Review

Exploring configuration space and path space of biomolecules using enhanced sampling techniques — Searching for mechanism and kinetics of biomolecular functions

Hiroshi Fujisaki [1,2]*, Kei Moritsugu [3] and Yasuhiro Matsunaga [4,5]

[1] Department of Physics, Nippon Medical School, 1-7-1 Kyonan-cho, Musashino, Tokyo 180-0023, Japan; fujisaki@nms.ac.jp

[2] AMED-CREST, Japan Agency for Medical Research and Development, 1-1-5 Sendagi, Bunkyo-ku, Tokyo 113-8603, Japan.

[3] Graduate School of Medical Life Science, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan; moritugu@yokohama-cu.ac.jp

[4] RIKEN Center for Computational Science, 7-1-26 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan; ymatsunaga@riken.jp

[5] JST PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan.

* Correspondence: fujisaki@nms.ac.jp; Tel.: +81-422-34-3409

Abstract: To understand functions of biomolecules such as proteins, not only structures but their conformational change and kinetics are important to be characterized but its atomistic details are hard to obtain both experimentally and computationally. We review our recent computational studies using novel enhanced sampling techniques for conformational sampling of biomolecules and calculations of their kinetics. For efficiently characterizing the free energy landscape of a biomolecule, we introduce the multiscale enhanced sampling method, which uses a combined system of atomistic and coarse-grained models. Based on the idea of Hamiltonian replica exchange, we can recover the statistical properties of the atomistic model without any biases. We next introduce the string method as a path search method to calculate the minimum free energy pathways along a multidimensional curve in high dimensional space. Finally we introduce novel methods to calculate kinetics of biomolecules based on the ideas of path sampling: One is the Onsager-Machlup action method, and the other is the weighted ensemble method. Some applications of above methods to biomolecular systems are also discussed and illustrated.

Keywords: Molecular dynamics simulation; Rare event; String method; Multiscale enhanced sampling; Weighted ensemble; Multidrug transporter; Onsager-Machlup action

1. Introduction

A protein (except disordered proteins) usually has a definite 3-dimensional structure (tertiary structure) formed by a sequence of amino-acids (primary structure), and since there is a structure-function relationship, it is extremely important to clarify and characterize the 3D protein structure in atomistic detail [1]. Experimentally x-ray diffraction crystallography, nuclear magnetic resonance (NMR), and recently cryo-electron microscopy have been usually employed to determine the protein structures and nowadays (as of August 2018) 143840 structures have been resolved and stored in protein data bank (PDB). Based on these structures, many chemical insights have been obtained and which are further utilized for the understanding of the biological function of a protein [1].

Even though the 3D structure is fully resolved with Angstrom lengthscale, however, there are missing information, protein structural dynamics, which is a main focus of this review article. Of course, recently more and more advanced experimental techniques, including time-resolved X-ray diffraction crystallography [2], time-resolved IR [3] and UV-Raman [4] spectroscopy, have been developed, but their applications to protein systems are still limited. On the other hand, molecular dynamics (MD) simulation method [5] among many computational methods has been also developed over the decades, and with the advance of computer power and numerical algorithms, it has become an essential tool for computational chemists and experimentalists.

Now (as of 2018) it is a routine to simulate a protein in explicit solvent or membrane with $\sim$ 100000 atoms for $\sim$ 1 micro seconds using MD simulations (if we can use the special-purpose computer Anton [6], it becomes $\sim$ 100 times faster), and the fluctuations of a protein around a naive structure might be fully characterized. However, there arise a difficulty due to *rare events* when we seek more biological consequences of protein dynamics. Rare event is a technical term [7,8] and researchers use it in two kinds of meanings. One is less likely phenomena such as huge earthquake, terror attack, stock market crash, which are categorized as very unlikely events. Corresponding to this definition in protein systems are metastable states which are less populated than a native state. The other is the rare transitions between such unlikely events. This corresponds to conformational change in protein systems because it crosses the (free) energy barrier A when conformational change occurs and is basically characterized by the Arrhenius law

$$k \sim \exp(-A/k_B T), \tag{1}$$

where k is the transition rate and $k_B T$ is the temperature T multiplied by the Boltzmann constant k_B. Unfortunately, sampling these rare events in protein systems are basically not feasible using a conventional MD simulations. This is because MD simulations tend to sample only a single basin around the initially starting structure and not the other basins. Of course, if we can wait a long time, other basins can be sampled but with a statistically insignificant amount. Metastable states and conformational change of proteins are, however, the most essential information of protein functions, and we need to overcome this problem.

There are two strategies: One is to use coarse-grained (CG) models [9–11]. Because such CG model should be simpler than the atomistic models with respect to computation of forces and energies, it is easier to calculate the dynamics and to sample the configuration space with less amount of computational resources. The drawbacks using CG models are that constructing CG models can be sometimes cumbersome and more importantly some detailed information is inevitably lost. In some cases, it is hard to extract time information such as kinetics from CG models.

The other is to use enhanced sampling techniques for atomistic models as described in this review. It is impossible to review all the enhanced sampling methods (see [12] and the references therein), but there are basically two categories: One is to modify the system parameters, and the other is to introduce artificial extended systems. The first category includes increasing the system temperature (e.g., temperature accelerated MD [13]) or boosting the potential bottoms (e.g., accelerated MD [14]). Since the protein dynamics is highly anisotropic [15], it is reasonable to utilize such "functional" directions [16] to modify the potential function (e.g. conformational flooding [17]). More general approaches are to modify the potential function so that the system feels (nearly) no force from such a modified potential, that is, the modified potential should be flat and guaranteed to easily sample the configuration space. The multicanonical method [18], the metadynamics method [19], and the adaptive biased force method [20] are such methods which have been successfully used for many molecular systems. The second category includes extended ensemble methods such as replica exchange methods and their variations. In this type of method, we prepare many replicas with different parameters such as temperatures or force constants, and exchange the parameters using Metropolis type criterion. (In terms of using replicas with different initial conditions, the PaCS MD method recently devised by Harada and Kitao [21] is also considered as a variant of this category.) The temperature replica exchange method was first devised by Hukushima and Nemoto [22], and introduced to the field of molecular dynamics by Hansmann [23], Sugita, and Okamoto [24], and has been routinely used nowadays. Another important variant was introduced independently by Sugita, Kitao, and Okamoto [25], and Fukunishi, Watanabe, and Takada [26], which are called the multidimensional replica exchange method, and the Hamiltonian replica exchange method, respectively. In these methods, we exchange some parameters in the potential function such as force constants. These methods were further extended to multiscale systems using coupling between fine and coarse-grained DoF by Moritsugu, Terada, and Kidera [27–35] as described below. (This method is also considered as an extension of resolution replica exchange [36].)

Even if we could use the most sophisticated enhanced sampling techniques augmented by parallel computation, however, sampling the whole free energy landscape is impossible for large molecular systems [36]. The choice of collective variables (CV) or order parameters can be another issue: if we cannot choose appropriate CVs, the convergence of the free energy calculation would be terribly slow. This is a hard problem still not solved, but to remedy these difficulties, the string method, especially its extension to finite temperatures situations [37–39] would be promising. Assuming that we know two metastable states and we are only concerned with the pathways connecting these metastable states, the string method is a powerful method to sample the pathways. We explain the basic principles of the method and several applications to biomolecular systems below.

Though many enhanced sampling techniques have been developed over the decades as mentioned above and the mechanisms of reaction or conformational change for biomolecules can be clarified, there is still something missing: kinetics or dynamics of biomolecules. If the local equilibrium and several other assumptions hold, a reaction rate (or transition rate) can be estimated from the transition state theory (TST) or Kramers' type formulas [8]. The resulting rate formulas basically look like Eq. (1), where A is the free energy barrier between a reactant and a transition state (more precisely we need to calculate a prefactor which depends on the shape of the potential energy surface and friction coefficient [8]). Hence if we can accurately calculate A then there seems to be no need to calculate the kinetics. Unfortunately this is wrong. Except the practical difficulties to calculate A, the free energy landscape depends on the choice of CVs, and so is the reaction rate. As recently shown by Nakamura [40], if we do not carefully choose CVs, the free energy landscape as a function of CVs only has vague meaning.

As such, many researchers have been pursuing "direct" approaches to calculate the kinetics without the TST or Kramers' type formulas. One such example is the Markov State Model (MSM) [41], and because of its simplicity for understanding and implementation, there have been many applications of this method. In MSM, we prepare several initial states, which are assumed to be located between a reactant and a product state, run a short-time MD simulations, and collect the resulting huge amount of trajectory data. Using some clustering algorithms or taking several dividing surfaces in some order parameter space for such data, we define so-called "micro"states in data space. We then count the numbers of transitions between "micro"states, and construct a transition matrix. (In the conventional MSM, we need to introduce a lag time, whereas in the milestoning [42], since the trajectory in order parameter space is assumed to be continuous, there is no need to introduce a lag time.) Manipulating thus-obtained transition matrix, we can calculate the population in each state or more importantly first passage time between such states. From the first passage time distribution, we can calculate the transition rates between microstates, which can be compared to experiment.

