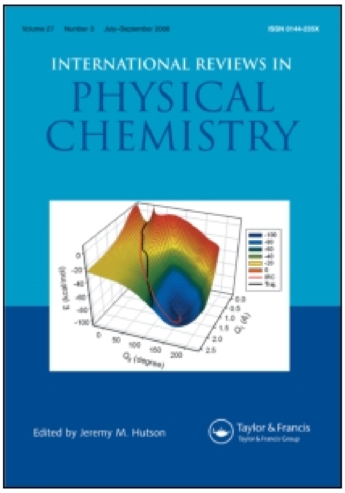

International Reviews in Physical Chemistry
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713724383

Thermodynamics and kinetics simulations of multi-time-scale processes for complex systems

Yi Qin Gao[a]; Lijiang Yang[a]; Yubo Fan[a]; Qiang Shao[a]
[a] Department of Chemistry, Texas A & M University, College Station, Texas 77843, USA

To cite this Article Gao, Yi Qin , Yang, Lijiang , Fan, Yubo and Shao, Qiang(2008) 'Thermodynamics and kinetics simulations of multi-time-scale processes for complex systems', International Reviews in Physical Chemistry, 27: 2, 201 — 227

To link to this Article: DOI: 10.1080/01442350801920334

URL: http://dx.doi.org/10.1080/01442350801920334

International Reviews in Physical Chemistry
Vol. 27, No. 2, April–June 2008, 201–227

REVIEW ARTICLE

Thermodynamics and kinetics simulations of multi-time-scale processes for complex systems

Yi Qin Gao*, Lijiang Yang, Yubo Fan and Qiang Shao

Department of Chemistry, Texas A & M University, College Station, Texas 77843, USA

(*Received 10 December 2007; final version received 14 January 2008*)

We discuss in this review paper the recent development of enhanced sampling methods for searching in energy and configuration space, as well as that for the reactive transition paths. These methods allow accelerated calculations of thermodynamics and kinetics. We summarize in this review the theoretical background and numerical implementation of a variety of methods. Examples of applications are given for each method.

Keywords: conformational change; enhanced sampling; molecular dynamics; Monte Carlo; protein folding; trajectory sampling

Contents

*Corresponding author. Email: yiqin@mail.chem.tamu.edu

ISSN 0144–235X print/ISSN 1366–591X online

DOI: 10.1080/01442350801920334
http://www.informaworld.com

1. Introduction

One of the greatest challenges of modern chemistry is the understanding of the multiple scale dynamics of complex systems, for example, those involved in solution chemistry and biological systems. Significant progress has been made over the past several decades on using molecular dynamics simulations (both with *ab initio* and empirical force fields) to understand these complex systems and their dynamics. Some of the methods focus on accelerating sampling over high energetic barriers. A variety of methods have been developed in this direction to enhance sampling over the phase or configuration space and thus allow fast calculations of thermodynamics properties. These methods include, but are not limited to, J-walking [1], adaptive umbrella sampling [2], replica exchange [3,4] (parallel tempering [5–8], multi-canonical simulations [9,10], metadynamics [11–13], conformational flooding [14], conformational space annealing [15], hyperdynamics [16,17], potential smoothing methods [18], Tsallis statistics [19], adaptive biasing force methods [20,21], and combinatorial usage of replica exchange and multi-canonical simulations [22,23]. In recent years, the Wang–Landau method [24–26] is becoming popular and a number of variants have been implemented, such as in statistical temperature sampling [27] and in the calculation of partition functions [28].

Many of the methods mentioned above are equivalent to biasing the system potentials, such as that suggested in umbrella sampling [29–31]. Recently, in a series of publications Hamelberg *et al.* [32–34] proposed and applied a method to alter the potential energy landscape, leading to an efficient sampling of the conformational space. In this method, sampling over the high energy range was increased at the cost of a largely reduced sampling over the states with lower energies. In an earlier publication, Barth *et al.* [35] pointed out the deficiency of the sampling of the low energy range by a number of methods, and the authors proposed a generalized variable temperature distribution method with applications to molecular dynamics simulations to achieve a uniform distribution of energy. The latter certainly is also achieved in multi-canonical simulations or through metadynamics simulation when the energy is used as the collective coordinate. More recently, we used an approach to accelerate molecular dynamics (MD) simulations in which a bias potential was added to control the enhanced sampling in a desired range of energy (e.g., near the energy barrier and in the entire energy range that is important at the given temperature) without under-sampling the low energy or over-sampling the high energy range, leading to a uniform distribution as a function of energy. The method was shown to be efficient in free energy simulations [36] as well as in protein folding [37]. It was also successfully applied in Monte Carlo simulations to study the coarse-grained models for secondary structures of polypeptides [38]. The temperature-based simulation method, such as replica exchange (or parallel tempering), as well as the modified Wang–Landau method by Zhang and Ma [28], can also be used to generate the desired uniform energy distribution, controlled through the sampling over a temperature range instead of an energy range. In a more recent paper, we took an approach that is based on a generalized non-Boltzmann distribution method. The method allows fast sampling of the configuration space and allows fast calculations of thermodynamic properties based on a fast calculation of the partition function [39].

Various other accelerated methods exist for molecular dynamics simulations, in particular to calculate the slow conformational changes of rather large systems, such as proteins which display a broad range of characteristic motions, ranging from the fast vibrations of covalent bonds to the slow large-scale folding transitions. The fast motions set the upper bound for the integration time step of MD simulation to fs, and, again, as a result, MD simulation can only explore typically nanosecond time-scales and its applicability is severely limited. A number of methods have been designed to improve MD's efficiency in studying the slow processes by treating the slow and fast degrees of freedom differently. In one class of methods, the fast motions are separated and treated differently from the slower ones, such as in the programs SHAKE [40] and its variant RATTLE [41], which allows a larger integration time step; or by using multiple time steps [42–46], the slow varying interactions are computed less frequently, such as in the RESPA method [43,44]. Alternately, to approach a longer simulation time, one can speed up slow events so that they take place fast enough to be accessed by a MD simulation. A large number of methods fall into this latter category, for example, the self-guided MD (SGMD) [47,48], in which a guiding force estimated from an average of the instantaneous forces over a period of normal MD simulations is introduced to enhance slow systematic motions. In a recent paper [49], Yang and Gao introduced an approximate molecular dynamics simulation method with the hope of extending the time-scale approachable by MD simulations through slowing down of the fast motions and speeding up the slow ones at the same time. One similarity between this method and the earlier SGMD is that in addition to the instantaneous forces and velocities, average forces and velocities obtained from trajectories of normal MD simulations for a non-equilibrium spontaneous process are introduced into the equations of motions to facilitate the computation of the slow protein conformational change. This method slows down the fast motions, allowing a larger integration step size than that of SGMD, in addition to the attempt to speed up the slow motions.

One of the common problems of the different methods mentioned above is that the dynamical as well as kinetic information of the original system is lost due to the change of the Hamiltonian. One of the most popular methods in calculating kinetics of complex systems using real trajectories is the transition path sampling method of Chandler and coworkers, which focuses on enhancing the sampling of transition paths [50–53]. This method generates an ensemble of trajectories connecting the reactant to the product using Monte Carlo (MC) procedures called shooting and shifting. Various forms of transition path sampling methods have been invented with the hope of further improving its efficiency. These efforts include the transition interface sampling (TIS) method [54,55] and the related partial path TIS (PPTIS) method [56]. On the other hand, in the application of TPS, several different approaches, for example, to initiate trajectories from the potential energy transition state [57], or from a high temperature [58], or targeted MD simulations [59], have been used to generate the initial trajectory, which is subsequently relaxed and used in the sampling for unbiased transition paths. The other methods that have been used to accelerate kinetic simulations include the weighted-ensemble [60] and milestoning methods [61].

In a recent paper [62], a combined approach was introduced to take advantages of both accelerated MD simulation and transition path sampling (shooting) methods. This method uses accelerated MD simulations over the configuration space to identify the active phase space (at a lower resolution thus a faster speed), which contains phase space points that are along transition paths of a given length in the real system following a Boltzmann distribution. Path sampling is then performed in these active portions of phase space using the original potential of the system ('high resolution' calculations) to obtain information such as reaction pathways and rate constants of the real system with higher transition barriers. This selective enhancement of sampling further increases the efficiency in obtaining the reaction pathways and kinetic data such as rate constants. Compared to the traditional transition path sampling method, there are three major differences: (1) the initial trajectories are easily generated, even for relatively large systems, such as proteins; (2) multiple independent initial trajectories are generated therefore the method avoids the entrapment of the trajectories in a particular pathway; (3) a pre-determined reaction coordinate is not required, and thus avoids the time-consuming and reaction coordinate dependent calculations of the reactive flux. This idea was further combined with Monte Carlo trajectory sampling which enhances the sampling of highly energetic and reactive configurations and phase space points as well as the sampling of trajectories initiated from these phase space points.

In the rest of this review, we focus mainly on the theoretical background and numerical implementation of the simulation methods that were recently developed in our laboratory and a few examples are given on the applications of these various methods. Two methods for enhanced sampling over energy and configurations are discussed in Section 2. The accelerated and approximate method for calculating large conformational changes is presented in Section 3 and the enhanced sampling methods in the trajectory space are summarized in Section 4. Possible future development and application of these various methods are discussed in Section 5.

