through the Excited State of T4 Lysozyme L99A Mutant. *Biophysical Journal* **2019**, *116* (2), 205-214.
50. Wang, J.; Miao, Y., Peptide Gaussian accelerated molecular dynamics (Pep-GaMD): Enhanced sampling and free energy and kinetics calculations of peptide binding. *J. Chem. Phys.* **2020**, *153* (15), 154109.
51. (a) Rocklin, G. J.; Boyce, S. E.; Fischer, M.; Fish, I.; Mobley, D. L.; Shoichet, B. K.; Dill, K. A., Blind Prediction of Charged Ligand Binding Affinities in a Model Binding Site. *Journal of Molecular Biology* **2013**, *425* (22), 4569-4583; (b) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R., Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *Journal of the American Chemical Society* **2015**, *137* (7), 2695-2703; (c) Perez, A.; Morrone, J. A.; Simmerling, C.; Dill, K. A., Advances in free-energy-based simulations of protein folding and ligand binding. *Current Opinion in Structural Biology* **2016**, *36*, 25-31.
52. (a) Wang, Y.; Papaleo, E.; Lindorff-Larsen, K., Mapping transiently formed and sparsely populated conformations on a complex energy landscape. *eLife* **2016**, *5*, e17505; (b) Rydzewski, J.; Valsson, O., Finding multiple reaction pathways of ligand unbinding. *The Journal of Chemical Physics* **2019**, *150* (22), 221101.
53. (a) Wang, Y.; Ribeiro, J. M. L.; Tiwary, P., Past-future information bottleneck for sampling molecular reaction coordinate simultaneously with thermodynamics and kinetics. *Nature Communications* **2019**, *10* (1), 3573; (b) Lamim Ribeiro, J. M.; Tiwary, P., Toward Achieving Efficient and Accurate Ligand-Protein Unbinding with Deep Learning and Molecular Dynamics through RAVE. *Journal of Chemical Theory and Computation* **2019**, *15* (1), 708-719.
54. (a) Schreiber, G.; Fersht, A. R., Energetics of protein-protein interactions: Analysis of the Barnase-Barstar interface by single mutations and double mutant cycles. *J. Mol. Biol.* **1995**, *248* (2), 478-486; (b) Schreiber, G.; Fersht, A. R., Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry* **1993**, *32* (19), 5145-50.
55. Ahmad, M.; Gu, W.; Geyer, T.; Helms, V., Adhesive water networks facilitate binding of protein interfaces. *Nat. Commun.* **2011**, *2*, 261.
56. (a) Banerjee, P.; Bagchi, B., Dynamical control by water at a molecular level in protein dimer association and dissociation. *Proc. Natl. Acad. Sci. USA* **2020**, *117* (5), 2302-2308; (b) Antoszewski, A.; Feng, C. J.; Vani, B. P.; Thiede, E. H.; Hong, L.; Weare, J.; Tokmakoff, A.; Dinner, A. R., Insulin Dissociates by Diverse Mechanisms of Coupled Unfolding and Unbinding. *J. Phys. Chem. B* **2020**, *124* (27), 5571-5587.
57. Saglam, A. S.; Chong, L. T., Protein-protein binding pathways and calculations of rate constants using fully-continuous, explicit-solvent simulations. *Chem Sci* **2019**, *10* (8), 2360-2372.

58. (a) Banerjee, P.; Mondal, S.; Bagchi, B., Insulin dimer dissociation in aqueous solution: A computational study of free energy landscape and evolving microscopic structure along the reaction pathway. *J. Chem. Phys.* **2018**, *149* (11), 114902; (b) Siebenmorgen, T.; Zacharias, M., Efficient Refinement and Free Energy Scoring of Predicted Protein-Protein Complexes Using Replica Exchange with Repulsive Scaling. *J. Chem. Inf. Model.* **2020**, *60* (11), 5552-5562.
59. (a) Abrams, C.; Bussi, G., Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy* **2014**, *16* (1), 163-199; (b) Miao, Y.; McCammon, J. A., Unconstrained Enhanced Sampling for Free Energy Calculations of Biomolecules: A Review. *Mol. Simul.* **2016**, *42* (13), 1046-1055.