Though MSM is a well-established and simple method, there are several assumptions which might hamper the justification for some applications. As such, some researchers have been developing different methods for kinetics, and path sampling methods are a general and sophisticated approach for this purpose [5,7,8]. In the path sampling approaches, a trajectory or "path" has a weight or probability as a whole, and based on such a weight we can devise a Monte

Carlo move to sample huge path space. A good thing about this approach is that if we know a reactant and product we can connect these states using the path sampling techniques without any biases and in principle we can calculate any dynamical quantities (as well as equilibrium properties) with thus-obtained path ensembles. The Onsager-Machlup (OM) action method is such a path sampling method for overdamped Langevin dynamics, and we explain the basics of the OM method below [35,43]. On the other hand, the conventional path sampling methods use molecular dynamics simulations, and if the process is very slow, it is not efficient to sample path space with these methods. Hence some modified types of path sampling techniques have been developed in the literature [8], and the weighted ensemble (WE) method is one such method [44]. We will introduce this method and discuss some applications to molecular systems below.

2. Multiscale enhanced sampling (MSES)

2.1. Overview of MSES

MSES is an enhanced sampling method for complex molecular systems, adopting an idea of multiscale simulations [27]. Coarse-grained (CG) models are used for MSES because they have been successfully applied for extracting functionally relevant motions of biomolecules [9,10]. Moritsugu and coworkers have developed various extensions of MSES [29,31] and applied them to many protein systems of biological importance [28,30,32–34] as described below.

In MSES, both a target physical system, e.g., an atomistic protein molecule in explicit solvent (we call it MM), and the corresponding CG model are coupled. See Fig. 1 (a). The potential energy of the multiscale system V is

$$V(\mathbf{r}_{\mathrm{MM}}, \mathbf{r}_{\mathrm{CG}}, k_{\mathrm{MMCG}}) = V_{\mathrm{MM}}(\mathbf{r}_{\mathrm{MM}}) + V_{\mathrm{CG}}(\mathbf{r}_{\mathrm{CG}}) + k_{\mathrm{MMCG}} V_{\mathrm{MMCG}}(\mathbf{r}_{\mathrm{MM}}, \mathbf{r}_{\mathrm{CG}}) \tag{2}$$

where V_{MM} and V_{CG} are the potential energies for MM and CG. The number of degrees of freedom (DoF) in the CG systems is M, and it is therefore much smaller than that of the MM system N. The coupling term between the MM and CG systems V_{MMCG} is described by a harmonic constraint,

$$V_{\mathrm{MMCG}} = [\chi_{\mathrm{MM}}(\mathbf{r}_{\mathrm{MM}}) - \chi_{\mathrm{CG}}(\mathbf{r}_{\mathrm{CG}})]^2 \tag{3}$$

with the associated force constant k_{MMCG}, where K coomponent vector χ_{CG} is arbitrarily defined by use of the CG coordinates and χ_{MM} is the projection of $\mathbf{r}_{\mathrm{MM}}$ onto the same K-dimensional space.

In order to obtain the structural ensemble of the intrinsic V_{MM}, the bias through V_{MMCG} needs be eliminated. For this purpose, the Hamiltonian replica exchange [26] is carried out, in which the replicated systems having various k_{MMCG} values from zero to a large value exchanges k_{MMCG} between the neighboring replicas. The exchange probability between replica m and replica n with different k^m_{MMCG} and k^n_{MMCG}, derived so as to satisfy the detailed balance condition, is $p_{mn} = \min(1, \exp(\Delta_{mn}))$ with

$$\Delta_{mn} = \beta\,(k^m_{\mathrm{MMCG}} - k^n_{\mathrm{MMCG}})\,[V_{\mathrm{MMCG}}\,(\mathbf{r}^m_{\mathrm{MM}}, \mathbf{r}^m_{\mathrm{CG}}) - V_{\mathrm{MMCG}}\,(\mathbf{r}^n_{\mathrm{MM}}, \mathbf{r}^n_{\mathrm{CG}})]\,, \tag{4}$$

where $\beta = 1/k_B T$.

It is noted that the exchange probability is in proportion to the squared difference between χ_{MM} and χ_{CG} which are described by the K-dimensional space of the CG coordinates. Because of $K \sim M \ll N$, the smallness of Δ_{mn} or a high exchange probability p_{mn} is assured irrespective of the number of the MM DoF N, leading to much higher scalability as compared with the conventional methods such as temperature replica exchange, where p_{mn} is determined by the difference in the potential energy of MM (scaling up to N^2).

We can determine V_{CG} arbitrarily from prior knowledge or experimental data, depending on which subspace is targeted for enhanced sampling. More importantly, since $K \gg 1$, MSES allows a "predictive" structural sampling in that a distribution is *roughly* defined by CG and then refined through the MM force field. This kind of flexibility is advantageous over other methods using only a few predefined CVs.

2.2. MSES extension using adiabatic separation

In applying MSES to large protein systems including a number of explicit solvents, it is often the case that the force on the CG system from the MM system $-\partial V_{\rm MMCG}/\partial \mathbf{r}_{\rm CG}$ overwhelms the CG intrinsic force $-\partial V_{\rm CG}/\partial \mathbf{r}_{\rm CG}$, leading to the confinement in a stable basin where MM is strongly trapped. To overcome this problem, we have recently developed an extension by use of the approximation of adiabatic separation and the high CG temperature limit. Here we present a brief summary and see [31] for detail.

$\mathbf{r}_{\rm MM}$, $\mathbf{r}_{\rm CG}$ and $k_{\rm MMCG}$ are now considered as the independent variables of the joint distribution, $\rho\left(\mathbf{r}_{MM}, \mathbf{r}_{CG}, k_{\rm MMCG}\right) \propto \exp\left(-\beta V\right)$, with V defined in Eq. 2: The unbiased MM ensemble is then obtained by extrapolating this ensemble to that with $k_{\rm MMCG} = 0$. The joint distribution can be calculated by use of Gibbs sampling, i.e., by sampling $\mathbf{r}_{\rm MM}$, $\mathbf{r}_{\rm CG}$ and $k_{\rm MMCG}$ separately in an iterative manner with their conditional probabilities [45]. Suppose the CG mass being much larger than MM so that CG moves much slower than MM, while the CG temperature, $1/\beta'$, is much higher than that of MM; i.e., $\beta' \ll \beta$. Under this approximation, named "adiabatic separation" [13,46,47], the conditional probabilities for $\mathbf{r}_{\rm MM}$ and $\mathbf{r}_{\rm CG}$ are written by

$$\rho\left(\mathbf{r}_{MM}\left|\mathbf{r}_{CG}, k_{\rm MMCG}\right.\right) \propto \exp\left(-\beta\left[V_{\rm MM}\left(\mathbf{r}_{\rm MM}\right) + k_{\rm MMCG} V_{\rm MMCG}\left(\mathbf{r}_{\rm MM}, \mathbf{r}_{\rm CG}\right)\right]\right), \tag{5}$$

$$\rho\left(\mathbf{r}_{CG}\left|k_{\rm MMCG}\right.\right) \propto \exp\left[-\beta' V_{\rm CG}\left(\mathbf{r}_{\rm CG}\right)\right] Z(\mathbf{r}_{\rm CG}, k_{\rm MMCG})^{\beta'/\beta}, \tag{6}$$

where the partition function,

$$Z\left(\mathbf{r}_{\rm CG}, k_{\rm MMCG}\right) \equiv \int dr_{MM} \exp\left(-\beta\left[V_{\rm MM}\left(\mathbf{r}_{\rm MM}\right) + k_{\rm MMCG} V_{\rm MMCG}\left(\mathbf{r}_{\rm MM}, \mathbf{r}_{\rm CG}\right)\right]\right), \tag{7}$$

appears as the potential of mean force because of the slowness of CG relative to MM. It is straightforward that, by further taking the high CG temperature limit, or $\beta'/\beta \rightarrow 0$, the CG conditional probability is simplified as

$$\rho\left(\mathbf{r}_{CG}\left|k_{\rm MMCG}\right.\right) \propto \exp\left[-\beta' V_{\rm CG}\left(\mathbf{r}_{\rm CG}\right)\right]. \tag{8}$$

These derivations show that while the sampling of $\mathbf{r}_{\rm MM}$ is performed by the MD simulation under the $V_{\rm MMCG}$ constraint, the CG DoF $\mathbf{r}_{\rm CG}$ is allowed to move freely, namely without the counterforce from MM, which results in the largest driving force for MM.

Finally, the sampling of $k_{\rm MMCG}$ is carried out by use of Markov chain Monte Carlo (MCMC) for discretized values of $k_{\rm MMCG}$. Here, we set the replicated systems of the MSES simulation to consist of many MMs which are coupled with a *single* copy of CG, i.e., $\mathbf{r}^{m}_{\rm CG} = \mathbf{r}^{n}_{\rm CG}$, and then the exchange probability for the extended MSES turns out to be the same as in the original MSES (see [31] for detail).

In summary, the simulation process consists of the following iterattion, (1) the MD simulations of the replicated MM models and one CG model, and (2) the MCMC simulations in terms of $k_{\rm MMCG}$. The temperature and mass of the CG model must then be set to satisfy the conditions of adiabatic separation and $\beta'/\beta \rightarrow 0$. To do this, e.g., the kinetic energies of the MM models need to be examined whether the energy flow from CG to MM is negligible [30,32].

