Ligand Gaussian accelerated molecular dynamics (LiGaMD): Characterization of ligand binding thermodynamics and kinetics

Yinglong Miao[*], **Apurba Bhattarai**, **Jinan Wang**

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

Abstract

Calculations of ligand binding free energies and kinetic rates are important for drug design. However, such tasks have proven challenging in computational chemistry and biophysics. To address this challenge, we have developed a new computational method "LiGaMD", which selectively boosts the ligand non-bonded interaction potential energy based on the Gaussian accelerated molecular dynamics (GaMD) enhanced sampling technique. Another boost potential could be applied to the remaining potential energy of the entire system in a dual-boost algorithm (LiGaMD_Dual) to facilitate ligand binding. LiGaMD has been demonstrated on host-guest and protein-ligand binding model systems. Repetitive guest binding and unbinding in the β-cyclodextrin host were observed in hundreds-of-nanosecond LiGaMD_Dual simulations. The calculated guest binding free energies agreed excellently with experimental data with < 1.0 kcal/mol errors. Compared with converged microsecond-timescale conventional molecular dynamics simulations, the sampling errors of LiGaMD_Dual simulations were also < 1.0 kcal/mol. Accelerations of ligand kinetic rate constants in LiGaMD simulations were properly estimated using Kramers' rate theory. Furthermore, LiGaMD allowed us to capture repetitive dissociation and binding of the benzamidine inhibitor in trypsin within 1 μs simulations. The calculated ligand binding free energy and kinetic rate constants compared well with the experimental data. In summary, LiGaMD provides a powerful enhanced sampling approach for characterizing ligand binding thermodynamics and kinetics simultaneously, which is expected to facilitate computer-aided drug design.

Graphical Abstract

[*]Corresponding: miao@ku.edu.
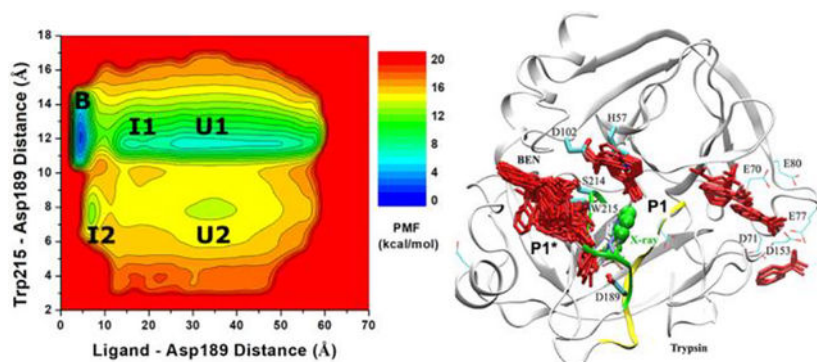
All-atom ligand Gaussian accelerated molecular dynamics (LiGaMD) simulations captured repetitive binding of the benzamidine (BEN) ligand to the trypsin model protein, which enabled us to characterize the ligand binding free energy profiles, pathways and kinetic rate constants.

Keywords

Ligand Gaussian accelerated molecular dynamics (LiGaMD); enhanced sampling; ligand binding; thermodynamics; free energy; kinetic rate constants

Introduction

Free energy calculations of protein-ligand binding have been central to computational chemistry and biophysics research for the last several decades[1]. This is largely motivated by designing potent drug molecules with high binding free energies in the pharmaceutical industry[2]. A number of computational methods that have been developed in the field include thermodynamic integration (TI)[3], free energy perturbation (FEP)[4], double decoupling method[5], umbrella sampling[6], steered molecular dynamics (SMD)[7], funnel metadynamics[8], molecular mechanics Poisson-Boltzmann surface area and generalized Born surface area (MM/PBSA and MM/GBSA)[9], and so on. Ligand binding free energy calculations have also been the focus of Statistical Assessment of Modeling of Proteins and Ligands (SAMPL)[10] and Drug Design Data Resource (D3R)[11] community challenges.

Kinetics of ligand binding have recently been recognized to be potentially more relevant for drug design. In particular, the dissociation rate constant that determines the drug residence time appears to better correlate with drug efficacy than the binding free energy[12]. However, ligand kinetic rates have proven even more difficult to compute than the binding free energies, largely due to slow processes of ligand binding and dissociation over long timescales[12b]. With remarkable advances in supercomputing and method developments, molecular dynamics (MD) simulations using specialized hardware[13], Markov state model (MSM)[14] and enhanced sampling[15] have extended our capabilities to capture spontaneous ligand binding to target proteins. A number of enhanced sampling methods, including metadynamics[16], random acceleration MD[17], SMD[18], weighted ensemble[19] and selectively scaled MD[20] have also been able to dissociate ligand molecules from the proteins. However, it remains extremely challenging to simulate repetitive ligand binding and dissociation processes, precluding accurate calculations of ligand kinetic rate constants.

With relatively small size and reduced complexity, host-guest binding systems has served as models of protein-ligand binding in SAMPL challenges[10]. The studied hosts include the cucurbit[7]uril (CB7), octa-acid, and CBClip and cyclodextrin (CD). These hosts are compounds that are typically much smaller than proteins but still large enough to bind guest molecules through non-covalent interactions, which share common characteristics as protein-ligand binding. Host-guest systems thus represent model systems for testing and improving free energy calculation methods. In addition to thermodynamics, Chang and co-workers have investigated kinetics of fast guest binding and dissociation in the β-CD host[21]. Tens to hundreds of repetitive guest binding and dissociation events were observed in microsecond-timescale MD simulations, which enabled comprehensive characterization of both thermodynamics and kinetics of the host-guest binding.

The MSM[22] and metadynamics[8, 23] have been applied to investigate the thermodynamics and kinetics of protein-ligand binding, using particularly the benzamidine inhibitor binding to trypsin as a model system. Multiple metadynamics trajectories with a total of 5 μs simulation time were obtained to predict the ligand unbinding pathways and dissociation rate constant, k_{off}. The calculated $k_{\text{off}} = 9.1 \pm 2.5 \text{ s}^{-1}$ was suggested to be in agreement of the experimental value $k_{off}^{exp} = 600 \pm 300 \text{ s}^{-1}$. Separate funnel metadynamics simulations also allowed calculations of ligand binding free energies, in particular -8.5 ± 0.7 kcal/mol for the trypsin-benzamidine system[8]. The MSM built with 150 μs MD simulation data provided a complex picture of ligand binding kinetics and protein conformational plasticity[22]. The simulation predicted $k_{\text{off}} = 131 \pm 109 \times 10^2 \text{ s}^{-1}$ appeared to be faster than the experimental value, while the predicted binding rate constant $k_{\text{on}} = 6.4 \pm 1.6 \times 10^7 \text{ M}^{-1}\cdot\text{s}^{-1}$ compared well with the experimental value of $k_{on}^{exp} = 2.9 \times 10^7 \text{ M}^{-1} \cdot \text{s}^{-1}$. Repetitive ligand binding and unbinding in trypsin were also captured using a selective integrated-tempering-sampling molecular dynamics (SITSMD) method recently[24]. The binding free energy of benzamidine in trypsin was estimated as -5.06 kcal/mol, but the ligand kinetic rates were not calculated from the SITSMD simulations.

Here, we present a new computational method called ligand Gaussian accelerated molecular dynamics ("LiGaMD"), which enables us to simulate repetitive ligand binding and unbinding, and thus characterize both thermodynamics and kinetics of ligand binding simultaneously. Gaussian accelerated molecular dynamics (GaMD) is an enhanced sampling computational technique that works by adding a harmonic boost potential to smooth the biomolecular potential energy surface.[25] GaMD greatly reduces energy barriers and accelerates biomolecular simulations by orders of magnitude.[26] GaMD does not require pre-defined collective variables or reaction coordinates. Compared with the enhanced sampling methods that rely on careful selection of the collective variables, GaMD is of particular advantage for studying complex biological processes such as ligand binding to proteins[27]. Moreover, because the boost potential follows a Gaussian distribution, biomolecular free energy profiles can be properly recovered through cumulant expansion to the second order.[25] GaMD builds on the previous accelerated MD (aMD) method[28], but solves its energetic reweighting problem[29] for free energy calculations of large biomolecules. GaMD has been implemented in the widely used AMBER[25, 30], NAMD[31] and GENESIS[32] packages. GaMD has successfully revealed physical pathways and mechanisms of protein folding and ligand

binding, which are consistent with experiments and long-timescale conventional MD (cMD) simulations.[25, 31, 33] It has also been applied to characterize protein-protein[34], protein-membrane,[35] and protein-nucleic acid[36] interactions.

Building upon GaMD, we have developed LiGaMD for more efficient sampling simulations of protein-ligand binding and unbinding processes. In LiGaMD, the non-bonded electrostatic and van der Waals interactions between the bound ligand and protein/environment are selectively boosted to enable ligand dissociation. In this context, selective acceleration has been found useful in previous enhanced sampling techniques, including the selective aMD[37], selectively scaled MD[20], essential energy space random walk[38], replica exchange solute tempering (REST)[39] and REST2[40]. In addition, a number of unbound ligand molecules are added in the solvent and another boost potential is applied on these ligand molecules, the protein and solvent in a dual-boost LiGaMD (LiGaMD_Dual) algorithm to facilitate ligand rebinding.

The LiGaMD method has been demonstrated on host-guest and protein-ligand binding model systems. Repetitive guest binding and unbinding in the β-CD host have been observed in hundreds-of-nanosecond LiGaMD simulations. The LiGaMD predicted guest binding free energies and kinetic rate constants agree well with those from previous cMD simulations[21] and experimental data. Furthermore, LiGaMD has also allowed us to capture multiple dissociation and rebinding events of the benzamidine inhibitor in trypsin within 1 μs simulations. The reweighted ligand binding free energy and kinetic rate constants compared well with the experimental data.

Methods

Ligand Gaussian Accelerated Molecular Dynamics (LiGaMD)

Gaussian accelerated molecular dynamics (GaMD) is an enhanced sampling technique that works by adding a harmonic boost potential to smooth biomolecular potential energy surface and reduce the system energy barriers[25]. Details of the GaMD method have been described in previous studies[25, 31, 33a]. A brief summary is provided in Appendix A. Here, we develop a new ligand GaMD (LiGaMD) method for more efficient sampling of protein-ligand binding.

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 $r \equiv \{\vec{r}_1, \cdots, \vec{r}_N\}$ and momenta $p \equiv \{\vec{p}_1, \cdots, \vec{p}_N\}$. The system Hamiltonian can be expressed as:

$$H(r, p) = K(p) + V(r), \tag{1}$$

where $K(p)$ and $V(r)$ are the system kinetic and total potential energies, respectively. Next, we decompose the potential energy into the following terms:

$$V(r) = V_{P,b}(r_P) + V_{L,b}(r_L) + V_{E,b}(r_E) + V_{PP,nb}(r_P) + V_{LL,nb}(r_L) \\ + V_{EE,nb}(r_E) + V_{PL,nb}(r_{PL}) + V_{PE,nb}(r_{PE}) + V_{LE,nb}(r_{LE}). \tag{2}$$

where $V_{P,b}$, $V_{L,b}$ and $V_{E,b}$ are the bonded potential energies in protein P, ligand L and environment E, respectively. $V_{PP,nb}$, $V_{LL,nb}$ and $V_{EE,nb}$ are the self non-bonded potential energies in protein P, ligand L and environment E, respectively. $V_{PL,nb}$, $V_{PE,nb}$ and $V_{LE,nb}$ are the non-bonded interaction energies between P-L, P-E and L-E, respectively. According to classical molecular mechanics force fields[41], the non-bonded potential energies are usually calculated as:

$$V_{nb} = V_{elec} + V_{vdW} . \tag{3}$$

Where V_{elec} and V_{vdW} are the system electrostatic and van der Waals potential energies. Presumably, ligand binding mainly involves the non-bonded interaction energies of the ligand, $V_{L,nb}(r) = V_{LL,nb}(r_L) + V_{PL,nb}(r_{PL}) + V_{LE,nb}(r_{LE})$. In LiGaMD, we add a boost potential selectively to the ligand non-bonded potential energy as:

$$\Delta V_{L,nb}(r) = \begin{cases} \frac{1}{2}k_{L,nb}\big(E_{L,nb} - V_{L,nb}(r)\big)^2, & V_{L,nb}(r) < E_{L,nb} \\ 0, & V_{L,nb}(r) \ge E_{L,nb} \end{cases} \tag{4}$$

where $E_{L,nb}$ is the threshold energy for applying boost potential and $k_{L,nb}$ is the harmonic constant. For simplicity, the subscript of $V_{L,nb}(r)$, $E_{L,nb}$ and $k_{L,nb}$ is dropped in the following. The LiGaMD simulation parameters are derived similarly as in the previous GaMD algorithm (Appendix A). 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:

$$k_0 = \min(1.0, k_0') = \min\left(1.0, \frac{\sigma_0}{\sigma_V}\frac{V_{max} - V_{min}}{V_{max} - V_{avg}}\right), \tag{5}$$

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 V (e.g., $10k_BT$) for proper reweighting. The harmonic constant is calculated as $k = k_0 \cdot \frac{1}{V_{max} - V_{min}}$ with $0 < k_0 \le 1$. Alternatively, when the threshold energy E is set to its upper bound $E = V_{min} + \frac{1}{k}$, k_0 is set to:

$$k_0 = k_0'' \equiv \left(1 - \frac{\sigma_0}{\sigma_V}\right)\frac{V_{max} - V_{min}}{V_{avg} - V_{min}}, \tag{6}$$

if k_0'' is found to be between 0 and 1. Otherwise, k_0 is calculated using (Eqn. 5).

Next, one can add multiple ligand molecules in the solvent to facilitate ligand binding to proteins in MD simulations[13b, 42]. This is based on the fact that the ligand binding rate constant k_{on} is inversely proportional to the ligand concentration. The higher the ligand concentration, the faster the ligand binds, provided that the ligand concentration is still within its solubility limit. In addition to selectively boosting the bound ligand to accelerate its dissociation, another boost potential could thus be applied on the unbound ligand molecules, protein and solvent to facilitate ligand rebinding. The second boost potential is

calculated using the total system potential energy other than the non-bonded potential energy of the bound ligand as:

$$\Delta V_D(r) = \begin{cases} \frac{1}{2}k_D(E_D - V_D(r))^2, V_D(r) < E_D \\ \\ 0, V_D(r) \geq E_D \end{cases} \tag{7}$$

where E_D and k_D are the corresponding threshold energy for applying the second boost potential and the harmonic constant, respectively. This leads to dual-boost LiGaMD (LiGaMD_Dual) with the total boost potential $V(r) = V_{L,nb}(r) + V_D(r)$.

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 following[14]:

$$\Delta G^0 = -\Delta W_{3D} - RTLn\frac{V_b}{V_0}, \tag{9}$$

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 W_{3D} is the depth of the 3D PMF. W_{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:

$$\Delta W_{3D} = -RTLn\frac{\int_u e^{-\beta W(r)}dr}{\int_u dr}, \tag{10}$$

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[14]. A python script PyReweighting-3D.py is freely available in the PyReweighting tool kit (http://miao.compbio.ku.edu/PyReweighting/) for calculating the 3D PMF and associated ligand binding free energies. It works for both cMD (without energetic reweighting) and enhanced sampling simulations using LiGaMD with energetic reweighting (Appendix B).

Ligand Binding Kinetics obtained from Reweighting of LiGaMD Simulations

Provided sufficient sampling of repetitive ligand dissociation and binding in simulations, one can 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 design[43]. Then the ligand dissociation and binding rate constants (k_{off} and k_{on}) were calculated as:

$$k_{off} = \frac{1}{\tau_B}. \tag{11}$$

$$k_{on} = \frac{1}{\tau_U \cdot [L]},$$ (12)

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

Two algorithms were implemented using the transition state theory (TST) and Kramers' rate theory for reweighting kinetics of the LiGaMD simulations. According to Kramers' Rate Theory (Appendix C), the rate of a chemical reaction in the large viscosity limit is calculated as[26]:

$$k_R \cong \frac{2\pi w_m w_b}{\xi} e^{-\Delta F / k_B T},$$ (13)

where w_m and w_b are frequencies of the approximated harmonic oscillators (also referred to as curvatures of free energy surface[44]) near the energy minimum and barrier, respectively, ξ is the frictional rate constant and F 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 equation[45] (Appendix C). In order to reweight ligand kinetics from the LiGaMD simulations using the Kramer's rate theory, the free energy barriers of ligand binding and dissociation are calculated from the original (reweighted, F) and modified (no reweighting, F^*) PMF profiles, similarly for curvatures of the reweighed (w) and modified (w^*, no reweighting) PMF profiles near the guest bound ("B") and unbound ("U") low-energy wells and the energy barrier ("Br"), and the ratio of apparent diffusion coefficients from LiGaMD simulations without reweighting (modified, D^*) and with reweighting (D). The resulting numbers are then plugged into Eq. (11) to estimate accelerations of the ligand binding and dissociation rates during LiGaMD simulations (Appendix C)[26], which allows us to recover the original kinetic rate constants.

