Extended phase-space methods for enhanced sampling in molecular simulations: a review

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

[1] Department of Physics, Nippon Medical School, Tokyo, Japan, [2] Graduate School of Medical Life Science, Yokohama City University, Yokohama, Japan, [3] Advanced Institute for Computational Science, Kobe, Japan, [4] National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan, [5] Center for Synaptic Neuroscience, Istituto Italiano di Tecnologia, Genova, Italy

Molecular Dynamics simulations are a powerful approach to study biomolecular conformational changes or protein–ligand, protein–protein, and protein–DNA/RNA interactions. Straightforward applications, however, are often hampered by incomplete sampling, since in a typical simulated trajectory the system will spend most of its time trapped by high energy barriers in restricted regions of the configuration space. Over the years, several techniques have been designed to overcome this problem and enhance space sampling. Here, we review a class of methods that rely on the idea of extending the set of dynamical variables of the system by adding extra ones associated to functions describing the process under study. In particular, we illustrate the Temperature Accelerated Molecular Dynamics (TAMD), Logarithmic Mean Force Dynamics (LogMFD), and Multiscale Enhanced Sampling (MSES) algorithms. We also discuss combinations with techniques for searching reaction paths. We show the advantages presented by this approach and how it allows to quickly sample important regions of the free-energy landscape via automatic exploration.

Keywords: molecular dynamics, rare events, free energy, biomolecular interactions, conformational transitions, protein–ligand binding

1. Introduction

Molecular dynamics (MD) simulations have become a fundamental tool to study biological systems at atomic scale (Perilla et al., 2015). MD allows to simulate the motion of each single atom of complex biomolecules in accurately modeled environments, thus providing information about molecular mechanisms at a scale which is still impossible to access experimentally. Such information can in turn be of extreme value to design new, more insight-driven experiments. A major difficulty in MD is, however, the well known time-scale issue: most of the phenomena of (biophysical) interest occur on times that are still inaccessible by standard simulations. Few remarkable exceptions exist (Freddolino et al., 2010; Lindorff-Larsen et al., 2011), although even in these cases it remains challenging to accrue enough statistics for a comprehensive understanding of the process.

The problem originates from the presence of dynamical hindrances of energetic or entropic nature, which confine the system in specific regions of phase space. Transitions among those metastable states, while being *rare events* on the simulation timescale, are often a mandatory requirement for biological function. This is the case, for example, of a receptor changing

conformation to accommodate a ligand, or of two associating proteins in solvent forming the encounter complex.

A thermodynamic quantity of pivotal importance in MD simulations of rare events is the free energy, or potential of mean force (Roux, 1995), since it provides information on the metastable states and the transitions between them. In particular, via the free energy it is possible to obtain the relative probability of different molecular conformations, or the rate at which dynamical transitions occur between them. From its statistical mechanics definition, the free energy is related to the probability density of a set of collective variables (CVs), which are typically used to study a reactive event. These are functions of the Cartesian variables of the original system, such as distances and angles between atoms, or more complicated functions. Unfortunately, due to the sampling issue described above, free-energy calculations are quite demanding. Many numerical methods have been developed in order to accelerate, or enhance, sampling in MD simulations and thus facilitate free-energy calculations. Notable examples include umbrella sampling (Torrie and Valleau, 1974), metadynamics (Laio and Parrinello, 2002), accelerated molecular dynamics (Hamelberg et al., 2004), adaptive biasing force (Darve and Pohorille, 2001), or replica exchange methods (Fukunishi et al., 2002). For a recent review, see Bernardi et al. (2015).

In this review, we illustrate a number of methods that were developed in recent years to reconstruct free-energy surfaces and transition pathways between regions over them. The distinctive feature of these methods is that they are all based on the introduction of an extended system where a set of extra variables is evolved dynamically together with the variables of the physical system. Extending the thermodynamic ensemble of the simulated system is a compelling strategy in computational chemistry/physics, and it has been exploited for example to control temperature and pressure (Nosé, 1984; Hoover, 1985), calculate chemical free energies (Kong and Brooks, 1996; Bitetti-Putzer et al., 2003), or enhance sampling by introducing *generalized ensembles* made of multiple replicas of the full physical system (Mitsutake et al., 2001). In the present context, the added variables are typically associated to the CVs, and evolved together with the physical variables in either a sequential or a concurrent setting. This strategy can be considered within the framework of multiscale computational approaches (E et al., 2007a), and presents several advantages, which are discussed in this review. In the next sections, we divide the discussed methods in two classes. First, we examine techniques designed to explore and reconstruct free-energy surfaces, while afterwards we illustrate methods devised to determine optimal reaction pathways between metastable states.

2. Methods to Explore and Reconstruct the Free-Energy Landscape

Let us consider a system whose configurational state is specified by $x \in \mathbb{R}^N$, and suppose we are interested in the statistical behavior of a set of collective variables $\theta(x) = (\theta_1(x), \ldots, \theta_n(x))$. If $z = (z_1, \ldots, z_n)$ is a particular realization of these variables, the free energy $F(z)$ of the system is defined as

$$F(z) = -\beta^{-1} \ln \left(Z^{-1} \int_{\mathbb{R}^N} e^{-\beta V(x)} \prod_{\alpha=1}^{n} \delta(\theta_\alpha(x) - z_\alpha) dx \right), \quad (1)$$

where $Z = \int_{\mathbb{R}^N} e^{-\beta V(x)} dx$, $V(x)$ is the potential energy of the system and $\beta = 1/k_B T$, where k_B is the Boltzmann constant and T the temperature. Note that $F(z)$ depends on a specific temperature β^{-1}, determined, for example, based on experimental information on the process to be studied. The definition equation (1) implies that $e^{-\beta F(z)}$ is the probability density function of the variables z.

A natural way to reconstruct $F(z)$ would be to consider an artificial dynamics where the variables z are evolved on the free-energy landscape $F(z)$ defined at β^{-1} using an artificial temperature $\bar{\beta}^{-1}$ higher than the physical one. This can be realized for example using the dynamics

$$\mu_\alpha \ddot{z}_\alpha = -\frac{\partial F(z)}{\partial z_\alpha} + \text{thermostat at } \bar{\beta}^{-1}, \quad (2)$$

where μ_α is the artificial mass of the fictitious particle α, and $\bar{\beta} = 1/k_B \bar{T}$ is the inverse of an artificial temperature $\bar{T}$, higher than physical temperature T (more details will be given below). The choice of the thermostat in equation (2) has no effect on the argument discussed here, and one could use Langevin, or Nosé-Hoover (Nosé, 1984; Hoover, 1985) or related techniques (Martyna et al., 1992; Liu and Tuckerman, 2000; Morishita, 2010). Since the equilibrium density of the dynamics equation (2) is proportional to $e^{-\bar{\beta} F(z)}$, by setting $\bar{\beta}^{-1} > \beta^{-1}$ the $z(t)$ solution of equation (2) will meander through $F(z)$ being able to cross energy barriers higher than β^{-1}, thus visiting also regions inaccessible to the physical system with thermal energy β^{-1}.

Such approach is indeed possible, at least in principle, even if the full $F(z)$ is not known, since the quantity $-\partial F(z)/\partial z_\alpha$ (called the *mean force*) can be calculated locally at any point z, via constrained simulations on $\theta(x) = z$ using the *blue moon* ensemble method (Carter et al., 1989; Ciccotti et al., 2005). In fact, this idea is at the basis of the methods that we will review in the next sections.

2.1. Temperature Accelerated Molecular Dynamics (TAMD)
2.1.1. Overview of the Method
As anticipated at the end of the previous section, integrating equation (2) can be an efficient way to explore the unknown free-energy surface $F(z)$, since the mean force $-\nabla F$ at a point z can be estimated via a conditional expectation on $\theta(x) = z$ (Carter et al., 1989; Ciccotti et al., 2005), even when $F(z)$ is unknown. This approach, however, has two difficulties: first, after each integration step providing the updated z values, the configuration of the physical system may be far from satisfying the new constraints [i.e., $\theta(x)$ will be different from z], which may lead to re-initialization problems; second, it requires to determine the length of the time-averaging window to compute ∇F, which is an extra parameter to adjust.