2.3. MSES applications to biomolecular systems

The applications to a broad range of complex protein systems of really biological relevance have been so far performed, demonstrating the potential of MSES. Here, these were briefly reviewed by focusing on the CG model used, as the selection of V_{CG} might be critical for the success of the enhanced samplings, and on the number of replicas to show the scalability of MSES.

The order-disorder transition of an intrinsically disordered protein, sortase (a transpeptidase in Gram positive bacteria), was investigated by MSES [28]. The large-scale structural sampling of the disordered region was achieved by use of 20 replicas, indicating the usefulness of the flexibility for the recognition of the substrate peptide. Both the substrate-bound and -unbound structures were used to construct the CG model which drives the order-disorder transition of the disordered region.

Protein-protein and protein-ligand interactions are the fundamental components in the interaction networks describing cellular processes such as signal transduction. The formation of barnase-barstar complex was simulated on atomic detail by use of MSES, demonstrating a funnel-like energy landscape of the strong binder (with high affinity) which is formed through the specific side-chain interactions and the desolvation (Fig. 1b) [30]. In contrast, the complex of di-ubiquitin (di-UB) and UB-binding domain (UBD) with relatively low affinity was found to have a subtle stability with an ensemble of highly dynamic structures, allowing dynamic recognition consisting of various binding modes [34]. In the applications, 12 and 16 replicas were used for barnase-barstar and di-UB/UBD complexes, respectively. To simulate the interaction process by the CG model, the Lennard-Jones potential was used for the protein-protein interactions, while the elastic network model [48] was applied for the rigidity of the intra-protein interactions.

The protein-ligand recognitions, especially in the case where the proteins undergo large structural changes between the open and closed forms, were also studied by use of MSES. In the CG model, the flexibility of the proteins was simulated by use of the double well model [49] which embeds the two open/closed structures, and the protein-ligand interactions, by the same Lennard-Jones potential as in the studies of protein-protein complexes. The full structural sampling of the glutamine binding to the glutamine binding protein, using 12 replicas for MSES, revealed the tight coupling between the protein structural change and the ligand interaction [32]. The derived energy landscape led to the determination of definite structural states, clarifying the dominant ligand interaction pathways in atomic detail. A pharmacological application to an important drug discovery target, glucokinase, was also carried out, demonstrating the drastic change of the energy landcapes depending on the glucose concentration [33]. This thermodynamics calculation, in combination with the weighted ensemble simulations (see Sec. 4.2), allowed the kinetics calculation of the millisecond-timescale structural transition that was found to be the same order as the experimental catalytic rate of glucokinase.

Recently, Hansmann and coworkers developed a new MSES method by adopting (1) a novel CG model using multiple Go-like potentials which can be switched like λ dynamics and (2) their replica-exchange-with-tunneling method for efficient MCMC [50,51]. This method was extensively applied to survey the structural dynamics of various amyloid fibrils using 16-32 replicas, such as the conversion of Aβ40 between parallel- and antiparallel-sheets [52], the formation and interconversion between fibril-like and barrel-like assemblies [53], and the transition of Aβ42 between the in-register and the out-of-register fibrils, and the barrel-shaped oligomers [54]. They also studied the fold-switch of the RfaH C-terminal domain between a double helix bundle and a β-barrel form [55]. Chen and coworkers also used the MSES method by adopting topology-based CG model, and simulated the folding/unfolding of kinase inducible domain of CREB using 16 replicas [56].

The general scheme of MSES will be applicable to broad classes of situations. For example, we adopted MSES to a path sampling method based on the Onsager-Machlup action, allowing a path ensemble to be efficiently sampled [35] (see also Sec. 4.1).

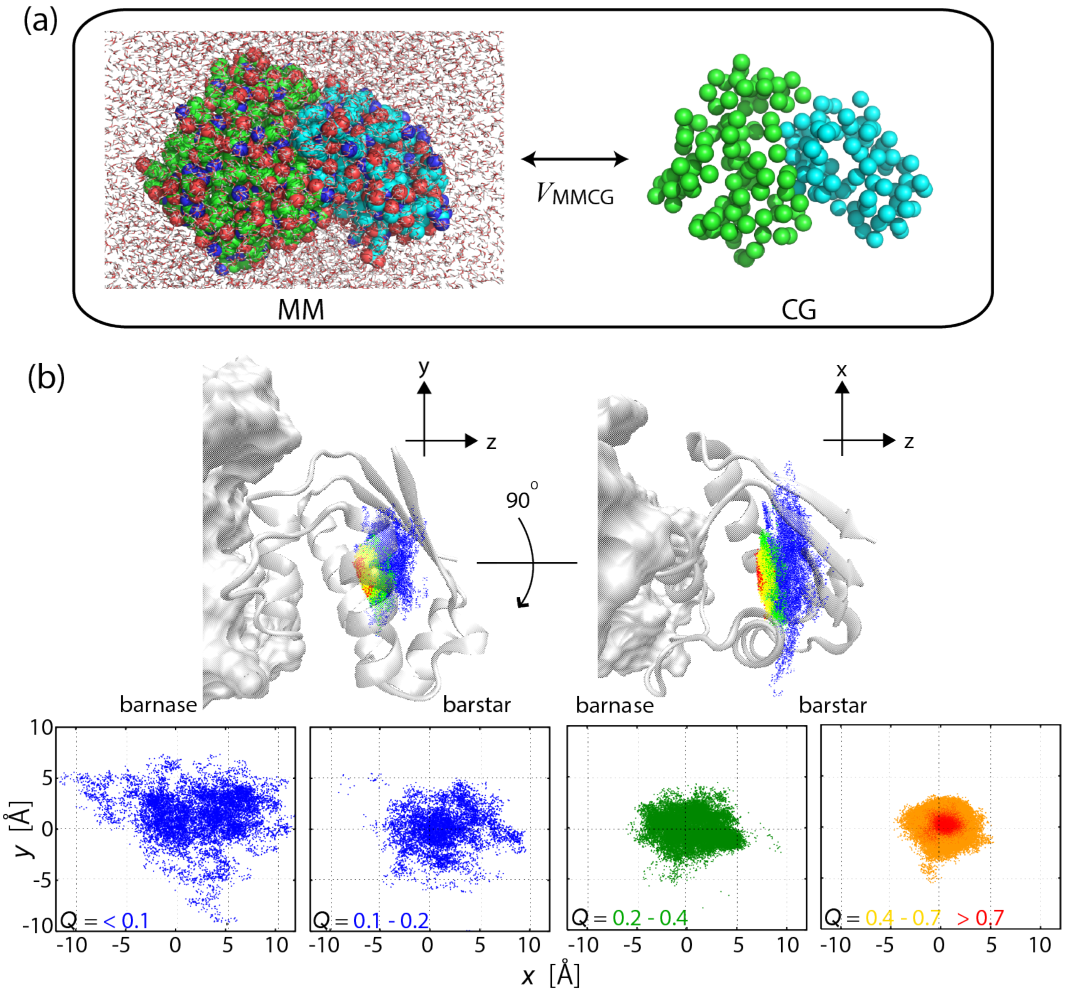

Figure 1. (**a**) Scheme of the MSES method. The structural sampling of MM is driven by CG model through the coupling V_{MMCG}. Hamiltonial replica exchange [26] is adopted to elimimate the bias via V_{MMCG} and then to obtain the unbiased MM structural ensemble. (**b**) Funnel landscape of the protein-protein interaction for barnase-barstar complex, seen in a narrowing of the configurational space with increasing the fraction of native inter-molecular contacts formed (Q). For details, see [30].

3. String method

3.1. Overview of string method

A number of different structures are frequently observed for a single protein in experiments, depending on their crystal or solution conditions (e.g., ligand-free, ligand-bound conditions). This kind of structural polymorphism generally implies an occurrence of structural changes in the cellular environment, which is often related to important biomolecular events (e.g., a transition from in-active to active state), and thus the mechanism of such a structural transition between observed structures attracts many researchers' interests. The string method [37–39], described in this section, is a powerful approach to find a reasonable conformational pathway connecting two known structures. The method efficiently searches the most probable pathway, and enables us to characterize the mechanism of the conformational change.

Suppose that we investigate conformational pathways of a transition from reactant A to product B. First, let us define a set of CVs $\mathbf{z}(\mathbf{r}) \in \mathbb{R}^N$ where $\mathbf{r} \in \mathbb{R}^{3n}$ is the configuration of the system and n is the number of atoms. Actually these are projections of $\mathbf{r}$ to a low-dimensional, coarse-grained space, and N is usually much smaller than $3n$. Typically, subsets of the Cartesian coordinates of protein atoms, or dihedral angles of backbone, or distance between specific atoms are chosen as CVs. The free energy profile or effective potential energy that $\mathbf{z}(\mathbf{r})$ "feels" at $\mathbf{z}^*$ in the CV space is given by,

$$F(\mathbf{z}^*) = -k_B T \ln Z^{-1} \int \delta\left(\mathbf{z}(\mathbf{r}) - \mathbf{z}^*\right) e^{-\beta U(\mathbf{r})} d\mathbf{r}, \tag{9}$$

where $Z = \int e^{-\beta U(\mathbf{r})} d\mathbf{r}$ is a partition function, and $U(\mathbf{r})$ is the potential energy of the system, and $\beta = 1/k_B T$.