2. Enhanced sampling in the energy and configuration space over high energy barriers

2.1. *Accelerated molecular dynamics simulations using biased potentials*

2.1.1. *Method*

We first introduce here an approach to enable efficient sampling over a chosen energy range(s) by making a simple transformation of potential energy surfaces,

$$\tilde{V}(\mathbf{r}) = V(\mathbf{r}) + f[V(\mathbf{r})], \tag{1}$$

In this method, a series of bias potentials (e.g., in the form of Gaussians) are added for several small energy ranges, with each additional one causing the energy distribution to be more even until a desired distribution is obtained:

$$f(V) = \sum_i a_i e^{-[(V-V_i)/\sigma_i]^2}, \tag{2}$$

where V_i is the energy of a state that is to be sampled with enhancement and a_i is a negative constant which defines how much the enhancement will be. The form of this function is therefore similar to that used in the metadynamics simulations [11]. A suitable bias potential is determined by an iterative procedure. Firstly, a short trajectory is obtained using the normal MD simulation at the desired temperature. The distribution function $P(V)$ as a function of the sampled potential energy is obtained, and $RT \ln P(V)$ is fitted to a Gaussian function. A Gaussian term $RT \ln P(V)$ is then added to the original potential. Similarly, a short normal MD simulation trajectory at a higher temperature (typically 400–600 K) is also obtained and analysed. The mostly sampled energy region at this temperature is also important and a negative bias potential should be introduced here to increase the sampling. Since we have obtained energy distributions at room temperature and a higher temperature, densities of states at room temperature $\rho_1(V)$ and that at a higher temperature $\rho_2(V)$ are fitted roughly to a function $\rho(V)$ using a polynomial. If V_1 is an energy at which the bias potential equals zero and V_2 is an energy at which the magnitude of the bias potential is the largest, $\Delta V = (V_2 - V_1) - RT \ln[\rho(V_1)/\rho(V_2)]$ is used as the initial estimation of the depth of the bias potential for the high energy range. Finally, the bias potential is added to the molecular system to obtain a short trial trajectory at room temperature. If the energy range is broadened to desirable high and low energy ranges, the bias potential is considered to be a suitable one and can be used in following accelerated MD simulations. Otherwise, the procedure described above will be repeated and new additional bias potentials will be generated and added to the system potential. After several rounds of this repeating procedure, a suitable bias potential should be determined. Although this strategy is not entirely quantitative, it is easy to implement and effective in practice.

A MD simulation with the scaled potential function $\tilde{V}(\boldsymbol{r})$ at equilibrium yields a distribution function over the configuration space [29],

$$\tilde{\rho}(\boldsymbol{r}) = e^{-\beta \tilde{V}(\boldsymbol{r})} / \tilde{Q} \tag{3}$$

where $\tilde{Q} = \int e^{-\beta \tilde{V}(\boldsymbol{r})} \mathrm{d}\boldsymbol{r}$. Since $\tilde{V}(\boldsymbol{r})$ is a function of $V(\boldsymbol{r})$, $\rho(\boldsymbol{r})$ can be recovered as:

$$\rho(r) = \frac{e^{-\beta V(\boldsymbol{r})}}{Q} = \frac{e^{-\beta \tilde{V}(\boldsymbol{r})} \times e^{-\beta[V(\boldsymbol{r}) - \tilde{V}(\boldsymbol{r})]}}{Q} = \frac{\tilde{\rho}(r) e^{-\beta[V(\boldsymbol{r}) - \tilde{V}(\boldsymbol{r})]} \tilde{Q}}{Q}. \tag{4}$$

Other thermodynamic properties, for example the average of any thermodynamic variable A, $\langle A \rangle$, can also be easily obtained:

$$\langle A \rangle = \int \rho(\boldsymbol{r}) A(\boldsymbol{r}) \, \mathrm{d}r = \frac{\tilde{Q}}{Q} \int \tilde{\rho}(r) e^{-\beta[V(r) - \tilde{V}(r)]} A(\boldsymbol{r}) \, \mathrm{d}r. \tag{5}$$

In particular, the potential of mean force (PMF) along a certain coordinate s, through the average over the rest of the coordinates $\boldsymbol{r}'(\boldsymbol{r} = (\boldsymbol{r}', \mathrm{s}))$, is

$$G(s) = -\frac{1}{\beta} \ln p(s) + \text{const.} \tag{6}$$

where $p(s)$, the probability of finding the coordinate s to be at its desired value, can be recovered easily from $\tilde{\rho}(\boldsymbol{r}', s)$, the probability of observing s with $\tilde{V}(\boldsymbol{r}', s)$:

$$p(s) = \int \rho(\boldsymbol{r}', s) \mathrm{d}\boldsymbol{r}' = \frac{\tilde{Q}}{Q} \int \tilde{\rho}(\boldsymbol{r}', s) e^{-\beta[V(\boldsymbol{r}', s) - \tilde{V}(\boldsymbol{r}', s)]} \mathrm{d}\boldsymbol{r}'. \tag{7}$$

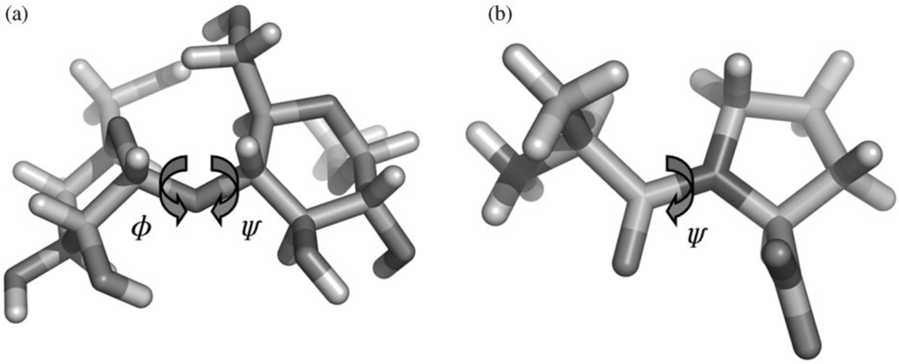

Figure 1. Structure of (a) methyl maltoside and (b) Ala-Pro dipeptide.

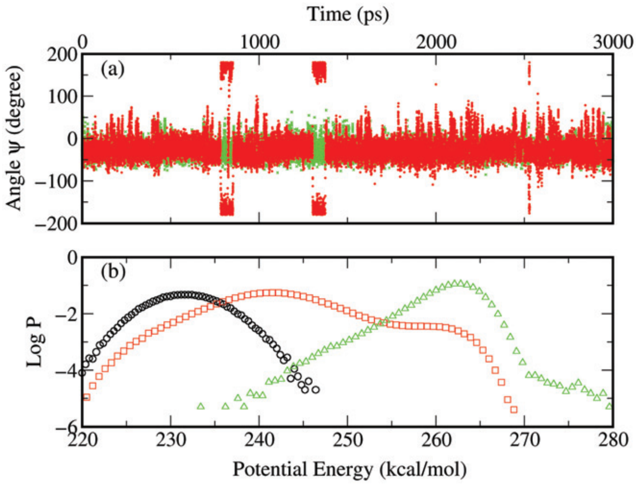

Figure 2. [Colour online] (a) The calculated trajectory of motions along the angle ψ for the maltose molecule (green: normal MD; red: accelerated MD). (b) The logarithm of the probability distribution function of energy obtained in normal MD (circles, 283 K; triangles, 400 K) and accelerated MD (squares) simulations.

2.1.2. *Studies of the isomerization of disaccharide and dipeptide*

The enhanced sampling method was first applied [36] to study the isomerization of small systems: a disaccharide, methyl maltoside and a dipeptide Ala-Pro (Figure 1). Samples of trajectories along the ψ dimension of the maltoside are given in Figure 2(a), where we compare the results obtained using the normal MD simulations and the accelerated MD simulations. It can be easily seen that the bias potential allows much wider distributions (data shown in red) along ψ and much more frequent transitions between different conformations. The results from the normal MD simulations are shown in green. The effect of the additional energy term $f(V)$ on the distribution over energy (the squares) is shown in Figure 2(b), which shows that a much wider range of energy is sampled in the accelerated MD simulation (the result from the normal MD simulation is shown in circles

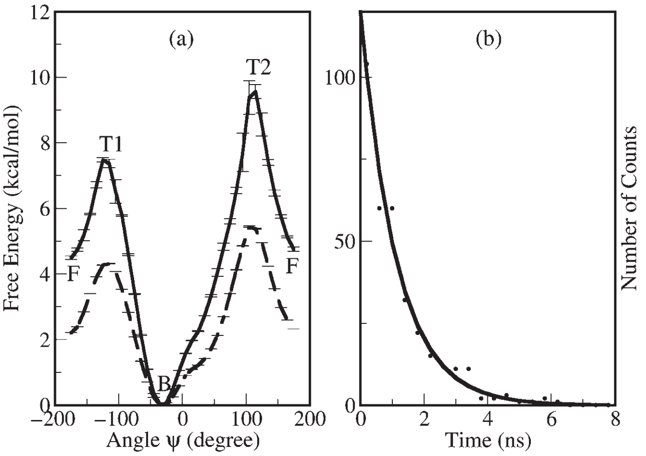

Figure 3. (a) The PMF plotted as a function of ψ of the maltose molecule (solid: normal MD; dashed: accelerated MD). (b) Observed occurrence of the life time of the state B. The largest count is at about 0.90 ns, which is corresponding to a $\tilde{k}$ of $1.1 \times 10^9\,\mathrm{s}^{-1}$.

for 298 K and in triangles for 400 K). Four trajectories, each of 200 ns, were used to construct the potential energy surfaces of the maltoside, see Figure 3(a). The results compare qualitatively well with earlier umbrella sampling results by Dimelow *et al* [63]. Furthermore, using the transition state theory (TST) and the calculated barrier height of 7.4 kcal/mol, the rate constant for the transition between the two main conformations B and F (see Figure 3), B → F, is estimated to be $2.3 \times 10^7\,\mathrm{s}^{-1}$. This result is in good agreement with earlier studies [63] (3.8×10^7 and $6.3 \times 10^6\,\mathrm{s}^{-1}$, estimated using TST and the TPS, respectively). We also calculated the rate constant by making use of the life time histograms of the state B (Figure 3b) in accelerated MD simulations, which is about 0.90 ns, corresponding to a $\tilde{k}$ of $1.1 \times 10^9\,\mathrm{s}^{-1}$. This value of $\tilde{k}$ and a $\Delta\Delta G^{\neq}$ of 3.1 kcal/mol lead to $k = 5.1 \times 10^6\,\mathrm{s}^{-1}$. Similar simulations were performed on a transition over a higher barrier, the *cis* to *trans* transition of a dipeptide, Ala–Pro. The calculated rate constant for the *cis* to *trans* transition at 283 K is $1.6{\times}10^{-3}\,\mathrm{s}^{-1}$ (estimated using TST and a barrier height of 20 kcal/mol). The corresponding experimental results is $6\text{–}9 \times 10^{-4}\,\mathrm{s}^{-1}$ [64]. These results demonstrated the applicability of the accelerated MD simulations in thermodynamics calculations and in the understanding of reaction mechanism without requiring pre-determined reaction coordinates.

