UC Riverside
UC Riverside Previously Published Works

Title
Gaussian accelerated molecular dynamics: Principles and applications

Permalink
https://escholarship.org/uc/item/0036f4tg

Journal
Wiley Interdisciplinary Reviews Computational Molecular Science, 11(5)

ISSN
1759-0876

Authors
Wang, Jinan
Arantes, Pablo R
Bhattarai, Apurba
et al.

Publication Date
2021-09-01

DOI
10.1002/wcms.1521

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

HHS Public Access
Author manuscript
Wiley Interdiscip Rev Comput Mol Sci. Author manuscript; available in PMC 2022 September 01.

Published in final edited form as:
Wiley Interdiscip Rev Comput Mol Sci. 2021 ; 11(5): . doi:10.1002/wcms.1521.

Gaussian accelerated molecular dynamics (GaMD): principles and applications

Jinan Wang[#],
Center for Computational Biology and Department of Molecular Biosciences, University of Kansas, 2030 Becker Dr., Lawrence, KS, 66047, United States

Pablo R. Arantes[#],
Department of Bioengineering, University of California Riverside, 900 University Avenue, Riverside, CA 92512, United States

Apurba Bhattarai,
Center for Computational Biology and Department of Molecular Biosciences, University of Kansas, 2030 Becker Dr, Lawrence, KS, 66047, United States

Rohaine V. Hsu,
Department of Bioengineering, University of California Riverside, 900 University Avenue, Riverside, CA 92512, United States

Shristi Pawnikar,
Center for Computational Biology and Department of Molecular Biosciences, University of Kansas, 2030 Becker Dr., Lawrence, KS, 66047, United States

Yu-ming M. Huang,
Department of Physics & Astronomy, Wayne State University, 666 W Hancock St, Detroit, MI 48207, USA

Giulia Palermo,
Department of Bioengineering and Department of Chemistry, University of California Riverside, 900 University Avenue, Riverside, CA 92512, United States

Yinglong Miao

Giulia Palermo, giulia.palermo@ucr.edu, Yinglong Miao, miao@ku.edu.

Author Contributions
Jinan Wang: Data curation; visualization; writing-original draft. **Pablo R. Arantes**: Data curation; visualization; writing-original draft. **Apurba Bhattarai**: Formal analysis; writing-review and editing; **Rohaine V. Hsu**: Formal analysis; writing-review and editing; **Shristi Pawnikar:** Formal analysis; writing-review and editing; **Yu-ming M. Huang**: Formal analysis; writing-review and editing; **Giulia Palermo:** Conceptualization; funding acquisition; project administration; resources; supervision; writing-original draft. **Yinglong Miao:** Conceptualization; funding acquisition; project administration; resources; supervision; writing-original draft.
[#]These authors contributed equally.

Conflict of Interest
The authors have declared no conflict of interest for this article.

Further Reading
A website is available for GaMD at: http://miao.compbio.ku.edu/GaMD/. GaMD has been implemented in the AMBER, NAMD and GENESIS software packages. User manuals and tutorials have been provided for using GaMD in these packages through the GaMD website. For method updates and discussions such as bug reports and simulation questions, users can subscribe to the GaMD mailing list: https://sourceforge.net/projects/gamd/lists/gamd-discuss, and send emails to gamd-discuss@lists.sourceforge.net.

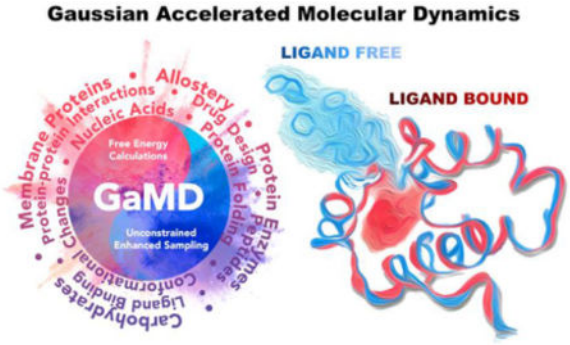
Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Center for Computational Biology and Department of Molecular Biosciences, University of Kansas, 2030 Becker Dr., Lawrence, Kansas 66047, United States

Abstract

Gaussian accelerated molecular dynamics (GaMD) is a robust computational method for simultaneous unconstrained enhanced sampling and free energy calculations of biomolecules. It works by adding a harmonic boost potential to smooth biomolecular potential energy surface and reduce energy barriers. GaMD greatly accelerates biomolecular simulations by orders of magnitude. Without the need to set predefined reaction coordinates or collective variables, GaMD provides unconstrained enhanced sampling and is advantageous for simulating complex biological processes. The GaMD boost potential exhibits a Gaussian distribution, thereby allowing for energetic reweighting via cumulant expansion to the second order (i.e., "Gaussian approximation"). This leads to accurate reconstruction of free energy landscapes of biomolecules. Hybrid schemes with other enhanced sampling methods, such as the replica exchange GaMD (rex-GaMD) and replica exchange umbrella sampling GaMD (GaREUS), have also been introduced, further improving sampling and free energy calculations. Recently, new "selective GaMD" algorithms including the ligand GaMD (LiGaMD) and peptide GaMD (Pep-GaMD) enabled microsecond simulations to capture repetitive dissociation and binding of small-molecule ligands and highly flexible peptides. The simulations then allowed highly efficient quantitative characterization of the ligand/peptide binding thermodynamics and kinetics. Taken together, GaMD and its innovative variants are applicable to simulate a wide variety of biomolecular dynamics, including protein folding, conformational changes and allostery, ligand binding, peptide binding, protein-protein/nucleic acid/carbohydrate interactions, and carbohydrate/nucleic acid interactions. In this review, we present principles of the GaMD algorithms and recent applications in biomolecular simulations and drug design.

Graphical Abstract



Gaussian accelerated molecular dynamics (GaMD) and its applications to a wide range of biological systems.

1. Introduction

Biological processes are mediated by biomolecules such as proteins, nucleic acids, lipids, and carbohydrates. Biomolecules often visit different functional conformations during

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

various biological functions, including cellular signaling, protein folding, gene translation/editing, and biomolecular recognition[1–4]. The underlying free energy landscapes of biomolecules determine their conformations[5, 6]. Molecular dynamics (MD) is an advanced technique that allows us to simulate biomolecular dynamics at an atomistic level[7]. It is now possible to run longer and cheaper MD simulations with remarkable advances in computing hardware (e.g., the Anton supercomputer and GPUs) and software developments[8]. Even so, conventional MD (cMD) is often limited to typically hundreds of nanoseconds to tens of microseconds[9–12]. On the other hand, many biological processes of interest take place over milliseconds or even longer timescales, due to high energy barriers (e.g., 8–12 kcal/mol)[1, 13, 14]. Due to this gap, it remains challenging to sufficiently sample different conformations and accurately calculate free energy profiles of biomolecules through cMD simulations.

To overcome the above challenges, numerous enhanced sampling techniques have been introduced since the dawn of MD as reviewed in a number of previous articles[15–18]. One class of these methods use predefined collective variables (CVs) or reaction coordinates, including umbrella sampling (US)[19, 20], metadynamics[21, 22], adaptive biasing force (ABF)[23, 24], steered MD (SMD)[25], conformational flooding[26, 27], and so on. Typical CVs include root-mean square deviation (RMSD) relative to a reference conformation, dihedrals, atom distances, eigenvectors of principal component analysis (PCA)[27], etc. These methods greatly improve the sampling of biomolecular dynamics and the accuracy of free energy calculations along the chosen CVs. However, it is rather challenging to define proper CVs *in prior* because the system needs to be studied in detail beforehand. Furthermore, the predefined CVs could largely limit sampling of the conformational space during the biasing simulations. This usually slows convergence of the simulations and suffers from the “hidden energy barrier” problem once crucial CVs are missing in the simulation setup[22].

Another kind of enhanced sampling techniques have been introduced without using predefined CVs, including replica exchange molecular dynamics (REMD)[28, 29] or parallel tempering[30], self-guided Langevin or molecular dynamics[31–34], essential energy space random walk[35–37] and accelerated molecular dynamics (aMD)[38]. In particular, Voter introduced aMD by adding a boost potential in non-barrier regions to accelerate infrequent transitions in solids[39]. Hamelberg et al. further developed this technique to perform biomolecular simulations[38]. The boost potential in aMD enables simulations to sample different low-energy conformational states by smoothing the system potential energy surface and reducing the energy barriers[38, 40]. Despite the advantage of unconstrained enhanced sampling, aMD can suffer from high statistical noise, affecting the description of the correct statistical ensemble[41]. In detail, the ensemble canonical average is reached by reweighting each point in the configuration space on the modified potential by the strength of the Boltzmann factor of the bias energy, at that particular point. By using the early aMD method, this has shown to lead to high statistical noise, since the points with the largest biases dominate the reweighted result[42–45]. In comparison with the CV-biasing methods, aMD has typically much higher boost potential with wider distributions (tens to hundreds of kcal/mol)[43], making it very challenging to accurately reweight free energies from aMD simulations, especially for biological macromolecules[44, 46, 47]. This issue can be severe for large biomolecular systems, such as transmembrane proteins and ribonucleoproteins, where

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the standard reweighting procedure has often been prohibitive, given the large statistical noise[43].

Gaussian accelerated molecular dynamics (GaMD) has been developed to smooth the surface of potential energy with a harmonic boost potential, following three newly formulated enhanced sampling principles[48]. Similar to the previous aMD, no predefined CV is needed for GaMD simulations. Furthermore, the new harmonic boost potential in GaMD exhibits a Gaussian distribution, which enables us to accurately recover the original biomolecular free energy landscapes by Gaussian approximation, *i.e.*, cumulant expansion to the second order. This useful scheme substantially reduces the statistical noise, thereby overcoming the limitations of the early aMD methodology (*vide supra*). Therefore, GaMD simultaneously enables enhanced sampling without any constraints and accurately calculates free energy landscapes of biomolecules. As previously reported[49, 50], GaMD has been successfully applied to simulate ligand binding[48, 51, 52], protein folding[48, 52], activation of G-protein-coupled receptors (GPCRs)[51], human dystonia related protein[53], ion channels[54], agonist and antagonist binding in the μ-OR[55, 56], virus enzymes[57, 58], bacterial effector proteins[59], etc.

In addition, GaMD has been combined with REMD to further improve conformational sampling and free energy calculations[60, 61]. More recently developed "selective GaMD" algorithms, including ligand GaMD (LiGaMD)[62] and peptide GaMD (Pep-GaMD)[63], have enabled unprecedented microsecond simulations to capture repetitive binding and dissociation of small-molecule ligands and highly flexible peptides. Accurate ligand/peptide binding free energies and kinetic rate constants are thus calculated through the selective GaMD simulations.

In this review, we will present the principles and the most recent applications of GaMD. Robust GaMD has been established for advanced simulation studies of a wide range of biomolecular systems, especially the protein-nucleic acid interactions[64–66] such as the CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 gene editing system[67, 68], protein-protein/peptide interactions[69–73], protein-ligand binding[56, 57, 74–77], protein folding[78], protein enzymes[53, 59, 79–93], membrane proteins (including GPCRs[69, 70, 74, 77, 94–96], ion channels[54, 97] and γ-secretase[98]) and carbohydrates[99–101], as well as drug design[102, 103].

2. Theory

2.1. Gaussian Accelerated Molecular Dynamics (GaMD)

A harmonic boost potential is added in GaMD to smooth the system potential energy surface and enhance the conformational sampling of biomolecules (Fig. 1)[48]. Consider a system with N atoms at positions $\vec{r} = \{\vec{r}_1, \cdots \vec{r}_N\}$, when the system potential $V(\vec{r})$ is lower than a threshold energy E, a boost potential is added as:

$$\Delta V(\vec{r}) = \frac{1}{2}k(E - V(\vec{r}))^2, \quad V(\vec{r}) < E, \tag{1}$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

where k is the harmonic force constant. The modified system potential, $V^*(\vec{r}) = V(\vec{r}) + \Delta V(\vec{r})$ is given by:

$$V^*(\vec{r}) = V(\vec{r}) + \frac{1}{2}k(E - V(\vec{r}))^2, \quad V(\vec{r}) < E\,. \tag{2}$$

Otherwise, when the system potential is above the threshold energy, i.e., $V(\vec{r}) \geq E$, the boost potential is set to zero and $V^*(\vec{r}) = V(\vec{r})$.

Three enhanced sampling principles are applied to the boost potential in GaMD to smooth the potential energy surface. First, for any two arbitrary potential values $V_1(\vec{r})$ and $V_2(\vec{r})$ found on the original energy surface, if $V_1(\vec{r}) < V_2(\vec{r})$, ΔV should be a monotonic function that does not change the relative order of the biased potential values, i.e., $V_1^*(\vec{r}) < V_2^*(\vec{r})$. By replacing $V^*(\vec{r})$ with Equation (2) and isolating E, we then obtain:

$$E < \frac{1}{2}[V_1(\vec{r}) + V_2(\vec{r})] + \frac{1}{k}\,. \tag{3}$$

Second, if $V_1(\vec{r}) < V_2(\vec{r})$, the potential difference observed on the smoothened energy surface should be smaller than that of the original, i.e., $V_2^*(\vec{r}) - V_1^*(\vec{r}) < V_2(\vec{r}) - V_1(\vec{r})$. Similarly, by replacing $V^*(\vec{r})$ with Equation (2), we can derive:

$$E > \frac{1}{2}[V_1(\vec{r}) + V_2(\vec{r})]\,. \tag{4}$$

With $V_{\min} \leq V_1(\vec{r}) < V_2(\vec{r}) \leq V_{\max}$, we need to set the threshold energy E in the following range by combining Equations (3) and (4):

$$V_{\max} \leq E \leq V_{\min} + \frac{1}{k}, \tag{5}$$

where $V_{\min}$ and $V_{\max}$ are the system minimum and maximum potential energies. To ensure that Equation (5) is valid, $V_{\max} \leq V_{\min} + \frac{1}{k}$ and k have to satisfy:

$$k \leq \frac{1}{V_{\max} - V_{\min}}\,. \tag{6}$$

Let us define $k \equiv k_0 \bullet \frac{1}{V_{\max} - V_{\min}}$, then $0 < k_0 \leq 1$. As illustrated in Fig. 1, k_0 determines the magnitude of the applied boost potential. With higher k_0, larger boost potential is added to the potential energy surface, which facilitates enhanced sampling of biomolecules across decreased energy barriers.

Third, in order to ensure accurate reweighting using cumulant expansion to the second order[43], the standard deviation of ΔV needs to be small enough (i.e., narrow distribution):

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\sigma_{\Delta V} = \sqrt{\left(\frac{\partial \Delta V}{\partial V}\bigg|_{V = V_{avg}}\right)^2 \sigma_V^2} = k\left(E - V_{avg}\right)\sigma_V \leq \sigma_0, \tag{7}$$

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., $10k_B$T) for accurate reweighting.

Provided Equation (5) that gives the range of threshold energy E, when E is set to the lower bound $E = V_{\max}$, we substitute in E and k, and obtain:

$$k_0 \leq \frac{\sigma_0}{\sigma_V} \bullet \frac{V_{\max} - V_{\min}}{V_{\max} - V_{avg}}. \tag{8}$$

Let us define the right-hand side in Equation (8) as $k_0' = \frac{\sigma_0}{\sigma_V} \bullet \frac{V_{\max} - V_{\min}}{V_{\max} - V_{avg}}$. For efficient enhanced sampling with the highest possible acceleration, k_0 can then be set to its upper bound as:

$$k_0 = \min(1.0, k_0') = \min\left(1.0, \frac{\sigma_0}{\sigma_V} \bullet \frac{V_{\max} - V_{\min}}{V_{\max} - V_{avg}}\right). \tag{9}$$

The larger $\sigma_{\Delta V}$ is obtained from the original potential energy surface (particularly for large biomolecules), the smaller k_0 may be applicable to allow for accurate reweighting. Alternatively, when the threshold energy E is set to its upper bound $E = V_{\min} + \frac{1}{k}$ according to Equation (5), we substitute in E and k in Equation (7) and obtain:

$$k_0 \geq \left(1 - \frac{\sigma_0}{\sigma_V}\right) \bullet \frac{V_{\max} - V_{\min}}{V_{avg} - V_{\min}}. \tag{10}$$

Let us define the right-hand side in Equation (10) as $k_0'' \equiv \left(1 - \frac{\sigma_0}{\sigma_V}\right) \bullet \frac{V_{\max} - V_{\min}}{V_{avg} - V_{\min}}$. Note that a smaller k_0 will give higher threshold energy E, but smaller force constant k. When $0 < k_0'' \leq 1$, k_0 can be set to either k_0'' for the highest threshold energy E or its upper bound 1.0 for the greatest force constant k. In this regard, $k_0 = k_0''$ is applied in the current GaMD. Otherwise, k_0 is calculated using Equation (9).

