Enhanced sampling and applications in protein folding in explicit solvent

Cheng Zhang[1] and Jianpeng Ma[1,2,a]

[1]*Department of Bioengineering and Applied Physics Program, Rice University Houston, Texas 77005, USA*
[2]*Verna and Marrs McLean Department of Biochemistry and Molecular Biology,*
Baylor College of Medicine, One Baylor Plaza, BCM-125 Houston, Texas 77030, USA

We report a single-copy tempering method for simulating large complex systems. In a generalized ensemble, the method uses runtime estimate of the thermal average energy computed from a novel integral identity to guide a continuous temperature-space random walk. We first validated the method in a two-dimensional Ising model and a Lennard-Jones liquid system. It was then applied to folding of three small proteins, trpzip2, trp-cage, and villin headpiece in explicit solvent. Within $0.5 \sim 1$ microsecond, all three systems were reversibly folded into atomic accuracy: the alpha carbon root mean square deviations of the best folded conformations from the native states were 0.2, 0.4, and 0.4 Å, for trpzip2, trp-cage, and villin headpiece, respectively. © *2010 American Institute of Physics*. [doi:10.1063/1.3435332]

I. INTRODUCTION

Molecular simulations at room temperature usually suffer from a slow dynamics for large complex systems, such as proteins in explicit solvent. A promising solution to the problem is to use tempering methods, either single-copy based methods[1,2] such as simulated tempering, or multiple-copy based methods such as parallel tempering, also known as replica exchange.[3] In either case, the system regularly changes its temperature in a way that is consistent to the underlying thermodynamics. The value of these methods lies in that they can efficiently overcome energy barriers by exploiting a fast dynamics at higher temperatures.

In traditional tempering methods, the temperature is a discrete random variable that can only assume a few predefined values. The success rate of transitions between two neighboring temperatures depends on the overlap of canonical energy distributions at the two temperatures and decays as the system size grows. Thus, to reach an optimal sampling efficiency for a large system, one needs to narrow down the temperature gap, and to increase the number of sampling temperatures. In simulated tempering, the number of weighting parameters to be estimated for simulation increases with the number of temperatures. In replica exchange, the node-node communication cost increases with the number of temperatures. Naturally, it is desirable to have an efficient tempering method that does not depend on a discrete-temperature setup.

In this paper, we report a single-copy tempering method in which the temperature is a continuous variable driven by a Langevin equation. It is based on an improved version of a previous method.[2] By employing improved estimators for thermodynamic quantities, one cannot only realize an efficient tempering but also correctly calculate thermodynamic quantities for the entire temperature spectrum. The essential feature of the method is the calculation of the thermal average energy $\widetilde{E}(\beta)$ along the simulation trajectory. After the convergence of $\widetilde{E}(\beta)$, the partition function and other thermodynamic quantities can be easily derived.

The paper is organized as follows. In Sec. II, we give a theoretical derivation of the method. In Sec. III, the method is verified on a two-dimensional Ising model and a Lennard-Jones liquid system, where the exact thermodynamic quantities are either known or accurately computable. In Sec. IV, we apply the method to the folding of three small proteins, trpzip2, trp-cage, and villin headpiece, in explicit solvent. The minimal alpha-carbon root mean square deviation (C_α-RMSDs) of the best folded conformations from the native states are 0.2, 0.4, and 0.4 Å, respectively (the last figure for villin headpiece is measured from an x-ray reference structure; and it should be 1.0 Å if it is measured from an NMR reference structure).

II. METHOD

Our method samples the system in a continuous temperature range and calculates thermodynamic properties as functions of the temperature. As we shall see, a random walk in temperature space only requires an estimate of the average energy at the current temperature in order to correctly populate the desired distribution. We present an efficient way for estimating the average energy and use it to perform a fast sampling along the temperature. In addition, an adaptive averaging scheme is used to improve convergence in early stages.

This section mainly concerns the detailed description of the method. A relatively self-contained outline is first presented in Sec. II A. The rest of the section is organized as follows. In Sec. II B, we review basics of sampling in a generalized ensemble where the temperature is a continuous random variable. In Secs. II C and II D, we present integral identities that help the ensemble to asymptotically reach the desired distribution. In Sec. II E, we present an adaptive averaging scheme to accelerate initial convergence.

a)Electronic mail: jpma@bcm.tmc.edu.

A. Brief outline of implementation

The method can be implemented as follows. It concerns the simulation of a system in a given temperature range $(\beta_{min}, \beta_{max})$ according to a predefined temperature distribution $w(\beta)$, which is usually proportional to $1/\beta$ for a molecular system (the choice is explained in Appendix C). Note, we work with the reciprocal temperature $\beta = 1/(k_B T)$, where k_B is the Boltzmann constant and T is the regular temperature, β here is a variable that continuously changes within the temperature range.

In each simulation step, we first change system configuration according either to constant temperature molecular dynamics or to Monte Carlo methods. Next, we update statistics about the potential energy and then compute an estimated average energy $\widetilde{E}(\beta)$. The energy fluctuation, measured from the difference between the instantaneous potential energy E and the averaged value $\widetilde{E}(\beta)$ at the current temperature, determines the amount of temperature change that the system can afford to maintain the distribution of the generalized ensemble. It thus can be used to drive a temperature-space random walk in a way that preserves the underlying thermodynamics at each individual temperature. Practically, we use the following Langevin equation to guide the temperature space random walk,

$$\frac{d(1/\beta)}{dt} = \frac{d(k_B T)}{dt} = E - \widetilde{E}(\beta) - \frac{\partial \ln w(\beta)}{\partial \beta} + \frac{\sqrt{2}}{\beta} \xi, \qquad (1)$$

where E is the current potential energy, and ξ is a Gaussian white noise that satisfies $\langle \xi(t) \cdot \xi(t') \rangle = \delta(t - t')$. Note, here t is the time scale for integrating the Langevin equation, which does not necessarily coincide with the actual time in molecular dynamics. Apart from the random noise, the rate of change of the temperature is determined by the difference between the instantaneous energy E and the average energy at the current temperature $\widetilde{E}(\beta)$. Thus, Eq. (1) tends to raise the temperature when the instantaneous energy E rises above the average value $\widetilde{E}(\beta)$, or to lower the temperature when E falls under $\widetilde{E}(\beta)$. A derivation of Eq. (1) can be found in Appendix A. In implementation, β is first converted to its reciprocal $k_B T = 1/\beta$; then $k_B T$ is updated according to Eq. (1); finally the new β is computed from the reciprocal of the updated $k_B T$. We use $k_B T$ instead of β as the variable of integration, because in this way, the magnitude from the random noise is proportional to the temperature value $\sqrt{2}/\beta = \sqrt{2} k_B T$. The effect of the temperature change can be realized as a scaling of velocity or force in molecular dynamics. We repeat the process for each simulation step until the simulation ends.