2.1.3. *Accelerated protein folding*

When the starting conformation of a simulation is far from equilibrium and has high energies, a bias potential that only favours the high energy states [37] will lead to an inefficient search for the low energy stable states. One of the possible examples is the protein folding problem, in which one normally starts from an extended and high energy conformation to find its native structure, which typically contains more local and global interactions and has a lower energy. When appropriate parameters can be chosen, this problem can be avoided by the accelerated MD simulation using a bias potential of the form of Equation (2). We have demonstrated [37] the applicability of Equation (2) by using it for the folding of a small protein, Trp-cage [65]. Using a bias potential of the form of Equation (2), we have successfully folded this protein to its native structure quickly and repeatedly. Figure 4 shows that the applied bias potential allows efficient sampling (black) over a much wider energy range compared to the normal MD (grey) and as a result allows a much faster folding of the protein (see the black curve in Figure 5 for the backbone root mean square deviation (RMSD) values as a function of time; a computed structure is shown in Figure 6 in comparison with the experimental structure). On the contrary, the 10 ns normal MD simulation is far from folding (a relatively flat RMSD curve around 5 Å). Actually, it takes longer to fold by using the normal MD simulation. Although Simmerling *et al.* obtained less than 1 Å C_α-RMSD in about 10 ns at 325 K, they also found that normal MD simulations at 300 K were kinetically trapped on the 100 ns time-scale [66]. Twenty-four independent trajectories were obtained using

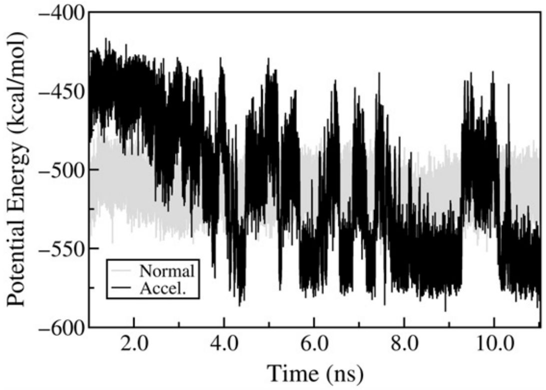

Figure 4. Potential energy distributions of the Trp-cage in different simulations (10 ns): normal MD (grey), and accelerated MD using bias potential of Equation (2) (black). The parameters used in Equation (2) are $n = 5$, $a_1 = -4.32$, $V_1 = -580$, $\sigma_1 = 25$; $a_2 = -0.792$, $V_2 = -500$, $\sigma_2 = 20$; $a_3 = -2.88$, $V_3 = -470$, $\sigma_3 = 32$; $a_4 = -3.96$, $V_4 = -440$, $\sigma_4 = 30$; $a_5 = -12.96$, $V_5 = -400$, $\sigma_5 = 50$. They are all in units of kcal/mol.

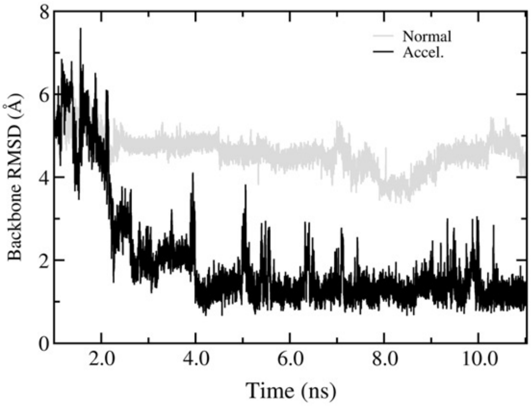

Figure 5. The backbone RMSD of the Trp-cage from the 10 ns normal MD (grey) and the accelerated MD simulation (black).

accelerated MD simulations and twenty two of them reached folded states within 10 ns. Therefore, the success folding rate is 91.6%. The typical computational folding time is 1–4 ns (75% of the trajectories fold in this time range) and the typical RMSD for the folded structure is around 1.5 Å for all heavy atoms.

To further test the applicability of the newly developed accelerated MD simulation method, protein folding simulations were also performed on other small proteins: the short β-hairpin polypeptide, trpzip2, trpzip4, and GB1; the three-strand β-sheet DPDP; the all-helical villin headpiece, and the designed small protein BBA5 with both α and β secondary structures. For example, the β-hairpin trpzip2 was folded (see the best folded structure in Figure 7) using the generalized Born implicit solvent model, with a smallest C_α-RMSD of 0.53 Å in comparison with the experimental structure (to the best of our knowledge, the best that has been achieved by all atom MD simulations). In addition, due to the fast reversible folding and unfolding events observed in the continuous trajectories (see Figure 8 for the multiple folding events observed in a 500 ns trajectory), the free energy surface and a folding mechanism of trpzip2 are obtained [67].

2.1.4. *Enhanced sampling for urea aqueous solution*

The applicability of the bias potential method was also demonstrated by its usage in simulating the aqueous solution of urea. A 7.5 M urea aqueous solution with 900 urea

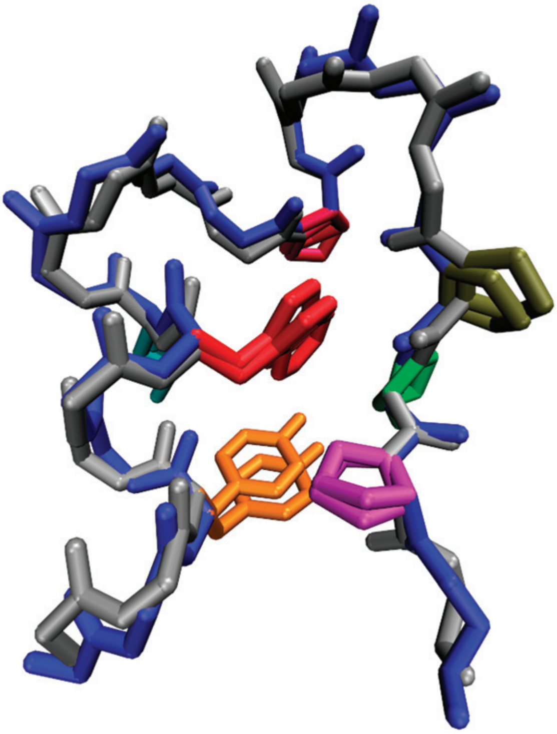

Figure 6. [Colour online] The best folded structure (blue) from accelerated MD simulations compared with the NMR structure of Trp-cage (silver). Only side chains forming the tryptophan cage are shown.

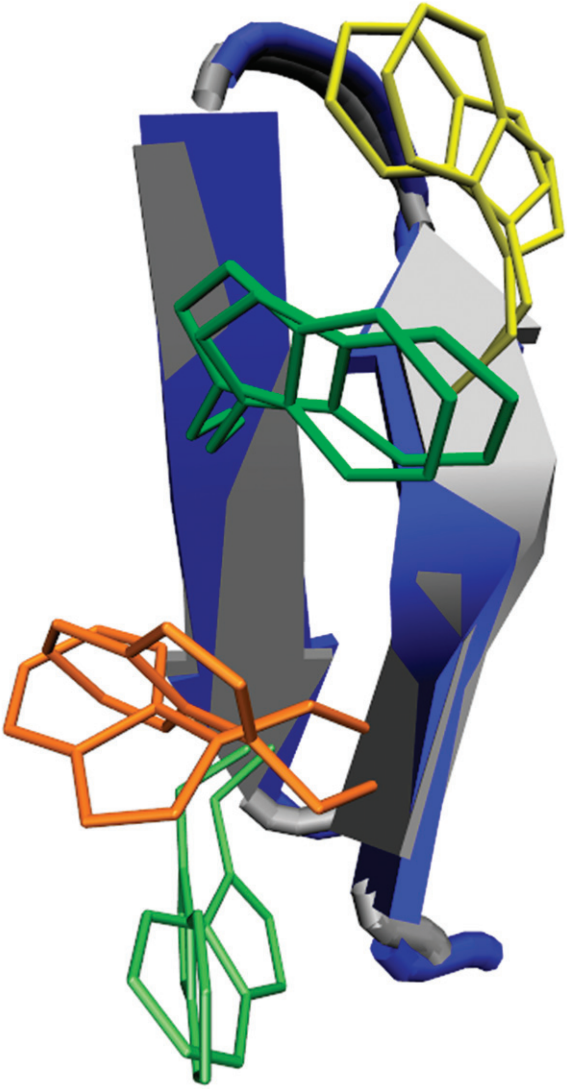

Figure 7. [Colour online] The best folded structure (blue) from accelerated MD simulations with 0.53 Å C_α-RMSD and 1.05 Å heavy atom RMSD compared with the NMR structure of 1LE1 (silver). The four tryptophan residues Trp_2, Trp_4, Trp_9, Trp_{11} are shown.

molecules and 4629 TIP3P water molecules under the periodic condition was studied. The essence of the accelerated MD method is to get an even sampling in a wide energy range. However, it is difficult to evenly broaden the energy distribution by adding bias potential to such a large energy range (in this case 750 kcal/mol), so that more steps are needed to obtain a broad and even energy distribution in explicit solvent simulations. The energy distribution of the normal MD simulation of the urea aqueous solution is shown in Figure 9(a) (see the solid line, the logarithm of probability *vs* potential energy)

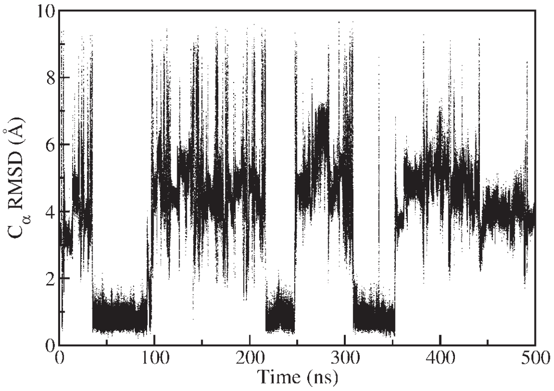

Figure 8. The C_α-RMSD distributions of the 500 ns accelerated MD simulations. Multiple folding/unfolding processes can be observed.