To overcome these difficulties, in TAMD (Maragliano and Vanden-Eijnden, 2006), it is proposed to simulate effectively equation (2) by using the following extended system of equations

$$\begin{cases} m_i \ddot{x}_i = -\frac{\partial V(x)}{\partial x_i} - \kappa \sum_{\alpha=1}^{n} (\theta_\alpha(x) - z_\alpha) \frac{\partial \theta_\alpha(x)}{\partial x_i} + \text{thermostat at } \beta^{-1}, \\ \mu_\alpha \ddot{z}_\alpha = \kappa (\theta_\alpha(x) - z_\alpha) + \text{thermostat at } \bar{\beta}^{-1}, \end{cases} \quad (3)$$

where κ is a parameter whose role will be specified later, and m_i is the mass of the atom i. We note that the specific dynamics of the x and z variables in equation (3) is not important. In the original formulation, TAMD was presented using overdamped Langevin in equations (3) and (2), while several further applications have shown the effectiveness of other dynamical schemes, such as second order Langevin (Monteferrante et al., 2008), or Nosé-Hoover (Trigila, 2007). For simplicity, we use here second order dynamics for both x and z leaving the thermostatting technique unspecified.

The system (3) describes the evolution of the extended system (x, z) under the effect of the potential energy

$$U_\kappa(x, z) = V(x) + \frac{1}{2}\kappa \sum_{\alpha=1}^{n} (z_\alpha - \theta_\alpha(x))^2. \qquad (4)$$

In the TAMD paper (Maragliano and Vanden-Eijnden, 2006), it was shown why a system like equation (3) can be used to sample the free energy $F(z)$. Here, we summarize the results and refer to the original paper the reader interested in the details. The key argument is that if in equation (3) the artificial masses μ_α are chosen such that $\mu_\alpha >> m_i$, then the variables $z(t)$ evolve much more slowly than the $x(t)$, feeling only their average effect. As a consequence, the variables $z(t)$ evolve according to an effective equation which is obtained by averaging the $z(t)$ equations (3) with respect to the probability density function for $x(t)$ from equation (3) at $z(t) = z$ fixed. More in detail, if we introduce the quantity

$$F_\kappa(z) = -\beta^{-1} \ln \left(Z_\kappa^{-1} \int_{\mathbb{R}^N} e^{-\beta U_\kappa(x,z)} dx \right), \qquad (5)$$

with $Z_\kappa = \int_{\mathbb{R}^N \times \mathbb{R}^n} e^{-\beta U_\kappa(x,z)} dx dz$, if $\mu_\alpha >> m_i$ the variables $z(t)$ solution of equation (3) satisfy approximately the effective equation

$$\mu_\alpha \ddot{z}_\alpha = -\frac{\partial F_\kappa(z)}{\partial z_\alpha} + \text{thermostat at } \bar{\beta}^{-1}, \qquad (6)$$

where

$$-\frac{\partial F_\kappa(z)}{\partial z_\alpha} = -Z_\kappa^{-1} \int_{\mathbb{R}^N} \kappa(z_\alpha - \theta_\alpha(x)) e^{-\beta U_\kappa(x,z)} dx. \qquad (7)$$

The quantity $F_k(z)$ in equation (5) is related to the free energy equation (1): it is a smoothed, also said mollified, version of the true free energy equation (1), which can be made as similar as desired to it by taking κ large in equation (3). How large, can be decided by observing that

$$e^{-\beta F_\kappa(z)} = \int_{\mathbb{R}^n} e^{-\beta F(z')} e^{-\beta \frac{1}{2}\kappa \sum_{\alpha=1}^{n} (z'_\alpha - z_\alpha)^2} dz', \qquad (8)$$

i.e., $F_k(z)$ is a filtered version of $F(z)$ with variance $1/\beta\kappa$. Hence, κ should be chosen considering that $1/\sqrt{\beta\kappa}$ sets the scale at which $F(z)$ will be resolved. Finally, since in the same condition of large κ we also have that $-\frac{\partial F_\kappa(z)}{\partial z_\alpha} \approx -\frac{\partial F(z)}{\partial z_\alpha}$, we obtain that the evolution equation (6) is an approximation of equation (2) [for explicit expressions of the errors this introduces, see Maragliano and Vanden-Eijnden (2006)].

The interesting property of equation (6) is that the probability distribution of the z variables obeying that dynamics is $e^{-\bar{\beta} F_\kappa(z)}$. This means that if we simulate the extended system equation (3) and we reconstruct the probability distribution of the z variables we can directly obtain the free energy $F_k(z)$. Since this holds at any artificial temperature $\bar{\beta}^{-1}$, we can choose it such that $\bar{\beta}^{-1} \gtrsim \Delta F$, where ΔF is an estimate of the free-energy barriers, so that the z variables will be able to cross barriers that are higher than the thermal energy at the physical temperature. Indeed, note that increasing the temperature of the CVs is equivalent to decreasing the height of the free-energy barriers.

The other important parameters to adjust to simulate equation (3) are the effective masses μ_α. In practice, the condition $\mu_\alpha >> m_i$ introduces an effective adiabatic separation between the z and the x variables, i.e., it guarantees that the x have time to equilibrate at the new value of z. This property relies on averaging theorems for systems with multiple time scales (Papanicolaou, 1977; Vanden-Eijnden, 2003; Pavliotis and Stuart, 2008), and it yields a useful criterion to choose the effective masses μ_α: specifically, the dynamics of the z must be damped so that it is possible to find a time interval over which the time average of the $\kappa(\theta_\alpha(x) - z_\alpha)$ term in equation (3) [see equation (7)] converges as if the z were essentially fixed.

TAMD shares ideas with other enhanced sampling techniques, in particular, for what concerns the extended phase-space formulation and the use of two temperatures. For the extended system, TAMD borrows from the extended Lagrangian version of metadynamics (Iannuzzi et al., 2003). It is also important to recall that a more recent version of metadynamics uses two different temperatures on the physical and collective variables (Barducci et al., 2008). As it is well-known, however, in metadynamics the CVs are pushed to exit the free-energy minima by flooding their space with Gaussian packets, which accumulate in a history-dependent term entering the equations of motion. Conversely, TAMD achieves barriers crossing by assigning a high temperature to the CVs. This simplifies the use of the technique, since in TAMD the acceleration is controlled by one parameter only ($\bar{\beta}^{-1}$), while in metadynamics one has to specify the width, height, and dropping frequency of the Gaussians. It also makes it more efficient in principle since due to the history-dependent term the cost of a metadynamics run increases rapidly with simulation time. This is particularly true in cases where the number of collective variables is large, since filling a multi-dimensional volume requires a lot of Gaussian packets and thus increases the computational burden to calculate the history-dependent term [for possible numerical strategies to facilitate the calculation of this term, see Babin et al. (2006), Smiatek and Heuer (2011), while an extensive discussion of the parameters of metadynamics can be found in Laio et al. (2005)].

As for the use of different temperatures, efficient control of double temperature systems based on adiabaticity was introduced long ago in MD (Blöchl and Parrinello, 1992), and revised later in the context of enhanced sampling by the work of Rothlisberger [CAFES method (VandeVondele and Rothlisberger, 2002)] and Tuckerman [AFED method (Rosso et al., 2002)]. In CAFES, however, there are no additional equations of motion for the CVs: the full system is divided in a reactive subsystem (S) and the

environment (E), and the dynamics of S is decoupled from E by using large masses and accelerated via a high temperature. The statistics of a CV is then extracted directly from the dynamics of S. AFED equations were also originally formulated in the space of the physical variables x only, however, collective variables are introduced in this case, and their equations of motion obtained via appropriate coordinate transformations. Using an extended system as in TAMD provides a significant simplification, especially for non-trivial $\theta(x)$, and, in fact, the most recent version of AFED, called driven-AFED (Abrams and Tuckerman, 2008), exploits the same idea and is essentially the same as TAMD.