In order to interpolate thermodynamic quantities on a continuous temperature range, we divide the entire range evenly into many small bins (β_i, β_{i+1}). The bin size is used for applying integral identities and is much smaller than the gap between neighboring sampling temperatures in traditional tempering methods, such as replica exchange or simulated tempering. During simulation, each bin i collects separate statistics on the potential energy and its variance along the trajectory for states with $\beta \in (\beta_i, \beta_{i+1})$. Statistics in dif-
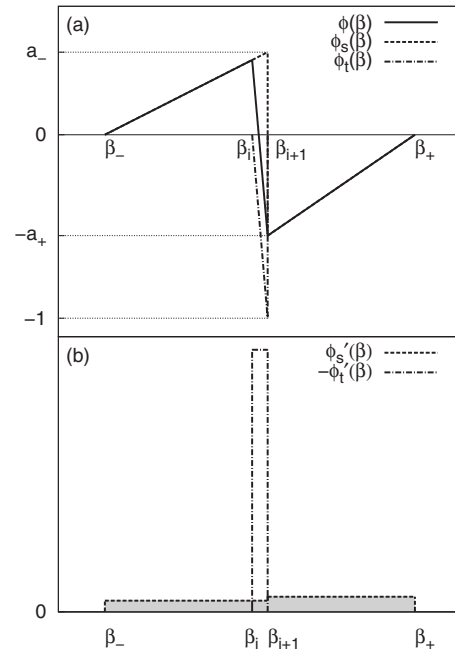

FIG. 1. (a) Schematic illustration of auxiliary functions $\phi(\beta)$, $\phi_s(\beta)$ and $\phi_t(\beta)$, which are used in the integral identity for estimating the thermal average energy $\widetilde{E}(\beta)$. The estimate computed in this way uses statistics from a large temperature window (β_-, β_+) instead of a single bin (β_i, β_{i+1}) and also avoids systematic bias. $\phi(\beta)$ is a combination of a smooth function $\phi_s(\beta)$ and a function $\phi_t(\beta)$ localized at (β_i, β_{i+1}). $\phi_s(\beta)$ is controlled by two parameters a_+ and a_- that satisfy $a_+ + a_- = 1$. (b) Schematic illustration of $\phi_s'(\beta)$ and $-\phi_t'(\beta)$. $\phi_s'(\beta)$ (shaded) spans over the whole temperature window (β_-, β_+) while $-\phi_t'(\beta)$ is localized in (β_i, β_{i+1}).

ferent bins are later combined together to form unbiased estimates of the average energy. For a complex molecular system, the adaptive averaging in Sec. II E can be used to improve the early convergence.

For statistical efficiency, $\widetilde{E}(\beta)$, $\beta \in (\beta_i, \beta_{i+1})$, is not calculated from the average energy of the bin, but instead from a large temperature window (β_-, β_+) containing the bin as

$$\widetilde{E}(\beta) = \sum_j \Delta \beta_j \langle E \rangle_j \phi_s'(\beta_j), \qquad (2)$$

where $\Delta \beta_j = \beta_{j+1} - \beta_j$ is the bin width; $\langle E \rangle_j$ is the average energy from the jth bin (β_j, β_{j+1}), $\phi_s'(\beta_j)$ is a modulating factor to ensure an unbiased estimate, see Fig. 1(b), and its computation is detailed in the next paragraph. The temperature window (β_-, β_+) is determined from the current temperature β in such a way that β is approximately at the center of the corresponding window (β_-, β_+).

The first step of computing $\widetilde{E}(\beta)$ is to determine two parameters a_+ and a_- by solving the following two equations:

$$a_+ + a_- = 1, \qquad (3)$$

$$\langle \phi(\beta') \langle \Delta E^2 \rangle_{\beta'} \rangle_{(\beta_-, \beta_+)} = 0. \qquad (4)$$

Here, a_+ and a_- are parameters defined in the equation $\phi(\beta) = \phi_s(\beta) + \phi_t(\beta)$, whose two components $\phi_s(\beta)$ and $\phi_t(\beta)$ are defined as

$$\phi_s(\beta) = \begin{cases} a_-(\beta - \beta_-)/(\beta_{j+1} - \beta_-), & \beta \in (\beta_-, \beta_{i+1}) \\ a_+(\beta - \beta_+)/(\beta_+ - \beta_{j+1}), & \beta \in (\beta_{i+1}, \beta_+), \end{cases} \quad (5)$$

and

$$\phi_t(\beta) = \begin{cases} -(\beta - \beta_i)/(\beta_{i+1} - \beta_i), & \beta \in (\beta_i . \beta_{i+1}) \\ 0, & \text{otherwise,} \end{cases} \quad (6)$$

respectively. Figure 1(a) schematically illustrates $\phi_s(\beta)$, $\phi_t(\beta)$, and $\phi(\beta)$. In Eq. (4), the inner bracket $\langle \cdots \rangle_{\beta'}$ is a configuration average of the energy fluctuation $\langle \Delta E^2 \rangle_{\beta'}$ at a fixed temperature β'. For $\beta' \in (\beta_j, \beta_{j+1})$, since the bin size is small, we compute the energy fluctuation from the states collected in the bin (β_j, β_{j+1}), and use it to approximate $\langle \Delta E^2 \rangle_{\beta'}$. The outer angular bracket $\langle \cdots \rangle_{(\beta_-, \beta_+)}$ denotes an average over temperature β' within the window (β_-, β_+), it is computed as a sum of the energy fluctuation from different bins within (β_-, β_+) with $\phi(\beta')$ being the coefficient of combination. After the averaging, Eq. (4) is a simple linear equation of a_+ and a_-. By solving Eqs. (3) and (4), a_+ and a_- are determined. In a physical solution, both a_+ and a_- are nonnegative. If the linear equations lead to a negative value for either a_+ or a_-, zero is used instead. This measure ensures the robustness of our estimator and is thus useful in early stages when the energy fluctuation $\langle \Delta E^2 \rangle_{\beta'}$ is unreliable. The determination of a_+ or a_- completely specifies $\phi_s(\beta)$ and its derivative $\phi_s'(\beta)$.