In comparison, the rate of a chemical reaction in the transition state theory (TST) is calculated as[46]:

$$k_{TST} \cong A e^{-\Delta F / k_B T},$$ (14)

where A is a pre-exponential coefficient or frequency factor. Only the energy barriers of ligand binding and dissociation need to be calculated from the original (reweighted, F) and modified (no reweighting, F^*) free energy profiles for estimating the accelerations and recovering the ligand kinetic rate constants.

Host-Guest Binding Simulations

GaMD simulations were performed on the binding of two guest molecules (aspirin and 1-butanol) to the β-CD host using the same input files as in a previous study by Tang and Chang[21]. The CD host were modeled with both the GAFF and q4MD force fields. The guest molecules were modeled with GAFF. The LiGaMD and dual-boost LiGaMD (LiGaMD_Dual) simulations were compared with GaMD simulations using the previous

algorithms[25, 33b], including the total potential boost GaMD (GaMD_Tot), dual-boost GaMD (GaMD_Dual)[25], non-bonded potential boost GaMD (GaMD_NB) and non-bonded dual-boost GaMD (GaMD_Dual_NB)[33b] simulations, as well as the previous long microsecond-timescale cMD simulations[21] (Tables 1 and S1).

With the same restart files as used for running the cMD production simulations[21], the GaMD simulations proceeded with 1 ns short cMD to collect potential statistics, 20 ns GaMD equilibration after adding the boost potential and then three independent 300 ns production runs. GaMD production frames were saved every 0.1 ps for analysis. The VMD[47] and CPPTRAJ[48] tools were used for simulation analysis. The number of host-guest dissociation and binding events (N_D and N_B) and the guest binding and unbinding time periods (τ_B and τ_U) were recorded from individual simulations (Table S1). With high fluctuations, τ_B and τ_U were recorded for only the time periods longer than 0.5 ns. For systems with more than one time of guest dissociation and binding in each of the individual simulations, 1D, 2D and 3D PMF profiles, as well as the host-guest binding free energies, were calculated through energetic reweighting of the GaMD simulations. The center-of-mass distances between the host and guests (d_{HG}) and the host radius of gyration (R_g) were chosen as reaction coordinates for calculating the 1D PMF profiles. 2D PMF profiles of (d_{HG}, R_g) were also calculated to analyze conformational changes of the CD host upon guest binding. The bin size was set to 0.5 Å for the d_{HG} and 0.05 Å for the R_g. The cutoff of the number of simulation frames in one bin for reweighting was set to 500 in 1D and 2D PMF calculations. The 3D PMF profiles of guest displacements from the CD host in the X, Y and Z directions were further calculated from the LiGaMD simulations. The bin sizes were set to 1 Å in the X, Y and Z directions. The cutoff for the number of simulation frames in one bin for reweighting the 3D PMF was set to 10 for LiGaMD simulations and 50 for LiGaMD_Dual simulations. The host-guest binding free energies (ΔG) were calculated using the reweighted 3D PMF profiles and compared with those from previous cMD simulations and experimental data[21] (Table 1). In addition, accelerations of the ligand dissociation and binding rate constants (k_{on} and k_{off}) in LiGaMD simulations were analyzed using the TST and Kramers' rate theory as described above (Table S2).

Protein-Ligand Binding Simulations

LiGaMD simulations using the dual-boost scheme were performed on benzamidine binding to the trypsin protein. The X-ray crystal structure of benzamidine-bound trypsin (PDB ID: 3PTB[49]) was used with the calcium ion and water molecules kept. The missing hydrogen atoms were added using by the *tleap* module in AMBER[50]. The general Amber force field[51] and the AMBER ff14SB force field[52] were used for the ligand and protein, respectively. Atomic partial charges of benzamidine were obtained through B3LYP/6–31G* quantum calculations of the electrostatic potential, for which the RESP charges[53] were fitted using the antechamber program[50]. The system was neutralized by adding a number of counter ions (Cl$^-$) and immersed in a cubic TIP3P water[54] box, which was extended 13 Å from the protein-ligand complex. Testing simulations were performed on protonation of the protein active-site residue His57 at the N_δ atom or the N_ε atom. Results showed that benzamidine could bind to the protein target site as in the X-ray conformation with the N_δ atom protonated, but not with atom N_ε protonated. Therefore, the protein residue His57 was protonated at the N_δ

atom. Moreover, a total of 10 ligand molecules (one in the X-ray bound conformation and another nine placed randomly in the solvent) were included in the system to facilitate ligand binding. This design was based on the fact that the ligand binding rate constant k_{on} is inversely proportional to the ligand concentration. The higher the ligand concentration, the faster the ligand binds, provided that the ligand concentration is still within its solubility limit.

The built simulation system was energy minimized with 1 kcal/mol/Å^2 constraints on the heavy atoms of the protein and ligand, including the steepest descent minimization of 5,000 steps followed by a conjugate gradient minimization of 5,000 steps. The system was then heated from 0 K to 300 K for 200 ps. It was further equilibrated using the NVT ensemble at 300K for 800 ps and the NPT ensemble at 300 K and 1 bar for 1 ns with 1 kcal/mol/Å^2 constraints on the heavy atoms of the protein and ligand, followed by 2 ns short cMD without any constraint. The LiGaMD simulations proceeded with 14 ns short cMD to collect the potential statistics, 54.6 ns GaMD equilibration after adding the boost potential and then five independent 1000 ns production runs. Initial testing simulations showed that when the threshold energy for applying boost potential to the ligand non-bonded energy was set to the lower bound (i.e., $E = V_{max}$), the bound ligand maintained the X-ray conformation and did not dissociate. In comparison, when the threshold energy was set to the upper bound (i.e., $E = V_{min}+1/k$), it enabled high enough boost potential to dissociate the ligand from the protein. Therefore, the threshold energy for applying the ligand boost potential was set to the upper bound in the LiGaMD_Dual simulations. For the second boost potential that was applied to the system total potential energy other than the ligand non-bonded potential energy, sufficient acceleration was obtained to sample ligand rebinding by setting the threshold energy to the lower bound. LiGaMD_Dual production simulation frames were saved every 0.2 ps for analysis.

The VMD[47] and CPPTRAJ[48] 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 (Table S3). With high fluctuations, τ_B and τ_U were recorded for only the time periods longer than 5 ns. The 1D, 2D and 3D PMF profiles, as well as the ligand binding free energy, were calculated through energetic reweighting of the LiGaMD_Dual simulations. The distance between charge centers at the amidine C atom in benzamidine and CG atom of Asp189 in trypsin was chosen as the reaction coordinate for calculating the 1D PMF profiles. The bin size was set to 1.0 Å. 2D PMF profiles of the benzamidine:C – Asp189:CG and Trp215:NE – Asp189:CG atom distances were also calculated to analyze conformational changes of the protein upon ligand binding. The bin size was set to 1.0 Å for the atom distances. 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 guest displacements from the CD host in the X, Y and Z directions were further calculated from the LiGaMD 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 1100–4000 for five individual LiGaMD_Dual simulations) was set to the minimum number below which the calculated 3D PMF minimum will be shifted. The ligand binding free energies (G) were calculated using the reweighted 3D PMF profiles. Furthermore, structural clustering was performed on frames of the diffusing

ligand molecules from each 1 μs LiGaMD_Dual simulation trajectory using the Density Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm[55] in CPPTRAJ[48]. The frames were sieved at a stride of 500 for clustering. The remaining frames were assigned to the closest cluster afterwards. The distance cutoff for DBSCAN clustering was set to 0.5 Å. The resulting structural clusters were reweighted to obtain energetically significant binding pathways of the ligand. In addition, the ligand dissociation and binding rate constants (k_{on} and k_{off}) were calculated from the LiGaMD_Dual simulations with their accelerations analyzed using the Kramers' rate theory as described above (Table S4).

Results

Thermodynamics of Host-Guest Binding

For the β-cyclodextrin (CD) host as modeled with both the GAFF and q4MD force fields, all-atom GaMD simulations were performed to investigate the dissociation and binding of two guest molecules aspirin (Fig. 1A) and 1-butanol (Fig. 1B) using a number of different potential boost algorithms (Table 1). The center-of-mass distances between the host and guest molecules were monitored as a function of time to record the number of dissociation and binding events (N_D and N_B) in each of the 300 ns GaMD simulations. The new LiGaMD method especially using the dual-boost algorithm (LiGaMD_Dual) showed significantly improved sampling of ligand binding compared with the previous GaMD algorithms, including the total potential boost GaMD (GaMD_Tot), dual-boost GaMD (GaMD_Dual), non-bonded potential boost GaMD (GaMD_NB) and non-bonded dual-boost GaMD (GaMD_Dual_NB) (Table 1). Repetitive binding and unbinding of guest molecules were observed in 300 ns LiGaMD and LiGaMD_Dual simulations, except for the binding of 1-butanol to CD that was modeled with the GAFF force field (Fig. S1). In comparison, no guest dissociation and/or binding was observed in one or more of the 300 ns GaMD simulations using the other algorithms (Table 1).

Provided improved sampling in the LiGaMD and LiGaMD_Dual simulations, we computed potential of mean force (PMF) free energy profiles to characterize the host-guest binding quantitatively (Fig. 1). The host-guest distance (d_{HG}) was chosen as a reaction coordinate. The PMF profiles calculated from three 300 ns LiGaMD or LiGaMD_Dual simulations were compared with those from microsecond-timescale cMD simulations of CD using the GAFF force field with aspirin (Fig. 1C), CD using the GAFF force field with 1-butanol (Fig. 1D), CD using the q4MD force field with aspirin (Fig. 1E), and CD using the q4MD force field with 1-butanol (Fig. 1F). For binding of 1-butanol to CD modeled with GAFF, while 300 ns LiGaMD and LiGaMD_Dual simulations sampled the system unbound ($d_{HG} > {\sim}15$ Å) and intermediate ($d_{HG} = {\sim}10$ Å) states, being closely similar to the 6500 ns cMD simulations, the bound state ($d_{HG} = {\sim}0.5$ Å) was poorly sampled in the LiGaMD simulations (Fig. 1D). This correlated with the above finding that the guest dissociation and binding were seldomly observed in these simulations (Table 1 and Figs. S1C–S1D). For the other three systems, 300 ns LiGaMD and LiGaMD_Dual simulations sampled the same bound ($d_{HG} = {\sim}0.5$–2 Å), intermediate ($d_{HG} = {\sim}10$ Å) and unbound ($d_{HG} > {\sim}15$ Å) low-energy states as in the microsecond timescale cMD simulations (Figs. 1C, 1E and 1F). Their global energy minima were all identified in the bound state. However, differences were found in the magnitudes of

PMF profiles near the intermediate, unbound and energy barrier regions. Overall, the LiGaMD_Dual simulations showed better agreements with the long-timescale cMD simulations than LiGaMD, especially for binding of 1-butanol to CD modeled with q4MD (Fig. 1F). The LiGaMD_Dual provided the most efficient sampling of host-guest binding and dissociation events (Table 1 and Fig. S1) and the closest free energy profiles as compared with the reference cMD simulations (Fig. 1).

Next, we analyzed conformational changes of the CD host modeled with GAFF (Fig. 2) and q4MD (Fig. 3) upon guest binding. PMF profiles were calculated for the radius of gyration (R_g) of the CD host modeled with GAFF from the cMD simulations in the ligand-free (*apo*), aspirin and 1-butanol binding forms (Fig. 2A). Two low-energy states, including "Compact" (R_g = ~5.7 Å) and "Open" (R_g = ~5.9 Å), were identified for the CD host. While the *apo* and 1-butanol bound CD predominantly adopted the compact conformational state, binding of aspirin biased conformational ensemble of CD towards the open state. In this regard, aspirin (Fig. 1A) appeared to be larger in size than 1-butanol (Fig. 1B). Furthermore, we calculated 2D profiles of (d_{HG}, R_g) and identified low-energy conformational states of the system. For aspirin binding, three low-energy states were found from the 9500 ns cMD simulation, including the "Bound (B)", "Intermediate (I)" and "Unbound (U)" states, in which the CD host adopted primarily the Open, Compact and Compact conformations, respectively (Figs. 2B and 2C). The 2D PMF profile of aspirin binding calculated from three 300 ns LiGaMD_Dual simulations combined was highly similar to that from 9500 ns cMD simulation, depicting the same three low-energy conformational states (Figs. 2C and 2D). In comparison, only the Bound and Unbound low-energy states were clearly identified for binding of 1-butanol in the 2D PMF profiles of both 6500 ns cMD (Fig. 2E) and three 300 ns LiGaMD_Dual simulations (Fig. 2F), while the intermediate state was hardly observed with apparently a very shallow energy well as shown in the 1D PMF profiles (Fig. 1D).

Furthermore, we evaluated the effects of different force field parameters for the CD host. With the q4MD force field, the CD host adopted predominantly the Open conformation with R_g = ~6.1 Å in PMF profiles of both the *apo* and guest-bound forms, although it still sampled the Compact conformation with a shallow and broad energy well in the *apo* form (Fig. 3A). This was in contrast to the above findings that the CD host using the GAFF force field preferred the Compact conformation in the *apo* form, while guest binding especially aspirin induced the host to open (Fig. 2). Therefore, the usage of different force fields generated distinct structural dynamics and free energy profiles of host-guest binding. Nevertheless, consistent results were obtained from LiGaMD_Dual and cMD simulations provided the same system setup and force field. In this context, LiGaMD only enhanced conformational sampling of the studied systems, yielding similar results as obtained from significantly longer cMD simulations.

In 6000 ns cMD simulation of aspirin binding to CD modeled with q4MD, three low-energy conformational states were identified from 2D PMF profiles of the host-guest distance d_{HG} and host R_g, including the "Bound (B)", "Intermediate (I)" and "Unbound (U)" states, in which the CD host all adopted primarily the Open conformation (Figs. 3B and 3C). A closely similar 2D PMF profile was obtained from three 300 ns LiGaMD_Dual simulations combined (Figs. 3D). Similar 2D profiles were obtained for 1-butanol binding from three

300 ns LiGaMD_Dual simulations combined and 5000 ns cMD simulation, during which CD was open in both the guest bound and unbound states (Figs. 3E and 3F).

In addition to the 1D and 2D PMF profiles, we computed 3D PMF profiles of guest binding to the CD host in the X, Y and Z directions and then the guest binding free energies (see details in Methods). The host-guest binding free energies calculated from LiGaMD and LiGaMD_Dual simulations were compared with those obtained from previous cMD simulations and experimental data (Tables 2 and S1). Compared with experimental data, LiGaMD_Dual provided generally more accurate estimates of the guest binding free energies than LiGaMD (Table 2A), being consistent with the finding that improved sampling with more dissociation and binding events was observed in the LiGaMD_Dual simulations (Table 1). For aspirin, the binding free energy errors were reduced from -1.16 ± 0.03 kcal/mol to 0.85 ± 0.18 kcal/mol with the CD host modeled from GAFF and from -1.00 ± 0.53 kcal/mol to 0.40 ± 0.15 kcal/mol with CD modeled using q4MD. For 1-butanol, the binding free energy error was calculated from only the simulations with CD modeled using q4MD, during each of which multiple dissociation and binding events were observed (Table 1). The binding free energy error of 1-butanol decreased from -1.17 ± 0.61 kcal/mol in LiGaMD simulations to 0.33 ± 0.21 kcal/mol in LiGaMD_Dual simulations (Table 2A). Overall, the guest binding free energy errors from the LiGaMD_Dual simulations were less than 1.0 kcal/mol in comparison with experimental data.

