CHAPTER FOUR

Enhanced sampling and free energy calculations for protein simulations

Qinghua Liao*
Science for Life Laboratory, Department of Chemistry-BMC, Uppsala University, Uppsala, Sweden
*Corresponding author: e-mail address: qinghua.liao@kemi.uu.se

Contents

Abstract

Molecular dynamics simulation is a powerful computational technique to study biomolecular systems, which complements experiments by providing insights into the structural dynamics relevant to biological functions at atomic scale. It can also be used to calculate the free energy landscapes of the conformational transitions to better understand the functions of the biomolecules. However, the sampling of biomolecular configurations is limited by the free energy barriers that need to be overcome, leading to considerable gaps between the timescales reached by MD simulation and those governing biological processes. To address this issue, many enhanced sampling methodologies have been developed to increase the sampling efficiency of molecular dynamics simulations and free energy calculations. Usually, enhanced sampling algorithms can be classified into methods based on collective variables (CV-based) and approaches which do not require predefined CVs (CV-free). In this chapter, the

Progress in Molecular Biology and Translational Science, Volume 170
ISSN 1877-1173
https://doi.org/10.1016/bs.pmbts.2020.01.006

theoretical basis of free energy estimation is briefly reviewed first, followed by the reviews of the most common CV-based and CV-free methods including the presentation of some examples and recent developments. Finally, the combination of different enhanced sampling methods is discussed.

1. Introduction

Thanks to the tremendous progress in both computing power and methodology, molecular dynamics (MD) simulation has been widely applied in computational modeling of biomolecules ranging from proteins to membrane lipids and nucleic acids since the first MD simulation of a protein, bovine pancreatic trypsin inhibitor in 1977.[1] In an MD simulation, the trajectory of the interacting atoms of a biological system is determined by solving Newton's equations of motion numerically, where the forces on atoms and the potential energies are usually calculated using molecular mechanics force fields (see chapters "Pairwise-additive and polarizable atomistic force fields for molecular dynamics simulations of proteins" by Justin Lemkul and "Scale-consistent approach to the derivation of coarse-grained force fields for simulating structure, dynamics, and thermodynamics of biopolymers" by Adam Liwo of this book for more information on force fields). Structural and dynamical information of the biomolecules can be extracted from MD trajectories, which can be compared to and used to interpret experimental results determined from, for example, neutron scattering,[2] nuclear magnetic resonance (NMR),[3–5] small-angle X-ray scattering (SAXS)[6] and cryogenic electron microscopy (cryo-EM).[7] Methods and protocols were also developed for using MD simulations to refine structures obtained from NMR spectroscopy,[8] cryo-EM,[9,10] X-ray crystallography,[11] and SAXS.[12]

Currently, MD simulations are limited to tens of microseconds for normal-size systems, and even shorter timescales for larger systems. However, the results obtained from MD simulations are reliable only if the simulations are long enough for the systems to visit all the relevant configurations in phase space. Unfortunately, MD trajectories are mostly not ergodic as many relevant regions in configurational space are left unexplored in MD simulations. In most cases the reason for this is that relevant configurations are separated by high free energy barriers, thus it would take milliseconds or even longer for configurational transitions across these energy barriers to occur. There is often a gap between the timescales than can be covered with MD simulations and those of biological events. This results in insufficient sampling and thus unconverged simulations of biomolecular systems, which also causes the free energy calculations to be unreliable.

To solve the problem of insufficient sampling, many enhanced sampling approaches have been developed during the last decades.[13–15] Usually, these methods can be divided into two classes: one class is based on adding bias potentials along predefined collective variables (CVs) of the system (CV-based), while the methods of other class do not need predefined CVs (CV-free). The CV-based sampling methods greatly improve the convergence of free energy calculations when the CVs are well selected. Umbrella sampling (US),[16] metadynamics (MetaD),[17] steered molecular dynamics (SMD),[18] adaptive biasing force (ABF),[19] conformational flooding,[20] string method,[21] and temperature accelerated MD (TAMD)[22] are widely applied methods belonging to this class of enhanced sampling methods. However, the selection of the proper CVs is a nontrivial task for many systems. In such cases, CV-free sampling methods are advantageous as they allow sampling biomolecular configurations without requiring a priori knowledge of the system. They are useful for exploring possible biomolecular structural transition pathways and determining possible unknown intermediate states of the systems. Replica exchange (RE),[23] or parallel tempering (PT), self-guided molecular Langevin dynamics (SGLD),[24,25] and accelerated molecular dynamics (aMD)[26] are well-known methods of this second class of sampling methods. For larger or more complex systems, a combination of enhanced sampling methods is a powerful approach for efficient sampling of conformational transitions, such as replica exchange umbrella sampling (REUS)[27,28] or bias-exchange metadynamics (BE-MetaD).[29]

In this chapter, three CV-based (US, MetaD, SMD) and two CV-free (REMD, aMD) enhanced sampling methods for biomolecular simulations are reviewed by providing introductions to their theoretical basis, presenting illustrative applications, and discussing recent developments. To better understand these methods, various methods for the calculation of free energies from simulations are introduced first. The end of the chapter is focused on methods that combine different enhanced sampling algorithms, followed by an overview of the availability of the different methods in the common biomolecular MD simulation engines.

2. Collective variable and free energy

For MD simulations, collective variables (CVs) instead of atomic coordinates are often used to quantify certain properties of the simulated system, thereby reducing the highly dimensional problem. A CV is a function of atomic coordinates which can describe a certain motion or transition of interest. For example, a distance between two atoms is a simple CV which

can be used to describe bond forming and breaking, while a dihedral angle is usually used to characterize an isomerization process as a CV. For a proton transfer process, the distances between the proton and each of the proton donor and acceptor could be two reasonable CVs. In short, a CV is coarse-grained description of the system.

With a defined CV s, the probability of observing the system at a given CV point s_i is the weight of all configurations q mapping to s_i:

$$P(s_i) = \langle \delta[s(q) - s_i] \rangle \tag{1}$$

where the Dirac δ function picks only s_i from all possible $s(q)$, and the angular bracket $\langle \cdot \rangle$ denotes an ensemble average. A free energy can be obtained from the probability by:

$$G(s_i) = -k_B T \ln \langle \delta[s(q) - s_i] \rangle \tag{2}$$

where k_B is Boltzmann constant and T is the temperature. Based on the free energy landscape $G(s)$, the energetic cost of a maximally efficient transition from one region of the CV space to another one can be measured. Fig. 1 shows an illustrative free energy landscape along a CV involving a conformational transition between states A and B. State A is the global minimum, and state B is a metastable state. States A and B are connected by an intermediate region with high energy, transition state C. The transition rate between A and B is heavily dependent on the height of the free energy barrier compared with the thermal energy $k_B T$, according to the Arrhenius-like equation:

$$\nu_{A \to B} = \nu_0 \exp\left(-\frac{\Delta G^{\ddagger}}{k_B T}\right) \tag{3}$$

where ν_0 is a prefactor, and $\Delta G^{\ddagger}$ is the free energy difference between state A and the transition state C:

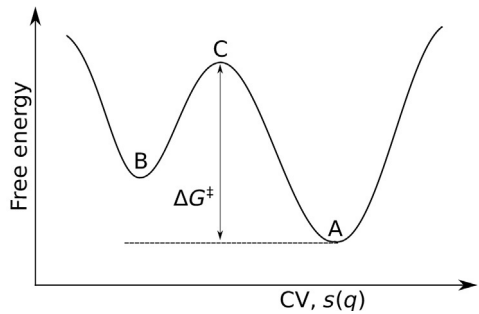

Fig. 1 A double-well potential model. The free energy is plotted against a CV. The stable state is termed as A, the metastable one is termed as B, while the transition state is C.

$$\Delta G^{\ddagger} = G_C - G_A = -k_B T \ln \frac{\langle \delta[s(q) - s_C] \rangle}{\langle \delta[s(q) - s_A] \rangle} \tag{4}$$

If an MD simulation is perfectly ergodic (long enough), all possible configurations in the CV space would be sampled. In the case of the double-well potential (Fig. 1) this would imply that enough transitions between states A and B crossing C were observed. This would allow one to determine how often each point s_i in the CV space is sampled, and the free energy can be computed by normalizing this histogram into a probability $P(s_i)$ via Eqs. (1) and (2). Histograms are concerned with computing CV distributions and averages. They are related to the fluctuations in the ensemble of interest, and further enable us to calculate free energies and entropies. There are many approaches developed to estimate free energy differences between discrete states (histograms).

The Zwanzig equation is the most well-known method historically for calculating free energies.[30] The classical free energy between two states characterized by Hamiltonians $H_i(q)$ and $H_j(q)$ can be estimated as:

$$\begin{aligned} \Delta G_{i \to j} &= -k_B T \ln \left\langle \exp \left[-\frac{H_j(q) - H_i(q)}{k_B T} \right] \right\rangle_i \\ &= -k_B T \ln \left\langle \exp \left[-\frac{\Delta H(q)}{k_B T} \right] \right\rangle_i \end{aligned} \tag{5}$$

This method is denoted as EXP, for exponential averaging. However, when the phase space overlap of the two states is too low, EXP becomes less accurate. The statistical uncertainty of EXP grows exponentially when the overlap drops. When the overlap between two states is poor, introducing a number of intermediate states can efficiently improve the computation of free energy differences.

A significantly more robust estimator was originally proposed by Bennett[31] in 1976, and further improved by Shirts et al.[32] The resulting Bennett acceptance ratio (BAR) method accounts for the expectations computed at both states:

$$\Delta G_{ij} = -\beta^{-1} \ln \frac{Q_j}{Q_i} = \beta^{-1} \ln \frac{\langle \alpha(q) \exp[-\beta H_i(q)] \rangle_j}{\langle \alpha(q) \exp[-\beta H_j(q)] \rangle_i} \tag{6}$$

where β equals to $(k_B T)^{-1}$, Q_i is the canonical partition function of state i, and $\alpha(q)$ is a weight function, which can be any everywhere-finite function of the coordinates q. Bennett used variational calculus to obtain the values of

$\alpha(q)$ minimizing the variance of the free energy.[31] The final result is an implicit function of the free energy and the number of samples n_i and n_j at each states have to fulfill the equation:

$$\sum_{i=1}^{n_i} \frac{1}{1+\exp\left(\ln\left(n_i/n_j\right)+\beta\Delta H_{ij}-\beta\Delta G\right)} - \sum_{j=1}^{n_j} \frac{1}{1+\exp\left(\ln\left(n_j/n_i\right)-\beta\Delta H_{ji}+\beta\Delta G\right)} = 0 \tag{7}$$

When multiple intermediate states are sampled, the weighted histogram analysis method (WHAM)[33] can be applied to combine information from all intermediates to obtain an optimal estimation of the free energy differences between all pairs of intermediates. WHAM is an improved version of the histogram weighted technique which was first proposed by Ferrenberg and Swendsen.[34] WHAM has been proven to be the lowest uncertainty method to calculate the free energy for samples collected from discrete intermediate states.[35] As the WHAM approach is widely used together with umbrella sampling, more details about WHAM are presented in the Umbrella Sampling section.

The multistate Bennett acceptance ratio (MBAR)[36] is a direct extension of BAR, which allows for accessing the data from all states and predicting the free energy at unsampled regions. MBAR is equivalent to BAR in the limit that only two states are assessed. The equation for the corresponding free energy calculation can be considered as a zero-width bin WHAM:

$$G_i = -\beta^{-1} \ln \sum_{k=1}^{K} \sum_{n=1}^{N_k} \frac{\exp\left[-\beta H_i(q_{kn})\right]}{\sum_{k'=1}^{K} N_{k'} \exp\left[\beta G_{k'} - \beta H_{k'}(q_{kn})\right]} \tag{8}$$

Here, k and k' are indexes of K, which is the number of states, and N_k and $N_{k'}$ are the number of configurations at state k and k', respectively. G_i is the free energy of state i, and $H_i(q_{kn})$ is the Hamiltonian at state i of configuration n at state k. To be noted, only the free energy difference $\Delta G_{ij} = G_j - G_i$ is meaningful.

