

Enhanced sampling simulations to construct free-energy landscape of protein–partner substrate interaction

Jinzen Ikebe¹ · Koji Umezawa² · Junichi Higo³

Received: 1 December 2015 / Accepted: 7 December 2015 / Published online: 11 January 2016

© International Union for Pure and Applied Biophysics (IUPAB) and Springer-Verlag Berlin Heidelberg 2016

Abstract Molecular dynamics (MD) simulations using all-atom and explicit solvent models provide valuable information on the detailed behavior of protein–partner substrate binding at the atomic level. As the power of computational resources increase, MD simulations are being used more widely and easily. However, it is still difficult to investigate the thermodynamic properties of protein–partner substrate binding and protein folding with conventional MD simulations. Enhanced sampling methods have been developed to sample conformations that reflect equilibrium conditions in a more efficient manner than conventional MD simulations, thereby allowing the construction of accurate free-energy landscapes. In this review, we discuss these enhanced sampling methods using a series of case-by-case examples. In particular, we review enhanced sampling methods conforming to trivial trajectory parallelization, virtual-system coupled multicanonical MD, and adaptive lambda square dynamics. These methods have been recently developed based on the existing method of multicanonical MD simulation. Their applications are reviewed with an emphasis on describing their practical implementation. In our concluding remarks we explore extensions of the enhanced sampling methods that may allow for even more efficient sampling.

Keywords Molecular dynamics simulation · Enhanced sampling · Multicanonical · Conformational ensemble · Protein interaction

Introduction

Interactions between biomolecules (such as that between a receptor and a ligand) are crucial and fundamental events in biological activity because a biomolecule, such as a protein, must bind to and recognize its target molecule for signal transduction, catalytic action, storage, among others. Therefore, the determining factors which regulate the strength of such interactions are pertinent research questions to be studied and answered. The interaction strength is characterized by the binding affinity between the receptor and the ligand, which is quantified by the free-energy difference between the isolated (i.e., free) and bound states. When the bound state is more stable than the isolated state at a given concentration of the receptor and ligand, then the free energy of the bound state is lower than that of the isolated state. Physical factors that stabilize the bound state originate from hydrogen bonding, salt bridges, van der Waals interactions, and hydrophobic interactions. The dominant factors among these have been identified by investigating the specific receptor–ligand complex structures using structural biology methods. Structurally ambiguous regions in some complex structures have been shown to contribute to binding due to the fact that the affinity of the complex is altered when the ambiguous region is removed. Along the same line, complexes, which are less stable than the most stable complex structure, may be formed in the ligand–receptor interaction process and may play an important role in exerting their biological functions. Those complexes possessing potentially important structural ambiguity or multiple complex forms have been referred to as “fuzzy

✉ Junichi Higo
higo@protein.osaka-u.ac.jp

¹ Molecular Modeling and Simulation Group, Japan Atomic Energy Agency, 8-1-7 Umegida, Kizugawa, Kyoto 619-0215, Japan

² Department of Pure and Applied Physics, Waseda University, Okubo 3-4-1, Shinjuku-Ku, Tokyo 169-8555, Japan

³ Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan

complexes” (Tompa and Fuxreiter 2008). In addition, such structural ambiguity/multiple complex forms may be related to each other, although uncovering such relations by experimental means is difficult. Molecular simulation is an extremely powerful tool for investigating the microscopic conformations involved in the binding process. Recently, efficient computational sampling methods have been applied to the study of free-energy and conformational profiles associated with the conversion from free to bound states.

Molecular simulation provides microscopic conformations as snapshots at an atomic resolution. Under exhaustive sampling conditions, the free energy is estimated from the ensemble of the snapshots. In such cases, the calculation of the binding affinity can be turned into a counting problem through enumeration of the number of snapshots in both the bound and isolated states. From knowledge of the density of states, the free-energy profile along the binding process can be then obtained. The free-energy profile, defined in relation to a specific reaction coordinate, is referred to as the free-energy landscape (FEL). One facile example of a reaction coordinate is the separation distance between the receptor and the ligand. In this case, the FEL is expressed as a function of the separation distance. Not only the bound state but also intermediate states or encounter complexes may be identified in the FEL, with a knowledge of all species helping to provide an understanding of the binding mechanism. A more rigorous means to describe the FEL is to use two structural parameters, such as the separation distance and the relative molecular orientation between the receptor and the ligand. In this case, the FEL is expressed two-dimensionally, which gives it a higher resolution than one-dimensional FELs and thereby allows easier identification of the intermediate or encounter complexes. Here, we note that the reliability of the FEL strongly depends on the quality of the simulation data. Typically, there are two practical problems in generating FELs: imperfect force fields and limited sampling times. The former results from the difficulty in determining physical parameters to calculate the potential energy. For example, the torsion energy for the protein backbone affects the secondary structure contents of the snapshots (Ikebe et al. 2007; Kamiya et al. 2005). Tuning of the torsion-energy parameters is still an extremely active research topic in this field (Maier et al. 2015; Sakae and Okamoto 2014). The second problem, which is related to the statistical accuracy of the resultant FEL, is the focus of this review. In the following sections we explain sampling techniques aimed at improving statistical accuracy and introduce the novel methodology we have developed.

We first consider the limitation of sampling time. With respect to biological molecules, the simulation of the equilibrium state at temperatures near that of room temperature is of primary concern. Conventional molecular dynamics (MD) simulations (i.e., canonical MD) at constant temperature and constant volume yield a conformational ensemble (ensemble

of snapshots) that will only cover part of the entire ensemble of conformations at equilibrium. In the ideal case in which simulation proceeds with exhaustive sampling, this entire ensemble is called the canonical ensemble. In reality, however, although conventional MD simulation using an all-atom model is a powerful and precise tool for producing biomolecular conformations, the simulation time is limited up to micro- or milliseconds, even with use of state-of-the-art computers. Given these time constraints, the conventional method is insufficient to sample the entire ensemble at room temperature to complete the FEL of the biomolecular system (Fig. 1). In general, there are many energy minima (or energy basins) in a space that express the conformation of the biomolecule(s). This space, called the “conformational space”, is characterized by a series of energy basins separated by energy barriers (Fig. 1). In a completely general manner we may suppose that our simulated system is in an energy basin. In order to access the surrounding conformational space the system must overcome an energy barrier, i.e., make it to the next energy basin. When the energy barrier is high, the system is trapped in the energy basin, and the system cannot escape from the basin quickly. Successful barrier crossing corresponds to a “rare event”. To search various conformations, barrier crossing should be facilitated (Fig. 1), as it is these time-consuming events which limit efficient sampling within the constraints of simulation time. Enhanced sampling methods have been developed to avoid such non-productive trapping. These representative enhanced sampling methods are discussed in the next section.

Enhanced conformational sampling methods

There are a number of enhanced sampling methods (Bernardi et al. 2015; Christen and van Gunsteren 2008; Mitsutake et al.

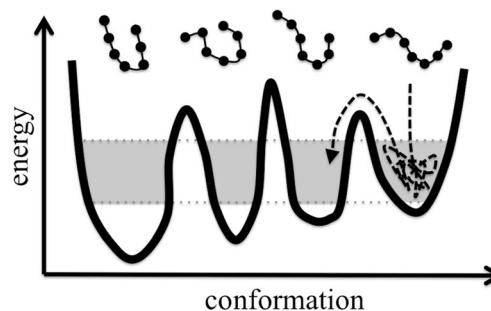


Fig. 1 Schematic view of the ragged free-energy landscape (FEL) of a biomolecule. There are three energy barriers and four basins. The four representative conformations at each of the four basins are shown as *chains of filled circles*. The *shaded regions* of each basin correspond to the respective conformation with high probability at room temperature. Within the constraint of simulation time for conventional molecular dynamics (MD), the MD trajectory (*dashed curved line*) spends most of the time in one basin, ultimately surpassing an energy barrier but unable to search all of the basins for the highlighted regions

2001). Here we only discuss those methods which produce the equilibrium ensemble at room temperature from the stock of snapshots without any additional simulation. Two such methods deserving particular attention are umbrella sampling and parallel tempering.

Umbrella sampling The umbrella sampling method (Chandler 1987; Deng and Roux 2009) facilitates sampling by introducing bias potential functions (so-called umbrella potential). During a simulation, the system is guided by the bias potential functions so that the system fluctuates in regions where the bias is low. Each umbrella potential is usually designed as a convex-downward harmonic potential, and the minima of these umbrella potential functions are typically set at points reflecting the high-free energy barriers. After umbrella sampling, the probability calculated from the sampled snapshots is, of course, affected by the umbrella potential functions. The probability independent of the umbrella potential can be estimated using the weighted histogram analysis method (WHAM; Kumar et al. 1992), and the FEL is reconstructed for the correct unbiased probability. Here, it is worth pointing out a practical aspect—that the choice of the variable (reaction coordinate) and the weight for the umbrella potential determine the sampling efficiency and that this choice is highly system-dependant.

Parallel tempering The parallel tempering method is a widely used enhanced sampling method (Earl and Deem 2005) and is also known as the temperature replica exchange method (tREM) (Sugita and Okamoto 1999). In tREM, a set of different temperatures is utilized that range from a lower temperature (i.e., room temperature) to a significantly higher temperature. Multiple conventional canonical MD runs are started simultaneously at these different temperatures such that each simulation has a different temperature. Each run is named a replica. During tREM, the temperature of a replica is exchanged with the temperature of one of its two neighboring replicas when the principle of detailed balance is satisfied to keep equilibrium. The temperature of a trajectory then fluctuates over a wide temperature range. At a higher temperature, the system can escape readily from a trapped state. As a practical check, when performing tREM, one might monitor how frequently exchange occurs. If no exchange occurs, the sampling is not enhanced because the run is simply a conventional canonical simulation. For the case of tREM, the fewer the transitions, the less effective is the sampling. Such situations generally occur when the difference in temperature between replicas is relatively large, as well as the system (i.e., many, many atoms). Very recently, the replica-permutation method (RPM) (Itoh and Okumura 2013b) and “heat bath like criteria” (HLC) method (Kondo and Taiji 2013) have been proposed. These methods uses the Suwa–Todo algorithm

(Suwa and Todo 2010) to obtain an equilibrium ensemble without imposing the detailed balance criterion, which increases the exchange ratio between the different temperatures among the replicas. With such developments the RPM and HLC may ultimately prove to be more efficient than tREM.

There are also other types of enhanced sampling methods which modify the potential energy directly to lower the energy barrier. Examples of this approach include an extended tREM, accelerated MD (aMD) (Doshi and Hamelberg 2015; Hamelberg et al. 2004), metadynamics (Laio and Parrinello 2002), and multicanonical MD methods. In the extended tREM, the replicas have different physical parameters [van der Waals radius (Itoh et al. 2010); coulomb potentials (Itoh and Okumura 2013a), model resolution (Lyman et al. 2006), and umbrella potentials (Okumura and Itoh 2013)] instead of different temperatures. Exchange between replicas involves transition between the different simulation runs under the detailed-balance condition. These extended methods of tREM could be generalized as the replica exchange umbrella sampling method (Sugita et al. 2000) and Hamiltonian replica exchange method (Fukunishi et al. 2002). In aMD, a particular potential energy is scaled by a user-defined factor, and the scaled energy is used throughout the simulation. After the aMD simulation, the probability with the genuine energy is reconstructed. In metadynamics, the bottom level of an energy basin along a reaction coordinate is raised gradually during simulation. Raising the bottom level pushes the conformation out of the energy basin where it is trapped and, consequently, the system can travel widely along the reaction coordinate. In the multicanonical method, the potential energy travels randomly and continuously in a wide potential energy range. How to achieve such sampling by the multicanonical method is described in the next section.