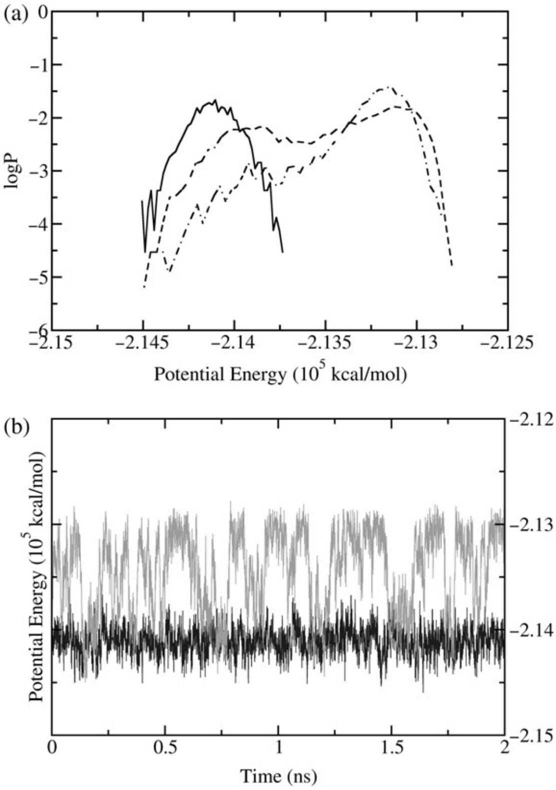

Figure 9. Potential energy distribution of the 7.5 M urea aqueous solution using normal and accelerated MD simulations. (a) The energy distribution of the normal MD simulation (solid line), the energy distribution of the accelerated MD simulation using an non-optimal bias potential (dash-dot line) and the energy distribution of the accelerated MD simulation using the final optimized bias potential (dash line), which is broad and even. (b) The black line represents the energy distribution of the normal MD simulation and the grey line represents the energy distribution of the accelerated MD simulation, which is two times wider than the normal MD simulation.

which is in the range of −214,500 kcal/mol to −213,750 kcal/mol and centred around −214,750 kcal/mol. The goal is to broaden the energy distribution as much as two times wider (i.e. width = 1500 kcal/mol, from −214,500 kcal/mol to −213,000 kcal/mol). At first, Gaussian functions were introduced to increase the sampling mainly at the less sampled high energy region around −213,000 kcal/mol. The energy distribution is then shown in Figure 9(a) as the dash-dot line. Another Gaussian term was then added and further parameter optimizations were conducted to make the sampling to be even for different energy regions. It is shown in the dash line of Figure 9(a) that the final bias potential enables us to obtain even samplings in low, high and intermediate energy regions. Similarly, it is shown in Figure 9(b) that the accelerated MD simulation by introducing bias potential (grey) has a two times wider energy distribution than the normal MD simulation (black), and the samplings of both low and high energy regions are even. Therefore, although normal MD simulations could be trapped for a very long time (>tens of nanoseconds) by the stability of locally crystallized water or urea clusters, with the help of the bias potential, not only the local minima can be escaped very easily, but also the transitions between low and high energy regions are increased significantly.

2.2. *Generalized non-Boltzmann distribution method for the enhanced sampling in the temperature space and fast partition function calculations*

2.2.1. *Method*

The method described above is, in essence, an energy-based enhanced sampling method. In the following, we review a different approach, which broadens the energy distribution in molecular dynamics (or Monte Carlo) simulations using a temperature-based method. To do this, we first write a generalized distribution function $p(V)$, which is a function of the potential energy V of the system, as an integration over β ($=1/k_\mathrm{B}T$, k_B being the Boltzmann constant and T being the temperature),

$$p(V) = \int_{\beta'} f(\beta')e^{-\beta' V}\,\mathrm{d}\beta'. \tag{8}$$

It is apparent that if one takes $f(\beta') = n(V)\delta(\beta' - \beta)$ with $n(V)$ being the density of states, Equation (8) reduces to the normal Boltzmann (after being normalized) distribution function $p_B(V) = n(V)e^{-\beta V}$. On the other hand, if one chooses to focus on a range of β (e.g. β_1 to β_N), separating this range of temperature into a series of smaller ranges $[(\beta_1, \beta_2), \ldots, (\beta_k, \beta_{k+1}), \ldots, (\beta_{N-1}, \beta_N)]$, writing $f(\beta') = n(V)\delta(\beta' - \beta_k)$ and making $n(V) = e^{\alpha_k}$ if V falls into one of the pre-selected energy ranges$V_k < V \leq V_{k+1}$ ($k = 1, 2, \ldots, N$), the function $p(V)$ then corresponds to that generated in a multi-canonical simulation [9,10]. If, as another example, independent simulations are performed at different β in parallel, and these trajectories are exchanged with the correct weighting factor, one can end up with replica exchange simulations.

In molecular dynamics simulations, a distribution function of Equation (8) can be obtained by running simulations on a modified potential V', which is a function of V, at the desired temperature (thus β). V' can be simply defined as

$$V' = -\frac{1}{\beta}\ln \int_{\beta'} f(\beta')e^{-\beta' V}\,\mathrm{d}\beta'. \tag{9}$$

and the biased forces F_b that are used in the Newtonian equations with the modified (biased) potential V' become

$$F_b = -\frac{\partial V'}{\partial r} = -\frac{\int_{\beta'} \beta' f(\beta') e^{-\beta' V} \, \mathrm{d}\beta'}{\beta \int_{\beta'} f(\beta') e^{-\beta' V} \, \mathrm{d}\beta'} \frac{\partial V}{\partial r} = \frac{\int_{\beta'} \beta' f(\beta') e^{-\beta' V} \, \mathrm{d}\beta'}{\beta \int_{\beta'} f(\beta') e^{-\beta' V} \, \mathrm{d}\beta'} F. \tag{10}$$

In Equation (10), F is the force vector calculated using the original potential function of the system under study. The remaining problem, unfortunately quite often a rather difficult one, is the determination or estimation of the functions $f(\beta')$ which can lead to a uniformly efficient sampling in the desired energy range, although, in principle, $f(\beta')$ could be a continuous equation and is directly related to the partition function. A quick and robust method is needed for calculating $f(\beta')$. For this purpose, and to keep the numerical implementation simple, in the following, we take a form of $f(\beta')$ that is a sum of δ functions:

$$f(\beta') = \sum_{k=1}^{N} n_k \delta(\beta' - \beta_k). \tag{11}$$

Equation (8) then becomes simply a summation of the Boltzmann factor $e^{-\beta_k V}$ of different temperatures and each Boltzmann factor carries a weighting factor to be determined. The resulted distribution takes the form of a distribution from a replica exchange with the so-far undetermined weighting factors

$$p(V) = \sum_{k=1}^{N} n_k e^{-\beta_k V}. \tag{12}$$

For the achievement of a smooth distribution function in the form of Equation (12) using simulations, the parameters n_k are then determined by the requirement that each term in the summation contributes a desired fraction to the total distribution. Namely, if we define

$$P_k = n_k \int_r e^{-\beta_k V(\mathbf{r})} \mathrm{d}\mathbf{r} \tag{13}$$

integrating over the entire configuration space, one can pre-select the ratios between the P_k's for all k's between 1 and N. Therefore one can define a set of fixed expectation values $\{p_k^0\}$ for the normalized quantities p_k:

$$p_k = \frac{P_k}{\sum_{k=1}^{N} P_k}. \tag{14}$$

For example, if equal contributions of the P_k's desired, $p_k^0 = 1/N$ for all k's. Generally speaking, $\{p_k^0\}$ need not be uniform and can be chosen to emphasize the sampling in a sub-temperature (energy) range.

An efficient approach for the calculations of n_k can be obtained by realizing that the overlap of sampling is maximal between nearest neighbouring temperatures. In this approach n_k is not updated according to the absolute values of p_k, but according to the ratios between neighbouring p_k's. Let us first define a series of numbers m_k,

$$m_k = \begin{cases} 1 & \text{if} \quad k = 1 \\ \dfrac{n_{k+1}}{n_k} & \text{if} \quad 1 < k \leq N. \end{cases} \tag{15}$$

The original series of n_k is then simply related to the m_k's by

$$n_k = n_1 \prod_{j=1}^{k} m_j. \tag{16}$$

After setting either $n_1 = 1$, or $\sum_{k=1}^{N} n_k = 1$, one can determine the n_k's by knowing the m_k's.

2.2.2. *Application to a model system*

To illustrate the working principle of the proposed method, we apply it to a modeled two-dimensional potential, which possesses two energy minima that are connected by two reaction pathways. A contour plot of the potential energy surface is given in Figure 10. The energy is given in units of k_BT. To calculate the rate constant and finding transition paths for this model system, we first apply the non-Boltzmann approach for the sampling of configuration space. In these calculations, 40 different temperatures between T and $21T$ with equal intervals are chosen. To test the robustness of this method, two dramatically different initial guesses were given for n_k, $n_k = 1/40$ and $n_k = 2^{-k}$. It is shown in Figure 11 that in both cases n_k converges quickly (in a few thousand Monte Carlo steps).