Table 1. Summary of LiGaMD2 simulations performed on ligand binding to the T4L mutants. ΔV is the total boost potential. N_D and N_B are the number of observed ligand dissociation and binding events, respectively. ΔG_{sim} and ΔG_{exp} are the ligand-T4L binding free energies obtained from LiGaMD2 simulations and experiments, respectively. ΔG_{sim}^a and ΔG_{sim}^b were calculated with 3D PMF and reweighted binding kinetic rates by $\Delta G = -RT\ln(k_{off}/k_{on})$, respectively.

T4L	Ligand	ID	N_B	N_D	ΔV (kcal/mol)	ΔG_{sim}^a (kcal/mol)	ΔG_{sim}^b (kcal/mol)	ΔG_{exp} (kcal/mol)
L99A	BEN	Sim1	6	6	107.77±9.67	-5.88±0.61	-5.17±0.72	-4.12
		Sim2	6	6	109.84±9.58			
		Sim3	5	5	109.63±9.59			
M102A	BEN	Sim1	6	6	108.10±10.04	-6.43±0.12	-5.01±0.73	-4.18
		Sim2	3	3	108.04±9.97			
		Sim3	7	7	109.21±9.49			
F104A	BEN	Sim1	4	4	70.96±8.11	-6.02±0.70	-3.42±0.72	-4.02
		Sim2	3	3	72.53±7.84			
		Sim3	3	3	70.45±7.81			
L99A	IND	Sim1	6	7	115.80±9.68	-7.40±0.39	-4.87±1.06	-4.82
		Sim2	3	3	118.16±9.79			
		Sim3	4	4	116.49±9.71			

Table 2 Comparison of kinetic rates obtained from experiments and LiGaMD2 simulations for ligand binding to T4L mutants. k_{on} and k_{off} are the kinetic dissociation and binding rate constants, respectively, from experimental data or LiGaMD2 simulations with reweighting using Kramers' rate theory. k_{on}^* and k_{off}^* are the accelerated kinetic dissociation and binding rate constants calculated directly from LiGaMD2 simulations without reweighting.

System	Method	$k_{on} (M^{-1} \cdot s^{-1})$	$k_{off} (s^{-1})$	$k_{on}^* (M^{-1} \cdot s^{-1})$	$k_{off}^* (s^{-1})$
T4L:L99A-BEN	Experiment	$0.8-1.0 \times 10^6$	9.50×10^2	-	-
	LiGaMD2	$7.42 \pm 4.81 \times 10^6$	$1.44 \pm 0.88 \times 10^3$	$8.22 \pm 5.48 \times 10^7$	$3.47 \pm 2.31 \times 10^5$
T4L:M102A-BEN	Experiment	$3.0-5.0 \times 10^6$	3.00×10^3		
	LiGaMD2	$9.57 \pm 6.29 \times 10^6$	$2.01 \pm 1.61 \times 10^3$	$7.79 \pm 1.36 \times 10^6$	$3.46 \pm 1.81 \times 10^9$
T4L:F104A-BEN	Experiment	$>10 \times 10^6$	$>1.00 \times 10^4$		
	LiGaMD2	$3.16 \pm 2.29 \times 10^9$	$1.38 \pm 0.67 \times 10^6$	$6.81 \pm 1.27 \times 10^7$	$1.72 \pm 1.44 \times 10^9$
T4L:L99A-IND	Experiment	$0.7-1.0 \times 10^6$	3.25×10^2		
	LiGaMD2	$2.99 \pm 2.87 \times 10^6$	$3.49 \pm 0.56 \times 10^3$	$1.67 \pm 0.67 \times 10^9$	$2.49 \pm 1.25 \times 10^7$