The basics of the multicanonical method were developed Berg and Neuhaus (1992) and are based on the Monte Carlo scheme and applied to the simple statistical mechanical Potts model. Hansmann and Okamoto (1993) applied the multicanonical method for searching the lowest energy conformation of the 5-residue peptide, Met-enkephalin. The multicanonical method was then incorporated into a MD scheme (Hansmann et al. 1996; Nakajima et al. 1997). The multicanonical MD (McMD) simulation has been improved and applied to larger protein systems with an all-atom model in an explicit solvent. The McMD method has been used to obtain the FELs for a chameleon sequence (Ikeda and Higo 2003), the binding between lysozyme and a saccharide (Kamiya et al. 2008), and domain-size protein folding (Ikebe et al. 2011a). In the following section, we provide a concise description of McMD. In subsequent sections, we introduce the theory and application of more efficient enhanced sampling methods, namely trivial trajectory parallelization (TTP) (Ikebe et al. 2011b), virtual-system coupled McMD (V-McMD) (Higo et al.

2013), and adaptive lambda square dynamics (ALSD) (Ikebe et al. 2014).

McMD simulation

To construct an accurate FEL, a well-equilibrated conformational ensemble must be obtained. McMD (Hansmann et al. 1996; Nakajima et al. 1997) provides a way to overcome energy barriers in the conformational space by scaling the potential energy, thereby allowing markedly faster sampling of the ensemble than is possible using conventional canonical MD. Here we briefly explain McMD simulation.

McMD is an MD simulation at constant temperature T_0 with a modified potential energy E_{mc} ,

$$\begin{aligned} E_{mc} &= \lambda_{mc}(E, T_0)E = \left(1 + \frac{RT_0 \ln P_c(E, T_0)}{E}\right)E \\ &= E + RT_0 \ln P_c(E, T_0), \end{aligned} \quad (1)$$

instead of the original potential energy, E . In this case, the forces acting on atoms are derived from derivatives of E_{mc} rather than E . We represent E_{mc} with three equivalent expressions in Eq. 1. The first expression means that McMD modulates E by a scaling factor λ_{mc} . When $0 < \lambda_{mc} < 1$, the potential energy is scaled down and, consequently, the forces are also scaled down. This scaling enhances conformational change of the simulated system because it weakens interaction energy to stabilize the conformations. As a result, the system explores a greater conformational space than canonical MD. The second expression provides details on λ_{mc} , where R is a gas constant and $P_c(E, T_0)$ is a probability distribution function of energy for a well-equilibrated canonical ensemble (canonical probability distribution) at T_0 on the E axis, although the functional form of $P_c(E, T_0)$ is unknown a priori. The factor λ_{mc} is designed to realize a random walk of the system on the E axis [for more details on the derivation, the reader is referred to the review of Higo et al. (2012)]. The random walk efficiently facilitates the overcoming of energy barriers and the search for various energy minima in the conformational space. In a practical McMD procedure, a canonical MD simulation run is first performed to approximate the unknown function $P_c(E, T_0)$, followed by gradual refinement of $P_c(E, T_0)$ by iterative McMD runs. After these iterative runs, a productive McMD simulation run using the refined $P_c(E, T_0)$ is performed to sample the ensemble. The productive run provides an ensemble (i.e., multicanonical conformational ensemble) which realizes a flat multicanonical probability distribution $P_{mc}(E, T_0)$ on the E axis due to the random walk. The third expression explains how the McMD formalism can be regarded as a type of complex umbrella sampling MD (Chandler 1987; Deng and Roux 2009) with an umbrella potential $RT_0 \ln P_c(E, T_0)$. A canonical ensemble at an arbitrary temperature T can be reconstructed

from the multicanonical conformational ensemble with a reweighting scheme (Higo et al. 2012) as

$$\begin{aligned} P_c(E, T) &= P_{mc}(E, T_0)W_{mc}(E, T_0, T) \\ &= P_{mc}(E, T_0)\exp\{\lambda_{mc}(E, T_0)E/RT_0\}\exp(-E/RT)C \end{aligned} \quad (2)$$

where W_{mc} is a reweighting factor for McMD and C is a normalization constant. Assigning a probability $P_c(E, T)$ to each snapshot, whose real energy is E in the multicanonical conformational ensemble, the canonical conformational ensemble at T is constructed, and through mapping of these reweighted snapshots to a conformational space, a FEL can be generated.

Trivial trajectory parallelization of McMD

Although McMD allows for more efficient conformational sampling than canonical MD it still requires long simulation times for large systems in order to generate statistically reliable data. Parallel computing of MD simulation is a commonly used technique to speed up conformational sampling. Parallel computing with MPI (Message Passing Interface Standard) and/or OpenMP (Open Multi-Processing) is generally implemented in MD simulation programs with multiple central processing units (CPUs). The MD technique with a general-purpose graphics processing unit (GPGPU) (Götz et al. 2012; Mashimo et al. 2013; Pall et al. 2014; Salomon-Ferrer et al. 2013) is growing in importance due to multiple processors being less expensive than CPUs. Hardware architectures specialized for MD, such as Anton (Shaw et al. 2009) and MD-GRAPe (Narumi et al. 2006), have enabled extremely long-time scale MD simulations (Kikugawa et al. 2009; Lindorff-Larsen et al. 2011; Shaw et al. 2010). These specialized architectures speed up a single MD simulation run and provide an increased number of conformations from the long MD trajectory. Combining these architectures and McMD would further enhance efficient generation of FELs.

Another form of parallel computing is to perform multiple MD runs. The benefit of this type of parallelization is that no additional device need be implemented in the MD code and no costly machine is required. The method provides an increased number of conformations, not from a single long trajectory, but from independent multiple trajectories, each of which may be short. The parallel computing method with multiple McMD runs, trivial trajectory parallelization of McMD (TTP-McMD) (Higo et al. 2009; Ikebe et al. 2011b), provides a conformational ensemble where the multiple McMD runs use a common E_{mc} and are started from different initial conformations that are widely distributed in conformational space (Fig. 2). Our group has empirically examined the efficiency of TTP-McMD for a short peptide system (Higo et al. 2009) and

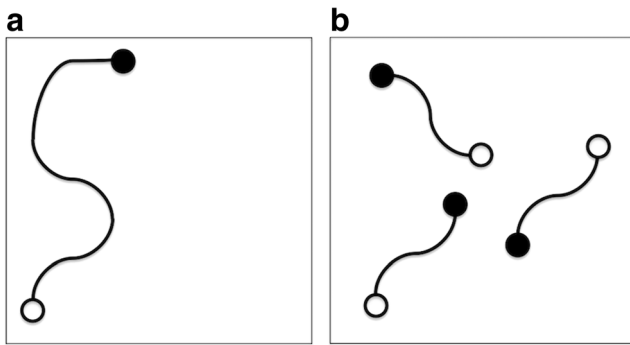


Fig. 2 Schematic figures showing sampling with multicanonical MD (McMD; **a**) and trivial trajectory parallelization-McMD (TTP-McMD; **b**). Boxes represent conformational spaces, curved lines capped at each end by circles show McMD simulation trajectories. Open and filled circles at the termini of the trajectories Initial and final snapshots of the McMD trajectories, respectively. McMD explores conformational space using of a single long trajectory. In contrast, TTP-McMD explores conformational space that by N -independent trajectories that are widely distributed in the conformational space. TTP-McMD with N runs generally constructs a more accurate FEL than a McMD run which is N -times longer than a single TTP-McMD run

have provided a theoretical framework for its interpretation (Ikebe et al. 2011b). The latter study demonstrated that TTP-McMD with N runs generates a more accurate FEL than a single McMD run N -times longer than each TTP-McMD run. In the iterative simulations of TTP-McMD, the final snapshot of each trajectory is used as the initial one for each following iterative run. This treatment is of essential importance for efficient sampling (Ikebe et al. 2011b). As mentioned above, this method requires neither special hardware nor highly adjusted software to improve sampling efficiency. Furthermore, its parallelization efficiency is always 100 %. The TTP procedure enables McMD to search FELs of complex systems, such as proteins and their substrates. Moreover, TTP is readily applicable to other sampling methods, such as V-McMD and ALS (to be discussed later in this review).

Applications of TTP-McMD to interactions between intrinsically disordered proteins and its partner protein

In this section, we describe the application of TTP-McMD to systems of intrinsically disordered proteins (IDPs) (Uversky 2013; Wright and Dyson 2015), which are involved in cellular signaling and transcriptional machinery (Dunker et al. 2002). In general, an ordinary protein folds into a specific and stable conformation (i.e., a specific tertiary structure) under physiological conditions, and the ordered conformation determines its unique function. We designate this specific structural conformation as C_{ss} in this review. In contrast, IDPs have no specific conformation and interconvert among semi-stable conformations in the free state (unbound state). When the

IDP interacts with its partner protein, it often folds into the C_{ss} form and the complex structure is formed. This folding mechanism of IDP accompanied by binding to the partner is referred to as the “coupled folding and binding” mechanism. Two representative schemes have been suggested to explain this mechanism, namely, population shift (Bosshard 2001; James and Tawfik 2003; Monod et al. 1965) and induced fit (Koshland 1960). In the former scheme, C_{ss} is intrinsically included even in the conformational ensemble of the unbound single-chain IDP, even though its population for C_{ss} may be small. The partner protein then selects C_{ss} when the complex is formed. In the latter scheme, IDP first binds to the partner with conformations different from C_{ss} , and then C_{ss} is induced. At the present time it is still unclear which scheme is appropriate. The detailed binding mechanisms have been investigated in two experimental settings where TTP-McMD was applied to two IDPs: the N-terminal repressor domain of neural restrictive silencer factor (NRSF) (Higo et al. 2011) and the phosphorylated kinase-inducible domain (pKID) of the cyclic-AMP response element binding protein (CREB) (Umezawa et al. 2012).

The complex structure of NRSF and its partner protein, the paired amphipathic helix (PAH) domain of mSin3 (Nomura et al. 2005), has been determined experimentally (Fig. 3a). For simplicity, we refer here to the molecule as mSin3. To obtain the conformational ensembles of NRSF, TTP-McMD of NRSF was conducted in the absence and presence of mSin3 with these two systems denoted as single-chain NRSF and the NRSF-mSin3 system, respectively. To initiate either simulation, the NRSF conformation was randomized. In the NRSF-mSin3 system, the two proteins were separated from each other. The proteins were also immersed in an explicit solvent for both systems. No artificial restraint to movement was placed on NRSF while the conformation of mSin3 was weakly restrained to the experimental structure throughout the simulation (discussed later in this review). From the TTP-McMD simulations, FELs of NRSF were constructed for both systems. The FEL of the NRSF-mSin3 system could be clustered into a number of states. Importantly, the largest cluster (i.e., thermodynamically most stable cluster or the lowest free-energy cluster) corresponded to the native-like complex structure. Three super clusters were apparent, one of which involved the largest cluster, and free-energy barriers existed among the super clusters. The existence of the super clusters led to the proposal of the following scenario for complex formation: once an encounter complex, which is involved in a cluster other than the main (and largest) cluster is formed, NRSF should change in conformation to overcome the free-energy barriers among the clusters to reach the native-complex structure.