2.2.3. *Search of the configurations for villin headpiece*

We further demonstrated how the sampling over the energy and configuration space is enhanced for a 35-residue protein, villin headpiece, and show that the enhanced sampling

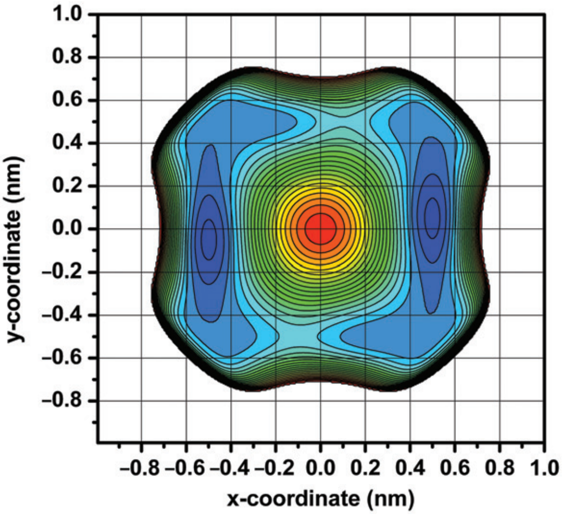

Figure 10. The contour plot of the potential energy surface of the model system:

$$V = 14.4\left\{e^{-\left[\frac{(x-0.4)^2}{0.02}+\frac{(y-0.5)^2}{0.05}\right]} + e^{-\left[\frac{(x+0.4)^2}{0.02}+\frac{(y+0.5)^2}{0.05}\right]}\right\} + 2.4\times 10^4\cdot(x-0.5)^2\cdot(x+0.5)^2$$
$$\times(y-0.5)^2\cdot(y+0.5)^2 + 12\left\{e^{-\left[(x-0.5)^2+\frac{(y-0.05)^2}{0.05}\right]} + e^{-\left[(x+0.5)^2+\frac{(y+0.05)^2}{0.05}\right]}\right\} + 12e^{20\cdot(x^2+y^2-0.6)}.$$

not only leads to its fast folding using an implicit model, but also speeds up the calculation of its partition function as a function of temperature [68]. Starting from an extended structure of the villin headpiece HP35, 20 different temperatures between 273 K and 417 K were chosen to enhance the sampling of the configuration space. In Figure 12 we show the potential energy distribution in the normal MD simulation (grey) and the generalized non-Boltzmann distribution simulation (black). It is shown in Figure 12 that the present method allows a largely uniform energy distribution to be obtained in a large energy range (in an energy range of ~280 kcal/mol, compared to the energy range of ~100 kcal/mol sampled by the normal MD simulation). In simulations, a large variety of configurations

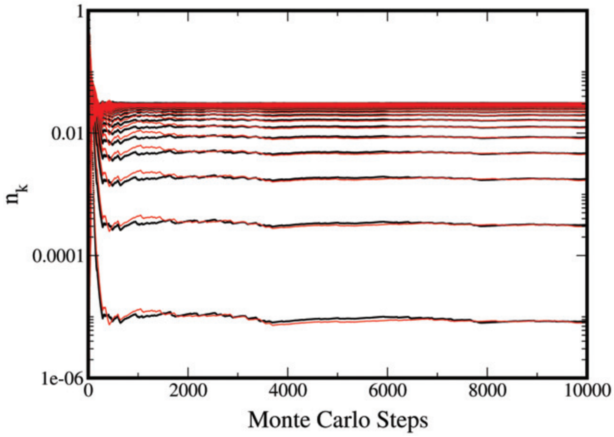

Figure 11. [Colour online] A plot of n_k as a function of iteration step (each 100 ps) obtained using generalized non-Boltzmann simulations. The black curve represents the initial guess, $n_k = 1/40$, and red curve represents the initial guess, $n_k = 2^{-k}$.

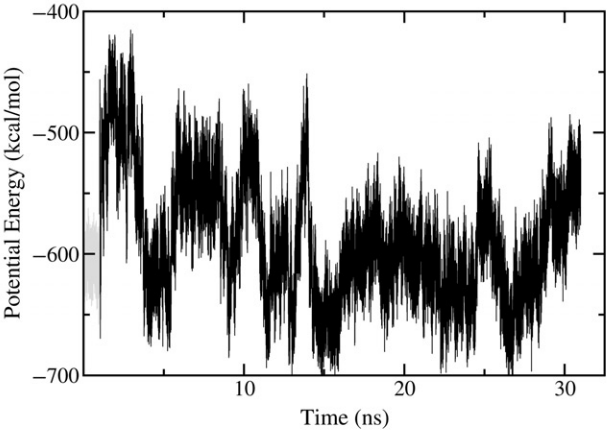

Figure 12. Potential energy distribution in the normal MD simulation (grey) and the generalized non-Boltzmann distribution simulation (black).

were observed and the best structure has a 2.4 Å backbone RMSD compared to experimental structure. In Figure 13, we show the change of n_k values as a function of time, from which it is evident that satisfactory convergence has been obtained within ~15 ns.

3. Separation of time-scales and accelerated calculation of conformational changes

3.1. *Method*

In the following, we switch gears to a different approach in treating the multi-scale motions of complex systems. As mentioned earlier, one of the central questions in molecular dynamic simulations is on the treatment of the multi-time-scale processes. In the following we describe a method that was recently developed in our group for the dynamic simulations of systems with a vast distribution of time-scales [49]. In this method, the slow degrees of freedom are treated as evolving in a potential of mean-force resulted from the motions of the fast degrees of freedom. Since the perturbation due to the chemical transitions in a motor protein is local and the energy associated with the perturbation is typically small (e.g., the total amount of energy that ATP hydrolysis provides under the cellular condition is about 20 k_BT), during the large-scale protein conformational changes, the fast degrees of freedom, including bond vibrations and angle bending, are close to local equilibrium. Assuming that we can separate the motions of a protein complex into collections of slow and fast degrees of freedom, x_1 and x_2, respectively, our task is thus to follow the motion along x_1 (to obtain a distribution function $f_1(x_1, t)$ along the coordinates x_1) under the influence of an almost equilibrium condition of the local environment x_2:

$$f_1(x_1, t) = \int_{x_2} f(x_1, x_2, t)\,\mathrm{d}x_2. \tag{17}$$

The above equation can be further written as

$$f_1(x_1, t) = \int_{x_2} f_2(x_2, t) f'(x_1; x_2, t)\,\mathrm{d}x_2 \tag{18}$$

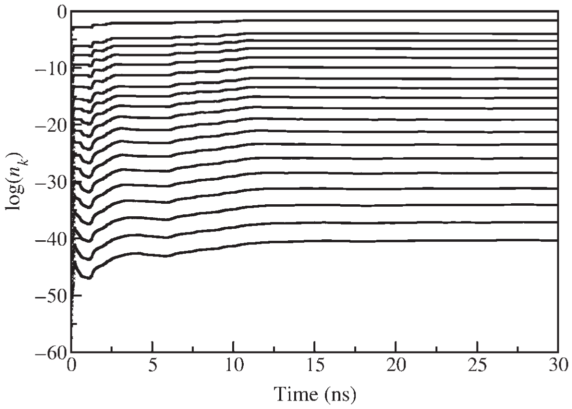

Figure 13. A plot of n_k as a function of simulation time in the generalized non-Boltzmann simulations of villin headpiece.

where $f_2(x_2, t)$ is the probability distribution function for x_2, and $f'(x_1;x_2, t)$ is a conditional probability function. Under the assumption that the motion in x_2 reaches equilibrium quickly and this equilibrium does not depend on x_1 sensitively, the $f_2(x_2, t)$ can be obtained from a normal molecular dynamics simulation for a relative short period of time, during which the fast degree of freedom approaches a local equilibrium. This distribution $f_2(x_2, t)$ is then used for the degrees of freedom in the propagation of dynamics to obtain $f_1(x_1, t)$.

In the implementation of this method, the separation between the fast and largely periodic degrees of freedom and slow directional degrees of freedom is not required to be known beforehand. The dynamic quantities (e.g., positions, momenta and forces) of the very fast degrees of freedom cancel out through the average over time much faster than the slow and directional motions of the protein do. Therefore, the average motion of the fast degrees of freedom is much slower than their real motions, although the distributions along these coordinates can be taken from the normal simulations that are used to obtain $f_2(x_2, t)$. Following this method, the dynamics along the slow degrees of freedom is calculated under the influence of an $f_2(x_2, t)$ obtained from a period of normal molecular dynamic simulations, although with a much slowed down motions in x_2. Due to the slowed motion in the fast degrees of freedom, the problem due to the large separation of time-scales is reduced and the integrations time step can be greatly prolonged. In the implementation of this method, a normal MD simulation is performed for a period of time, followed by a simulation using the multi-time-scale technique. The energy of the system is carefully followed. When the energy increases to a certain limit, the multi-time-scale simulation will be stopped and a new normal MD simulation will be performed using the new protein conformation as the starting point, which is followed again by another multi-time-scale simulation. These procedures are repeated, until a satisfactory result is obtained, judged either by the structure or by the energy.

The idea of 'filtering' the fast motions out through dynamic averaging in our method is similar to that used in self-guided molecular dynamics. However, there are several important features that are different: (1) In our method no additional forces that are added with a somehow arbitrary magnitude. (2) Instead of speeding up the motions along the slow degrees of freedom, we slow down the fast degrees of freedom, which are treated in a form of potential of mean force. (3) As a result of (2), the real dynamics of the slow degrees of freedom, including its dependence on the real time, is largely conserved. (4) The addition of a too large extra force has the potential of increasing the speed of motion and thus requires a smaller integration time step. Consequently, there is a limit to how much one can speed up the simulations in SGMD. By slowing down the fast degrees of freedom, however, one does not create such a problem.

3.2. *Applications*

To demonstrate how the multi-time-scale method works, the method was applied to study the conformational change of calmodulin in solution, a calcium binding protein with 148 amino acids, upon the removal of the calcium ion that was originally bound. Using the multi-time-scale method we were able to observe the conformational change from the calcium bound conformation to the calcium-free conformation of calmodulin due to Ca^{2+} release. Figure 14 shows the change of one of the bond lengths (the C_β–C_γ bond of the residue Leu_{18}) and one of the bond angles (the C_β–C_γ–$C_{\delta 1}$ bond angle of the same residue) during the normal MD simulation and the multi-time-scale simulation. It is seen that the application of multi-time-scale technique slows down the motions of the bond and bond angle vibrations, while keeping the distributions of the bond length and bond angle. Therefore, a relatively large integration time step of 40 fs (20 times larger than that of the normal MD simulation) can be used in the multi-time-scale simulation. Figure 15 below shows the backbone RMSD of the N-terminal domain of calmodulin, which indicates the open to close conformation change of the protein. It shows that multi-time-scale method

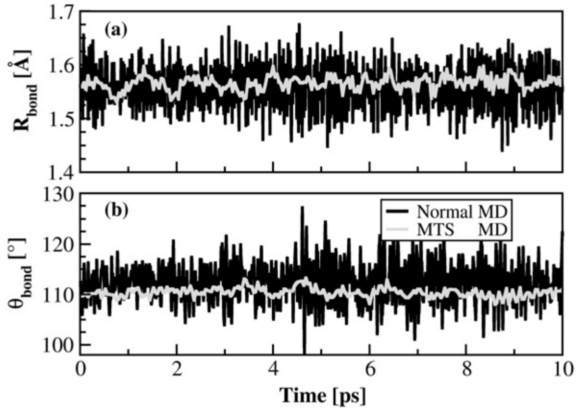

Figure 14. Leu$_{18}$ (a) C_β–C_γ bond length and (b) C_β–C_γ–$C_{\delta1}$ bond angle comparison between the normal MD and the multi-time-scale MD over 10 ps of each.