In many cases, simulations are trapped in minima (like A and B in Fig. 1)—and often the configurations from where simulations were started are minima—and barely cross the barriers to visit other metastable states. Therefore, normal MD simulations are insufficient for free energy calculations. MD simulations with enhanced sampling are thus necessary to speed up the crossing of free energy barriers with limited computational resources.

3. CV-based sampling

CV-based enhanced sampling is also called constrained enhanced sampling. With CV-based enhanced sampling, a bias potential as a function of chosen collective variables (CVs) is added to the potential energy of the simulated system to improve the sampling efficiency. As a result, the sampled space is largely constrained to the desired CVs space. The selection of CVs is critical to the biased sampling. More details about selecting CVs are provided in previous reviews.[37,38] Bias potentials can be added to the simulated system in a variety of ways to obtain efficient sampling. In this section, three widely used CV-based methods are discussed.

3.1 Umbrella sampling

Umbrella sampling (US) was developed by Torrie and Valleau[16] in 1977. It is one of the most widely used approaches to overcome free energy barriers in free energy calculation. As shown in Fig. 2, umbrella sampling uses a series of independent windows along a selected CV to model a conformational transition. For each of the windows, a simple harmonic potential $\Delta V_i(q)$, whose strength defined by the force constant k is chosen depending on the potential energy at each of the reference points s_i along the CV $s(q)$, is added to the system's Hamiltonian:

$$\Delta V_i(q) = \frac{k}{2} \times (s(q) - s_i)^2 \tag{9}$$

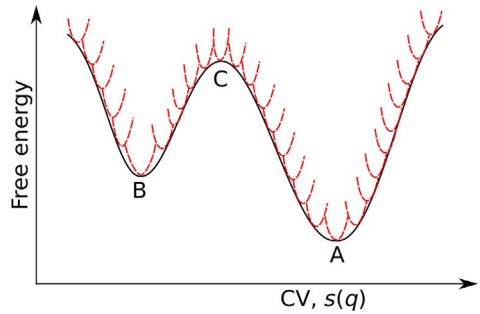

Fig. 2 Schematic illustration of the umbrella sampling method. The *red dash lines* represent the harmonic bias potentials that are added to the system Hamiltonian at different CV points (windows) along the CV space.

and MD simulations are then performed for all windows. When these simulations are finished, WHAM[33] is the most popular method to combine the statistics from all the independent windows. The WHAM equations are expressed as follows:

$$P_u(s) = \frac{\sum_{l=1}^{N} h_l(s)}{\sum_{k=1}^{N} n_k \exp\left([F_k - V(s)]/k_B T\right)} \tag{10}$$

$$F_k = -k_B T \ln \sum_{q} P_u(s) \exp\left(-V(s)/k_B T\right) \tag{11}$$

where $P_u(s)$ is the unbiased probability distribution along s, N in the number of umbrella windows, l and k are indexes of the umbrella windows, $h(s)$ is the counts at bin s, n_k is the number of data points from window k, and F_k is the factor to be determined for window k. The two equations are coupled and solved iteratively until a self-consistent solution is reached. More details of the WHAM iteration method were previously described by Kumar et al.[33] and Roux.[39] WHAM was shown to be equivalent to BAR when only two states are assessed. Moreover, other WHAM variants were also developed based on maximum-likelihood[40] and Bayesian methods.[41,42] Alternatively, the umbrella integration (UI)[43] method and multistate Bennett acceptance ratio (MBAR)[36] can also be used to recover the unbiased free energy landscape from umbrella sampling. More details about umbrella sampling are discussed by Kästner.[35]

Due to its quick convergence and the fact that the MD simulations for the different windows can be run independently from each other, allowing to add further windows to the system to improve convergence, umbrella sampling has been demonstrated great successes by a tremendous amount of studies that employed umbrella sampling in a wide range of biological systems including protein folding,[44] conformational changes of large proteins, protein–protein interactions,[45] protein–lipid association,[46] ligand–protein binding,[47] DNA conformational change,[48] enzyme catalysis.[49,50] A recent study where lipid-protein binding was studied using US[46] is chosen as illustrative example. The authors of this study applied US to characterize the free energy profile, also called potential of mean force (PMF), for the binding of a cardiolipin (CL) molecule to the mitochondrial transport protein ANT within membranes—modeled by a POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) membrane—using coarse-grained (CG) MD simulations. The CV was defined as the distance between the CG glycerol particle of CL and the backbone bead of the proline residue in a conserved motif of the protein family (Fig. 3A). MD simulations of 7 μs were performed for 32 windows linearly spaced between 1.6 and 4.5 nm along the chosen CV with a force constant

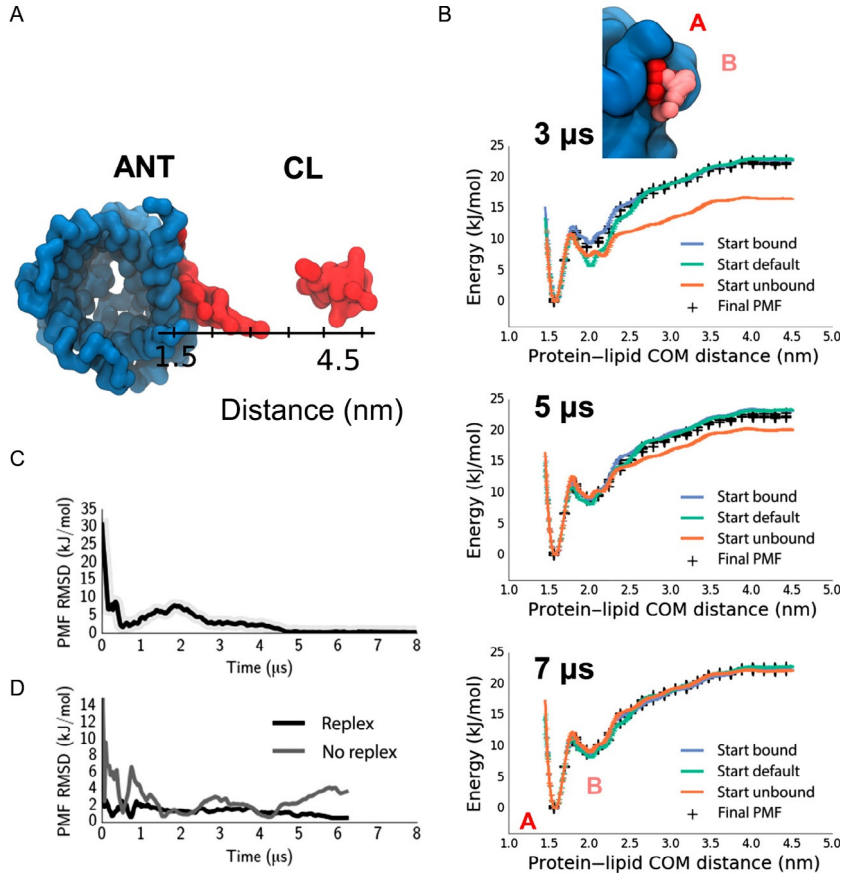

Fig. 3 Cardiolipin binding to ANT, providing an example of calculating a free energy profile for a lipid interacting with a specific binding site on an integral membrane protein. (A) Representative structures of the closest ($d = 1.5$ nm) and furthest ($d = 4.5$ nm) protein–lipid COM distance in the US simulation. (B) PMFs obtained from three different US schemes with different initial configurations (default in *cyan*, unbound in *orange*, bound in *blue*), computed after 3, 5, and 7 μs of the simulation (within each case the first half of the simulation was discarded as equilibration). The final PMF (indicated by *black crosses*) was calculated from the two US replica exchange simulations, corresponding to the starting bound and unbound initial conditions. (C) RMSD between the two PMFs started from all bound and all unbound configurations as a function of time. (D) RMSD between the final PMF and US simulations started from the default initial condition and either allowing for replica exchange or without exchange. *Reprinted with permission from Domański J, Hedger G, Best RB, et al. Convergence and sampling in determining free energy landscapes for membrane protein association. J Phys Chem B. 2017;121:3364–3375, copyright 2017 American Chemistry Society.*

of $k = 1000$ kJ·mol^{-1} ·nm^{-2} (see Eq. 9). Three US schemes were tested, termed as default, bound and unbound. The default scheme corresponds to windows starting from the bias center of each window, which were generated by steered MD simulations (see Section 3.3 for details). The other two schemes were started with all windows being in the bound state (protein–lipid center of mass distance $d = 1.5$ nm) or in the unbound state ($d = 4.5$ nm). The latter two schemes are special as here every 200 ps it was attempted to exchange the configurations between windows according to the Boltzmann criterion, i.e., the same idea as applied in replica exchange MD (see Section 4.1 for details) was used. The unbiased free energy profiles from the three US simulations were determined by WHAM. Two distinct minima were observed, at $d = 1.5$ nm and $d = 2.0$ nm with well-depths of 20 and 10 kJ·mol^{-1}, respectively (Fig. 3B). After 7 μs, the PMF profiles had well converged. The progress of convergence was also monitored by calculating the root-mean-square difference (RMSD) between the PMFs derived from the US simulation started with all windows bound and that started with all windows unbound. This PMF RMSD becomes less than 1 kJ·mol^{-1} after 5 μs (Fig. 3C). Up to this difference, the PMF is thus insensitive to the initial configuration. The authors also demonstrated that exchange between the windows greatly improved the sampling efficiency and speed up the convergence of US (Fig. 3D).

In practice, the harmonic potentials for all the windows need to be tuned manually, so that reasonable positional overlap between the windows along the CVs can be obtained. This implies that more windows are needed, the higher the force constants of the harmonic potentials are, which is especially required to ensure the sampling at high energy regions. The results obtained from US can be very sensitive to the starting configurations as the sampling is restrained to a simple pathway, other important pathways could be neglected while the free energy profiles may appear to be converged. Furthermore, when the number of CVs becomes larger ($\geq$3) or the system of interest is complicated, the computation will be demanding, and it will be difficult to determine accurate results. Thus, self-learning adaptive US[51] was proposed. Moreover, the replica exchange algorithm was combined with umbrella sampling (REUS)[27,28] to further improve the sampling efficiency.