The single-chain NRSF system produced a conformational ensemble consisting of various conformational clusters characterized by α -helix and β -strand secondary structural

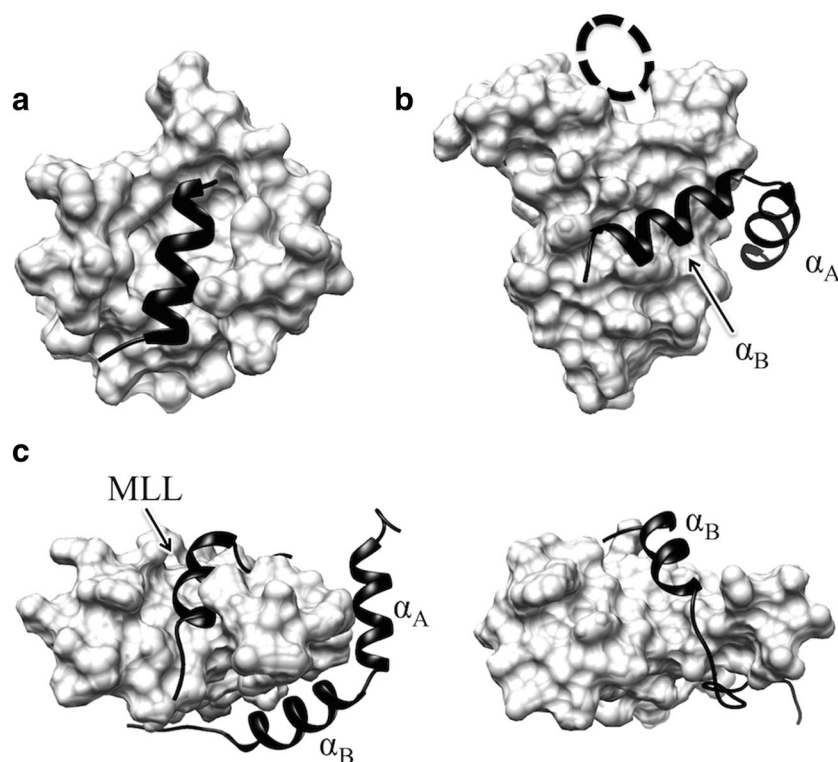


Fig. 3 Folded structures of target intrinsically disordered proteins (IDPs) [N-terminal repressor domain of neural restrictive silencer factor (NRSF) and phosphorylated kinase-inducible domain (pKID) of CREB] in the complex with their partner proteins. **a** Native complex structure of NRSF is shown as *black ribbon* on the mSin3 surface (PDB ID: 2CZY). NRSF is bound to the cleft of mSin3. **b** *Black ribbon* represents the native structure of pKID on the KIX surface (PDB ID: 1KDX). The two helices (α_A , α_B) are attached in the shallow concave on the surface of KIX. *Dashed-line circle* MLL binding site. **c** *Left panel*

Native structure of MLL in the triple complex with KIX and pKID (PDB ID: 2LXT), corresponding to the top view of **b**. The MLL and pKID structures are represented as *black ribbons* and the KIX as its surface. *Right panel* Conformation of pKID binding to the MLL-binding site of KIX, which was taken from the TTP-McMD ensemble of the pKID-KIX system. The viewing angle at right is the same as that of the left. The shapes of the KIX surfaces are different between the left and the right views because KIX has dynamic side chains and its C-terminus is flexible in the simulation

elements as well as other structural elements. Interestingly, a majority of those NRSF conformations were found in the ensemble of the NRSF–mSin3 coexisting system. Therefore, the population shift paradigm may readily incorporate the existence of various encounter complexes. However, the conformation of the encounter complex should undergo sufficient change to reach the native complex, as mentioned above. In this context, the induced fit takes place and folding and binding in the NRSF–mSin3 system proceeds in a way that the conformational shift and induced fit are coupled.

The other application of TTP-McMD targeted pKID. pKID in complex with its partner, the KIX domain of CREB binding protein (CBP), adopts a specific conformation composed of N- and C-terminal helices (these regions are denoted as α_A and α_B , respectively, even when the segments do not adopt helical conformation in a simulation snapshot) and a loop region spacing these helices (Radhakrishnan et al. 1997) (Fig. 3b). Using the same procedure as for the NRSF systems, we conducted TTP-McMD of two systems of pKID in the absence and presence of the KIX domain, denoted as the single-chain pKID and pKID-KIX systems, respectively.

The FEL of the single-chain pKID system showed that the helical propensity for the α_B region is considerably smaller than that of the α_A region, which is in agreement with experimental results (Radhakrishnan et al. 1998). In the pKID-KIX coexisting system, FEL illustrated a number of semi-stable bound forms, including the native-like complex conformation. Subsequent analyses determined that the coupled folding and binding mechanism of the α_A and the α_B regions are different. The α_A region bound with KIX exhibited large fluctuations, and therefore discernment of the appropriate mechanism for the α_A region could not be concluded. When categorized, the binding of the α_A region could be seen to follow the population shift mechanism. To the contrary, the α_B region follows the induced fit mechanism. The α_B region folds into a helix when it binds to either one of two distinct hydrophobic sites on the KIX domain: the native binding site (genuine binding site) in the native structure and another binding site, whose location is far from the native binding site (Fig. 3b, c). Interestingly, the latter corresponds to the binding site for another KIX-binding transcription factor, the mixed-lineage leukemia protein (MLL) (Fig. 3b, c), where a segment of MLL

adopts a helical conformation (Brüschweiler et al. 2013; De Guzman et al. 2006) (PDB IDs: 2AGH, 2LXS, 2LXT). Experimental results suggest that pKID also binds to the MLL-binding site (Sugase et al. 2007), although a detailed structure of the complex has not been determined. A benefit of molecular simulation is that the simulation can provide atomistic information for the molecular structure. Such simulations suggested a conformation of pKID on the MLL-binding site that was helical and overlapped well with the bound form of the MLL segment (Fig. 3c). Thus, TTP-McMD yielded pictures that furthers our understanding of the binding mechanism and conformational characteristics of IDPs; such information cannot be obtained from experiments.

However, we should add a note of caution regarding the technical treatment of the partner proteins in both the NRSF-mSin3 and pKID-KIX coexisting systems. The partner molecules, mSin3 and KIX, respectively, are ordered proteins. Since the McMD simulation scales the whole energy of the systems (i.e., the whole energy is elevated to a high temperature during the simulation), the partner proteins may undergo unfolding. Therefore, the conformation of the ordered partner proteins is restrained weakly around the native structure to avoid such undesired sampling for the unfolded state of the partner, while the targeted IDP is not restrained. Herein, our cautionary note is that very strong structural restraints on the partner can yield artificial outputs. In the native NRSF-mSin3 complex, the natively bound NRSF is buried in the closed crevice of the mSin3 surface. Thus, if the structural restraints are so strong as to freeze the open/close fluctuations of the crevice, NRSF might not reach the natively bound form in the simulation. Also in the pKID-KIX system, the strong restraint seems to be inadequate if the research is focused on an allosteric mechanism of the KIX domain (Law et al. 2014). Therefore, a method free from the application of structural restraints is useful to remove any artificial effects induced within the sampling. Such a restraint-free method is described in the section **Adaptive Lambda Square Dynamics** of this review.

Theory of virtual-system coupled McMD

Enhanced sampling methods, such as McMD, provide a flat probability distribution over a wide range of energy or reaction coordinates, meaning that the system walks randomly over a wide range. In McMD, the system can search low-energy conformations by overcoming the energy barriers. If we imagine that the McMD simulation starts from a non-equilibrated conformation in a low-energy region, then the energy of the system transitions to a high-energy region when climbing an energy barrier, and to a low energy region when

the barrier is overcome. Since the McMD simulation is a method to obtain an equilibrated ensemble at an arbitrarily temperature, these up-and-down motions of energy lead the system to a relaxed conformation in the low-energy range. Thus, traffic between the low- and high-energy regions is very important for the system to rapidly reach a well-equilibrated conformational ensemble. However, some systems may have an energy surface where the low-energy regions and high-energy regions are connected by a very narrow pathway. In such cases simple McMD sampling may not sufficiently improve the sampling efficiency (Higo et al. 2012; Higo and Nakamura 2012).

To achieve the efficient traffic in multicanonical sampling, a variant of McMD, named “virtual-system coupled multicanonical molecular dynamics” (V-McMD), was developed (Higo et al. 2013). First, a brief introduction of the virtual system: imagine an abstract degree of freedom (i.e., virtual degree of freedom) v , which expresses a system (virtual system). Because v does not actually exist, we need not define a specific system for the virtual system. It is assumed that v varies discretely ($v=v_1, v_2, \dots$), and the discrete values are termed “virtual states”. The number and the energy of the virtual states are denoted as Nv and E_{v_i} ($i=1, 2, 3, \dots, Nv$), respectively.

Note that the entire system (the molecular system + the virtual system) is specified by the potential energy, E , of the molecular system and the virtual state, v_i . If there are no direct interactions between the molecular system and the virtual system, E is apparently expressed only by the coordinates of the molecular system. We then design the energy of the entire system as

$$E_{v_i}(E) = E + RT_0 \ln \left\{ \frac{P_c(E)}{g_i(E)} \right\} = RT_0 \ln \left\{ \frac{n(E)}{g_i(E)} \right\}, \quad (3)$$

where $P_c(E)$ is the canonical probability distribution function of E and $g_i(E)$ is a arbitrary function of E under the condition that the virtual system is in the i -th virtual state. Note that $g_i(E)$ is defined by both E and the virtual state number i , which means that E_{v_i} involves interference between the molecular and virtual systems despite E not involving v_i . The resulting probability distribution of E in the i -th virtual state, is given as:

$$P_{v_i}(E) \propto n(E) \exp \left(-\frac{E_{v_i}}{RT_0} \right) \propto g_i(E). \quad (4)$$

When $g_i(E)$ is a constant, a flat energy distribution results, similar to that for the conventional McMD simulation.

Up to this point, the method for time development of the system has not been given. In V-McMD, the time development for the molecular system is performed by integrating the Newtonian equations as usually done for MD. However, the time development for the virtual system is done with the Monte Carlo scheme, with the transition probability from the

i -th to j -th virtual states set so that the transition satisfies the detailed balance, as per Eq. 5.

$$P_{V_i \rightarrow V_j}(E) = \min \left[1, \exp \left(-\frac{E_{V_j} - E_{V_i}}{RT_0} \right) \right] = \min \left[1, \frac{g_j(E)}{g_i(E)} \right]. \quad (5)$$

In actual sampling, the molecular system moves according to the multicanonical MD scheme for a time interval of τ , during which the virtual system is kept in a virtual state. Then, at the end of the interval, the transition among virtual states is attempted according to the Monte Carlo scheme, at which point the molecular system is then fixed at the current conformation. After the trial, the molecular system time-develops for another interval of τ , and so on. What is important is that we can set the function shape for g_i arbitrarily because the virtual system is literally a virtual object. $g_i(E)$ and $g_{i+1}(E)$ are designed as

$$g_i(E) = \begin{cases} 0 & (E_{V_i}^{Max} < E) \\ 1 & (E_{V_i}^{Min} \leq E \leq E_{V_i}^{Max}) \\ 0 & (E < E_{V_i}^{Min}) \end{cases} \text{ and } g_{i+1}(E) = \begin{cases} 0 & (E_{V_{i+1}}^{Max} < E) \\ 1 & (E_{V_{i+1}}^{Min} \leq E \leq E_{V_{i+1}}^{Max}) \\ 0 & (E < E_{V_{i+1}}^{Min}) \end{cases}, \quad (6)$$

respectively, where the boundary parameters are aligned as $E_{V_i}^{Min} < E_{V_{i+1}}^{Min} < E_{V_i}^{Max} < E_{V_{i+1}}^{Max}$. We call the overlapping range $[E_{V_{i+1}}^{Min}, E_{V_i}^{Max}]$ between non-zero ranges of $g_i(E)$ and $g_{i+1}(E)$ as the “zone”, then the transition probability from the i -th to j -th virtual states is 1 in the zone, and 0 outside the zone (illustrated in Fig. 4). Thus, the multicanonical distributions at the i -th and j -th virtual states are connected, thereby maintaining the detailed balance criterion. Here, the key point of V-McMD is that the connection is achieved with the probability of 1 so that the traffic is driven with keeping the equilibrium condition (for details, see Higo et al. 2013).

In practical terms, multiple runs (i.e., TTP procedure) can be applied to V-McMD in a straightforward manner—a process referred to as “TTP-V-McMD”. One trajectory taken from the multiple V-McMD runs migrates to another trajectory, where the migration occurs unconditionally. Therefore, it should be once again noted that this migration facilitates the traffic (the low-to-high and the opposite transition of the energy) during simulation, and it yields more rapid equilibration than conventional MD simulations. The equilibration speed is discussed in the following section.