In the string method, we assume that most of reactive trajectories which undergo conformational transitions from A to B, once projected in the CV space, go through a single thin tube (called the *reactive tube*). If the free energy barrier during the transition is much higher than $k_B T$, it was mathematically shown that the center of the reaction tube is well approximated by the minimum free energy pathway (MFEP) [37]. The geometry of the MFEP is represented by

$$\left(\mathrm{M}(\mathbf{z}^*)\nabla F(\mathbf{z}^*)\right)^{\perp} = \mathbf{0}, \tag{10}$$

where the superscript $\perp$ denotes the orthogonal component to the curve., $\nabla F(\mathbf{z}^*(s))$ is a gradient of free energy, which is proportional to the mean force acting at $\mathbf{z}^*$, and $\mathrm{M}(\mathbf{z})$ is a metric tensor which accounts for the curvilinear nature of the CVs, given by the following conditional expectation [37],

$$M_{ij}(\mathbf{z}^*) = \left\langle \sum_{k=1}^{3n} \frac{1}{m_k} \frac{\partial z_i(\mathbf{r})}{\partial r_k} \frac{\partial z_j(\mathbf{r})}{\partial r_k} \right\rangle_{\mathbf{z}(\mathbf{r})=\mathbf{z}^*}. \tag{11}$$

The string method is an algorithm to efficiently search the MFEP using a set of replicated simulation systems instead of performing long brute-force simulations.

The MFEP, obtained by the string method, is able to capture the mechanism of the conformational transition because it allows us to determine the committor function [8,36,37]. The committor function, which is known to be *the best* reaction coordinate, is the probability that a trajectory initiated at an arbitrary point will reach first the product B state without going back to the reactant A state. This function allows us to derive various quantities of the transition, including the probability density of reactive trajectories, their probability current, and the rate of the reaction [57]. With properly chosen CVs, the MFEP is expected to be orthogonal to isocommittor surfaces [57]. Since the isocommittor surface of $\frac{1}{2}$ defines the transition state, the MFEP allows us to identify such a state. It is noted, however, that the accuracy of the MFEP in approximating the committor function crucially depends on the choice of CVs. This point will be discussed in the next subsection.

Currently, there are three major algorithms in the string method: (i) string method with mean forces [37], (ii) on-the-fly string method [38], and (iii) string method with swarms-of-trajectories [39]. In all of these methods, a pathway is represented by m discretized CV values (called *images*) connecting the A and B states (Fig. 2a). In the mean force method, a short MD simulation samples conformations around each image with restraint potentials and computes a mean force and an average metric tensor. In the swarms-of-trajectories, a set of restraint-free simulations are initiated around each image and the average drift is computed. Then, each image is evolved using the calculated mean force and average metric tensor, (i.e, $\mathrm{M}(\mathbf{r}^*)\nabla F(\mathbf{r}^*)$ in Eq. 10) or the mean drift. After the evolution of the images, a piecewise linear curve is interpolated through the images and new images are distributed along this curve in an equidistant manner. This procedure is iterated until the pathway converges. The convergences of the pathway implies that the orthogonal component of $\mathrm{M}(\mathbf{r}^*)\nabla F(\mathbf{r}^*)$ becomes zero for all images, thus the converged pathway is shown to be the MFEP.

The string method is available in several MD software packages. The swarms-of-trajectories string method is available in NAMD [58] and the SANDER and PMEMD modules of AMBER. GENESIS supports the string method with mean-forces [59,60]. For a quantum string method [61] using the idea of centroid, PIMD code can be used [62].

3.2. Impact of the CV choice on the accuracy of pathways

The choice of CVs is important because it determines not only the convergence rate of the string method calculation but also the accuracy of the MFEP in terms of the committor function. Then, are there any principles for choosing better CVs? For large-scale conformational changes of proteins, such as open-to-close motions of multi-domain proteins, several choices of CVs have been proposed and demonstrated with a number of CV-based enhanced sampling methods. For example, Abrams and Vanden-Eijnden sampled conformational changes of the GroEL subunit and HIV-1 gp120 by using the temperature-accelerated molecular dynamics (TAMD) [63]. As CVs, they chose Cartesian coordinates of centers of contiguous subdomains, composed of 9 subdomains for GroEL and 14 subdomains for gp120. Vashisth and Brooks applied a potential energy bias in the direction of displacement calculated from the crystal structures and facilitated functional motions in their TAMD simulations [64]. In the context of the string method, Maragliano *et al.* showed, for accurate description of conformational transition of alanine-dipeptide, that four dihedral angles (rather than two) are required by evaluating committor functions [37]. Pan *et al.* showed that, through the comparison with brute-force simulations by Anton, root-mean-square deviations (RMSDs) of the flexible region are better CVs for capturing an accurate transition pathway of EGFR kinase rather than RMSDs of the entire molecule [65]. Recently, Matsunaga *et al.* examined how many CVs are required to capture the correct transition state during the open-to-close motion of Adenylate kinase's CG model in the string method [66]. They tested various numbers of large amplitude principal components as CVs. Using the Bayesian statistics measure, they showed that incorporation of local coordinates into CVs, which is possible in higher dimensional CV spaces, is important for capturing a reliable transition state. All taken together, in the string method, it is better to choose flexible regions and multi-dimensional coordinates as putative CVs.

3.3. String method applications to biomolecular systems

Thus far, the string method has been successfully applied for finding the conformational transition pathways of various biomolecular systems, such as c-Src kinase [67], myosin VI [68], ion channels [69,70], 2-microglobulin [71], V1-ATPase [72], calcium pump [73], and membrane transporters [74,75]. Among them, two application cases, Adenylate kinase [76] and the multidrug transporter AcrB [77], conducted by the authors will be illustrated below.

Adenylate kinase is the best-studied biomolecule exhibiting a large conformational transition [78]. It is an enzyme which catalyzes the reversible phosphoryl transfer reaction: ATP + AMP ↔ 2ADP. Its crystal structures suggests that, upon ligand binding, this enzyme undergoes a transition from the inactive open form to the catalytically competent closed structure. This transition is mediated by large-scale closure motions of three rigid-body domains (LID, AMPbd, and CORE domains). This functional open-to-close motion was investigated by the string method [76]. The MFEPs of the open-to-close transition were calculated by the string method using 20 largest-amplitude principal coordinates as CVs, under ligand-free and ligand-bound conditions. By comparing the two pathways, it was found that the LID domain was able to partially close without the ligand, while the closure of the AMPbd domain required the substrate binding. The transition state of the substrate bound form was identified as a highly specific binding state of the substrate to the AMPbd domain, and was validated by a committor test (see Sec. 3.1) with restraint-free MD simulations. These findings suggest that the interplay of the two different types of domain motion is an essential feature in the conformational transition of the enzyme.

The multidrug transporter AcrB transports a broad range of drugs out of the bacterial cell by means of the proton-motive force [79]. The drug transportation by transporters is one of the main causes of multidrug resistance in bacteria. Thus, understanding the mechanism of the drug transportation is important for the treatment of bacterial infections. The asymmetric crystal structure of trimeric AcrB suggests that a large conformational change in AcrB (called the functionally rotation) is coupled to the drug transport. Despite various supportive data from biochemical and simulation studies for this functional rotating mechanism, the link between the functional rotation and proton-motive force remained elusive. By calculating the MFEPs of the functional rotation for the complete AcrB trimer, the authors described the molecular basis behind the coupling between the functional rotation and the proton translocation [77]. Free energy calculations along the pathways showed that protonation of Asp408 in the transmembrane domain of the drug-bound protomer drives the functional rotation. The conformational pathway identifies vertical shear motions among several transmembrane helices, which regulate alternate access of water in the transmembrane as well as peristaltic motions that pump drugs in the periplasmic domain (Figs. 2(b) and (c)).