3.2 Metadynamics

Metadynamics (MetaD) was originally developed by Parrinello and coworkers to enhance the sampling of rare events.[17] In MetaD simulation, an external

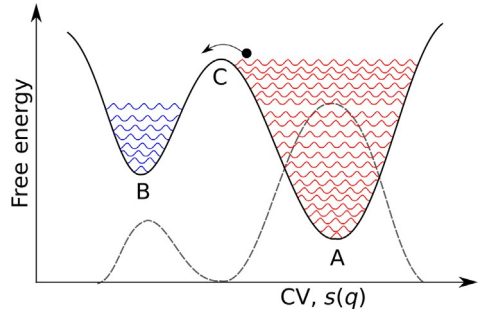

Fig. 4 Schematic illustration of metadynamics simulation. In MetaD, history-dependent Gaussian-type bias potentials (first *red* Gaussians, then *blue* Gaussians) are deposited along the CV space. The deposited potentials encourage the system to explore disfavored configurations. As shown here, when basin A is filled up with Gaussians, the system can explore configurations at with high energies, followed by crossing the transition state C and going to basin B. After basin B is also fully filled up, the free energy landscape along CV space is flattened. The summation of the deposited bias potentials provides an estimation of the negative free energy landscape (*gray dash line*).

history-dependent bias potential is imposed on the system during the simulation, the bias potential which is a function of the CVs can be written as a sum of Gaussians added along the CVs space to encourage the system to visit configurations which have not already been sampled (Fig. 4). The Gaussian-type bias potential $V(s(q), t)$ can be expressed as:

$$V(s(q),t) = \sum_{k\tau<t} W(k\tau) \exp\left(-\sum_{i=1}^{d} \frac{(s_i - s_i(q(k\tau)))^2}{2\sigma_i^2}\right) \tag{12}$$

where τ is the Gaussian deposition rate, σ_i is the width of the Gaussian for the ith CV, and $W(k\tau)$ is the height of the Gaussian at simulation time $k\tau$. The effect of the positive bias potential is to push the system out of any local minimum in the CVs space, to encourage visiting unexplored configurations. In the long time limit, the bias potential will converge to the negative free energy as a function of the CVs, thus the free energy landscapes can be reconstructed by reversing the imposed bias potential:

$$V(s, t \to \infty) = -G(s) + C \tag{13}$$

where C is a constant. The added bias potential leads to highly efficient sampling in which the energy barriers separating different local minima can be easily crossed. In standard MetaD, the height of the deposited Gaussian-type bias potential is constant during the simulation. As a result, the estimation of

the free energy landscape calculated from the bias potential oscillates around the real values, and there is also a risk of driving the system into physically irrelevant regions of phase space. To solve this problem of convergence difficulties, the well-tempered metadynamics (WT-MetaD)[52] was developed, where the height of the added Gaussian potentials decreases with simulation time according to:

$$W(k\tau) = W_0 \exp\left(-\frac{V(s(q(k\tau)), k\tau)}{k_B \Delta T}\right) \tag{14}$$

where W_0 is the initial height and ΔT is a parameter with the dimension of a temperature. This rescaling technique indeed ensures that the bias potential converges more smoothly though it does not compensate the underlying free energy:

$$V(s, t \to \infty) = -\frac{\Delta T}{T + \Delta T} G(s) + C \tag{15}$$

where T is the temperature of the system. When $\Delta T = 0$, it is identical to standard molecular dynamics, while $\Delta T \to \infty$ corresponds to standard MetaD. For WT-MetaD as implemented in PLUMED,[53,54] the corresponding input parameter is termed "bias factor" γ:

$$\gamma = \frac{T + \Delta T}{\Delta T} \tag{16}$$

MetaD is robust and easy to use, it is one of the most widely used enhanced sampling methods for free energy calculations, and continues to be applied in simulations by the community. Comparing to methods such as US,[16] a major advantage of MetaD is that no a priori knowledge of the end states is required and that multiple CVs can be used, it is very helpful to explore possible pathways while umbrella sampling is strictly constrained to certain pathways. However, the efficiency of MetaD scales poorly with the number of CVs used and in practice is limited to three CVs. To some extent, the accuracy and convergence of MetaD are also depending on the choices of the parameters of the Gaussians ($W(k\tau)$, σ in Eq. (12) and γ in Eq. (16)), but there is no universal method to choose the parameters, which is one disadvantage of MetaD.

A lot of variants were further developed to address the issues and to further improve sampling efficiency including metadynamics with parallel tempering (PTMetaD),[55,56] metadynamics with multiple walker (MW-MetaD),[57] funnel metadynamics,[58] and bias-exchange metadynamics (BE-MetaD).[29]

More recently, parallel bias metadynamics (PBMetaD)[59] was developed to enhance the crossing of high energy barriers by applying multiple parallel bias potentials in low dimensionality. A newly proposed infrequent metadynamics (InMetaD)[60] can be used to accurately predict the long time scale kinetics, and it was successfully used to estimate ligand unbinding kinetics.[61–64] Wang et al. further developed this scheme termed as frequency-adaptive metadynamics (FaMetaD).[65] More details on the theory and applications of MetaD simulation were presented in previous reviews.[37,38,66]

MetaD has been successfully applied to a wide range of biological problems, including unbinding kinetics of drug molecules[61–64], exploration of conformational changes of biomolecules,[67,68] predictions of ligand binding poses[69] and enzyme catalysis.[70,71] In one of our recent studies,[68] we applied both Hamiltonian replica exchange molecular dynamics (HREX)[72] and bias-exchange metadynamics (BE-MetaD)[29] simulations to study the conformational dynamics of loop 6 of triosephosphate isomerase (TIM), which is critical to the catalytic efficiency of this enzyme (5A and B). TIM catalyzes a simple reversible isomerization of dihydroxyacetone phosphate (DHAP) and (R)-glyceraldehyde-3-phosphate (GAP).[73,74] Loop 6 prefers the closed state when the substrate DHAP is bound to TIM, but the open state when no substrate is bound. For the reaction to occur loop 6 needs to be closed, and it has to open to release the product when the reaction is finished.[73,74] In the HREX simulations, loops 6 (Pro166-Ala176) and 7 (Tyr208-Ser211) were chosen as the hot regions (see Section 4.1 for the theory of HREX). The simulations were carried out using 8 replicas with scaling factors exponentially ranging from 1.0 to 0.6, corresponding to an effective temperature range from 300 to 500 K. Each replica was sampled for 200 ns for apo-TIM and for 300 ns for DHAP-TIM. Our HREX simulations failed to capture the closed state of loop 6. The closed conformation of loop 6 appeared to be briefly sampled in the DHAP-TIM complex (Fig. 5D), due to the fact that the simulations were started from the closed state, and the loop opened within very short simulation time and did not close again. Then we turned to BE-MetaD simulations, in which 7 biased and 1 unbiased replicas were simulated. The distances between the backbone atoms (N, C_α, C, O) of each of the 7 residues Ala169-Ala175 of loop 6 and the backbone atoms of the residues of loop 7 (Tyr208-Ser211) were used to train the native contacts[75] (ranging from 1 to 0) as the 7 CVs in the 7 biased replicas. For the MetaD simulation on each of the 7 replicas, a time-dependent Gaussian-type potential was added to the 7 CVs every 2 ps with an initial height of 0.1 kcal·mol^{-1}, which was gradually decreased with a

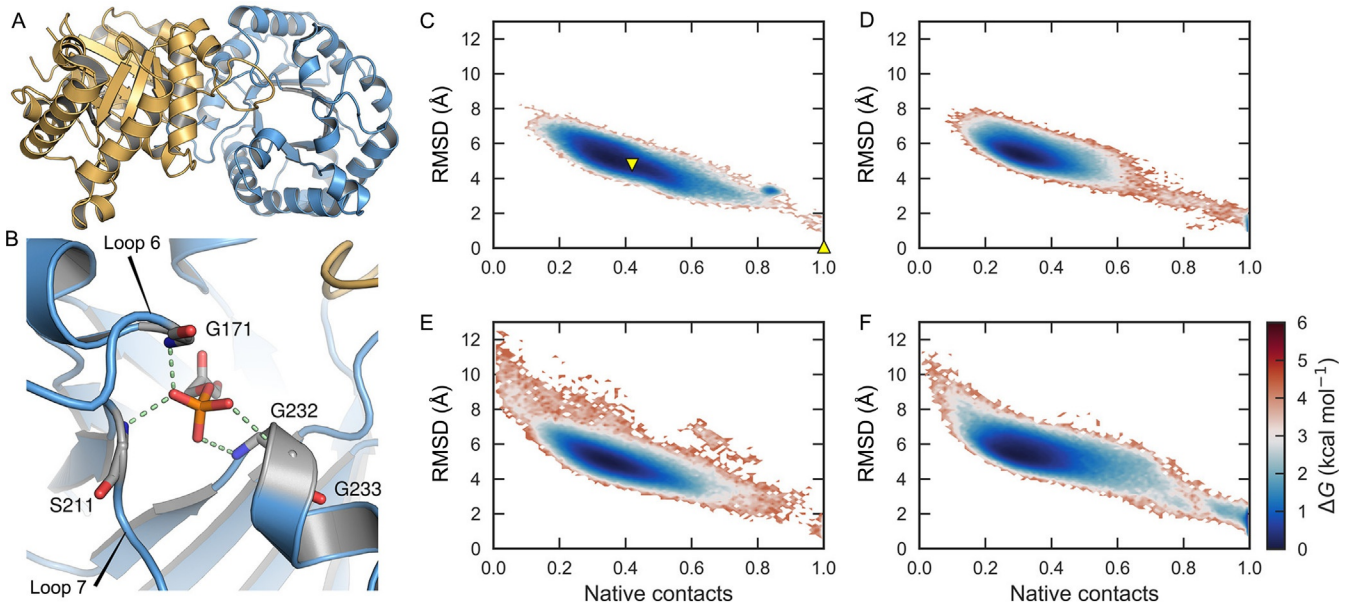

Fig. 5 (A) The overall structure and (B) active site with bound DHAP of TIM (PDB ID: 1NEY[76,77]), (C) Relative free energy surfaces ($T = 300$ K) calculated from (C) and (D) HREX and (E) and (F) BE-MetaD simulations of (C, E) apo-TIM and (D) and (F) DHAP-TIM. The surfaces are defined in terms of the native contacts between loops 6 and 7 and the RMSD of loop 6 relative to its conformation in the crystal structure of DHAP-TIM. In (C) the open and closed states of the loops are projected from chains A of crystal structures 1YPI (▽) and 1NEY (△), respectively. *Reprinted with permission from Liao Q, Kulkarni Y, Sengupta U, et al. Loop motion in triosephosphate isomerase is not a simple open and shut case. J Am Chem Soc. 2018;140:15889–15903, copyright 2018 American Chemistry Society.*

bias factor γ of 10. The width of the Gaussians was set at 0.05. Exchanges between replicas were attempted every 2 ps, and each replica was sampled for 500 ns. Unlike the HREX simulation, the BE-MetaD simulations captured the opening and closing of loop 6 in the presence of DHAP (Fig. 5F). Based on the BE-MetaD sampling, we also explored the influence of the loop 6 conformational dynamics on the catalytic activity of TIM.[68]

3.3 Steered molecular dynamics

Inspired by atomic force microscopy (AFM) experiments, steered molecular dynamics (SMD) was developed.[18,78,79] SMD is often applied to drive the system from one initial state to a final state by pulling in one or multiple directions. In SMD simulations, an external bias potential, which is usually a function of the selected CVs and simulation time, is added to the Hamiltonian of the system to drive it from one state to another:

$$V_\lambda(q,t) = \frac{k(t)}{2}(s(q,t) - \lambda(t))^2 = \frac{k(t)}{2}(s(q,t) - s_0 - \nu t)^2 \qquad (17)$$

where $k(t)$ is the force constant of the harmonic potential, ν is the pulling speed, $s(q)$ are the chosen CVs to define the pulling directions, λ is an external parameter that is correlated with $s(q)$. SMD is a nonequilibrium sampling method, but the free energy differences between states i and j driven by SMD can also be calculated by measuring the work $W_{i\to j}$ done on the system as a result of pulling.[78,80,81] To this end, the Jarzynski equation[80] is used:

$$\Delta G = -\beta^{-1} \ln \left\langle \exp\left(-\beta W_{i\to j}\right)\right\rangle_0 \qquad (18)$$

Later on, Crooks proposed a new version[82]:

$$\Delta G = -\beta^{-1} \ln \frac{\left\langle f(W_{i\to j})\right\rangle_i}{\left\langle f(W_{j\to i})\right\rangle_j} \qquad (19)$$

where $f(W_{i\to j})$ denotes any finite function of the work. The pulling causes an instantaneous transformation, where the instantaneous work is simply the change in potential energy. Thus, the Jarzynski equation can be considered identical to EXP in Eq. (5), and the Crooks relation can be used to derive the BAR equation (see Eq. (6)) for nonequilibrium work.[32,82]

The SMD approach is widely applied in studies of conformational changes of proteins,[83] unbinding processes of ligands from their binding pockets, exploring possible ligand unbinding pathways,[84,85] and the free energy profiles along dissociation pathways can be evaluated.[86,87] Its role in rational

drug design has been reviewed.[88] Furthermore, SMD is powerful in exploring the CVs space in short simulation time, thus it is widely applied together in combination with US and MetaD simulations, for example, to prepare starting conformations for umbrella sampling.[46,89,90] An adaptive SMD method was also proposed,[91] where the idea is that the initial configuration for a given step is obtained from the previous step of the trajectories. To be noted, when the CV is a root-mean-square deviation (RMSD), SMD is then also called targeted MD (TMD),[92] which was initially proposed to search pathways of conformational transitions.