Applications of virtual-system coupled methods for a peptide homodimer

In the study of Hoh et al. (2004), the target for TTP-V-McMD was a bioactive peptide, Endothelin-1 derivative (KR-CSH-

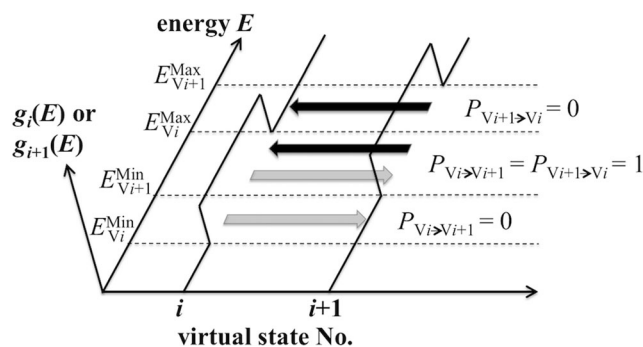


Fig. 4 The transition probabilities between the i -th and $(i+1)$ -th virtual states in virtual-system coupled McMD (V-McMD). For definition of the symbols, see text in section Theory of virtual-system coupled McMD. Functions $g_i(E)$ and $g_{i+1}(E)$ are defined in Eq. 6. Dashed lines denote the energy boundaries for the two virtual states: $E_{V_i}^{Min}$ and $E_{V_i}^{Max}$ for $g_i(E)$, and $E_{V_{i+1}}^{Min}$ and $E_{V_{i+1}}^{Max}$ for $g_{i+1}(E)$. When the system with energy E attempts to transfer from the i -th to $(i+1)$ -th state (gray arrow), the transition probability $P_{V_i \rightarrow V_{i+1}}(E)$ equals 1 in the zone ($E_{V_{i+1}}^{Min} \leq E \leq E_{V_i}^{Max}$) and 0 outside the zone. On the other hand, from the $(i+1)$ -th to i -th state (black arrow), the transition probability $P_{V_{i+1} \rightarrow V_i}(E)$ equals 1 in the zone and 0 outside the zone. Thus, the transition between the two states succeeds undoubtedly in the zone. If three virtual states exist ($N_V = 3$), there are zones between the first and second states and between the second and third states. The system can then travel from the first/third to the third/first states via two sequential transitions as from the first/third to the second and from the second to the third/first

ET1), which forms a stable antisymmetric homodimer (Fig. 5a). The dimerization of KR-CSH-ET1 was investigated by Higo et al. (2013) to test the equilibration speed of TTP-V-McMD as compared with TTP-McMD. Two molecules of KR-CSH-ET1 were separated from each other and immersed in an explicit solvent for the initial conformation of TTP-V-McMD, and both TTP-V-McMD and TTP-McMD were performed with the same number of multiple runs. The protocol is described in detail in Higo et al. (2013). Although both

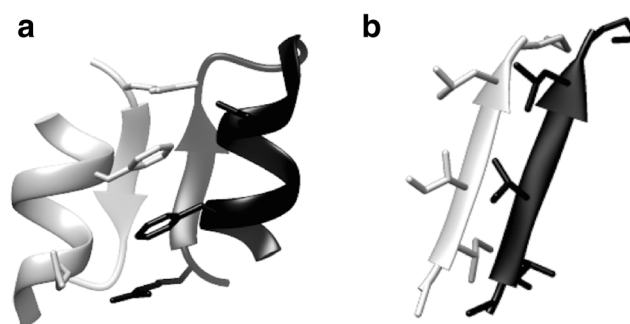


Fig. 5 Native structures of target homodimers of the Endothelin-1 variant (KR-CSH-ET1). **a** The X-ray crystal structure of KR-CSH-ET1 (PDB ID: 1T7H), with one chain shown in gray and the other in black. The inter-molecular hydrophobic interaction of phenylalanine and the salt bridge between glutamate and arginine stabilizes the homodimer. Their side chains are displayed as the stick model. The inter-molecular main-chain interaction with the anti-parallel β -sheet also contributes to the stability. **b** The two segments of Amyloid- β peptide (sequence: ALA-ILE-ILE-GLY-LEU-MET) shown in gray and black, respectively, as taken from amyloid fibril structure (PDB ID: 2Y3J)

TTP-V-McMD and TTP-McMD produced the native-like dimer for the most thermodynamically stable conformation, the convergence speed of TTP-V-McMD was faster than that of TTP-McMD.

The virtual-system and trajectory-parallelization procedures are extendable to other enhanced sampling methods. Here we describe the method for adaptive umbrella sampling (AUS), which is abbreviated as TTP-V-AUS (Higo et al. 2015). The TTP-AUS (i.e., trajectory-parallelization of the conventional AUS) yields a random walk over a wide range of reaction coordinates. In TTP-V-AUS, the range of the reaction coordinate is divided into the virtual states, as done for the energy axis in TTP-V-McMD, and the transitions among the virtual states are defined as the same way as for TTP-V-McMD (for details, see Higo et al. 2015). The traffic was compared between TTP-V-AUS and the conventional TTP-AUS with sampling association/dissociation of two Amyloid- β peptides in an explicit solvent, where the separation distance between the two peptides was adopted as the variable for the sampling enhancement. This peptide is known to form an amyloid fibril (Fig. 5b). The free-energy landscape shows that various complex forms, such as the antiparallel β sheet, parallel β sheets, α -helix, and other complex forms, are distributed in a conformational space. In comparison between TTP-V-McMD and TTP-McMD, TTP-V-AUS realized quicker convergence than TTP-AUS. Thus, it would appear that the virtual-system method improves the traffic along the reaction coordinate in comparison to methods without the virtual system (Higo et al. 2015).

Adaptive lambda square dynamics simulations

Adaptive lambda square dynamics (Ikebe et al. 2014) has been developed from two free energy calculation methods, namely, adaptive umbrella sampling (Mezei 1987) and λ dynamics (Kong and Brooks CL III 1996), and is an effective method for sampling the ensemble of a complex system that is not dependent upon the application of structural restraints. As previously described, the imposition of strong restraints may cause artifacts. ALSD enhances sampling only for a part of the system. For this purpose, the potential energy E is divided into a number of terms, and these terms are differently scaled. The division of E is accomplished through two stages, energy term division (ETD) and spatial energy division (SED), which are sequentially processed.

A potential energy is decomposed usually as

$$E = E_{\text{bond}} + E_{\text{angle}} + E_{\text{tors}} + E_{\text{imp}} + E_{\text{ele}} + E_{\text{vdW}}, \quad (7)$$

where the subscripts “bond”, “angle”, “tors”, “imp”, “ele”, and “vdW” stand for the bond length, bond angle, torsion angle, improper torsion angle, electrostatic, and van der

Waals potential energies, respectively. In the ETD stage, the terms are classified into two groups: interesting terms ($E_{\text{int_term}}$) and others ($E_{\text{rest_term}}$). In the original article on ALSD (Ikebe et al. 2014), three terms, namely, E_{tors} , E_{ele} , and E_{vdW} , were selected for $E_{\text{int_term}}$ because these three terms may affect the global conformation of the system more than the other terms do: $E_{\text{int_term}} = E_{\text{tors}} + E_{\text{ele}} + E_{\text{vdW}}$. E is then expressed as:

$$E = E_{\text{int_term}} + E_{\text{rest_term}}, \quad (8)$$

In SED, the system is divided spatially into an interesting region (region A) and the other region (region B) in order to selectively enhance the sampling of region A. For a complex system, it would be reasonable that the substrate (ligand) and the other components (receptor protein, ions, and solvent molecules) are set as region A and B, respectively (Fig. 6a). According to this spatial division, each interaction term in $E_{\text{int_term}}$ is assigned to “intra-A”, “inter-A–B”, or “rest”. For a pairwise interaction between two atoms, the interaction is assigned to intra-A or rest, when both atoms are in region A or B. When the atoms belong to different regions, the interaction is done to inter-A–B. To assign a torsion angle interaction that is specified by a straight chain of four atoms, A–B–C–D, bound with covalent bonds, atoms A and D are considered to interact. All interactions in $E_{\text{rest_term}}$ are assigned to rest. As a result, E is decomposed into three terms:

$$E = E_{\text{intAA}} + E_{\text{intAB}} + E_{\text{rest}}, \quad (9)$$

where the subscripts “intAA”, “intAB”, and “rest” denote intra-A, inter-A–B, and rest, respectively.

ALSD is an MD simulation on an extended coordinate space (\mathbf{r}, λ) at constant temperature T_0 with a modified Hamiltonian:

$$H_{\text{ALSD}}(\mathbf{r}, \dot{\mathbf{r}}, \lambda, \dot{\lambda}) = \lambda^2 E_{\text{intAA}}(\mathbf{r}) + \lambda E_{\text{intAB}}(\mathbf{r}) + E_{\text{rest}}(\mathbf{r}) + K(\dot{\mathbf{r}}) + \frac{1}{2} m_{\lambda} \dot{\lambda}^2 + RT_0 \ln P_{\text{ex}}(\lambda, T_0), \quad (10)$$

where \mathbf{r} and $\dot{\mathbf{r}}$ are sets of atomic coordinates and their velocities, respectively, λ is a scaling factor for ALSD as an extra dynamic variable with the fictitious mass m_{λ} , K is the kinetic energy of the system, $\dot{\lambda}$ is the velocity of λ , and $P_{\text{ex}}(\lambda, T_0)$ is a canonical probability distribution on the λ axis obtained from a well-equilibrated canonical MD simulation with the scaled potential energy $E_{\text{ex}} (= \lambda^2 E_{\text{intAA}} + \lambda E_{\text{intAB}} + E_{\text{rest}})$ in the (\mathbf{r}, λ) space. During an ALSD simulation, the scaling factor λ moves as a variable obeying H_{ALSD} and scales only the potential energy terms with respect to region A (i.e., E_{intAA} and E_{intAB}). When $0 < \lambda < 1$, E_{intAA} and E_{intAB} are scaled down and only the conformational changes of region A are enhanced. Meanwhile, the conformation of region B is

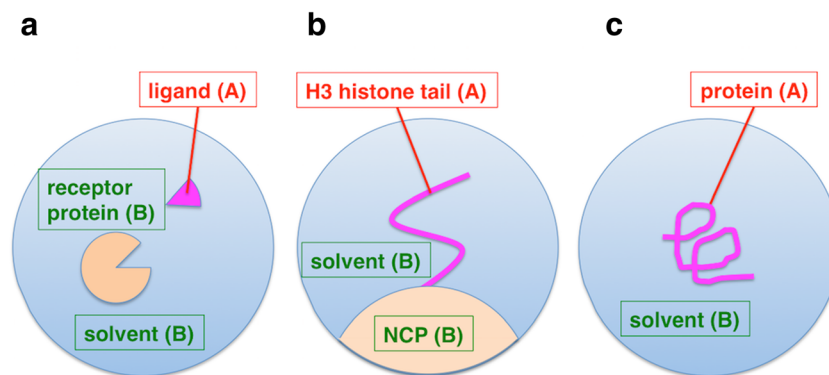


Fig. 6 Examples of spatial energy division (SED) in adaptive lambda square dynamics (ALSD). ALSD enhances conformational sampling of region A while maintaining the conformation of region B in thermal fluctuations at the simulation temperature. **a** An example of SED for a complex system consisting of receptor protein (*region B*) and the ligand (*region A*). McMD tends to enhance conformational changes not only for the ligand but also for the protein. Thus, in McMD for such complex system, weak restraint forces are applied to the protein to maintain the conformation. On the other hand, ALSD can maintain the protein conformation without the restraint forces. **b** SED for an H3 histone tail on a nucleosome system. The histone tail and the nucleosome core particle (NCP) were assigned as region A and B, respectively. This system is highly polarized: negatively charged DNA covers the surface of the NCP, and the histone tail includes a many positively charged lysine and arginine residues. McMD could not sufficiently sample conformations of the histone tail because the histone tail is stuck to DNA by strong electrostatic interactions. To the contrary, SED in ALSD realized dissociation of the histone tail from DNA and sampled

the various conformations. See Section [Adaptive lambda square dynamics simulations](#) for details. **c** SED in ALSD, as originally described in Ikebe et al. (2014). ALSD was originally developed as a more efficient conformational sampling method than McMD. In the original article, ALSD was applied to a short peptide (poly-lysine decapeptide, region A) system in explicit solvent (region B) to compare its sampling efficiency with that by McMD. To sample various conformations of the peptide, McMD must explore a wide potential energy range of the whole system, composed of the peptide and many water molecules. To the contrary, ALSD selectively enhances the conformational sampling of the peptide by exploration only in a narrower potential energy range with respect to the peptide. This focused sampling reduces the conformational space to be sampled for the solvent and concomitantly increases the sampling efficiency for the peptide. This system was also used in this review for ALSD simulations to investigate the efficient operation procedure of energy term division (ETD) on the sampling efficiency

maintained in thermal fluctuations at T_0 without structural restraint forces. As with McMD, H_{ALSD} is designed to realize a random walk on the λ axis. Iterative ALSD runs are required to estimate a priori the unknown $P_{\text{ex}}(\lambda, T_0)$. A canonical ensemble at $\lambda = 1$ is reconstructed by a reweighting scheme. The reader is referred to the original article by Ikebe et al. (2014) for more details.