Given E and k_0, we can calculate the boost potential as:

$$\Delta V(\vec{r}) = \frac{1}{2} k_0 \frac{1}{V_{\max} - V_{\min}} (E - V(\vec{r}))^2, \quad V(\vec{r}) < E. \tag{11}$$

GaMD provides different options to add only the total potential boost ΔV_P, only dihedral potential boost ΔV_D, or the dual potential boost (both ΔV_P and ΔV_D). The dual-boost GaMD generally provides higher acceleration than the other two types of simulations for enhanced sampling[40]. The simulation parameters comprise of the threshold energy values and the effective harmonic force constants, k_{0P} and k_{0D} for the total and dihedral potential boost, respectively.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2.2. Energetic Reweighting of GaMD for Free Energy Calculations

For simulations of a biomolecular system, the probability distribution along a selected reaction coordinate $A(\boldsymbol{r})$ is written as $p^*(A)$, where $\boldsymbol{r}$ denotes the atomic positions $\{\boldsymbol{r}_1,\ldots,\boldsymbol{r}_N\}$. Given the boost potential $\Delta V(\boldsymbol{r})$ of each frame, $p^*(A)$ can be reweighted to recover the canonical ensemble distribution, $p(A)$, as:

$$p(A_j) = p^*(A_j)\frac{\left\langle e^{\beta\Delta V(\boldsymbol{r})}\right\rangle_j}{\sum_{i=1}^{M}\left\langle p^*(A_i)e^{\beta\Delta V(\boldsymbol{r})}\right\rangle_i}, j = 1, \cdots, M, \tag{12}$$

where M is the number of bins, $\beta = k_B T$ and $\left\langle e^{\beta\Delta V(\boldsymbol{r})}\right\rangle_j$ is the ensemble-averaged Boltzmann factor of $\Delta V(\boldsymbol{r})$ for simulation frames found in the jth bin. In order to reduce the energetic noise, the ensemble-averaged reweighting factor can be approximated using cumulant expansion[104, 105]:

$$e^{\beta\Delta V} = \exp\left\{\sum_{k=1}^{\infty}\frac{\beta^k}{k!}C_k\right\}, \tag{13}$$

where the first three cumulants are given by:

$$\begin{aligned} C_1 &= \langle\Delta V\rangle, \\ C_2 &= \left\langle\Delta V^2\right\rangle - \langle\Delta V\rangle^2 = \sigma_{\Delta V}^2, \\ C_3 &= \left\langle\Delta V^3\right\rangle - 3\left\langle\Delta V^2\right\rangle\langle\Delta V\rangle + 2\langle\Delta V\rangle^3. \end{aligned} \tag{14}$$

When the boost potential follows near-Gaussian distribution, cumulant expansion to the second order (or "Gaussian Approximation") provides the accurate approximation for free energy calculations[43]. The reweighted free energy $F(A) = -k_B T\ln p(A)$ is calculated as:

$$F(A) = F^*(A) - \frac{1}{\beta}\sum_{k=1}^{2}\frac{\beta^k}{k!}C_k + F_c, \tag{15}$$

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

To characterize the extent to which ΔV follows a Gaussian distribution, its distribution anharmonicity γ is calculated as[43]:

$$\gamma = S_{\max} - S_{\Delta V} = \frac{1}{2}\ln\left(2\pi e\sigma_{\Delta V}^2\right) + \int_0^{\infty} p(\Delta V)\ln\left(p(\Delta V)\right)d\Delta V, \tag{16}$$

where ΔV is dimensionless as divided by $k_B T$ with k_B and T being the Boltzmann constant and system temperature, respectively, and $S_{\max} = \frac{1}{2}\ln\left(2\pi e\sigma_{\Delta V}^2\right)$ is the maximum entropy of ΔV[43]. When γ is zero, ΔV follows exact Gaussian distribution with sufficient sampling. Reweighting by approximating the exponential average term with cumulant expansion to the second order is able to accurately recover the original free energy landscape. As γ increases, the ΔV distribution becomes less harmonic and the reweighted free energy profile

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

obtained from cumulant expansion to the second order would deviate from the original. The anharmonicity of ΔV distribution serves as an indicator of the enhanced sampling convergence and accuracy of the reweighted free energy. Nevertheless, with the GaMD theoretical framework, the GaMD boost potential does not change shape of the biomolecular overall energy landscape. A near Gaussian distribution is achieved for the GaMD boost potential. A toolkit of Python scripts for GaMD/aMD reweighting "PyReweighting"[43] is developed and distributed free of charge at http://miao.compbio.ku.edu/PyReweighting/.

2.3. Replica Exchange-GaMD

Replica exchange and GaMD have been combined in a rex-GaMD approach to further improve the sampling and free energy calculations of biomolecules[60]. According to Equation (11), both the threshold energy E and the effective force constant k_0 could adjust the boost potential. Therefore, two versions of rex-GaMD were proposed: force constant rex-GaMD and threshold energy rex-GaMD. During simulations of force constant rex-GaMD, the boost potential can be exchanged between replicas, in which the threshold energy is fixed and harmonic force constants are different. Whereas the algorithm of threshold energy rex-GaMD tends to switch the threshold energy between lower and upper bounds for generating different levels of boost potential.

The rex-GaMD simulations allow replicas exchanged between each pair of neighboring σ_{0P} or threshold energy based on the probability that meets the Metropolis criterion. In the simulation system, each state x can be weighted by the Boltzmann factor,

$$W_B(x) = \exp\left(\frac{1}{k_B T} H(x)\right), \tag{17}$$

where k_B is the Boltzmann constant, T is the system temperature and $H(x)$ is the system Hamiltonian. The weight factor for the state X here is given by the product of the Boltzmann factor of each replica:

$$W_{RE}(X) = \exp\left(-\sum_{i=1}^{N} \frac{1}{k_B T} H(x_i)\right), \tag{18}$$

where N is the number of total states. Thus, the replica exchange probability can be written as $w(X_i \rightarrow X_j)$, which needs to meet the Metropolis criterion to calculate the exchange probability:

$$w(X_i \rightarrow X_j) = \min(1.0, e^{\Delta}), \tag{19}$$

where X_i and X_j are the states of the two nearby replicas, and $\Delta = \frac{1}{k_B T}\left(V_i^* - V_j^*\right)$ and V^*_i and V^*_j are the total modified system potential energies calculated from the last conformation of the GaMD simulations at replica i and j. These exchange processes will keep repeating until the end of the simulation. The rex-GaMD simulations were tested on three model systems, including the alanine dipeptide, chignolin, and HIV protease, demonstrating that the distribution width of the boost potential is narrowed down, and the system conformational space is enhanced sampled.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Recently, Sugita et al.[61] proposed another approach (GaREUS) that combined GaMD with replica exchange umbrella sampling (REUS). GaREUS was successfully demonstrated on accurate calculations of free energy landscapes underlying the N-glycan equilibration, conformational change of adenylate kinase and chignolin folding. The computational resource for GaREUS was the same as that required for REUS, while the sampling in GaREUS was more efficient than REUS or GaMD.

2.4. Ligand Gaussian Accelerated Molecular Dynamics (LiGaMD)

Based on GaMD, LiGaMD[62] has been proposed to more efficiently simulate both binding and dissociation of small-molecule ligands for calculating the ligand binding free energies and kinetics. For such simulations, the system contains ligand L, protein P and the 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:

$$H(r, p) = K(p) + V(r), \tag{20}$$

where $K(p)$ and $V(r)$ are the system kinetic and total potential energies, respectively. Then, the potential energy could be decomposed into the following terms:

$$\begin{aligned} V(r) = {} & V_{P,b}(r_P) + V_{L,b}(r_L) + V_{E,b}(r_E) + V_{PP,nb}(r_P) + V_{LL,nb}(r_L) \\ & + V_{EE,nb}(r_E) + V_{PL,nb}(r_{PL}) + V_{PE,nb}(r_{PE}) + V_{LE,nb}(r_{LE}). \end{aligned} \tag{21}$$

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 molecular mechanics force fields,[106, 107] the non-bonded potential energies are usually calculated as:

$$V_{nb} = V_{elec} + V_{vdW}, \tag{22}$$

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, $V_{L,nb}(r) = V_{LL,nb}(r_L) + V_{PL,nb}(r_{PL}) + V_{LE,nb}(r_{LE})$. Therefore, we add a boost potential selectively to the ligand non-bonded potential energy according to the GaMD algorithm:

$$\Delta V_{L,nb}(r) = \begin{cases} \frac{1}{2} k_{L,nb}\left(E_{L,nb} - V_{L,nb}(r)\right)^2, & V_{L,nb}(r) < E_{L,nb} \\ 0, & V_{L,nb}(r) \geq E_{L,nb} \end{cases} \tag{23}$$

where $E_{L,nb}$ is the threshold energy for applying boost potential and $k_{L,nb}$ is the harmonic constant. These parameters in LiGaMD are derived similarly as in the GaMD algorithm.

Next, one can add multiple ligand molecules in the solvent to facilitate ligand binding to proteins in MD simulations. This is based on the fact that the average ligand unbound time

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

τ_U is inversely proportional to the ligand concentration $[L]$, i.e., $\tau_U = \frac{1}{k_{on} \bullet [L]}$ with k_{on} being the ligand binding rate constant. 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, another boost potential could thus be applied on the unbound ligand molecules, protein, and solvent to facilitate both ligand dissociation and 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:

$$\Delta V_D(r) = \begin{cases} \frac{1}{2} k_D (E_D - V_D(r))^2, & V_D(r) < E_D \\ 0, & V_D(r) \geq E_D \end{cases} \tag{24}$$

where V_D is the total system potential energy other than the non-bonded potential energy of the bound ligand, E_D is the corresponding threshold energy for applying the second boost potential and k_D is the harmonic constant. This leads to dual-boost LiGaMD (LiGaMD_Dual) with the total boost potential $\Delta V(r) = \Delta V_{L,nb}(r) + \Delta V_D(r)$.

2.5. Peptide Gaussian Accelerated Molecular Dynamics (Pep-GaMD)

Large conformational changes of peptides often occur via peptides binding to the target proteins, being distinct from small-molecule ligand binding or protein-protein interactions (PPIs). We have developed another algorithm called peptide GaMD or “Pep-GaMD” that enhances sampling of peptide-protein interactions[63].

In Pep-GaMD, we consider a system of ligand peptide L binding to a target protein P in a biological environment E. We decompose the potential energy into similar terms as in Equation (21). Presumably, peptide binding mainly involves in both the bonded and non-bonded interaction energies of the peptide since peptides often undergo large conformational changes during binding to the target proteins. Thus, the essential peptide potential energy is $V_L(r) = V_{LL,b}(r_L) + V_{LL,nb}(r_L) + V_{PL,nb}(r_{PL}) + V_{PL,nb}(r_{LE})$. In Pep-GaMD, we add boost potential selectively to the essential peptide potential energy according to the GaMD algorithm:

$$\Delta V_L(r) = \begin{cases} \frac{1}{2} k_L (E_L - V_L(r))^2, & V_L(r) < E_L \\ 0, & V_L(r) \geq E_L \end{cases} \tag{25}$$

where E_L is the threshold energy for applying boost potential and k_L is the harmonic constant.

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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\Delta V_D(r) = \begin{cases} \frac{1}{2}k_D(E_D - V_D(r))^2, & V_D(r) < E_D \\ 0, & V_D(r) \geq E_D \end{cases} \quad (26)$$

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

3. Applications

Without the need to set predefined reaction coordinates or CVs, GaMD enables a wide range of applications in enhanced sampling of biomolecules. Furthermore, accurate reweighting using cumulant expansion to the 2nd order could be achieved in GaMD simulations because the boost potential exhibits a Gaussian distribution, allowing recovery of the original free energy landscapes even for large biomolecules[48, 51, 52]. Depending on the system size, orders of magnitude speedup for biomolecular simulations could be achieved in GaMD. As demonstrated on alanine dipeptide, GaMD simulations achieved ~36–67 times speedup for sampling of the backbone dihedral transitions compared with the long cMD simulations[108]. Higher acceleration could be potentially achieved for larger systems with greater boost potential applied in the GaMD simulations. Hundreds-of-nanosecond to microsecond GaMD simulations could capture millisecond timescale events. Here, we summarize recent application studies of GaMD.

3.1 Protein-nucleic acid interactions

CRISPR-Cas9 system is a bacterial immune system that has introduced a powerful genome editing technology, which has revolutionized life sciences[109]. At the molecular level, CRISPR-Cas9 is a protein/nucleic acid complex, composed of the Cas9 protein associated with a guide RNA and matching sequences of DNA[110]. Cas9 site-specifically recognizes the DNA by binding its Protospacer-Adjacent Motif (PAM), a short trinucleotide that enables the selection of the DNA across the genome. Upon PAM binding, the DNA binds Cas9 by matching the RNA with one strand (the target strand, TS), such forming an RNA:DNA heteroduplex structure. The second non-target strand (NTS) of the DNA gets displaced and also accommodated within the protein. Structures of the *Streptococcus Pyogenes* Cas9 (SpCas9) revealed a bilobed architecture (Fig. 2A). One lobe – viz., the recognition lobe (REC) – includes three regions that mediate nucleic acid binding (REC1–3), while the second is the nuclease lobe (NUC)[111–113]. The latter comprises two catalytic domains, HNH and RuvC, which cleave the DNA TS and NTS, respectively. X-ray crystallography and cryo-EM studies portrayed the structure of SpCas9 in different states, as *apo* protein[113], in complex with RNA[114] and upon DNA binding[111, 112]. These structural studies have been a stepping stone to understand the mechanism of action of CRISPR-Cas9. However, although critical, this information could not access the dynamics and the complex conformational transitions of this ribonucleoprotein, raising fundamental questions on the system's biological function. In this regard, we have successfully applied GaMD to decipher

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the molecular mechanism of nucleic acid processing and selectivity of this genome editing tool.

3.1.1 Conformational changes underlying RNA binding to CRISPR-Cas9— Based on structural data, large structural transitions of the protein have been hypothesized to enable RNA binding[113, 114]. To characterize this process, we applied a GaMD in combination with Targeted MD (TMD) approach, which reduces the RMSD between an initial and final target conformations[115]. By using TMD, we obtained an initial pathway of the conformational change from the *apo* protein to the RNA-bound form. We observed that the REC1–3 regions of the protein moved in opposite directions relative to each other, leading to the closure of the REC lobe to accommodate RNA. This observation agreed well with previous hypotheses based on cryo-EM[113, 114]. Then, we used GaMD to precisely describe the energetic landscape associated to this conformational change (Fig. 2B). The free energy landscape described three local minima: M1 corresponds to the crystallographic *apo* structure, M2 is the RNA-bound structure, while M3 is an intermediate state, characterized by the solvent exposure of an arginine-rich helix. The latter directly binds the RNA guide in both RNA-bound and DNA-bound structures of Cas9[114], suggesting a mechanism for the recruitment of RNA, in which the electrostatics could play a key role. By further computing the ensemble averaged electrostatic potential, we found that a positively charged cavity is formed at the level of the arginine-rich helix and is suitable for RNA binding (Fig. 2B). Overall, these simulations indicated that the arginine helix is critical for the recruitment of RNA, and that the formation of positively charged cavity allows for the formation of the Cas9:RNA binary complex.

3.1.2 Conformational activation of the Cas9 protein for DNA cleavage—The process of conformational activation of the Cas9 protein toward DNA cleavages involves a critical transition of the catalytic HNH domain. The latter undergoes a structural change from an inactive form to the active state prone to perform DNA cleavages[111, 112, 116]. Our first study based on cMD highlighted a “striking plasticity” of the catalytic HNH nuclease[117]. That study revealed the critical dynamic interplay between the DNA NTS and HNH, suggesting that the binding of the NTS would allow increased dynamics of HNH and, thereby, its activation toward DNA cleavage. After this first computational study, single-molecule Förster Resonance Energy Transfer (FRET) experiments have investigated the large scale dynamics of the system, showing that the dynamical docking of HNH at the cleavage site critically requires the presence of the NTS,[118] and thereby confirming the predictions of molecular simulations. Overall, that early MD study has been instrumental in characterizing the atomic level details of the CRISPR-Cas9 dynamics. Nevertheless, considering that cMD simulations are limited to short timescales (i.e., ns–to–μs), that investigation could not fully address the activation mechanism of the catalytic HNH domain. To overcome the inherent timescale limits of cMD and characterize the HNH activation process, we performed GaMD simulations, capturing multiple states of the HNH conformational landscape[115]. As expected, the simulations broadly sampled various possible configurations of the HNH domain. Energetic reweighting of the conformational landscape revealed that the energetic minima identified through GaMD correspond to the conformational states found by FRET experiments[112, 114, 116, 119] and structural studies[111]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(Fig. 3A). Notably, this extensive GaMD sampling (collecting >20 μs) identified a "bona-fide" conformation of the active state (*viz.*, active #4 in Fig. 3A), which was shown to be thermodynamically stable. This conformation predicted the active state two years before structural data was made available[111], allowing also to start in-depth studies of the catalysis though hybrid quantum-classical methods[120, 121]. Moreover, to better describe the conformational change from the pre-active state (captured in PDB ID: 5f9r) to the active configuration where HNH catalytic residue (H983) docks at the DNA TS (Fig. 3B, **top panel**), the specialized supercomputer Anton-2[122] has been used to carry out continuous multi-μs cMD simulations[123]. These simulations captured the late step of HNH activation over ~16 μs of continuous simulations (Fig. 3B, **bottom panel**). The dynamical docking of HNH at the cleavage site on the TS occurred by following the same pathway previously observed over multiple GaMD replicas (Fig. 3B, **central panel**). Indeed, while the continuous simulation performed on Anton-2 recovered the transition over ~16 μs, GaMD captured the conformational change by running ~400 ns and in three replicas. This finding indicates that GaMD reliably captures structural transitions of biomolecules that occur over longer time scales. Finally, it is notable that the activated state, which was early identified though GaMD[115] and later refined using Anton-2 simulations[123], resulted in notable agreement with the cryo-EM structure of the active complex[111]. This showed the reliability of the early predictions that have been obtained based on GaMD.