Furthermore, we compared the LiGaMD and LiGaMD_Dual simulations with previous microsecond-timescale (5–9.5 μs) cMD simulations that could be considered as "converged" for sampling with tens to hundreds of guest binding and unbinding events captured (Table 1B). Two different algorithms were applied to compute the guest binding free energies from the cMD simulations in the previous study: G_{comp1} was computed using the estimated guest binding enthalpy and entropy, while G_{comp2} was computed using the guest dissociation and binding rate constants[21]. All the G_{comp2} estimates exhibited significantly smaller standard deviations (error bars) than the G_{comp1} (Table 1B). Compared with experimental data, the guest binding free energies calculated using the q4MD force field for CD were more accurate than using the GAFF force field with the only exception of the G_{comp1} estimate for aspirin (Table 2A). Therefore, the guest binding free energies obtained from cMD simulations using the q4MD force field for CD and G_{comp2} estimates were used for further comparison, for which the errors mostly due to force field inaccuracy were -0.37 ± 0.05 kcal/mol for aspirin and -0.60 ± 0.19 kcal/mol for 1-butanol (Table 2A). Provided the same computational model and force field, the LiGaMD and LiGaMD_Dual simulations were then compared with the converged cMD simulations to examine their sampling errors. Relative to cMD simulations using the q4MD force field for CD and G_{comp2} estimates, LiGaMD_Dual simulations showed < 1.0 kcal/mol guest binding free energy differences with 0.77 ± 0.16 kcal/mol for aspirin and 0.93 ± 0.10 kcal/mol for 1-butanol, and their standard deviations were significantly smaller than those of LiGaMD simulations (Table 2B). Taken together, LiGaMD_Dual achieved improved sampling than LiGaMD and the sampling errors were < 1.0 kcal/mol compared with converged cMD simulations.

In summary, the PMF profiles and binding free energies of guest molecules in the CD host calculated from hundreds-of-nanosecond LiGaMD_Dual simulations agreed excellently with

those from experimental data and previous microsecond-timescale cMD simulations[21]. Compared with the experimental data, errors in the guest binding free energies were smaller than 1.0 kcal/mol from the LiGaMD_Dual simulations. Compared with converged microsecond-timescale cMD simulations, the errors of LiGaMD_Dual simulations resulted from sampling were also smaller than 1.0 kcal/mol (Table 2). Therefore, both efficient enhanced sampling and accurate free energy calculations of host-guest binding were achieved through the LiGaMD_Dual simulations.

Kinetics of Host-Guest Binding

In addition to the thermodynamic free energies, kinetic dissociation and binding rate constants (k_{off} and k_{on}) were further derived from the relevant LiGaMD and LiGaMD_Dual simulations and compared with those from cMD simulations (Table 3). Note that variations were observed in the kinetic rate constants, especially the k_{off} (Table 1), which appeared to result from high statistical fluctuations in recording of the relatively short binding time periods (τ_B) of the guest molecules (Table S1). While the binding rate constants of guest molecules decreased to ~30%–90% of those calculated from cMD simulations[21], the dissociation rate constants increased by ~1.7–15 times in LiGaMD simulations and ~11–18 times in LiGaMD_Dual simulations (Table 3). In this context, long-timescale cMD simulations with repetitive host-guest binding and unbinding were available for comparison. However, in most protein-ligand binding studies (e.g., the trypsin-benzamidine binding described below) such cMD simulations are often not available, due to the extreme challenge of sampling free ligand binding and dissociation over long timescales. In this regard, we sought to reweight kinetics of LiGaMD and recover the original kinetic dissociation and binding rate constants from only the LiGaMD enhanced sampling simulations as follows.

Two algorithms were implemented using the transition state theory (TST) and Kramers' rate theory for reweighting kinetics of the LiGaMD simulations (see details in Methods). In both algorithms, the energy barriers of ligand binding and dissociation were calculated from the original (reweighted) and modified (no reweighting) free energy profiles of the LiGaMD simulations. Since the reweighted PMF profiles were already obtained for the host-guest distances from the LiGaMD and LiGaMD_Dual simulations (Fig. 1), we also calculated the corresponding modified PMF profiles without energetic reweighting for comparative analysis (Fig. 4). Note that the bound state of 1-butanol in CD modeled with GAFF was poorly sampled in the LiGaMD and LiGaMD_Dual simulations (Figs. 1D, S1 and S2). This system was thus excluded for both binding free energy and kinetics calculations. For the other three systems, the energy barriers were significantly reduced in the modified PMF profiles of both the LiGaMD and LiGaMD_Dual simulations for guest dissociation (Fig. 4 and Table S2). In LiGaMD_Dual simulations of aspirin binding to CD modeled with GAFF, the dissociation free energy barrier (F_{off}) decreased by 73% from 2.06 ± 0.28 kcal/mol in the reweighted PMF profile to 0.59 ± 0.37 kcal/mol in the modified PMF profile (Fig. 4B). In the other LiGaMD and LiGaMD_Dual simulations, F_{off} mostly decreased by ~50% (Fig. 4 and Table S2). Correspondingly, k_{off} estimated using the TST increased by ~9–22 times in the LiGaMD simulations of aspirin and 1-butanol binding to CD and ~6–41 times in the LiGaMD_Dual simulations (Table 3). On the other hand, the free energy barrier for ligand binding (F_{on}) actually increased in all modified PMF profiles of the LiGaMD and

LiGaMD_Dual simulations than in the reweighted profiles (Fig. 4 and Table S2). According the TST, the corresponding binding rate constant (k_{on}) decreased by a factor ~0.1–0.4 in LiGaMD simulations and ~0.1–0.2 in LiGaMD_Dual simulations (Table 3). Despite the slower binding, significant accelerations were achieved in the rate-limiting dissociation of guest molecules during LiGaMD and LiGaMD_Dual simulations. Therefore, the guest dissociation and binding appeared to be more balanced in LiGaMD than in cMD, thereby resulting in overall improved sampling.

For applying the Kramers' rate theory, curvatures of the reweighed (w) and modified (w^*, no reweighting) free energy profiles were further calculated near the guest bound ("B") and unbound ("U") energy minima and the energy barrier ("Br"), as well as the ratio of apparent diffusion coefficients calculated from the LiGaMD and LiGaMD_Dual simulations with reweighting (D) and without reweighting (modified, D^*) (Table S2). The resulting numbers were plugged into the Kramers' rate equation to estimate accelerations of the guest binding and dissociation rate constants. Particularly, the k_{off} increased by ~8–12 times in LiGaMD simulations and ~6–23 times in LiGaMD_Dual simulations (Table 3). The k_{on} decreased by a factor of 0.02–0.55 in LiGaMD simulation. In LiGaMD_Dual simulations, k_{on} remained the same for 1-butanol binding to CD modeled with q4MD, but also decreased by factors of 0.06 and 0.23 for aspirin binding to CD modeled with GAFF and q4MD force fields, respectively (Table 3).

In summary, accelerations of host-guest dissociation and binding obtained from LiGaMD and LiGaMD_Dual simulations were derived using the TST and Kramers' Rate Theory. Compared with cMD simulations, while the binding rates decreased to a certain extent, guest unbinding was significantly accelerated by ~1.7–15 times in LiGaMD and 11–18 times in LiGaMD_Dual. Overall, the Kramers' rate theory provided more accurate estimates of the ligand kinetic rate accelerations in the LiGaMD simulations than the TST (Table 3). It was thus applied for reweighting of ligand kinetics in analysis of further LiGaMD simulations.

Thermodynamics of Ligand Binding to the Trypsin Model Protein

In addition to host-guest binding, LiGaMD was further tested on protein-ligand binding using trypsin as a model system. Initial testing simulations suggested the following setup for proper simulations of benzamidine binding to trypsin: First, residue His57 at the protein active site was protonated at the N_δ atom rather than the default N_e atom, which was also depicted for the catalytic triad at the active site of proteases[56]. Testing simulations with the N_e atom protonated in residue His57 showed that the benzamidine inhibitor could not bind to the protein target site as in the X-ray crystal conformation with ~4.3 Å minimum root-mean square deviation (RMSD). In comparison, simulations with N_δ atom could capture repetitive binding of benzamidine to the target site with ~1.0 Å minimum RMSD compared with the X-ray conformation (Fig. S3). Second, the threshold energy for applying boost potential to the ligand non-bonded interaction energy was set to the upper bound (i.e., $E = V_{min}+1/k$). This enabled high enough boost potential to dissociate the ligand from the protein active site. In comparison, the bound ligand maintained the X-ray conformation during ~200 ns testing simulations with the threshold energy set to the lower bound (i.e., $E = V_{max}$). Moreover, higher acceleration was observed for the bound ligand as the input

parameter σ_{0P} was increased from 1.0 kcal/mol to 6.0 kcal/mol and the ligand started to dissociate from the target site during the LiGaMD equilibration simulation with $\sigma_{0P} = 4.0$ kcal/mol (Fig. S3), which was thus used for production simulations. Third, a total of 10 ligand molecules (one in the X-ray bound conformation and another nine placed randomly in the solvent) were included in the system to facilitate ligand rebinding. This design was based on the fact that the ligand binding rate constant k_{on} is inversely proportional to the ligand concentration. The higher the ligand concentration, the faster the ligand binds, provided that the ligand concentration is still within its solubility limit. By applying the second boost potential to the unbound ligand molecules, protein and solvent, ligand rebinding could be sampled even during the LiGaMD equilibration simulations (Figs. S3D and S3F).

With the above settings, LiGaMD_Dual simulations were able to capture repetitive dissociation and binding of the benzamidine inhibitor in trypsin within 1 μs simulation time (Figs. 5A, S4 and S5 and Movies S1–S5). In five independent LiGaMD_Dual simulations as summarized in Table 4, the average of the GaMD boost potential V was ~21–22 kcal/mol with ~4.2–4.3 kcal/mol standard deviation. The ligand dissociated for 3–11 times and rebound for 3–10 times during the five 1 μs LiGaMD_Dual simulations. In the X-ray bound conformation, the distance between the amidine C atom in benzamidine and CG atom of Asp189 in trypsin was 3.9 Å. During the LiGaMD_Dual simulations, the bound ligand dissociated from the protein and diffused into the bulk solvent with the benzamidine:C – ASP189:CG distance increased up to ~60 Å (Figs. 5A and S5), similarly for the ligand RMSD relative to the X-ray conformation (Fig. S4). Then after sufficient sampling of the bulk solvent space, one of the ten ligand molecules rebound to the protein with a salt bridge formed between the charged benzamidine and side chain of protein residue Asp189, for which the benzamidine:C – Asp189:CG distance and ligand RMSD dropped back to ~3.9 Å (Fig. 5A) and ~1.0 Å (Fig. S4), respectively. When the benzamidine:C – Asp189:CG distance dropped below a threshold value (3.7 Å defined here), atomic coordinates, velocities and forces of the bound ligand were swapped with those of the original bound ligand in the simulation starting structure (denoted "Lig0"). Upon dissociation of Lig0 from the protein, the high concentration of ten ligand molecules in the solvent facilitated ligand rebinding to the protein. Through such cycles, repetitive dissociation and binding of the benzamidine ligand in trypsin were efficiently sampled in the LiGaMD_Dual simulations.

Next, we calculated PMF free energy profiles to characterize ligand binding to trypsin. Since plots of the benzamidine:C – Asp189:CG distance and ligand RMSD depicted the ligand binding processes similarly (Figs. S4 and S5) and extra efforts were needed to calculate symmetry-corrected RMSD of benzamidine, the PMF profiles were calculated for the benzamidine:C – Asp189:CGdistance. The resulting reweighted and modified 1D PMF profiles were shown in Fig. 5B. In the reweighted PMF profile, three low-energy wells were identified for the ligand in the Bound, Intermediate and Unbound states, for which the benzamidine:C – Asp189:CG distance was centered around ~4.5 Å, ~15 Å and ~35 Å, respectively. A high energy barrier of 12.17 ± 1.54 kcal/mol was observed for the ligand dissociation (Table S4). In the modified PMF profile without reweighting, the energy barrier decreased to 1.37 ± 0.56 kcal/mol for ligand dissociation and the energy wells became significantly shallower for the ligand in the Bound, Intermediate and Unbound states. This

justified enhanced sampling of protein-ligand binding and unbinding in the LiGaMD_Dual simulations.

Furthermore, we computed 2D PMF profiles to analyze conformational changes of the protein upon ligand binding. Upon close examination of the system trajectories, residue Trp215 and its associated loop in trypsin underwent the largest conformational changes during ligand binding (Movies S1–S5). In this regard, a distance between the N_e atom in the Trp215 side chain and CG atom of Asp189 was selected as another reaction coordinate. The calculated reweighted and modified 2D PMF profiles of the benzamidine:C – Asp189:CG and Trp215:NE – Asp189:CG atom distances are plotted in Fig. 5C and 5D, respectively. Five 1000 ns LiGaMD_Dual simulations were combined for calculating the PMF profiles. The reweighted and modified 2D PMF profiles calculated from five individual LiGaMD_Dual simulations separately are also shown in Figs. S6 and S7, respectively.

In the reweighted 2D PMF, low-energy wells were identified for the trypsin-benzamidine system in the Bound, Intermediate I1, Intermediate I2, Unbound U1 and Unbound U2 states, for which the (benzamidine:C – Asp189:CG, Trp215:NE – Asp189:CG) atom distances were centered around (4.0 Å, 12.5 Å), (17.0 Å, 12.5 Å), (7.5 Å, 7.5 Å), (35.0 Å, 12.5 Å) and (35.0 Å, 8.0 Å), respectively (Fig. 5C). In the modified 2D PMF profile (Fig. 5D), low-energy wells were also identified for the system in the Bound, Intermediate I1, Unbound U1 and Unbound U2 states in similar locations as in the reweighted 2D PMF. A distinct low-energy well was identified for the system in an Intermediate I3 state in the modified PMF with (14.0 Å, 9.0 Å) for the (benzamidine:C – Asp189:CG, Trp215:NE – Asp189:CG) atom distances, rather than the I2 state in the reweighted 2D PMF. These low-energy wells were also observed in the reweighed and modified PMF profiles of five individual LiGaMD_Dual simulations (Figs. S6 and S7).

Next, we extracted six representative low-energy conformations of the trypsin-benzamidine system as identified from the PMF profiles, including the "Bound" (Fig. 6A), "Intermediate I1" (Fig. 6B), "Intermediate I2" (Fig. 6C), "Intermediate I3" (Fig. 6D), "Unbound U1" (Fig. 6E), and "Unbound U2" (Fig. 6F). In the Bound conformational state, the benzamidine ligand bound to protein residue Asp189 in the S1 pocket as in the X-ray crystal structure. Protein residue Trp215 closed the S1* pocket, but left the S1 pocket open, which was located between the Asp189 loop and the Trp215 loop (Fig. 6A). In the Intermediate I1 state, the benzamidine ligand bound to a site out of the Trp215 "gate", being close to the His57-Asp102-Ser214 catalytic triad (Fig. 5B). In the Intermediate I3 state, residue Trp215 flipped its side chain to close the S1 pocket, while the S1* pocket became open with rearrangement of the Trp215 loop compared with the X-ray structure. The benzamidine ligand bound to the open S1* pocket (Fig. 6C). In the Intermediate I3 state, the side chain of protein residue Trp215 rotated to a conformation perpendicular to the X-ray conformation such that both S1 and S1* pockets became open and the benzamidine ligand bound to a site in close proximity of Trp215 and catalytic residues Asp102 and Ser214 (Fig. 6D). In the Unbound U1 state, trypsin adopted a conformation similar to the X-ray structure with the S1 pocket open and the S1* pocket closed by Trp215, while the benzamidine ligand was found far away from the protein (Fig. 6E). In the Unbound U2 state, trypsin flipped its side chain to close the S1

pocket, similar to the Intermediate I2 conformation, while the benzamidine ligand stayed far away from the protein (Fig. 6F).

In addition to the 1D and 2D PMF profiles, 3D PMF was calculated from each individual 1 μs LiGaMD_Dual simulation of benzamidine binding to trypsin in terms of displacements of the amidine C atom of benzamidine from the CG atom in protein residue Asp189 in the X, Y and Z directions. We then calculated the ligand binding free energy from each 3D reweighted PMF (see Methods). The average of resulting free energy values was −6.13 kcal/mol and the standard deviation was 0.35 kcal/mol. This was in excellent agreement with the experimental value of −6.2 kcal/mol for the ligand binding free energy of benzamidine in trypsin[57] (Table 5).