An application of ALSD to a highly polarized system

Ikebe et al. (2015) applied ALSD to conformational sampling of the histone tails on a nucleosome system (Fig. 7). The nucleosome is a fundamental conformational unit of DNA in a eucaryotic nucleus and is composed of DNA wrapped around histone proteins. Although the structures of the core region of the nucleosome have been determined by X-ray crystallography, the flexible N-terminal regions of the histone proteins, referred to as histone tails, have no specific stable conformation. It is known that chemical modifications on the histone tails induce conformational changes of the histone tail itself, DNA, and nucleosome. These conformational changes concomitantly regulate biologically important DNA functions, such as transcription, duplication, splicing, and DNA repair (Jenuwein and Allis 2001; Sidoli et al. 2012; Strahl and Allis 2000). To elucidate the mechanism of the

conformational changes by the chemical modifications, it is necessary to obtain the conformational ensemble of the histone tails on a nucleosome.

However, it had been difficult to sample the ensemble because the system is highly polarized compared to general protein–substrate complex systems: negatively charged DNA covers the surface of the nucleosome, and the histone tails include a lot of positively charged lysine and arginine residues. The attractive electrostatic interactions between them make the histone tails stick to DNA and constrain the conformational changes of the histone tails. To realize efficient conformational sampling, it is important to dissociate the histone tails from the DNA during the simulation. An McMD simulation of a histone tail (H3 histone tail) on the nucleosome system could not sufficiently sample a variety of conformations of the histone tail (data not shown). The H3 histone tail has 13 positively and no negatively charged amino acids in the N-terminal 40 residues. To dissociate the histone tail from the DNA, the histone tail must break all electrostatic contacts with DNA simultaneously. Although the McMD equally scaled down all interactions, including electrostatic interactions between DNA and the histone tail, when $\lambda_{\text{mc}} < 1$, the scaling was not sufficient to dissociate them even when the λ_{mc} was quite small ($=0.5$).

To the contrary, ALSD, which sets the histone tail to region A (Fig. 6b) and $E_{\text{int_term}} = E$, was able to dissociate them and

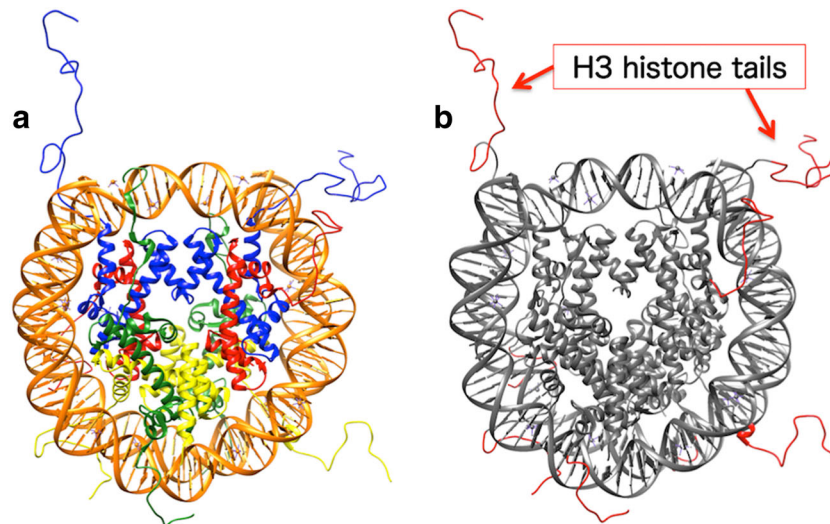


Fig. 7 **a** An X-ray crystal structure of a nucleosome (PDB ID: 1KX5). The nucleosome is composed of 147 base pairs of DNA (*orange*) wrapped around a histone octamer, which is composed of two copies of H3 (*blue*), H4 (*red*), H2A (*green*), and H2B (*yellow*) histone proteins. **b** Locations of histone tails in the nucleosome. Although crystal structures

of the nucleosome core region (*gray*) have already been determined, the N-terminal regions (histone tails, *red*) of the histone proteins have not yet been determined due to their innate structural flexibility. The histone tail conformations in 1KX5 are modeled ones. In the applied research of ALSD, conformational sampling of an H3 histone tail was performed

sampled a variety of conformations of the H3 histone tail. Although the ALSD scaled down interactions of DNA with the histone tail when $\lambda < 1$, it did not scale down those with ions and solvent, unlike McMD: in ALSD, DNA preferentially interacted with positive ions and positively polarized hydrogen atoms of solvent water rather than the histone tail, which allowed the histone tail to dissociate from DNA. This result shows that ALSD is a promising method for highly polarized systems, which pose difficulties to sampling by other methods, such as McMD.

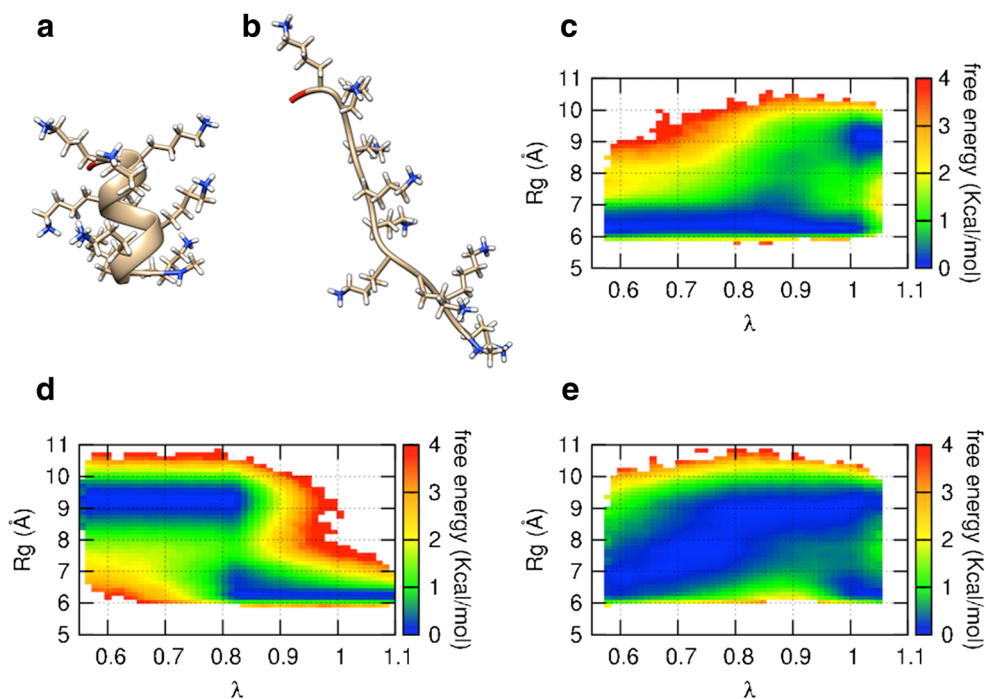
Efficient operation procedures for ALSD

Here we note efficient operation procedures for ALSD. Although $E_{\text{int_term}}$ can be set arbitrarily in theory, E_{ele} and E_{vdw} should not be individually treated in the ETD stage. To present the problem, we performed three ALSD simulations of a short peptide (poly-lysine decapeptide), which is the same one as that studied in the original ALSD article (Ikebe et al. 2014), at different settings of ETD: $E_{\text{int_term}} = E_{\text{ele}}$ (referred to as “ALSD_{ele}”), $E_{\text{int_term}} = E_{\text{vdw}}$ (“ALSD_{vdw}”), and $E_{\text{int_term}} = E_{\text{ele}} + E_{\text{vdw}}$ (“ALSD_{both}”). In these simulations, the peptide and the solvent were selected as region A and B, respectively (Fig. 6c). These simulations were performed under the same simulation condition: seven iterative runs to refine $P_{\text{ex}}(\lambda, T_0)$ and the productive run to sample the ensemble with TTP (10 ns \times 64 trajectories). Note that the total simulation time for each simulation was shorter than that in the original ALSD article (30 ns \times 72 trajectories). In theory, these ALSD simulations provide the equivalent canonical ensembles at $\lambda = 1$ if the simulation times are sufficiently long.

However, the different setting of ETD affects the sampling efficiency when the simulation times are insufficient.

The correlations between λ and the radius of gyration (R_g) of the peptide are shown as FESs in Fig. 8. The free energy was calculated as $-RT_0 \ln P(\lambda, R_g)$ and normalized for each λ where T_0 was 300 K and $P(\lambda, R_g)$ was a probability distribution function at (λ, R_g) . The peptide is known to have two stable conformations: a compact α -helix and an elongated polyproline II helix-like conformation (JiJi et al. 2006; Tiffany and Krimm 1968) corresponding to an R_g of approximately 6.5 and 9 Å, respectively (Fig. 8a, b). Although ALSD_{ele} sampled these two stable conformations at $\lambda = 1$, the peptide was compact at small λ (Fig. 8c). Scaling-down of E_{ele} corresponds to scaling-down of point charge parameters for region A. For more detail, the reader is referred to the original article (Ikebe et al. 2014). Thus, the scaling-down of E_{ele} without the scaling-down of E_{vdw} makes the peptide more hydrophobic and concomitantly compact. In contrast, ALSD_{vdw} sampled elongated conformations for small λ (Fig. 8d). It should be remembered that the conformational ensemble at 300 K is the same as that at $\lambda = 1$. The scaling-down of E_{vdw} weakens the collision energy of atoms in region A, and exposure of the peptide to the solvent induces a decrease of the potential energy. Thus, the scaling-down of E_{vdw} without any scaling-down of E_{ele} makes the peptide elongate with decreasing λ . The authors of the original ALSD article (Ikebe et al. 2014) have suggested that such unbalanced sampling at $\lambda < 1$ provides a worse sampling efficiency. However, ALSD_{both} sampled various conformations of the peptide in a wide range of R_g with decreasing λ (Fig. 8e) because the compaction by scaling-down of E_{ele} and the elongation by that of E_{vdw} offset each other. The ALSD method extends the

Fig. 8 Representative structures showing compact α -helix (a) and elongated polyproline II helix-like conformations of poly-lysine decapeptide (b) obtained from $\text{ALSD}_{\text{both}}$ at a λ of 1. It is known that the peptide adopts these two conformations as the thermodynamically stable ones in the canonical ensemble. The terminal regions colored in red are the N-termini of the peptide. c–e Free energy landscapes on a space composed of λ and the radius of gyration (R_g) of the peptide obtained from ALSD_{ele} (c), ALSD_{vdW} (d), and $\text{ALSD}_{\text{both}}$ (e). See text in section [Efficient operation procedures for ALS](#) for explanation of terms



potential surface by including λ . The force field parameters for E_{ele} and E_{vdW} work in a complementary manner, with the result that the extended potential surface is simple to sample. We strongly recommend that E_{ele} and E_{vdW} should be scaled together for efficient sampling.