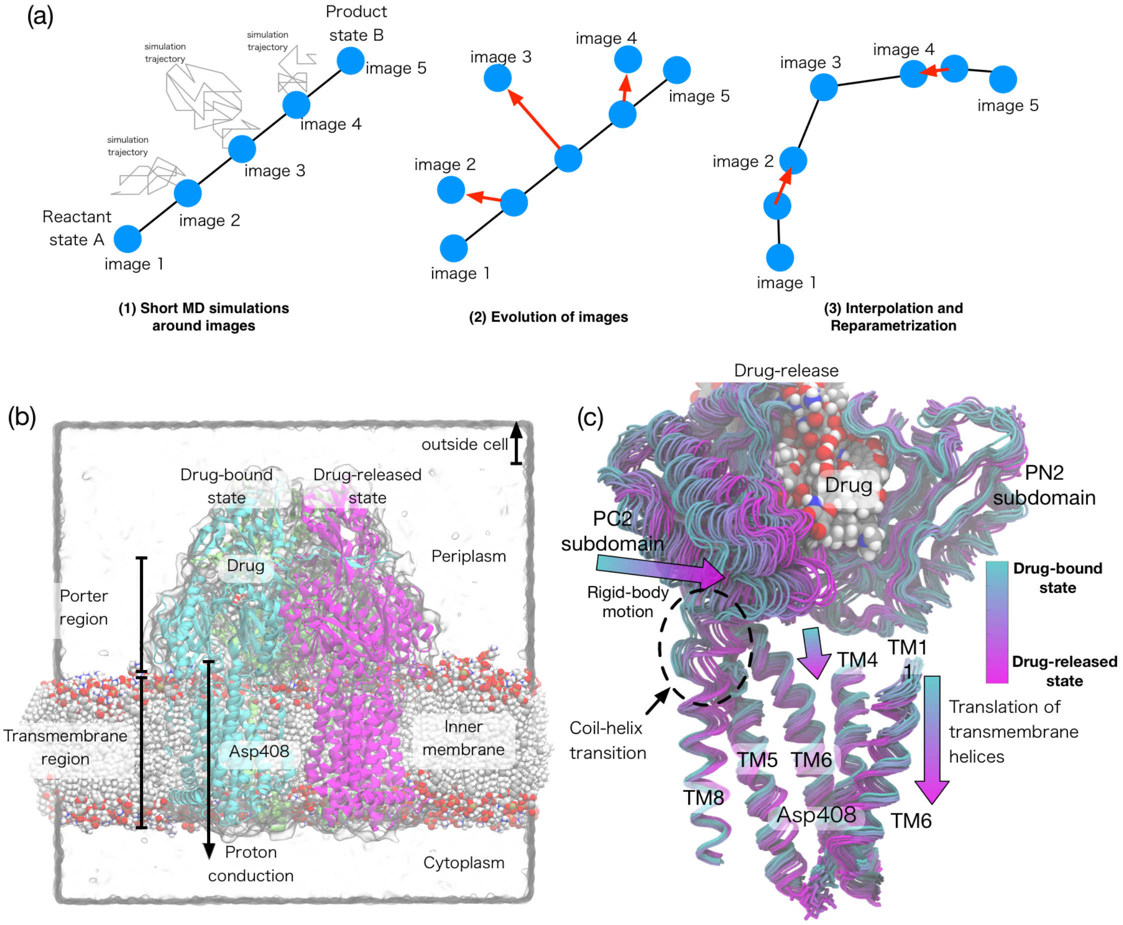

Figure 2. (**a**) Scheme of the string method algorithm. (**b**) Crystal structure of multidrug transporter AcrB embedded in the lipid bilayer. (**c**) Conformational transition (called the functional rotation) pathway which links the proton binding/unbinding in the transmembrane region and the drug transportation in the periplasmic region. Modified and reprinted from Ref. [77] under the terms of the Creative Commons Attribution license.

4. Calculation of kinetics for biomolecules

4.1. Onsager-Machlup action method

The Onsager-Machlup (OM) action method is a genuine path sampling method based on the action principle. As well known in classical mechanics, the Newton equation is derived from the action calculated by a Lagrangian from the least action principle. The same is true for the overdamped Langevin dynamics

$$\frac{dx}{dt} = \frac{1}{\zeta}F(x) + \sqrt{2D}\eta(t) \tag{12}$$

where x is a system variable, $F(x)$ is a force, ζ is a friction coefficient, $D = k_B T/\zeta$ is a diffusion coefficient, and $\eta(t)$ is a Gaussian white noise, satisfying $\langle \eta(t)\eta(0)\rangle = \delta(t)$. In this case, we can define an action called the OM action as [36]

$$S[x(t)] = \frac{1}{2}\int_0^T dt \left(\frac{dx}{dt} - \frac{1}{\zeta}F(x)\right)^2 \tag{13}$$

where T is the total time for a numerical simulation. This is for the overdamped case, but for the underdamped case a similar action can be defined [80]. The important thing is that the path weight is determined by this action as

$$P[x(t)] \propto e^{-S[x(t)]/2D} \tag{14}$$

and since this looks like a Boltzmann weight for a configuration, we can use all the gimmicks in equilibrium statistical mechanics such as replica exchange, reweighting and so on here in path space. Of course, for numerical simulations, we use a discretized form of the action, and we can map the path space onto a connected beads system with some effective potential energy. The situation is very much similar to mapping a quantum system onto a connected classical beads system by the Chandler-Wolynes mapping [5].

The OM action and the other actions were combined with temperature replica exchange and used for sampling path space [43,81]. As mentioned above, MSES can be combined with the OM action to sample path space of a model polymer [35]. Some researchers used the OM action and the other actions for reweighting in path space [80,82–84]. By minimizing or optimizing the OM action and the other actions, we can obtain a "most probable" path, and such a strategy was used in [85,86].

4.2. Weighted ensemble method

The weighted ensemble (WE) method, which was originally devised by Huber and Kim [87], and further elucidated by Zuckerman and coworkers [44], is a simple method for path sampling [5,8] in some CV space. The conventional path sampling methods (such as the OM action method mentioned above) utilize a weight for a path, whereas the WE method considers a weight for a configuration in CV space. The basic procedure of the WE method is as follows (see Fig. 3(a)). For simplicity, we explain a nonequilibrium type simulation using the WE method, and for the other types of the WE method, see [44] and the references therein.

We start from an initial state represented by several particles (trajectories). Let the number of the particles M. Initially each particle has a weight $1/M$ and summing them up leads to one. In a conventional setting, we divide the CV space into several cells (the hexagon in Fig. 3(a)), and we check each cell every τ second in a MD simulation. We run a normal MD simulation using particles in an initial cell, and some particle might visit the other cells during τ. We then make multiple copies of such a particle in the other cells until the total number of particles becomes M. An important thing is that the total weight over all the cells should become 1 because of the conservation of probability. For example, if a particle with weight $1/M$ enters an empty cell, then $M-1$ particles should be generated and each particle should have a weight $1/M^2$.

If we further run the MD simulation, it is the case where particles enter a fully occupied cell. In such a case, we need to eliminate some of the particles such that the total number of that cell becomes M. And we also need to modify the particle weights accordingly. Hence this procedure can be regarded as a time evolution of a distribution function in CV space such as that calculated by the Fokker-Planck equation [36], and each CV variable is associated with a configuration (of biomolecules). Of course, if we prepare many particles in an initial state and run a normal MD simulation, and make a histogram in CV space, we basically get the same result. An important difference is that we divide the CV space into small cells and we basically monitor the transitions among such cells. If the distance between cells are "small" then the transition is faster than the transition between a reactant and a product state which might be distant each other. The acceleration of the transition is compensated by the smaller particle weights associated with the transition. This is the basic principle of the WE method and why this method works for the rare event problem.

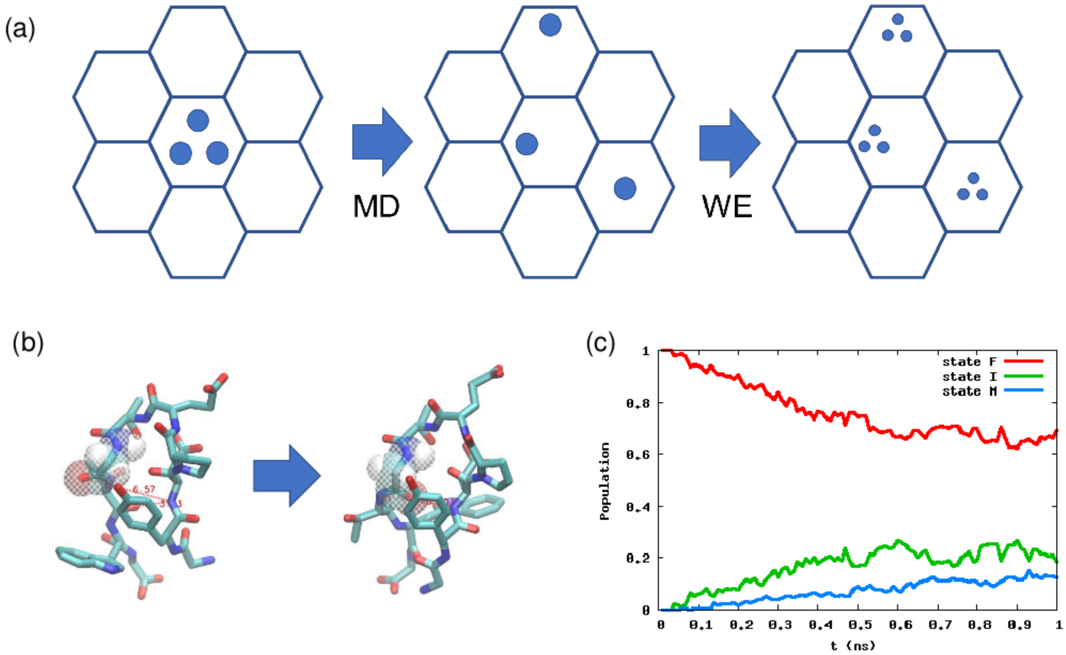

Figure 3. (a) Schematic picture of the weighted ensemble (WE) method. There are two phases: MD phase and WE phase. In the former, we run a normal MD simulation, and in the latter, we make multiple copies of a particle (trajectory) but we modify the particle weights accordingly. (b) Native and misfolded configurations of chignolin. (c) Population dynamics calculated by the WE method. Here the order parameters are the two hydrogen bond distances, which can discriminate the native and misfolded states. For details, see [88].