B. Generalized ensemble with a continuous temperature

We start our method by constructing a generalized ensemble in which the temperature β is a *continuous* variable in a given range $(\beta_{\min}, \beta_{\max})$. To sample the system correctly, we also need to make sure that the configurational distribution at a particular β is identical to that of the canonical ensemble at the same temperature. The aim of the method is to correctly populate states in the generalized ensemble and to extract thermodynamic properties for the entire temperature range.

First of all, the generalized ensemble is completely specified by an overall β-distribution $p(\beta)$. Once $p(\beta)$ is given, the complete distribution of atomic configuration X as well as the temperature β is also determined

$$p(\beta, X) = \frac{\exp[-\beta E(X)]}{Z(\beta)} p(\beta), \quad (7)$$

where $E(X)$ is the potential energy of configuration X, and $Z(\beta) = \int \exp[-\beta E(X)] dX$ is the canonical partition function. Equation (7) is due to the requirement of preserving a canonical distribution at each temperature. It is easily verified that $p(\beta)$ is recovered after we integrate $p(\beta, X)$ over all configurations.

The usefulness of the joint distribution $p(\beta, X)$ in Eq. (7) lies in that it specifies a temperature distribution under a fixed configuration X. For a fixed configuration X, we can perform temperature-space sampling according to $p(\beta, X)$ and replace $p(\beta)$ by any desired temperature distribution $w(\beta)$, which is fixed during simulation, in Eq. (7). If the configurational space is sampled according to the Boltzmann

distribution, the resulting overall temperature distribution after an infinitely long simulation trajectory, must be identical to the desired one $w(\beta)$.

However, the exact $Z(\beta)$ is usually unknown in advance. We therefore use a modified version of Eq. (7)

$$p(\beta, X) = \frac{\exp[-\beta E(X)]}{\widetilde{Z}(\beta)} w(\beta), \quad (8)$$

to conduct sampling in the temperature space, where an approximate partition function $\widetilde{Z}(\beta)$ is used in place of $Z(\beta)$. Note, if $\widetilde{Z}(\beta)$ differs from $Z(\beta)$ in Eq. (8), $w(\beta)$ is no longer the overall temperature distribution, but only a parameter that specifies $p(\beta, X)$, which is used in guiding the temperature-space sampling. The overall temperature distribution $p(\beta)$ is calculated from integrating the joint distribution $p(\beta, X)$ over configurations

$$p(\beta) = \int p(\beta, X) dX. \quad (9)$$

Using Eq. (8) in Eq. (9), we have

$$p(\beta) = \frac{Z(\beta)}{\widetilde{Z}(\beta)} w(\beta). \quad (10)$$

In simulation, $\widetilde{Z}(\beta)$ is adaptively adjusted and $w(\beta)$ is fixed. Therefore, the overall distribution $p(\beta)$ varies according to Eq. (10). Upon convergence, $\widetilde{Z}(\beta) \rightarrow Z(\beta)$, the overall temperature distribution $p(\beta)$ converges to the desired one $w(\beta)$.

Given a configuration X, as well as the joint distribution $p(\beta, X)$ Eq. (8), sampling along the temperature can be performed by the Langevin equation Eq. (1), in which the estimated average energy $\widetilde{E}(\beta)$ relates to the estimated partition function as $\widetilde{E}(\beta) \equiv -\partial \ln \widetilde{Z}(\beta)/\partial \beta$. One can demonstrate its correctness by solving the corresponding Fokker–Planck equation, see Appendix A.

The remaining task is to make sure the convergence of the estimated partition function $\widetilde{Z}(\beta)$ to the correct one $Z(\beta)$. As evidenced by Eq. (10), the current overall temperature distribution $p(\beta)$ is close to the desired one $w(\beta)$ only if the estimated the partition function $\widetilde{Z}(\beta)$ is sufficiently accurate. It is interesting to note that the Langevin equation Eq. (1) does not involve the estimated partition function $\widetilde{Z}(\beta)$ itself, but its derivative $\widetilde{E}(\beta)$ instead. We should therefore exploit this feature and focus on a technique that adaptively improves the estimate, $\widetilde{E}(\beta)$.

C. Unbiased estimate of a partition function

The asymptotic convergence of the partition function requires that $\widetilde{Z}(\beta)$ approaches to $Z(\beta)$ at any β in the range $(\beta_{\min}, \beta_{\max})$. In implementation, we divide the temperature range to many narrow bins (β_i, β_{i+1}). Thus we lower the requirement to that at any bin boundary β_i, $\widetilde{Z}(\beta_i)$ should equal to $Z(\beta_i)$. By choosing a small bin size, we can ensure that the deviation from $\widetilde{Z}(\beta)$ to $Z(\beta)$ is negligible for all practical purposes. Further, to remove the dependence on a

reference value of the partition function, the convergence condition is unambiguously expressed as a condition on ratios

$$\tilde{Z}(\beta_i)/\tilde{Z}(\beta_{i+1}) = Z(\beta_i)/Z(\beta_{i+1}). \qquad (11)$$

Equation (11) can be rearranged as[2]

$$\frac{Z(\beta)}{\tilde{Z}(\beta)}\Bigg|_{\beta_i}^{\beta_{i+1}} = \int_{\beta_i}^{\beta_{i+1}} d\beta \left[\frac{d\ln Z(\beta)}{d\beta} - \frac{d\ln \tilde{Z}(\beta)}{d\beta}\right]\frac{Z(\beta)}{\tilde{Z}(\beta)}$$

$$= \int_{\beta_i}^{\beta_{i+1}} d\beta [\tilde{E}(\beta) - \langle E\rangle_\beta]\frac{p(\beta)}{w(\beta)}$$

$$= \left\langle \frac{\tilde{E}(\beta) - \langle E\rangle_\beta}{w(\beta)}\right\rangle_i,$$

where we have used Eq. (10) on the second line; the inner bracket $\langle E\rangle_\beta$ denotes the average energy at a particular energy β; on the last line, we convert the integral over the small temperature bin (β_i, β_{i+1}) to a temperature average in the same bin, as represented by the outer bracket $\langle\cdots\rangle_i$, which is formally defined as $\langle A\rangle_i \equiv \int_{\beta_i}^{\beta_{i+1}} d\beta p(\beta)A$.