4. CV-free sampling

Although CV-based enhanced sampling algorithms can greatly extend the timescales of MD simulations, predefined CVs are required to precisely describe the actual biological events in these methods, and they usually suffer from hidden energy barriers. It is nontrivial to determine the optimal CVs in MD simulations, especially when the transition processes are not straightforward. The simulations can be very difficult to converge when the CVs are not chosen properly. An example of a hidden barrier is shown in Fig. 6; there is only one basin along CV1, but two basins along CV2. If only CV1 is chosen for biasing sampling, the convergence of CV2 cannot be guaranteed while CV1 can be well sampled. The issue of hidden barriers and thus hidden CVs is explained in more details in other reviews.[14,38] CV-free enhanced sampling techniques, such as replica exchange molecular dynamics and

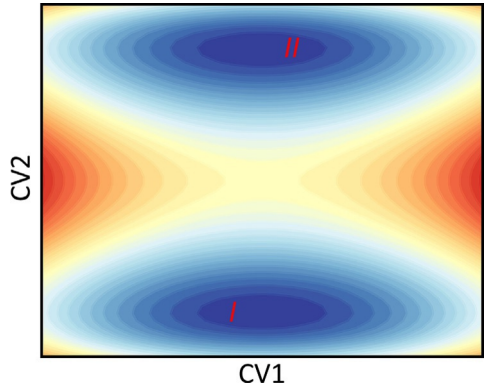

Fig. 6 An illustration of a hidden energy barrier along CV2 between the two basins, I and II.

accelerated molecular dynamics can be used to avoid this problem. Opposite to CV-based enhanced sampling methods, predefined CVs are not required in CV-free enhanced sampling approaches, thus a priori knowledge about the biological process to be studied is not necessarily needed. In Sections 4.1 and 4.2, brief overviews of the just mentioned CV-free enhanced sampling approaches are presented.

4.1 Replica exchange molecular dynamics

Replica exchange molecular dynamics (REMD),[23] also termed as parallel tempering (PT) was developed to enhance conformational exploration. In REMD, as shown in Fig. 7, several independent copies of the target system, which are called replicas, are simulated in parallel at different temperatures. At regular time intervals, exchanges of either the temperatures or the configurations of neighbored replicas are attempted. An exchange between replicas i and j is only accepted when the Metropolis criterion is satisfied:

$$P(q_i \leftrightarrow q_j) = \min\left(1, \exp\left[\left(\frac{1}{k_B T_j} - \frac{1}{k_B T_i}\right)[H(q_j) - H(q_i)]\right]\right) \qquad (20)$$

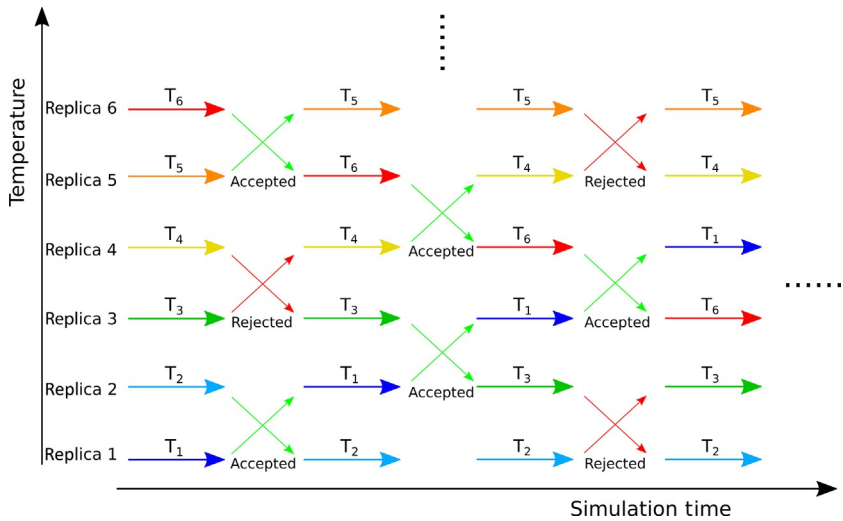

Fig. 7 Schematic of the replica exchange molecular dynamics protocol. In REMD, a number of MD simulations called replicas (six shown here) are performed in parallel and with different temperatures. At recurring intervals, temperatures, or coordinates of neighboring replicas are exchanged with a probability that meets the Metropolis criterion.

where $H(q_i)$ and and $H(q_j)$ are the potential energies for replicas i and j, respectively, and T_i and T_j are the temperatures of replicas i and j, respectively. Random walks of the replicas in the temperature ladder are realized, each replica experiences both high and low temperatures. The sampling of the conformational space is thus enhanced by transferring configurations which are only accessible at high temperatures to the replicas at lower temperatures, which in turn allows one to derive accurate thermodynamics information or free energy landscapes. The main issue of temperature REMD (T-REMD) is that the number of replicas greatly increases as the simulated systems become large, thus it is computationally demanding for complicated biological systems.

In T-REMD, the replicas differ in their temperature at which the MD simulations are performed. However, differences in the replicas are not limited to that. Any control parameter can be changed; for instance, also the Hamiltonian of the simulated system can be changed for the different replicas.[93,94] In a more generalized way, every replica is simulated at a different Hamiltonian and a different temperature (HT-REMD); the acceptance criterion is then expressed as:

$$P(q_i \leftrightarrow q_j) = \min\left[1, \exp\left(\frac{H_i(q_i) - H_i(q_j)}{k_B T_i} + \frac{H_j(q_j) - H_j(q_i)}{k_B T_j}\right)\right] \tag{21}$$

Various ways for modifying the Hamiltonian in replica exchange simulations have been proposed[72,95–99] to address the issues of T-REMD and promote the sampling efficiency. Solute tempering for simulations of solvated systems was an excellent idea among of them. In replica exchange with solute scaling (REST),[96,97] the solute–solute and solute–solvent interactions are scaled by different scaling factors in the different replicas, while solvent–solvent interactions are left unchanged. Thus a relative small number of replicas are sufficient to obtain good exchange probabilities, for which sufficient overlap in the energy contributions of the replicas to be exchanged is required. This algorithm was further improved in REST2,[99] in which all replicas are simulated at the same temperature, while the conformational sampling is enhanced through scaling of the intramolecular potential energy of the solute. Compared to T-REMD, REST2 showed significantly lower computational cost and better sampling efficiency. Later on, a new variant of REST2 with optimized scaling and more flexibility with respect to which parts of the system the scaling is applied was developed by Bussi and termed

as Hamiltonian replica exchange molecular dynamics (also known as REST2 or HREX).[72] The Metropolis criterion of HREX for exchange attempts between two replicas i and j is

$$P(q_i \leftrightarrow q_j) = \min\left[1, \exp\left(\frac{H_i(q_i) - H_i(q_j)}{k_B T} + \frac{H_j(q_j) - H_j(q_i)}{k_B T}\right)\right] \quad (22)$$

When all replicas are simulated at the same temperature, Eq. (21) becomes Eq. (22). In HREX, all replicas are simulated at the same temperature, and the Hamiltonian of the solute is scaled by different scaling factors ($0 < \lambda \leq 1$) at different replicas, but the solvent–solvent interactions are not modified. With HREX, one can limit the scaling to a certain part of the simulated system which is considered to be interesting. For instance, in the example introduced in Section 3.2, HREX simulations were also mentioned. In that example only the interactions of loops 6 and 7 of TIM were modified during the HREX simulation.[68]

As REMD enhances unrestrained sampling, this method including its variants are widely applied to problems where the definition of a CV, especially prior sampling, is difficult or even impossible. Such problems include the sampling of peptides and their aggregation,[100,101] protein–protein recognition mechanisms,[102] or the identification of ligand binding sites and poses,[103–105] which could be useful for ligand identification in computational drug discovery. We used HREX to study the dynamics and dimerization of the amyloid peptide Aβ_{1-42} under varying external conditions, such as different pH values, the presence of Cu^{2+} that binds to Aβ, and oxidation of Aβ (oxidized Gly25).[106,107] For each dimer system, 24 replicas with scaling factors λ ranging from 1.0 to 0.4 (corresponding to effective temperatures of 300–750 K) were used in the HREX simulations, and each replica was subjected to 500 ns sampling, leading to 12 μs of cumulative simulation time for each dimer system. As shown in Fig. 8, we could characterize the different structural preferences of the four Aβ_{1-42} dimer systems at different conditions. Moreover, transition network analysis[108,109] on the HREX trajectories revealed that the acidic pH considerably increased the β-sheet content, while Cu^{2+} binding increased the exposed hydrophobic surface area. Both of the two factors can contribute to an increased oligomerization and toxicity of Aβ.

Replica exchange is also widely combined with other computational methods, such as, constant pH replica exchange,[110] REMD with elastic

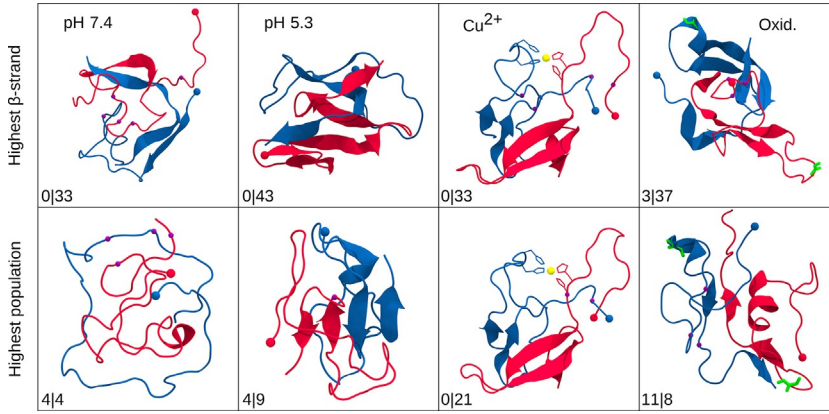

Fig. 8 Representative conformations with highest β-sheet content (*top*) and the most abundant conformations (*bottom*) of the Aβ_{1-42} dimer at different conditions. The peptides are rendered as a cartoon with one peptide of the dimer shown in *red* and the other one in *blue*. The N-terminus of each peptide is indicated by a *sphere* and β-bridges are marked by *violet spheres*. In the Cu^{2+} system, Cu^{2+} is shown as *yellow sphere* and the His13 and His14 residues bound to it as sticks. In the oxidized system, the oxidized Gly25 residues are shown as *green sticks*. In the *lower left corner* of each panel the number of residues in the $\alpha|\beta$ conformation is given. *Reproduced with permission from Liao Q, Owen MC, Bali S, et al. Aβ under stress: the effects of acidosis, Cu^{2+}-binding, and oxidation on amyloid β-peptide dimers. Chem Commun. 2018;54:7766–7769, The Royal Society of Chemistry.*

network analysis,[111] REMD with FEP,[112–115] multiplexed-REMD,[116] replica exchanging self-guided Langevin dynamics[117,118] and quantum mechanics/molecular mechanics (QM/MM)[119] replica exchange.[49] Moreover, multidimensional REMD has also been applied.[93,120,121]

4.2 Accelerated molecular dynamics

In REMD, the sampling is enhanced by exchanging configurations sampled at high temperature replicas with configurations from low temperature replicas. In 2004, the McCammon group developed another promising CV-free enhanced sampling method, accelerated molecular dynamics (aMD).[26] In aMD, as shown in Fig. 9, a boost potential is added to the potential energy surface of the studied system to facilitate the crossing of energy barriers between different conformational states. The boost potential $\Delta V(q)$ is activated when the potential energy of the system is below a threshold energy E:

$$\begin{aligned} V^*(q) &= V(q) && V(q) \geq E \\ V^*(q) &= V(q) + \Delta V(q) && V(q) < E \end{aligned} \tag{23}$$