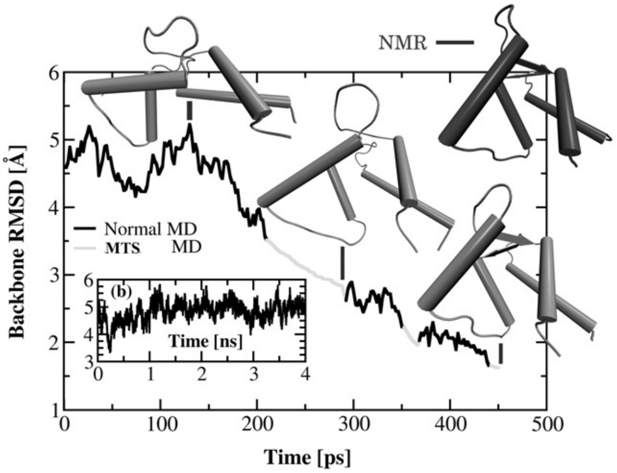

Figure 15. (a) Backbone RMSD between simulated structures and the NMR apo structure 1F70 for calmodulin N-terminal domain. Black denotes the normal simulation stage which is used to accumulate information for calculating the average. Grey denotes the multi-time-scale MD simulation stage. (b) Backbone RMSD of a 4 ns normal MD simulation using the same initial structure as that used to obtain (a).

captures the slow motion of the protein. The best structure that we obtain as a result of these simulations has a backbone RMSD of 1.56 Å in comparison to the NMR structure of the N-terminal domain of the apo calmodulin.

4. Enhanced sampling in the trajectory space

4.1. *Accelerated MD simulations guided transition path sampling*

4.1.1. *Method*

As discussed earlier, MD simulations performed using a biased potential allow fast convergence for the sampling of the configuration space and since the simulations are performed at a constant temperature, the momenta follow a Maxwell distribution. Although using the bias potential method, one can easily generate successful trajectories connecting these different states without the requirement of knowing the reaction coordinate, they do not represent the transition paths for the original system. In the following, a systematic method of selectively generating reactive transition paths for the original unbiased system is proposed. The basic assumptions of this method are: (1) the convergence in the free energy calculation indicates that a correct Boltzmann distribution in the configuration space has been achieved when they are weighted by the corresponding bias potential term; [36] (2) the total energy of the system is kept constant. As a result of the above two assumptions, (3) trajectories generated on the *real* potential, using the phase space points sampled for the *biased* system as initial points, constitute a sample of *real* and *unbiased* transition paths for the *real* system when the initial points are weighted by their correct Boltzmann distribution. And further, (4) the trajectories obtained using the bias potential provide a biased sampling of the phase space, from which the most 'reactive' phase space region, which contains phase space points more likely to be on a reactive trajectory of a given (short) length, can be identified. If a phase space point in a system with a lowered barrier belongs to an unsuccessful transition trajectory (which is short in time), it is less likely to be on a successful transition path of the same length in the original system with a higher barrier. As a result, the phase space points which are more likely to be 'reactive' will be chosen with a high probability for the forward/backward trajectory shooting. This procedure will further reduce the computational cost by reducing the sampling over the unsuccessful transition paths (the ratio between the successful and unsuccessful trajectories are easily recovered, see below).

The method described above is illustrated in Figure 16 using a double well potential. The two states corresponding to the two potential wells are labelled as states A and B (the reaction coordinate is labelled as R.C.). First, a simulation is performed for the biased potential and successful trajectories such as the one shown in Figure 16(b) are obtained. Many small segments of the trajectory which contain the transition path (such as the one enclosed in the rectangular region in Figure 16b) will be chosen for further simulations. Phase space points (configuration and momentum) will be randomly (or every certain number of points) taken from these segments, e.g., the points 1, 2, and 3 in Figure 16(b). In the next step, MD simulations will be performed on the original potential, using each of the chosen phase space points as the starting point and the system will be propagated both forward and backward for a given short time period. If the resulting trajectory ends at potential well A (e.g., the backward trajectory) and B (e.g., the forward trajectory), it is counted successful (e.g., trajectory 2 in Figure 16d). Otherwise, it is considered

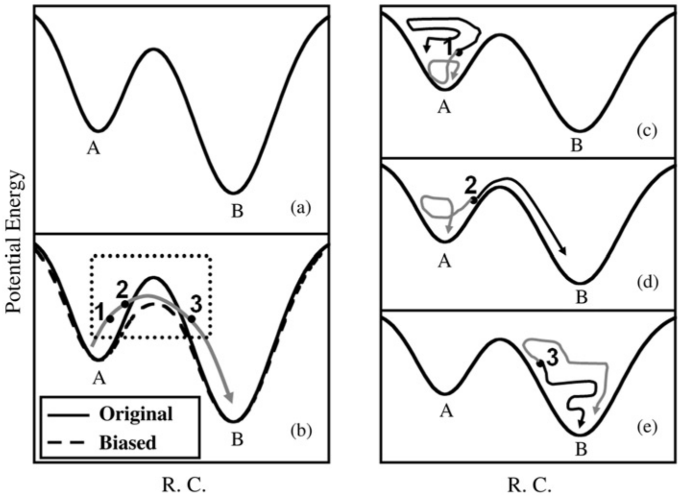

Figure 16. Scheme of the selective sampling of transition path method: (a) A double well potential model system. (b) The barrier in the original potential (black solid line) is lowered in the biased potential (black broken line). A successful transition trajectory under bias potential is shown with the grey arrow line. Points such as 1, 2, 3 in the dotted rectangular region were chosen for trajectory shooting. (c), (d), (e) Shooting trajectories starting from points 1, 2, 3 in (b). Grey arrow lines represent the backward shooting trajectories and black arrow lines denote the forward shooting trajectories.

unsuccessful (the same as in the standard TPS approach). With a sufficient sampling of these trajectories and phase space points, the successful transition paths can also be adequately sampled. The number of corresponding unsuccessful trajectories, N_{non}, that is needed for the rate calculations can be estimated as $N_S = N_{\text{non}}\eta$, where η is the ratio between N_S, the re-weighted number of phase points that leads to successful transition paths in the given propagating time and that of the unsuccessful ones,

$$\eta = \frac{\sum_{\text{succ}} e^{\beta f(V)}}{\sum_{\text{unsuc}} e^{\beta f(V)}}. \tag{19}$$

In Equation (19) the summation is over the initial shooting points corresponding to the successful and unsuccessful transition paths, respectively. As usual, $\beta = 1/k_B T$ where k_{B} is the Boltzmann constant and T is the temperature. This weighting factor, $\exp[f(V(r))/k_B T]$, where $f(V(r))$ is the bias potential, recovers the Boltzmann distribution of the original system. The rate constant is then calculated in the standard way of transition path sampling based on the numbers of successful and unsuccessful trajectories

$$k = \left(\frac{N_s}{N_{\text{non}} + N_s}\right)\Big/\tau = \left(\frac{\eta}{1+\eta}\right)\Big/\tau \tag{20}$$

where τ is the length of the trajectories. Equation (20) is obtained by assuming a first-order reaction and when the length of the trajectories is much shorter than the reaction half time.

4.1.2. *Application to isomerization of methyl maltoside*

In the following, as an example, we apply the selective sampling method of transition paths introduced above to study the conformational change of methyl maltoside, and compare our results with those from studies using the standard transition path sampling method [63]. Firstly, six independent accelerated MD simulations similar to the ones mentioned in Section 2.1.2 were conducted. Each one of the six simulations was extended to 400 ns to guarantee that a correct sampling over the configuration space has been obtained. In order to combine accelerated MD simulations and transition path shooting method, potential energies, coordinates and velocities were recorded every 50 steps in all six simulations (every one out of 50 steps is used to reduce the correlation of the selected points). In practice, choosing fewer points will further reduce the correlation between different trajectories but also reduces the number of successful trajectories. In the simulations, there were in total 4.8×10^7 conformations which consisted of the ensemble of starting points for shooting trajectories using the selective sampling method of the transition path. Certainly, there is no need to shoot trajectories from all these points. The benefit of using the selective sampling of transition path method is that the most 'reactive' phase space region can be identified through the biased sampling of the phase space. Focusing on these points in the identified 'reactive' phase space portions will reduce the computational cost on the sampling over the unsuccessful transition paths. In our simulations, 205,792 points in total (0.43%) were selected out of the ensemble consisted of 4.8×10^7 phase space points. After shooting backward/forward from these 205,792 points, 10,729 successful transition trajectories were obtained. The total simulation time is then approximately 3.2 μs (the sum of the time needed by accelerated MD simulations: $6 \times 400\,\text{ns} = 2.4\,\mu\text{s}$ and trajectory shooting time: $205792 \times 4\,\text{ps} \sim 823\,\text{ns}$), while, during the same simulation time, only ~20 transitions could be found in normal MD simulations. Although not all of the 10,729 transition paths are important at room temperature, the improvement of the efficiency in sampling transition paths is prominent.