During simulation, if we adaptively enforce the right hand side of the above equation to be zero, i.e.,

$$\left\langle \frac{\tilde{E}(\beta)}{w(\beta)}\right\rangle_i = \left\langle \frac{\langle E\rangle_\beta}{w(\beta)}\right\rangle_i, \qquad (12)$$

the left hand side naturally vanishes as well, i.e., $\tilde{Z}(\beta_{i+1})/Z(\beta_{i+1}) = \tilde{Z}(\beta_i)/Z(\beta_i)$. In this way, the partition function as a function of the temperature can be obtained.[2]

We shall proceed by assuming a sufficiently small bin size and (i) $\tilde{E}(\beta)$ being a constant $\tilde{E}_i$ within a bin i, and correspondingly, $\ln\tilde{Z}(\beta)$ varies linearly with β, and (ii) $p(\beta)/w(\beta)$ can be treated as a constant. Equation (12) is then simplified as

$$\tilde{E}_i = \frac{\int_{\beta_i}^{\beta_{i+1}} d\beta \langle E\rangle_\beta p(\beta)/w(\beta)}{\int_{\beta_i}^{\beta_{i+1}} d\beta p(\beta)/w(\beta)} \approx \langle\langle E\rangle_\beta\rangle_i, \qquad (13)$$

and the ratio of the partition function is estimated as

$$\ln[Z(\beta_i)/Z(\beta_{i+1})] = \tilde{E}_i \Delta\beta_i,$$

where $\Delta\beta_i = \beta_{i+1} - \beta_i$.

A direct implication of Eq. (13) is that one can estimate $\tilde{E}_i$ from averages from statistics accumulated in bin i. Such an approach (which is similar to those used in the force averaging method[4]) is, although correct, ineffective when the bin size is small because the amount of statistics within a bin shrinks with the bin size. On the other hand, a large bin size can lead to a significant deviation from the desired temperature distribution, since we assumed a constant $\tilde{E}(\beta)$ within a bin. This dilemma can be resolved by using integral identities that remove the bin size dependence.

D. Estimators based on integral identities

We now present a method for drawing an unbiased estimate $\tilde{E}_i$ from a large temperature window instead of a small

temperature bin. The method removes bin size dependence by combining statistics from neighboring bins in a way that avoids systematic error. A similar technique of employing integral identities to improve statistics was previously used in improving statistical distributions.[5,6]

We aim at transforming the right hand side of Eq. (13) from an average over a single bin (β_i, β_{i+1}) to an average over a larger temperature range (β_-, β_+) that encloses the bin. To do so, we use the following integral identity:

$$0 = \phi(\beta)\langle E\rangle_\beta|_{\beta_-}^{\beta_+} = \int_{\beta_-}^{\beta_+} d\beta \phi'(\beta)\langle E\rangle_\beta - \int_{\beta_-}^{\beta_+} d\beta \phi(\beta)\langle\Delta E^2\rangle_\beta, \qquad (14)$$

where $\phi(\beta)$ is a function that vanishes at the two boundaries, i.e., $\phi(\beta_-) = \phi(\beta_+) = 0$; on the second line, we convert the difference between the two boundaries to an integral within the temperature range; the identity $\partial\langle E\rangle_\beta/\partial\beta = -\langle\Delta E^2\rangle_\beta$ from statistical mechanics is also used.

We choose $\phi(\beta)$ as a superposition of a smoothly varying $\phi_s(\beta)$ that spans over the entire window (β_-, β_+) and a localized function $\phi_t(\beta)$ limited within the bin $(\beta_i. \beta_{i+1})$, see Fig. 1 and Eqs. (5) and (6), in which a_+ and a_- are two non-negative parameters that sum to unity, i.e., $a_+ + a_- = 1$. Note at $\beta = \beta_{i+1}$, the sudden jump in $\phi_s(\beta)$ is exactly cancelled by that in $\phi_t(\beta)$, thus we can ignore the δ-functions in $\phi_s'(\beta)$ and $\phi_t'(\beta)$ in actual computation.

The purpose of the decomposition of $\phi(\beta)$ into $\phi_s(\beta)$ and $\phi_t(\beta)$ is to use the localized function $\phi_t'(\beta)$ to create an integral exactly equal to the right hand side of Eq. (13). In this way, the integral over the small bin is transformed to another one, but over a larger temperature window

$$\tilde{E}_i = \frac{1}{\Delta\beta_i}\int_{\beta_i}^{\beta_{i+1}} d\beta\langle E\rangle_\beta = \int_{\beta_-}^{\beta_+} d\beta[\phi_s'(\beta)\langle E\rangle_\beta - \phi(\beta)\langle\Delta E^2\rangle_\beta]. \qquad (15)$$

In early stages of simulation, the energy fluctuation $\langle\Delta E^2\rangle_\beta$ in the second term of the right hand side of Eq. (15) can be unreliable for a complex system. To avoid direct inclusion of energy fluctuation, we choose the combination parameters a_+ and a_- in such a way that the fluctuation term vanishes,

$$\int_{\beta_-}^{\beta_+} d\beta\phi(\beta)\langle\Delta E^2\rangle_\beta = 0. \qquad (16)$$

Equation (16) and $a_+ + a_- = 1$ yield a solution of a_+ and a_-, which also determines $\phi_s'(\beta)$. Thus we have,

$$\tilde{E}_i = \int_{\beta_-}^{\beta_+} d\beta\phi_s'(\beta)\langle E\rangle_\beta. \qquad (17)$$

Since $\phi_s'(\beta)$ is a constant within a single bin, it can be factored out of the integral when integrating each individual bin. The integral is thus converted to a sum over averages,

$$\widetilde{E}_i = \sum_j \phi_s'(\beta_j) \int_{\beta_j}^{\beta_{j+1}} d\beta \langle E \rangle_\beta = \sum_j \phi_s'(\beta_j) \Delta\beta_j \langle\langle E \rangle_\beta\rangle_j.$$

(18)

A comparison with Eq. (13) shows that Eq. (18) is merely a linear combination of $\widetilde{E}_i$'s obtained from different bins. The auxiliary function $\phi_s'(\beta)$ serves as a set of coefficients of combination, whereas Eq. (16) ensures the asymptotical convergence.

The above technique of extracting an estimate from an integral identity can be employed in computing other thermodynamic quantities, such as the average energy, heat capacity, and energy histogram. Thus, we are able to calculate these quantities at a particular temperature, even though the simulation is performed in an ensemble where the temperature is continuous.

The result for the average energy is most easily obtained. Consider a limiting case where $\beta_{i+1} \to \beta_i$ (with $\phi(\beta)$ modified accordingly), Eq. (15) immediately becomes an unbiased estimate of the thermal average energy $\langle E \rangle_{\beta_i}$ exactly at β_i.