2.1.2. Generalizations and Biophysical Applications of TAMD

As already noted, the equilibrium probability distribution of the $z(t)$ variables that are solution of the system equation (3) is $\rho_\kappa(z) \propto e^{-\bar{\beta}F_\kappa(z)}$, where $F_k(z)$ is given by equation (5). This implies that if we simulate equation (3) and we bin the values of $z(t)$ to get an estimate of $\rho_\kappa(z)$, we can directly reconstruct the free-energy associated to these variables (an operation called Boltzmann inversion). This was indeed the original idea for TAMD. Later (Maragliano and Vanden-Eijnden, 2008), it was demonstrated that TAMD can be used as an efficient means to rapidly browse through the relevant regions of the free-energy landscape (including those surrounded by high energy barriers), generating points where the mean forces are accurately computed at a second stage and used in a variational procedure to reconstruct the free-energy globally. This approach, termed Single-Sweep, has been successfully applied to calculate free-energy surfaces for CO (Maragliano et al., 2010), H_2O (Lapelosa and Abrams, 2013), and O_2 (Bucci and Abrams, 2014) diffusion in myoglobin and sarcosine oxidase.

TAMD has also been used to explore the free-energy space associated to a large number of CVs in several applications: from structure determination of solvated polymers (Lucid et al., 2012) to large-scale protein conformational transitions, such as those of the GroEL chaperonin subunit and the HIV-1 envelope glycoprotein gp120 (Abrams and Vanden-Eijnden, 2010), a maltose transporter (Vashisth and Brooks, 2012), the β2-adrenergic receptor (Nygaard et al., 2013), allosteric inhibition (Vashisth et al., 2013), and nucleotide release (Dror et al., 2015) in G proteins, the insulin receptor kinase (Vashisth et al., 2012a), an adenyl cyclase (Selwa et al., 2014), and the acetylcholine binding protein (Mohammad Hosseini Naveh et al., 2014).

Other attractive applications of TAMD are the determination of coarse-grained force-field parameters (Abrams and Vanden-Eijnden, 2012) and protein structure refinement from electron microscopy maps (Vashisth et al., 2012b). In particular, in the latter case, TAMD was used to enhance the flexible fitting of all-atom protein and RNA models into low-resolution density maps. In Yamamori and Kitao (2013), an interesting combination of TAMD and Replica Exchange Umbrella Sampling (REUS) was proposed. More recently, TAMD was combined with a soft-ratcheting algorithm to achieve a focused exploration of the CV space by relying on low-resolution or even qualitative experimental information (Cortes-Ciriano et al., 2015). Such an approach was successful in providing previously unavailable conformations

of calmoduline-free adenyl cyclase toxin from the causative agent of whooping cough.

2.2. Logarithmic Mean Force Dynamics (LogMFD)

LogMFD was introduced in Morishita et al. (2012) and further discussed in Morishita et al. (2013). In Nakamura et al. (2014), it was combined with a Density Functional Theory approach to study the conformational dynamics of a glycine peptide. Similarly to what was suggested with equation (2), in LogMFD the collective variables are considered dynamical variables, and they are evolved using the following equation of motion

$$\mu_\alpha \ddot{z}_\alpha = -\left(\frac{\sigma\lambda}{\sigma F(z)+1}\right)\frac{\partial F(z)}{\partial z_\alpha} + \text{thermostat at } \bar{\beta}^{-1}, \quad (9)$$

where σ and λ are real positive parameters discussed below, and $\bar{\beta} = 1/k_B\bar{T}$ is again the inverse of an effective temperature, which is in general different from the physical temperature, although in LogMFD is often set as equal to it. By comparing equations (9) with (2), we see that in LogMFD the force acting on the CVs is not simply the gradient of the free energy $F(z)$, but instead that of its logarithmic form $\lambda\log(\sigma F(z)+1)$. Using a logarithmic scaling for $F(z)$ is convenient since it is a non-linear transformation, i.e., it preserves the details of the landscape at low $F(z)$ while it smooths the regions at high values. To avoid $(\sigma F + 1) < 1$, $F(z)$ is shifted by a constant c, i.e., $F = F' + c$.

The advantage of using equation (9) becomes clearer if we consider the Nosé-Hoover thermostat and we introduce the conserved quantity associated to it (Nosé, 1984; Hoover, 1985),

$$\hat{H} = \sum_{\alpha=1}^{n}\frac{1}{2}\mu_\alpha\dot{z}_\alpha^2 + \lambda\log(\sigma F(z)+1) + \frac{1}{2}m_\eta\dot{\eta}^2 + n\bar{\beta}^{-1}, \quad (10)$$

where η is the thermostat variable and m_η its fictitious mass. From equation (10), it is possible to obtain an expression for the free energy $F(z)$,

$$F(z) = \frac{1}{\sigma}\left\{\exp\left[\frac{1}{\lambda}\left(\hat{H} - \sum_{\alpha=1}^{n}\frac{1}{2}\mu_\alpha\dot{z}_\alpha^2 - \frac{1}{2}m_\eta\dot{\eta}^2 - n\bar{\beta}^{-1}\right)\right] - 1\right\}. \quad (11)$$

This indicates that we do not need to explicitly set the shifting constant c, but instead the conserved quantity $\hat{H}$ needs to be set at the beginning of a LogMFD run to guarantee $F(= F' + c) \geq 0$. In particular, $\hat{H}$ should be chosen to satisfy

$$\left(\hat{H} - \sum_{\alpha=1}^{n}\frac{1}{2}\mu_\alpha\dot{z}_\alpha^2 - \frac{1}{2}m_\eta\dot{\eta}^2 - n\bar{\beta}^{-1}\right) \equiv F^*(z) \geq 0. \quad (12)$$

Since the average of the kinetic energy of z and the terms associated with the thermostat variables can be easily determined for a defined temperature $\bar{\beta}^{-1}$, one can roughly estimate an appropriate value of $\hat{H}$. In order to take full advantage of the logarithmic form, F^*/λ should be close to 0 at the bottom of the free-energy landscape. Obviously, we do not know the exact position of the minimum of $F(z)$ before running a LogMFD calculation, but since results are not sensitive to the choice of $\hat{H}$, only a rough estimate of the conserved quantity is sufficient.

Equation (11) shows that a free-energy profile $F(z)$ can be reconstructed directly as the $z(t)$ evolve in a LogMFD run obtained by solving equation (9). The key to this direct estimation is the existence of the conserved quantity $\hat{H}$, and hence, in equation (9), it is possible to use any Nosé-Hoover (NH) type thermostat [NH chains (Martyna et al., 1992), recursive NH (RNH) (Morishita, 2010), or Gaussian moment thermostats (Liu and Tuckerman, 2000)], as long as it exists a conserved quantity analogous to $\hat{H}$ in equation (10). Similarly to equation (2), the mean force in equation (9) can be estimated via conditional averages using the blue moon ensemble or with harmonic restraints as in equation (7).

Moreover, to avoid the problems with averaging and re-initialization discussed in Section 2.1.1, a concurrent evolution scheme for $z(t)$ and $x(t)$ as in TAMD system of equations equation (3) can be envisioned, where the $z(t)$ are damped with respect to the $x(t)$ by larger masses.