Pathways and kinetics of Ligand Binding to Trypsin

With accurate prediction of the ligand binding free energy, we analyzed the LiGaMD_Dual simulations further to determine the pathways and kinetic rate constants of benzamidine binding to trypsin. For each 1 μs simulation trajectory, structural clustering was performed on snapshots of the diffusing ligand molecules using the DBSCAN algorithm and the resulting structural clusters were reweighted to obtain energetically significant pathways of the ligand (see Methods). Fig. 7 depicts the dissociation and binding pathways of the benzamidine ligand in trypsin obtained from the five individual LiGaMD_Dual simulations. The ligand clusters along each pathway were colored according to the reweighted PMF values in a blue (0 kcal/mol)-white (7.5 kcal/mol)-red (15.0 kcal/mol) scale. The lowest-energy ligand cluster was consistently identified at the target site with benzamidine forming ionic interaction with trypsin residue Asp189 as determined in the X-ray crystal structure (Fig. 7).

During the LiGaMD_Dual simulations, benzamidine first dissociated predominantly through the protein opening between the Asp189 loop and the Trp215 loop, denoted pathway "P1" (Fig. 7). Then benzamidine molecules could rebind spontaneously to trypsin. Negatively charged residues on the protein surface, including Glu70, Asp71, Glu77, Glu80 and Asp153, appeared to steer the positively charged ligand towards the enzyme catalytic site formed by residues His57-Asp102-Ser214. This was consistent with previous findings of "electrostatic steering" in ligand recognition by proteins[58]. Subsequently, the ligand bound to the target site via two pathways, one being the pathway "P1" and the other connecting the S1* pocket and the Trp215 side chain gate as closed in the X-ray structure (denoted pathway "P1*") (Fig. 7). In Sim1 and Sim2, because the Trp215 side chain maintained the X-ray conformation and closed the S1* pocket, the benzamidine ligand dissociated and bound to the S1 pocket repetitively through the P1 pathway (Figs. 7A and 7B and Movies S1 and S2). In the other three simulations Sim3-Sim5, the Trp215 side chain underwent conformational changes to open either the S1 or S1* pocket or both pockets. Benzamidine dissociated and bound to the protein via either of the P1 and P1* pathways (Figs. 7C–7E and Movies S3–S5). Notably, binding of the ligand was able to switch from pathway P1* to pathway P1 via the S1* binding pocket to the final target site in the S1 pocket as observed near 286 ns in the Sim5 trajectory (Movie S5). In addition, two ligand molecules were found to bind both the

S1 and S1* pockets simultaneously during 310.5 ns-333.5 ns in the Sim5 trajectory (Movie S5 and Fig. S5E).

To calculate kinetic rate constants of benzamidine binding to trypsin, we recorded the time periods for the ligand found in the bound (τ_B) and unbound (τ_U) states throughout the LiGaMD_Dual simulations (Table S3). Since a total of 10 ligand molecules and 10,478 water molecules were included in the simulation system, the ligand concentration was 0.053 M in the simulations. Without reweighting of the LiGaMD_Dual simulations, the benzamidine ligand binding (k_{on}*) and dissociation (k_{off}*) rate constants were calculated as $1.58 \pm 1.09 \times 10^8$ M·s^{-1} and $3.40 \pm 1.36 \times 10^7$ s^{-1}, respectively.

Following the same protocol as described for analyzing the host-guest binding simulations, we reweighted the trypsin-benzamidine binding simulations to calculate acceleration factors of the ligand binding and dissociation (Table S4) and recover the original kinetic rate constants using the Kramers' rate theory. The original (reweighted) and modified PMF profiles of protein-ligand distance are shown in Fig. 5B. The dissociation free energy barrier (F_{off}) decreased by ~90% from 12.17 ± 1.54 kcal/mol in the reweighted PMF profile to 1.37 ± 0.56 kcal/mol in the modified PMF profile (Table S4). On the other hand, the free energy barrier for ligand binding (F_{on}) decreased from 3.04 ± 2.04 kcal/mol in the reweighted profile to 2.40 ± 0.41 kcal/mol in the modified PMF profile (Fig. 5B and Table S4). Furthermore, curvatures of the reweighed (w) and modified (w*, no reweighting) free energy profiles were calculated near the guest bound ("B") and unbound ("U") low-energy wells and the energy barrier ("Br"), as well as the ratio of apparent diffusion coefficients calculated from the LiGaMD_Dual simulations with reweighting (D) and without reweighting (modified, D*) (Table S4). According to the Kramers' rate theory, the ligand binding and acceleration were accelerated by 13.76 and 9.62×10^7 times, respectively. Therefore, the reweighted k_{on} and k_{off} were calculated as $1.15 \pm 0.79 \times 10^7$ M^{-1}·s^{-1} and 3.53 ± 1.41 s^{-1}, respectively. They were comparable to the experimental data[57] of $k_{on}^{exp} = 2.9 \times 10^7$ M$^{-1} \cdot$ s^{-1} and $k_{off}^{exp} = 600 \pm 300$ s^{-1} (Table 5).

Discussion

A new LiGaMD method has been developed to selectively boost the ligand non-bonded interaction potential energy, which enables enhanced sampling simulations of repetitive ligand dissociation and binding as demonstrated on the host-guest and protein-ligand binding model systems. LiGaMD provides a promising approach for simultaneously calculating the free energy and kinetic rate constants of ligand binding.

For host-guest binding, LiGaMD and dual-boost LiGaMD (LiGaMD_Dual) significantly improved the sampling efficiency compared with the previous GaMD_Tot, GaMD_Dual, GaMD_NB and GaMD_NB_Dual algorithms (Table 1). For simulations that captured multiple events of ligand binding and dissociation, the accuracy of calculated ligand binding free energies and kinetic rate constants was similar to those obtained from much longer cMD simulations as compared with the experimental data. Notably, the CD host exhibited distinct structural dynamics when modeled with the different GAFF and q4MD force fields. Nevertheless, similar results were obtained from LiGaMD and cMD simulations provided

the same force field. Therefore, LiGaMD only enhanced conformational sampling of the studied systems, but did not change statistics related to the force field. Apart from enhanced sampling, it remains critical to develop accurate force fields for systems of our interest such as the proteins and ligand molecules.

Detailed analysis showed that the guest binding free energies calculated using the q4MD force field for CD appeared to be more accurate than using the GAFF force field. LiGaMD_Dual provided more powerful sampling than LiGaMD with significantly smaller standard deviations of free energy estimates. Compared with the experimental data, errors in the guest binding free energies calculated from the LiGaMD_Dual simulations were smaller than 1.0 kcal/mol. Compared with converged microsecond-timescale cMD simulations, the sampling errors of LiGaMD_Dual simulations were also smaller than 1.0 kcal/mol. With the availability of both experimental data and converged cMD simulations, simulations on the host-guest binding model systems thus enabled us to examine sampling accuracy of the new LiGaMD method. Moreover, accelerations of the guest dissociation and binding rate constants in LiGaMD simulations were found to be properly estimated using the Kramers' rate theory. Compared with cMD simulations, while the binding rates decreased to a certain extent, guest unbinding was significantly accelerated by ~1.7–15 times in LiGaMD and 11–18 times in LiGaMD_Dual.

For the trypsin model protein-ligand binding system, the threshold energy for applying LiGaMD boost potential to non-bonded potential energy of the bound ligand was set to the upper bound so that high enough acceleration was obtained to allow for ligand dissociation. Multiple ligand molecules (e.g., 10 in the present simulations) were included in the system to facilitate binding. This was based on the fact that higher ligand concentration would lead to faster binding rate constant k_{on}. Multiple events of ligand dissociation and binding were observed in each of the five independent 1 μs LiGaMD_Dual simulations of the trypsin-benzamidine system. Trypsin residue Trp215 appeared to be a gate for ligand binding through the P1 and P1* pathways.

The low-energy conformational states (Fig. 6) and ligand pathways (Fig. 7) of the trypsin-benzamidine system identified from LiGaMD_Dual simulations were mostly consistent with those obtained from previous simulation studies, especially the MSM[22]. However, LiGaMD_Dual simulations further revealed two novel findings. First, two benzamidine molecules were able to bind both the S1 and S1* pockets simultaneously during one of the five LiGaMD_Dual simulation trajectories (Movie S5). While this finding could result from a relatively high ligand concentration (0.053 M) with 10 benzamidine molecules in the simulation system, such rare event is worthy further investigation in the future. Second, binding of the benzamidine ligand was able to switch from the P1* pathway to the P1 pathway in the LiGaMD_Dual simulations. The ligand bound to the S1 pocket via an intermediate site in the S1* pocket. LiGaMD_Dual appeared to provide improved sampling compared with MD simulations used in the MSM so that such binding process could be captured. This LiGaMD_Dual simulation finding also suggested that the final target binding site of benzamidine is located in the S1 pocket as determined in the X-ray crystal structure. Ligand binding in the X-ray conformational state corresponded the global free energy minimum in the calculated PMF profiles.

Furthermore, we compared the trypsin ligand binding free energy and kinetic rate constants with experimental data[57] and computational predictions from previous enhanced sampling[8, 16a, 24] and MSM[14, 22] studies since converged cMD simulations were not available yet on this system. The ligand binding free energy calculated from five 1 μs LiGaMD_Dual simulations $G^0 = -6.13 \pm 0.35$ kcal/mol was in excellent agreement with the experimental value $\Delta G^0_{exp} = -6.2$ kcal/mol. The ligand kinetic rate constants calculated by reweighting of the LiGaMD_Dual simulations were $k_{on} = 1.15 \pm 0.79 \times 10^7\,M^{-1} \cdot s^{-1}$ and $k_{off} = 3.53 \pm 1.41\,s^{-1}$. The binding rate constant k_{on} agreed excellently with the experimental value $k_{on}^{exp} = 2.9 \times 10^7\,M^{-1} \cdot s^{-1}$, while the dissociation rate constant k_{off} appeared to be slower than the experimental value $k_{off}^{exp} = 600 \pm 300\,s^{-1}$[57] (Table 5). In comparison, the ligand dissociation rate constant estimated from five 1 μs LiGaMD_Dual simulations $k_{off} = 3.53 \pm 1.41\,s^{-1}$ was similar to the value of $k_{off} = 9.1 \pm 2.5\,s^{-1}$ obtained from previous metadynamics simulations also performed at a total of 5 μs length[16a]. However, the benzamidine binding free energy calculated from LiGaMD_Dual simulations $G^0 = -6.13 \pm 0.35$ kcal/mol was more accurate than the values of -8.5 ± 0.7 kcal/mol obtained from separate funnel metadynamics simulations[8] and -5.06 kcal/mol from the SITSMD simulations[24], as compared with experimental data $\Delta G^0_{exp} = -6.2$ kcal/mol. In this regard, the more recent MSM predicted accurate ligand binding free energy $G^0 = -6.05 \pm 1$ kcal/mol and binding rate constant $k_{on} = 6.4 \pm 1.6 \times 10^7\,M^{-1} \cdot s^{-1}$, while the model predicted $k_{off} = 131 \pm 109 \times 10^2\,s^{-1}$ appeared to be faster than the experimental value[22]. Both MSM and LiGaMD_Dual could be applied to calculate the ligand binding free energy and kinetic rate constants simultaneously, but more expensive and significantly longer simulations were needed for the MSM (~150 μs in total) than for the LiGaMD_Dual simulations (1 μs × 5).

In summary, LiGaMD provides a powerful enhanced sampling approach to calculating both free energies and kinetic rate constants of ligand binding simultaneously. By physically capturing repetitive ligand binding and unbinding through unconstrained enhanced sampling, LiGaMD allows us to calculate ligand binding free energies from 3D PMF profiles, being distinct from previous methods such as the TI[3], FEP[4] and funnel metadynamics[8]. In particular, funnel metadynamics also provides an efficient approach to calculating ligand binding free energy based on physical PMF profiles[8], but it requires carefully selected, pre-defined collective variables and relies on restraint potential to confine the ligand in a funnel shaped space (increase local ligand concentration) to enhance sampling of ligand binding. In comparison, LiGaMD does not require pre-defined collective variables. Multiple ligand molecules are added to directly increase ligand concentration in the solvent and enhance ligand binding in LiGaMD simulations. LiGaMD thus allows for free sampling of the ligand binding and unbinding processes without any constraints. Notably, low-energy wells could be clearly identified for the ligand in the unbound state from LiGaMD simulations of trypsin (Fig. 5), but not in the funnel metadynamics simulations[8]. Therefore, LiGaMD appears to simulate diffusion of the unbound ligand in solvent more properly. Beyond thermodynamics, LiGaMD also presents, to the best of our knowledge, the first approach to estimate the ligand kinetic dissociation and binding rate constants using the Kramers' rate theory. In this context, metadynamics simulations have

been applied to calculate the ligand dissociation rate constant[16], but not for the binding rate constant. Funnel metadynamics has not been applied to study ligand binding kinetics[8]. Taken together, a more comprehensive picture of ligand binding involving thermodynamics, kinetics and the newly revealed ligand pathways and mechanisms have been obtained with the new LiGaMD method.

In this study, LiGaMD has been demonstrated on host-guest and trypsin protein-ligand binding model systems. Nevertheless, the method awaits further testing on many other ligand binding systems, notably membrane proteins. Although a total number of 10 ligand molecules has been included in the trypsin-benzamidine simulations to facilitate ligand binding, it will be more systematic to optimize the number of ligand molecules for future simulations according to the ligand solubility. Furthermore, the distance has been used in this study to determine when a ligand molecule binds to the protein target site and higher boost potential will be applied accordingly to enable ligand dissociation. Mean square displacement (MSD) has also been implemented to track ligand diffusion for LiGaMD implemented in AMBER 20[59] (SI Methods). More metrics that can serve in this purpose for a wide range of ligand binding systems with different chemical and physical properties will be implemented in the future. These developments are expected to further improve the LiGaMD method for applications in characterizing both ligand binding thermodynamics and kinetics, which should facilitate computer-aided drug design.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Appendix A: Gaussian accelerated molecular dynamics (GaMD)

Consider a system with N atoms at positions $r \equiv \{\overrightarrow{r}_1, \cdots, \overrightarrow{r}_N\}$. When the system potential $V(r)$ is lower than a reference energy E, the modified potential $V^*(r)$ of the system is calculated as:

$$V^*(r) = V(r) + \Delta V(r),$$

$$\Delta V(r) = \begin{cases} \frac{1}{2}k(E - V(r))^2, V(r) < E \\ 0, V(r) \geq E \end{cases} \tag{A1}$$

where k is the harmonic force constant. The two adjustable parameters E and k are automatically determined based on three enhanced sampling principles[25]. The reference energy needs to be set in the following range:

$$V_{max} \leq E \leq V_{min} + \frac{1}{k}, \tag{A2}$$

where V_{max} and V_{min} are the system minimum and maximum potential energies. To ensure that Eqn. (A2) is valid, k has to satisfy: $k \leq \frac{1}{V_{max} - V_{min}}$ Let us define $k \equiv k_0 \cdot \frac{1}{V_{max} - V_{min}}$, then $0 < k_0 \leq 1$. The standard deviation of ΔV needs to be small enough (i.e., narrow distribution) to ensure proper energetic reweighting[60]: $\sigma_{\Delta V} = k(E - V_{avg}) \sigma_V \leq \sigma_0$ where V_{avg} and σ_V are the average and standard deviation of the system potential energies, $\sigma_{\Delta V}$ is the standard deviation of ΔV with σ_0 as a user-specified upper limit (e.g., $10k_BT$) for proper reweighting. When E is set to the lower bound $E = V_{max}$, k_0 can be calculated as:

$$k_0 = \min(1.0, k_0') = \min\left(1.0, \frac{\sigma_0}{\sigma_V} \cdot \frac{V_{max} - V_{min}}{V_{max} - V_{avg}}\right). \tag{A3}$$

Alternatively, when the threshold energy E is set to its upper bound $E = V_{min} + \frac{1}{k}$, k_0 is set to:

$$k_0 = k_0'' \equiv \left(1 - \frac{\sigma_0}{\sigma_V}\right)\frac{V_{max} - V_{min}}{V_{avg} - V_{min}}, \tag{A4}$$

if k_0'' is found to be between 0 and 1. Otherwise, k_0 is calculated using Eqn. (A3).