Similarly, the heat capacity $C_V(\beta)$ can be calculated from the energy fluctuation as $\beta^2 \langle \Delta E^2 \rangle_\beta$. The energy fluctuation at a particular temperature β is computed as

$$\langle \Delta E^2 \rangle_\beta = \int_{\beta_-}^{\beta_+} d\beta' \, \phi'(\beta') \langle \Delta E^2 \rangle_{\beta'} - \int_{\beta_-}^{\beta_+} d\beta' \, \phi(\beta') \langle \Delta E^3 \rangle_{\beta'},$$

(19)

where we have substituted $-\langle \Delta E^3 \rangle_\beta$ for $\partial \langle \Delta E^2 \rangle_\beta / \partial\beta$.

Estimating the energy histogram requires additional care in choosing the function $\phi(\beta)$, in order to ensure that the histogram is non-negative everywhere. It is shown in Appendix D that the optimal estimate of an energy histogram at temperature β is

$$h_\beta(E, E+\Delta E)$$

$$\equiv \int_E^{E+\Delta E} p_\beta(E') dE'$$

$$\approx \frac{\int_{\beta_-}^{\beta_+} \int_E^{E+\Delta E} p(\beta', E') dE' d\beta'}{\int_{\beta_-}^{\beta_+} w(\beta') \widetilde{Z}(\beta) / \widetilde{Z}(\beta') \exp[-(\beta'-\beta)(E+\Delta E/2)] d\beta'},$$

(20)

where on the first line, the energy histogram $h_\beta(E, E+\Delta E)$ at β is defined from the constant temperature energy distribution $p_\beta(E)$ at the same temperature; on the second line, it is converted to an average over the joint temperature-energy $p(\beta, E)$ distribution in the generalized ensemble, which can be measured from the temperature-energy histogram in simulation trajectory. Note, for simplicity, we have assumed that the energy bin size is small compared with the energy fluctuation of a typical Boltzmann distribution. Equation (20) resembles the result from the multiple histogram method,[7] and therefore can be treated as its counterpart in a continuous temperature ensemble.

For a quantity whose statistics are not fully accumulated during simulation (e.g., it is calculated from periodically saved trajectory snapshots after simulation), we use the following reweighting formula to obtain its value at a particular temperature β

$$\langle A \rangle_\beta = \frac{\int_{\beta_-}^{\beta_+} A \exp[-(\beta-\beta')E] \widetilde{Z}(\beta') / \widetilde{Z}(\beta) d\beta'}{\int_{\beta_-}^{\beta_+} \exp[-(\beta-\beta')E] \widetilde{Z}(\beta') / \widetilde{Z}(\beta) d\beta'},$$

(21)

where $\exp[-(\beta-\beta')E] \widetilde{Z}(\beta') / \widetilde{Z}(\beta)$ serves as a weighting function for borrowing statistics from β' to β. Note, Eq. (21) is unbiased estimator even if $\widetilde{Z}(\beta)$ contains error.

E. Adaptive averaging

We now introduce an adaptive averaging scheme for accelerating the convergence in initial stages of a simulation. Since we start from zero statistics, the error in initially estimated $\widetilde{E}(\beta)$ can lead to a slow random walk in the temperature space. The adaptive averaging scheme[2] overcomes the problem by assigning larger weights toward recent statistics, and thus encourages a faster random walk in the temperature space.

Usually, for a statistical sample of size n, the average of a quantity A is calculated as an arithmetic mean $\langle A \rangle = S_A^{(n)} / S_1^{(n)}$, where $S_A^{(n)} = \sum_{i=1}^n A_i$ is the sum of A and $S_1^{(n)}$ is the sample size n. In simulation, since statistics is collected along the trajectory, the sample size n increases as simulation progresses. To increase the weight toward recent A_i's, we redefine $S_A^{(n)}$ and $S_1^{(n)}$ as

$$S_A^{(n)} = A_1 \gamma^{n-1} + A_2 \gamma^{n-2} + \cdots + A_n,$$

$$S_1^{(n)} = \gamma^{n-1} + \gamma^{n-2} + \cdots + 1,$$

where $\gamma (<1)$ is used to gradually damp out old statistics. Such an average can be easily implemented as recursions,

$$S_A^{(n)} = \gamma S_A^{(n-1)} + A_n,$$

$$S_1^{(n)} = \gamma S_1^{(n-1)} + 1.$$

(22)

However, if a constant γ is used, an average derived from Eq. (22) does not asymptotically reduce its error due to that the sample size S_1 ultimately saturates to a fixed value $1/(1-\gamma)$. Therefore, we use

$$\gamma = 1 - C_\gamma / n,$$

(23)

where C_γ is a numerical constant, to gradually reduce the difference between γ and 1.0. In this way, both a fast random walk in early stages and an asymptotical convergence can be achieved.

III. NUMERICAL RESULTS

A. Ising model

As the first example, we test our method on a 32×32 Ising model, which is a nontrivial system with exactly known thermodynamic properties.[8] Results from the alternative method of estimating $\widetilde{E}(\beta)$, described in Appendix B, are also included. Parameters common to the two methods were set to be the same.
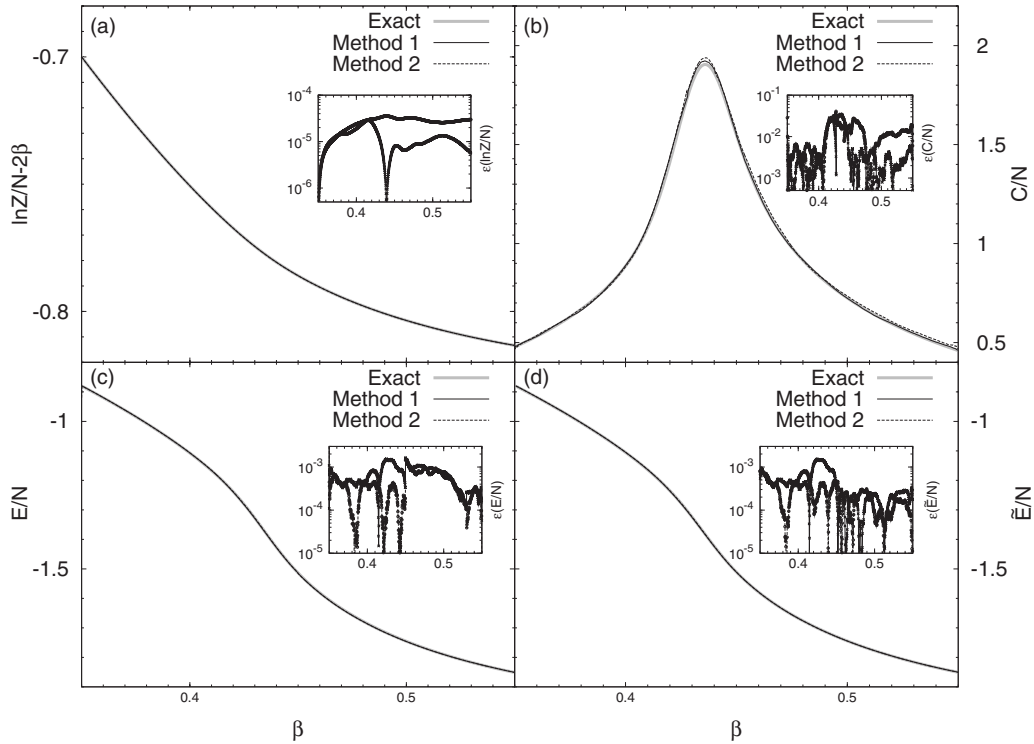
FIG. 2. Thermodynamic quantities as functions of β for the 32×32 Ising model. Results from the method introduced in Sec. II are labeled as method 1 (thin solid lines, cross for errors), and those from the method introduced in Appendix B are labeled as method 2 (dashed lines, circles for errors); (a) the partition function, (b) the heat capacity, (c) the average energy, and (d) $\tilde{E}(\beta)$.