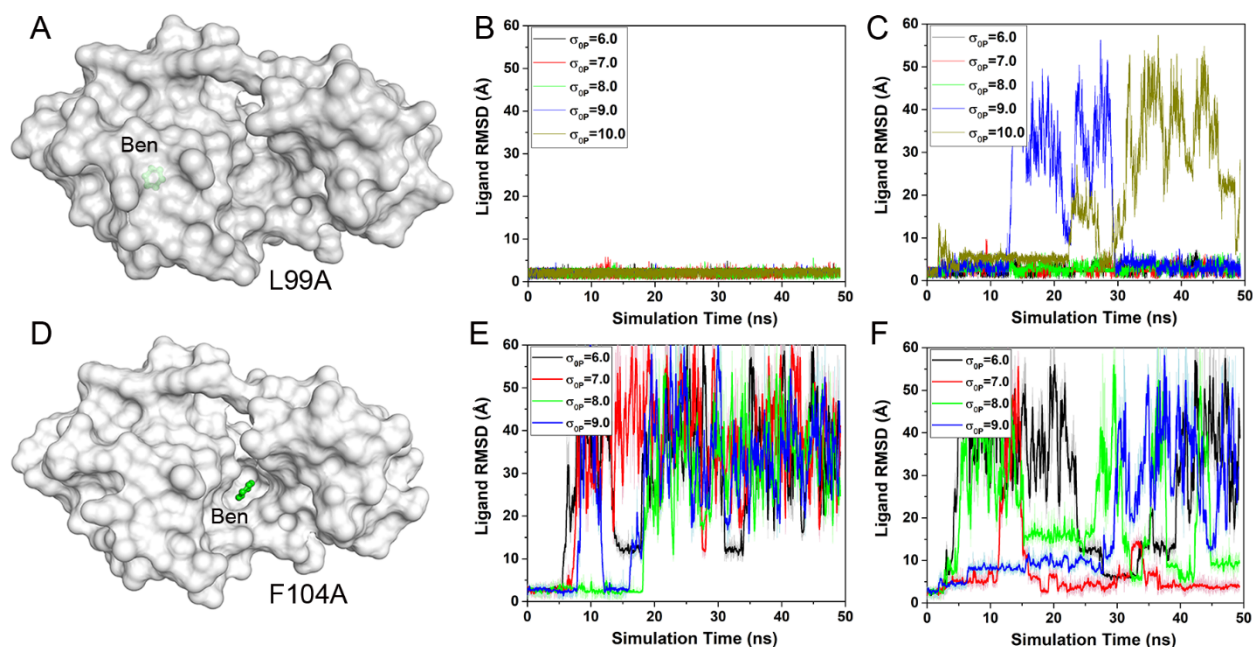


Figure 1. Comparison of LiGaMD and LiGaMD2 simulations on the T4L mutant systems with buried and open binding pockets: Computational models of benzene binding to the T4L:L99A with a burred binding pocket (**A**) and T4L:F104A with an open binding pocket (**D**); Time courses of ligand root-mean-square deviation (RMSD) in T4L:L99A calculated from 49.2 ns LiGaMD (**B**) and LiGaMD2 (**C**) equilibration simulations, respectively. Time courses of ligand RMSD in F104A T4L calculated from 49.2 ns LiGaMD (**E**) and LiGaMD2 (**F**) equilibration simulations, respectively.