GaMD provides options to boost only the total potential boost (GaMD_Tot), only the dihedral potential energy (GaMD_Dih), both the total and dihedral potential energies (GaMD_Dual), the non-bonded potential energy (GaMD_NB), both the non-bonded potential and dihedral energies (GaMD_Dual_NB), only non-bonded potential energy of the bound ligand (LiGaMD), and both the ligand non-bonded potential energy of the bound ligand and the remaining potential energy of the entire system (LiGaMD_Dual). The dual-boost simulation generally provides higher acceleration than the other single-boost simulations for enhanced sampling. The simulation parameters comprise of settings for calculating the threshold energy values and the effective harmonic force constants of the boost potentials.

Appendix B: Energetic reweighting of GaMD simulations

For energetic reweighting of GaMD simulations to calculate potential of mean force (PMF), the probability distribution along a reaction coordinate is written as $p^*(A)$. Given the boost

potential $V(r)$ of each frame, $p^*(A)$ can be reweighted to recover the canonical ensemble distribution, $p(A)$, as:

$$p(A_j) = p^*(A_j)\frac{\left\langle e^{\beta\Delta V(r)}\right\rangle_j}{\sum_{i=1}^M \left\langle p^*(A_i)e^{\beta\Delta V(r)}\right\rangle_i}, j = 1, \ldots, M, \tag{A5}$$

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

$$\left\langle e^{\beta\Delta V(r)}\right\rangle = exp\left\{\sum_{k=1}^{\infty}\frac{\beta^k}{k!}C_k\right\}, \tag{A6}$$

where the first two cumulants are given by:

$$C_1 = \langle\Delta V\rangle$$
$$C_2 = \left\langle\Delta V^2\right\rangle - \langle\Delta V\rangle^2 = \sigma_v^2. \tag{A7}$$

The boost potential obtained from GaMD simulations usually follows near-Gaussian distribution[33a]. Cumulant expansion to the second order thus provides a good approximation for computing the reweighting factor[25, 60]. The reweighted free energy $F(A) = -k_B T\ln p(A)$ is calculated as:

$$F(A) = F^*(A) - \sum_{k=1}^2 \frac{\beta^k}{k!}C_k + F_c, \tag{A8}$$

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

Appendix C: Reweighting of Biomolecular Kinetics with Kramers' Rate Theory

A brief summary is provided here for reweighting of biomolecular kinetics from GaMD simulations with Kramers rate theory as described recently[26]. For a particle climbing over potential energy barriers, Kramers showed that the reaction rate depends on temperature and viscosity of the host medium[61]. The reaction rates were derived for both limiting cases of small and large viscosity. In the context of biomolecular simulations in aqueous medium, it is relevant for us to focus on the large viscosity limiting case. Biomolecules move in the high friction ("overdamping") regime and energy barriers are much greater than $k_B T$ (k_B is the Boltzmann's constant and T is temperature). In this case, the reaction rate is calculated as:

$$k_R \cong \frac{2\pi w_m w_b}{\xi}e^{-\Delta F/k_B T}, \tag{A9}$$

where w_m and w_b are frequencies of the approximated harmonic oscillators (also referred to as curvatures of free energy surface[44]) near the energy minimum and barrier, respectively, ξ is the friction constant and $\quad F$ is the free energy barrier of transition.

Without the loss of generality, we consider a 1D potential of mean force (PMF) free energy profile of a reaction coordinate $F(A)$. Near minimum at A_m, the free energy can be approximated by a harmonic oscillator[61] of frequency w_m, i.e., $F(A) = \frac{1}{2}(2\pi w_m)^2(A - A_m)^2$.

Near barrier at A_b, the free energy is approximated as $F(A) = F_b - \frac{1}{2}(2\pi w_b)^2(A - A_b)^2$, where F_b is the free energy at A_b and w_b is the frequency of the approximated harmonic oscillator. Then we can calculate w_m and w_b as:

$$w = \sqrt{\frac{|F''(A)|}{2\pi}}, \tag{A10}$$

where $F'(A)$ is the second-order derivative of the PMF profile.

The friction constant ξ or diffusion coefficient D with $\xi = k_B T/D$ can be estimated as follows. First, we calculate a survival function $S(t)$ as the probability that the system remains in an energy well longer than time t. In a direct approach[45], we count the events that the system visits the energy well throughout a simulation. We record and measure the time intervals of each visiting event until the system escapes over an energy barrier. Then we have a time series T_i, where $i=1, 2, \ldots, N$, and N is the total number barrier transitions observed in the simulation. The time series is subsequently ordered such that $\hat{T}_1 \leq \hat{T}_2 \leq \cdots \leq \hat{T}_N$. With that, the survival function is estimated as $S(\hat{T}_i) \approx 1 - i/N$, which is the probability that the system is trapped in the energy well for time longer than $\hat{T}_i$. Alternatively, we can numerically calculate the time-dependent probability density of reaction coordinate A, $\rho(A,t)$ by solving the Smoluchowski equation along 1D PMF profile of the reaction coordinate:

$$\frac{\partial \rho(A,t)}{\partial t} = D\frac{\partial}{\partial A}\left[e^{-F(A)/k_B T}\frac{\partial}{\partial A}\left(e^{F(A)/k_B T}\rho\right)\right]. \tag{A11}$$

Then the survival function is calculated as $S(t) = \int_t^\infty \int_{A_{b1}}^{A_{b2}} \rho(A,t)dAdt$, where A_{b1} and A_{b2} are two boundaries of the energy well. The initial condition is often set as the Boltzmann distribution of reaction coordinate A in the energy well, i.e., $\rho(A,0) = e^{-F(A)/k_B T}$.

Second, using the above survival functions, we estimate the effective kinetic rates as the negative of the slopes in linear fitting of the $\ln[S(t)]$ versus t, i.e., $k = -d\ln[S(t)]/dt$. This is based on the assumption that the survival function exhibits exponential decay as observed in earlier studies[45, 62]. Finally, the apparent diffusion coefficient D is obtained by dividing the kinetic rate calculated directly using the transition time series collected from the simulation by that using the probability density solution of the Smoluchowski equation[45].

The curvatures and energy barriers of the reweighted and modified free energy profiles, as well as the apparent diffusion coefficients, are calculated and used in Kramers' rate equation to determine accelerations of biomolecular kinetics in the GaMD simulations[26].

References

1. Gilson MK; Zhou HX, Calculation of protein-ligand binding affinities. Annu Rev Bioph Biom 2007, 36, 21–42.

2. Jorgensen WL, The many roles of computation in drug discovery. Science 2004, 303 (5665), 1813–8. [PubMed: 15031495]

3. Kirkwood JG, Statistical Mechanics of Fluid Mixtures. The Journal of Chemical Physics 1935, 3 (5), 300–313.

4. Zwanzig RW, High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. The Journal of Chemical Physics 1954, 22 (8), 1420–1426.

5. Gilson MK; Given JA; Bush BL; McCammon JA, The statistical-thermodynamic basis for computation of binding affinities: A critical review. Biophysical Journal 1997, 72 (3), 1047–1069. [PubMed: 9138555]

6. Torrie GM; Valleau JP, Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. Journal of Computational Physics 1977, 23 (2), 187–199.

7. Villarreal OD; Yu L; Rodriguez RA; Chen LY, Computing the binding affinity of a ligand buried deep inside a protein with the hybrid steered molecular dynamics. Biochem Biophys Res Commun 2017, 483 (1), 203–208. [PubMed: 28034750]

8. Limongelli V; Bonomi M; Parrinello M, Funnel metadynamics as accurate binding free-energy method. Proc Natl Acad Sci 2013, 110 (16), 6358–63. [PubMed: 23553839]

9. Srinivasan J; Cheatham TE; Cieplak P; Kollman PA; Case DA, Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices. J Am Chem Soc 1998, 120 (37), 9401–9409.

10. (a)Yin J; Henriksen NM; Slochower DR; Shirts MR; Chiu MW; Mobley DL; Gilson MK, Overview of the SAMPL5 host-guest challenge: Are we doing better? J Comput Aided Mol Des 2017, 31 (1), 1–19; [PubMed: 27658802] (b)Muddana HS; Fenley AT; Mobley DL; Gilson MK, The SAMPL4 host-guest blind prediction challenge: an overview. J Comput Aided Mol Des 2014, 28 (4), 305–17; [PubMed: 24599514] (c)Muddana HS; Varnado CD; Bielawski CW; Urbach AR; Isaacs L; Geballe MT; Gilson MK, Blind prediction of host-guest binding affinities: a new SAMPL3 challenge. J Comput Aided Mol Des 2012, 26 (5), 475–87. [PubMed: 22366955]

11. (a)Gaieb Z; Parks CD; Chiu M; Yang H; Shao C; Walters WP; Lambert MH; Nevins N; Bembenek SD; Ameriks MK; Mirzadegan T; Burley SK; Amaro RE; Gilson MK, D3R Grand Challenge 3: blind prediction of protein-ligand poses and affinity rankings. J Comput Aided Mol Des 2019; (b)Gathiaka S; Liu S; Chiu M; Yang H; Stuckey JA; Kang YN; Delproposto J; Kubish G; Dunbar JB Jr.; Carlson HA; Burley SK; Walters WP; Amaro RE; Feher VA; Gilson MK, D3R grand challenge 2015: Evaluation of protein-ligand pose and affinity predictions. J Comput Aided Mol Des 2016, 30 (9), 651–668; [PubMed: 27696240] (c)Gaieb Z; Liu S; Gathiaka S; Chiu M; Yang H; Shao C; Feher VA; Walters WP; Kuhn B; Rudolph MG; Burley SK; Gilson MK; Amaro RE, D3R Grand Challenge 2: blind prediction of protein-ligand poses, affinity rankings, and relative binding free energies. J Comput Aided Mol Des 2018, 32 (1), 1–20. [PubMed: 29204945]

12. (a)Schuetz DA; de Witte WEA; Wong YC; Knasmueller B; Richter L; Kokh DB; Sadiq SK; Bosma R; Nederpelt I; Heitman LH; Segala E; Amaral M; Guo D; Andres D; Georgi V; Stoddart LA; Hill S; Cooke RM; De Graaf C; Leurs R; Frech M; Wade RC; de Lange ECM; IJzerman AP; Muller-Fahrnow A; Ecker GF, Kinetics for Drug Discovery: an industry-driven effort to target drug residence time. Drug Discov Today 2017, 22 (6), 896–911; [PubMed: 28412474] (b)Tonge PJ, Drug-Target Kinetics in Drug Discovery. ACS chemical neuroscience 2018, 9 (1), 29–39. [PubMed: 28640596]

13. (a)Shan Y; Kim ET; Eastwood MP; Dror RO; Seeliger MA; Shaw DE, How does a drug molecule find its target binding site? J Am Chem Soc 2011, 133 (24), 9181–3; [PubMed: 21545110] (b)Dror RO; Pan AC; Arlow DH; Borhani DW; Maragakis P; Shan Y; Xu H; Shaw DE, Pathway and

mechanism of drug binding to G-protein-coupled receptors. Proc Natl Acad Sci 2011, 108 (32), 13118–23. [PubMed: 21778406]

14. 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. [PubMed: 21646537]

15. (a)Kappel K; Miao Y; McCammon JA, Accelerated Molecular Dynamics Simulations of Ligand Binding to a Muscarinic G-protein Coupled Receptor. Quarterly Reviews of Biophysics 2015, 48 (04), 479–487; [PubMed: 26537408] (b)Saglam AS; Chong LT, Protein-protein binding pathways and calculations of rate constants using fully-continuous, explicit-solvent simulations. Chemical Science 2019, 10 (8), 2360–2372. [PubMed: 30881664]

16. (a)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; [PubMed: 25605901] (b)Casasnovas R; Limongelli V; Tiwary P; Carloni P; Parrinello M, Unbinding Kinetics of a p38 MAP Kinase Type II Inhibitor from Metadynamics Simulations. J Am Chem Soc 2017, 139 (13), 4780–4788. [PubMed: 28290199]

17. (a)Kokh DB; Amaral M; Bomke J; Gradler U; Musil D; Buchstaller HP; Dreyer MK; Frech M; Lowinski M; Vallee F; Bianciotto M; Rak A; Wade RC, Estimation of Drug-Target Residence Times by tau-Random Acceleration Molecular Dynamics Simulations. J Chem Theory Comput 2018, 14 (7), 3859–3869; [PubMed: 29768913] (b)Wang T; Duan Y, Chromophore channeling in the G-protein coupled receptor rhodopsin. J. Am. Chem. Soc 2007, 129 (22), 6970–6971. [PubMed: 17500517]

18. (a)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;(b)Gonzalez A; Perez-Acle T; Pardo L; Deupi X, Molecular Basis of Ligand Dissociation in beta-Adrenergic Receptors. Plos One 2011, 6 (9), e23815. [PubMed: 21915263]

19. Dickson A; Lotz SD, Multiple Ligand Unbinding Pathways and Ligand-Induced Destabilization Revealed by WExplore. Biophys J 2017, 112 (4), 620–629. [PubMed: 28256222]

20. Deb I; Frank AT, Accelerating Rare Dissociative Processes in Biomolecules Using Selectively Scaled MD Simulations. J Chem Theory Comput 2019, 15 (11), 5817–5828. [PubMed: 31509413]

21. Tang ZY; Chang CEA, Binding Thermodynamics and Kinetics Calculations Using Chemical Host and Guest: A Comprehensive Picture of Molecular Recognition. J Chem Theory Comput 2018, 14 (1), 303–318. [PubMed: 29149564]

22. Plattner N; Noe F, Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. Nat Commun 2015, 6, 7653. [PubMed: 26134632]

23. Tiwary P; Parrinello M, From Metadynamics to Dynamics. Phys Rev Lett 2013, 111 (23).

24. Shao Q; Zhu W, Exploring the Ligand Binding/Unbinding Pathway by Selectively Enhanced Sampling of Ligand in a Protein-Ligand Complex. J Phys Chem B 2019, 123 (38), 7974–7983. [PubMed: 31478672]

25. Miao Y; Feher VA; McCammon JA, Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. J Chem Theory Comput 2015, 11 (8), 3584–3595. [PubMed: 26300708]

26. Miao Y, Acceleration of Biomolecular Kinetics in Gaussian Accelerated Molecular Dynamics. J Chem Phys 2018, 149 (7), 072308. [PubMed: 30134710]

27. Miao Y; McCammon JA, Unconstrained enhanced sampling for free energy calculations of biomolecules: a review. Molecular simulation 2016, 42 (13), 1046–1055. [PubMed: 27453631]

28. (a)Hamelberg D; Mongan J; McCammon JA, Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. J Chem Phys 2004, 120 (24), 11919–11929; [PubMed: 15268227] (b)Voter AF, Hyperdynamics: Accelerated molecular dynamics of infrequent events. Physical Review Letters 1997, 78 (20), 3908.

29. Shen TY; Hamelberg D, A statistical analysis of the precision of reweighting-based simulations. J Chem Phys 2008, 129 (3), 034103. [PubMed: 18647012]

30. Case D; Babin V; Berryman J; Betz R; Cai Q; Cerutti D; Cheatham III T; Darden T; Duke R; Gohlke H, Amber 14, University of California, San Francisco 2014.

31. Pang YT; Miao Y; Wang Y; McCammon JA, Gaussian Accelerated Molecular Dynamics in NAMD. J Chem Theory Comput 2017, 13 (1), 9–19. [PubMed: 28034310]

32. Oshima H; Re S; Sugita Y, Replica-Exchange Umbrella Sampling Combined with Gaussian Accelerated Molecular Dynamics for Free-Energy Calculation of Biomolecules. J Chem Theory Comput 2019, 15 (10), 5199–5208. [PubMed: 31539245]

33. (a)Miao Y; McCammon JA, Gaussian Accelerated Molecular Dynamics: Theory, Implementation and Applications. Annu Rep Comp Chem 2017, 13, 231–278; [PubMed: 29720925] (b)Pawnikar SP; Miao Y, Pathway and Mechanism of Drug Binding to Chemokine Receptors Revealed by Accelerated Molecular Simulations. Future Med Chem 2020, In Press.