In the sampling of transition paths, initial phase space points for trajectory shooting are taken from the sampled points in the accelerated MD simulations, for both forward and backward (reverse the direction of stored velocities) trajectory calculations (each with a length of 2 ps). These new paths were then collected and determined whether they were successful or not. According to the free energy calculation and Figure 3, the reactant region was defined as $-70^\circ < \psi < 40^\circ$ and the product region was defined as $\psi > 150^\circ$ or $\psi < -150^\circ$. We show in Figure 17 two typical successful transition paths, the starting points of which are at time $t = 0$. Backward trajectories are represented by grey solid lines and forward trajectories are represented by black solid lines. Connecting a pair of grey and black lines shows a 4 ps transition path. In Figure 17(a), starting from point at $\psi = -107^\circ$, the forward segment ends at the product region F in 2 ps and the backward segment ends at the reactant region B in 2 ps, resulting in a successful transition path passing through the transition state T_1. Figure 17(b) shows a successful 4 ps transition path passing through transition state T_2. The rate constant of conformational change between B state and F state was estimated through these simulations to be $1.26(\pm 0.31) \times 10^7\,\text{s}^{-1}$, which is in reasonable agreement with the results obtained by transition state theory and the transition path sampling method [63]. The logarithm of the visiting probability as a function of φ and ψ was calculated from the successful transition path ensemble and shown in Figure 18.

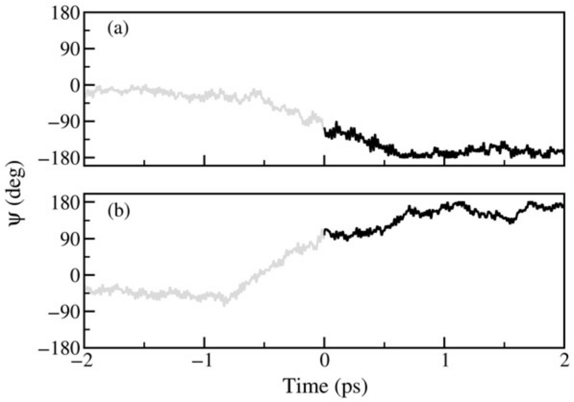

Figure 17. Typical successful transition trajectories: (a) successful transition through the transition state T_1 (see text); (b) successful transition through the transition state T_2 (see text).

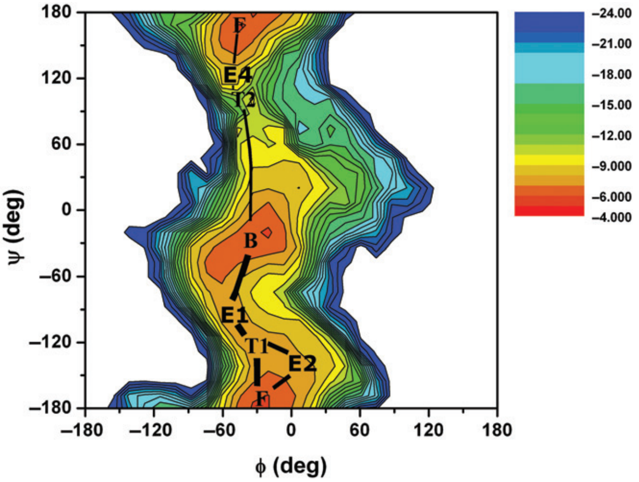

Figure 18. Logarithm of probabilities of states over φ and ψ calculated from successful transition path ensembles. Three transition path ensembles are labelled as E1, E2, E4.

All three transition pathways found in the previous TPS study [63] were also captured in this study. The most favourable transition path is the transition path E1 which goes through a rotation about ψ in the negative direction and passed through the transition state T_1. A less favourable transition path ensemble is E2 passing through a metastable region G (+30°, −130°) instead of T_1. The third transition path ensemble is the transition path E4 in which transitions propagated from the state B to the state F through a rotation passing through the transition state T_2 (see Figure 17b). Since the free energy barrier between B and F through a rotation in a positive direction is higher than that through negative direction, transition path E4 has the lowest probability density.

4.2. *Self-adaptive sampling of transition paths*

4.2.1. *Method*

In the method described below, we achieve the goal of fast identification of reaction pathways and calculations of kinetics for systems with high barriers and without precisely defined reaction coordinate by combining two strategies: (1) adaptive enhanced sampling in the energy and configuration space so that the highly energetic states (configurations) are sampled more frequently in a controlled way; (2) adaptive enhanced sampling of the phase space points and enhanced trajectory shooting initiated from these points. These two techniques lead to the enhanced sampling of reactive transition paths in a controlled manner. The enhanced sampling in the energy and configuration space was already discussed in Section 2. The sampling of reaction paths is performed in an adaptive way. The method is straightforward. Firstly, collective coordinates which are most likely involved in the reaction of interest will be selected, in the same way as is done in metadynamics for free energy calculations. Secondly, the range of coordinates is divided into small regions in a coarse-grained way (see Figure 10). The reactant and product states are also defined using the collective coordinates. Let us use two-dimensional collective coordinates (x, y) as an example, with the space of these two dimensions divided into $N_x \times N_y$ rectangular areas each with sides Δx and Δy (therefore the total space is a rectangle with sides of $N_x\Delta x$ and $N_y\Delta y$). Each of the small areas is labelled (e.g., (I_x, I_y)). An initial probability of 1 is assigned to each of the small areas. These probabilities are denoted as $X(I_x, I_y)$ and they will be updated as the simulations proceeds to encourage the sampling and trajectory shooting of the more reactive phase space points.

4.2.2. *Test of the adaptive transition path sampling method*

To illustrate the working principle of the proposed method, we apply it to the same two-dimensional potential model used in Section 2.2.2 (Figure 10), which possesses two energy minima that are connected by two reaction pathways. To calculate the rate constant and finding transition paths for this model system, we first apply the non-Boltzmann approach for the sampling of configuration space. In these calculations, 40 different temperatures between T and $3T$ with equal intervals are chosen. After the converged n_k were obtained, iteration are repeated for 100,000 Monte Carlo moves on the effective potential and the normalized probabilities $X(J_x, J_y)$ are updated self-adaptively during the simulations. The trajectory shooting on the real potential energy surface is performed using the Newtonian and Langevin equations with the velocity–Verlet algorithm. The mass is taken as 100 atomic units and the time step for the propagation is 1 fs. Figure 19 shows the sampled points (black) and the sampled points which can initiate successful trajectories (red). In Figure 20, we show the results for a system with zero friction and with trajectory length of 1 ps. It is seen from this figure that as the adaptive sampling proceeds, the successful rate of the trajectories increases.

5. Summary and future studies

In this review, we summarized a number of methods that have been developed and used in our laboratory for the studies of the thermodynamics and kinetics of complex systems. Methods used to enhance sampling in the energy, configuration, as well as trajectory

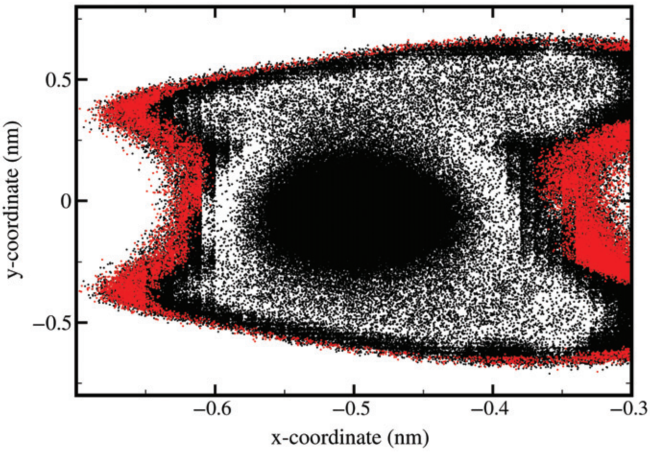

Figure 19. [Colour online] All sampled points (black) and the sampled points which can initiate successful trajectories (red) in adaptive transition path sampling of the model system.

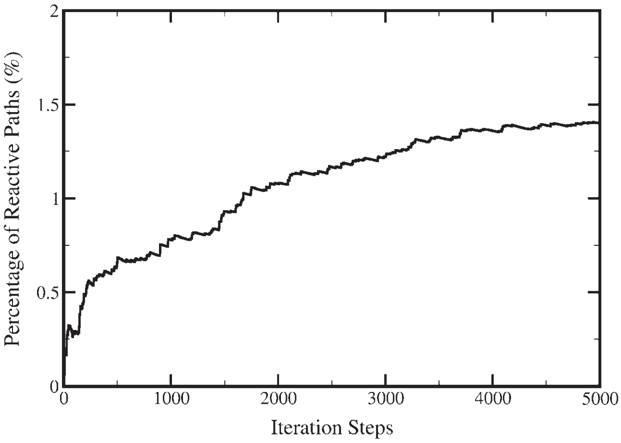

Figure 20. The ratio between the number of successful trajectories and the number of total trajectories as a function of iteration steps.

spaces were discussed. The existing applications of these methods have shown their efficiency in searching configurations and in calculating thermodynamics for both small and relatively large systems. In particular, these methods allow fast folding and unfolding trajectories to be obtained for small proteins and thus provide valuable information on their folding mechanisms. In the meantime, transition path sampling and rate constant calculations can be obtained with lowered computational cost when the transition path sampling method is used in combination with the enhanced sampling simulations. We expect that these methods will have broad applications in a large variety of complex systems. As a summary of our research effort in the field of enhanced sampling, the developed methodologies, studied systems and properties are listed in Table 1.

Table 1. Summary of our effort in the field of enhanced sampling.