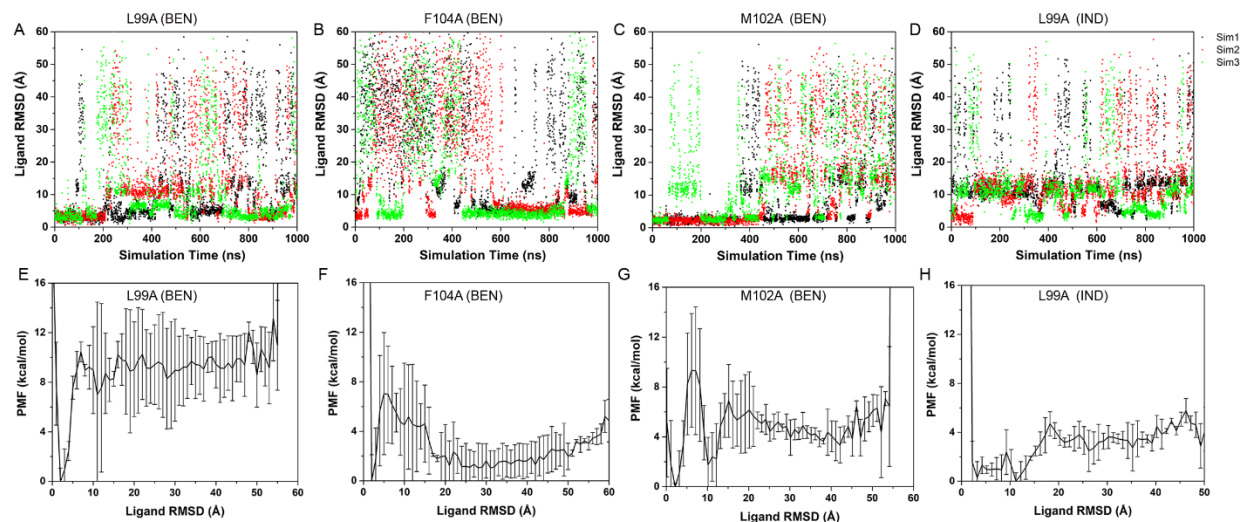


Figure 2. LiGaMD2 simulations captured repetitive dissociation and binding of two different ligands (benzene and indole) to T4L mutants: (A–D) time courses of ligand heavy atom RMSDs relative to X-ray structures calculated from three independent 1 μ s LiGaMD2 simulations of (A) benzene binding to T4L:L99A, (B) benzene binding to T4L:F104A, (C) benzene binding to the T4L:M102A and (D) indole binding to the T4L:L99A. (E–H) The corresponding PMF profiles of the ligand RMSDs averaged over three LiGaMD2 simulations of (E) benzene binding to T4L:L99A, (F) benzene binding to T4L:F104A, (G) benzene binding to the T4L:M102A and (H) indole binding to the T4L:L99A. Error bars are standard deviations of the free energy values calculated from three LiGaMD2 simulations.