34. (a)Wang J; Miao Y, Mechanistic Insights into Specific G Protein Interactions with Adenosine Receptors. The Journal of Physical Chemistry B 2019, 123 (30), 6462–6473; [PubMed: 31283874] (b)Miao Y; McCammon JA, Mechanism of the G-Protein Mimetic Nanobody Binding to a Muscarinic G-Protein-Coupled Receptor. Proc Natl Acad Sci 2018, 115 (12), 3036–3041. [PubMed: 29507218]

35. Bhattarai A; Wang J; Miao Y, G-Protein-Coupled Receptor-Membrane Interactions Depend on the Receptor Activation State. J Comput Chem 2019, 41, 460–471. [PubMed: 31602675]

36. (a)East KW; Newton JC; Morzan UN; Narkhede YB; Acharya A; Skeens E; Jogl G; Batista VS; Palermo G; Lisi GP, Allosteric Motions of the CRISPR-Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics. J Am Chem Soc 2020, 142 (3), 1348–1358; [PubMed: 31885264] (b)Ricci CG; Chen JS; Miao Y; Jinek M; Doudna JA; McCammon JA; Palermo G, Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics. ACS central science 2019, 5 (4), 651–662. [PubMed: 31041385]

37. Wereszczynski J; McCammon JA, Using Selectively Applied Accelerated Molecular Dynamics to Enhance Free Energy Calculations. J Chem Theory Comput 2010, 6 (11), 3285–3292. [PubMed: 21072329]

38. (a)Zheng LQ; Yang W, Essential energy space random walks to accelerate molecular dynamics simulations: Convergence improvements via an adaptive-length self-healing strategy. J Chem Phys 2008, 129 (1), 014105; [PubMed: 18624468] (b)Li H; Min D; Liu Y; Yang W, Essential energy space random walk via energy space metadynamics method to accelerate molecular dynamics simulations. J Chem Phys 2007, 127 (9), 094101. [PubMed: 17824726]

39. Liu P; Kim B; Friesner RA; Berne BJ, Replica exchange with solute tempering: A method for sampling biological systems in explicit water. Proc Natl Acad Sci 2005, 102 (39), 13749–13754. [PubMed: 16172406]

40. Wang L; Friesner RA; Berne BJ, Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). J Phys Chem B 2011, 115 (30), 9431–8. [PubMed: 21714551]

41. (a)Vanommeslaeghe K; MacKerell AD Jr., CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. Biochimica et biophysica acta 2014;(b)Duan Y; Wu C; Chowdhury S; Lee MC; Xiong GM; Zhang W; Yang R; Cieplak P; Luo R; Lee T; Caldwell J; Wang JM; 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. [PubMed: 14531054]

42. Miao Y; Bhattarai A; Nguyen ATN; Christopoulos A; May LT, Structural Basis for Binding of Allosteric Drug Leads in the Adenosine A1 Receptor. Sci. Rep 2018, 8 (1), 16836. [PubMed: 30442899]

43. Ferruz N; De Fabritiis G, Binding Kinetics in Drug Discovery. Mol Inform 2016, 35 (6–7), 216–26. [PubMed: 27492236]

44. (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; [PubMed: 26596291] (b)Frank AT; Andricioaei I, Reaction Coordinate-Free Approach to Recovering Kinetics from Potential-Scaled Simulations: Application of Kramers' Rate Theory. The Journal of Physical Chemistry B 2016, 120 (33), 8600–8605. [PubMed: 27220565]

45. Hamelberg D; Shen T; McCammon JA, Relating kinetic rates and local energetic roughness by accelerated molecular-dynamics simulations - art. no. 2411003. J Chem Phys 2005, 122 (24).

46. Truhlar DG; Garrett BC; Klippenstein SJ, Current Status of Transition-State Theory. The Journal of Physical Chemistry 1996, 100 (31), 12771–12800.

47. Humphrey W; Dalke A; Schulten K, VMD: Visual molecular dynamics. Journal of Molecular Graphics & Modelling 1996, 14 (1), 33–38.

48. Roe DR; Cheatham TE, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. J Chem Theory Comput 2013, 9 (7), 3084–3095. [PubMed: 26583988]

49. Marquart M; Walter J; Deisenhofer J; Bode W; Huber R, The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. Acta Crystallographica Section B 1983, 39 (4), 480–490.

50. Case DA; Cheatham TE; Darden T; Gohlke H; Luo R; Merz KM; Onufriev A; Simmerling C; Wang B; Woods RJ, The Amber biomolecular simulation programs. J Comput Chem 2005, 26 (16), 1668–1688. [PubMed: 16200636]

51. Wang J; Wolf RM; Caldwell JW; Kollman PA; Case DA, Development and testing of a general amber force field. Journal of computational chemistry 2004, 25 (9), 1157–1174. [PubMed: 15116359]

52. Maier JA; Martinez C; Kasavajhala K; Wickstrom L; Hauser KE; Simmerling C, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput 2015, 11 (8), 3696–3713. [PubMed: 26574453]

53. Cieplak P; Cornell WD; Bayly C; Kollman PA, Application of the multimolecule and multiconformational RESP methodology to biopolymers: charge derivation for DNA, RNA, and proteins. Journal of Computational Chemistry 1995, 16 (11), 1357–1377.

54. Jorgensen WL; Chandrasekhar J; Madura JD; Impey RW; Klein ML, Comparison of Simple Potential Functions for Simulating Liquid Water. J Chem Phys 1983, 79 (2), 926–935.

55. Ester M; Kriegel H-P; Sander J; Xu X, A density-based algorithm for discovering clusters in large spatial databases with noise. Knowledge Discovery and Data Mining 1996, 96 (34), 226–231.

56. Voet D; Voet JG; Pratt CW, Fundamentals of Biochemistry, 4th ed. Ch. 11, 339–349.

57. Guillain F; Thusius D, The use of proflavin as an indicator in temperature-jump studies of the binding of a competitive inhibitor to trypsin. J Am Chem Soc 1970, 92 (18), 5534–6. [PubMed: 5449454]

58. Wade RC; Gabdoulline RR; Ludemann SK; Lounnas V, Electrostatic steering and ionic tethering in enzyme-ligand binding: insights from simulations. Proc Natl Acad Sci 1998, 95 (11), 5942–9. [PubMed: 9600896]

59. Case DA; Belfon K; Ben-Shalom IY; Brozell SR; Cerutti DS; Cheatham TE I.; Cruzeiro VWD; Darden TA; Duke RE; Giambasu G; Gilson MK; Gohlke H; Goetz AW; Harris R; Izadi S; jhala KK–; Kovalenko A; Krasny R; Kurtzman T; Lee TS; LeGrand S; Li P; Lin C; Liu J; Luchko T; Luo R; Man V; Merz KM; Miao Y; Mikhailovskii O; Monard G; Nguyen H; Onufriev A; Pan F; Pantano S; Qi R; Roe DR; Roitberg A; Sagui C; Schott-Verdugo S; Shen J; Simmerling CL; Skrynnikov N; Smith J; Swails J; Walker RC; Wang J; Wilson L; Wolf RM; Wu X; York DM; Kollman PA, AMBER 20, University of California, San Francisco 2020.

60. Miao Y; Sinko W; Pierce L; Bucher D; McCammon JA, Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. J Chem Theory Comput 2014, 10 (7), 2677–2689. [PubMed: 25061441]

61. Kramers HA, Brownian motion in a field of force and the diffusion model of chemical reactions. Physica 1940, 7, 284–304.

62. Xin Y; Doshi U; Hamelberg D, Examining the limits of time reweighting and Kramers' rate theory to obtain correct kinetics from accelerated molecular dynamics. J Chem Phys 2010, 132 (22), 224101. [PubMed: 20550384]
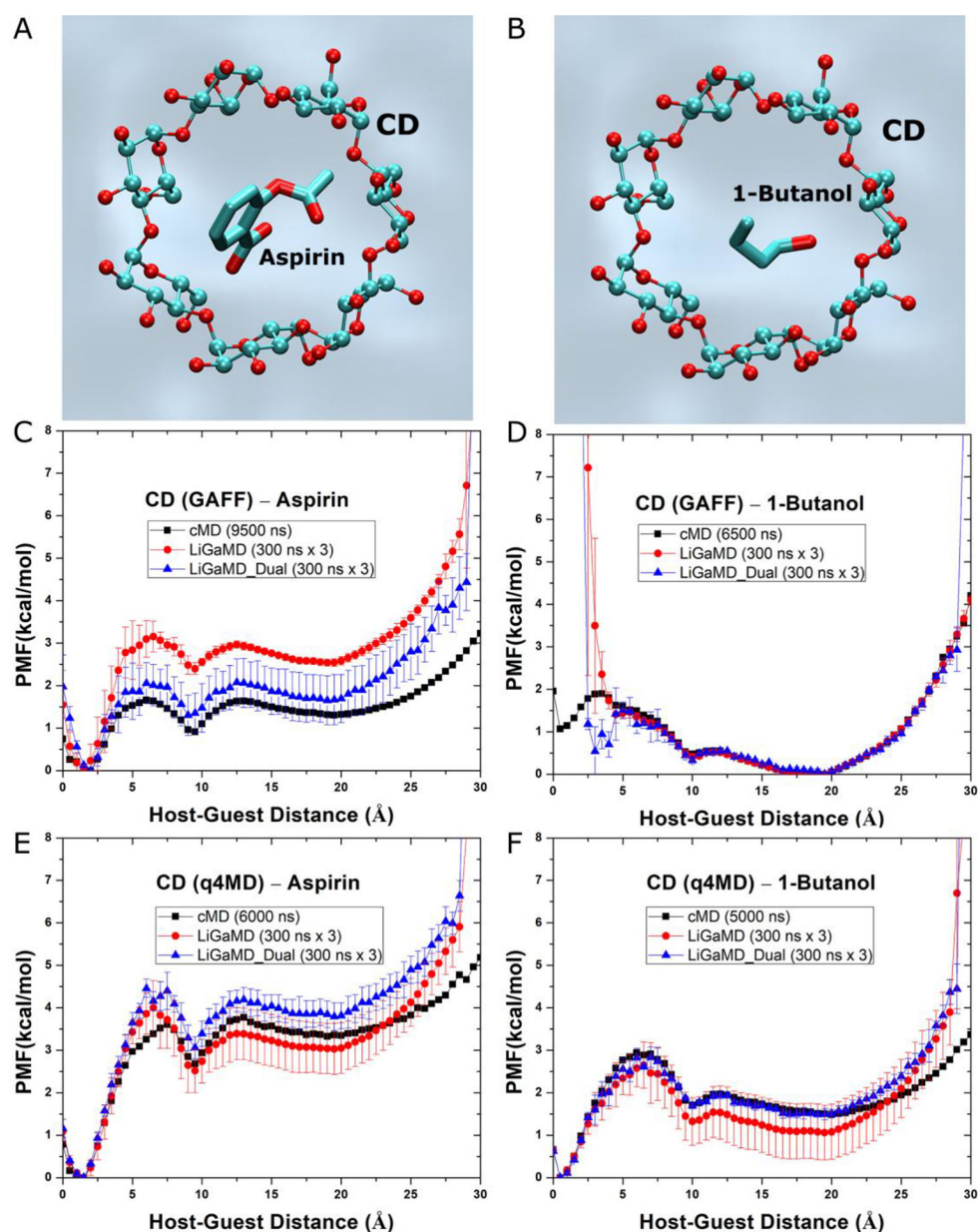
Fig. 1.

Comparison of potential of mean force (PMF) free energy profiles calculated from conventional MD (cMD), LiGaMD and dual-boost LiGaMD (LiGaMD_Dual) simulations of host-guest binding: (A-B) Computational models of the β-cyclodextrin (CD) host (balls-and-sticks) in the presence of guest (A) aspirin and (B) 1-butanol (thick sticks) in aqueous medium (cyan). (C-F) PMF profiles calculated from microsecond-timescale cMD simulations and three independent 300 ns LiGaMD and LiGaMD_Dual simulations of (C) CD using the GAFF force field with aspirin, (D) CD using the GAFF force field with 1-butanol, (E) CD using the q4MD force field with aspirin, (F) CD using the q4MD force field with 1-butanol.
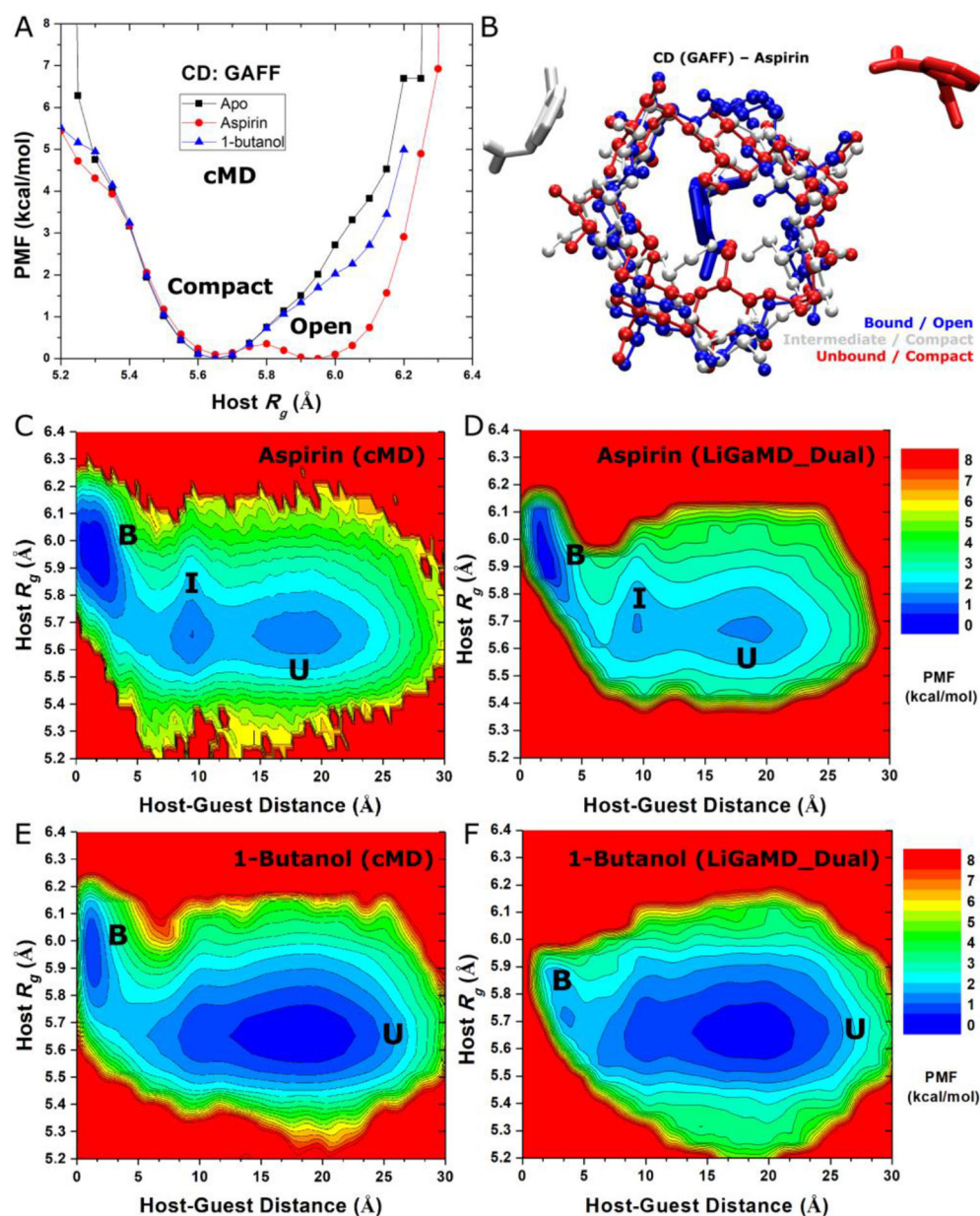
Fig. 2.