Of course, the SED procedure also affects the sampling efficiency. An arbitrary setting of region A does not necessarily realize efficient sampling. For an extreme example, we consider an ALS simulation of a complex system at a temperature that is sufficiently low to freeze the solvent. When the substrate is set as region A, the substrate is not sampled sufficiently because the frozen solvent suppresses the conformational motions of the substrate. Note that ALS presupposes that conformational fluctuations in region B are sufficiently large to allow efficient sampling for region A.

For a practical case of protein–ligand binding, another precaution may be useful: a protein is relatively rigid and the conformational changes are slow. As such, a long simulation time is required to realize binding and dissociation of the ligand during a simulation. It may therefore be useful to set region A not only to the ligand but also to the prospective binding sites on the protein.

Future perspective of enhanced sampling methods

We have described enhanced sampling methods in general, as well as our specific enhanced methods to sample the equilibrium ensemble with all-atom models in an explicit solvent. Despite the limitations of simulation time, these methods could provide the reliable conformational ensembles to

construct the accurate and fine-grained FELs. In this final section, we introduce studies that have gone beyond the fine-grained model and which possibly represent new strategies for further development of enhanced sampling methods.

The MD simulation evolves the system at each small time-step. The size of the time-step is set to safely simulate the fastest motion (i.e., avoid numerical violation). An atom with a light mass usually moves rapidly, and the vibrational motions of a chemical bond are rapid. The gold standard method of the SHAKE algorithm maintains the length of the covalent bonds with hydrogen atoms so that the time-step elongates by approximately 2 fs, which is the time-step elongation that we used in the studies described in the preceding sections of this review. If we could take a larger time-step, however, the MD simulation would concomitantly be generated over a longer time-scale for the same computational output. To achieve this, early researchers simply increased the masses of all atoms (or alternatively only hydrogens) to slow these rapid motions, resulting in an increase in the total mass of the system. An unwanted side-effect of the increase in total mass, however, was the scale down of the whole motion of the system (not only the fast, but also the meaningful dynamics). Subsequent studies led to the exploration of the repartitioning of masses (Feenstra et al. 1999), such that a heavier mass was assigned to hydrogen atoms and smaller masses were assigned to heavy atoms, with the result being that the total mass was kept constant. The long-time MD simulation with the hydrogen-mass repartitioning (HMR) method (Hopkins et al. 2015) reported that a 4-fs time-step provided a numerically safe result and did not influence thermodynamic properties. Furthermore, reducing the solvent viscosity by rescaling the mass of water

molecules was effective (Gee and van Gunsteren 2006; Lin and Tuckerman 2010). Under low-viscosity conditions, the biomolecule moves smoothly. Another method for a larger time-step is the reversible multiple time-scale algorithm (RESPA) (Tuckerman et al. 1992), which enables a larger time-step of about 4 fs for the long-distant interaction terms because the long-distant interactions vary slowly. Combining such longer time-step methods with the enhanced sampling method may realize less computational cost to obtain the conformational ensemble.

The energy and the forces on atoms are calculated in MD simulations. Rapid calculation methods have been developed by approximating long-distance interactions, with the particle mesh Ewald (PME) (Darden et al. 1993) for periodic boundary condition and multipole method (Ding et al. 1992) for spherical solvent boundary (droplet) being frequently used examples. Cutoff methods, where the interactions over a threshold distance are neglected or approximated, are also used. A zero-dipole method (Fukuda et al. 2011) (or zero-multipole method generally) (Fukuda 2013) for the periodic boundary condition has recently been developed as an extension of the cutoff method. The zero-dipole method is implemented to take advantage of massively parallelized computational resource rather than PME because it needs less communication for parallelization. Importantly, the electrostatic accuracy of the zero-dipole method is compatible with that of the PME method.

Computational costs can be lowered by using alternate methods. Most of the computational cost of such studies results from including the interaction terms for the solvent molecules. When solvation effects are treated approximately using the pair-wise function among solute atoms, the computational cost can be smaller. In such treatments, the system does not contain solvent atoms explicitly, rather it exhibits the solvation effect and is thus described as implicit solvent. In the implicit solvent, solute atoms move smoothly because collisions between solute and solvent atoms vanish. There are several implicit solvent models based on the continuum solvent model, namely, the EEF (effective energy function) 1 model (Lazaridis and Karplus 1999), the generalized-born (GB) model (Still et al. 1990), and their variants (Kleinjung and Fraternali 2014). A distance-dependent dielectric treatment had been used for many years as a convenient approximation. However, it is a crude approximation for the all-atom model of biomolecules as the solvent effect is of critical importance in determining the stability of the peptide conformation and FEL (Mitomo et al. 2006; Shell et al. 2008; Zhou and Berne 2002). As such implicit solvent models should be employed carefully. Rigorous methods for the solvation free energy have been developed by the three-dimensional reference-interaction-site-model (3D-RISM) theory (Kovalenko and Hirata 2000a; Kovalenko and Hirata 2000b) and the morphometric approach (Roth et al. 2006), which are

based on statistical physics. Although the more rigorous methods require computational cost, they are especially powerful for distinguishing the native structure outside of the pools, including decoy (close but not native) conformations (Yasuda et al. 2011). The combination of conventional McMD and the RISM theory was attempted to calculate the ensemble of Met-enkephalin (Mitsutake et al. 2000).

Restriction on the translational space of substrate is also useful when the aim is to compute the binding free energy. Without the restriction, the substrate tends to travel through all space of the system volume. When the substrate is significantly far from the receptor, the potential of mean force should be homogeneous. Then, confining the substrate to a spherical or cone-shaped space by the restriction can reduce the computation cost for the homogeneous part. This confinement method has been combined with metadynamics to obtain the binding/unbinding kinetics as well as the binding free energy (Limongelli et al. 2013; Tiwary et al. 2015). It should be noted that this confinement space should be designed so that its center locates at the determinate binding site. If the binding site is known in advance, this method can therefore be selected. Similar to reducing sampling space, there is a method to sample the conformations and FEL along a certain binding pathway: the filling potential method (Fukunishi et al. 2003). Application of the confinement method without any knowledge of the binding site has also been reported with the use of the replica exchange method (Anselmi and Pisabarro 2015), where the ligand resides inside a receptor-shaped closed surface larger than the receptor surface. This procedure can be useful for finding the binding mode and binding affinity at the same time.

Coarse-grained (CG) models for biomolecules (Saunders and Voth 2013; Takada 2012; Tozzini 2005) have been used to investigate the behavior of large biomolecules. In the CG models, the biomolecule is modeled with a particle that represents a sum of several atoms. For example, one particle represents one amino acid residue of protein, and the chain of the particles depicts the protein. The time-step for the CG model can be larger than that for the all-atom model. The energy function for the CG model may involve the solvation effect implicitly. A type of CG model uses a specific structure (such as the experimental structure or native structure) of biomolecule to determine the bottom of the energy funnel with this approach, also known as a Go-like model or native structure-based model (Clementi et al. 2000; Karanicolas and Brooks 2003). Studies with the Go-like model have revealed a protein folding mechanism (Onuchic and Wolynes 2004), and a binding mechanism of biomolecular interaction has been proposed (Levy et al. 2007). Energy functions for the CG model that do not require the specific structure have been developed and used for prediction of protein folding (Gront et al. 2008). Such energy functions are referred to as physics-based or sequence-dependent potentials. The energy functions

of these models are defined in various ways, with physics-based potentials using physical relations specified between atoms (or coarser united atom representations) while statistical potentials infer relations from the examination of a large number of structures gathered in a structural database. Statistics-based energy functions are also referred to as knowledge-based (Sippl 1995). A combination of physics-based models and native structure-based models have been used to investigate protein–protein (Ganguly et al. 2012; Kim and Hummer 2008; Okazaki et al. 2012), protein–DNA (Terakawa et al. 2012; Vuzman et al. 2010), and protein receptor–small compound association/dissociation mechanisms (Negami et al. 2014). The inter-molecular interactions were treated by the physics-based model, while the intra-molecular interactions were treated with native structure-based model. Research into the use of CG models is focused on the kinetic behaviors of binding/unbinding as well as thermodynamics. The tREM and McMD methods have been used to sample the conformational ensemble by the CG models (Li et al. 2012; Naniyas et al. 2006). The improved sampling methods introduced in this review may contribute to even more rapid simulation procedures based on the use of CG models.

The multicanonical algorithm is specific for the flat distribution method. The interested reader may realize that the iteration procedure is crucial to both the refinement of the canonical distribution function in Eq. 1 and equilibration of the conformational ensemble. Although V-McMD requires fewer iterations, it still takes time. However, if no iteration technique is available, it is the generally more useful method. There are two types of non-iteration techniques: on-the-fly and one-time. The former technique, also known as the Wang–Landau method (Shimoyama et al. 2011; Wang and Landau 2001), involves updating the estimation of the canonical probability at every step of the simulation. When the probability function converges, the Wang–Landau simulation is finished. After that, the conventional McMD with the converged probability is performed to obtain an equilibrated conformational ensemble. The one-time technique was proposed by Terada et al. (2003). The canonical probability distribution function in a wide energy range is estimated only once from multiple canonical runs at different temperatures. Such techniques can be useful for small systems. However, when the system is large and complex, the techniques are less efficient, especially for the low-energy range because of shortage of covering conformational space. We use the one-time technique in V-McMD and ALSMD to produce the first guess of the probability distribution function with subsequent iterative improvement. The estimation of correct canonical probability is inextricably associated with an efficient search for the conformational space. The rapid conformational search method (Harada et al. 2015; Klvana et al. 2009; Lüdemann et al. 2000) can then be useful for generating initial widespread conformations although their conformations are non-equilibrated.

Summary

Molecular dynamics simulations using all-atom models in explicit solvent can provide meaningful biomolecular conformations that can be hard/impossible to determine experimentally. However, it is still difficult to correctly sample the conformational ensemble, with the result that errors are introduced into all estimates of the FEL of biomolecular interactions. These difficulties are due to the problem of overcoming numerous time-consuming energy barriers in the entire conformational space. This has led to the development of enhanced sampling method to achieve efficient sampling of the total conformational space. It is these methods which we have reviewed in this article. We also introduced the theory and application of three improved enhanced sampling methods, namely, TTP-McMD, V-McMD, and ALSMD. In the final section, we introduced computational techniques, such as larger time-step and implicit solvent, which can be combined with the enhanced sampling methods. These enhanced sampling methods can be extrapolated from the all-atom model to the CG models for larger biomolecules. Such enhanced sampling methods are important applications to determine the thermodynamic properties of biomolecules by computational methods.

Acknowledgments This research was funded by Ministry of Education, Culture, Sports, Science and Technology (MEXT) Strategic Programs for Innovative Research, Computational Life Science and Application in Drug Discovery and Medical Development (hp120309, hp130003, hp140029). KU is supported by Research Fellowships of JSPS for Young Scientists. JH was supported by a Grant-in-Aid for Scientific Research on Innovative Areas (21113006) received from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) Japan and by the New Energy and Industrial Technology Development Organization (NEDO) Japan. We thank Dr. Damien Hall for a thorough reading of the text prior to submission.

Compliance with ethical standards

Conflict of interest Jinzen Ikebe, Koji Umezawa, and Junichi Higo declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human or animal subjects performed by any of the authors.