3.1.3 Molecular mechanism of off-target effects of CRISPR-Cas9—An important mechanistic question relates to the onset of off-target effects, which arise from the binding of DNA sequences that do not fully match the guide RNA, resulting in RNA:DNA hybrids containing mismatched pairs. Off-target effects result in cleavages at DNA sites, representing a limitation for the application of CRISPR-Cas9 for *in vivo* and *ex vivo* genome editing. Kinetic and single-molecule FRET studies provided critical hints on the molecular basis of off-target effects. Indeed, it has been shown that DNAs containing one to three mismatches located at the RNA:DNA hybrid ends result in a flexible and catalytically active HNH domain[118, 124, 125]. Contrariwise, four (or more) mismatches result in decreased flexibility of HNH and in its catalytic inactivation. The single-molecule experiments, however, could not explain how a different number of DNA mismatches at the RNA:DNA hybrid ends could affect the activation of HNH. Knowing the molecular basis of this mechanism is of critical importance, as it could help in developing more specific CRISPR-Cas9 systems, in which a single base pair mismatch is sufficient for reducing the HNH dynamics and catalytic function, thereby inhibiting the cleavage of incorrect DNA sequences. Considering that GaMD has been successful in describing the activation mechanism of HNH[115] (Fig. 3), we employed the method to investigate the effect of base pair mismatches on its conformational dynamics[68, 126]. For this application, we employed GaMD without carrying out energetic reweighting, increasing the sampling of low-energy states and providing a semi-quantitative ranking of the associated probabilities. This enabled us to broadly explore the system's conformational dynamics in the presence of base pair mismatches. The simulations revealed that four or more mismatches induce a broad opening of the RNA:DNA hybrid (Fig. 4A), which results in newly formed interactions between the TS and the L2 loop. These interactions importantly reduce the HNH flexibility, hampering its conformational activation. On the other hand, one to three

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

base pair mismatches do not result in sensible openings of the heteroduplex, as evinced by the minor groove width (measured at position 17, Fig. 4B), resulting in a negligible effect on the HNH conformational dynamics and thereby not affecting its activation for cleavage. Overall, the simulations could discriminate the different effects of base pair mismatches on the HNH activation, providing a mechanistic rationale to previous kinetic and single-molecule experiments. Building on the outcomes of GaMD simulations, we suggested that altering the TS-L2 interactions could reduce off-target binding. This speculation has been supported by the experimental engineering of the L2 loop in several variants of the Cas9 enzyme[125, 127, 128], which increase the system's specificity toward on-target sequences.

3.1.4 Allosteric effects across the CRISPR-Cas9 complex—Multiple evidences including experiments and computations have indicated that CRISPR-Cas9 is also an intriguing "allosteric engine"[125, 129–131]. Indeed, CRISPR-Cas9 requires an intricate allosteric activation to accomplish DNA cleavages. Biochemical experiments have indicated that the central element of the CRISPR-Cas9 allosteric signaling is the HNH domain, since its high flexibility can allow the signal transmission. To describe the allosteric signaling across HNH and how it transfers the information of DNA binding (occurring within the REC lobe) to the catalytic sites for cleavage, GaMD was combined with graph theory[132]. This combination allowed inclusion of long timescale motions in the calculation of the allosteric pathways[67]. Specifically, while GaMD characterized the long timescale system's dynamics, network models derived from graph theory accurately described the allosteric network and information transfer. This approach revealed the existence of a millisecond timescale dynamic pathway across HNH, which connects the RuvC nuclease domain to the recognition lobe REC. This allosteric route was validated through NMR relaxation experiments, showing that a contiguous pathway of slow residues overlaps with the prediction from GaMD and graph theory-based analysis. In summary, the combination of GaMD simulations with graph theory provided a useful approach for determining the signal transduction in CRISPR-Cas9, laying the foundations for characterizing allostery in other protein/nucleic acid complexes whose biological function relies on slow dynamical motions associated to the (re)organization of protein domains and long-range effects.

3.2 Protein-protein/peptide interactions

Protein–protein interactions (PPIs) and protein-peptide interactions are central to biological functions and have thus been targeted to design novel therapeutic drugs[133–137]. Here, we will summarize recent GaMD applications in simulation studies of PPIs and protein-peptide interactions.

3.2.1 Protein-protein interactions—The recognition of T cell receptor (TCR) and peptides presented by major histocompatibility molecules (pMHC) initiates adaptive immune responses. The pMHC binding affinity often correlates with the TCR-signaling strength. However, frequent high-affinity of pMHC in the human T cell repertoire are not stimulatory. Recently, enhanced sampling methods including GaMD, ABF and SMD were performed to distinguish stimulators from non-stimulatory ligands by simulating the TCR-pMHC disengagement[71]. The GaMD was first performed to reveal the structural flexibility of the complex and identify important CVs, including the orientation angle of the TCR about

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the pMHC assembly, salt bridges and hydrogen bonds. Then, the identified CVs were used for free energy calculation using the ABF method. Constant velocity SMD simulations were performed starting from the free energy minima identified by ABF. The simulations revealed that dynamic interactions in the TCR-pMHC interface play a critical role in determining the TCR specificity. One collective property of the entire TCR-pMHC interface is the formation of a catch or slip bond, being consistent with the results from single-molecule force measurements. In addition to simulations of PPIs with globular proteins such as the TCR-pMHC complex, GaMD have also been successfully applied to investigate PPIs with membrane proteins such as GPCR-G protein interactions[69, 70], which will be described in Section 3.5. In summary, GaMD is suitable to study large biomolecular complexes and provide important insights into functionally important PPIs.

3.2.2 Protein-peptide interactions—Petrizzelli et. al.[73] applied GaMD to investigate the pathogenic mechanisms caused by missense mutations of KDM6A on the histone H3, including P941S, D980V, S1025G, H1060L, L1200F, G1223D, Q1248R and R1255W. GaMD simulations showed that the interaction between the linker and JmjC domains was significantly impacted by residue mutations, leading to a loss of function. All mutants exhibited movements of the disordered linker domain, leading to increased flexibility of the KDM6A-H3 complex, which induced wrong exposure and orientation of the trimethylated lysine in the catalytic site. Therefore, GaMD simulations revealed important pathogenic mechanisms of the KDM6A-H3 interaction[73].

We developed a novel approach, namely *PeptiDock*+GaMD, in which the global peptide docking-*ClusPro PeptiDock* and GaMD simulations were combined for improving modeling of protein-peptide interactions[138]. For three model peptides (peptide 1–3), docking models generated with *PeptiDock*[139] showed 3.3 Å, 3.5 Å and 4.8 Å RMSD of the peptide backbone relative to their experimental structures. The peptide docking poses were refined by GaMD simulations. Then, the *PyReweighting* toolkit[43] was applied to reweight and calculate free energies of the peptide structural clusters obtained from GaMD simulations. RMSDs of Peptides 1 and 2 in the 1st top-ranked cluster were 0.9 Å and 0.6 Å, respectively. The 3rd top-ranked cluster in Peptide 3 exhibited the smallest RMSD of 2.7 Å. Thus, the *PeptiDock*+GaMD could be used to accurately predict the peptide-protein interaction. In comparison, cMD simulations with same simulation time were much less efficient in refining the peptide docking poses.[139] Only 1 among 4 cMD simulations of Peptide 2 improved the peptide binding pose. RMSD decrease was not observed in any of cMD simulations of Peptide 3. The top-ranked models obtained by clustering of cMD snapshots were of high quality for only Peptide 1 but medium for both Peptides 2 and 3. Therefore, GaMD simulations refined peptide docking poses and provided significantly improved sampling than cMD.

The *PeptiDock*+GaMD approach was further applied to model interactions of cyclic peptides with proteins, including peptide binding to MDM2/MDMX[140] and Tsg101 UEV protein[72]. Compared with the linear peptides, cyclic peptides often possess longer lifetime and better biological activity. To facilitate cyclic peptide design, the *PeptiDock*+GaMD approach was applied to investigate binding interactions between the UEV domain protein and three cyclic peptides[72]. The predicted peptide binding mode identified from GaMD

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

simulations was further validated by binding free energy calculations, which agreed well with the experimental binding affinities. Therefore, GaMD simulations provided important insights to protein-peptide interactions and were applicable to both linear and cyclic peptides.

3.2.3 Binding thermodynamics and kinetics of peptide—Pep-GaMD[63] has been developed to simulate both peptide binding and dissociation, which allows us to calculate the binding free energies and kinetics of flexible peptides. It has been demonstrated on binding of three model peptides to the SH3 domains[141, 142], which include "PPPVPPRR" (PDB: 1CKB), "PPPALPPKK" (PDB: 1CKA) and "PAMPAR" (PDB: 1SSH) (Figs. 5A–5C). Repetitive peptide binding and unbinding events were captured in independent 1 μs Pep-GaMD simulations, allowing us to calculate peptide binding thermodynamics and kinetics (Figs. 5D–5I). Peptide kinetics especially the dissociation rate was accelerated by ~3–4 orders of magnitude in the Pep-GaMD simulations. The predicted values from Pep-GaMD were in good agreement with available experimental data. Furthermore, the Pep-GaMD simulations revealed the important role of long-range electrostatics in peptide binding and the binding mainly followed a conformational selection model.

3.3 Protein-ligand binding

3.3.1 Protein-ligand interactions—Use of EGFR tyrosine kinase inhibitors is a promising approach to improve progression-free survival in cancer patients. Osimertinib was approved as a third generation EGFR mutant selective inhibitor. Unfortunately, resistances were detected against the osimertinib therapy. Brown et al.[80] identified G724S as an osimertinib resistance mutation. Microsecond GaMD simulations were performed on EGFR mutants in presence and absence of osimertinib to investigate the underlying mechanism. The GaMD simulations showed that the G724S mutation disrupts the osimertinib binding to the enzyme with exon 19 in-frame deletion (Ex19Del) mutation, while does not affect the enzyme with exon 21 missense mutation (L858R). The G724S mutation induces hyper stabilization of glycine-rich P-loop in β-bend conformation. It disrupts the interaction between indole ring of osimertinib and phenyl ring of F723. These results were further verified in animal cell culture experiments and in cancer patients. Overall, GaMD simulations elucidated the molecular mechanisms of ligand binding in EGFR mutations for treatment of non-small cell lung carcinomas, as well as many other protein-ligand interactions[56, 57, 74–77].

3.3.2 Ligand binding thermodynamics and kinetics characterized by LiGaMD —LiGaMD has been proposed to quantitatively characterize ligand binding thermodynamics and kinetics[62]. Host-guest and protein-ligand binding model systems have been used to validate the LiGaMD algorithm. Hundreds-of-nanosecond LiGaMD simulations captured repetitive guest binding and unbinding in the β-cyclodextrin host. The calculated guest binding free energies were in good agreement with experimental data, for which the errors were <1.0 kcal/mol. The sampling errors of LiGaMD simulations were < 1.0 kcal/mol in comparison with converged μs-timescale cMD simulations. Additionally, ligand kinetic rate constants were accurately predicted using Kramers' rate theory. Furthermore, repetitive binding and unbinding of the benzamidine inhibitor in trypsin was observed in 1 μs LiGaMD

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

simulations, allowing us to accurately calculate ligand binding free energy and kinetic rate constants. The ligand dissociation rate was remarkably accelerated by 7 orders of magnitude in the LiGaMD simulations. The predicted values were in excellent agreement with the experimental data[62].

3.4 Protein enzymes

3.4.1 Structural dynamics of protein kinases—Casein kinase 1δ (CK1δ) has been regarded as an important component in metazoan circadian rhythms regulation. Despite its importance, little was known about substrate selectivity and activity of the enzyme in molecular detail. Philpott et. al.[90] performed GaMD simulations on wildtype and tau mutant CK1δ systems and discovered a conformational switching mechanism of the activation loop. The switch regulates two different regions of the PER2 protein, which in turn regulates the protein stability and circadian timings in eukaryotes. The GaMD simulations further revealed that anion binding to a highly conserved site monitors the conformation in the activation loop and thereby regulating the overall conformation of the substrate binding cleft. The tau mutant, on the other hand, disrupts the allosteric regulation between the anionic sites. This disturbs the conformational flexibility of the activation loop and affects the stability of the PER2 protein. GaMD simulations thus provided molecular basis of the decreased activity in the tau mutant CK1δ.

Brassinosteroid insensitive 1-associated kinase 1 (BAK1) is an important receptor like kinase which initiates numerous immune and growth signaling pathways in plants. Moffett et al.[89] applied GaMD simulations to explore physiochemical basis of BAK1 activation through phosphorylation. GaMD simulations revealed the effects of various phosphorylation patterns and ATP binding on the enzyme conformation. GaMD simulations identified a metastable inactive enzyme conformation using activation-loop cracking. This activation loop conformation had been also found in other kinases like the ERK2. Phosphorylation of residues T450 and T455 played important roles in stabilizing the active-like activation loop without cracking. During the GaMD simulations, phosphorylation helped αC helix of the enzyme maintain its position near the N-lobe. In contrast, the αC helix of the unphosphorylated systems switched to an inactive state as the activation loop changed into a “cracked” conformation. Overall, GaMD simulations revealed the mechanism of phosphorylation controlled BAK1 activation. In another study, Koh et. al.[88] performed GaMD simulations for mechanistic insights into the flux-dependent transport signaling by Bce-like antibiotic resistance systems. They found that the transport activity is directly related to histidine kinase activity even with different antibiotic concentrations.

3.4.2 Active site dynamics of protein enzymes—The oncoprotein AlkB homolog 5 (Alkbh5) is involved in cancers such as leukemia, brain cancer and breast cancer. NMR experiments and GaMD simulations were combined to generate the structural model of the *apo* human Alkbh5[81]. The Alkbh5 active site was observed to be more disordered than that in the x-ray structure (PDB:4NJ4). It was likely due to the absence of the Cys^{230}–Cys^{267} disulfide bond in solution, which limited the protein conformational accessibility. GaMD simulations captured breathing motions of the protein, which expands the α-ketoglutarate binding pocket and permits binding of small molecules.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

FabA and FabZ are two *Escherichia coli* dehydratases involved in production of the unsaturated fatty acids (UFAs) from fatty acid biosynthesis[83]. Both FabA and FabZ are known to catalyze dehydration reactions, however, only FabA can further catalyze isomerization reaction. A combined approach involving chemical biology, structural biology and GaMD simulations was applied to understand the substrate selectivity and divergent activity of the two enzymes. Cross-linking experiments were performed to produce the acyl-AcpP•FabA and acyl-AcpP•FabZ complexes, which were used for GaMD simulations to elucidate the dehydration mechanism catalyzed by FabA and FabZ. GaMD simulations revealed dynamic mechanism of the unique isomerase activity of FabA and successfully differentiated the substrate preferences of FabA and FabZ. Moreover, GaMD simulations showed that only FabA selectively sampled the (−) *gauche* conformer of *trans*-2-decenoyl-AcpP for allylic rearrangement.

Furthermore, GaMD was successfully applied on simulations of protein enzymes for structure-based drug design of anti-malarial drugs,[82] design of inhibitors targeting *Staphylococcus aureus* enzyme MnaA,[85] investigation of soybean lecithin–gallic acid complex formation to aid in alcoholic liver disease (ALD),[87] usefulness of antioxidative agent for treating vascular endothelial deficits,[86] and understanding of drug resistance mechanism of rifampin[92]. Simulations using GaMD and replica exchange solute tempering (REST2)[143] were performed to understand mechanism of the transactivation of estrogen receptor[79].

3.4.3 Protein allostery—The interaction between HCV NS5A-D2 and human prolyl isomerase cyclophilin A (CypA) plays an essential role in viral RNA replication. Dujardin et al.[84] employed GaMD simulations and NMR to investigate the role of a short structural motif PW-turn (^{314}P*X*WA^{317}) on the structural disorder in NS5A-D2. There is a conformational equilibrium between folded and disordered states in the PW-turn motif, which is allosterically regulated by the *cis*/*trans* isomerization of 5 prolines residues (P306, P310, P315, P319 and P320). Moreover, the HCV RNA replication efficiency correlates well with the fraction of the structured PW-turn obtained from GaMD simulations.

Another study by Sztain et al.[91] employed GaMD to identify cryptic pockets of the SARS-CoV-2 main protease (Mpro), which are far away from the active site. Four systems including the monomer and dimer of Mpro in the absence and presence of the co-crystalized N3 inhibitor were built to perform GaMD simulations. Three regions including the distal allosteric site, active site and dimer interface region were identified as potential drug pockets using the PockDrug webserver.[144] Virtual screening against the above-mentioned pockets allowed to identify more hit molecules than using only the active site in the crystal structure. Furthermore, correlation analysis suggested that the three pockets could be allosterically regulated by each other. Therefore, the above identified pockets could be useful in virtual screening to identify novel inhibitors of SARS-CoV-2.

3.5 Membrane proteins

Membrane proteins including GPCRs, intramembrane proteases and ion channels play essential roles in cellular signaling and serve as important drug targets. Here, we will

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

summarize recent applications of GaMD in studies of GPCRs (including muscarinic acetylcholine, adenosine, opioid and chemokine receptors), γ-secretase, etc.

3.5.1 Binding mechanism of G protein mimetic nanobody to M_2 muscarinic GPCR—

GaMD simulations were performed to capture the spontaneous binding of nanobody Nb9–8, a G protein mimetic, to the M_2 muscarinic GPCR[145]. The agonist and nanobody in the X-ray structure of the active M_2 receptor were placed to be >20 Å far away from the receptor to build the starting model. Five independent GaMD simulations lasting ~4500 *ns* were performed. One GaMD simulation successfully captured the binding of the nanobody to the receptor G-protein coupling site of M_2 with a minimum RMSD of 2.48 Å in the nanobody core domain relative to the X-ray conformation, although the agonist still not reached its binding site (Figs. 6A and 6B). Both the orthosteric ligand-binding pocket and intracellular domains of the M2 receptor involved conformational change along with the binding of the nanobody (Fig. 6B). The orthosteric pocket in the X-ray structures of antagonist-bound and agonist nanobody-bound receptor are “open” and “closed”, respectively. Binding of the nanobody induced the orthosteric pocket from the “open” to “closed” state. Moreover, activation of the M_2 receptor was occurred during the binding of the nanobody, as measured by the distance between intracellular transmembrane helix 3 and 6 (TM3–TM6 distance) (Fig. 6B). Free energy profile of the nanobody RMSD relative to the 4MQS X-ray conformation and the receptor Arg121$^{3.50}$-Thr386$^{6.34}$ distance were calculated to characterize the nanobody binding pathways (Fig. 6C). Three low-energy conformational states including the unbound (“U”), intermediate 1 (“I1”), and intermediate 2 (“I2”) were identified from the potential of mean force (PMF) profile. The bound (“B”) conformation identified in the GaMD simulations is similar to that sampled in previous simulations of the 4MQS X-ray structure.[145] On the intracellular side, the nanobody core domain especially the β2, β3, β6, β7 and β8 strands overlapped well with the 4MQS X-ray structure when the nanobody RMSD decreased to 2.48 Å (Fig. 6D). Therefore, the GaMD provided important insights into the binding mechanism of the nanobody to the M_2 receptor.