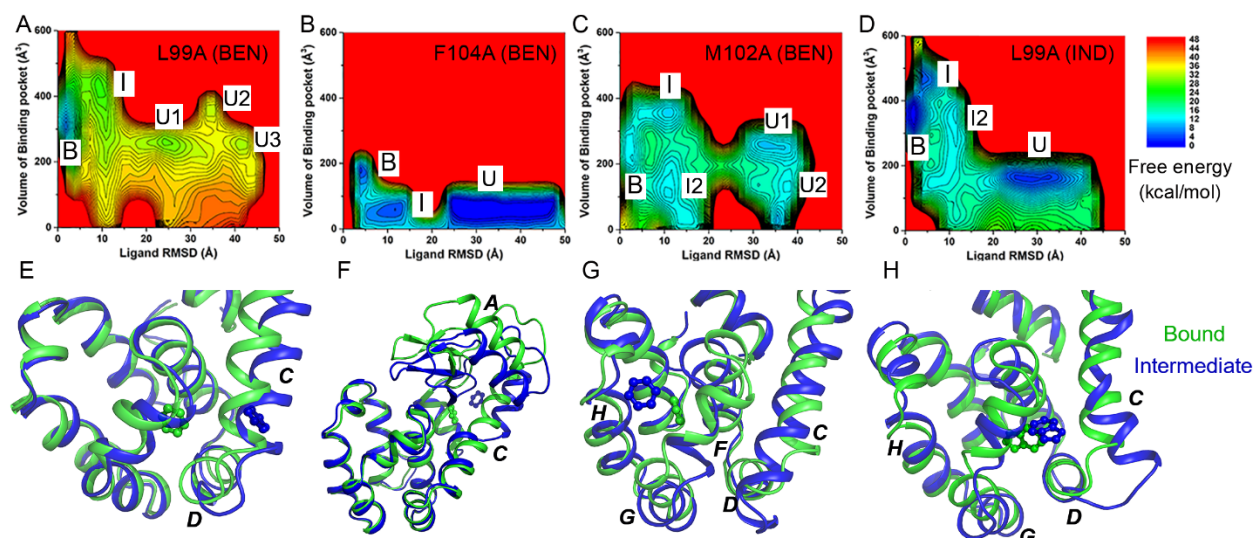


Figure 3. 2D Free energy profiles and low-energy intermediate conformational states of ligand binding to the T4L mutants: (A-D) 2D PMF profiles regarding the ligand heavy atom RMSD and the pocket volume in LiGaMD2 simulations of (A) benzene binding to T4L:L99A, (B) benzene binding to T4L:F104A, (C) benzene binding to T4L:M102A, (D) indole binding to T4L:F99A. (E-H) Low-energy “Intermediate” (“I”) conformations (blue) as identified from the 2D PMF profiles of (E) benzene binding to T4L:L99A, (F) benzene binding to T4L:F104A, (G) benzene binding to T4L:M102A and (H) indole binding to T4L:L99A. X-ray structures of the ligand-bound complexes (“Bound”) are shown in green. The ligands are shown in balls and sticks, and the helix are shown in cartoon. The helix C, D, F and G are labeled as they show significant changes between the “Bound” and “Intermediate” conformational states.

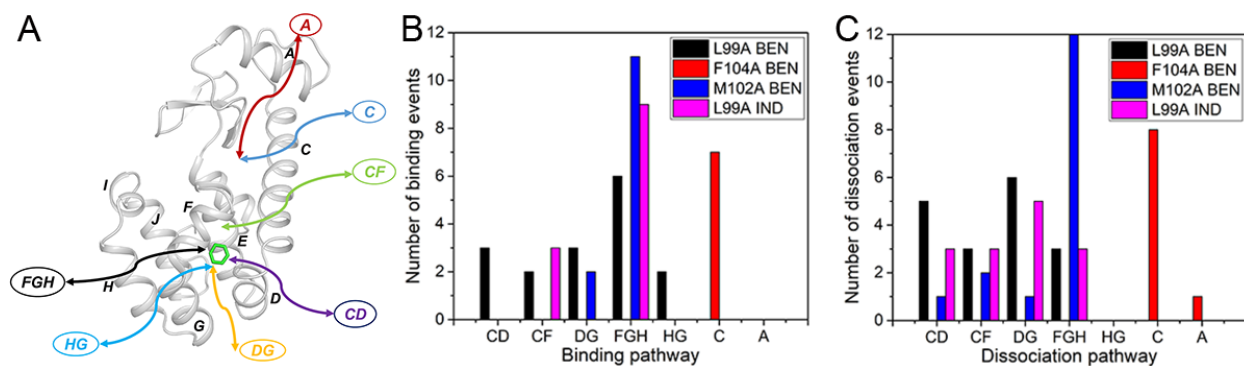
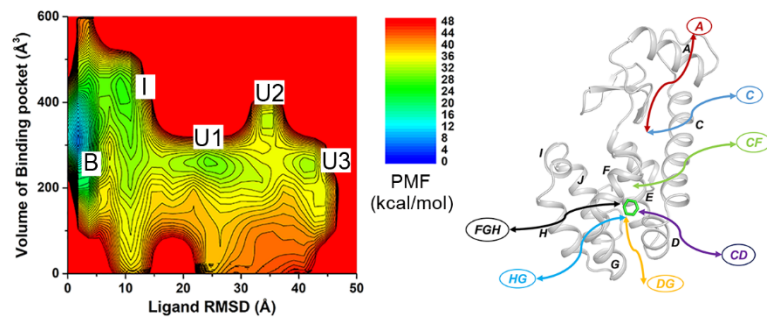


Figure 4. Pathways of ligand binding and dissociation in the T4L mutants. (A) Cartoon representation of the protein with helices labelled. Binding and dissociation pathways are denoted by the arrow lines. Number of binding (B) and dissociation (C) events through the different pathways captured by the LiGaMD2 simulations.