References

- Anselmi M, Pisabarro MT (2015) Exploring Multiple Binding Modes Using Confined Replica Exchange Molecular Dynamics. *J Chem Theory Comput* 11:3906–3918. doi:10.1021/acs.jctc.5b00253
- Berg BA, Neuhaus T (1992) Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Phys Rev Lett* 68:9–12. doi:10.1103/PhysRevLett.68.9
- Bernardi RC, Melo MCR, Schulten K (2015) Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim Biophys Acta* 1850:872–877 doi:10.1016/j.bbagen.2014.10.019

- Bosshard HR (2001) Molecular recognition by induced fit: how fit is the concept? *Physiology* 16:171–173
- Brüschweiler S, Konrat R, Tollinger M (2013) Allosteric communication in the KIX domain proceeds through dynamic repacking of the hydrophobic core. *ACS Chem Biol* 8:1600–1610. doi:10.1021/cb4002188
- Chandler D (1987) Introduction to modern statistical mechanics. Oxford University Press, Oxford
- Christen M, van Gunsteren WF (2008) On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: A review. *J Comput Chem* 29:157–166. doi:10.1002/jcc.20725
- Clementi C, Nymeyer H, Onuchic JN (2000) Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? an investigation for small globular proteins. *J Mol Biol* 298:937–953. doi:10.1006/jmbi.2000.3693
- Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092. doi:10.1063/1.464397
- De Guzman RN, Goto NK, Dyson HJ, Wright PE (2006) Structural Basis for Cooperative Transcription Factor Binding to the CBP Coactivator. *J Mol Biol* 355:1005–1013. doi:10.1016/j.jmb.2005.09.059
- Deng Y, Roux B (2009) Computations of standard binding free energies with molecular dynamics simulations. *J Phys Chem B* 113:2234–2246. doi:10.1021/jp807701h
- Ding HQ, Karasawa N, Goddard WA (1992) Atomic level simulations on a million particles: The cell multipole method for Coulomb and London nonbond interactions. *J Chem Phys* 97:4309–4315. doi:10.1063/1.463935
- Doshi U, Hamelberg D (2015) Towards fast, rigorous and efficient conformational sampling of biomolecules: Advances in accelerated molecular dynamics. *Biochim Biophys Acta* 1850:878–888. doi:10.1016/j.bbagen.2014.08.003
- Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradović Z (2002) Intrinsic disorder and protein function. *Biochemistry* 41:6573–6582. doi:10.1021/bi012159+
- Earl DJ, Deem MW (2005) Parallel tempering: Theory, applications, and new perspectives. *Phys Chem Chem Phys* 7:3910–3916. doi:10.1039/B509983H
- Feenstra KA, Hess B, Berendsen HJC (1999) Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J Comput Chem* 20:786–798. doi:10.1002/(SICI)1096-987X(199906)20:8<786::AID-JCC5>3.0.CO;2-B
- Fukuda I (2013) Zero-multipole summation method for efficiently estimating electrostatic interactions in molecular system. *J Chem Phys* 139:174107. doi:10.1063/1.4827055
- Fukuda I, Yonezawa Y, Nakamura H (2011) Molecular dynamics scheme for precise estimation of electrostatic interaction via zero-dipole summation principle. *J Chem Phys* 134:164107. doi:10.1063/1.3582791
- Fukunishi H, Watanabe O, Takada S (2002) On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J Chem Phys* 116:9058–9067. doi:10.1063/1.1472510
- Fukunishi Y, Mikami Y, Nakamura H (2003) The filling potential method: a method for estimating the free energy surface for protein–ligand docking. *J Phys Chem B* 107:13201–13210. doi:10.1021/jp035478e
- Ganguly D, Otieno S, Waddell B, Iconaru L, Kriwacki RW, Chen J (2012) Electrostatically accelerated coupled binding and folding of intrinsically disordered proteins. *J Mol Biol* 422:674–684. doi:10.1016/j.jmb.2012.06.019
- Gee PJ, van Gunsteren WF (2006) Numerical simulation of the effect of solvent viscosity on the motions of a β -peptide heptamer. *Chem A Eur J* 12:72–75. doi:10.1002/chem.200500587
- Götz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC (2012) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1 Generalized born. *J Chem Theory Comput* 8:1542–1555
- Gront D, Latek D, Kurcinski M, Kolinski A (2008) Template-free predictions of three-dimensional protein structures: From first principles to knowledge-based potentials. In: Bujnicki J (ed) Prediction of protein structures, functions, and interactions. John Wiley & Sons, New York, pp 117–141. doi:10.1002/9780470741894.ch5
- Hamelberg D, Mongan J, McCammon JA (2004) Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J Chem Phys* 120:11919–11929. doi:10.1063/1.1755656
- Hansmann UHE, Okamoto Y (1993) Prediction of peptide conformation by multicanonical algorithm: New approach to the multiple-minima problem. *J Comput Chem* 14:1333–1338. doi:10.1002/jcc.540141110
- Hansmann UHE, Okamoto Y, Eisenmenger F (1996) Molecular dynamics Langevin and hybrid Monte Carlo simulations in a multicanonical ensemble. *Chem Phys Lett* 259:321–330. doi:10.1016/0009-2614(96)00761-0
- Harada R, Takano Y, Baba T, Shigeta Y (2015) Simple, yet powerful methodologies for conformational sampling of proteins. *Phys Chem Chem Phys* 17:6155–6173. doi:10.1039/C4CP05262E
- Higo J, Nakamura H (2012) Virtual states introduced for overcoming entropic barriers in conformational space. *Biophysics* 8:139–144. doi:10.2142/biophysics.8.139
- Higo J, Kamiya N, Sugihara T, Yonezawa Y, Nakamura H (2009) Verifying trivial parallelization of multicanonical molecular dynamics for conformational sampling of a polypeptide in explicit water. *Chem Phys Lett* 473:326–329
- Higo J, Nishimura Y, Nakamura H (2011) A free-energy landscape for coupled folding and binding of an intrinsically disordered protein in explicit solvent from detailed all-atom computations. *J Am Chem Soc* 133:10448–10458. doi:10.1021/ja110338e
- Higo J, Ikebe J, Kamiya N, Nakamura H (2012) Enhanced and effective conformational sampling of protein molecular systems for their free energy landscapes. *Biophys Rev* 4:27–44. doi:10.1007/s12551-011-0063-6
- Higo J, Umezawa K, Nakamura H (2013) A virtual-system coupled multicanonical molecular dynamics simulation: Principles and applications to free-energy landscape of protein–protein interaction with an all-atom model in explicit solvent. *J Chem Phys* 138:184106. doi:10.1063/1.4803468
- Higo J, Dasgupta B, Mashimo T, Kasahara K, Fukunishi Y, Nakamura H (2015) Virtual-system-coupled adaptive umbrella sampling to compute free-energy landscape for flexible molecular docking. *J Comput Chem* 36:1489–1501. doi:10.1002/jcc.23948
- Hoh F, Cerdan R, Kaas Q et al. (2004) High-Resolution X-ray Structure of the unexpectedly stable Dimer of the [Lys(-2)-Arg(-1)-des(17–21)]Endothelin-1 peptide. *Biochemistry* 43:15154–15168. doi:10.1021/bi049098a
- Hopkins CW, Le Grand S, Walker RC, Roitberg AE (2015) Long-time-step molecular dynamics through hydrogen mass repartitioning. *J Chem Theory Comput* 11:1864–1874. doi:10.1021/ct5010406
- Ikebe J, Kamiya N, Ito J-I, Shindo H, Higo J (2007) Simulation study on the disordered state of an Alzheimer's β amyloid peptide A β (12–36) in water consisting of random-structural, β -structural, and helical clusters. *Protein Sci* 16:1596–1608. doi:10.1110/ps.062721907
- Ikebe J, Standley DM, Nakamura H, Higo J (2011a) Ab initio simulation of a 57-residue protein in explicit solvent reproduces the native conformation in the lowest free-energy cluster. *Protein Sci* 20:187–196. doi:10.1002/pro.553
- Ikebe J, Umezawa K, Kamiya N et al. (2011b) Theory for trivial trajectory parallelization of multicanonical molecular dynamics and application to a polypeptide in water. *J Comput Chem* 32:1286–1297. doi:10.1002/jcc.21710

- Ikebe J, Sakuraba S, Kono H (2014) Adaptive lambda square dynamics simulation: An efficient conformational sampling method for biomolecules. *J Comput Chem* 35:39–50. doi:10.1002/jcc.23462
- Ikebe J, Sakuraba S, Kono H (2015) Conformational sampling of unmodified and acetylated H3 histone tails on a nucleosome by all-atom model molecular dynamics simulations. *Biophys J* 108:540a–541a
- Ikeda K, Higo J (2003) Free-energy landscape of a chameleon sequence in explicit water and its inherent α/β bifacial property. *Protein Sci* 12:2542–2548. doi:10.1110/ps.03143803
- Itoh SG, Okumura H (2013a) Coulomb replica-exchange method: Handling electrostatic attractive and repulsive forces for biomolecules. *J Comput Chem* 34:622–639. doi:10.1002/jcc.23167
- Itoh SG, Okumura H (2013b) Replica-permutation method with the Suwa-Todo algorithm beyond the replica-exchange method. *J Chem Theory Comput* 9:570–581. doi:10.1021/ct3007919
- Itoh SG, Okumura H, Okamoto Y (2010) Replica-exchange method in van der Waals radius space: Overcoming steric restrictions for biomolecules. *J Chem Phys* 132:134105. doi:10.1063/1.3372767
- James LC, Tawfik DS (2003) Conformational diversity and protein evolution—a 60-year-old hypothesis revisited. *Trends Biochem Sci* 28:361–368. doi:10.1016/S0968-0004(03)00135-X
- Jenuwein T, Allis CD (2001) Translating the histone code. *Science* 293:1074–1080
- Ji Ji RD, Balakrishnan G, Hu Y, Spiro TG (2006) Intermediacy of poly (L-proline) II and β -strand conformations in poly (L-lysine) β -sheet formation probed by temperature-jump/UV resonance Raman spectroscopy. *Biochemistry* 45:34–41
- Kamiya N, Watanabe YS, Ono S, Higo J (2005) AMBER-based hybrid force field for conformational sampling of polypeptides. *Chem Phys Lett* 401:312–317. doi:10.1016/j.cplett.2004.11.070
- Kamiya N, Yonezawa Y, Nakamura H, Higo J (2008) Protein-inhibitor flexible docking by a multicanonical sampling: Native complex structure with the lowest free energy and a free-energy barrier distinguishing the native complex from the others. *Proteins: Structure. Funct Bioinforma* 70:41–53. doi:10.1002/prot.21409
- Karanicolas J, Brooks CL III (2003) Improved G \ddot{o} -like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. *J Mol Biol* 334:309–325. doi:10.1016/j.jmb.2003.09.047
- Kikugawa G, Apostolov R, Kamiya N et al. (2009) Application of MDGRAPE-3, a special purpose board for molecular dynamics simulations, to periodic biomolecular systems. *J Comput Chem* 30:110–118
- Kim YC, Hummer G (2008) Coarse-grained models for simulations of multiprotein complexes: Application to ubiquitin binding. *J Mol Biol* 375:1416–1433. doi:10.1016/j.jmb.2007.11.063
- Kleinjung J, Fraternali F (2014) Design and application of implicit solvent models in biomolecular simulations. *Curr Opin Struct Biol* 25:126–134. doi:10.1016/j.sbi.2014.04.003
- Klvana M, Pavlova M, Koudelakova T et al. (2009) Pathways and mechanisms for product release in the engineered haloalkane dehalogenases explored using classical and random acceleration molecular dynamics simulations. *J Mol Biol* 392:1339–1356. doi:10.1016/j.jmb.2009.06.076
- Kondo HX, Taiji M (2013) Enhanced exchange algorithm without detailed balance condition for replica exchange method. *J Chem Phys* 138:244113. doi:10.1063/1.4811711
- Kong X, Brooks CL III (1996) λ dynamics: A new approach to free energy calculations. *J Chem Phys* 105:2414
- Koshland DE (1960) The active site and enzyme action. *Adv Enzymol Relat Subj Biochem* 22:45–97
- Kovalenko A, Hirata F (2000a) Potentials of mean force of simple ions in ambient aqueous solution. I. Three-dimensional reference interaction site model approach. *J Chem Phys* 112:10391–10402. doi:10.1063/1.481676
- Kovalenko A, Hirata F (2000b) Potentials of mean force of simple ions in ambient aqueous solution. II Solvation structure from the three-dimensional reference interaction site model approach, and comparison with simulations. *J Chem Phys* 112:10403–10417. doi:10.1063/1.481677
- Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem* 13:1011–1021. doi:10.1002/jcc.540130812
- Laio A, Parrinello M (2002) Escaping free-energy minima. *Proc Natl Acad Sci USA* 99:12562–12566. doi:10.1073/pnas.202427399
- Law SM, Gagnon JK, Mapp AK, Brooks CL (2014) Prepaying the entropic cost for allosteric regulation in KIX. *Proc Natl Acad Sci USA* 111:12067–12072. doi:10.1073/pnas.1405831111
- Lazaridis T, Karplus M (1999) Effective energy function for proteins in solution proteins: structure. *Funct Bioinforma* 35:133–152. doi:10.1002/(SICI)1097-0134(19990501)35:2<133::AID-PROT1>3.0.CO;2-N
- Levy Y, Onuchic JN, Wolynes PG (2007) Fly-casting in protein–DNA binding: frustration between protein folding and electrostatics facilitates target recognition. *J Am Chem Soc* 129:738–739. doi:10.1021/ja065531n
- Li W, Terakawa T, Wang W, Takada S (2012) Energy landscape and multiroute folding of topologically complex proteins adenylate kinase and 2ouf-knot. *Proc Natl Acad Sci* 109:17789–17794. doi:10.1073/pnas.1201807109
- Limongelli V, Bonomi M, Parrinello M (2013) Funnel metadynamics as accurate binding free-energy method. *Proc Natl Acad Sci USA* 110:6358–6363. doi:10.1073/pnas.1303186110
- Lin IC, Tuckerman ME (2010) Enhanced conformational sampling of peptides via reduced side-chain and solvent masses. *J Phys Chem B* 114:15935–15940. doi:10.1021/jp109865y
- Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How fast-folding proteins fold. *Science* 334:517–520
- Lüdemann SK, Lounnas V, Wade RC (2000) How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms. *J Mol Biol* 303:797–811. doi:10.1006/jmbi.2000.4154
- Lyman E, Ytreberg FM, Zuckerman DM (2006) Resolution exchange simulation. *Phys Rev Lett* 96:028105. doi:10.1103/PhysRevLett.96.028105
- Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C (2015) ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* 11:3696–3713. doi:10.1021/acs.jctc.5b00255
- Mashimo T, Fukunishi Y, Kamiya N, Takano Y, Fukuda I, Nakamura H (2013) Molecular dynamics simulations accelerated by GPU for biological macromolecules with a non-Ewald scheme for electrostatic interactions. *J Chem Theory Comput* 9:5599–5609
- Mezei M (1987) Adaptive umbrella sampling: Self-consistent determination of the non-Boltzmann bias. *J Comput Phys* 68:237–248
- Mitomo D, Watanabe YS, Kamiya N, Higo J (2006) Explicit and GB/SA solvents: Each with two different force fields in multicanonical conformational sampling of a 25-residue polypeptide. *Chem Phys Lett* 427:399–403. doi:10.1016/j.cplett.2006.06.116
- Mitsutake A, Kinoshita M, Okamoto Y, Hirata F (2000) Multicanonical algorithm combined with the RISM theory for simulating peptides in aqueous solution. *Chem Phys Lett* 329:295–303. doi:10.1016/S0009-2614(00)01018-6
- Mitsutake A, Sugita Y, Okamoto Y (2001) Generalized-ensemble algorithms for molecular simulations of biopolymers. *Pept Sci* 60:96–123. doi:10.1002/1097-0282(2001)60:2<96::AID-BIP1007>3.0.CO;2-F
- Monod J, Wyman J, Changeux J-P (1965) On the nature of allosteric transitions: A plausible model. *J Mol Biol* 12:88–118. doi:10.1016/S0022-2836(65)80285-6