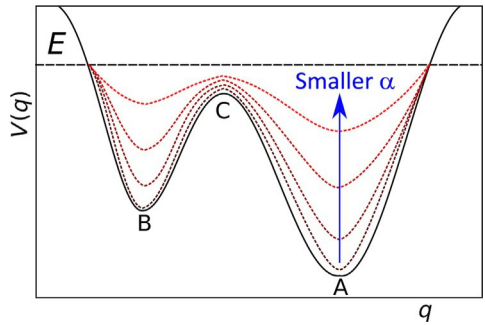

Fig. 9 An illustration of aMD. A nonnegative boost potential is imposed on the system potential energy surface when it is lower than a threshold energy *E*. As the acceleration factor α gets smaller, the potential energy surface becomes flatter and the number of transitions between states A and B increases. To be noted, *q* is for coordinates of the simulated system, but not a collective variable.

where $V(q)$ is the original potential energy of the system at configuration q, E is the threshold energy, and $V^*(q)$ is the modified potential. $\Delta V(q)$ is the boost potential, which is given by:

$$\Delta V(q) = \frac{(E - V(q))^2}{\alpha + E - V(q)} \tag{24}$$

where α is a predefined acceleration factor which controls the flatness of the potential energy. The boost potential can be either acting on the torsional potential,[26] the whole potential, or even both of them.[122–124] Similarly to scaling only parts of the solute in HREX, the boost potential can also be added only to regions of the biomolecule to enhance specific conformational transitions.[125]

aMD is a powerful approach for enhancing conformational sampling, and in theory, the free energy landscape can be recovered by a reweighting method termed "exponential average"[126] by applying the Boltzmann factor $e^{\Delta V(q)/k_B T}$ accounting for the boost potential acting on each conformational state. However, in practice, this algorithm is known to result in big statistical noise[126–128] as the Boltzmann reweighting factors are usually dominated by a small number of configurations with high boost potentials. The RaMD[129] (only the rotatable dihedrals are subjected to aMD) and the selective aMD[130] methods were developed to reduce the noise through applying boost potentials only to rotatable dihedrals, or certain regions of the simulated systems. The statistical noise can also be reduced by adapting a different reweighting algorithm. For example, less noise was obtained when the exponential term in the Boltzmann factor was replaced by a summation of the

Maclaurin series expansion of the boost potential,[123] which was further improved by a cumulant expansion to the second order.[131] A new variant called Gaussian accelerated MD (GaMD) was developed[132] to reduce the statistical noise by applying a harmonic boost potential instead to flatten the potential energy surface of the system. In GaMD, the system potential $V(q)$ is altered by adding a boost potential when it goes below a threshold energy E:

$$\Delta V(q) = \frac{k}{2} \times (E - V(q))^2 \tag{25}$$

where k is the force constant of the harmonic potential. The boost potential in GaMD follows a Gaussian distribution, and accurate free energies can be recovered through reweighting using a cumulant expansion to the second order.[132]

The boost potential in aMD simulations can be on the order of tens to hundreds of kcal·mol^{-1}, which is much larger in magnitude and wider in distribution than the boost potentials applied in CV-based biasing simulations, thus millisecond-timescale events can be captured by hundreds of nanoseconds aMD simulations.[123,128] A large range of applications has shown the success of aMD in enhancing the sampling of biomolecular systems, including peptide–protein binding,[133] ligand–protein binding,[134] conformational change of proteins.[135] aMD was also used to explore the most likely binding poses of ligands to their targets.[136] To be noted, GaMD was shown to be more efficient than aMD in characterizing protein folding and ligand binding pathways.[132,137] In a recent study, Palermo[138] used GaMD to study the conformational change of R976 in the CRISPR-Cas9 complex. For GaMD, the boost potential was applied in a dual-boost scheme, in which two acceleration potentials are applied simultaneously to the dihedral terms and entire potential of the system. As shown in Fig. 10, free energy landscapes were obtained by reweighting over the aggregated trajectories of four individual runs (~1.6 μs). The results clearly show the conformational preference of R976 in the RuvC domain (which performs DNA cleavages) of the activated CRISPR-Cas9 complex.

A number of further variants of aMD were developed to improve the configurational sampling efficiency, including the windowed aMD,[139] the integrated aMD,[140] and adaptive aMD.[141] Moreover, aMD was combined with many other computational techniques including replica exchange[98],[142–144] and thermodynamics integration[145] for better sampling. More details about the development and applications of aMD can be found in a previous review.[14]

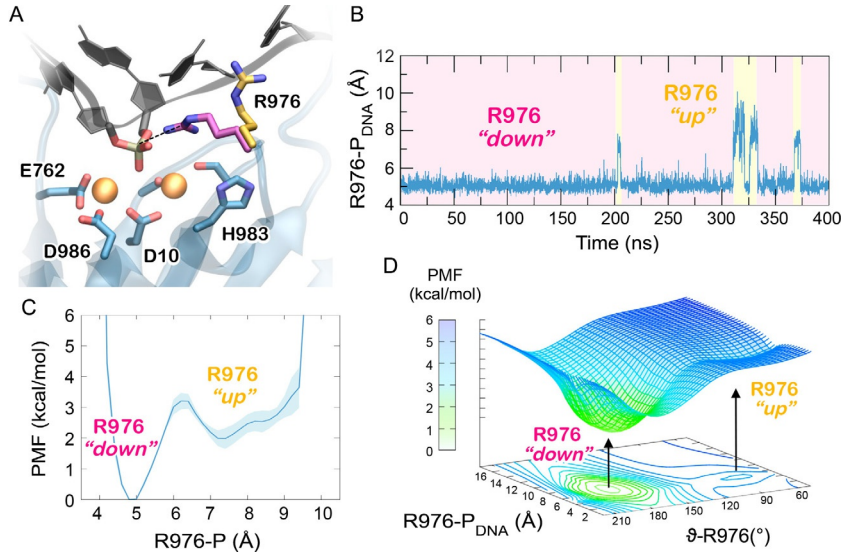

Fig. 10 (A) Catalytic site of the RuvC domain in the CRISPR-Cas9 complex, displaying two conformations of R976. In the activated CRISPR-Cas9 complex, R976 assumes a "down" conformational state (*magenta*), bringing its positive side chain in close contact to the scissile phosphate. In the inactive state of the CRISPR-Cas9 complex, R976 moves away from the scissile phosphate and assumes an "up" conformation (*yellow*). (B) Time evolution of the distance between the terminal carbon of R976 and the scissile phosphate (R976-P_{DNA}) during a ~400 ns GaMD simulation of the activated CRISPR-Cas9 complex. Time windows indicate the "down" (*magenta*) and "up" (*yellow*) conformations of R976. (C) PMF of the R976-P_{DNA} distance computed over the aggregated data from four individual runs for ~1.6 μs of total sampling. The PMF is expressed in kcal·mol^{-1}. (D) Two-dimensional PMF of the R976-P_{DNA} distance in combination with the dihedral angle between the C_α-C_β-C_γ-C_δ atoms of R976 (ϑ-R976). *Reprinted with permission from Palermo G. Structure and dynamics of the CRISPRCas9 catalytic complex. J Chem Inf Model. 2019;59:2394–2406, copyright 2019 American Chemistry Society.*

5. Combination of enhanced sampling approaches

The features of several popular enhanced sampling approaches have been described above. Due to the methodology differences between CV-based and CV-free approaches, methods from these classes can be easily combined to further optimize the efficiency of conformational space sampling. As mentioned above, any control parameter can be used in Eq. (21) to calculate the exchange probability between replicas in REMD (e.g., bias potentials in REUS, BE-MetaD, etc.). In this way, other enhanced sampling

methods can be combined with replica exchange. For example, replica exchange (parallel tempering) can enhance the sampling of all degrees of freedom by simulating the system with multiple replicas at different temperatures simultaneously, while metadynamics is used to enhance the sampling along predefined CVs. The combination of the two approaches leads to the methodology of PTMetaD,[55] however, PTMetaD is still suffering from the poor scaling of computational resources with the system size similar to T-REMD. This issue can be solved by applying the well-tempered ensemble (WTE) algorithm[146] during the PTMetaD simulation, which is then called PTMetaD-WTE scheme.[56] In this ensemble, the average energy value is still close to the canonical average value, but the enhancement of its fluctuations can be tuned to improve the overlap, thus improve the sampling efficiency. PTMetaD-WTE is widely used to study conformational changes of proteins.[147,148]

Inspired by the success of PTMetaD, another combination was also introduced as bias-exchange metadynamics (BE-MetaD) simulations. In BE-MetaD,[29] multiple metadynamics simulations are carried out in parallel, bias potentials are imposed on different CVs in different replicas, configuration exchange between different replicas is attempted at regular time intervals according to the bias potential. It is a powerful method to study protein folding,[149,150] ligand binding[149] and conformational change of disordered peptides.[151–154] As mentioned above, in our study of the loop motion of TIM only BE-MetaD but not standard MD or HREX were able to characterize the full conformational space of loop 6 of TIM with bound substrate.[68]

Further combinations involving replica exchange were developed. For instance, Bussi introduced replica exchange with collective-variable tempering (RECT),[155] which improved the efficiency of exploring the conformational space of RNA. Umbrella sampling was combined with replica exchange (REUS), and it was shown to be powerful in the study of the association and dissociation of a ligand-DNA system by Zeller and Zacharias.[28] Fajer et al.[98] combined replica exchange and aMD, leading to a replica exchange of various degrees of acceleration (REXAMD) to obtain better reweighting statistics while achieving enhanced sampling. This technique was validated and benchmarked on two simple gas-phase model systems. Roe et al.[142] evaluated the advantages of the combination of accelerated molecular dynamics and Hamiltonian replica exchange methods, which was applied to an RNA tetranucleotide system. Huang et al.[143] developed the replica exchange Gaussian aMD, and it was tested on three systems:

alanine dipeptide, chignolin, and HIV protease. Very recently, a combination of replica exchange umbrella sampling with Gaussian accelerated molecular dynamics was proposed.[156]

6. Programs and tutorials

In order to perform enhanced sampling simulations and free energy calculations, one has to be proficient with at least one MD program. Various popular MD programs are available and under active development, such as Amber,[160,161] Gromacs,[162,163] NAMD,[164] OpenMM,[165,166] and LAMMPS.[167] Several enhanced sampling methods are implemented in these programs. For example, Amber Gromacs support US, SMD, REMD, and REUS, while Amber also offers aMD and GaMD. The Colvars module in NAMD is powerful for performing biasing simulations, and very flexible with regard to the definition of specific CVs through Tcl scripts. The plugin PLUMED[53,54] can be interfaced with most biomolecular MD programs allowing to perform enhanced sampling simulation and free energy calculations. It should be noted that it is also a useful tool for analyzing MD simulations. There are a lot of functions for defining CVs available in PLUMED. A more complete overview of the enhanced sampling approaches that are provided by the different MD programs are listed in Table 1. For each program, there is also one, and in some cases even several tutorials provided for the users to learn specific methods. Among them, the hands-on PLUMED tutorials need mentioning as they are well explained and provide details

Table 1 Availability of enhanced sampling and free energy calculation methods in MD programs.