The WE method has been applied to many systems including folding of a CG protein model, ligand binding, and chemical reaction networks [44]. Here we show some examples recently studied by us. One is conformational change of a small peptide chignolin, which has a native state and misfolded state (Fig. 3(b))[88]. In this case, two hydrogen bond distances are known to be good order parameters, so we divide this 2D space into several cells, and calculate the kinetics using the WE method starting from the native state. The result is shown in Fig. 3(c), and by linear fitting of the misfolded state population, we can estimate the transition time is $\simeq$ 10 ns, which is similar to the mean first passage time calculated by non-Markov type analysis and milestoning in [89].

The other example is the kinetics calculations of the structural change in glucokinase (GCK) [33]. The enhanced sampling using MSES (see Sec. 2) revealed that both the domain motion (between the open and closed forms) and the folding of an inter-domain helix (between the helical and coiled forms) are related to the regulation of the GCK function. Since the derived energy landscape also clarified the pathway of the structural change as, firstly the domain opening from the closed/helical (CH) to open/helical (OH) basins (path1) and secondly the helix collapse from the open/helical (OH) to open/coiled (OC) basins (path2), the WE calculations were carried out for the two paths

Peer-reviewed version available at Int. J. Mol. Sci. 2018, 19, 3177; doi:10.3390/ijms19103177

respectively, and the overall rate constant between the reactant and the product ($k_{\mathrm{CH/OC}}$) was derived in combination with the 3-basin kinetic model and the free energies of CH, OH and OC. As described, the definition of the cells is essential for the success of the WE method: for path1, 22 cells were defined based on the first principal component representing the domain motion, and for path2, the root mean square derivation from the helical form was used to define equally-divided 52 cells. For the two paths, 50-ps MD simulations were carried out for $M = 64$ starting at each bin, 100 iterations were used (the total simulation time being 5 ns), and three WE runs were performed to make the error estimation for the derived rate constants. In the end, we obtained $k_{\mathrm{CH/OC}} = 1.1$ ms^{-1}, indicating the GCK structural change in the timescale of milliseconds. This quantity results in a similar range to the experimental turnover rate of GCK phosphorylation, $k_{\mathrm{cat}} = 0.22$ ms^{-1}, suggesting the energy barrier between the closed/helical basin and the open/coiled basin as the origin of the GCK positive cooperativity.

5. Concluding remarks

In this review article, after explaining the importance of enhanced sampling techniques for configuration and path space of biomolecules, we introduced and discussed the basic principles of the multiscale-enhanced sampling (MSES) method, the string method, the Onsager-Machlup (OM) action method, and the weighted ensemble method, all of which have been used by us for enhance sampling of biomolecules. We also illustrated several applications using above mentioned methods to some biomolecular systems. Our motivations for using these methods are their simplicity for understanding and implementation especially for large biomolecular systems. We hope that this review will facilitate further uses of these methods by the researchers in the field of computational protein dynamics.

Finally we mention our expectation of the future directions for computational protein dynamics.

(a) Larger molecular systems: Obviously, these enhanced sampling techinques will be used for much larger systems; recent focus of computational studies is on protein-protein, protein-DNA, protein-RNA, complexes, proteins in membrane, proteins in crowded environment, so the signaling pathways in a cell would be a next target [90] as already studied in [30,34]. Though there are several attempts to model atomistic details in a cell [91], multiscale type methods such as MSES would be quite useful here. A signaling pathway represents sequential molecular processes including proteins associations, dissociations, and associated chemical reactions and sampling all the molecular processes is not feasible, so the path sampling ideas such as the string method and the OM method should play a role. Multicellular dynamics would be another target [92]. Here it is quite unrealistic to model all the atomistic details, so a coarse-grained model for multiple cells such as cellular Potts models would be combined with MD simulations representing the associated molecular processes.

(b) More efficient methods: Recent advance in computational resources such as GPGPU and Anton [6] is very promising, but it is still necessary to devise efficient numerical algorithms. In particular, for sampling dynamics or kinetics, we need to combine the path sampling algorithms with conventional MD simulations, so the slowness of the latter would be a bottleneck. In material science, hyperdynamics [93] is usually used to accelerate the barrier crossing processes, but it assumes the transition state theory (TST), which works best if the barrier height is much larger than $k_B T$. It is, however, not always the case for protein dynamics and we need to use the path sampling ideas to accelerate the dynamics. Also it would be promising to accelerate MD simulations using some novel ideas of machine learning [94–97].

(c) Right CVs or reaction coordinates: It is always the issue how to choose "right" CVs and the quality of the calculations heavily relies on it. We usually take intuitive and chemically "reasonable" variables such hydrogen bond distances or geometrical measures such as root mean square displacement (RMSD) from a reference structure, or radius of gyration (R_G) etc. As mentioned above (Sec. 3.2), the committor function is the best reaction coordinate, and for its calculation we better know the transition states in advance, but it is always not the case. The string method can extract an importance multidimensional curve connecting a reactant and product, but for diffusive pathways, the application of the string method encounters some difficulties. Hence it has been a trend to borrow some ideas from recently developed statistics methods to this field of protein dynamics. Principal component analysis (PCA) [15] has been used for extracting "large" functional motions of biomolecules as principal modes, but it is a linear and static analysis, and recently several other methods have been developed. Relaxation mode analysis (RMA) [98] or time-structure based independent component analysis (tICA) [99] are such methods and it can extract the slowest motions of biomolecules. Manifold learning techniques such as ISOMAP [100,101] and diffusion map [88,102,103] have been applied to a CG model or atomistic protein systems. Other machine learning techniques are also promising and wait for further applications to biomolecular systems [94–97].

Acknowledgments: We would like to thank Prof. Akinori Kidera, Prof. Tohru Terada, and Dr. Yuji Sugita for their guidance and collaborations. This research was partly supported by JST PRESTO (Grant No. JPMJPR1679) (to YM), AMED-CREST, AMED (to HF), JSPS (KAKEN 16K00059, 17KT0101) (to HF) and (25840060, 15K18520, 17KT0101) (to KM) , and the Platform Project for Supporting Drug Discovery and Life Science Research (BINDS) from AMED (JP18am0101109) (to KM).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MM: Molecular mechanics
CG: Coarse grained
CV: Collective variable
MSES: Multiscale enhanced sampling
MFEP: Minimum free energy path
OM: Onsager-Machlup
WE: Weighted ensemble

Bibliography

1. Petsko, G.A.; Ringe, D., *Protein Structure and Function (Primers in Biology)*, Oxford University Press (2008).
2. Srajer, V.; Schmidt, M., Watching proteins function with time-resolved x-ray crystallography. *J. Phys. D: Appl. Phys.*, **2017**, *50*, 373001.
3. Vos, M.H.; Liebl, U. Time-resolved infrared spectroscopic studies of ligand dynamics in the active site from cytochrome c oxidase. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, **2015**, *1847*, 79-85.
4. Balakrishnan, G.; Weeks, C.L.; Ibrahim, M.; Soldatova, A.V.; Spiro, T.G. Protein Dynamics from Time-Resolved UV Raman Spectroscopy. *Curr Opin Struct Biol.* **2008**, *18*, 623–629.
5. Allen, M.P.; Tildesley, D.J. *Computer Simulation of Liquids, Second Edition*, Oxford University Press (2017).
6. Shaw, D.E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R.O.; Eastwood, M.P.; Bank, J.A.; Jumper, J.M.; Salmon, J.K.; Shan, Y.; Wriggers, W. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science*, **2010**, *330*, 341–346.
7. E, W. *Principles of Multiscale Modeling*, Cambridge Univ. Press (2011).
8. Peters, B. *Reaction Rate Theory and Rare Events*, Elsevier (2017).
9. Takada, S. Coarse-grained molecular simulations of large biomolecules. *Curr. Opin. Struct. Biol.* **2012**, *22*, 130.
10. Ingólfsson, H. I.; Lopez, C. A.; Uusitalo, J. J.; de Jong, D. H.; Gopal, S. M.; Periole, X.; Marrink, S. J. The power of coarse graining in biomolecular simulations. *WIREs Comput. Mol. Sci.* **2014**, *4*, 225–248.
11. Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A.E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.*, **2016**, *116*, 7898–7936.
12. Pietrucci, F. Strateties for the exploraion of free energy landscapes: Unity in diversity and challenges ahead. *Reviews in Physics*, **2017**, *2*, 32-45.
13. Maragliano, L.; Vanden-Eijnden, E. A Temperature Accelerated Method for Sampling Free Energy and Determining Reaction Pathways in Rare Events Simulations. *Chem. Phys. Lett.* **2006**, *426*, 168-175.