Free energy profiles and low-energy conformational states of guest binding to the CD host that was modeled with the GAFF force field: (A) PMF profiles of the host radius of gyration (R_g) calculated from cMD simulations in the ligand-free (apo), aspirin and 1-butanol binding forms. Two low-energy states ("Compact" and "Open") are identified for the CD host. (B) Three representative conformational states observed in simulations of guest aspirin binding to CD modeled with GAFF: the "Bound (B)", "Intermediate (I)" and "Unbound (U)", in which the CD host adopted primarily the Open, Compact and Compact conformations, respectively. (C-E) 2D PMF profiles of the host R_g versus the center-of-mass distance between host-guest of (C) aspirin binding from 9500 ns cMD simulation, (D) aspirin binding from three 300 ns LiGaMD_Dual simulation combined, (E) 1-butanol binding from 6500 ns

cMD simulation, and (F) 1-butanol binding from three 300 ns LiGaMD_Dual simulation combined. The low-energy states are labeled.
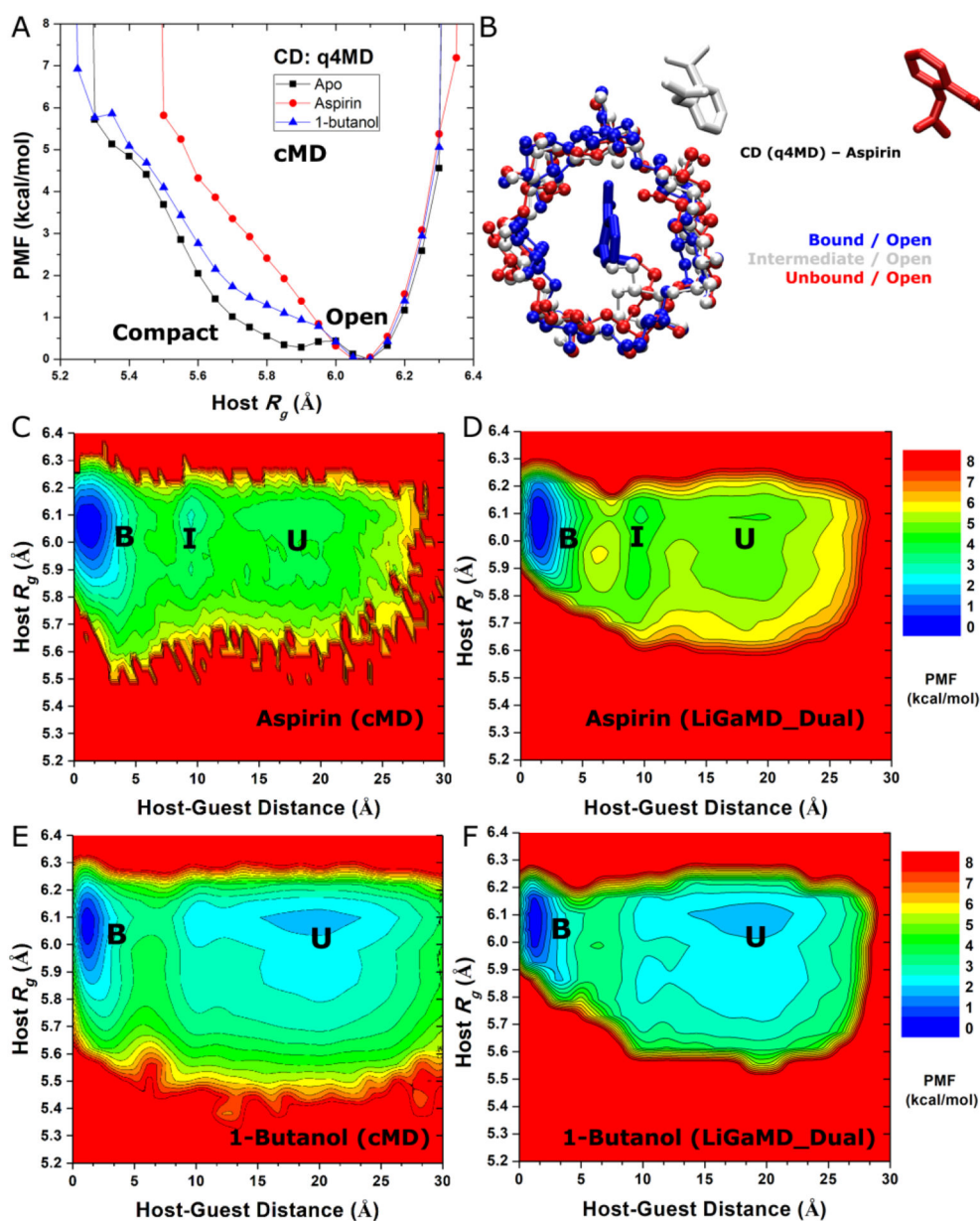
Fig. 3.

Free energy profiles and low-energy conformational states of guest binding to the CD host that was modeled with the q4MD force field: (A) PMF profiles of the host radius of gyration (R_g) calculated from cMD simulations in the ligand-free (apo), aspirin and 1-butanol binding forms. Two low-energy states ("Compact" and "Open") are labeled for the CD host. (B) Three representative conformational states observed in simulations of aspirin binding to CD modeled with q4MD: the "Bound (B)", "Intermediate (I)" and "Unbound (U)", in which the CD host all adopted primarily the Open conformation. (C-E) 2D PMF profiles regarding the host R_g versus center-of-mass distance between host-guest of (C) aspirin binding from 6000 ns cMD simulation, (D) aspirin binding from three 300 ns LiGaMD_Dual simulation

combined, (E) 1-butanol binding from 5000 ns cMD simulation, and (F) 1-butanol binding from three 300 ns LiGaMD_Dual simulation combined. The low-energy states are labeled.
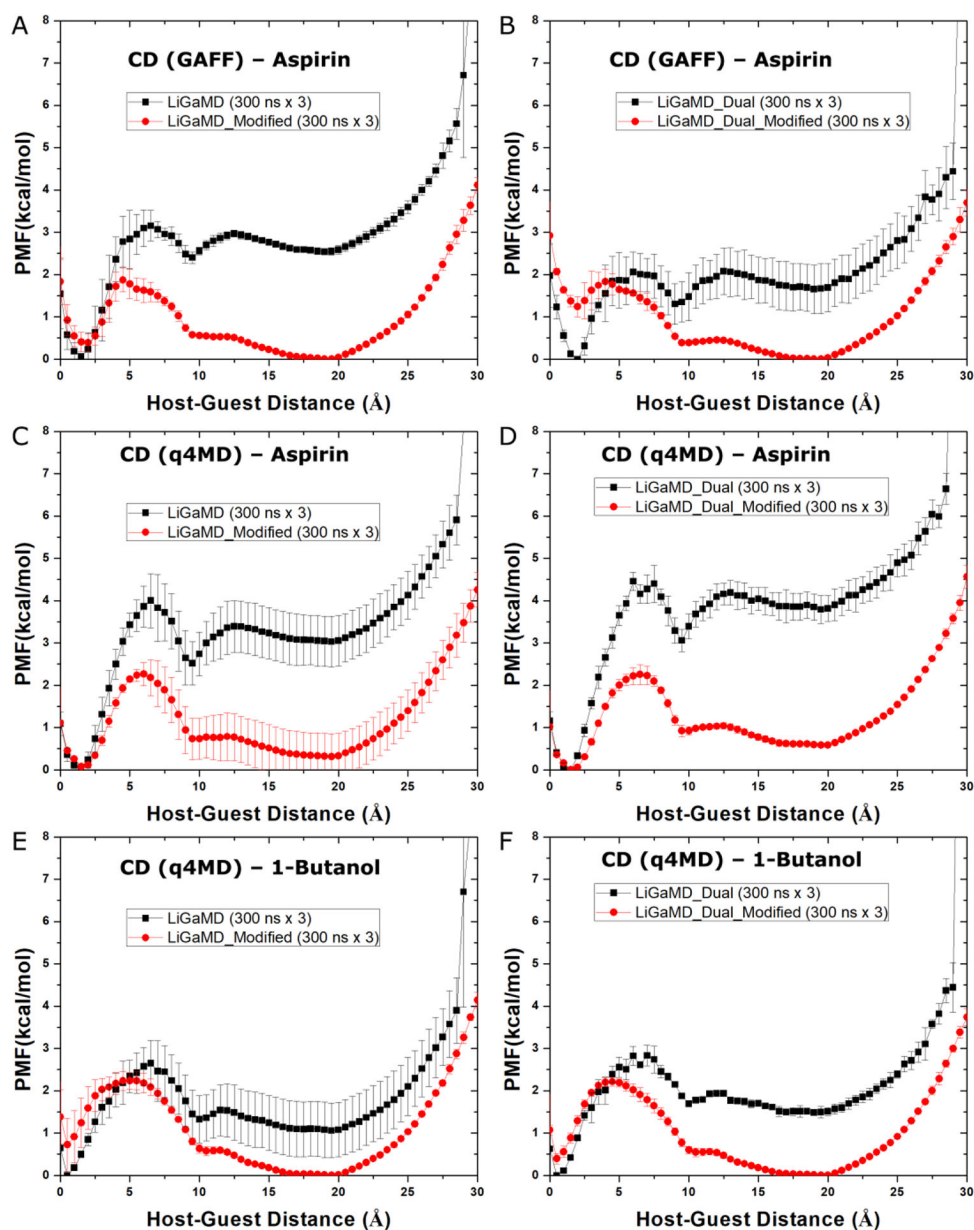
Fig. 4.
(A-F) The original (reweighted) and modified (no reweighting) PMF profiles of (A)
LiGaMD simulations of CD using the GAFF force field with aspirin, (B) LiGaMD_Dual
simulations of CD using the GAFF force field with aspirin, (C) LiGaMD simulations of CD
using the q4MD force field with aspirin, (D) LiGaMD_Dual simulations of CD using the
q4MD force field with aspirin, (E) LiGaMD simulations of CD using the q4MD force field
with 1-butanol, and (F) LiGaMD_Dual simulations of CD using the q4MD force field with
1-butanol.
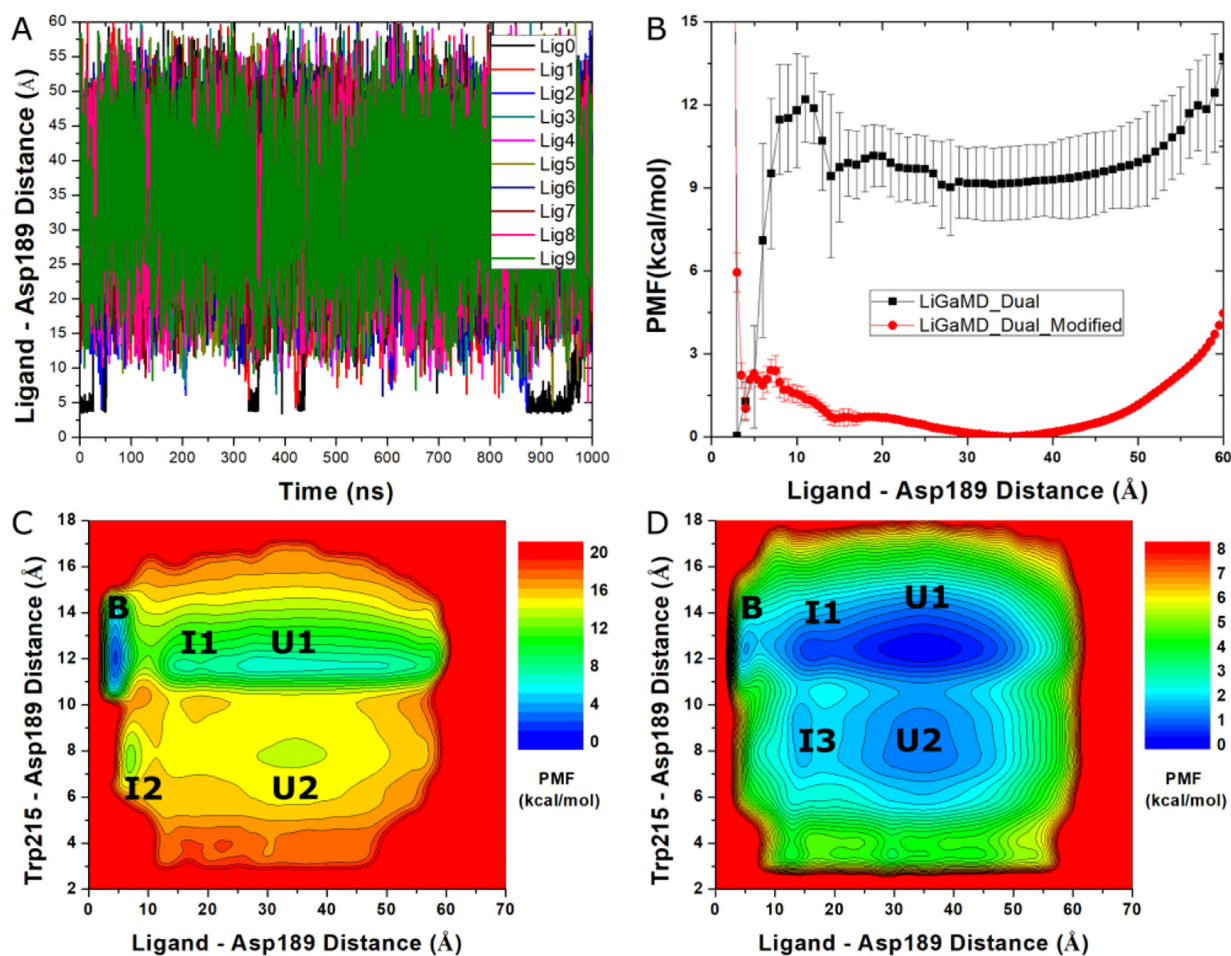
Fig. 5.
LiGaMD_Dual simulations have captured repetitive binding and unbinding of the benzamidine ligand in the trypsin protein: (A) time courses of distances between the charge centers at the amidine C atom in benzamidine and CG atom of Asp189 in trypsin calculated from a representative 1 μs LiGaMD_Dual simulation (Sim2 in Table 5). Distance plots of the other simulations are provided in Figure S5. (B) The original (reweighted) and modified (no reweighting) PMF profiles of the distance between the amidine C atom of benzamidine ligand and CG atom of protein residue Asp189. (C) Reweighted and (D) modified 2D PMF profiles of the benzamidine:C – Asp189:CG and Trp215:NE – Asp189:CG atom distances. The PMF profiles were calculated by combining all five 1000 ns LiGaMD_Dual simulations.
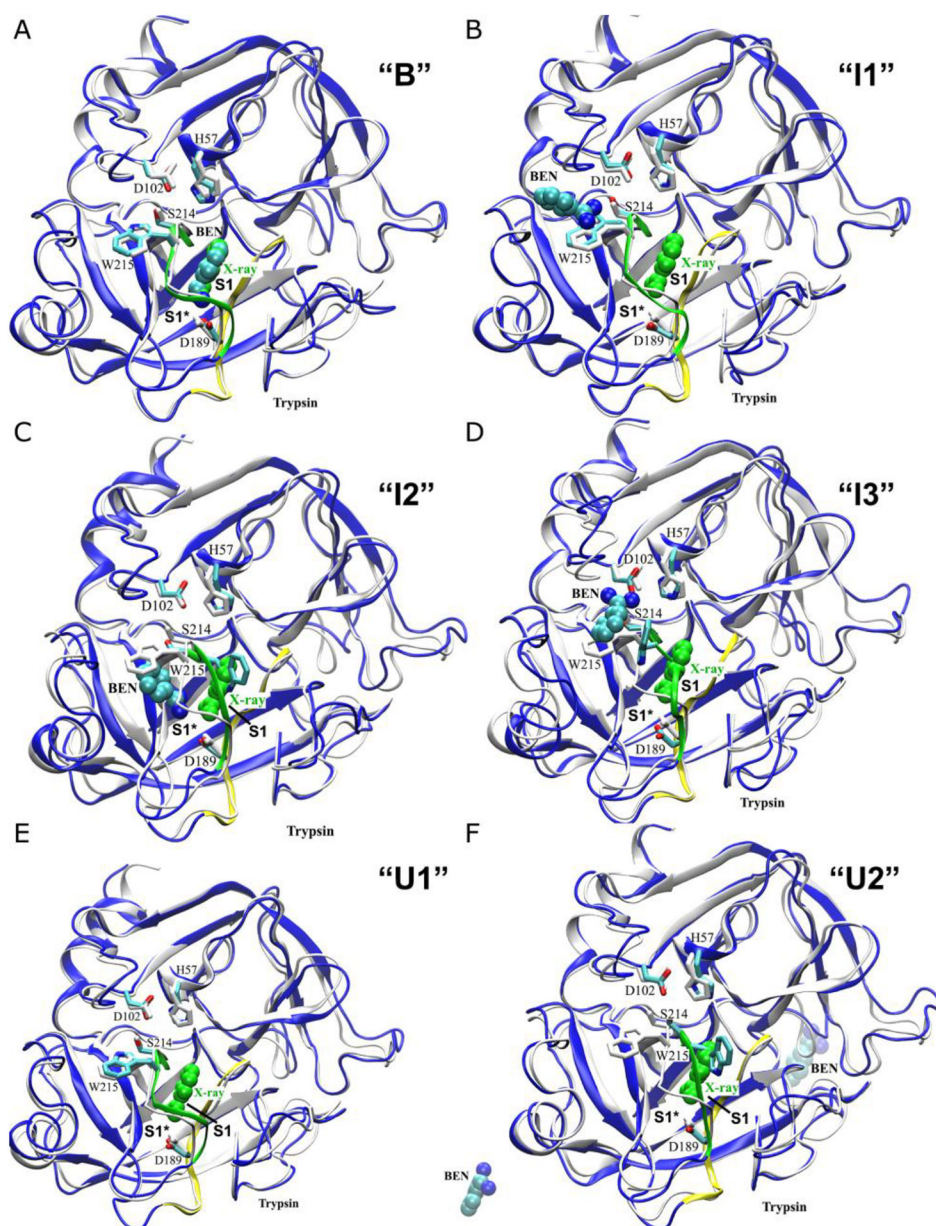
Fig. 6.
Six representative low-energy conformational states as identified from PMF profiles of benzamidine (BEN) ligand (spheres) binding to trypsin (blue ribbons): (A) "Bound (B)", (B) "Intermediate 1 (I1)", (C) "Intermediate 2 (I2)", (D) "Intermediate 3 (I3)", (E) "Unbound (U1)", and (F) "Unbound (U2)". Reference X-ray conformations of the ligand and protein (PDB: 3PTB) are shown in green spheres and grey ribbons, respectively. Protein residues Asp189 and Trp215 that are important for ligand binding and those in the catalytic triad (His57, Asp102 and Ser214) are highlighted in thick sticks.
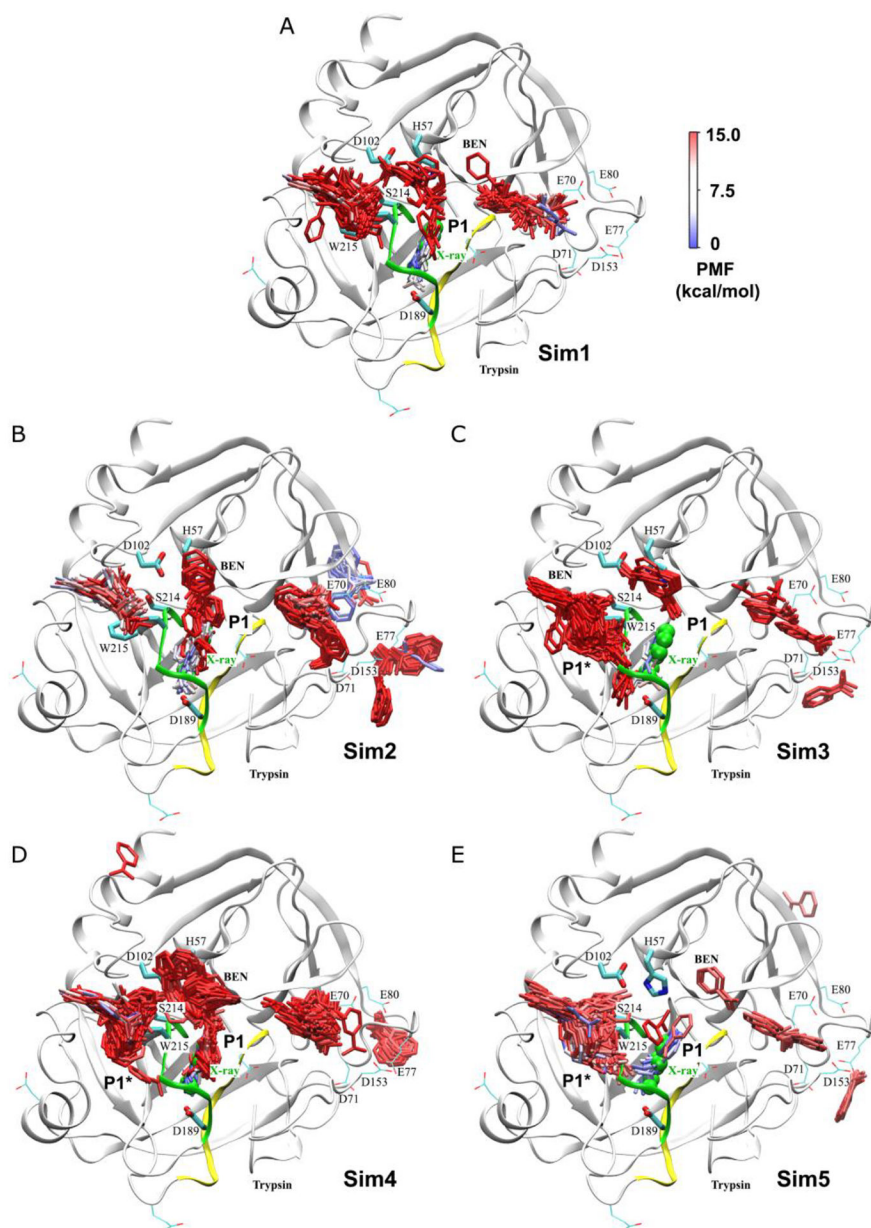
Fig. 7.
Pathways of the benzamidine (BEN) ligand (sticks) in trypsin (ribbons) obtained from structural clustering of five 1000 *ns* LiGaMD_Dual simulations: (A) Sim1, (B) Sim2, (C) Sim3, (D) Sim4 and (E) Sim5. The ligand is represented by sticks that are colored by reweighted PMF values of the ligand clusters in a blue (0 kcal/mol)-white (7.5 kcal/mol)-red (15.0 kcal/mol) scale. X-ray conformation of the ligand is shown in green spheres for reference. Protein residues Asp189 and Trp215 that are important for ligand binding and those in the catalytic triad (His57, Asp102 and Ser214) are highlighted in thick sticks. Other negatively charged residues in the protein including Glu70, Asp71, Glu77, Glu80 and