3.5.2 Mechanism of specific G protein coupling to adenosine receptors—

There are four subtypes of adenosine receptors (A_1, A_{2A}, A_{2B}, and A_3) in human, which mediate the effects of adenosine (ADO). The odd ARs including A_1AR and A_3AR mainly couple to the $G_{i/o}$ proteins, while the even ARs (A_{2A}AR and A_{2B}AR) preferentially couple to the G_s proteins. In one of our recent studies[69], we employed GaMD simulations on four AR-G protein models, including the native structures of ADO-A_1AR-G_i with ADO and G_i protein bound[146] and the NECA-A_{2A}AR-G_s with 5’-N-ethylcarboxamidoadenosine (NECA) and an engineered G_s protein bound[147], as well as “decoy” complexes ADO-A_1AR-G_s and NECA-A_{2A}AR-G_i generated by switching the G proteins. In the ADO-A_1AR-G_i and NECA-A_{2A}AR-G_s complexes, GaMD identified only one stable low-energy conformation, which is similar to the cryo-EM structure (Figs. 7A and 7B). Similarly, only one low-energy conformation was identified in the NECA-A_{2A}AR-G_i complex, suggesting that agonist NECA binding in the A_{2A}AR could be still stabilized by coupling to the G_i protein (Fig. 7C). While the binding of G_s protein in the A_1AR led to increased fluctuations of both the receptor and ADO (Fig. 7D). The ADO agonist exhibited high fluctuations and sampled two different binding poses (“L1” and “L2”) in the ADO-A_1AR-G_s complex. In the “L2”

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

binding pose, ADO interacted with residues Tyr12$^{1.35}$ and Tyr271$^{7.36}$ in the sub-pocket 2 of the A_1AR, which is described earlier[148]. GaMD simulations indicated that coupling with the G_i protein was preferred to the G_s in the A_1AR (Fig. 7E), while both the G_s and G_i proteins could be coupled with the A_{2A}AR (Fig. 7F), being well agreement with experimental data of the ARs[149–151]. In summary, the dynamic mechanism of specific GPCR-G protein interactions could be obtained from the GaMD simulations.

3.5.3 GPCR-membrane interactions depend on the receptor activation state

—The phospholipid membrane bilayer plays an important role in GPCR transiting among different conformational states. Elucidation of the lipid-protein interactions could facilitate to understand the functional mechanism of GPCRs. In one of our recent study[96], the cryo-EM structure of the active ADO-A_1AR-Gi[146] and the X-ray structure[148, 152] of the inactive antagonist PSB36-bound A_1AR (PSB36-A_1AR) were used to perform GaMD simulations. They were embedded in a 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (POPC) lipid bilayer. GaMD simulations revealed important role of the membrane lipids in stabilizing different states of the A_1AR. Different structural flexibility profiles of the inactive and active A_1AR were obtained by the GaMD simulations. In comparison with the inactive state, higher fluctuations of the A_1AR ECL2 region, intracellular ends of TM6 and TM5 were found in the active state. Furthermore, the -S_{CD} order parameter values obtained from GaMD simulations were consistent with experimental data.[153] Particularly, the inactive and active A_1AR systems exhibited similar values of the -S_{CD} order parameters of sn-2 acyl chains of POPC in the upper leaflet. However, the active A_1AR exhibited smaller value of the same -S_{CD} order parameters in the lower leaflet than those in the inactive A_1AR, suggesting that POPC lipids in the lower leaflet of the active A_1AR system were more fluid than in the inactive A_1AR system. One reliable explanation is that the outward movement of TM6 in the active A_1AR could induce higher inclination of the C-H bonds to be aligned along the bilayer normal. Thus, GaMD simulations showed that the protein-membrane interactions depended on different conformational states of the A_1AR.

3.5.4 Mechanism of allosteric drug lead binding to an adenosine GPCR—

Preclinical studies suggest that the A_1AR is an important drug target for treating diseases including reduce neuropathic pain and ischemia-reperfusion injury[154–156]. However, off-target side effects have hindered the therapeutic development of A_1AR agonists, which mainly originated from the high conservation of the endogenous agonist binding ("orthosteric") site across different AR subtypes[157]. Positive allosteric modulators (PAMs), which bind to a less conserved "allosteric" site, have the potential to develop high subtype selective A_1AR therapeutics[158]. Using the X-ray structure of the A_1AR (PDB: 5UEN)[148] as a model, GaMD simulations[159] were performed to investigate binding mechanisms of two PAMs, VCP171 and PD81723. Each PAM was initially placed at >20 Å away from the receptor. Spontaneous binding of PAMs to the A_1AR was captured in GaMD simulations using both AMBER[48] and NAMD[52] at different acceleration levels (dihedral and dual boost).

GaMD simulations identified similar binding modes of PD81723 and VCP171 that bound to a site around ECL2 in the A_1AR from different acceleration levels of GaMD simulations

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

performed using different software packages. They were in highly agreement with experimental results of site-directed mutagenesis obtained by the Christopoulos group[160]. In the mutagenesis experiments, numerous ECL2 residues including Asn148^{ECL2}, Glu153^{ECL2}, Ser161^{ECL2}, Ile167^{ECL2} and Glu172^{ECL2} were mutated to alanine, which caused significant alterations in PAM binding affinity, efficacy, and cooperativity[160, 161]. These residues were found to interact with the bound PAMs in the GaMD simulations. Additionally, agonist binding affinity was enhanced by PAM binding. In the absence of PAM, the agonist sampled a large conformational space in the receptor orthosteric pocket without binding of PAM. Upon PAM binding at the ECL2, movement of agonist was significantly reduced.

3.5.5 Mechanism of drug binding to a chemokine GPCR—Chemokine receptors are regarded as one of the important GPCRs with implications in human health and therapeutics. CXCR4 is an important subtype with involvement in different human diseases including cancer and HIV infection. Despite its important, less is known about the mechanism of drug interaction with the receptor. GaMD simulations[74] were performed to study the binding mechanism of the drug Plerixafor (PLX) and its pathway to CXCR4. Simulation systems were built by placing ten unbound ligand molecules at a distance >15 Å away from the receptor. GaMD_Dual_NB boost scheme was used in which system nonbonded and dihedral energy terms were boosted. The GaMD_Dual_NB simulations captured spontaneous binding of the PLX from the bulk solvent to the receptor orthosteric site in one of five production runs. The complete binding of PLX was observed at ~480 ns timescale with minimum RMSD relative to the bound conformation of 2.76 Å. In the binding pocket of CXCR4, the positively charged PLX formed stable salt bridges with residues Asp97$^{2.63}$, Asp262$^{6.58}$ and Glu288$^{7.39}$ occupying both the minor and major sub-pockets of the receptor. In 2D PMF profile calculated from the GaMD simulations, “unbound”, “intermediate 1” (I1), “intermediate 2” (I2) and “bound” PLX low-energy conformational states were identified. In the intermediate conformational states I1 and I2, same polar and charged residues in the receptor ECL2-TM5-TM6 region, namely Asp187^{ECL2}, Asp193$^{5.32}$ and Asp262$^{6.58}$, formed favorable interactions with the positively charged nitrogen atoms of PLX. Thus, the ECL2-TM5-TM6 region of CXCR4 formed a novel intermediate drug binding site. Furthermore, GaMD simulations identified PLX drug binding pathway to the CXCR4. These studies are expected to greatly facilitate drug design of CXCR4.

3.5.6 Mechanisms of γ-secretase activation and substrate processing—The mechanism of activation of γ-secretase bound to amyloid precursor protein (APP) was investigated using GaMD simulations[98]. The Cryo-EM structures of two substrates App and Notch bounded γ-secretase were simulated to investigate substrate processing by γ-secretase of wildtype and APP mutant causing the familial Alzheimer’s disease (FAD)[162] (Fig. 8A). Mutations in the cryo-EM structure introduced unnatural enzyme-substrate interactions and hindered the activation process. With a combined study of GaMD simulations, mass spectrometry and western blotting, a model highlighting the process of intramembrane proteolysis of APP by γ-secretase was presented.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Spontaneous activation of γ-secretase in complex with wildtype APP in the presence of a water molecule was captured by GaMD simulations (Figs. 8A and 8D). The water molecule that entered the presenilin active site was trapped between two catalytic Asp residues that were ~7 Å apart forming stable hydrogen bonds. A carbonyl oxygen of the scissile amide bond present between Leu49 and Val50 residues of the APP substrate formed a hydrogen bond with γ-secretase residue Asp257. During activation of γ-secretase, TM1, TM2 and TM8 helices of catalytic PS1 subunit showed some flexibility while TM6a was observed interacting directly with the substrate. Free energy profiles revealed four low-energy conformations of γ-secretase bound to wildtype APP, namely "Inactive", "Intermediate", "Inhibited" and "Active" (Fig. 8D). The active conformational state resembled the activation of the enzyme whereas the inactive state correlated with the starting cryo-EM structure (Figs. 8A & 8B). The active conformational state resembled the activation of the enzyme whereas the inactive state correlated with the starting cryo-EM structure. The inhibited conformational state closely resembled the γ-secretase structure in complex with DAPT inhibitor[163]. The intermediate conformational state resembled the transitional structure in between these states.

In addition to wildtype, GaMD simulations were performed on the APP mutants namely I45F, T48P and M51F that lead to FAD. The I45F and T48P system systems revealed faster activation compared to the wildtype system which were in good agreement with the experimental mass spectrometry data of APP intracellular domain (AICD) proteolytic products that showed greater AICD50–99/AICD49–99 ratio in comparison to the wildtype. Four low-energy conformations, similar to those observed in wildtype, were identified in the FAD mutant systems. Furthermore, the GaMD simulation of M51F mutant shifted the ε cleavage between Thr48-Leu49 (Figs. 8C & 8E). The mass spectrometry analysis validated the AICD products formed as a result of proteolysis cleavage between Thr48-Leu49 was observed higher. *In vitro* assay showed significant high production of AICD in M51F system as compared to other systems. This was consistent with the GaMD free energy landscape as the inhibited state was observed in wildtype, I45F and T48P system but not in M51F system. The experimental validations strongly correlated with the GaMD simulation model of γ-secretase.

GaMD simulations of the wildtype and FAD mutant APPs of the different γ-secretase enzyme systems were analyzed with respect to their secondary structures. In particular, the M51F mutant APP shifted ~4 Å downwards with Thr48 and Leu49 residues flipping its side chain. The C-terminus of APP lost its β-sheet conformation as required for local rearrangements during the shift in ε cleavage. The sub-pockets S1', S2' and S3' in the active site could also be visualized in the wildtype and FAD mutant APPs bound to γ-secretase via GaMD simulations. Overall, the combination of GaMD simulations, mass spectrometry and western blots enabled deep understanding of substrate processing by γ-secretase and its activation.

3.5.7 Structural dynamics of cytochrome P450—Cytochrome P450 3A4 (CYP34A) enzyme plays a crucial role in mammalian metabolic pathways including synthesis and breakdown of fatty acids and hormones. It undergoes large conformational changes in the active site and other structural components of the enzyme. Redhair et.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

al.[93] performed GaMD simulations to understand the protein dynamics and protein-ligand interactions induced by allosteric drug benzodiazepine midazolam (MDZ) in a lipid bilayer. The GaMD simulations showed that F- and G-helical regions could be a possible allosteric site for MDZ drug, which were further verified by hydrogen-deuterium exchange mass spectrometry. The GaMD simulations showed that the local environment at the Phe-cluster comprising the region between the active site and the lipid bilayer was dynamic and could be a possible allosteric site. Even in the presence of a ligand at the active site, the enzyme generated flexible allosteric site nearby. Overall, GaMD simulations provided the molecular basis of enzyme activity and allosteric drug interaction in cytochrome CYP34A.

3.6 Carbohydrates

Here, we summarize recent GaMD studies of carbohydrates, fundamental components of cells that engage in energy functions and form structural components. We review how GaMD has been applied to describe the dynamic interplay of carbohydrates with DNA[101] and proteins,[59] and how the method has been used to facilitate the development of carbohydrate force field parameters.[99]

3.6.1 The importance of carbohydrates conformation on DNA triplex—DNA triplexes are higher-order structural arrangements important for gene regulation and biotechnological applications. GaMD simulations were recently employed to examine the impact of substituting deoxyribose sugars by conformationally locked sugars on the DNA triplex structure.[101] Multiple GaMD simulations were performed on both 3’–5’ and 5’–3’ modified triplexes, as well as on an unmodified DNA triplex, which was used as a control. The simulations revealed that the DNA triplexes, in which the deoxyribose was replaced by locked sugar, lost their structural integrity, and disintegrated resembling the structure of a duplex. On the other hand, the control DNA triplex preserved the structural integrity during the simulations. As a notable observation, both modified triplexes changed conformation reaching duplex structures containing a modified strand and a regular strand, while the third DNA strand was dissociated from the complex. In-depth analysis of the trajectories indicated a significant reduction in the major groove width and diminished solvent accessible surface area in the modified triplexes, as compared to reference systems. On this basis, the authors suggested that the newly introduced locked sugars impose a remarkable steric constraint, which alters the DNA structure and results in the inefficient binding of the third DNA strand. Overall, the authors concluded that knowledge of the structural changes induced by modified sugars could be leveraged for the design of new antisense oligonucleotides, as well as to understand the role of modified oligonucleotides in anticancer therapy.

3.6.2 Carbohydrate-protein interactions—Protein glycosylation is a post-translational modification that is involved in several cellular and biological processes.[59] Glycosyltransferases (GTs) catalyze the transfer of a sugar moiety to acceptor amino acids, such as serine, threonine (O-linked glycosylation) and asparagine (N-linked glycosylation). The glycosylation process also occurs in bacterial proteins, where it plays a critical role in the immune response against pathogens. Some bacterial effectors leverage the glycosylation process to suppress the nuclear factor NF-κB, which is central in regulating the immune response. The non-locus of enterocyte effacement effector protein B (NleB)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

has glycosyltransferase activity and inhibits NF-κB by transferring N-acetyl glucosamine (GlcNAc). To further understand the glycosylation process, Park and co-authors[59] focused on the SseK1 and SseK2 effectors, which are orthologs of NleB from *Salmonella typhimurium*. The authors combined X-ray crystallography, NMR, enzyme kinetics, GaMD simulations and *in vivo* experiments. Structural evidences revealed a glycosyltransferase architecture displaying a helix-loop-helix (HLH) domain relevant to protein substrate recognition and a catalytic core, which includes the conserved catalytic triad (His-Glu-Asn) critical for enzyme catalysis and bacterial virulence. GaMD simulations showed large amplitude motions of the HLH domain, with significant differences in SseK1 and SseK2 that affect the HLH approach toward the substrate binding pocket. Specifically, SseK1 was considerably more flexible than SseK2 in the loop region connecting the HLH, in line with the increased specificity toward the substrate that has been measured experimentally.[59] The simulations also suggested a possible conformation of the catalytically competent active site, showing that the binding of GlcNAc properly orients the substrate, *viz.*, an arginine residue target of the glycosylation process, for the chemical reaction.

3.6.3 Development of carbohydrate force field parameters with GaMD simulations—

Heparin is a highly sulphated, linear polysaccharide belonging to the family of glycosaminoglycans. Endogenous heparin critically regulates blood coagulation by interacting with the small protein antithrombin (AT), through the pentasaccharide domain responsible for the heparin activity.[99, 164] The capability of a heparin penta-saccharide to bind AT is determined by the conformational dynamics of the sugar rings, and particular by the conformation of the L-iduronic acid residue. On this basis, idraparinux derivatives, which are non-glycosaminoglycan analogues of the heparin penta-saccharide, are promising anticoagulant drug candidates. However, computational simulations of carbohydrates for drug discovery are difficult due to their high flexibility. Moreover, the difficulty of computationally modelling idraparinux derivatives is increased by the presence of sulfonato-methyl moieties, which are highly charged. Therefore, to attain a proper description of heparin and idraparinux derivatives, Balogh and co-authors assessed the performance of the GAFF1,[165] GLYCAM06[166, 167] and CHARMM[168, 169] force fields using GaMD simulations.[99] This enabled enhancing the conformational landscape of the pentasaccharide domain obtaining agreement with NMR experiments. The analysis of simulations and their comparison with NMR demonstrated that the CHARMM force field was better reproducing best the experimental data on the ring conformations, producing also good agreement with the Nuclear Overhauser Effects (NOE) distances on L-iduronic acid ring conformations. Therefore, the use of the CHARMM force field was proposed for the exhaustive and comparative conformational analyses of idraparinux derivatives.

3.7 Drug Design

3.7.1 Retrospective ensemble docking of allosteric modulators of A_1AR—

Virtual screening has been widely used for agonist/antagonist design targeting GPCRs.[170] The success rate for virtual screening of GPCRs in the orthosteric pocket is mostly >20%, which is even higher than that of globular proteins.[171] However, it is rather challenging to apply virtual screening to identify allosteric modulators due to their low affinity compared with the agonist/antagonist. In a recent study[103], we tested whether receptor structural

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

ensembles obtained from GaMD simulations could be used to increase docking performance of known PAMs using the A_1AR as a model GPCR. Retrospective ensemble docking calculations of PAMs to the A_1AR combining GaMD simulations and *Autodock*[172] were performed.