The temperature range was $\beta \in (0.35, 0.55)$ or $T = (1.818, 2.857)$, which covered the critical temperature $T \approx 2.27$ of the phase transition, and the temperature distribution was defined by a constant $w(\beta)$, or a flat-β histogram. The bin size for collecting statistics was $\delta\beta = 0.0002$, and thus there were 1000 bins in the entire temperature range. For each temperature bin (β_i, β_{i+1}), the temperature window for applying integral identities was given by $(\beta_-, \beta_+) = (\beta_i - \Delta\beta, \beta_{i+1} + \Delta\beta)$ with $\Delta\beta = 0.02$ in evaluating Eq. (18); thus 201 bins were included in a window (except at temperatures near the boundaries, where the largest possible value of $\Delta\beta$ is used). The Langevin equation was integrated after every 100 Monte Carlo moves, with an integrating step $\Delta t = 2.0 \times 10^{-5}$. The whole simulation stopped after 10^6 Monte Carlo moves per site.

The calculated partition function, average energy and heat capacity are shown in Fig. 2. The thin solid lines and dashed lines are for the method introduced in Sec. II D (method 1), and that introduced in Appendix B (method 2), respectively. In most cases, both results for the partition function and the average energy coincide with exact results,[8] see Figs. 2(a), 2(c), and 2(d). Even for the heat capacity, which is the second order derivative of the partition function and is harder to compute, the deviations in both cases are small. This indicates that our method is unbiased and it can produce exact thermodynamic quantities asymptotically.

We now show that the correct energy distribution can be reconstructed from Eq. (20). Since the energy levels are discrete and the energy bin size we used was equal to the smallest gap between energy levels, a normalized energy histogram is equivalent to the energy distribution. During

simulation, the current potential energy was registered into the histogram every 100 Monte Carlo moves, thus there were roughly 10^5 samples in the entire histogram. The reconstructed energy distributions at a few representative temperatures are shown in Figure 3. The temperature window for estimating the energy distribution at β was $(\beta - \Delta\beta, \beta + \Delta\beta)$ with $\Delta\beta = 0.02$. One can see a good agreement between the integral identity Eq. (20) (the thin solid line), and the exact distribution (thick solid line), which was computed from the exact density of states.[9] For comparison, distributions constructed by averaging energy distributions of two adjacent
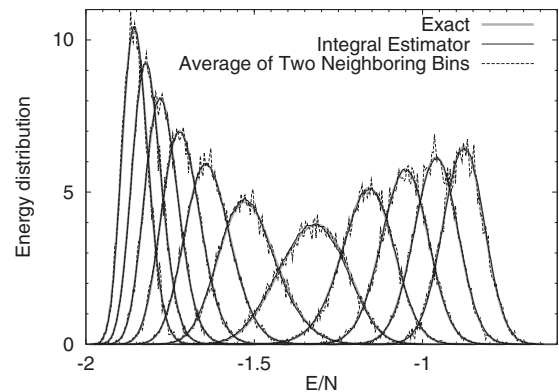

FIG. 3. Reconstructed energy distribution at a few temperatures using Eq. (20) with a window size $\Delta\beta = 0.02$ for the 32×32 Ising model. For comparison, we also show the resulting distributions from averaging energy distributions of two adjacent temperature bins, with the bin size $\delta\beta = 0.0002$. The energy distributions constructed from Eq. (20) with a larger window are more precise than those from simple averages of two adjacent bins.

bins are shown as dashed lines. It is apparent that the simple average yields more noisy results than the integral identity Eq. (20). The reduced error is due to Eq. (20) being able to access statistics from a much larger temperature window than from two neighboring bins.

B. Lennard-Jones system

As the second example, we test the method on an 864-particle Lennard-Jones liquid. In reduced units, the Lennard-Jones potential for a particle pair separated by distance r is

$$u(r) = 4\left(\frac{1}{r^{12}} - \frac{1}{r^6}\right),$$

where the units of energy, mass, and length are 1.0. In the simulation, the density was 0.8; the cutoff was 2.5; and the temperature range was $\beta = (0.48, 1.02)$, corresponding to $T = (0.98, 2.08)$. We used Monte Carlo to generate configuration changes. In each step, a random particle was displaced randomly in each of x, y, and z directions according to a uniform distribution in $(-0.1, 0.1)$. After a Monte Carlo step, we applied the Langevin equation Eq. (1) with an integration time step $\Delta t = 0.0002$. The system was simulated for 10^5 sweeps (a sweep=a step per particle). Coordinates were saved every ten sweeps for data analysis. The overall temperature distribution $w(\beta)$ was proportional to $1/\beta$ (the choice of optimal $w(\beta)$ is discussed in Appendix C). The temperature bin for collecting statistics was $\delta\beta = 0.0005$. The window size $\Delta\beta$ for applying the integral identity Eq. (18) was 0.05.

The simulation results for the estimated average energy $\langle E \rangle_\beta$ and $\widetilde{E}(\beta)$ are shown in Fig. 4(a). Due to a small bin size, the difference between the two is invisible. The dots on the figures represent results from independent constant temperature simulations, each of which uses the same amount of simulation time. A good agreement between the two indicates that our method can produce exact thermodynamic quantities for the entire temperature spectrum with less simulation time.

Furthermore, we reconstructed the radial distribution function $g(r)$ at a particular temperature according to Eq. (21). The reconstruction was performed after simulation and was based on the saved coordinates along the trajectory. From Fig. 4(b), we see a good agreement between reconstructed $g(r)$'s and those from constant temperature simulations at two different temperatures $\beta = 0.5$ and 1.0, or $T = 2.0$ and 1.0, respectively.