Two important parameters are introduced in LogMFD, σ and λ. For simplicity, λ is usually set as $1/\sigma$. The role of σ is to flatten the free-energy landscape during the evolution of $z(t)$. The choice of optimal σ comes from a trade-off between a large value which would enhance barrier crossing and a small value to reduce numerical errors originating from the exponential form of $F(z)$ equation (11). A reasonable choice is to set σ such that $\bar{\beta}^{-1}$ is of the order of $\log(\sigma\Delta F + 1)/\sigma$, where ΔF is an estimate of the free-energy barrier and we have used $\lambda = 1/\sigma$. Note that, as a consequence, the temperature $\bar{\beta}^{-1}$ in LogMFD is usually lower than then the one needed in TAMD for barrier crossing ($\bar{\beta}^{-1} \approx \Delta F$), and it can be even lower than the physical temperature. However, this comes at the price of setting two more parameters, σ and λ. For an extensive investigation on the role of the different LogMFD parameters, we refer the reader to Morishita et al. (2013).

2.3. Multiscale Enhanced Sampling (MSES)

The MSES technique was introduced in Moritsugu et al. (2010) and applied so far to study the disorder to order transition in a sortase protein (Moritsugu et al., 2012), and the formation of a Barnase–Barstar complex (Moritsugu et al., 2014a). In MSES, a physical system x is again coupled to a set of extra variables z via the use of CVs functions $\theta(x)$. The equations of motion for MSES can be written as

$$\begin{cases} m_i \ddot{x}_i = -\dfrac{\partial V(x)}{\partial x_i} - \kappa \displaystyle\sum_{\alpha=1}^{n} (\theta_\alpha(x) - z_\alpha) \dfrac{\partial \theta_\alpha(x)}{\partial x_i} + \text{thermostat at } \beta^{-1}, \\ \mu_\alpha \ddot{z}_\alpha = \kappa(\theta_\alpha(x) - z_\alpha) - \dfrac{\partial V^{CV}(z)}{\partial z_\alpha} + \text{thermostat at } \bar{\beta}^{-1}. \end{cases}$$
$$(13)$$

At variance with TAMD equations (3), the system equation (13) contains also a force on the z variables coming from a potential energy acting on these variables, $V^{CV}(z)$ [a version of MSES using multiple and mutually interacting z sets was presented in Moritsugu et al. (2014b)]. This potential is arbitrarily chosen based on prior knowledge or experimental information. In general, it is designed to guide sampling by relying on well-established coarse-grained models. For example, in the study of the Barnase–Barstar complex (Moritsugu et al., 2014a), $C\alpha$–$C\alpha$ distances were used as CVs, and the CV potential $V^{CV}(z)$ was prepared as the sum of two terms representing

intra- and inter-molecular interactions. For intra-molecular interactions the potential function of the $C\alpha$ elastic network model was used (Tirion, 1996), while for inter-molecular interactions the Lennard–Jones potential was applied to selected $C\alpha$ atom pairs.

Another difference with TAMD is that in MSES the adiabatic separation between x and z variables is not necessarily assumed, although in practice the values of μ_α and $\bar{\beta}^{-1}$ can be set to achieve maximal sampling efficiency. Rather, it is suggested to use equation (13) in a Hamiltonian replica exchange scheme (Fukunishi et al., 2002), by introducing a set of replicas of the extended (x, z) system, each characterized by a different value of the coupling constant κ. Harvesting data from the replica with $\kappa = 0$ enable the elimination of the biasing coupling potential, recovering a canonical probability distribution for the x and thus the possibility to reconstruct free-energy surfaces from any other CV by simply monitoring their probability distribution. In practice, 20 and 12 replicas were used to study the sortase disorder to order transition and the Barnstar–Barnase complex, respectively. Note that in principle any kind of biasing potential can be used. However, using stronger bias will lead to a larger number of replicas for unbiasing to $\kappa = 0$.

3. Methods to Determine Reactive Paths

In many cases, rather than reconstructing large portions of the free-energy space, one is interested in directly finding paths connecting different minima in this landscape. This different perspective was first introduced by Pratt (1986) using a "chain-of-states" linking two minima, and successively developed by Chandler and collaborators in the transition path sampling (TPS) technique (Bolhuis et al., 2002), where the basic idea is to reconstruct the ensemble of transition paths directly from pieces of dynamical trajectories joining the two minima. These pioneer methods and many others such as those of Elber and collaborators [see Májek et al. (2008) for a comprehensive review], the Nudged Elastic Band (Jonsson et al., 1998) or the zero temperature string method (E et al., 2002, 2007b), were developed to work in the full Cartesian space, thus searching for transition paths in the potential energy space. This approach becomes difficult when the dimensionality of the system is large (as it is often encountered in relevant biophysical problems), since in this case the potential energy space is very noisy (i.e., rugged on a scale smaller than the thermal energy), and as a consequence the reactive paths are highly twisted. To get around this hurdle, the string method in collective variables algorithm was introduced in Maragliano et al. (2006) to find reaction paths in free-energy space, by identifying them as Minimum Free Energy Paths (MFEPs). The main advantage of this approach with respect to those working in potential energy space is that the free energy is usually much smoother. Moreover, with respect to techniques for sampling and reconstructing globally the free-energy space, the string method permits the use of a larger number of collective variables.

The basic idea of the string method is to represent the transition path between two free-energy minima in the space of a set of CVs $\theta(x) = (\theta_1(x), \ldots, \theta_n(x))$ as a curve (the string), discretized in a collection of points called images. The positions of these

images are evolved until convergence using information harvested from MD simulations of replicas of the full system, constrained or restrained at the CV values that define the images. To avoid that all images collapse in the two minima, a curve parametrization is enforced during the evolution. The path obtained at convergence depends on the version of the algorithm used. The original version in Maragliano et al. (2006) converges to the MFEP, while other successive versions yield a slightly different path (more details will be given below). A detailed comparison of different string method algorithms and of the paths they converge to is presented in Maragliano et al. (2014) [for a comprehensive review, see also E and Vanden-Eijnden (2010)]. In the next paragraph, we briefly review the definition of these paths and their relevance in the study of reactive events, while in Section 3.1.2, we describe the main features of the on-the-fly version of the string method in CVs.

3.1. The On-The-Fly String Method in Collective Variables

3.1.1. The Minimum Free-Energy Path and its Relevance

Given two minima A and B of the free energy $F(z)$, the MFEP was defined in Maragliano et al. (2006) as the curve that connects A and B and to which the vector $M(z) \cdot \bigtriangledown F(z)$ is everywhere tangent. Hence, it is the curve satisfying the equation

$$[M(z) \cdot \nabla F(z)]^{\perp} = 0 \qquad (14)$$

where the superscript $\perp$ indicates projection in the direction perpendicular to the curve itself. In equation (14), $\bigtriangledown F(z)$ is the gradient of the free energy (the mean force), and $M(z)$ is a metric tensor which accounts for the curvilinear nature of the CVs, defined component-wise by the conditional average

$$M_{\alpha\beta}(z) = \left\langle \sum_{i=1}^{N} \frac{1}{m_i} \frac{\partial \theta_\alpha(x)}{\partial x_i} \frac{\partial \theta_\beta(x)}{\partial x_i} \right\rangle_{\theta(x)=z} \qquad (15)$$

or explicitly

$$M_{\alpha\beta}(z) = e^{F(z)/k_B T} Z^{-1} \int dx \sum_{i=1}^{N} \frac{1}{m_i} \frac{\partial \theta_\alpha(x)}{\partial x_i} \frac{\partial \theta_\beta(x)}{\partial x_i}$$
$$e^{-V(x)/k_B T} \delta(\theta(x) - z). \qquad (16)$$

As already anticipated in Sections 1 and 2.1.1, the mean force can be obtained using the *blue moon* ensemble technique (Ciccotti et al., 2005) or by harmonic restraints [see equation (7)]. The same is valid for the tensor $M(z)$.

The MFEP equation (14) is the curve obtained by the original string method algorithm in CVs described in Maragliano et al. (2006). The importance of the MFEP in order to understand the mechanism of the transitions between a reactant state A and the product state B is established within the framework of Transition Path Theory (Vanden-Eijnden, 2007; E and Vanden-Eijnden, 2010). Indeed, using TPT, it is possible to demonstrate that in appropriate regimes, and when the reaction is described in the

space of a set of collective variables, the MFEP is the most probable transition path between A and B.