Table of content



**Ligand Gaussian accelerated molecular dynamics 2 (LiGaMD2):
Improved calculations of ligand binding thermodynamics and
kinetics with closed protein pocket**

Jinan Wang and Yinglong Miao*

Center for Computational Biology and Department of Molecular Biosciences, University of
Kansas, Lawrence, KS 66047, USA

* Corresponding email: miao@ku.edu

Table S1 The ligand bound and unbound time periods (τ_B and τ_U) recorded from LiGaMD2 simulations of the ligand-T4L binding system.

System	ID	τ_B (ns)	τ_U (ns)
L99A-BEN	Sim1	85.3,77.3,70.5,21.0,140.4,41.9	42.0,35.3,61.8,70.7,170.6,83.3
	Sim2	212.49,26.2,19.6,13.0,111.2,32.9	218.80,106.3,38.0,64.6,13.2,43.8
	Sim3	113.9,35.5,14.3,94.4,278.2	19.7,136.34,173.3,110.1,25.17
M102A-BEN	Sim1	12.99,187,6,45.8,26.2	490.7,56.1,156.2,24.2
	Sim2	54.94,36.0,362.6	240.01,281.4,25.5
	Sim3	112.9,419.21,40.9	87.78,240.1,101.1
F104A-BEN	Sim1	360.7,191.6,64.29,10.5,32.9,17.1	85.2,17.1,26.3,40.5,43.3,101.5
	Sim2	456.4,47.2,18.4	261.1,110.6,96.3
	Sim3	68.17,147.0,57.6,10.5,9.8,45.0,8.9	129.8,35.4,56.6,45.4,108.2,59.2,221.0
L99A-IND	Sim1	6.84,26.7,14.8,100.8,20.7,9.9	11.85,329.0,127.5,97.8,23.0,41.5,189.9
	Sim2	89.14,151.3,26.7	347.5,88.9,412.3
	Sim3	16.87,81.4,21.0,171.7	284.8,73.9,217.7,122.7

Table S2 Energy barriers of ligand dissociation (“off”) and binding (“on”) calculated from the reweighed (ΔF) and modified (no reweighting, ΔF^*) free energy profiles, curvatures of the reweighed (w) and modified (w^*) free energy profiles near the ligand Bound (“B”), Barrier (“Br”) and Unbound (“U”) states, and the ratio of apparent diffusion coefficients calculated from the LiGaMD2 simulations without reweighting (modified, D^*) and with reweighting (D).

Sim	ΔF (kcal/mol)		ΔF^* (kcal/mol)		w			w^*			D^*/D	
	Off	On	Off	On	B	Br	U	B	Br	U	Off	On
L99A-BEN	10.5±0.86	4.93±1.38	2.55±0.43	0.57±0.14	9.78±9.20	0.95±0.29	0.07±0.03	0.87±0.81	0.077±0.034	0.02±0.001	4.42	17.67
F104A-BEN	9.00 ± 1.07	6.03±0.11	2.35 ± 0.34	1.15 ± 0.50	22.38 ± 11.42	2.05±0.75	0.15 ± 0.05	0.50 ± 0.11	0.054 ± 0.037	0.035 ± 0.0017	6.16	813.11
M102A-BEN	8.55 ± 0.54	5.40±1.07	2.36 ± 0.35	0.54 ± 0.053	3.11±0.29	1.98±1.23	0.57 ± 0.14	0.57 ± 0.14	0.018 ± 0.014	0.0028 ± 0.0112	86.51	3519.25
L99A-IND	7.11 ± 0.34	4.37±0.33	2.23 ± 0.26	0.53 ± 0.10	2.12±0.83	0.14±0.09	0.27 ± 0.15	0.21 ± 0.06	0.14±0.07	0.0075 ± 0.0017	26.89	467.56