14. Hamelberg, D.; Mongan, J.; McCammon, J.A. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.*, **2004**, *120*, 11919-29.
15. Kitao, A.; Takemura, K. High anisotropy and frustration: the keys to regulating protein function efficiently in crowded environments. *Curr. Opin. Struct. Biol.*, **2017**, *42*, 50-58.
16. Fuchigami, S.; Matsunaga, Y.; Fujisaki, H.; Kidera, A. Protein Functional Motion: Basic Concepts and Computational Methodologies. *Adv. Chem. Phys.*, **2011**, *145*, 35-82.
17. Grubmüller, H. Predicting slaw structural transitions in macromolecular systems: Conformational flooding. *Phys. Rev. E*, **1995**, *52*, 2893-2906.
18. Nakajima, N.; Nakamura, H.; Kidera, A. Multicanonical Ensemble Generated by Molecular Dynamics Simulation for Enhanced Conformational Sampling of Peptides. *J. Phys. Chem. B*, **1997**, *101*, 817–824.
19. Marinelli, F.; Faraldo-Gómez, J.D. Ensemble-Biased Metadynamics: A Molecular Simulation Method to Sample Experimental Distributions. *Biophys. J.*, **2015**, *108*, 2779–2782.
20. Comer,J.; Gumbart, J.C.; Hénin, J.; Lelièvre, T.; Pohorille, A.; Chipot, C. The Adaptive Biasing Force Method: Everything You Always Wanted To Know but Were Afraid To Ask. *J. Phys. Chem. B*, **2015**, *119*, 1129-1151.
21. Harada, R.; Kitao, A. Parallel Cascade Selection Molecular Dynamics (PaCS-MD) to generate conformational transition pathway. *J. Chem. Phys.*, **2013**, *139* 035103.
22. Hukushima, K.; Nemoto, K. Exchange Monte Carlo Method and Application to Spin Glass Simulations. *J. Phys. Soc. Jpn.*, **1996**, *65*, 1604-1608.
23. Hansmann, U.H.E. Parallel Tempering Algorithm for Conformational Studies of Biological Molecules. *Chem. Phys. Lett.*, **1997**, *281*, 140-150.
24. Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, **1999**, *314*, 141-151.
25. Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.*, **2000**, *113*, 6042–6051.
26. Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian Replica Exchange Method for Efficient Sampling of Biomolecular Systems: Application to Protein Structure Prediction. *J. Chem. Phys.* **2002**, *116*, 9058-9067.
27. Moritsugu, K.; Terada, T.; Kidera, A. Scalable Free Energy Calculation of Proteins via Multiscale Essential Sampling. *J. Chem. Phys.* **2010**, *133*, 224105.
28. Moritsugu, K.; Terada, T.; Kidera, A. Disorder-to-order Transition of an Intrinsically Disordered Region of Sortase Revealed by Multiscale Enhanced Sampling. *J. Am. Chem. Soc.* **2012**, *134*, 7094-7101.
29. Moritsugu, K.; Terada, T.; Kidera, A. Multiscale Enhanced Sampling Driven by Multiple Coarse-grained Models. *Chem. Phys. Lett.* **2014**, *616-617*, 20-24.
30. Moritsugu, K.; Terada, T.; Kidera, A. Energy Landscape of All-Atom Protein-Protein Interactions Revealed by Multiscale Enhanced Sampling. *Plos Comp. Biol.* **2014**, *10*, e1003901.
31. Moritsugu, K.; Terada, T.; Kidera, A. Multiscale Enhanced Sampling for Protein Systems: An Extension via Adiabatic Separation. *Chem. Phys. Lett.* **2016**, *661*, 279-283.
32. Moritsugu, K.; Terada, T.; Kidera, A. Free Energy Landscape of Protein-Ligand Interactions Coupled with Protein Structural Changes. *J. Phys. Chem. B* **2017**, *121*, 731-740.
33. Moritsugu, K.; Terada, T.; Kokubo, H.; Endo, S.; Tanaka, T.; Kidera, A. Multiscale enhanced sampling of glucokinase: Regulation of the enzymatic reaction via a large scale domain motion. *J. Chem. Phys.* **2018**, *149*, 072314.
34. Moritsugu, K.; Nishi, H.; Inariyama, K.; Kobayashi, M.; Kidera, A. Dynamic recognition and linkage specificity in K63 di-ubiquitin and TAB2 NZF domain complex. *Sci. Rep. submitted.*
35. Fujisaki, H.; Shiga, M.; Moritsugu, K.; Kidera, A. Multiscale enhanced path sampling based on the Onsager-Machlup action: Application to a model polymer. *J. Chem. Phys.* **2013**, *139*, 054117.
36. Zuckerman, D. M., *Statistical Physics of Biomolecules: An Introduction*, CRC Press (2010).
37. Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. String method in collective variables: minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.* **2006**, *125*, 024106.
38. Maragliano, L.; Vanden-Eijnden, E. On-the-fly string method for minimum free energy paths calculation. *Chem. Phys. Lett.* **2007**, *446*, 182–190.
39. Pan, A. C.; Sezer, D.; Roux, B. Finding transition pathways using the string method with swarms of trajectories. *J. Phys. Chem. B* **2008**, *112*, 3432–3440.

40. Nakamura, T. Diffeomorphism invariance requirement on free-energy landscape to describe reaction phenomena. arXiv:1803.09034.
41. Bowman, G.R.; Pande, V.S.; Noé, F. (editors), *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, Springer (2014).
42. Schütte, C.; Noé, F; Lu, J.; Sarich, M.; Vanden-Eijnden, E. Markov state models based on milestoning. *J. Chem. Phys.* **2011**, *134*, 204105.
43. Fujisaki, H.; Shiga, M.; Kidera, A. Onsager–Machlup action-based path sampling and its combination with replica exchange for diffusive and multiple pathways. *J. Chem. Phys.* **2010**, *132*, 134101.
44. Zuckerman, D.M.; Chong, L.T. Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. *Annu. Rev. Biophys.* **2017**, *46*, 43-57.
45. Chodera, J. D.; Shirts, M. R. Replica Exchange and Expanded Ensemble Simulations as Gibbs Sampling: Simple Improvements for Enhanced Mixing. *J. Chem. Phys.* **2011**, *135*, 194110.
46. Rosso, L.; Tuckerman, M. E. An Adiabatic Molecular Dynamics Method for the Calculation of Free Energy Profiles. *Mol. Simul.* **2002**, *28*, 91-112.
47. Morishita, T.; Itoh, S. G.; Okumura, H.; Mikami, M. Free-energy Calculation via Mean-force Dynamics Using a Logarithmic Energy Landscape. *Phys. Rev. E* **2012**, *85*, 066702.
48. Tirion, M. M. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905-1908.
49. Zheng, W.; Brooks, B. R.; Hummer, G. Protein Conformational Transitions Explored by Mixed Elastic Network Models. *Proteins* **2007**, *69*, 43-57.
50. Yasar, F.; Bernhardt, N. A.; Hansmann, U. H. E. Replica-Exchange-with-Tunneling for Fast Exploration of Protein Landscapes. *J. Chem. Phys.* **2015**, *143*, 224102.
51. Bernhardt, N. A.; Xi, W.; Wang, W.; Hansmann, U. H. E. Simulating Protein Fold Switching by Replica-Exchange-with-Tunneling. *J. Chem. Theor. Comp.* **2016**, *12*, 5656-5666.
52. Zhang, H.; Xi, W.; Hansmann, U. H. E.; Wei, Y. Fibril-Barrel Transitions in Cylindrin Amyloids. *J. Chem. Theor. Comp.* **2017**, *13*, 3936–3944.
53. Xi, W.; Hansmann, U. H. E. Conversion between parallel and antiparallel β-sheets in wild type and Iowa mutant Aβ40 fibrils. *J. Chem. Phys.* **2018**, *148*, 045103.
54. Xi, W.; Vanderford, E. K.; Hansmann, U. H. E. Out-of-Register Aβ42 Assemblies as Models for Neurotoxic Oligomers and Fibrils. *J. Chem. Theor. Comp.* **2018**, *14*, 1099-1110.
55. Bernhardt, N. A.; Hansmann, U. H. E. Simulating Protein Fold Switching by Replica Exchange with Tunneling. *J. Phys. Chem. B* **2018**, *122*, 1600-1607.
56. Lee, K. H.; Chen, J. H. Multiscale enhanced sampling of intrinsically disordered protein conformations. *J. Comput. Chem.* **2016**, *37*, 550-557.
57. Vanden-Eijnden, E.: Transition Path Theory. In *Computer Simulations in Condensed Matter Systems: From Materials to CHemical Biology* Vol. 1; Ferrario, M., Ciccotti, G., Binder, K., Eds.; Springer: Berlin, 2007; pp. 453-493.
58. Jiang, W.; Phillips, J. C.; Huang, L.; Fajer, M.; Meng, Y.; Gumbart, J. C.; Luo, Y.; Schulten, K.; Roux, B. Generalized scalable multiple copy algorithms for molecular dynamics simulations in NAMD. *Comput. Phys. Commun.* **2014**, *185*, 908–916.
59. Jung, J.; Mori, T.; Kobayashi, C.; Matsunaga, Y.; Yoda, T.; Feig, M.; Sugita, Y. GENESIS: a hybrid-parallel and multi-scale molecular dynamics simulator with enhanced sampling algorithms for biomolecular and cellular simulations. *WIREs Comput. Mol. Sci.* **2015**, *5*, 310-323.
60. Kobayashi, C.; Jung, J.; Matsunaga, Y.; Mori, T.; Ando, T.; Tamura, K.; Kamiya, M.; Sugita, Y. GENESIS 1.1: A hybrid-parallel molecular dynamics simulator with enhanced sampling algorithms on multiple computational platforms. *J. Comput. Chem.* **2017**, *38*, 2193–2206.
61. Shiga, M.; Fujisaki, H. A quantum generalization of intrinsic reaction coordinate using path integral centroid coordinate. *J. Chem. Phys.*, **2012**, *136*, 184103.
62. Ruiz-Barragan, S.; Ishimura, K.; Shiga, M. On the hierarchical parallelization of ab initio simulations. *Chem. Phys. Lett.* **2016**, *646*, 130-135.
63. Abrams, C. F.; Vanden-Eijnden, E. Large-scale conformational sampling of proteins using temperature-accelerated molecular dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 4961–4966.