Let us recall here briefly this result, while for more details we refer any interested reader to Vanden-Eijnden (2007) and E and Vanden-Eijnden (2010). From TPT, the key quantity to describe the statistical properties of the reactive trajectories is the committor $q(x, v)$, a function of the positions and velocities of the physical system, which gives the probability that the trajectory initiated at (x, v) will reach the product B before the reactant A. Via the committor function it is possible to obtain analytic expressions for the probability density of reactive trajectories, their probability current, and the rate of the reaction, and for this reason, it is the *optimal* reaction coordinate. The committor satisfies a Fokker–Planck equation, which is too complicated to be solved by standard numerical methods when the dimensionality of the system is large. To get around this difficulty, in the string method it is assumed that $q(x, v)$ can be approximated by a function that depends only on the positions x, via a set of collective variables $\theta(x) = (\theta_1(x), \ldots, \theta_n(x))$, i.e., $q(x, v) \approx Q(\theta(x))$.

As shown in Maragliano et al. (2006), $Q(z)$ satisfies a much simpler Fokker–Planck equation, the equation for the committor function associated to the following overdamped equation for the collective variables ($\alpha = 1, \ldots, n$)

$$\gamma_z \frac{dz_\alpha}{dt} = \sum_{\beta=1}^{n} \left(-M_{\alpha\beta}(z) \frac{\partial F(z)}{\partial z_\beta} + k_B T \frac{\partial}{\partial z_\beta} M_{\alpha\beta}(z) \right)$$
$$+ \sqrt{2k_B T} \sum_{\beta} M_{\alpha\beta}^{1/2}(z) \eta_\beta \qquad (17)$$

where $M_{\alpha\beta}$ is the tensor defined in equation (16) and γ_z is an artificial friction coefficient. Notably, this reasoning does not require to specify a value for γ_z. The discussion above implies that we can use equation (17) to study the mechanism of the reaction from A to B. Indeed, if the temperature is low compared to the relevant free-energy barriers, the trajectories solution of equation (17) concentrate around a curve in collective variable space which (i) connects the two minima A and B of the free-energy surface via a saddle point and (ii) is everywhere tangent to the vector field $M(z)\bigtriangledown F(z)$ [consider the limit $T \to 0$ of equation (17)]. In other words, the reactive trajectories in CV space lie along the MFEP defined as in equation (14). Hence, the string method is a practical procedure to identify globally these transition paths by evolving their representative curve in collective variable space.

It is also possible to include in the definition of the reaction path equation (14) the extra term in the drift from equation (17), which contains the divergence of $M(z)$, $k_B T \Sigma_\beta \partial M_{\alpha,\beta}/\partial z_\beta$. In this case, the relevant path becomes the curve solution of

$$0 = [-M(z)\nabla F(z) + k_B T \nabla \cdot M(z)]^{\perp}. \qquad (18)$$

where $\bigtriangledown \cdot M(z)$ is compact notation for the divergence of $M(z)$. The path solution of equation (18) is the one identified by a particular implementation of the on-the-fly string method (Maragliano and Vanden-Eijnden, 2007) (see below), and also

by other successive variants of the method (Pan et al., 2008; Johnson and Hummer, 2012). Note that the solution of equation (18) does not go through the same critical points as the MFEP, i.e., unlike the MFEP it cannot be used to identify the saddle points on the free-energy landscape. However, because the extra term in equation (18) is proportional to $k_B T$ it should not have much influence, and the differences between the MFEP solution of equation (14) and the curve solution of equation (18) are expected to be small and difficult to distinguish from numerical fluctuations. This was indeed the result observed for the specific cases of solvated (Maragliano and Vanden-Eijnden, 2007) and gas-phase (Maragliano et al., 2014) alanine dipeptide, using dihedral angles as collective variables.

3.1.2. The on-the-Fly String Method Algorithm

In the original formulation of the string method in CVs (Maragliano et al., 2006), one MD replica per image is used, and the force driving the evolution of the images is the product of the mean force (the gradient of free energy) and the tensor $M(z)$, both of which can be expressed in terms of expectations conditional on $\theta(x) = z$. This procedure requires that once the new values of the images are obtained, the MD system is re-initialized to have that at each image the $\theta(x)$ are close to the new z.

Successively (Maragliano and Vanden-Eijnden, 2007), a different version of the string method, called on-the-fly, was developed, which exploits the same ideas at the basis of TAMD that were described in Section 2.1.1. The basic idea of this other algorithm is to evolve the string images concurrently with the attached replicas of the MD system, which provide on-the-fly the data needed to the evolution of the images. This is attained by introducing an extended system comprising the physical and collective variables, which are evolved concurrently. Similarly to TAMD, the motion of the string images is dampened with respect to the MD replicas, so that, by relying on the same averaging theorems for systems with multiple time scales that are at the basis of TAMD, one can bypass the computation of the string driving force via averaging. This approach also eliminates the need for re-initializing the MD systems after every string update, thus providing a simpler, more stable and more efficient algorithm.

There are two versions of the on-the-fly string method algorithm, which differ by the number of independent MD replicas attached to each image. Indeed, it was shown in Maragliano and Vanden-Eijnden (2007) that when two replicas per image are used, the algorithm converges to the MFEP equation (14), while when only one replica is used the algorithm converges to the curve defined by equation (18). We illustrate here for simplicity the one-replica version, and we address the reader to Maragliano and Vanden-Eijnden (2007) for more details about the one with two replicas and the method in general.

It is useful to present the method in the continuous formulation, i.e., by considering the moving string as a time-dependent curve parametrized as $\{z(s, t) : s \in [0, 1]\}$. Such curve is evolved concurrently with a one-parameter family of replicas of the MD system $\{x(s, t) : s \in [0, 1]\}$. Written component-wise for $\alpha = 1, \ldots, n$ and $i = 1, \ldots, N$, the equations for the concurrent

dynamics are

$$
\begin{cases}
\gamma_z \dot{z}_\alpha(s, t) = \sum_{\beta=1}^{n} \tilde{M}_{\alpha\beta}(x(s, t))\kappa(\theta_\beta(x(s, t) - z_\beta(s, t))) \\
\qquad\qquad + \Lambda(s, t)\, z'_\alpha(s, t) \\
m_i \ddot{x}_i(s, t) = -\dfrac{\partial V(x(s, t))}{\partial x_i} - \kappa \sum_{\beta=1}^{n} (\theta_\beta(x(s, t) \\
\qquad\qquad - z_\beta(s, t)))\dfrac{\partial \theta_\beta(x(s, t))}{\partial x_i} + \text{thermostat at } \beta^{-1}
\end{cases}
$$

$$(19)$$

where γ_z is an effective friction term acting on the collective variables and

$$
\tilde{M}_{\alpha\beta}(x) = \sum_{i=1}^{N} \frac{1}{m_i} \frac{\partial \theta_\alpha(x)}{\partial x_i} \frac{\partial \theta_\beta(x)}{\partial x_i}. \tag{20}
$$

In equation (19), $\Lambda(s, t)z'_\alpha(s, t)$, with $z'_\alpha = \partial z_\alpha/\partial s$, is a Lagrange multiplier term used to enforce the particular parametrization chosen for the curve (for example, normalized arc length). Note that, at variance with equation (3), in equation (19), we use first order dynamics for the z variables. This is done for simplicity and because the evolution of $z(s, t)$ in equation (19) has a different purpose than that of $z(t)$ in equation (3), i.e., it is not an exploratory dynamics but rather a steepest descent along the driving force. Analogously to equation (3), however, the system equation (19) describes the evolution of the extended system (x, z) under the effect of the potential energy equation (4). Here, the friction coefficient γ_z plays the role of the masses μ_α in equation (3).