Methods	Bias potential	Free energy estimator	Variants	Programs
US	Eq. (9)	UI, WHAM, MBAR	Adaptive US	Gromacs, Amber, PLUMED[a], Colvars[b]
MetaD	Eq. (12)	Sum up Gaussians	WT-MetaD	PLUMED, Colvars
SMD	Eq. (17)	Jarzynski Eq. (18)	Adaptive SMD	Gromacs, Amber, Colvars
REMD	N/A	WHAM, MBAR	HREX	Gromacs, Amber, NAMD
aMD	Eq. (24)	Reweighting	GaMD	Amber NAMD

Continued

Table 1 Availability of enhanced sampling and free energy calculation methods in MD programs.—cont'd

Methods	Bias potential	Free energy estimator	Variants	Programs
Methods		**Codes**		
UI	https://github.com/ATB-UQ/umbrella_integration[c]			
WHAM	Gromacs (g_wham),[157] PLUMED, PyEMMA[158], wham[159]			
MBAR	PyEMMA,[158] pymbar (https://github.com/choderalab/pymbar)[36]			
Programs		**Web Pages & Tutorials**		
Gromacs	www.gromacs.org & www.mdtutorials.com/gmx			
Amber	https://ambermd.org & https://ambermd.org/tutorials			
NAMD	www.ks.uiuc.edu/Research/namd & www.ks.uiuc.edu/Training/Tutorials			
Plumed	www.plumed.org & www.plumed.org/doc-v2.5/user-doc/html/tutorials.html			

[a]PLUMED is a plugin for analyzing and biasing simulations; it can be interfaced with multiple MD programs, including Gromacs, Amber, NAMD, openMM, and LAMMPS.
[b]Colvars is a module available in NAMD and LAMMPS for biasing simulations.
[c]https://doi.org/10.5281/zenodo.164996.

ranging from theory, how to run simulations to analyzing them. Especially beginners are recommended to go through these tutorials to get familiar with the different enhanced sampling methods.

7. Conclusion and outlook

In the last decades, many methodologies have been proposed to increase the efficiency of sampling in MD simulations and of free energy calculations. They have been found to be highly powerful and successful for exploring the conformational space of biomolecular systems and improving the convergence of free energy calculations. In this chapter, comprehensive reviews of three CV-based and two CV-free enhanced sampling methods were provided, presenting their theories, recent developments and example applications. CV-based methods (e.g., US, MetaD, SMD) can effectively guide the simulations along predefined CVs, while CV-free methods (e.g., REMD, aMD) do not require predefined CVs for the simulations. Normally, there is no absolute standard for the decision which of these

methods should be chosen for an enhanced sampling biomolecular simulation. When good CVs exist for describing the biological events of interest, US and MetaD are good choices. When little information is known for the process to be simulated (e.g., protein folding or exploring the dynamics of an intrinsically disordered protein), REMD (or HREX) or aMD should be used. It should be noted that enhanced sampling methods and free energy calculations can also combined with the QM/MM (quantum mechanics/molecular mechanics) method[119] in computational enzymology.[168]

Although enhanced sampling and free energy calculations have demonstrated great efficiency and successes in biomolecular simulations, there are still some challenges. Future developments will be on the following directions to soften the challenges. It is known that CV-based methods suffer from the hidden CV problem, which may lead to inefficient samplings and inaccurate free energy calculations. The same consequence may result from selecting improper CVs that ignore certain dimensions of freedom in biasing simulations of complex or flexible systems. Furthermore, CV-based methods do not scale well with the number of CVs being biased, normally, US is limited to 2 CVs and MetaD is limited to 3 CVs. Although CV-free methods do not have this problem, they may still not assure sufficient sampling along the CVs of interest, as the sampling is enhanced over the whole system. To address these issues, new smart CVs which could describe high dimensional complex processes are proposed continually. For example, principal components (PCs),[169] native contacts,[75] conformational entropy[170] and time-lagged independent component analysis (tICA)[171,172] are very helpful for studies of conformational changes and folding of proteins, while eRMSD[173] was proposed to study RNA folding. Moreover, combinations of CV-based and CV-free methods (e.g., REUS, PTMetaD, BE-MetaD) can enhance the sampling along the interested CVs and other relevant degrees of freedom simultaneously, thus further improve the sampling efficiency and speed up the convergence. To be noted, path collective variables (PCV)[174] and the string method[21] are more advanced approaches to study more complex biological processes. The sampling with PCV or the string method are focused along the minimal free energy paths in high dimensional configuration space, thus promoting the efficiency of sampling. The string method was used to calculate binding free energies of protein–protein systems,[175] which is extremely challenging for other methods.

In general, longer timescales and more massive computational resources are required for more complex systems with millions of atoms, and simulating them with REMD (even HREX) would further increase the

computational expense. CV-based methods working well for smaller systems may not be efficient in simulations of huge systems due to the complex and large-scale nature of the relevant CVs. Enhanced sampling methods combined with coarse-grained force fields[176] or multiple time-stepping methods[177] can offer help by greatly accelerating the sampling efficiency for large systems. In all-atom simulation, the use of larger time steps (e.g., 4 fs) is advisable, which can be accomplished by employing virtual sites for hydrogens or hydrogen mass repartition[178] to remove fastest degrees of freedom. Machine learning has become very popular in recent years, more and more powerful deep unsupervised and reinforced learning techniques were developed, which in turn provided powerful tools to develop more intelligent sampling methods.[179–181] Machine learning was also used to train smart CVs.[182,183] It would be interesting to see more developments in this direction, in addition to more applications of machine-learning related methodologies in biomolecular simulations for validation of these methods.

Acknowledgments

The author gratefully acknowledges the help from the editor (BS) in providing detailed feedbacks on the draft of this chapter.

References

1. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. *Nature*. 1977;267:585–590.
2. Smith JC, Kneller GR. Combination of neutron scattering and molecular dynamics to determine internal motions in biomolecules. *Mol Sim*. 1993;10:363–375.
3. Ma P, Schillinger O, Schwarten M, et al. Conformational polymorphism in autophagy-related protein GATE-16. *Biochemistry*. 2015;54:5469–5479.
4. Narayanan C, Bafna K, Roux LD, Agarwal PK, Doucet N. Applications of NMR and computational methodologies to study protein dynamics. *Arch Biochem Biophys*. 2017;628:71–80.
5. Chan-Yao-Chong M, Durand D, Ha-Duong T. Molecular dynamics simulations combined with nuclear magnetic resonance and/or small-angle X-ray scattering data for characterizing intrinsically disordered protein conformational ensembles. *J Chem Inf Model*. 2019;59:1743–1758.
6. Boldon L, Laliberte F, Liu L. Review of the fundamental theories behind small angle X-ray scattering, molecular dynamics simulations, and relevant integrated application. *Nano Rev*. 2015;6:25661.
7. Lundborg M, Narangifard A, Wennberg CL, Lindahl E, Daneholt B, Norlén L. Human skin barrier structure and function analyzed by cryo-EM and molecular dynamics simulation. *J Struct Biol*. 2018;203:149–161.
8. Dolenc J, Missimer JH, Steinmetz MO, van Gunsteren WF. Methods of NMR structure refinement: molecular dynamics simulations improve the agreement with measured NMR data of a C-terminal peptide of GCN4-p1. *J Biomol NMR*. 2010; 47:221–235.

9. Perilla JR, Zhao G, Lu M, et al. CryoEM structure refinement by integrating NMR chemical shifts with molecular dynamics simulations. *J Phys Chem B*. 2017;121: 3853–3863.
10. Igaev M, Kutzner C, Bock LV, Vaiana AC, Grubmüller H. Automated cryo-EM structure refinement using correlation-driven molecular dynamics. *eLife*. 2019;8:e43542.
11. Srivastava A, Nagai T, Srivastava A, Miyashita O, Tama F. Role of computational methods in going beyond X-ray crystallography to explore protein structure and dynamics. *Int J Mol Sci*. 2018;19:3401.
12. Shevchuk R, Hub JS. Bayesian refinement of protein structures and ensembles against SAXS data using molecular dynamics. *PLoS Comput Biol*. 2017;13:1–27.
13. Abrams C, Bussi G. Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. *Entropy*. 2014;16:163–199.
14. Miao Y, McCammon JA. Unconstrained enhanced sampling for free energy calculations of biomolecules: a review. *Mol Simul*. 2016;42:1046–1055.
15. Yang YI, Shao Q, Zhang J, Yang L, Gao YQ. Enhanced sampling in molecular dynamics. *J Chem Phys*. 2019;151:070902.
16. Torrie GM, Valleau JP. Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J Comput Phys*. 1977;23:187–199.
17. Laio A, Parrinello M. Escaping free-energy minima. *Proc Natl Acad Sci USA*. 2002;99:12562–12566.
18. Lu H, Schulten K. Steered molecular dynamics simulations of force-induced protein domain unfolding. *Proteins*. 1999;35:453–463.
19. Darve E, Rodríguez-Gómez D, Pohorille A. Adaptive biasing force method for scalar and vector free energy calculations. *J Chem Phys*. 2008;128:144120.
20. Grubmüller H. Predicting slow structural transitions in macromolecular systems: conformational flooding. *Phys Rev E*. 1995;52:2893–2906.
21. Weinan E, Ren W, Vanden-Eijnden E. String method for the study of rare events. *Phys Rev B*. 2002;66:052301.
22. Maragliano L, Vanden-Eijnden E. A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chem Phys Lett*. 2006;426:168–175.
23. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett*. 1999;314:141–151.
24. Wu X, Wang S. Self-guided molecular dynamics simulation for efficient conformational search. *J Phys Chem B*. 1998;102:7238–7250.
25. Wu X, Brooks BR, Vanden-Eijnden E. Self-guided Langevin dynamics via generalized Langevin equation. *J Comput Chem*. 2016;37:595–601.
26. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J Chem Phys*. 2004;120: 11919–11929.
27. Sabri Dashti D, Roitberg AE. Optimization of umbrella sampling replica exchange molecular dynamics by replica positioning. *J Chem Theory Comput*. 2013;9: 4692–4699.
28. Zeller F, Zacharias M. Adaptive biasing combined with Hamiltonian replica exchange to improve umbrella sampling free energy simulations. *J Chem Theory Comput*. 2014;10:703–710.
29. Piana S, Laio A. A bias-exchange approach to protein folding. *J Phys Chem B*. 2007;111:4553–4559.
30. Zwanzig RW. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J Chem Phys*. 1954;22:1420–1426.
31. Bennett CH. Efficient estimation of free energy differences from Monte Carlo data. *J Comput Phys*. 1976;22:245–268.

32. Shirts MR, Bair E, Hooker G, Pande VS. Equilibrium free energies from non-equilibrium measurements using maximum-likelihood methods. *Phys Rev Lett*. 2003;91:140601.
33. 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.
34. Ferrenberg AM, Swendsen RH. Optimized Monte Carlo data analysis. *Phys Rev Lett*. 1989;63:1195–1198.
35. Kästner J. Umbrella sampling. *WIREs Comput Mol Sci*. 2011;1:932–942.
36. Shirts MR, Chodera JD. Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys*. 2008;129:124105.
37. 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.
38. Barducci A, Bonomi M, Parrinello M. Metadynamics. *WIREs Comput Mol Sci*. 2011;1:826–843.
39. Roux B. The calculation of the potential of mean force using computer simulations. *Comput Phys Commun*. 1995;91:275–282.
40. Bartels C, Karplus M. Multidimensional adaptive umbrella sampling: Applications to main chain and side chain peptide conformations. *J Comput Chem*. 1997;18:1450–1462.
41. Gallicchio E, Andrec M, Felts AK, Levy RM. Temperature weighted histogram analysis method, replica exchange, and transition paths. *J Phys Chem B*. 2005;109:6722–6731.
42. Ferguson AL. BayesWHAM: a Bayesian approach for free energy estimation, reweighting, and uncertainty quantification in the weighted histogram analysis method. *J Comput Chem*. 2017;38:1583–1605.
43. Kästner J, Thiel W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "umbrella integration". *J Chem Phys*. 2005;123:144104.
44. Meshkin H, Zhu F. Thermodynamics of protein folding studied by umbrella sampling along a reaction coordinate of native contacts. *J Chem Theory Comput*. 2017;13:2086–2097.
45. Lemkul JA, Bevan DR. Assessing the stability of Alzheimer's amyloid protofibrils using molecular dynamics. *J Phys Chem B*. 2010;114:1652–1660.
46. Domański J, Hedger G, Best RB, Stansfeld PJ, Sansom MSP. Convergence and sampling in determining free energy landscapes for membrane protein association. *J Phys Chem B*. 2017;121:3364–3375.
47. Lee MS, Olson MA. Calculation of absolute protein-ligand binding affinity using path and endpoint approaches. *Biophys J*. 2006;90:864–877.
48. Banavali NK, Roux B. Free energy landscape of A-DNA to B-DNA conversion in aqueous solution. *J Am Chem Soc*. 2005;127:6866–6876.
49. Fedorov DG, Sugita Y, Choi CH. Efficient parallel implementations of QM/MM-REMD (quantum mechanical/molecular mechanics-replica-exchange MD) and umbrella sampling: isomerization of H_2O_2 in aqueous solution. *J Phys Chem B*. 2013;117:7996–8002.
50. Lence E, Van der Kamp MW, González-Bello C, Mulholland AJ. QM/MM simulations identify the determinants of catalytic activity differences between type II dehydroquinase enzymes. *Org Biomol Chem*. 2018;16:4443–4455.
51. Wojtas-Niziurski W, Meng Y, Roux B, Bernéche S. Self-learning adaptive umbrella sampling method for the determination of free energy landscapes in multiple dimensions. *J Chem Theory Comput*. 2013;9:1885–1895.