The GaMD simulations implemented in AMBER and NAMD were applied to generate receptor ensemble. The flexible docking and rigid-body docking at different levels (short, medium, and long) were all evaluated. Docking scores corrected by the GaMD reweighted free energy of the receptor structural cluster further improved the docking performances. The calculated docking enrichment factors (EFs) and the area under the receiver operating characteristic curves (AUC) are increased using ranking by the average binding energy (BE_{avg}) in comparison with the minimum binding energy (BE_{min}). Ensemble obtained from AMBER dual-boost GaMD simulations of the VCP171-bound ADO-A_1AR-Gi complex outperformed other ensembles for docking. Interactions between the PAM and receptor ECL2 in the VCP171-bound ADO-A_1AR-Gi complex might induce more suitable conformations for PAM binding, which were difficult to be sampled in the simulations of PAM-free (*apo*) A_1AR. Dual-boost GaMD with higher boost potential was observed to perform better than the dihedral-boost GaMD for ensemble docking. Overall, flexible docking performed significantly better than the rigid-body docking at different levels with *AutoDock*, suggesting that the flexibility of protein side chains is also important in ensemble docking. In summary, the docking performance has been highly improved by combining GaMD simulations and flexible docking, which effectively account for the flexibility of backbone and side chain in receptors. Such an ensemble docking protocol will greatly facilitate future PAM design of the A_1AR and other GPCRs.

3.7.2 Discovery of Novel Small-Molecule Calcium Sensitizers for Cardiac Troponin C—Cardiac troponin C (TnC) is a calcium-dependent protein in the troponin complex responsible for the activation of muscle contraction. Disorder of TnC may trigger heart diseases and then cause death. One of the current therapies[173] is to design small molecules that can stabilize an open structure of the TnC and facilitate binding of the TnC switch peptide. To identify potential small molecules for the treatment, Coldren *et al.*[102] combined GaMD and high-throughput virtual screening to predict binding conformations and affinities of small molecules in TnC. The simulations were compared with experiments for the TnC protein structures in complex with calcium sensitivity modulators. 300 ns GaMD simulations were performed on each system to obtain protein conformations. The simulation trajectory snapshots were clustered into ten most representative conformations based on an agglomerative hierarchical algorithm, which were the structures used for virtual screening and docking studies. The work identified a number of novel compounds that reduced the calcium dissociation rate and showed an overall calcium sensitization effect. One of the compounds exhibited high binding affinity in TnC and was further verified by the stopped-flow kinetic experiment.

4. Software

GaMD has been implemented in the common MD simulation packages, including both GPU and CPU versions of AMBER[48, 174], NAMD[52] and GENESIS[61] (Table 1). All the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

above-mentioned GaMD algorithms have been incorporated into the latest GPU version of AMBER 20[175].

5. Conclusions

Without the need to set predefined reaction coordinates or collective variables, GaMD can be advantageous for exploring biomolecular conformational space and complex biomolecular interactions without *a priori* knowledge or constraints. Additionally, the boost potential in GaMD simulations exhibits a Gaussian distribution, allowing for accurate reweighting of the simulations using cumulant expansion to the second order. Thus, GaMD is applicable to a wide range of biological systems as described in this review.

Based on GaMD, novel approaches including rex-GaMD[60], GaREUS[61] and selective GaMD (LiGaMD and Pep-GaMD)[62, 63] have been developed. The selective LiGaMD and Pep-GaMD algorithms appear to be more efficient and easier to use as compared with other existing methods, including the cMD[176, 177], replica exchange[178–180] and metadynamics[180, 181]. For example, a binding event of drug Dasatinib to its binding site of Src Kinase was captured in a total of 35 μs Anton cMD simulation[177]. Repetitive binding and unbinding of an IDP peptide was captured in 200 μs Anton cMD simulations at elevated temperature (400 K)[176]. Replica exchange algorithm needs to simultaneously simulate many replicas to model ligand/peptide binding and dissociation[178–180]. In comparison, LiGaMD and Pep-GaMD simulations was able to capture multiple events of ligand/peptide binding and unbinding within microsecond simulation time. These highly efficient simulations allowed us to accurately characterize the ligand/peptide binding thermodynamics and kinetics[62, 63].

Finally, more efficient GaMD algorithms and enhanced sampling methods in general are still needed to characterize the thermodynamics and kinetics of important protein-protein/nucleic acid interactions, which could involve difficult simulations of both binding and dissociation of large biomolecular complexes. Moreover, structural dynamics in systems of increasing sizes such as viruses and cells present grand challenges for computational modelling and enhanced sampling simulations. Continued innovations in both supercomputing hardware and method developments may help us to address these challenges in the future.

Acknowledgments

This work used supercomputing resources with allocation awards TG-MCB180049 (to YM) and TG-MCB160059 (to GP) through the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562 and project M2874 through the National Energy Research Scientific Computing Center (NERSC, to YM), which is a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231. It also used computational resources provided by the Research Computing Cluster at the University of Kansas, the University of California Riverside High-Performance Computing Center (HPCC) and the Triton Shared Computing Cluster (TSCC) of the San Diego Supercomputing Center. This work was supported in part by the American Heart Association (Award 17SDG33370094), the National Institutes of Health (R01GM132572) and the startup funding in the College of Liberal Arts and Sciences at the University of Kansas to YM. This work was also supported in part by the National Science Foundation under Grant No. CHE-1905374, and startup funding from the University of California Riverside to GP. We are grateful to the startup grant from Wayne State University (to YH).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

References

1. Henzler-Wildman K, Kern D. Dynamic personalities of proteins. Nature 2007, 450:964–972. [PubMed: 18075575]
2. Hatoum-Aslan A, Marraffini LA. Impact of CRISPR immunity on the emergence and virulence of bacterial pathogens. Curr Opin Microbiol 2014, 17:82–90. [PubMed: 24581697]
3. Englander SW, Mayne L. The nature of protein folding pathways. Proc Natl Acad Sci U S A 2014, 111:15873–15880. [PubMed: 25326421]
4. Ritter SL, Hall RA. Fine-tuning of GPCR activity by receptor-interacting proteins. Nat Rev Mol Cell Biol 2009, 10:819–830. [PubMed: 19935667]
5. Onuchic JN, and L-S Z, Wolynes PG. THEORY OF PROTEIN FOLDING: The Energy Landscape Perspective. Annu Rev Phys Chem 1997, 48:545–600. [PubMed: 9348663]
6. Deupi X, Kobilka BK. Energy Landscapes as a Tool to Integrate GPCR Structure, Dynamics, and Function. Physiology 2010, 25:293–303. [PubMed: 20940434]
7. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nat Struct Biol 2002, 9:646–652. [PubMed: 12198485]
8. Hollingsworth SA, Dror RO. Molecular Dynamics Simulation for All. Neuron 2018, 99:1129–1143. [PubMed: 30236283]
9. Harvey MJ, Giupponi G, Fabritiis GD. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. J Chem Theory Comput 2009, 5:1632–1639. [PubMed: 26609855]
10. Johnston JM, Filizola M. Showcasing modern molecular dynamics simulations of membrane proteins through G protein-coupled receptors. Curr Opin Struct Biol 2011, 21:552–558. [PubMed: 21764295]
11. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan Y, et al. Atomic-level characterization of the structural dynamics of proteins. Science 2010, 330:341–346. [PubMed: 20947758]
12. Lane TJ, Shukla D, Beauchamp KA, Pande VS. To milliseconds and beyond: challenges in the simulation of protein folding. Curr Opin Struct Biol 2013, 23:58–65. [PubMed: 23237705]
13. Vilardaga J-P, Bünemann M, Krasel C, Castro M, Lohse MJ. Measurement of the millisecond activation switch of G protein–coupled receptors in living cells. Nat Biotechnol 2003, 21:807–812. [PubMed: 12808462]
14. Miao Y, Ortoleva PJ. Viral structural transitions: An all-atom multiscale theory. J Chem Phys 2006, 125:214901. [PubMed: 17166043]
15. Christen M, van Gunsteren WF. On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: A review. J Comput Chem 2008, 29:157–166. [PubMed: 17570138]
16. Spiwok V, Sucur Z, Hosek P. Enhanced sampling techniques in biomolecular simulations. Biotechnol Adv 2015, 33:1130–1140. [PubMed: 25482668]
17. Abrams C, Bussi G. Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. Entropy 2014, 16:163–199.
18. Miao Y, McCammon JA. Unconstrained Enhanced Sampling for Free Energy Calculations of Biomolecules: A Review. Mol Simul 2016, 42:1046–1055. [PubMed: 27453631]
19. Torrie GM, Valleau JP. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. J Comput Phys 1977, 23:187–199.
20. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comput Chem 1992, 13:1011–1021.
21. Laio A, Gervasio FL. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. Rep Prog Phys 2008, 71:126601.
22. Bešker N, Gervasio FL. Using metadynamics and path collective variables to study ligand binding and induced conformational transitions. In: Computational Drug Discovery and Design: Springer; 2012, 501–513.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

23. Darve E, Rodríguez-Gómez D, Pohorille A. Adaptive biasing force method for scalar and vector free energy calculations. J Chem Phys 2008, 128:144120. [PubMed: 18412436]
24. Darve E, Wilson MA, Pohorille A. Calculating free energies using a scaled-force molecular dynamics algorithm. Mol Simul 2002, 28:113–144.
25. Isralewitz B, Baudry J, Gullingsrud J, Kosztin D, Schulten K. Steered molecular dynamics investigations of protein function. J Mol Graph Model 2001, 19:13–25. [PubMed: 11381523]
26. Grubmuller H Predicting slow structural transitions in macromolecular systems: Conformational flooding. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics 1995, 52:2893–2906. [PubMed: 9963736]
27. Bouvier B, Grubmuller H. A molecular dynamics study of slow base flipping in DNA using conformational flooding. Biophys J 2007, 93:770–786. [PubMed: 17496048]
28. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 1999, 314:141–151.
29. Okamoto Y Generalized-ensemble algorithms: enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. J Mol Graph Model 2004, 22:425–439. [PubMed: 15099838]
30. Hansmann UH. Parallel tempering algorithm for conformational studies of biological molecules. Chem Phys Lett 1997, 281:140–150.
31. Wu X, Wang S. Crossing the time scale of protein folding through self-guided molecular dynamics simulation. In: J Mol Graphics Modell: ELSEVIER SCIENCE INC 655 AVENUE OF THE AMERICAS, NEW YORK, NY 10010 USA; 1998.
32. Wu X, Wang S. Self-guided molecular dynamics simulation for efficient conformational search. J Phys Chem B 1998, 102:7238–7250.
33. Wu X, Brooks BR. Self-guided Langevin dynamics simulation method. Chem Phys Lett 2003, 381:512–518.
34. Wu X, Brooks BR, Vanden-Eijnden E. Self-guided Langevin dynamics via generalized Langevin equation. J Comput Chem 2016, 37:595–601. [PubMed: 26183423]
35. 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, 127:094101. [PubMed: 17824726]
36. Zheng L, 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, 129:014105. [PubMed: 18624468]
37. Lv C, Zheng L, Yang W. Generalized essential energy space random walks to more effectively accelerate solute sampling in aqueous environment. J Chem Phys 2012, 136:044103. [PubMed: 22299857]
38. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J Chem Phys 2004, 120:11919–11929. [PubMed: 15268227]
39. Voter AF. Hyperdynamics: Accelerated molecular dynamics of infrequent events. Phys Rev Lett 1997, 78:3908.
40. Hamelberg D, de Oliveira CAF, McCammon JA. Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. J Chem Phys 2007, 127:10B614.
41. Shen T, Hamelberg D. A statistical analysis of the precision of reweighting-based simulations. J Chem Phys 2008, 129:034103. [PubMed: 18647012]
42. Miao Y, Feixas F, Eun C, McCammon JA. Accelerated molecular dynamics simulations of protein folding. J Comput Chem 2015, 36:1536–1549. [PubMed: 26096263]
43. Miao Y, Sinko W, Pierce L, Bucher D, Walker RC, McCammon JA. Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. J Chem Theory Comput 2014, 10:2677–2689. [PubMed: 25061441]
44. Jiang W, Thirman J, Jo S, Roux B. Reduced Free Energy Perturbation/Hamiltonian Replica Exchange Molecular Dynamics Method with Unbiased Alchemical Thermodynamic Axis. J Phys Chem B 2018, 122:9435–9442. [PubMed: 30253098]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

45. Fajer M, Hamelberg D, McCammon JA. Replica-Exchange Accelerated Molecular Dynamics (REXAMD) Applied to Thermodynamic Integration. J Chem Theory Comput 2008, 4:1565–1569. [PubMed: 19461870]
46. Miao Y, Nichols SE, McCammon JA. Free energy landscape of G-protein coupled receptors, explored by accelerated molecular dynamics. Phys Chem Chem Phys 2014, 16:6398–6406. [PubMed: 24445284]
47. Kappel K, Miao YL, McCammon JA. Accelerated molecular dynamics simulations of ligand binding to a muscarinic G-protein-coupled receptor. Q Rev Biophys 2015, 48:479–487. [PubMed: 26537408]
48. Miao Y, Caliman Alisha D, McCammon JA. Allosteric Effects of Sodium Ion Binding on Activation of the M3 Muscarinic G-Protein-Coupled Receptor. Biophys J 2015, 108:1796–1806. [PubMed: 25863070]
49. Miao Y, McCammon JA. Chapter Six - Gaussian Accelerated Molecular Dynamics: Theory, Implementation, and Applications. In: Dixon DA, ed. Annu Rep Comput Chem. Vol. 13: Elsevier; 2017, 231–278. [PubMed: 29720925]
50. Bhattarai A, Miao Y. Gaussian accelerated molecular dynamics for elucidation of drug pathways. Expert Opin on Drug Discovery 2018, 13:1055–1065.
51. Miao Y, McCammon JA. Graded activation and free energy landscapes of a muscarinic G-protein–coupled receptor. Proc Natl Acad Sci U S A 2016, 113:12162–12167. [PubMed: 27791003]
52. Pang YT, Miao Y, Wang Y, McCammon JA. Gaussian Accelerated Molecular Dynamics in NAMD. J Chem Theory Comput 2017, 13:9–19. [PubMed: 28034310]
53. Salawu EO. The Impairment of TorsinA's Binding to and Interactions With Its Activator: An Atomistic Molecular Dynamics Study of Primary Dystonia. Front Mol Biosci 2018, 5:64. [PubMed: 30042949]
54. Zhang J, Wang N, Miao Y, Hauser F, McCammon JA, Rappel W-J, Schroeder JI. Identification of SLAC1 anion channel residues required for CO2 bicarbonate sensing and regulation of stomatal movements. Proc Natl Acad Sci U S A 2018.
55. Wang Y-T, Chan Y-H. Understanding the molecular basis of agonist/antagonist mechanism of human mu opioid receptor through gaussian accelerated molecular dynamics method. Sci Rep 2017, 7:7828. [PubMed: 28798303]
56. Liao JM, Wang YT. In silico studies of conformational dynamics of Mu opioid receptor performed using gaussian accelerated molecular dynamics. J Biomol Struct Dyn 2019, 37:166–177. [PubMed: 29277141]
57. Chuang CH, Chiou SJ, Cheng TL, Wang YT. A molecular dynamics simulation study decodes the Zika virus NS5 methyltransferase bound to SAH and RNA analogue. Sci Rep 2018, 8:6336. [PubMed: 29679079]
58. Miao Y, Huang Y-mM, Walker RC, McCammon JA, Chang C-eA. Ligand Binding Pathways and Conformational Transitions of the HIV Protease. Biochemistry 2018, 57:1533–1541. [PubMed: 29394043]
59. Park JB, Kim YH, Yoo Y, Kim J, Jun SH, Cho JW, El Qaidi S, Walpole S, Monaco S, Garcia-Garcia AA, et al. Structural basis for arginine glycosylation of host substrates by bacterial effector proteins. Nat Commun 2018, 9:4283. [PubMed: 30327479]
60. Huang Y-mM McCammon JA, Miao Y. Replica Exchange Gaussian Accelerated Molecular Dynamics: Improved Enhanced Sampling and Free Energy Calculation. J Chem Theory Comput 2018, 14:1853–1864. [PubMed: 29489349]
61. 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, 15:5199–5208. [PubMed: 31539245]
62. Miao Y, Bhattarai A, Wang J. Ligand Gaussian Accelerated Molecular Dynamics (LiGaMD): Characterization of Ligand Binding Thermodynamics and Kinetics. J Chem Theory Comput 2020, 16:5526–5547. [PubMed: 32692556]
63. Wang J, Miao Y. Peptide Gaussian accelerated molecular dynamics (Pep-GaMD): Enhanced sampling and free energy and kinetics calculations of peptide binding. J Chem Phys 2020, 153:154109. [PubMed: 33092378]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