Following the same reasoning illustrated in Section 2.1.1, in Maragliano and Vanden-Eijnden (2007), it was shown that by simulating the system equation (19) with κ and γ_z sufficiently large, the $z(s, t)$ evolve according to an effective equation obtained by averaging the equations for $z(s, t)$ in equation (19) with respect to the probability density function for the equation for $x(s, t)$ in equation (19) at $z(s, t) = z(s)$ fixed. In this case, the effective equation is

$$
\gamma_z \dot{z}_\alpha(s, t) = \sum_{\beta=1}^{n} \left(-M_{\alpha\beta}(z(s, t))\frac{\partial F(z(s, t))}{\partial z_\beta} \right.
$$
$$
\left. + k_B T \frac{\partial}{\partial z_\beta} M_{\alpha\beta}(z(s, t)) \right) + \lambda(s, t)z'_\alpha(s, t). \tag{21}
$$

It can be easily seen that the steady state solution of equation (21) is precisely the curve equation (18). Hence, if we evolve a string $\{z(s) : s \in [0, 1]\}$ by simulating equation (19) using large but finite values of κ and γ_z, this will converge to the curve defined by equation (18).

As a final comment, note that, at variance with equation (3), the system equation (19) does not include a thermostat acting on the $z(s, t)$ variables. This generates a steepest descent dynamics for $z(s, t)$ on the free-energy surface, which is appropriate when the surface is smooth. In cases where the free-energy landscape is rugged, albeit on a scale smaller than the barrier defining the transition state, a thermostat term can be added to the $z(s, t)$ equation in order to let the string overcome local free-energy asperities. This idea was explored in Vanden-Eijnden and Venturoli (2009) and Stober and Abrams (2012). However, it is important to

note that in practical implementations of the system equation (19) the timescale separation between x and z is usually approximate, which means that $z(s, t)$ satisfy the limiting equation (21) only up to a residual error. This error produces a noise on the evolution of $z(s, t)$, which is comparable to a low-temperature thermostat. Basing on our experience, we believe that it is possible to impose a separation of timescales such that $z(s, t)$ obey equation (21) to a satisfactory accuracy, but with enough residual noise to navigate over the local ruggedness of $F(z)$.

3.1.3. Biophysical Applications of the OTF String Method

So far, the OTF string method has been successfully utilized to study the ligand-induced conformational change of adenylate kinase (Matsunaga et al., 2012), the formation of a stalk between two apposed membrane bilayers (Müller et al., 2012), and the normal-to-amyloidogenic isomerization of β2-microglobulin (Stober and Abrams, 2012). Notably, in Stober and Abrams (2012), the authors introduce a clever implementation of the method by simulating a single MD system comprising all the string images embedded in a single, large solvent box. In Ovchinnikov et al. (2011), an intermediate variant between the original and on-the-fly string method was employed to study the conformational transition of myosin VI. Finally, in Zinovjev et al. (2013) the on-the-fly string method was used in combination with a Quantum Mechanics/Molecular Mechanics approach to study the reaction catalyzed by guanidinoacetate methyltransferase.

3.2. The Onsager–Machlup Multiscale Enhanced Sampling Method

An enticing idea to sample transition paths has been developed in Fujisaki et al. (2013) by combining the MSES technique presented in Section 2.3 with an Onsager–Machlup (OM) action approach (MSES-OM). In MSES-OM, an extended system is again considered, comprising the physical and CV variables (x, z), and a global OM action is defined as the sum of the OM actions for the x and z variables and an interaction term, $\mathcal{S}(x, z) = S(x) + S^{CV}(z) + \kappa S^{\text{int}}(x, z)$. The basic idea is that the CG variables z move rather freely in path space, dragging with them the x variables. Similarly to MSES, a replica exchange framework is introduced by using many replicas of the system (x, z), each one with a different values of k. The canonical OM path weight $\exp(-S[(x)])$ is then recovered from the replica with $\kappa = 0$. The MSES-OM method has been extensively investigated in

Fujisaki et al. (2013) by applying it to study the unlooping kinetics of a model polymer, revealing a substantial gain in computational time with respect to conventional OM method.

The advantages in using action methods are that they provide dynamical information on the transition studied, and from a numerical standpoint that the time step for the action discretization can be taken much larger than the typical MD step ($\approx$ 1fs). On the other hand, known limitations of the approach are that one is required to know beforehand the total time of the transition, and numerical difficulties due to the use of high order derivatives of the potential (Vanden-Eijnden and Heymann, 2008).

4. Concluding Remarks

We have reviewed a few enhanced sampling methods that allow to accelerate MD simulations and to study rare reactive events. The common feature of these methods is that they rely on extending the phase space of the physical system under study by adding a set of extra variables also considered as dynamical ones, linked to the original system via CV functions of the physical coordinates. This setting allows efficient exploration of the CV space, i.e., of the free-energy landscape of the original system with respect to the CVs. The various methods achieve this exploration via different means. In TAMD, the CVs are evolved at a higher temperature than the physical variables, so that they can efficiently cross free-energy barriers higher than the physiological temperature. In LogMFD, a logarithmic transformation of the free energy is employed together with the conserved quantity from deterministic dynamics to reconstruct directly the free energy during CV evolution. In MSES, a model potential is introduced for the CVs also, and different replicas of the extended system are used in a replica exchange framework, each with a different value of the coupling parameter between the CVs and the original system. The statistical properties of the unbiased physical system are then reconstructed from the replica with zero coupling. We have also reviewed path sampling methods based on the same extended phase-space ideas, namely as the on-the-fly string method and the Onsager–Machlup action MSES.

Acknowledgments

References

Abrams, C., and Vanden-Eijnden, E. (2010). Large-scale conformational sampling of proteins using temperature-accelerated molecular dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4961–4966. doi:10.1073/pnas.0914540107

Abrams, C., and Vanden-Eijnden, E. (2012). On-the-fly free energy parameterization via temperature accelerated molecular dynamics. *Chem. Phys. Lett.* 547, 114–119. doi:10.1016/j.cplett.2012.07.064

Abrams, J., and Tuckerman, M. (2008). Efficient and direct generation of multidimensional free energy surfaces via adiabatic dynamics without coordinate transformations. *J. Phys. Chem. B* 112, 15742–15757. doi:10.1021/jp805039u

Babin, V., Roland, C., Darden, T., and Sagui, C. (2006). The free energy landscape of small peptides as obtained from metadynamics with umbrella sampling corrections. *J. Chem. Phys.* 125, 204909. doi:10.1063/1.2393236

Barducci, A., Bussi, G., and Parrinello, M. (2008). Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* 100, 020603. doi:10.1103/PhysRevLett.100.020603

Bernardi, R. C., Melo, M. C., and 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

Bitetti-Putzer, R., Yang, W., and Karplus, M. (2003). Generalized ensembles serve to improve the convergence of free energy simulations. *Chem. Phys. Lett.* 377, 633–641. doi:10.1016/S0009-2614(03)01057-1

Blöchl, P. E., and Parrinello, M. (1992). Adiabaticity in first-principles molecular dynamics. *Phys. Rev. B Condens. Matter* 45, 9413–9416. doi:10.1103/PhysRevB.45.9413

Bolhuis, P. G., Chandler, D., Dellago, C., and Geissler, P. L. (2002). Transition path sampling: throwing ropes over Rough Mountain passes, in the dark. *Annu. Rev. Phys. Chem.* 53, 291–318. doi:10.1146/annurev.physchem.53.082301.113146

Bucci, A., and Abrams, C. F. (2014). Oxygen pathways and allostery in monomeric sarcosine oxidase via single-sweep free-energy reconstruction. *J. Chem. Theory Comput.* 10, 2668–2676. doi:10.1021/ct500088z

Carter, E. A., Ciccotti, G., Hynes, J. T., and Kapral, R. (1989). Constrained reaction coordinate dynamics for the simulation of rare events. *Chem. Phys. Lett.* 156, 472–477. doi:10.1016/S0009-2614(89)87314-2