IV. APPLICATIONS IN FOLDING SMALL PROTEINS

In this section, we report applications of the method in folding of several small proteins. The method was implemented in a modified GROMACS 4.0.5,[10] using AMBER force field ports[11] with TIP3P water model.[12] In all cases, the particle-meshed Ewald method[13] was used for handling long range electrostatic interaction, and the velocity-rescaling method was used as thermostat.[14] For constraints, we used the SETTLE algorithm for water molecules,[15] and the parallel LINCS algorithm[16] for proteins. Since proteins drastically changed their configurations during simulations, dynamic
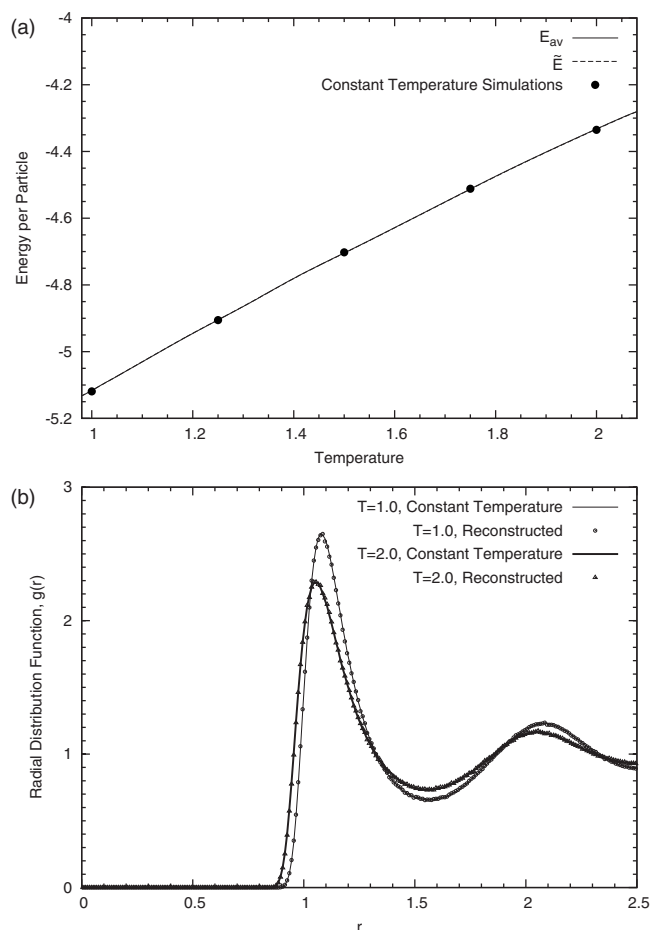

FIG. 4. (a) $\widetilde{E}(\beta)$ of an 864-particle Lennard-Jones system. As a comparison, the values of the average energy from several constant temperature simulations are shown as dots. (b) The reconstructed radial distribution functions of an 864-particle Lennard-Jones system at two selected temperatures $T = 1.0$ and 2.0 are shown as points. As a comparison, the corresponding radial distribution functions from independent constant temperature simulations are shown as lines.

load balancing was turned on when using domain decomposition.

For configuration-space sampling under a fixed temperature, the canonical ensemble, i.e., the constant (N, V, T) ensemble, was used. A 10 Å cutoff was used for Lennard-Jones interaction, electrostatic interactions and neighboring list. Since the temperature in our method is a variable, the temperature change was realized by scaling the force according to $\mathbf{F}' = (\beta/\beta_0)\mathbf{F} = (T_0/T)\mathbf{F}$, with $\mathbf{F}'$ and $\mathbf{F}$ being the scaled and the original force, respectively. In this way, the scaled potential energy transforms a canonical ensemble at T_0 to another temperature T, for sampling in the configurational space. On the other hand, the thermostat temperature T_0, which controls the kinetic energy, is unaffected and can be maintained at a fixed value $T_0 = 480$ K. Three separate thermostats were applied to protein, solvent and ion groups, and the coupling time τ_T for the thermostats was 0.1 ps.

The time step for molecular dynamics was 0.002 ps. The center of mass motion was removed every step. Trajectory snapshots were saved every 2 ps. Since the force and energy calculation was much more time consuming than estimating $\widetilde{E}(\beta)$, we applied the Langevin equation in every molecular
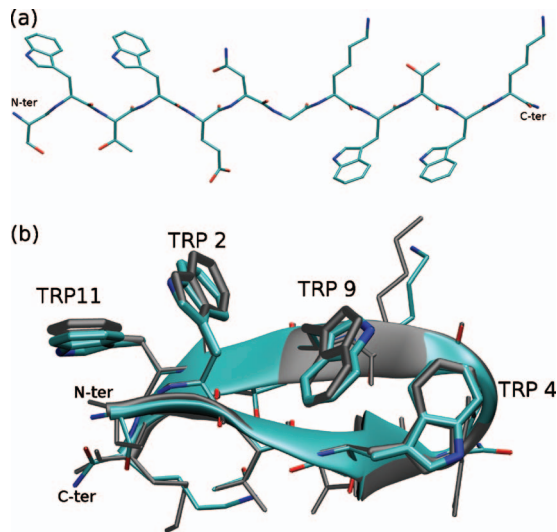
FIG. 5. Trpzip2: (a) the initial fully extended conformation, (b) a typical folded structure from trajectory 2. The C_α-RMSD and heavy atom RMSD are 0.25 and 1.08 Å, respectively. Gray: reference structure (PDB ID: 1LE1).

dynamics step using an integration step of $\Delta t = 10^{-4}$. The parameter C_γ in the adaptive averaging Eq. (23) was 0.1 in all cases.

A. Trpzip-2

The first system is a 12 amino acid β-hairpin tryptophan zipper, whose Protein Data Bank (PDB) ID code is 1LE1, and its sequence is SWTWENGKWTWK.[17] A unique feature of this short hairpin is that its four tryptophan side chains are locked against each other to stabilize the structure.