64. Gao Y, Cao D, Pawnikar S, John KP, Ahn HM, Hill S, Ha JM, Parikh P, Ogilvie C, Swain A, et al. Structure of the Human Respiratory Syncytial Virus M2–1 Protein in Complex with a Short Positive-Sense Gene-End RNA. Structure 2020, 28:979–990.e974. [PubMed: 32697936]
65. Wang J, Lan L, Wu X, Xu L, Miao Y. Mechanism of RNA recognition by a Musashi RNA-binding protein. bioRxiv 2020:2020.2010.2030.362756.
66. Roy R, Mishra A, Poddar S, Nayak D, Kar P. Investigating the mechanism of recognition and structural dynamics of nucleoprotein-RNA complex from Peste des petits ruminants virus via Gaussian accelerated molecular dynamics simulations. J Biomol Struct Dyn 2020:1–14.
67. 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, 142:1348–1358. [PubMed: 31885264]
68. 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 Cent Sci 2019.
69. Wang J, Miao Y. Mechanistic Insights into Specific G Protein Interactions with Adenosine Receptors. J Phys Chem B 2019, 123:6462–6473. [PubMed: 31283874]
70. Miao Y, McCammon JA. Mechanism of the G-protein mimetic nanobody binding to a muscarinic G-protein-coupled receptor. Proc Natl Acad Sci U S A 2018, 115:3036–3041. [PubMed: 29507218]
71. Sibener LV, Fernandes RA, Kolawole EM, Carbone CB, Liu F, McAffee D, Birnbaum ME, Yang X, Su LF, Yu W, et al. Isolation of a Structural Mechanism for Uncoupling T Cell Receptor Signaling from Peptide-MHC Binding. Cell 2018, 174:672–687 e627. [PubMed: 30053426]
72. Lin W-W, Wang Y-J, Ko C-W, Cheng T- L, Wang Y-T. Cyclic Peptide Inhibitors of the Tsg101 UEV Protein Interactions Refined through Global Docking and Gaussian Accelerated Molecular Dynamics Simulations. Polymers 2020, 12:2235.
73. Petrizzelli F, Biagini T, Barbieri A, Parca L, Panzironi N, Castellana S, Caputo V, Vescovi AL, Carella M, Mazza T. Mechanisms of pathogenesis of missense mutations on the KDM6A-H3 interaction in type 2 Kabuki Syndrome. Comput Struct Biotechnol J 2020, 18:2033–2042. [PubMed: 32802275]
74. Pawnikar S, Miao Y. Pathway and mechanism of drug binding to chemokine receptors revealed by accelerated molecular simulations. Future Med Chem 2020, 12:1213–1225. [PubMed: 32515227]
75. Chen J, Wang W, Sun H, Pang L, Yin B. Mutation-mediated influences on binding of anaplastic lymphoma kinase to crizotinib decoded by multiple replica Gaussian accelerated molecular dynamics. J Comput-Aided Mol Des 2020, 34:1289–1305. [PubMed: 33073300]
76. Paco L, Zarate-Perez F, Clouser AF, Atkins WM, Hackett JC. Dynamics and Mechanism of Binding of Androstenedione to Membrane-Associated Aromatase. Biochemistry 2020, 59:2999–3009. [PubMed: 32786398]
77. Chen Y, Liu T, Xi Q, Jia W, Yin D, Wang X. A Computational Approach to the Study of the Binding Mode of S1P1R Agonists Based on the Active-Like Receptor Model. J Chem Inf Model 2019, 59:1624–1633. [PubMed: 30856323]
78. Tyagi C, Marik T, Vagvolgyi C, Kredics L, Otvos F. Accelerated Molecular Dynamics Applied to the Peptaibol Folding Problem. Int J Mol Sci 2019, 20.
79. Peng Y, Cao S, Kiselar J, Xiao X, Du Z, Hsieh A, Ko S, Chen Y, Agrawal P, Zheng W, et al. A Metastable Contact and Structural Disorder in the Estrogen Receptor Transactivation Domain. Structure 2019, 27:229–240 e224. [PubMed: 30581045]
80. Brown BP, Zhang Y-K, Westover D, Yan Y, Qiao H, Huang V, Du Z, Smith JA, Ross JS, Miller VA, et al. On-target Resistance to the Mutant-Selective EGFR Inhibitor Osimertinib Can Develop in an Allele-Specific Manner Dependent on the Original EGFR-Activating Mutation. Clin Cancer Res 2019, 25:3341–3351. [PubMed: 30796031]
81. Purslow JA, Nguyen TT, Egner TK, Dotas RR, Khatiwada B, Venditti V. Active Site Breathing of Human Alkbh5 Revealed by Solution NMR and Accelerated Molecular Dynamics. Biophys J 2018, 115:1895-1905. [PubMed: 30352661]
82. Venkatramani A, Gravina Ricci C, Oldfield E, McCammon JA. Remarkable similarity in Plasmodium falciparum and Plasmodium vivax geranylgeranyl diphosphate synthase dynamics

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and its implication for antimalarial drug design. Chem Biol Drug Des 2018, 91:1068–1077. [PubMed: 29345110]

83. Dodge GJ, Patel A, Jaremko KL, McCammon JA, Smith JL, Burkart MD. Structural and dynamical rationale for fatty acid unsaturation in Escherichia coli. Proc Natl Acad Sci U S A 2019, 116:6775–6783. [PubMed: 30872475]
84. Dujardin M, Madan V, Gandhi NS, Cantrelle FX, Launay H, Huvent I, Bartenschlager R, Lippens G, Hanoulle X. Cyclophilin A allows the allosteric regulation of a structural motif in the disordered domain 2 of NS5A and thereby fine-tunes HCV RNA replication. J Biol Chem 2019, 294:13171–13185. [PubMed: 31315928]
85. de Azevedo EC, Nascimento AS. Energy landscape of the domain movement in Staphylococcus aureus UDP-N-acetylglucosamine 2-epimerase. J Struct Biol 2019, 207:158–168. [PubMed: 31088716]
86. Han J, Shi X, Du Y, Shi F, Zhang B, Zheng Z, Xu J, Jiang L. Schisandrin C targets Keap1 and attenuates oxidative stress by activating Nrf2 pathway in Ang II-challenged vascular endothelium. Phytother Res 2019, 33:779–790. [PubMed: 30632210]
87. Wu X, Wang Y, Jia R, Fang F, Liu Y, Cui W. Computational and biological investigation of the soybean lecithin-gallic acid complex for ameliorating alcoholic liver disease in mice with iron overload. Food Funct 2019, 10:5203–5214. [PubMed: 31380553]
88. Koh A, Gibbon MJ, Van der Kamp MW, Pudney CR, Gebhard S. Conformation control of the histidine kinase BceS of Bacillus subtilis by its cognate ABC-transporter facilitates need-based activation of antibiotic resistance. Mol Microbiol 2020.
89. Moffett AS, Shukla D. Structural Consequences of Multisite Phosphorylation in the BAK1 Kinase Domain. Biophys J 2020, 118:698–707. [PubMed: 31962105]
90. Philpott JM, Narasimamurthy R, Ricci CG, Freeberg AM, Hunt SR, Yee LE, Pelofsky RS, Tripathi S, Virshup DM, Partch CL. Casein kinase 1 dynamics underlie substrate selectivity and the PER2 circadian phosphoswitch. Elife 2020, 9.
91. Sztain T, Amaro R, McCammon JA. Elucidation of cryptic and allosteric pockets within the SARS-CoV-2 protease. bioRxiv 2020:2020.2007.2023.218784.
92. Zhang Q, Tan S, Xiao T, Liu H, Shah SJA, Liu H. Probing the Molecular Mechanism of Rifampin Resistance Caused by the Point Mutations S456L and D441V on Mycobacterium tuberculosis RNA Polymerase through Gaussian Accelerated Molecular Dynamics Simulation. Antimicrob. Agents Chemother 2020, 64:e02476–02419. [PubMed: 32393493]
93. Redhair M, Hackett JC, Pelletier RD, Atkins WM. Dynamics and Location of the Allosteric Midazolam Site in Cytochrome P4503A4 in Lipid Nanodiscs. Biochemistry 2020, 59:766–779. [PubMed: 31961139]
94. An X, Bai Q, Bing Z, Zhou S, Shi D, Liu H, Yao X. How Does Agonist and Antagonist Binding Lead to Different Conformational Ensemble Equilibria of the kappa-Opioid Receptor: Insight from Long-Time Gaussian Accelerated Molecular Dynamics Simulation. ACS Chem Neurosci 2019, 10:1575–1584. [PubMed: 30372027]
95. Conrad M, Soldner CA, Miao Y, Sticht H. Agonist Binding and G Protein Coupling in Histamine H2 Receptor: A Molecular Dynamics Study. Int J Mol Sci 2020, 21.
96. Bhattarai A, Wang J, Miao Y. G-Protein-Coupled Receptor-Membrane Interactions Depend on the Receptor Activation State. J Comput Chem 2020, 41:460–471. [PubMed: 31602675]
97. Zhao Y, Ung PM, Zahoranszky-Kohalmi G, Zakharov AV, Martinez NJ, Simeonov A, Glaaser IW, Rai G, Schlessinger A, Marugan JJ, et al. Identification of a G-Protein-Independent Activator of GIRK Channels. Cell Rep 2020, 31:107770. [PubMed: 32553165]
98. Bhattarai A, Devkota S, Bhattarai S, Wolfe MS, Miao Y. Mechanisms of gamma-Secretase Activation and Substrate Processing. ACS Cent Sci 2020, 6:969–983. [PubMed: 32607444]
99. Balogh G, Gyöngyösi T, Timári I, Herczeg M, Borbás A, Fehér K, Kövér KE. Comparison of Carbohydrate Force Fields Using Gaussian Accelerated Molecular Dynamics Simulations and Development of Force Field Parameters for Heparin-Analogue Pentasaccharides. J Chem Inf Model 2019, 59:4855–4867. [PubMed: 31593467]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

100. Balogh G, Komaromi I, Bereczky Z. The mechanism of high affinity pentasaccharide binding to antithrombin, insights from Gaussian accelerated molecular dynamics simulations. J Biomol Struct Dyn 2020, 38:4718–4732. [PubMed: 31686597]
101. Pant P, Fisher M. DNA triplex with conformationally locked sugar disintegrates to duplex: Insights from molecular simulations. Biochem Biophys Res Commun 2020, 532:662–667. [PubMed: 32907714]
102. Coldren WH, Tikunova SB, Davis JP, Lindert S. Discovery of Novel Small-Molecule Calcium Sensitizers for Cardiac Troponin C: A Combined Virtual and Experimental Screening Approach. J Chem Inf Model 2020, 60:3648–3661. [PubMed: 32633957]
103. Bhattarai A, Wang J, Miao Y. Retrospective ensemble docking of allosteric modulators in an adenosine G-protein-coupled receptor. Biochim Biophys Acta Gen Subj 2020:129615. [PubMed: 32298791]
104. Hummer G Fast-growth thermodynamic integration: Error and efficiency analysis. J Chem Phys 2001, 114:7330–7337.
105. Eastwood MP, Hardin C, Luthey-Schulten Z, Wolynes PG. Statistical mechanical refinement of protein structure prediction schemes: Cumulant expansion approach. J Chem Phys 2002, 117:4602–4615.
106. Vanommeslaeghe K, MacKerell AD Jr., CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. Biochim. Biophys. Acta, Gen. Subj 2015, 1850:861–871.
107. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 2003, 24:1999–2012. [PubMed: 14531054]
108. Miao Y Acceleration of biomolecular kinetics in Gaussian accelerated molecular dynamics. J Chem Phys 2018, 149:072308. [PubMed: 30134710]
109. Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science 2014, 346:1258096. [PubMed: 25430774]
110. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 2012, 337:816–821. [PubMed: 22745249]
111. Zhu X, Clarke R, Puppala AK, Chittori S, Merk A, Merrill BJ, Simonović M, Subramaniam S. Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. Nat Struct Mol Biol 2019, 26:679–685. [PubMed: 31285607]
112. Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, Thompson AJ, Nogales E, Doudna JA. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. Science 2016, 351:867–871. [PubMed: 26841432]
113. Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science 2014, 343:1247997.
114. Jiang F, Zhou K, Ma L, Gressel S, Doudna JA. STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. Science 2015, 348:1477–1481. [PubMed: 26113724]
115. Palermo G, Miao Y, Walker RC, Jinek M, McCammon JA. CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. Proc Natl Acad Sci U S A 2017, 114:7260–7265. [PubMed: 28652374]
116. Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell 2014, 156:935–949. [PubMed: 24529477]
117. Palermo G, Miao Y, Walker RC, Jinek M, McCammon JA. Striking Plasticity of CRISPR-Cas9 and Key Role of Non-target DNA, as Revealed by Molecular Simulations. ACS Cent Sci 2016, 2:756–763. [PubMed: 27800559]
118. Dagdas YS, Chen JS, Sternberg SH, Doudna JA, Yildiz A. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. Sci Adv 2017, 3:eaao0027. [PubMed: 28808686]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

119. Sternberg SH, LaFrance B, Kaplan M, Doudna JA. Conformational control of DNA target cleavage by CRISPR-Cas9. Nature 2015, 527:110–113. [PubMed: 26524520]
120. Casalino L, Nierzwicki Ł, Jinek M, Palermo G. Catalytic Mechanism of Non-Target DNA Cleavage in CRISPR-Cas9 Revealed by Ab-Initio Molecular Dynamics. ACS Catal. DOI: 10.1021/acscatal.0c03566. 2020.
121. Palermo G. Structure and dynamics of the CRISPR–Cas9 catalytic complex. J Chem Inf Model 2019, 59:2394–2406. [PubMed: 30763088]
122. Shaw DE, Grossman J, Bank JA, Batson B, Butts JA, Chao JC, Deneroff MM, Dror RO, Even A, Fenton CH. Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. In: SC'14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis: IEEE; 2014.
123. Palermo G, Chen JS, Ricci CG, Rivalta I, Jinek M, Batista VS, Doudna JA, McCammon JA. Key role of the REC lobe during CRISPR–Cas9 activation by 'sensing','regulating', and 'locking'the catalytic HNH domain. Q Rev Biophys 2018, 51.
124. Yang M, Peng S, Sun R, Lin J, Wang N, Chen C. The Conformational Dynamics of Cas9 Governing DNA Cleavage Are Revealed by Single-Molecule FRET. Cell Rep 2018, 22:372–382. [PubMed: 29320734]
125. Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, Harrington LB, Sternberg SH, Joung JK, Yildiz A, Doudna JA. Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. Nature 2017, 550:407–410. [PubMed: 28931002]
126. Mitchell BP, Hsu RV, Medrano MA, Zewde NT, Narkhede YB, Palermo G. Spontaneous embedding of DNA mismatches within the RNA: DNA hybrid of CRISPR-Cas9. Front Mol Biosci 2020, 7:39. [PubMed: 32258048]
127. Schmid-Burgk JL, Gao L, Li D, Gardner Z, Strecker J, Lash B, Zhang F. Highly Parallel Profiling of Cas9 Variant Specificity. Mol Cell 2020, 78:794–800 e798. [PubMed: 32187529]
128. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. Science 2016, 351:84–88. [PubMed: 26628643]
129. Nierzwicki Ł, Arantes PR, Saha A, Palermo G. Establishing the allosteric mechanism in CRISPR-Cas9. WIREs Comput. Mol. Sci 2020, n/a:e1503.
130. Palermo G, Ricci CG, Fernando A, Basak R, Jinek M, Rivalta I, Batista VS, McCammon JA. Protospacer Adjacent Motif-Induced Allostery Activates CRISPR-Cas9. J Am Chem Soc 2017, 139:16028–16031. [PubMed: 28764328]
131. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 2014, 507:62–67. [PubMed: 24476820]
132. Sethi A, Eargle J, Black AA, Luthey-Schulten Z. Dynamical networks in tRNA:protein complexes. Proc Natl Acad Sci U S A 2009, 106:6620–6625. [PubMed: 19351898]
133. Ahrens VM, Bellmann-Sickert K, Beck-Sickinger AG. Peptides and peptide conjugates: therapeutics on the upward path. Future Med Chem 2012, 4:1567–1586. [PubMed: 22917246]
134. Fosgerau K, Hoffmann T. Peptide therapeutics: current status and future directions. Drug Discov Today 2015, 20:122–128. [PubMed: 25450771]
135. Kahler U, Fuchs JE, Goettig P, Liedl KR. An unexpected switch in peptide binding mode: from simulation to substrate specificity. J Biomol Struct Dyn 2018, 36:4072–4084. [PubMed: 29210603]
136. Andreani J, Guerois R. Evolution of protein interactions: from interactomes to interfaces. Arch Biochem Biophys 2014, 554:65–75. [PubMed: 24853495]
137. Arkin MR, Wells JA. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. Nat Rev Drug Discov 2004, 3:301–317. [PubMed: 15060526]
138. Wang J, Alekseenko A, Kozakov D, Miao Y. Improved Modeling of Peptide-Protein Binding Through Global Docking and Accelerated Molecular Dynamics Simulations. Front Mol Biosci 2019, 6:112. [PubMed: 31737642]
139. Porter KA, Xia B, Beglov D, Bohnuud T, Alam N, Schueler-Furman O, Kozakov D. ClusPro PeptiDock: efficient global docking of peptide recognition motifs using FFT. Bioinformatics 2017, 33:3299–3301. [PubMed: 28430871]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