Ciccotti, G., Kapral, R., and Vanden-Eijnden, E. (2005). Blue moon sampling, vectorial reaction coordinates, and unbiased constrained dynamics. *Chemphyschem* 6, 1809–1814. doi:10.1002/cphc.200400669

Cortes-Ciriano, I., Bouvier, G., Nilges, M., Maragliano, L., and Malliavin, T. E. (2015). Temperature accelerated molecular dynamics with soft-ratcheting criterion orients enhanced sampling by low-resolution information. *J. Chem. Theory Comput.* 11, 3446–3454. doi:10.1021/acs.jctc.5b00153

Darve, E., and Pohorille, A. (2001). Calculating free energies using average force. *J. Chem. Phys.* 115, 9169–9183. doi:10.1063/1.1410978

Dror, R. O., Mildorf, T. J., Hilger, D., Manglik, A., Borhani, D. W., Arlow, D. H., et al. (2015). Structural basis for nucleotide exchange in heterotrimeric G proteins. *Science* 348, 1361–1365. doi:10.1126/science.aaa5264

E, W., Engquist, B., Li, X., Ren, W., and Vanden-Eijnden, E. (2007a). Heterogeneous multiscale methods: a review. *Commun. Comput. Phys.* 2, 367–450. doi:10.1.1.225.9038

E, W., Ren, W., and Vanden-Eijnden, E. (2007b). Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *J. Chem. Phys.* 126, 164103. doi:10.1063/1.2720838

E, W., Ren, W., and Vanden-Eijnden, E. (2002). String method for the study of rare events. *Phys. Rev. B* 66, 052301. doi:10.1103/PhysRevB.66.052301

E, W., and Vanden-Eijnden, E. (2010). Transition-path theory and path-finding algorithms for the study of rare events. *Annu. Rev. Phys. Chem.* 61, 391–420. doi:10.1146/annurev.physchem.040808.090412

Freddolino, P. L., Harrison, C. B., Liu, Y., and Schulten, K. (2010). Challenges in protein-folding simulations. *Nat. Phys.* 6, 751–758. doi:10.1038/nphys1713

Fujisaki, H., Shiga, M., Moritsugu, K., and Kidera, A. (2013). Multiscale enhanced path sampling based on the Onsager-Machlup action: application to a model polymer. *J. Chem. Phys.* 139, 054117. doi:10.1063/1.4817209

Fukunishi, H., Watanabe, O., and 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

Hamelberg, D., Mongan, J., and McCammon, A. J. (2004). Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* 120, 11919–11929. doi:10.1063/1.1755656

Hoover, W. (1985). Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* 31, 1695–1697. doi:10.1103/PhysRevA.31.1695

Iannuzzi, M., Laio, A., and Parrinello, M. (2003). Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics. *Phys. Rev. Lett.* 90, 238302. doi:10.1103/PhysRevLett.90.238302

Johnson, M. E., and Hummer, G. (2012). Characterization of a dynamic string method for the construction of transition pathways in molecular reactions. *J. Phys. Chem. B* 116, 8573–8583. doi:10.1021/jp212611k

Jonsson, H., Mills, G., and Jacobsen, K. W. (1998). "Nudged elastic band method for finding minimum energy paths of transitions," in *Classical and Quantum Dynamics in Condensed Phase Simulations*, eds B. J. Berne, G. Ciccotti, and D. F. Coker (Singapore: World Scientific), 385–404.

Kong, X., and Brooks, C. L. III (1996). λ-dynamics: a new approach to free energy calculations. *J. Chem. Phys.* 105, 2414–2423. doi:10.1063/1.472109

Laio, A., and Parrinello, M. (2002). Escaping free energy minima. *Proc. Natl. Acad. Sci. U.S.A.* 99, 12562–12566. doi:10.1073/pnas.202427399

Laio, A., Rodriguez-Fortea, A., Gervasio, F., Ceccarelli, M., and Parrinello, M. (2005). Assessing the accuracy of metadynamics. *J. Phys. Chem. B* 109, 6714. doi:10.1021/jp045424k

Lapelosa, M., and Abrams, C. (2013). A computational study of water and CO migration sites and channels inside myoglobin. *J. Chem. Theory Comput.* 9, 1265–1271. doi:10.1021/ct300862j

Lindorff-Larsen, K., Piana, S., Dror, R. O., and Shaw, D. E. (2011). How fast-folding proteins fold. *Science* 334, 517–520. doi:10.1126/science.1208351

Liu, Y., and Tuckerman, M. E. (2000). Generalized Gaussian moment thermostatting: a new continuous dynamical approach to the canonical ensemble. *J. Chem. Phys.* 112, 1685–1700. doi:10.1063/1.480769

Lucid, J., Meloni, S., MacKernan, D., Spohr, E., and Ciccotti, G. (2012). Probing the structures of hydrated nafion in different morphologies using temperature accelerated molecular dynamics simulations. *J. Phys. Chem. C* 117, 774–782. doi:10.1021/jp309038n

Májek, P., Elber, R., and Weinstein, H. (2008). "Pathways of conformational transitions in proteins," in *Coarse-Graining of Condensed Phase and Biomolecular Systems*, ed. G. A. Voth (Boca Raton, FL: CRC Press), 185.

Maragliano, L., Cottone, G., Ciccotti, G., and Vanden-Eijnden, E. (2010). Mapping the network of pathways of CO diffusion in myoglobin. *J. Am. Chem. Soc.* 132, 1010–1017. doi:10.1021/ja905671x

Maragliano, L., Fischer, A., Vanden-Eijnden, E., and Ciccotti, G. (2006). String method in collective variables: minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.* 125, 024106. doi:10.1063/1.2212942

Maragliano, L., Roux, B., and Vanden-Eijnden, E. (2014). Comparison between mean forces and swarms-of-trajectories string methods. *J. Chem. Theory Comput.* 10, 524–533. doi:10.1021/ct400606c

Maragliano, L., and Vanden-Eijnden, E. (2006). A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chem. Phys. Lett.* 426, 168–175. doi:10.1016/j.cplett.2006.05.062

Maragliano, L., and Vanden-Eijnden, E. (2007). On-the-fly string method for minimum free energy paths calculation. *Chem. Phys. Lett.* 446, 182–190. doi:10.1016/j.cplett.2007.08.017

Maragliano, L., and Vanden-Eijnden, E. (2008). Single-sweep methods for free energy calculations. *J. Chem. Phys.* 128, 184110. doi:10.1063/1.2907241

Martyna, G. J., Klein, M. L., and Tuckerman, M. (1992). Nosé-hoover chains: the canonical ensemble via continuous dynamics. *J. Chem. Phys.* 97, 2635–2643. doi:10.1063/1.463940

Matsunaga, Y., Fujisaki, H., Terada, T., Furuta, T., Moritsugu, K., and Kidera, A. (2012). Minimum free energy path of ligand-induced transition in adenylate kinase. *PLoS Comput. Biol.* 8:e1002555. doi:10.1371/journal.pcbi.1002555

Mitsutake, A., Sugita, Y., and Okamoto, Y. (2001). Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers* 60, 96–123. doi:10.1002/1097-0282(2001)60:2<96::AID-BIP1007>3.0.CO;2-F

Mohammad Hosseini Naveh, Z., Malliavin, T., Maragliano, L., Cottone, G., and Ciccotti, G. (2014). Conformational changes in acetylcholine binding protein investigated by temperature accelerated molecular dynamics. *PLoS ONE* 9:e88555. doi:10.1371/journal.pone.0088555

Monteferrante, M., Bonella, S., Meloni, S., Eijnden, E. V., and Ciccotti, G. (2008). Calculations of free energy barriers for local mechanisms of hydrogen diffusion in alanates. *Sci. Model. Simul.* 15, 187–206. doi:10.1007/978-1-4020-9741-6_13

Morishita, T. (2010). From Nosé-Hoover chain to Nosé-Hoover network: design of non-Hamiltonian equations of motion for molecular-dynamics with multiple thermostats. *Mol. Phys.* 108, 1337–1347. doi:10.1080/00268971003689923