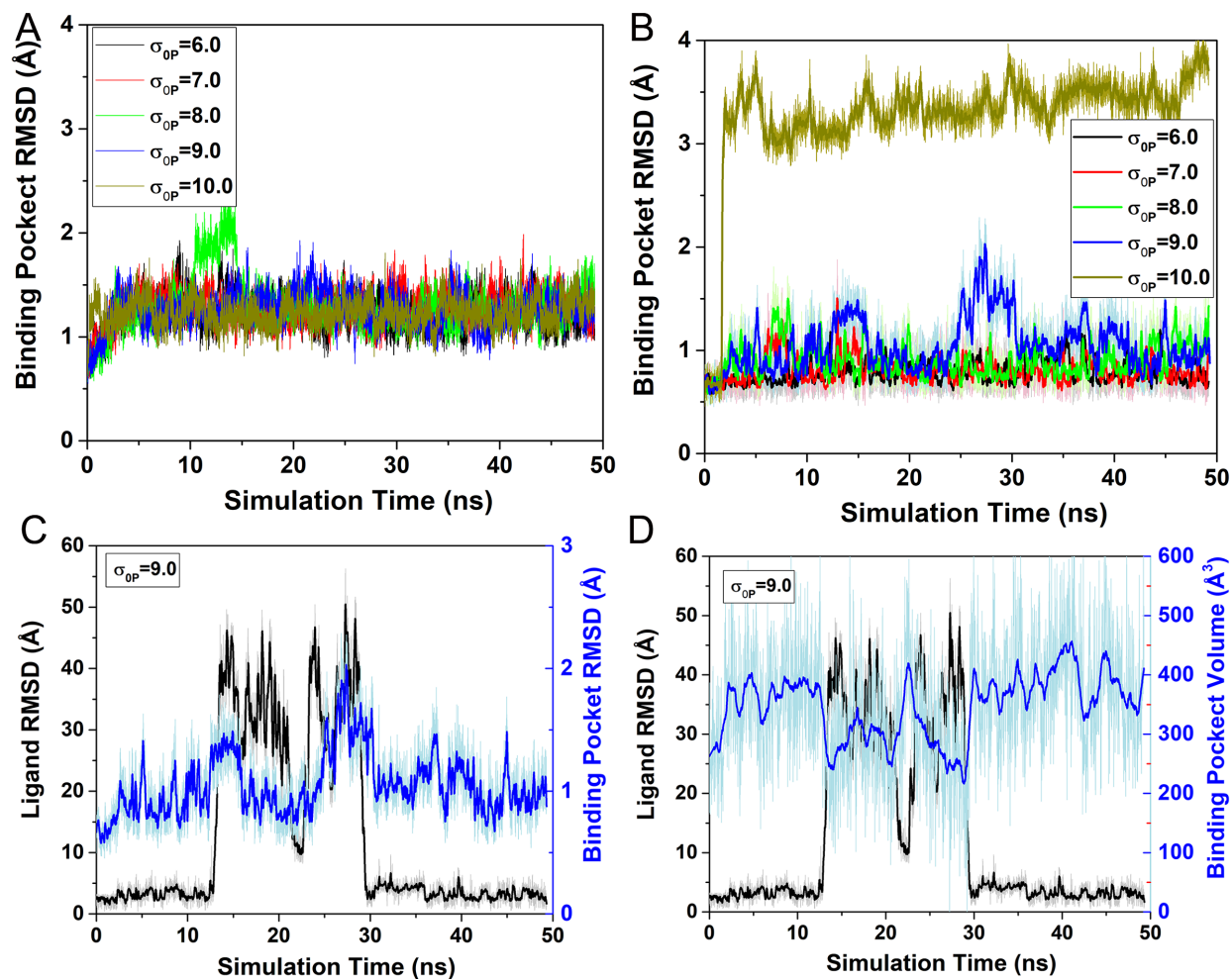


Figure S1. Time courses of RMSD of the ligand binding pocket relative to the X-ray structure in L99A T4L calculated from 49.2 ns LiGaMD (A) and LiGaMD2 (B) equilibration simulations; (C) Time courses of RMSD of the ligand relative to the X-ray structure and ligand binding pocket relative to the X-ray structure in L99A T4L calculated from 49.2 ns LiGaMD2 with σ_{OP} at 9.0 kcal/mol. (D) Time courses of RMSD of ligand relative to the X-ray structure in L99A T4L and pocket volume calculated from 49.2 ns LiGaMD2 with σ_{OP} at 9.0 kcal/mol.

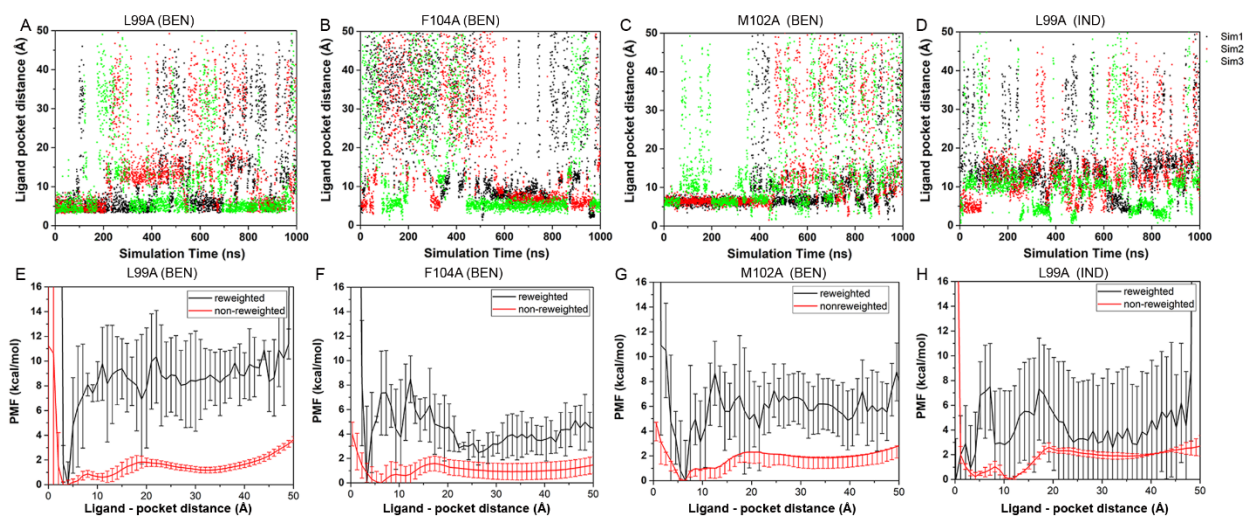


Figure S2. (A-D) time courses of the center-of-mass distance between the ligand and the protein pocket (defined by protein residues within 5 Å of ligand) calculated from three independent 1 μs LiGaMD2 simulations of (A) benzene binding to the L99A T4L, (B) benzene binding to the F104A T4L, (C) benzene binding to the M102A T4L, and (D) indole binding to the L99A T4L. (E-H) The corresponding PMF profiles of the ligand-pocket distance averaged over three LiGaMD2 simulations of (E) benzene binding to L99A T4L, (F) benzene binding to F104A T4L, (G) benzene binding to M102A T4L, and (H) indole binding to L99A T4L. Error bars are standard deviations of the free energy values calculated from three LiGaMD2 simulations.