140. Wang Y-T, Cheng T-L. Computational modeling of cyclic peptide inhibitor–MDM2/MDMX binding through global docking and Gaussian accelerated molecular dynamics simulations. J Biomol Struct Dyn 2020:1–10.
141. Ahmad M, Helms V. How do proteins associate? A lesson from SH3 domain. Chem Cent J 2009, 3:O22.
142. Ball LJ, Kuhne R, Schneider-Mergener J, Oschkinat H. Recognition of proline-rich motifs by protein-protein-interaction domains. Angew Chem Int Ed Engl 2005, 44:2852–2869. [PubMed: 15880548]
143. 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, 115:9431–9438. [PubMed: 21714551]
144. Hussein HA, Borrel A, Geneix C, Petitjean M, Regad L, Camproux A-C. PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins. Nucleic Acids Res 2015, 43:W436–W442. [PubMed: 25956651]
145. Kruse AC, Ring AM, Manglik A, Hu J, Hu K, Eitel K, Hübner H, Pardon E, Valant C, Sexton PM, et al. Activation and allosteric modulation of a muscarinic acetylcholine receptor. Nature 2013, 504:101–106. [PubMed: 24256733]
146. Draper-Joyce CJ, Khoshouei M, Thal DM, Liang YL, Nguyen ATN, Furness SGB, Venugopal H, Baltos JA, Plitzko JM, Danev R, et al. Structure of the adenosine-bound human adenosine A1 receptor-Gi complex. Nature 2018, 558:559–563. [PubMed: 29925945]
147. Garcia-Nafria J, Lee Y, Bai X, Carpenter B, Tate CG. Cryo-EM structure of the adenosine A2A receptor coupled to an engineered heterotrimeric G protein. Elife 2018, 7.
148. Glukhova A, Thal DM, Nguyen AT, Vecchio EA, Jorg M, Scammells PJ, May LT, Sexton PM, Christopoulos A. Structure of the Adenosine A1 Receptor Reveals the Basis for Subtype Selectivity. Cell 2017, 168:867–877 e813. [PubMed: 28235198]
149. Gao ZG, Inoue A, Jacobson KA. On the G protein-coupling selectivity of the native A2B adenosine receptor. Biochem Pharmacol 2018, 151:201–213. [PubMed: 29225130]
150. Cordeaux Y, Ijzerman AP, Hill SJ. Coupling of the human A1 adenosine receptor to different heterotrimeric G proteins: evidence for agonist-specific G protein activation. Br J Pharmacol 2004, 143:705–714. [PubMed: 15302686]
151. Cordeaux Y, Briddon SJ, Megson AE, McDonnell J, Dickenson JM, Hill SJ. Influence of receptor number on functional responses elicited by agonists acting at the human adenosine A(1) receptor: evidence for signaling pathway-dependent changes in agonist potency and relative intrinsic activity. Mol Pharmacol 2000, 58:1075–1084. [PubMed: 11040056]
152. Cheng RKY, Segala E, Robertson N, Deflorian F, Dore AS, Errey JC, Fiez-Vandal C, Marshall FH, Cooke RM. Structures of Human A1 and A2A Adenosine Receptors with Xanthines Reveal Determinants of Selectivity. Structure 2017, 25:1275–1285 e1274. [PubMed: 28712806]
153. Perly B, Smith IC, Jarrell HC. Acyl chain dynamics of phosphatidylethanolamines containing oleic acid and dihydrosterculic acid: deuteron NMR relaxation studies. Biochemistry 1985, 24:4659–4665. [PubMed: 4063348]
154. Shneyvays V, Leshem D, Zinman T, Mamedova LK, Jacobson KA, Shainberg A. Role of adenosine A1 and A3 receptors in regulation of cardiomyocyte homeostasis after mitochondrial respiratory chain injury. Am J Physiol Heart Circ Physiol 2005, 288:H2792–2801. [PubMed: 15681707]
155. Schutte F, Burgdorf C, Richardt G, Kurz T. Adenosine A1 receptor-mediated inhibition of myocardial norepinephrine release involves neither phospholipase C nor protein kinase C but does involve adenylyl cyclase. Can J Physiol Pharmacol 2006, 84:573–577. [PubMed: 16902603]
156. Liang BT. Protein kinase C-mediated preconditioning of cardiac myocytes: role of adenosine receptor and KATP channel. Am J Physiol Heart Circ Physiol 1997, 273:H847–H853.
157. Kiesman WF, Elzein E, Zablocki J. A 1 adenosine receptor antagonists, agonists, and allosteric enhancers. In: Adenosine receptors in health and disease: Springer; 2009, 25–58.
158. Romagnoli R, Baraldi PG, Tabrizi AM, Gessi S, Borea PA, Merighi S. Allosteric enhancers of A1 adenosine receptors: state of the art and new horizons for drug development. Curr Med Chem 2010, 17:3488–3502. [PubMed: 20738250]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

159. 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, 8:16836. [PubMed: 30442899]

160. Nguyen AT, Vecchio EA, Thomas T, Nguyen TD, Aurelio L, Scammells PJ, White PJ, Sexton PM, Gregory KJ, May LT, et al. Role of the Second Extracellular Loop of the Adenosine A1 Receptor on Allosteric Modulator Binding, Signaling, and Cooperativity. Mol Pharmacol 2016, 90:715–725. [PubMed: 27683013]

161. Peeters MC, Wisse LE, Dinaj A, Vroling B, Vriend G, Ijzerman AP. The role of the second and third extracellular loops of the adenosine A1 receptor in activation and allosteric modulation. Biochem Pharmacol 2012, 84:76–87. [PubMed: 22449615]

162. Zhou R, Yang G, Guo X, Zhou Q, Lei J, Shi Y. Recognition of the amyloid precursor protein by human gamma-secretase. Science 2019, 363.

163. Bai XC, Rajendra E, Yang G, Shi Y, Scheres SH. Sampling the conformational space of the catalytic subunit of human gamma-secretase. Elife 2015, 4.

164. Cui JY, Zhang F, Nierzwicki L, Palermo G, Linhardt RJ, Lisi GP. Mapping the Structural and Dynamic Determinants of pH-Sensitive Heparin Binding to Granulocyte Macrophage Colony Stimulating Factor. Biochemistry 2020, 59:3541–3553. [PubMed: 32897051]

165. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. J Comput Chem 2004, 25:1157–1174. [PubMed: 15116359]

166. Singh A, Tessier MB, Pederson K, Wang X, Venot AP, Boons GJ, Prestegard JH, Woods RJ. Extension and validation of the GLYCAM force field parameters for modeling glycosaminoglycans. Can J Chem 2016, 94:927–935. [PubMed: 28603292]

167. Kirschner KN, Yongye AB, Tschampel SM, Gonzalez-Outeirino J, Daniels CR, Foley BL, Woods RJ. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. J Comput Chem 2008, 29:622–655. [PubMed: 17849372]

168. Mallajosyula SS, Guvench O, Hatcher E, Mackerell AD Jr., CHARMM Additive All-Atom Force Field for Phosphate and Sulfate Linked to Carbohydrates. J Chem Theory Comput 2012, 8:759–776. [PubMed: 22685386]

169. Guvench O, Mallajosyula SS, Raman EP, Hatcher E, Vanommeslaeghe K, Foster TJ, Jamison FW 2nd, Mackerell AD Jr., CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling. J Chem Theory Comput 2011, 7:3162–3180. [PubMed: 22125473]

170. Wang J, Bhattarai A, Ahmad WI, Farnan TS, John KP, Miao Y. Chapter 15 - Computer-aided GPCR drug discovery. In: Jastrzebska B, Park PSH, eds. GPCRs: Academic Press; 2020, 283–293.

171. Shoichet BK, Kobilka BK. Structure-based drug screening for G-protein-coupled receptors. Trends Pharmacol Sci 2012, 33:268–272. [PubMed: 22503476]

172. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem 2009, 30:2785–2791. [PubMed: 19399780]

173. Remme WJ, Swedberg K, Task Force for the D, Treatment of Chronic Heart Failure ESoC. Guidelines for the diagnosis and treatment of chronic heart failure. Eur Heart J 2001, 22:1527–1560. [PubMed: 11492984]

174. Case DSC D.A., Cheatham TE III, Darden TA, Duke RE, Giese TJ, Gohlke H, Goetz AW, Greene D, Homeyer N, Izadi S, Kovalenko A, Lee TS, LeGrand S, Li P, Lin C, Liu J, Luchko T, Luo R, Mermelstein D, Merz KM, Monard G, Nguyen H, Omelyan I, Onufriev A, Pan F, Qi R, Roe DR, Roitberg A, Sagui C, Simmerling CL, Botello-Smith WM, Swails J, Walker RC, Wang J, Wolf RM, Wu X, Xiao L, York DM and Kollman PA (2020), AMBER 2020, University of California, San Francisco.

175. Case DA, Belfon K, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham TE I, Cruzeiro VWD, Darden TA, Duke RE, Giambasu G, et al. AMBER 20, University of California, San Francisco. 2020.

176. Robustelli P, Piana S, Shaw DE. Mechanism of Coupled Folding-upon-Binding of an Intrinsically Disordered Protein. J Am Chem Soc 2020, 142:11092–11101. [PubMed: 32323533]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

177. 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, 133:9181–9183. [PubMed: 21545110]

178. Morrone JA, Perez A, MacCallum J, Dill KA. Computed Binding of Peptides to Proteins with MELD-Accelerated Molecular Dynamics. J Chem Theory Comput 2017, 13:870–876. [PubMed: 28042966]

179. Morrone JA, Perez A, Deng Q, Ha SN, Holloway MK, Sawyer TK, Sherborne BS, Brown FK, Dill KA. Molecular Simulations Identify Binding Poses and Approximate Affinities of Stapled α-Helical Peptides to MDM2 and MDMX. J Chem Theory Comput 2017, 13:863–869. [PubMed: 28042965]

180. Zou R, Zhou Y, Wang Y, Kuang G, Ågren H, Wu J, Tu Y. Free Energy Profile and Kinetics of Coupled Folding and Binding of the Intrinsically Disordered Protein p53 with MDM2. J Chem Inf Model 2020, 60:1551–1558. [PubMed: 32053358]

181. Tiwary P, Limongelli V, Salvalaglio M, Parrinello M. Kinetics of protein–ligand unbinding: Predicting pathways, rates, and rate-limiting steps. Proc Natl Acad Sci U S A 2015, 112:E386–E391. [PubMed: 25605901]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

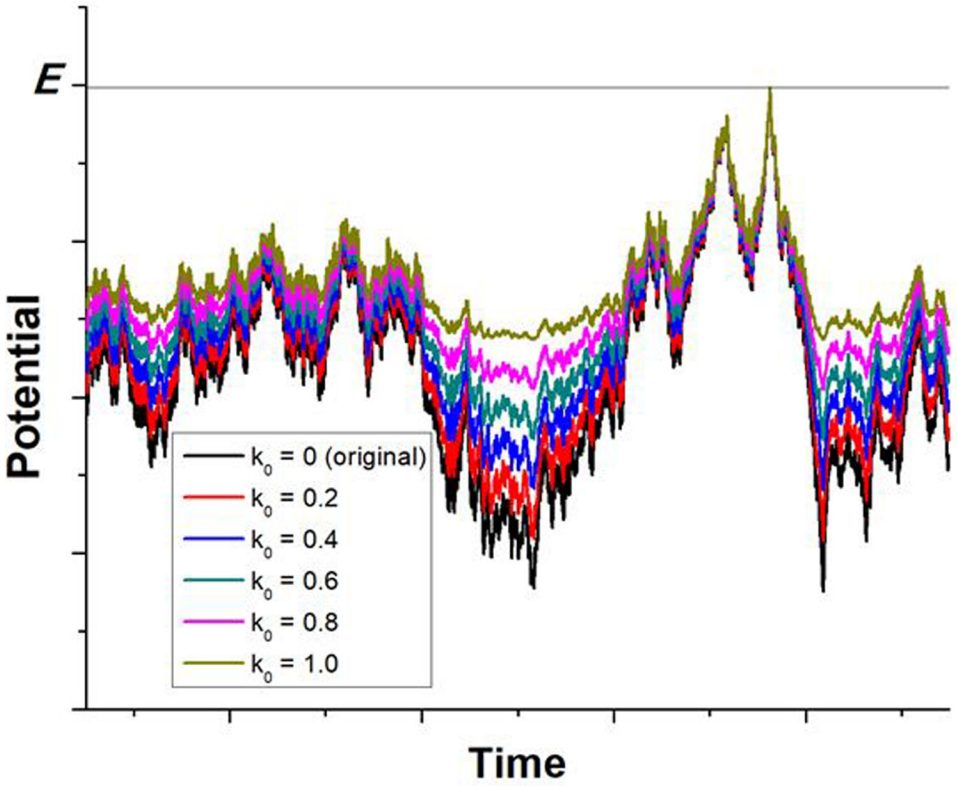


Figure 1.

Scheme illustration of Gaussian accelerated molecular dynamics (GaMD). When the threshold energy is set to the maximum potential (E=Vmax), the system potential energy surface is smoothened by adding a harmonic boost potential that follows Gaussian distribution. The coefficient k_0 in the range of 0−1 determines the magnitude of the applied boost potential. With greater k_0, higher boost potential is added to the original energy surface in conventional molecular dynamics (cMD), which provides enhanced sampling of biomolecules across decreased energy barriers. Adapted with permission from Miao et al. (2015). Copyright 2015 American Chemical Society. https://pubs.acs.org/doi/abs/10.1021/acs.jctc.5b00436. Further permissions related to the material excerpted should be directed to the American Chemical Society.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

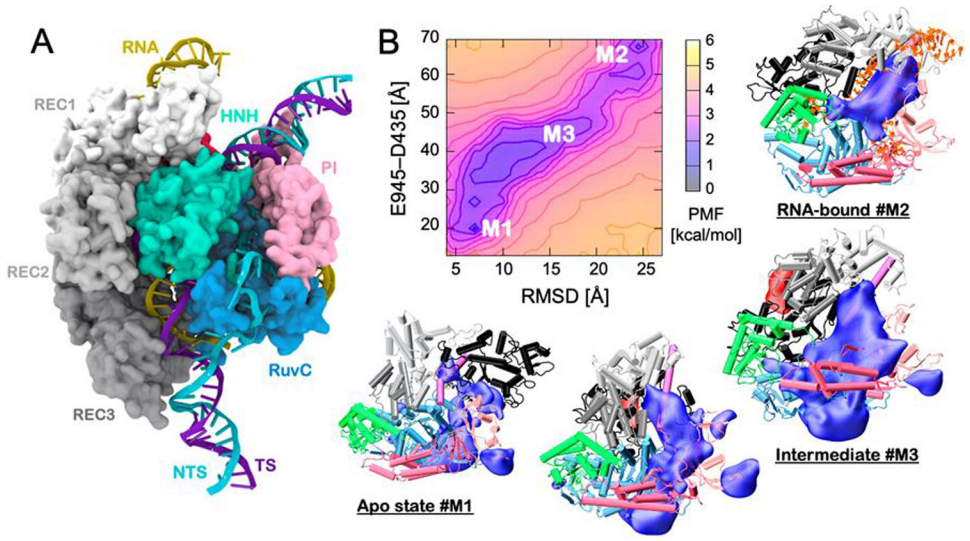


Figure 2.

(**A**) Overview of the Streptococcus pyogenes CRISPR-Cas9 system. The Cas9 protein is represented in molecular surface, showing individual domains in different colors. The RNA (yellow), target DNA (TS, violet), and non-target DNA (NTS, cyan) are also shown. (**B**) Energetic landscape associated to the conformational transition of the Cas9 protein from the apo form to the RNA-bound state, computed using GaMD. The potential of mean force (PMF), which describes the free energy landscape, was computed along the E945-D435 FRET distance and the root mean square deviation (RMSD) with respect to the apo state. The simulations identified three minima: M1 corresponding to the crystallographic apo, M2 that is the RNA-bound structure and M3, which is an intermediate characterized by the solvent exposure of an arginine-rich helix. For selected states, the ensemble averaged electrostatic potential has been computed, revealing the formation of a positively charged cavity (blue) in the intermediate states. Adapted with permission from Palermo et al. (2017). https://www.pnas.org/content/114/28/7260. Copyright 2017 National Academy of Sciences.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

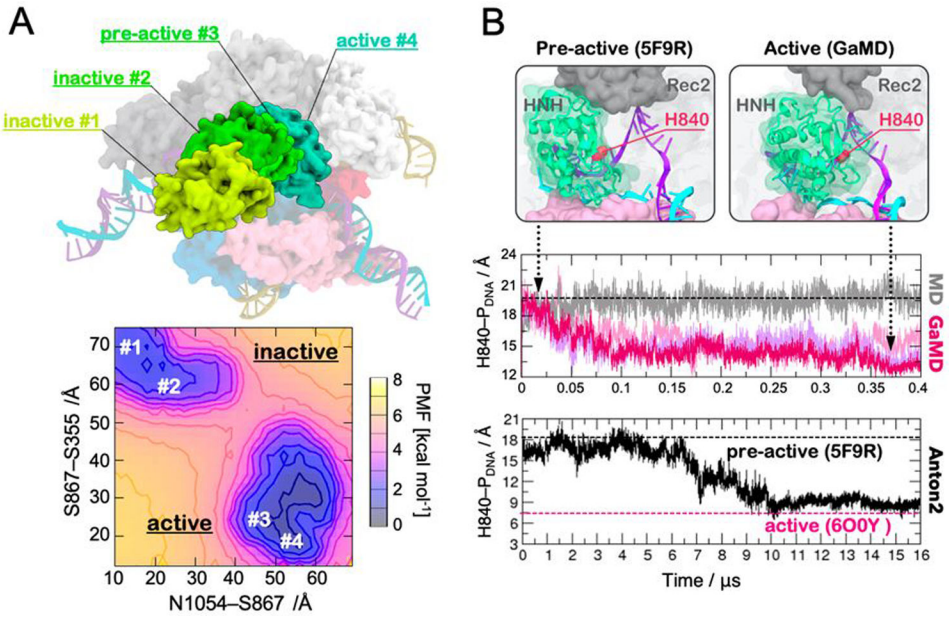


Figure 3.