Morishita, T., Itoh, S. G., Okumura, H., and Mikami, M. (2012). Free-energy calculation via mean-force dynamics using a logarithmic energy landscape. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 85, 066702–066707. doi:10.1103/PhysRevE.85.066702

Morishita, T., Itoh, S. G., Okumura, H., and Mikami, M. (2013). On-the-fly reconstruction of free-energy profiles using logarithmic mean-force dynamics. *J. Comput. Chem.* 34, 1375–1384. doi:10.1002/jcc.23267

Moritsugu, K., Terada, T., and Kidera, A. (2010). Scalable free energy calculation of proteins via multiscale essential sampling. *J. Chem. Phys.* 133, 224105. doi:10.1063/1.3510519

Moritsugu, K., Terada, T., and Kidera, A. (2012). Disorder-to-order transition of an intrinsically disordered region of sortase revealed by multiscale enhanced sampling. *J. Am. Chem. Soc.* 134, 7094–7101. doi:10.1021/ja3008402

Moritsugu, K., Terada, T., and Kidera, A. (2014a). Energy landscape of all-atom protein-protein interactions revealed by multiscale enhanced sampling. *PLoS Comput. Biol.* 10:e1003901. doi:10.1371/journal.pcbi.1003901

Moritsugu, K., Terada, T., and Kidera, A. (2014b). Multiscale enhanced sampling driven by multiple coarse-grained models. *Chem. Phys. Lett.* 61, 20–24. doi:10.1016/j.cplett.2014.10.009

Müller, M., Smirnova, Y. G., Marelli, G., Fuhrmans, M., and Shi, A.-C. (2012). Transition path from two apposed membranes to a stalk obtained by a combination of particle simulations and string method. *Phys. Rev. Lett.* 108, 228103. doi:10.1103/PhysRevLett.108.228103

Nakamura, M., Obata, M., Morishita, T., and Oda, T. (2014). An ab initio approach to free-energy reconstruction using logarithmic mean force dynamics. *J. Chem. Phys.* 140, 184110. doi:10.1063/1.4874654

Nosé, S. (1984). A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* 52, 255–268. doi:10.1080/00268978400101201

Nygaard, R., Zou, Y., Dror, R., Mildorf, T., Arlow, D., Manglik, A., et al. (2013). The dynamic process of β(2)-adrenergic receptor activation. *Cell* 152, 532–542. doi:10.1016/j.cell.2013.01.008

Ovchinnikov, V., Karplus, M., and Vanden-Eijnden, E. (2011). Free energy of conformational transition paths in biomolecules: the string method and its application to myosin VI. *J. Chem. Phys.* 134, 085103. doi:10.1063/1.3544209

Pan, A. C., Sezer, D., and Roux, B. (2008). Finding transition pathways using the string method with swarms of trajectories. *J. Phys. Chem. B* 112, 3432–3440. doi:10.1021/jp0777059

Papanicolaou, G. C. (1977). "Introduction to the asymptotic analysis of stochastic equations," in *Lectures in Applied Mathematics*, Vol. 16, ed. R. DiPrima (Providence, RI: American Mathematical Society), 109–147.

Pavliotis, G., and Stuart, A. (2008). *Multiscale Methods: Averaging and Homogenization*. New York, NY: Springer.

Perilla, J. R., Goh, B. C., Cassidy, C. K., Liu, B., Bernardi, R. C., Rudack, T., et al. (2015). Molecular dynamics simulations of large macromolecular complexes. *Curr. Opin. Struct. Biol.* 31, 64–74. doi:10.1016/j.sbi.2015.03.007

Pratt, L. R. (1986). A statistical method for identifying transition states in high dimensional problems. *J. Chem. Phys.* 85, 5045–5048. doi:10.1063/1.451695

Rosso, L., Minary, P., Zhu, Z., and Tuckerman, M. (2002). On the use of the adiabatic molecular dynamics technique in the calculation of free energy profiles. *J. Chem. Phys.* 116, 4389–4403. doi:10.1063/1.1448491

Roux, B. (1995). The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* 91, 275–282. doi:10.1016/0010-4655(95)00053-I

Selwa, E., Huynh, T., Ciccotti, G., Maragliano, L., and Malliavin, T. E. (2014). Temperature-accelerated molecular dynamics gives insights into globular conformations sampled in the free state of the ac catalytic domain. *Proteins* 107, 2483–2496. doi:10.1002/prot.24612

Smiatek, J., and Heuer, A. (2011). Calculation of free energy landscapes: a histogram reweighted metadynamics approach. *J. Comput. Chem.* 32, 2084–2096. doi:10.1002/jcc.21790

Stober, S. T., and Abrams, C. F. (2012). Energetics and mechanism of the normal-to-amyloidogenic isomerization of β2-microglobulin: on-the-fly string method calculations. *J. Phys. Chem. B* 116, 9371–9375. doi:10.1021/jp304805v

Tirion, M. M. (1996). Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 77, 1905–1908. doi:10.1103/PhysRevLett.77.1905

Torrie, G. M., and Valleau, J. P. (1974). Monte Carlo free energy estimates using non-Boltzmann sampling: application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.* 28, 578–581. doi:10.1016/0009-2614(74)80109-0

Trigila, G. (2007). Master's thesis, Department of Physics, University of Rome La Sapienza.

Vanden-Eijnden, E. (2003). Numerical techniques for multi-scale dynamical systems with stochastic effects. *Commun. Math. Sci.* 1, 385–391. doi:10.4310/CMS.2003.v1.n2.a11

Vanden-Eijnden, E. (2007). "Transition path theory," in *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1*, eds M. Ferrario, G. Ciccotti, and K. Binder (Berlin: Springer), 453–493.

Vanden-Eijnden, E., and Heymann, M. (2008). The geometric minimum action method for computing minimum energy paths. *J. Chem. Phys.* 128, 061103. doi:10.1063/1.2833040

Vanden-Eijnden, E., and Venturoli, M. (2009). Markovian milestoning with Voronoi tessellations. *J. Chem. Phys.* 130, 194103. doi:10.1063/1.3129843

VandeVondele, J., and Rothlisberger, U. (2002). Canonical adiabatic free energy sampling (cafes): a novel method for the exploration of free energy surfaces. *J. Phys. Chem. B* 106, 203–208. doi:10.1021/jp013346k

Vashisth, H., and Brooks, C. L. (2012). Conformational sampling of maltose-transporter components in Cartesian collective variables is governed by the low-frequency normal modes. *J. Phys. Chem. Lett.* 3, 3379–3384. doi:10.1021/jz301650q

Vashisth, H., Maragliano, L., and Abrams, C. (2012a). "DFG-flip" in the insulin receptor kinase is facilitated by a helical intermediate state of the activation loop. *Biophys. J.* 102, 1979–1987. doi:10.1016/j.bpj.2012.03.031

Vashisth, H., Skiniotis, G., and Brooks, C. L. III (2012b). Using enhanced sampling and structural restraints to refine atomic structures into low-resolution electron microscopy maps. *Structure* 20, 1453–1462. doi:10.1016/j.str.2012.08.007

Vashisth, H., Storaska, A., Neubig, R., and Brooks, C. (2013). Conformational dynamics of a regulator of G-protein signaling protein reveals a mechanism of allosteric inhibition by a small molecule. *ACS Chem. Biol.* 8, 2778–2784. doi:10.1021/cb400568g

Yamamori, Y., and Kitao, A. (2013). Mustar md: multi-scale sampling using temperature accelerated and replica exchange molecular dynamics. *J. Chem. Phys.* 139, 145105. doi:10.1063/1.4823743

Zinovjev, K., Ruiz-Pernía, J. J., and Tuñón, I. N. (2013). Toward an automatic determination of enzymatic reaction mechanisms and their activation free energies. *J. Chem. Theory Comput.* 9, 3740–3749. doi:10.1021/ct400153r