52. Barducci A, Bussi G, Parrinello M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys Rev Lett*. 2008;100:020603.
53. Bonomi M, Branduardi D, Bussi G, et al. PLUMED: a portable plugin for free-energy calculations with molecular dynamics. *Comp Phys Commun*. 2009;180:1961–1972.
54. Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G. PLUMED 2: new feathers for an old bird. *Comp Phys Commun*. 2014;185:604–613.
55. Bussi G, Gervasio FL, Laio A, Parrinello M. Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics. *J Am Chem Soc*. 2006; 128:13435–13441.
56. Deighan M, Bonomi M, Pfaendtner J. Efficient simulation of explicitly solvated proteins in the well-tempered ensemble. *J Chem Theory Comput*. 2012;8:2189–2192.
57. Raiteri P, Laio A, Gervasio FL, Micheletti C, Parrinello M. Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *J Phys Chem B*. 2006;110:3533–3539.
58. Limongelli V, Bonomi M, Parrinello M. Funnel metadynamics as accurate binding free-energy method. *Proc Natl Acad Sci USA*. 2013;110:6358–6363.
59. Pfaendtner J, Bonomi M. Efficient sampling of high-dimensional free-energy landscapes with parallel bias metadynamics. *J Chem Theory Comput*. 2015;11:5062–5067.
60. Tiwary P, Parrinello M. From metadynamics to dynamics. *Phys Rev Lett*. 2013; 111:230602.
61. Tiwary P, Limongelli V, Salvalaglio M, Parrinello M. Kinetics of protein-ligand unbinding: predicting pathways, rates, and rate-limiting steps. *Proc Natl Acad Sci USA*. 2015;112:E386–E391.
62. Tiwary P, Mondal J, Berne BJ. How and when does an anticancer drug leave its binding site? *Sci Adv*. 2017;3:e1700014.
63. Casasnovas R, Limongelli V, Tiwary P, Carloni P, Parrinello M. Unbinding kinetics of a p38 MAP kinase type II inhibitor from metadynamics simulations. *J Am Chem Soc*. 2017;139:4780–4788.
64. Wang Y, Martins JM, Lindorff-Larsen K. Biomolecular conformational changes and ligand binding: from kinetics to thermodynamics. *Chem Sci*. 2017;8:6466–6473.
65. Wang Y, Valsson O, Tiwary P, Parrinello M, Lindorff-Larsen K. Frequency adaptive metadynamics for the calculation of rare-event kinetics. *J Chem Phys*. 2018;149: 072309.
66. Valsson O, Tiwary P, Parrinello M. Enhancing important fluctuations: rare events and metadynamics from a conceptual viewpoint. *Ann Rev Phys Chem*. 2016;67:159–184.
67. Barrozo A, Liao Q, Esguerra M, et al. Computer simulations of the catalytic mechanism of wild-type and mutant β-phosphoglucomutase. *Org Biomol Chem*. 2018;16: 2060–2073.
68. Liao Q, Kulkarni Y, Sengupta U, et al. Loop motion in triosephosphate isomerase is not a simple open and shut case. *J Am Chem Soc*. 2018;140:15889–15903.
69. Clark AJ, Tiwary P, Borrelli K, et al. Prediction of protein-ligand binding poses via a combination of induced fit docking and metadynamics simulations. *J Chem Theory Comput*. 2016;12:2990–2998.
70. Ensing B, De Vivo M, Liu Z, Moore P, Klein ML. Metadynamics as a tool for exploring free energy landscapes of chemical reactions. *Acc Chem Res*. 2006;39:73–81.
71. Hsu WL, Furuta T, Sakurai M. ATP hydrolysis mechanism in a maltose transporter explored by QM/MM metadynamics simulation. *J Phys Chem B*. 2016;120: 11102–11112.
72. Bussi G. Hamiltonian replica exchange in GROMACS: a flexible implementation. *Mol Phys*. 2014;112:379–384.
73. Knowles JR, Albery WJ. Perfection in enzyme catalysis: the energetics of triosephosphate isomerase. *Acc Chem Res*. 1977;10:105–111.

74. Wierenga RK, Kapetaniou EG, Venkatesan R. Triosephosphate isomerase: a highly evolved biocatalyst. *Cell Mol Life Sci*. 2010;67:3961–3982.
75. Best RB, Hummer G, Eaton WA. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc Natl Acad Sci USA*. 2013;110:17874–17879.
76. Berman HM, Westbrook J, Feng Z, et al. The protein data bank. *Nucleic Acids Res*. 2000;28:235–242.
77. Jogl G, Rozovsky S, McDermott AE, Tong L. Optimal alignment for enzymatic proton transfer: structure of the Michaelis complex of triosephosphate isomerase at 1.2-Å resolution. *Proc Natl Acad Sci USA*. 2003;100:50–55.
78. Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J Chem Phys*. 2003;119:3559–3566.
79. Park S, Schulten K. Calculating potentials of mean force from steered molecular dynamics simulations. *J Chem Phys*. 2004;120:5946–5961.
80. Jarzynski C. Nonequilibrium equality for free energy differences. *Phys Rev Lett*. 1997;78:2690–2693.
81. Jarzynski C. Equilibrium free-energy differences from nonequilibrium measurements: a master-equation approach. *Phys Rev E*. 1997;56:5018–5035.
82. Crooks GE. Path-ensemble averages in systems driven far from equilibrium. *Phys Rev E*. 2000;61:2361–2366.
83. Xiao BL, Ning YN, Niu NN, et al. Steered molecular dynamic simulations of conformational lock of Cu, Zn-superoxide dismutase. *Sci Rep*. 2019;9:4353.
84. Capelli AM, Costantino G. Unbinding pathways of VEGFR2 inhibitors revealed by steered molecular dynamics. *J Chem Inf Model*. 2014;54:3124–3136.
85. Jin H, Zhu J, Dong Y, Han W. Exploring the different ligand escape pathways in acylaminoacyl peptidase by random acceleration and steered molecular dynamics simulations. *RSC Adv*. 2016;6:10987–10996.
86. Patel JS, Berteotti A, Ronsisvalle S, Rocchia W, Cavalli A. Steered molecular dynamics simulations for studying protein-ligand interaction in cyclin-dependent kinase 5. *J Chem Inf Model*. 2014;54:470–480.
87. Colizzi F, Perozzo R, Scapozza L, Recanatini M, Cavalli A. Single-molecule pulling simulations can discern active from inactive enzyme inhibitors. *J Am Chem Soc*. 2010;132:7361–7371.
88. Do PC, Lee EH, Le L. Steered molecular dynamics simulation in rational drug design. *J Chem Inf Model*. 2018;58:1473–1482.
89. Marzinek JK, Bond PJ, Lian G, et al. Free energy predictions of ligand binding to an α-helix using steered molecular dynamics and umbrella sampling simulations. *J Chem Inf Model*. 2014;54:2093–2104.
90. Zhou JK, Yang DY, Sheu SY. The molecular mechanism of ligand unbinding from the human telomeric G-quadruplex by steered molecular dynamics and umbrella sampling simulations. *Phys Chem Chem Phys*. 2015;17:12857–12869.
91. Ozer G, Valeev EF, Quirk S, Hernandez R. Adaptive steered molecular dynamics of the long-distance unfolding of neuropeptide Y. *J Chem Theory Comput*. 2010;6:3026–3038.
92. Schlitter J, Engels M, Krüger P. Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. *J Mol Graph*. 1994;12:84–89.
93. Sugita Y, Kitao A, Okamoto Y. Multidimensional replica-exchange method for free-energy calculations. *J Chem Phys*. 2000;113:6042–6051.
94. Sugita Y, Okamoto Y. Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape. *Chem Phys Lett*. 2000;329:261–270.

95. Fukunishi H, Watanabe O, Takada S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: application to protein structure prediction. *J Chem Phys*. 2002;116:9058–9067.
96. Liu P, Kim B, Friesner RA, Berne BJ. Replica exchange with solute tempering: a method for sampling biological systems in explicit water. *Proc Natl Acad Sci USA*. 2005;102:13749–13754.
97. Huang X, Hagen M, Kim B, Friesner RA, Zhou R, Berne BJ. Replica exchange with solute tempering: efficiency in large scale systems. *J Phys Chem B*. 2007;111:5405–5410.
98. Fajer M, Hamelberg D, McCammon JA. Replica-exchange accelerated molecular dynamics (REXAMD) applied to thermodynamic integration. *J Chem Theory Comput*. 2008;4:1565–1569.
99. 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.
100. Laghaei R, Mousseau N, Wei G. Effect of the disulfide bond on the monomeric structure of human amylin studied by combined Hamiltonian and temperature replica exchange molecular dynamics simulations. *J Phys Chem B*. 2010;114:7071–7077.
101. Laghaei R, Mousseau N, Wei G. Structure and thermodynamics of amylin dimer studied by Hamiltonian-temperature replica exchange molecular dynamics simulations. *J Phys Chem B*. 2011;115:3146–3154.
102. Leahy CT, Kells A, Hummer G, Buchete NV, Rosta E. Peptide dimerization-dissociation rates from replica exchange molecular dynamics. *J Chem Phys*. 2017;147:152725.
103. Wang K, Chodera JD, Yang Y, Shirts MR. Identifying ligand binding sites and poses using GPU-accelerated Hamiltonian replica exchange molecular dynamics. *J Comput Aid Mol Des*. 2013;27:989–1007.
104. Luitz MP, Zacharias M. Protein-ligand docking using Hamiltonian replica exchange simulations with soft core potentials. *J Chem Inf Model*. 2014;54:1669–1675.
105. Ostermeir K, Zacharias M. Accelerated flexible protein-ligand docking using Hamiltonian replica exchange with a repulsive biasing potential. *PLoS ONE*. 2017;12:1–17.
106. Liao Q, Owen MC, Olubiyi OO, Barz B, Strodel B. Conformational transitions of the amyloid-β peptide upon copper(II) binding and pH changes. *Isr J Chem*. 2017;57:771–784.
107. Liao Q, Owen MC, Bali S, Barz B, Strodel B. Aβ under stress: the effects of acidosis, Cu^{2+}-binding, and oxidation on amyloid β-peptide dimers. *Chem Commun*. 2018;54:7766–7769.
108. Barz B, Wales DJ, Strodel B. A kinetic approach to the sequence-aggregation relationship in disease-related protein assembly. *J Phys Chem B*. 2014;118:1003–1011.
109. Barz B, Olubiyi OO, Strodel B. Early amyloid β-protein aggregation precedes conformational change. *Chem Commun*. 2014;50:5373–5375.
110. Meng Y, Roitberg AE. Constant pH replica exchange molecular dynamics in biomolecules using a discrete protonation model. *J Chem Theory Comput*. 2010;6:1401–1412.
111. Ostermeir K, Zacharias M. Hamiltonian replica exchange combined with elastic network analysis to enhance global domain motions in atomistic molecular dynamics simulations. *Proteins*. 2014;82:3410–3419.
112. Jiang W, Hodoscek M, Roux B. Computation of absolute hydration and binding free energy with free energy perturbation distributed replica-exchange molecular dynamics. *J Chem Theory Comput*. 2009;5:2583–2588.
113. Jiang W, Roux B. Free energy perturbation Hamiltonian replica-exchange molecular dynamics (FEP/H-REMD) for absolute ligand binding free energy calculations. *J Chem Theory Comput*. 2010;6:2559–2565.