Asp153 are shown in thin sticks. The protein loops containing residues Asp189 and Trp215 are highlighted in yellow and green, respectively.

Table 1

Summary of host-guest binding thermodynamics and kinetics obtained from (A) experimental data, (B) cMD simulations from Ref.[21] and (C) GaMD simulations on the binding of two guest molecules (aspirin and 1-butanol) to the β-cyclodextrin (CD) host. In simulations, the CD host has been modeled with both the GAFF and q4MD force fields. G is the ligand binding free energy. The G_{comp1} and G_{comp2} are ligand binding free energies with G_{comp1} computed using the guest binding enthalpy and entropy and G_{comp2} computed using the guest dissociation and binding rate constants estimated from the cMD simulations[21]. k_{on} and k_{off} are the kinetic dissociation and binding rate constants, respectively, with N_D and N_B being the number of host-guest dissociation and binding events collected from individual simulations. V is the system boost potential applied in GaMD simulations. $k_{on}*$ and $k_{off}*$ are the kinetic dissociation and binding rate constants, which are calculated from only the GaMD simulations with $N_D > 0$ and $N_B > 0$.

Host	Ligand	G (kcal/mol)	k_{on} ($\times 10^8$ M·s^{-1})	k_{off} ($\times 10^6$ s^{-1})
CD	Aspirin	−3.74 ± 0.00	7.2 ± 0.04	1.3 ± 0.03
	1-Butanol	−1.67 ± 0.19	2.8 ± 0.8	38 ± 6
(A) Experimental data				

Host	Ligand	cMD	N_D	N_B	G_{comp1} (kcal/mol)	G_{comp2} (kcal/mol)	k_{on} ($\times 10^8$ M·s^{-1})	k_{off} ($\times 10^6$ s^{-1})
CD: GAFF	Aspirin	9500 ns	133	133	−3.84 ± 0.35	−2.27 ± 0.06	11 ± 0.1	24 ± 3
	1-Butanol	6500 ns	42	42	1.01 ± 0.59	−0.41 ± 0.10	2.2 ± 0.1	110 ± 20
CD: q4MD	Aspirin	6000 ns	17	18	−6.33 ± 0.37	−4.11 ± 0.05	32 ± 3	3.1 ± 0.9
	1-Butanol	5000 ns	89	89	−1.33 ± 0.34	−2.27 ± 0.02	15 ± 0.3	33 ± 0.7
(B) cMD simulations								

Host	Ligand	GaMD (300 ns × 3)	V (kcal/mol)	N_D	N_B	G (kcal/mol)	$k_{on}*$ ($\times 10^8$ M^{-1}·s^{-1})	$k_{off}*$ ($\times 10^6$ s^{-1})
CD: GAFF	Aspirin	GaMD_Tot	6.48 ± 2.59	0, 2, 1	0, 1, 0	-	-	-
		GaMD_Dual	10.37 ± 2.98	0, 3, 0	0, 3, 0	-	-	-
		GaMD_NB	4.15 ± 2.26	1, 1, 2	1, 0, 1	-	-	-
		GaMD_NB_Dual	8.75 ± 2.82	0, 1, 0	0, 1, 0	-	-	-
		LiGaMD	2.60 ± 1.27	3, 1, 2	2, 1, 1	−4.90 ± 0.03	3.05 ± 0.92	40.67 ± 24.25
		LiGaMD_Dual	9.18 ± 2.84	2, 4, 2	2, 3, 2	−2.89 ± 0.18	3.93 ± 0.98	275.6 ± 57.5
	1-Butanol	LiGaMD	2.54 ± 1.00	1, 0, 0	1, 0, 0	-	-	-
		LiGaMD_Dual	9.51 ± 2.85	1, 0, 1	1, 0, 1	-	-	-
CD: q4MD	Aspirin	GaMD_Tot	6.22 ± 2.52	0, 2, 0	0, 2, 0	-	-	-
		GaMD_Dual	10.77 ± 3.02	1, 0, 0	1, 0, 0	-	-	-
		GaMD_NB	4.13 ± 2.17	0, 2, 2	0, 2, 2	-	-	-
		GaMD_NB_Dual	8.78 ± 2.83	0, 2, 2	0, 2, 2	-	-	-
		LiGaMD	2.84 ± 1.34	4, 3, 2	3, 2, 2	−4.74 ± 0.53	8.99 ± 5.02	42.82 ± 15.41
		LiGaMD_Dual	10.06 ± 3.15	8, 7, 4	8, 6, 3	−3.34 ± 0.15	10.37 ± 2.60	56.61 ± 19.05
	1-Butanol	LiGaMD	2.87 ± 1.12	9, 4, 7	9, 4, 7	−2.84 ± 0.58	8.76 ± 2.93	482.8 ± 246.3

Host	Ligand	GaMD (300 ns × 3)	V (kcal/mol)	N_D	N_B	G (kcal/mol)	$k_{on}*$ (×10^8 M^{-1}·s^{-1})	$k_{off}*$ (×10^6 s^{-1})
		LiGaMD_Dual	9.74 ± 2.93	9, 13, 8	9, 13, 8	−1.34 ± 0.10	13.53 ± 2.15	436.8 ± 90.7
(C) GaMD simulations.								

Table 2

Comparison of the calculated host-guest binding free energies: (A) between simulations and experimental data[21] and (B) between LiGaMD and cMD simulations[21].

Host	Ligand	$G_{sim} - G_{exp}$ (kcal/mol)			
		cMD (G_{comp1})	cMD (G_{comp2})	LiGaMD	LiGaMD_Dual
CD: GAFF	Aspirin	−0.10 ± 0.35	1.47 ± 0.06	−1.16 ± 0.03	0.85 ± 0.18
	1-Butanol	2.68 ± 0.62	1.26 ± 0.21	-	-
CD: q4MD	Aspirin	−2.59 ± 0.37	−0.37 ± 0.05	−1.00 ± 0.53	0.40 ± 0.15
	1-Butanol	0.34 ± 0.39	−0.60 ± 0.19	−1.17 ± 0.61	0.33 ± 0.21
(A) Comparison of simulations with experimental data.					

Host	Ligand	$G - G_{comp1}$ (kcal/mol)		$G - G_{comp2}$ (kcal/mol)	
		LiGaMD	LiGaMD_Dual	LiGaMD	LiGaMD_Dual
CD: GAFF	Aspirin	−1.06 ± 0.35	0.95 ± 0.39	−2.63 ± 0.07	−0.62 ± 0.19
CD: q4MD	Aspirin	1.59 ± 0.65	2.99 ± 0.40	−0.63 ± 0.53	0.77 ± 0.16
	1-Butanol	−1.51 ± 0.67	−0.01 ± 0.35	−0.57 ± 0.58	0.93 ± 0.10
(B) Comparison of LiGaMD with cMD simulations.					

Table 3

Accelerations of host-guest dissociation and binding rate constants obtained from (A) LiGaMD and (B) LiGaMD_Dual simulations as compared directly with previous cMD simulations[21] and those derived using the Transition State Theory and Kramers' Rate Theory.

Host	Ligand	LiGaMD vs. cMD		Transition State Theory		Kramer's Rate Theory	
		$k_{on}*/k_{on}$	$k_{off}*/k_{off}$	$k_{on}*/k_{on}$	$k_{off}*/k_{off}$	$k_{on}*/k_{on}$	$k_{off}*/k_{off}$
CD: GAFF	Aspirin	0.28 ± 0.08	1.69 ± 1.03	0.12 ± 0.09	15.07 ± 12.94	0.20 ± 0.16	12.19 ± 10.47
CD: q4MD	Aspirin	0.28 ± 0.16	13.81 ± 6.39	0.19 ± 0.16	21.59 ± 16.11	0.02 ± 0.02	12.36 ± 9.22
	1-Butanol	0.58 ± 0.20	14.63 ± 7.47	0.43 ± 0.34	8.77 ± 6.72	0.55 ± 0.43	7.91 ± 6.06
(A) LiGaMD simulations							

Host	Ligand	LiGaMD_Dual vs. cMD		Transition State Theory		Kramer's Rate Theory	
		$k_{on}*/k_{on}$	$k_{off}*/k_{off}$	$k_{on}*/k_{on}$	$k_{off}*/k_{off}$	$k_{on}*/k_{on}$	$k_{off}*/k_{off}$
CD: GAFF	Aspirin	0.36 ± 0.09	11.48 ± 2.79	0.09 ± 0.07	11.89 ± 9.36	0.06 ± 0.05	7.97 ± 6.27
CD: q4MD	Aspirin	0.32 ± 0.09	18.26 ± 8.12	0.18 ± 0.14	41.37 ± 22.17	0.23 ± 0.18	22.82 ± 12.23
	1-Butanol	0.90 ± 0.14	13.24 ± 2.76	0.23 ± 0.11	5.54 ± 2.73	1.03 ± 0.47	5.96 ± 2.94
(B) LiGaMD_Dual simulations							

Table 4

Summary of LiGaMD_Dual production simulations performed on the benzamidine ligand binding to trypsin. V is the GaMD boost potential. N_D and N_B are the number of host-guest dissociation and binding events recorded from individual simulations.

System	Method	ID	Length (ns)	V (kcal/mol)	N_D	N_B
Trypsin - benzamidine	LiGaMD_Dual	Sim1	1000	20.93 ± 4.24	3	3
		Sim2	1000	21.64 ± 4.67	5	4
		Sim3	1000	20.99 ± 4.23	11	10
		Sim4	1000	20.87 ± 4.18	4	4
		Sim5	1000	20.94 ± 4.18	4	4

Table 5

Comparison of trypsin protein-ligand binding free energy and kinetic rates obtained from experimental data and LiGaMD_Dual simulations. G is the ligand binding free energy. k_{on} and k_{off} are the kinetic dissociation and binding rate constants, respectively, from experimental data or LiGaMD_Dual simulations with reweighting using Kramers' rate theory. $k_{on}*$ and $k_{off}*$ are the accelerated kinetic dissociation and binding rate constants calculated directly from LiGaMD_Dual simulations without reweighting.

Method	G (kcal/mol)	k_{on} (M^{-1}·s^{-1})	k_{off} (s^{-1})	$k_{on}*$ (M^{-1}·s^{-1})	$k_{off}*$ (s^{-1})
Experiment	−6.2	2.9×10^7	600 ± 300	-	-
LiGaMD_Dual	-6.13 ± 0.35	$1.15 \pm 0.79 \times 10^7$	3.53 ± 1.41	$1.58 \pm 1.09 \times 10^8$	$3.40 \pm 1.36 \times 10^7$