64. Vashisth, H.; Brooks, C. L., III Conformational Sampling of Maltose-Transporter Components in Cartesian Collective Variables Is Governed by the Low-Frequency Normal Modes. *J. Phys. Chem. Lett.* **2012**, *3*, 3379–3384.
65. Pan, A. C.; Weinreich, T. M.; Shan, Y.; Scarpazza, D. P.; Shaw, D. E. Assessing the Accuracy of Two Enhanced Sampling Methods Using EGFR Kinase Transition Pathways: The Influence of Collective Variable Choice. *J. Chem. Theory Comput.* **2014**, *10*, 2860–2865.
66. Matsunaga, Y.; Komuro, Y.; Kobayashi, C.; Jung, J.; Mori, T.; Sugita, Y. Dimensionality of Collective Variables for Describing Conformational Changes of a Multi-Domain Protein. *J. Phys. Chem. Lett.* **2016**, *7*, 1446–1451.
67. Gan, W.; Yang, S.; Roux, B. Atomistic view of the conformational activation of Src kinase using the string method with swarms-of-trajectories. *Biophys. J.* **2009**, *97*, L8–L10.
68. Ovchinnikov, V.; Karplus, M.; Vanden-Eijnden, E. Free energy of conformational transition paths in biomolecules: the string method and its application to myosin VI. *J. Chem. Phys.* **2011**, *134*, 085103.
69. Zhu, F.; Hummer, G. Pore opening and closing of a pentameric ligand-gated ion channel. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 19814–19819.
70. Lev, B.; Murail, S.; Poitevin, F.; Cromer, B. A.; Baaden, M.; Delarue, M.; Allen, T. W. String method solution of the gating pathways for a pentameric ligand-gated ion channel. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, E4158–E4167.
71. Stober, S. T.; Abrams, C. F. Energetics and mechanism of the normal-to-amyloidogenic isomerization of beta2-microglobulin: on-the-fly string method calculations. *J. Phys. Chem. B* **2012**, *116*, 9371–9375.
72. Singharoy, A.; Chipot, C.; Moradi, M.; Schulten, K. Chemomechanical Coupling in Hexameric Protein–Protein Interfaces Harnesses Energy within V-Type ATPases. *J. Am. Chem. Soc.* **2017**, *139*, 293–310.
73. Das, A.; Rui, H.; Roux, B. Conformational Transitions and Alternating-Access Mechanism in the Sarcoplasmic Reticulum Calcium Pump. *J. Mol. Biol.* **2017**, *429*, 647–666.
74. Moradi, M.; Tajkhorshid, E. Computational Recipe for Efficient Description of Large-Scale Conformational Changes in Biomolecular Systems. *J. Chem. Theory Comput.* **2014**, *10*, 2866–2880.
75. Moradi, M.; Enkavi, G.; Tajkhorshid, E. Atomic-level characterization of transport cycle thermodynamics in the glycerol-3-phosphate:phosphate antiporter. *Nat. Commun.* **2015**, *6*, 8393.
76. Matsunaga, Y.; Fujisaki, H.; Terada, T.; Furuta, T.; Moritsugu, K.; Kidera, A. Minimum Free Energy Path of Ligand-Induced Transition in Adenylate Kinase. *PLoS Comput. Biol.* **2012**, *8*, e1002555.
77. Matsunaga, Y.; Yamane, T.; Terada, T.; Moritsugu, K.; Fujisaki, H.; Murakami, S.; Ikeguchi, M.; Kidera, A. Energetics and conformational pathways of functional rotation in the multidrug transporter AcrB. *eLife* **2018**, *7*, 243.
78. Wolf-Watz, M.; Thai, V.; Henzler-Wildman, K.; Hadjipavlou, G.; Eisenmesser, E. Z.; Kern, D. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nat. Struct. Mol. Biol.* **2004**, *11*, 945–949.
79. Du, D.; van Veen, H. W.; Murakami, S.; Pos, K. M.; Luisi, B. F. Structure, mechanism and cooperation of bacterial multidrug transporters. *Curr. Opin. Struct. Biol.* **2015**, *33*, 76–91.
80. Grazioli, G.; Andricioaei, I. Advances in milestoning. I. Enhanced sampling via wind-assisted reweighted milestoning (WARM). *J. Chem. Phys.* **2018**, *149*, 084103.
81. Chodera, J.D.; Swope, W.C.; Noe, F.; Prinz, J.-H.; Shirts, M.R.; Pande, V.S. Dynamical reweighting: Improved estimates of dynamical properties from simulations at multiple temperatures. *J. Chem. Phys.* **2011**, *134*, 244107.
82. Zuckerman, D.M.; Woolf, T.B. Efficient dynamic importance sampling of rare events in one dimension. *Phys. Rev. E*, **2001**, *63*, 016702.
83. Takayanagi S.; Iba, Y. Backward Simulation of Stochastic Process using a Time Reverse Monte Carlo method. arXiv:1708.08045.
84. Donati, L.; Hartmann, C.; Keller, B.G. Girsanov reweighting for path ensembles and Markov state models. *J. Chem. Phys.* **2017**, *146*, 244112.
85. Beccara, S.; Škrbić, T.; Covino, R.; Faccioli, P. Dominant folding pathways of a WW domain. *Proc. Natl. Acad. Sci. U.S.A.*, **2012**, *109*, 2330-2335.
86. Lee, J.; Lee, I.-H.; Joung, I.; Lee, J.; Brooks, B.R. Finding multiple reaction pathways via global optimization of action. *Nat. Commun.*, **2017**, *8*, 15443.

87. Huber, G.A.; Kim, S. Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophys. J.*, **1996**, *70*, 97-110.
88. Fujisaki, H.; Moritsugu, K.; Mitsutake, A.; Suetani, H. Conformational change of a biomolecule studied by the weighted ensemble method: Use of the diffusion map method to extract reaction coordinates. submitted to J. Chem. Phys.
89. Fujisaki, H.; Mitsutake, A.; Maragliano, L. Numerical investigation of kinetic properties of a small peptide using non-Markov type analysis and milestoning. unpublished.
90. Marks, F.; Klingmüller, U.; Müller-Decker, K. *Cellular Signal Processing: An Introduction to the Molecular Mechanisms of Signal Transduction*, 2nd ed. Garland Science (2017)
91. Trovato, F.; Fumagalli, G. Molecular simulations of cellular processes. *Biophys. Rev.*, **2017**, *9*, 941-958
92. Mak, M.; Kim, T.; Zaman, M.H.; Kamm, R.D. Multiscale mechanobiology: computational models for integrating molecules to multicellular systems. *Integr. Biol.*, **2015**, DOI:10.1039/c5ib00043b.
93. Chena, L.Y.; Horing, N.J.M. An exact formulation of hyperdynamics simulations. *J. Chem. Phys.*, **2007**, *126*, 224103.
94. Sultan, M.M.; Wayment-Steele, H.K.; Pande, V.S. Transferable Neural Networks for Enhanced Sampling of Protein Dynamics. *J. Chem. Theory Comput.*, **2018**, *14*, 1887–1894.
95. Sultan, M. M.; Pande, V. S. Automated design of collective variables using supervised machine learning. *J. Chem. Phys.*, **2018**, *149*, 094106.
96. Endo, K.; Tomobe, K.; Yasuoka, K. Multi-Step Time Series Generator for Molecular Dynamics. *The Thirty-Second AAAI Conference on Artificial Intelligence*, AAAI Publications (2018).
97. Brandt, S.; Sittel, F.; Ernst, M.; Stock, G. Machine Learning of Biomolecular Reaction Coordinates. *J. Phys. Chem. Lett.*, **2018**, *9*, 2144-2150.
98. Mitsutake, A.; Takano, H. Relaxation mode analysis and Markov state relaxation mode analysis for chignolin in aqueous solution near a transition temperature. *J. Chem. Phys.*, **2015**, *143*, 124111.
99. Naritomi, Y.; Fuchigami, S. Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: the case of domain motions. *J. Chem. Phys.*, **2011**, *134*, 065101.
100. Suetani, H.; Soejima, K.; Matsuoka, R.; Parlitz, U.; Hata, H. Manifold learning approach for chaos in the dripping faucet. *Phys. Rev. E*, **2012**, *86*, 036209.
101. Ito, R.; Yoshidome, T. An Accurate Computation of an Order Parameter with a Markov State Model Constructed using a Manifold-Learning Technique. *Chem. Phys. Lett.*, **2018**, *691*, 22-27.
102. Rohrdanz, M.A.; Zheng, W.; Maggioni, M.; Clementi, C. Determination of reaction coordinates via locally scaled diffusion map. *J. Chem. Phys.*, **2011**, *134*, 124116.
103. Nedialkova, L.V.; Amat, M.A.; Kevrekidis, I.G.; Hummer, G. Diffusion maps, clustering and fuzzy Markov modeling in peptide folding transitions. *J. Chem. Phys.*, **2014**, *141*, 114102.