114. Jiang W, Luo Y, Maragliano L, Roux B. Calculation of free energy landscape in multi-dimensions with Hamiltonian-exchange umbrella sampling on petascale supercomputer. *J Chem Theory Comput*. 2012;8:4672–4680.
115. 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.
116. Rhee YM, Pande VS. Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys J*. 2003;84:775–786.
117. Lee MS, Olson MA. Protein folding simulations combining self-guided Langevin dynamics and temperature-based replica exchange. *J Chem Theory Comput*. 2010; 6:2477–2487.
118. Wu X, Hodoscek M, Brooks BR. Replica exchanging self-guided Langevin dynamics for efficient and accurate conformational sampling. *J Chem Phys*. 2012;137:044106.
119. Warshel A, Levitt M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol*. 1976;103:227–249.
120. Bergonzo C, Henriksen NM, Roe DR, Swails JM, Roitberg AE, Cheatham TE. Multidimensional replica exchange molecular dynamics yields a converged ensemble of an RNA tetranucleotide. *J Chem Theory Comput*. 2014;10:492–499.
121. Cruzeiro VWD, Roitberg AE. Multidimensional replica exchange simulations for efficient constant pH and redox potential molecular dynamics. *J Chem Theory Comput*. 2019;15:871–881.
122. Hamelberg D, de Oliveira CAF, McCammon JA. Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *J Chem Phys*. 2007;127:155102.
123. Pierce LCT, Salomon-Ferrer R, de Oliveira CAF, McCammon JA, Walker RC. Routine access to millisecond time scale events with accelerated molecular dynamics. *J Chem Theory Comput*. 2012;8:2997–3002.
124. Miao Y, Feixas F, Eun C, McCammon JA. Accelerated molecular dynamics simulations of protein folding. *J Computat Chem*. 2015;36:1536–1549.
125. Lan P, Tan M, Zhang Y, et al. Structural insight into precursor tRNA processing by yeast ribonuclease P. *Science*. 2018;362:eaat6678.
126. Shen T, Hamelberg D. A statistical analysis of the precision of reweighting-based simulations. *J Chem Phys*. 2008;129:034103.
127. Sinko W, Miao Y, de Oliveira CAF, McCammon JA. Population based reweighting of scaled molecular dynamics. *J Phys Chem B*. 2013;117:12759–12768.
128. 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.
129. Doshi U, Hamelberg D. Improved statistical sampling and accuracy with accelerated molecular dynamics on rotatable torsions. *J Chem Theory Comput*. 2012;8:4004–4012.
130. Wereszczynski J, McCammon JA. Using selectively applied accelerated molecular dynamics to enhance free energy calculations. *J Chem Theory Comput*. 2010;6: 3285–3292.
131. 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.
132. Miao Y, Feher VA, McCammon JA. Gaussian accelerated molecular dynamics: unconstrained enhanced sampling and free energy calculation. *J Chem Theory Comput*. 2015;11:3584–3595.
133. 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.

134. Miao Y, Huang YM, Walker RC, McCammon JA, Chang CA. Ligand binding pathways and conformational transitions of the HIV protease. *Biochemistry*. 2018;57:1533–1541.
135. Miao Y, Feixas F, Eun C, McCammon JA. Accelerated molecular dynamics simulations of protein folding. *J Comput Chem*. 2015;36:1536–1549.
136. Fratev F, Steinbrecher T, Jónsdóttir SO. Prediction of accurate binding modes using combination of classical and accelerated molecular dynamics and free-energy perturbation calculations: an application to toxicity studies. *ACS Omega*. 2018;3:4357–4371.
137. Huang YM, Raymundo MAV, Chen W, Chang CA. Mechanism of the association pathways for a pair of fast and slow binding ligands of HIV-1 protease. *Biochemistry*. 2017;56:1311–1323.
138. Palermo G. Structure and dynamics of the CRISPRCas9 catalytic complex. *J Chem Inf Model*. 2019;59:2394–2406.
139. Sinko W, de Oliveira CAF, Pierce LCT, McCammon JA. Protecting high energy barriers: a new equation to regulate boost energy in accelerated molecular dynamics simulations. *J Chem Theory Comput*. 2012;8:17–23.
140. Peng X, Zhang Y, Li Y, et al. Integrating multiple accelerated molecular dynamics to improve accuracy of free energy calculations. *J Chem Theory Comput*. 2018;14:1216–1227.
141. Markwick PRL, Pierce LCT, Goodin DB, McCammon JA. Adaptive accelerated molecular dynamics (Ad-AMD) revealing the molecular plasticity of P450cam. *J Phys Chem Lett*. 2011;2:158–164.
142. Roe DR, Bergonzo C, Cheatham TE. Evaluation of enhanced sampling provided by accelerated molecular dynamics with Hamiltonian replica exchange methods. *J Phys Chem B*. 2014;118:3543–3552.
143. Huang YM, 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.
144. Fajer M, Swift RV, McCammon JA. Using multistate free energy techniques to improve the efficiency of replica exchange accelerated molecular dynamics. *J Comput Chem*. 2009;30:1719–1725.
145. de Oliveira CAF, Hamelberg D, McCammon JA. Coupling accelerated molecular dynamics methods with thermodynamic integration simulations. *J Chem Theory Comput*. 2008;4:1516–1525.
146. Bonomi M, Parrinello M. Enhanced sampling in the well-tempered ensemble. *Phys Rev Lett*. 2010;104:190601.
147. Sutto L, Gervasio FL. Effects of oncogenic mutations on the conformational free-energy landscape of EGFR kinase. *Proc Natl Acad Sci USA*. 2013;110:10616–10621.
148. Papaleo E, Sutto L, Gervasio FL, Lindorff-Larsen K. Conformational changes and free energies in a proline isomerase. *J Chem Theory Comput*. 2014;10:4169–4174.
149. Baftizadeh F, Cossio P, Pietrucci F, Laio A. Protein folding and ligand-enzyme binding from bias-exchange metadynamics simulations. *Curr Phys Chem*. 2012;2:79–91.
150. Granata D, Camilloni C, Vendruscolo M, Laio A. Characterization of the free-energy landscapes of proteins by NMR-guided metadynamics. *Proc Natl Acad Sci USA*. 2013;110:6817–6822.
151. Baftizadeh F, Pietrucci F, Biarnés X, Laio A. Nucleation process of a fibril precursor in the C-terminal segment of amyloid-β. *Phys Rev Lett*. 2013;110:168103.
152. Granata D, Baftizadeh F, Habchi J, et al. The inverted free energy landscape of an intrinsically disordered peptide by simulations and experiments. *Sci Rep*. 2015;5:15449.
153. Liu N, Duan M, Yang M. Structural properties of human IAPP dimer in membrane environment studied by all-atom molecular dynamics simulations. *Sci Rep*. 2017;7:7915.
154. Guo AZ, Fluitt AM, de Pablo JJ. Early-stage human islet amyloid polypeptide aggregation: mechanisms behind dimer formation. *J Chem Phys*. 2018;149:025101.

155. Gil-Ley A, Bussi G. Enhanced conformational sampling using replica exchange with collective-variable tempering. *J Chem Theory Comput*. 2015;11:1077–1085.
156. 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.
157. Hub JS, de Groot BL, van der Spoel D. g_wham—a free weighted histogram analysis implementation including robust error and autocorrelation estimates. *J Chem Theory Comput*. 2010;6:3713–3720.
158. Scherer MK, Trendelkamp-Schroer B, Paul F, et al. PyEMMA 2: a software package for estimation, validation, and analysis of Markov models. *J Chem Theory Comput*. 2015;11:5525–5542.
159. Grossfield A. WHAM: the weighted histogram analysis method, version 2.0.9. http://membrane.urmc.rochester.edu/wordpress/?page_id=126.
160. Case DA, Cheatham IIITE, Darden T, et al. The amber biomolecular simulation programs. *J Comput Chem*. 2005;26:1668–1688.
161. Salomon-Ferrer R, Case DA, Walker RC. An overview of the amber biomolecular simulation package. *WIREs Comput Mol Sci*. 2013;3:198–210.
162. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: fast, flexible, and free. *J Comput Chem*. 2005;26:1701–1718.
163. Abraham MJ, Murtola T, Schulz R, et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015;1-2:19–25.
164. Phillips JC, Braun R, Wang W, et al. Scalable molecular dynamics with NAMD. *J Comput Chem*. 2005;26:1781–1802.
165. Eastman P, Friedrichs MS, Chodera JD, et al. OpenMM 4: a reusable, extensible, hardware independent library for high performance molecular simulation. *J Chem Theory Comput*. 2013;9:461–469.
166. Eastman P, Swails J, Chodera JD, et al. OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol*. 2017;13:1–17.
167. Plimpton S. Fast parallel algorithms for short-range molecular dynamics. *J Comput Phys*. 1995;117:1–19.
168. Van der Kamp MW, Mulholland AJ. Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry*. 2013;52:2708–2728.
169. Spiwok V, Lipovová P, Králová B. Metadynamics in essential coordinates: free energy simulation of conformational changes. *J Phys Chem B*. 2007;111:3073–3076.
170. Palazzesi F, Valsson O, Parrinello M. Conformational entropy as collective variable for proteins. *J Phys Chem Lett*. 2017;8:4752–4756.
171. Pérez-Hernández G, Noé F. Hierarchical time-lagged independent component analysis: computing slow modes and reaction coordinates for large molecular systems. *J Chem Theory Comput*. 2016;12:6118–6129.
172. Sultan MM, Pande VS. tICA-Metadynamics: accelerating metadynamics by using kinetically selected collective variables. *J Chem Theory Comput*. 2017;13:2440–2447.
173. Bottaro S, Di Palma F, Bussi G. The role of nucleobase interactions in RNA structure and dynamics. *Nucleic Acids Res*. 2014;42:13306–13314.
174. Branduardi D, Gervasio FL, Parrinello M. From A to B in free energy space. *J Chem Phys*. 2007;126:054103.
175. Suh D, Jo S, Jiang W, Chipot C, Roux B. String method for protein-protein binding free-energy calculations. *J Chem Theory Comput*. 2019;15:5829–5844.
176. Singh N, Li W. Recent advances in coarse-grained models for biomolecules and their applications. *Int J Mol Sci*. 2019;20:3774.

177. Chen PY, Tuckerman ME. Molecular dynamics based enhanced sampling of collective variables with very large time steps. *J Chem Phys*. 2018;148:024106.
178. Hopkins CW, Le Grand S, Walker RC, Roitberg AE. Long-time-step molecular dynamics through hydrogen mass repartitioning. *J Chem Theory Comput*. 2015;11:1864–1874.
179. Ribeiro JML, Bravo P, Wang Y, Tiwary P. Reweighted autoencoded variational Bayes for enhanced sampling (RAVE). *J Chem Phys*. 2018;149:072301.
180. Noé F, Olsson S, Köhler J, Wu H. Boltzmann generators: sampling equilibrium states of many-body systems with deep learning. *Science*. 2019;365:eaaw1147.
181. Bonati L, Zhang YY, Parrinello M. Neural networks-based variationally enhanced sampling. *Proc Natl Acad Sci USA*. 2019;116:17641–17647.
182. Sultan MM, Pande VS. Automated design of collective variables using supervised machine learning. *J Chem Phys*. 2018;149:094106.
183. Trapl D, Horvacanin I, Mareska V, Ozcelik F, Unal G, Spiwok V. Anncolvar: approximation of complex collective variables by artificial neural networks for analysis and biasing of molecular simulations. *Front Mol Biosci*. 2019;6:25.