- Nakajima N, Nakamura H, Kidera A (1997) Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides. *J Phys Chem B* 101:817–824. doi:10.1021/jp962142e
- Nanias M, Czaplewski C, Scheraga HA (2006) Replica exchange and multicanonical algorithms with the coarse-grained united-residue (UNRES) force field. *J Chem Theory Comput* 2:513–528. doi:10.1021/ct050253o
- Narumi T, Ohno Y, Okimoto N, Suenaga A, Yanai R, Tajji M (2006) A high-speed special-purpose computer for molecular dynamics simulations: MDGRAPE-3. *NIC Series* 34:29–36
- Negami T, Shimizu K, Terada T (2014) Coarse-grained molecular dynamics simulations of protein–ligand binding. *J Comput Chem* 35:1835–1845. doi:10.1002/jcc.23693
- Nomura M, Uda-Tochio H, Murai K, Mori N, Nishimura Y (2005) The neural repressor NRSF/REST binds the PAH1 domain of the Sin3 corepressor by using its distinct short hydrophobic helix. *J Mol Biol* 354:903–915. doi:10.1016/j.jmb.2005.10.008
- Okazaki K-i, Sato T, Takano M (2012) Temperature-enhanced association of proteins due to electrostatic interaction: A coarse-grained simulation of actin–myosin binding. *J Am Chem Soc* 134:8918–8925. doi:10.1021/ja301447j
- Okumura H, Itoh SG (2013) Transformation of a design peptide between the [small alpha]-helix and [small beta]-hairpin structures using a helix-strand replica-exchange molecular dynamics simulation. *Phys Chem Chem Phys* 15:13852–13861. doi:10.1039/C3CP44443K
- Onuchic JN, Wolynes PG (2004) Theory of protein folding. *Curr Opin Struct Biol* 14:70–75. doi:10.1016/j.sbi.2004.01.009
- Pall S, Abraham MJ, Kutzner C, Hess B, Lindahl E (2014) Tackling exascale software challenges in molecular dynamics simulations with GROMACS. In: Markidis S, Laure E (eds) *Solving software challenges for exascale*, vol. 8759. Springer, Switzerland, pp 3–27
- Radhakrishnan I, Pérez-Alvarado GC, Parker D, Dyson HJ, Montminy MR, Wright PE (1997) Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: A model for activator:coactivator interactions. *Cell* 91:741–752. doi:10.1016/S0092-8674(00)80463-8
- Radhakrishnan I, Pérez-Alvarado GC, Dyson HJ, Wright PE (1998) Conformational preferences in the Ser133-phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of CREB. *FEBS Lett* 430:317–322. doi:10.1016/S0014-5793(98)00680-2
- Roth R, Harano Y, Kinoshita M (2006) Morphometric approach to the solvation free energy of complex molecules. *Phys Rev Lett* 97:078101. doi:10.1103/PhysRevLett.97.078101
- Sakae Y, Okamoto Y (2014) Optimizations of Protein Force Fields. In: Liwo A (ed) *Computational methods to study the structure and dynamics of biomolecules and biomolecular processes*, vol 1, Springer Series in Bio-/Neuroinformatics. Springer, Berlin, pp 195–247. doi:10.1007/978-3-642-28554-7_7
- Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC (2013) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J Chem Theory Comput* 9:3878–3888
- Saunders MG, Voth GA (2013) Coarse-Graining Methods for Computational Biology. *Annu Rev Biophys* 42:73–93. doi:10.1146/annurev-biophys-083012-130348
- Shaw DE, Dror RO, Salmon JK et al. (2009) Millisecond-scale molecular dynamics simulations on Anton. In: *High performance computing networking, storage and analysis*, Proceedings of the Conference on 14–20 Nov. 2009. Institute of Electrical and Electronics Engineers (IEEE), New York, pp 1–11. doi:10.1145/1654059.1654126
- Shaw DE, Maragakis P, Lindorff-Larsen K et al. (2010) Atomic-level characterization of the structural dynamics of proteins. *Science* 330:341
- Shell MS, Ritterson R, Dill KA (2008) A test on peptide stability of AMBER force fields with implicit solvation. *J Phys Chem B* 112:6878–6886. doi:10.1021/jp800282x
- Shimoyama H, Nakamura H, Yonezawa Y (2011) Simple and effective application of the Wang–Landau method for multicanonical molecular dynamics simulation. *J Chem Phys* 134:024109. doi:10.1063/1.3517105
- Sidoli S, Cheng L, Jensen ON (2012) Proteomics in chromatin biology and epigenetics: Elucidation of post-translational modifications of histone proteins by mass spectrometry. *J Proteome* 75:3419–3433
- Sippl MJ (1995) Knowledge-based potentials for proteins. *Curr Opin Struct Biol* 5:229–235. doi:10.1016/0959-440X(95)80081-6
- Still WC, Tempczyk A, Hawley RC, Hendrickson T (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc* 112:6127–6129. doi:10.1021/ja00172a038
- Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* 403:41–45
- Sugase K, Dyson HJ, Wright PE (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447:1021–1025. doi:10.1038/nature05858
- Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 314:141–151. doi:10.1016/S0009-2614(99)01123-9
- Sugita Y, Kitao A, Okamoto Y (2000) Multidimensional replica-exchange method for free-energy calculations. *J Chem Phys* 113:6042–6051. doi:10.1063/1.1308516
- Suwa H, Todo S (2010) Markov Chain Monte Carlo method without detailed balance. *Phys Rev Lett* 105:120603. doi:10.1103/PhysRevLett.105.120603
- Takada S (2012) Coarse-grained molecular simulations of large biomolecules. *Curr Opin Struct Biol* 22:130–137. doi:10.1016/j.sbi.2012.01.010
- Terada T, Matsuo Y, Kidera A (2003) A method for evaluating multicanonical potential function without iterative refinement: Application to conformational sampling of a globular protein in water. *J Chem Phys* 118:4306–4311. doi:10.1063/1.1541613
- Terakawa T, Kenzaki H, Takada S (2012) p53 searches on DNA by rotation-uncoupled sliding at C-terminal tails and restricted hopping of core domains. *J Am Chem Soc* 134:14555–14562. doi:10.1021/ja305369u
- Tiffany ML, Krimm S (1968) New chain conformations of poly (glutamic acid) and polylysine. *Biopolymers* 6:1379–1382
- Tiwary P, Limongelli V, Salvalaglio M, Parrinello M (2015) Kinetics of protein–ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proc Natl Acad Sci USA* 112:E386–E391. doi:10.1073/pnas.1424461112
- Tomba P, Fuxreiter M (2008) Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends Biochem Sci* 33:2–8. doi:10.1016/j.tibs.2007.10.003
- Tozzini V (2005) Coarse-grained models for proteins. *Curr Opin Struct Biol* 15:144–150. doi:10.1016/j.sbi.2005.02.005
- Tuckerman M, Berne BJ, Martyna GJ (1992) Reversible multiple time scale molecular dynamics. *J Chem Phys* 97:1990–2001. doi:10.1063/1.463137
- Umezawa K, Ikebe J, Takano M, Nakamura H, Higo J (2012) Conformational ensembles of an intrinsically disordered protein Pkld with and without a KIX domain in explicit solvent investigated by all-atom multicanonical molecular dynamics. *Biomolecules* 2:104–121. doi:10.3390/biom2010104
- Uversky VN (2013) A decade and a half of protein intrinsic disorder: Biology still waits for physics. *Protein Sci* 22:693–724. doi:10.1002/pro.2261

- Vuzman D, Azia A, Levy Y (2010) Searching DNA via a “Monkey Bar” mechanism: The significance of disordered tails. *J Mol Biol* 396: 674–684. doi:[10.1016/j.jmb.2009.11.056](https://doi.org/10.1016/j.jmb.2009.11.056)
- Wang F, Landau DP (2001) Efficient Multiple-Range Random Walk algorithm to calculate the density of states. *Phys Rev Lett* 86:2050–2053. doi:[10.1103/PhysRevLett.86.2050](https://doi.org/10.1103/PhysRevLett.86.2050)
- Wright PE, Dyson HJ (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol* 16:18–29. doi:[10.1038/nrm3920](https://doi.org/10.1038/nrm3920)
- Yasuda S, Yoshidome T, Harano Y et al. (2011) Free-energy function for discriminating the native fold of a protein from misfolded decoys Proteins: Structure. *Funct Bioinforma* 79:2161–2171. doi:[10.1002/prot.23036](https://doi.org/10.1002/prot.23036)
- Zhou R, Berne BJ (2002) Can a continuum solvent model reproduce the free energy landscape of a β -hairpin folding in water? *Proc Natl Acad Sci USA* 99:12777–12782. doi:[10.1073/pnas.142430099](https://doi.org/10.1073/pnas.142430099)