Property	Methodology	System
Structure/Thermodynamics	Accelerated MD	Protein folding (Trp-cage, Trpzip2, Trpzip4, GB1, Villin, BBA5, etc.) Protein conformational change Sugar/membrane interactions
	Accelerated MC Coarse-grained protein model	Protein aggregation
Kinetics/Dynamics	Accelerated MD, MC/enhanced TPS	Protein conformational change Sugar/membrane interactions

One of the advantages of the biasing potential, either in the form of an added bias potential such as given by Equation (2) or in the form of the effective potential (Equation 9) obtained in the form of a non-Boltzmann distribution is that different domains of the system can be treated differently. In certain cases, this separated treatment of subsystems is essential. As shown from the discussion and the application, this method has the advantages of both replica exchange and multi-canonical simulations in that it broadens the energy distribution by making use of Boltzmann distributions in a large range of temperatures and achieves a largely uniform (or other desired) energy distribution. It avoids the usage of multiple parallel trajectory calculations and the constant exchange between them, which are required for the replica exchange method. On the other hand, it also avoids selecting energy ranges (which is sometimes difficult, in particular when slow non-equilibrium processes are involved) as the criterion for assigning different distribution functions, which is typically required in a multi-canonical simulation. In addition, due to the simultaneous presence of all the different Boltzmann distribution functions (of various temperatures), the random walk in the entire chosen energy space is largely facilitated and the resulted enhancement of the overlap between different distribution functions allows fast convergence to a reasonably good biased potential (and thus allows one to obtain the density of states in terms of multi-canonical simulations). As a result, the method is robust and allows fast convergence largely independent of the initial guesses. Given these merits, the method is expected to be helpful in complex systems for which the energy of interest is difficult to identify (e.g., a system far away from equilibrium) and/or a good initial guess of the density of states is hard to obtain. In the present review, we focused on the application of the enhanced sampling method in MD simulations. The application to Monte Carlo simulations is more straightforward and has shown to significantly accelerate thermodynamics simulations in the studies of polypeptide secondary structures and their aggregation with coarse-grained amino acid models [38].

We note that the methods presented here provide quick and reliable ways of generating bias potentials in the accelerated MD simulations. One advantage of using biased simulations is that one can choose to add different bias potentials to different parts of the system, and to activate only the degrees of freedom that are involved in the reaction of interest while keeping the rest of the system at more ‘real’ (less perturbed) conditions. In some simulations, this separated treatment is essential. For example, to study protein folding in explicit water, one would like to accelerate the motions of the proteins to sample its conformations quickly but does not want to perturb the structure of water (which evaporates easily) too much. On the other hand, in a simulation of trehalose protected DNA, a large bias potential is needed to enhance the sampling of the trehalose (which is close to a solid) configurations and the bias on DNA should be minimal, due to the easy deformation and destruction of the latter.

The present methods provide a framework for the acquisition of thermodynamic and kinetic information of complex systems with small rate constant. The combination of enhanced sampling in the configuration and trajectory spaces allows fast calculations of the rate constant. The main assumption of these methods is that the system follows ergodically and therefore thermodynamic information of the real systems can be recovered from the biased ones. The methods are thus expected to fail in the lack of ergodicity and are incapable of yielding slow dynamic information (long correlation times). More theoretical development is needed to understand the slow dynamics, which is characteristic of many complex systems.

Acknowledgements

The authors acknowledge financial support from the Texas A&M University, the ACS-Petroleum Research Foundation, and the Welch Foundation. YQG is a recipient of a Camile and Henry Dreyfus New Faculty Award and a 2006 Searle Scholar.

References

[1] D. D. Frantz, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **93**, 2769 (1990).
[2] C. Bartels and M. Karplus, J. Phys. Chem. B **102**, 865 (1998).
[3] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).
[4] Y. Sugita, A. Kitao, and Y. Okamoto, J. Chem. Phys. **113**, 6042 (2000).
[5] U. H. E. Hansmann, Chem. Phys. Lett. **281**, 140 (1997).
[6] M. Falcioni and M. W. Deem, J. Chem. Phys. **110**, 1754 (1999).
[7] Q. L. Yan and J. J. de Pablo, J. Chem. Phys. **111**, 9509 (1999).
[8] Q. L. Yan and J. J. de Pablo, J. Chem. Phys. **113**, 1276 (2000).
[9] B. A. Berg and T. Neuhaus, Phys. Lett. B **267**, 249 (1991).
[10] S. G. Itoh and Y. Okamoto, J. Chem. Phys. **124**, 104103 (2006).
[11] G. Bussi, A. Laio, and M. Parrinello, Phys. Rev. Lett. **96**, 090601 (2006).
[12] D. H. Min, Y. S. Liu, I. Carbone, *et al.*, J. Chem. Phys. **126**, 194104 (2007).
[13] V. Babin, C. Roland, T. A. Darden, *et al.*, J. Chem. Phys. **125**, 204009 (2006).
[14] H. Grubmuller, Phys. Rev. E **52**, 2893 (1995).
[15] J. Lee, H. A. Scheraga, and S. Rackovsky, J. Comput. Chem. **18**, 1222 (1997).
[16] A. F. Voter, Phys. Rev. Lett. **78**, 3908 (1997).
[17] A. F. Voter, J. Chem. Phys. **106**, 4665 (1997).
[18] L. Piela, J. Kostrowicki, and H. A. Scheraga, J. Phys. Chem. **93**, 3339 (1989).
[19] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).
[20] E. Darve and A. Pohorille, J. Chem. Phys. **115**, 9169 (2001).
[21] D. Rodriguez-Gomez, E. Darve, and A. Pohorille, J. Chem. Phys. **120**, 3563 (2004).
[22] A. Mitsutake, Y. Sugita, and Y. Okamoto, J. Chem. Phys. **118**, 6664 (2003).
[23] A. Mitsutake, Y. Sugita, and Y. Okamoto, J. Chem. Phys. **118**, 6676 (2003).
[24] F. G. Wang and D. P. Landau, Phys. Rev. E **64**, 056101 (2001).

[25] F. G. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).
[26] D. P. Landau and F. Wang, Comput. Phys. Commun. **147**, 674 (2002).
[27] J. Kim, J. E. Straub, and T. Keyes, Phys. Rev. Lett. **97**, 050601 (2006).
[28] C. Zhang and J. P. Ma, Phys. Rev. E **76**, 036708 (2007).
[29] G. M. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).
[30] A. Warmflash, P. Bhimalapuram, and A. R. Dinner, J. Chem. Phys. **127**, 154112 (2007).
[31] K. Arora and C. L. Brooks, Proc. Natl. Acad. Sci. U. S. A. **104**, 18496 (2007).
[32] D. Hamelberg, J. Mongan, and J. A. McCammon, J. Chem. Phys. **120**, 11919 (2004).
[33] D. Hamelberg, T. Shen, and J. A. McCammon, J. Chem. Phys. **122**, 241103 (2005).
[34] D. Hamelberg, T. Shen, and J. A. McCammon, J. Am. Chem. Soc. **127**, 1969 (2005).
[35] E. J. Barth, B. B. Laird, and B. J. Leimkuhler, J. Chem. Phys. **118**, 5759 (2003).
[36] Y. Q. Gao and L. J. Yang, J. Chem. Phys. **125**, 114103 (2006).
[37] L. J. Yang, M. P. Grubb, and Y. Q. Gao, J. Chem. Phys. **126**, 125102 (2007).
[38] Y. Mu and Y. Q. Gao, J. Chem. Phys. **127**, 105102 (2007).
[39] Y. Q. Gao, J. Chem. Phys. **128**, 064105 (2008).
[40] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).
[41] H. C. Andersen, J. Comput. Phys. **52**, 24 (1983).
[42] R. D. Swindoll and J. M. Haile, J. Comput. Phys. **53**, 289 (1984).
[43] M. E. Tuckerman, G. J. Martyna, and B. J. Berne, J. Chem. Phys. **93**, 1287 (1990).
[44] M. Tuckerman, B. J. Berne, and G. J. Martyna, J. Chem. Phys. **97**, 1990 (1992).
[45] J. A. Izaguirre, S. Reich, and R. D. Skeel, J. Chem. Phys. **110**, 9853 (1999).
[46] P. Minary, M. E. Tuckerman, and G. J. Martyna, Phys. Rev. Lett. **93**, 150201 (2004).
[47] X. W. Wu and S. M. Wang, J. Phys. Chem. B **102**, 7238 (1998).
[48] X. W. Wu and S. M. Wang, J. Chem. Phys. **110**, 9401 (1999).
[49] L. J. Yang and Y. Q. Gao, J. Phys. Chem. B **111**, 2969 (2007).
[50] C. Dellago, P. G. Bolhuis, and F. S. Csajka, J. Chem. Phys. **108**, 1964 (1998).
[51] P. G. Bolhuis, C. Dellago, and D. Chandler, Faraday Discuss. **421**, 110 (1998).
[52] C. Dellago, P. G. Bolhuis, and D. Chandler, J. Chem. Phys. **110**, 6617 (1999).
[53] P. G. Bolhuis, D. Chandler, and C. Dellago, Ann. Rev. Phys. Chem. **53**, 291 (2002).
[54] T. S. van Erp, D. Moroni, and P. G. Bolhuis, J. Chem. Phys. **118**, 7762 (2003).
[55] D. Moroni, T. S. van Erp, and P. G. Bolhuis, Physica A **340**, 395 (2004).
[56] D. Moroni, P. G. Bolhuis, and T. S. van Erp, J. Chem. Phys. **120**, 4055 (2004).
[57] P. G. Bolhuis, C. Dellago, and D. Chandler, Proc. Natl. Acad. Sci. U. S. A. **97**, 5877 (2000).
[58] P. G. Bolhuis, Proc. Natl. Acad. Sci. U. S. A. **100**, 12129 (2003).
[59] R. Radhakrishnan and T. Schlick, Proc. Natl. Acad. Sci. U. S. A. **101**, 5970 (2004).
[60] G. A. Huber and S. Kim, Biophys. J. **70**, 97 (1996).
[61] A. M. A. West, R. Elber, and D. Shalloway, J. Chem. Phys. **126**, 145104 (2007).
[62] X. B. Fu, L. J. Yang, and Y. Q. Gao, J. Chem. Phys. **127**, 154106 (2007).
[63] R. J. Dimelow, R. A. Bryce, and A. J. Masters, J. Chem. Phys. **124**, 114113 (2006).
[64] C. I. D. Newman and V. L. McGuffin, Electrophoresis **27**, 542 (2006).
[65] J. W. Neidigh, R. M. Fesinmeyer, and N. H. Andersen, Nature Structural Biology **9**, 425 (2002).
[66] C. Simmerling, B. Strockbine, and A. E. Roitberg, J. Am. Chem. Soc. **124**, 11258 (2002).
[67] L. J. Yang, Q. Shao, and Y. Q. Gao, (to be published).
[68] L. J. Yang and Y. Q. Gao (to be published).