(**A**) Conformations of the HNH domain (green) in its inactive (#1, #2), pre-active (#3) and active (#4) states, as experimentally determined through single-molecule FRET and structural approaches (top panel). The free energy landscape (i.e., Potential of Mean Force, PMF) associated to the conformational changes of the HNH domain from its inactive to active states is shown in the bottom panel. The minima correspond to the four states experimentally found (top). The PMF was computed along the S867-S355 and N1054-S867 FRET distances. Adapted with permission from Palermo et al. (2017). https://www.pnas.org/content/114/28/7260. Copyright 2017 National Academy of Sciences. (**B**) Conformational change of the HNH domain from its pre-active conformation (captured in the PDB ID: 5F9R, left) to the active state identified through GaMD (right). The active state displays the catalytic residue H983 close to the DNA target strand (TS). The distance between the catalytic H840 and the scissile phosphate (H840–PDNA) has been computed along ~400 ns GaMD (central panel) and ~16 µs of continuous MD using the specialized supercomputer Anton-2 (bottom panel). The black dashed line indicates the pre-active conformation (PDB ID: 5F9R) used as a starting point for MD simulations, while the magenta dashed line indicates the active conformation more recently captured through cryo-EM (PDB ID: 6O0Y). Reprinted with permission from Palermo et al. (2018). Copyright 2018 Cambridge University Press. https://doi.org/10.1017/S0033583518000070.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

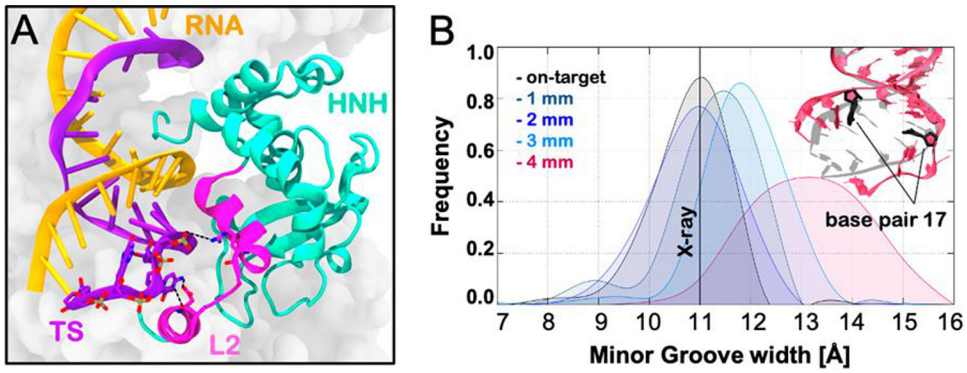


Figure 4.

(**A**) Extended opening of the RNA:DNA hybrid and newly formed interactions with the L2 loop (magenta) of the HNH domain (green), observed during GaMD simulations of CRISPR-Cas9 in the presence of four base pair mismatches at the RNA:DNA hybrid ends. **(B)** RNA:DNA minor groove width computed along MD simulations of CRISPR-Cas9 bound to an on-target DNA (black) and in the presence of one to four mismatches at the hybrid ends. A vertical bar indicates the experimental minor groove width (i.e., 11 Å from X-ray crystallography). The minor groove width has been measured at the level of base pair 17 (shown on the right). Adapted with permission from Ricci et al. (2019). Copyright 2019 American Chemical Society. https://pubs.acs.org/doi/full/10.1021/acscentsci.9b00020. Further permissions related to the material excerpted should be directed to the American Chemical Society.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

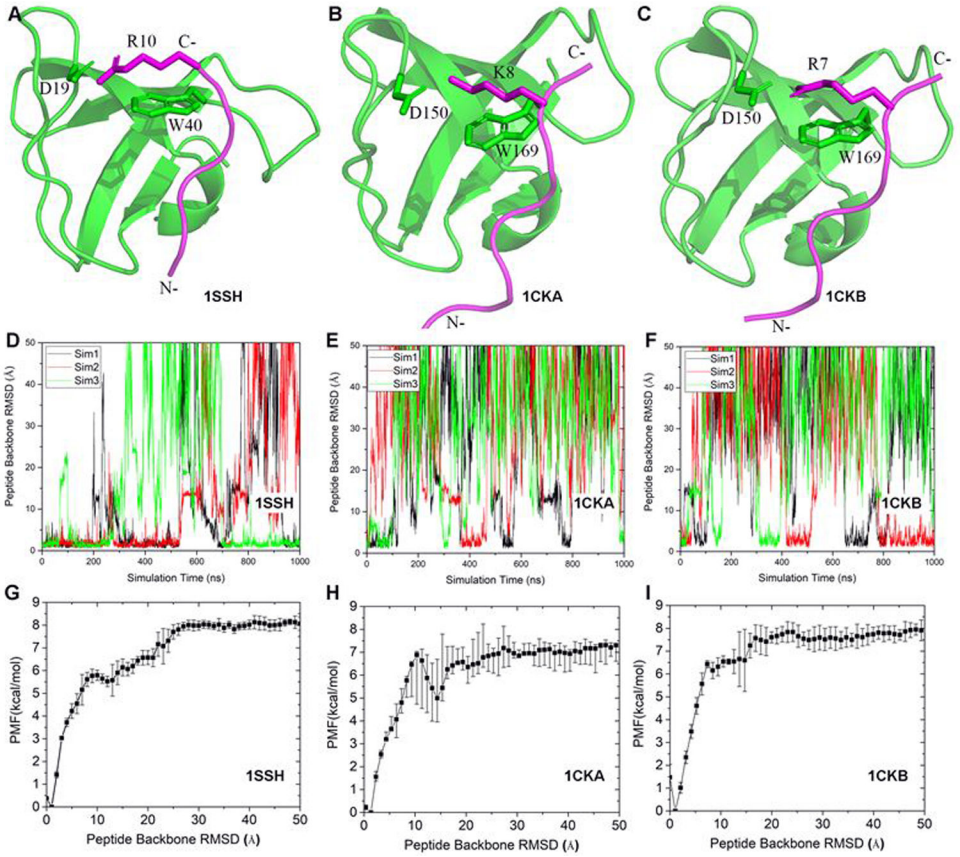


Figure 5.

Pep-GaMD simulations have captured repetitive dissociation and binding of three model peptides to the SH3 domains: (**A**-**C**) X-ray structures of the SH3 domains bound by peptides (**A**) "PAMPAR" (PDB: 1SSH), (**B**) "PPPALPPKK" (PDB: 1CKA) and (**C**) "PPPVPPRR" (PDB: 1CKB). The SH3 domains and peptides are shown in green and magenta cartoon, respectively. Key protein residues Asp19 and Trp40 in the 1SSH structure and Asp150 and Trp169 in the 1CKA and 1CKB structures, and peptide residues Arg10 in the 1SSH structure, Lys8 in the 1CKA structure and Arg7 in the 1CKB structure are highlighted in sticks. The "N" and "C" labels denote the N-terminus and C-terminus of the peptides. (**D**–**F**) time courses of peptide backbone RMSDs relative to X-ray structures with the protein aligned calculated from three independent 1 μs Pep-GaMD simulations of the (**D**) 1SSH, (**E**) 1CKA and (**F**) 1CKB structures. (**G**–**I**) The corresponding PMF profiles of the peptide backbone RMSDs averaged over three Pep-GaMD simulations of the (**G**) 1SSH, (**H**) 1CKA and (**I**) 1CKB structures. Error bars are standard deviations of the free energy values calculated from three Pep-GaMD simulations. Reprinted from "Jinan Wang, Yinglong Miao, *J Chem Phys* 2020, 153:154109", with the permission of AIP Publishing.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

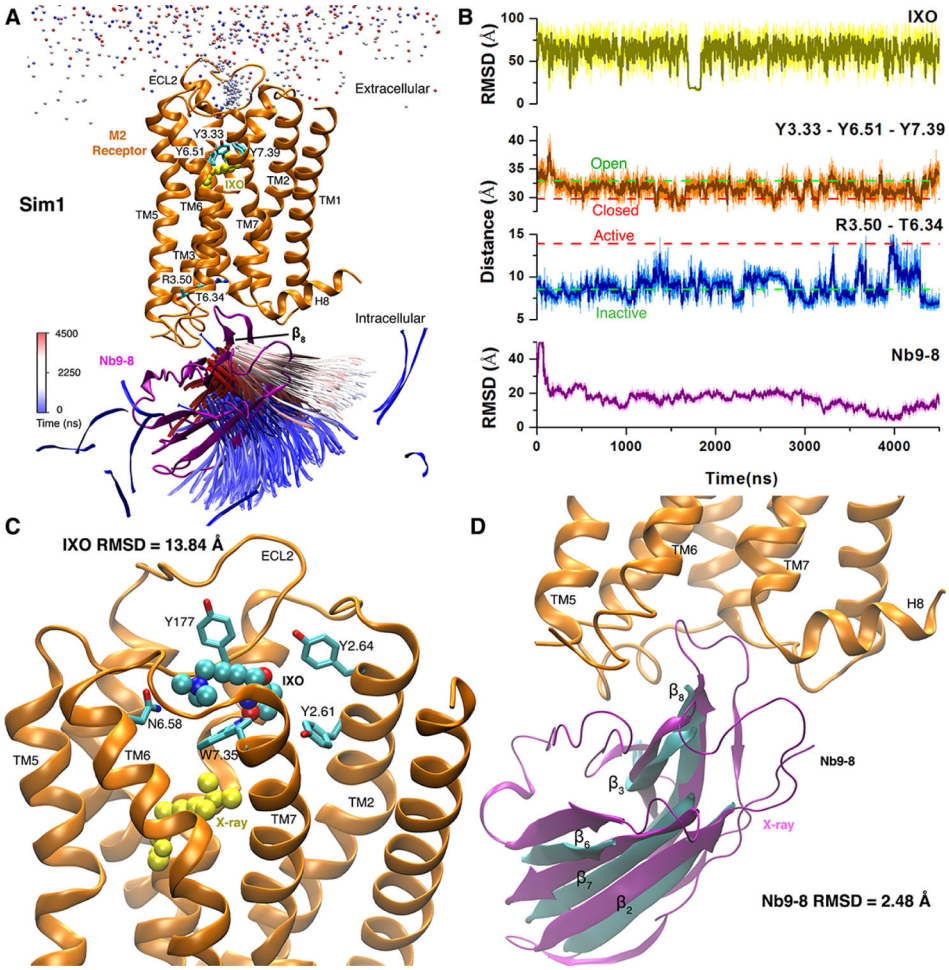


Figure 6.

Binding of agonist IXO and G_i protein mimetic nanobody Nb9–8 to the M2 muscarinic GPCR was captured in one of five GaMD simulations: (**A**) Trajectories of a nitrogen atom in IXO (beads) and the β8 strand of Nb9–8 (ribbons) colored by simulation time in a blue (0 ns)–white (2250 ns)–red (4500 ns) scale. (**B**) RMSDs of the IXO and Nb9–8 relative to the X-ray structure, Tyr104$^{3.33}$-Tyr403$^{6.51}$-Tyr426$^{7.39}$ triangle perimeter and Arg121$^{3.50}$-Thr386$^{6.34}$ distance calculated from the simulation. Dashed lines indicate X-ray structural values of the M2 receptor (3UON: green and 4MQS: red). (**C**) Binding pose of IXO (spheres) in the receptor extracellular vestibule with 13.84 Å RMSD relative to the X-ray conformation (yellow spheres). Residues found within 5 Å of IXO are highlighted in sticks. (**D**) Binding of Nb9–8 (cyan), which exhibits only 2.48 Å RMSD in the protein core (the β2, β3, β6, β7 and β8 strands). X-ray conformations of the M2 receptor and nanobody are shown in orange and purple ribbons, respectively. Adapted with permission from Miao et al. (2018). https://www.pnas.org/content/115/12/3036. Copyright 2018 National Academy of Sciences.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

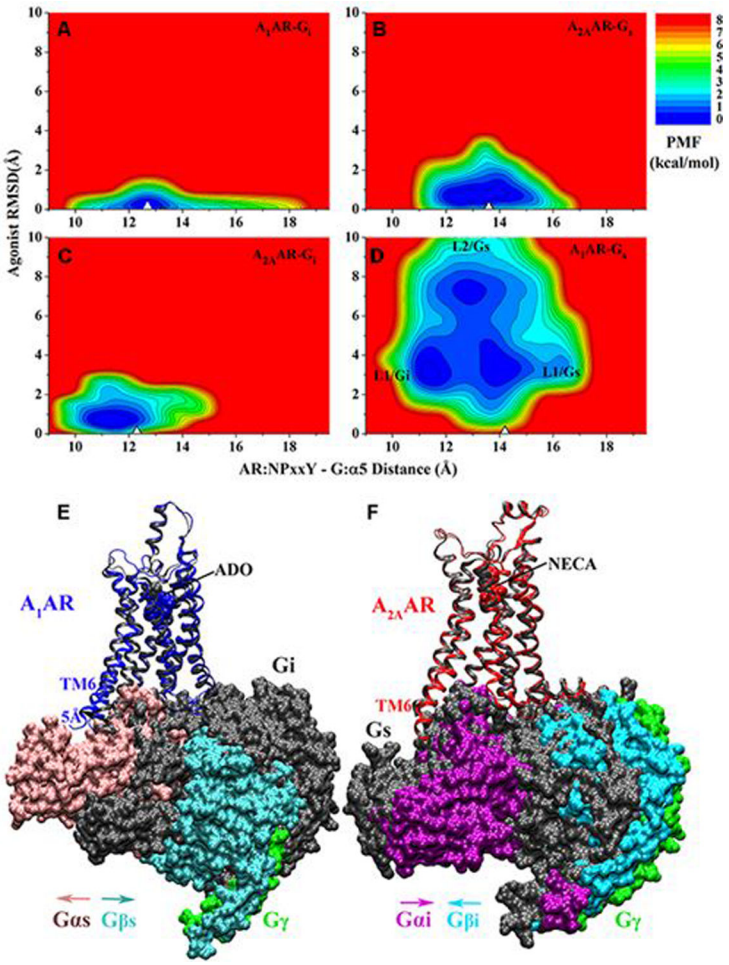


Figure 7.

2D potential of mean force (PMF) profiles of the **(A)** ADO-bound A_1AR-G_i, **(B)** NECA bound $A_{2A}AR$-G_s, **(C)** NECA bound $A_{2A}AR$-G_i and **(D)** ADO bound A_1AR-G_s complex systems regarding the agonist RMSD relative to the cryo-EM conformation and AR:NPxxY-G:α5 distance. The white triangles indicate the cryo-EM or simulation starting structures. Summary of specific AR-G protein interactions: **(E)** the ADO-bound A_1AR prefers to bind the G_i protein to the G_s. The latter could not stabilize binding of agonist ADO in the A_1AR and tended to dissociate from the receptor. **(F)** The $A_{2A}AR$ could bind both the G_s and G_i proteins, which adopted distinct conformations in the complexes. Adapted with permission from Wang et al. (2019). Copyright 2019 American Chemical Society. https://pubs.acs.org/doi/10.1021/acs.jpcb.9b04867. Further permissions related to the material excerpted should be directed to the American Chemical Society.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

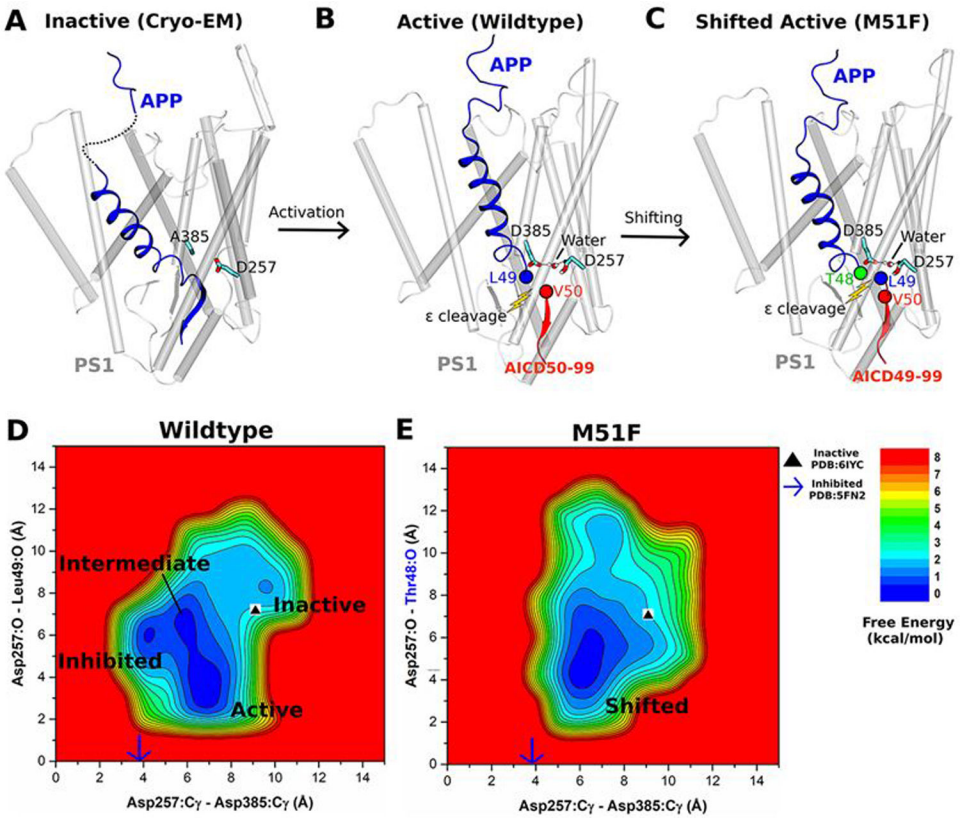


Figure 8.

GaMD simulations revealed the activation and its ε cleavage mechanisms of γ-secretase in the wildtype and mutant APP substrates. Summary of the (**A**) inactive cryo-EM, (**B**) active (wildtype), and (**C**) shifted active (M51F) conformational states of the APP-bound γ-secretase. Distinct AICD products were generated from the wildtype and M51F mutant APP. GaMD free energy profiles of (**D**) wildtype and (**E**) M51F APP-bound γ-secretase regarding the Asp257:Cγ–Asp385:Cγ and Asp257:protonated O–Leu49:O distances. Adapted with permission from Bhattarai et al. (2020). Copyright 2020 American Chemical Society. https://pubs.acs.org/doi/abs/10.1021/acscentsci.0c00296. Further permissions related to the material excerpted should be directed to the American Chemical Society.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

The implemented GaMD algorithms in different MD software packages.

	AMBER CPU version	AMBER GPU version	NAMD	GENESIS
Dihedral GaMD	X	X	X	X
Total boost GaMD	X	X	X	X
Dual boost GaMD	X	X	X	X
Nonbonded dual-boost GaMD	X	X		
Selective GaMD (LiGaMD & Pep-GaMD)		X		