Previously, the system was intensively studied both in explicit solvent[18] and in implicit solvent.[19,20] However, *de novo* folding in explicit solvent from an extended chain, see Fig. 5(a), is much more challenging. We used AMBER99SB as the force field, which is a modified version of AMBER99 force field[21] with updated φ and ψ torsions.[22]

We report four simulation trajectories. All of them reached atomic accuracy within in a time scale of 1 μs. In all the cases, we used a cubic $45 \times 45 \times 45$ Å^3 box, filled with 2968 water molecules as well as two Cl$^-$ ions. The temperature range was $\beta \in (0.20, 0.41)$, which corresponded to $T = 293.6 - 601.9$ K temperature range. The grid spacing for Fourier transform was 1.15 Å, and the alpha parameter was 0.3123 Å^{-1} for the Ewald method. In application of integral identities equations (18)–(20), the temperature

windows size was 8% of temperature value, e.g., at $\beta = 1.0/(k_B 500 \text{ K}) \approx 0.241$ the temperature window is $(\beta - \Delta\beta, \beta + \Delta\beta)$, with $\Delta\beta = 4\% \times \beta \approx 0.010$, which can be translated to $T = (480.8, 520.8)$ K. At boundaries, we used the largest possible size that allows a symmetrical window.

A typical folded structure is shown in Fig. 5(b), with its root mean square deviations (RMSD) for alpha-carbon (C_α) and heavy atoms being 0.25 and 1.08 Å, respectively. The lowest C_α-RMSD and heavy atom RMSD found in the four trajectories are listed in Table I, in which we also list the approximate first time of stably reaching the atomically accurate native structure (the criteria were C_α-RMSD <0.5 Å). It is interesting to note that even among structures with lowest RMSDs, tryptophan side chains can still adopt different conformations. For example, in trajectories 1 and 3, we found nativelike structures with one of the tryptophan side chain (TRP9 or TRP4) flipped 180° with respect to the native configuration. This suggests that the free energy change involved in flipping a TRP side chain is small.

In Fig. 6, C_α-RMSD, radius of gyration, instantaneous temperature, and potential energy along trajectory are shown in (a)–(d), respectively, for the two independent simulations, trajectory 1 and 3. In trajectory 1, the folded state was reached within 20 ns. This suggests the possibility of fast folding. However, in other trajectories, it took longer for the system to reach the native structure. In trajectory 3, upon reaching of the native structure, the system lingered in a state of low-temperature, low-energy state, and small radius of gyration for about 100 ns. This corresponds to the fact that the folded state has a lower energy than the unfolded state and thus occupies a larger fraction in a low temperature Boltzmann distribution. This feature serves as a signature of the system reaching a native structure, and can be useful in folding prediction where the native structure is unknown.

Figure 7(a) shows the distribution along the C_α-RMSD at three different temperatures calculated from trajectory 3. The distribution demonstrates two well-separated peaks, corresponding to roughly defined folded and the unfolded states, respectively. The average C_α-RMSD from the native structure is roughly 0.8 and 5.5 Å for folded and unfolded states, respectively. As the temperature increases, the first peak gradually diminishes, whereas the second peak dominates.

The folding temperature can be computed by assuming a two-state model of folding. The folding fraction P was first calculated as the fraction of configurations with C_α-RMSD less 2.0 Å at different temperatures. The curve $P(\beta)$ as a

TABLE I. Lowest RMSDs and folding time from four independent folding trajectories of trpzip2. Snapshots of reaching the lowest C_α-RMSDs can differ slightly from those of reaching the lowest heavy atom RMSDs.

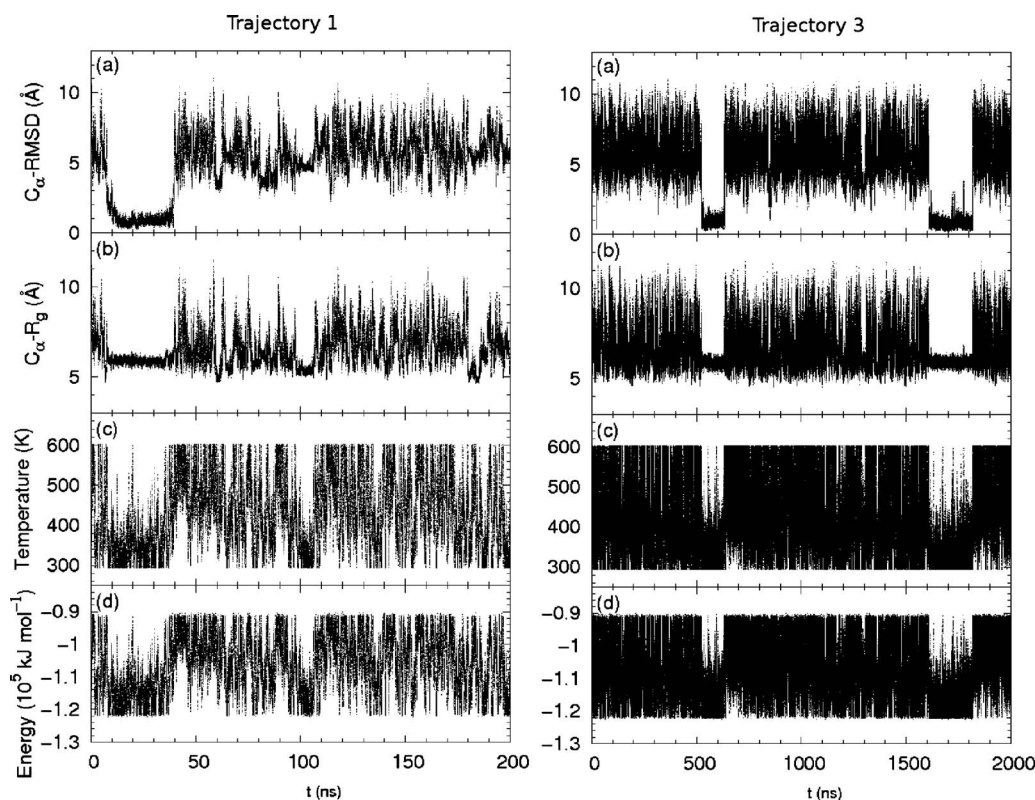| Traj. ID	Lowest C_α-RMSD (Å)	Lowest heavy atom RMSD (Å)	First time of stably reaching atomic accuracy of the native state (ns)
1	0.20	1.00	20
2	0.25	0.84	520
3	0.20	0.88	530
4	0.25	1.31	1080

FIG. 6. Trpzip2: quantities along two independent simulation trajectories. Left: the first 200 ns of a fast-folding trajectory (trajectory 1). Right: 2 μs of trajectory 3. Panels from top to bottom: (a) C_α-RMSD from native structure, (b) C_α radius of gyration, (c) temperature, and (d) potential energy.

function of temperature β was then fitted against a two-state model in the range $\beta \in (0.24, 0.4)$, equivalently $T \in (300.9 \text{ K}, 501.5 \text{ K})$,

$$P = \frac{P_0}{1 + \exp[\Delta E(\beta_\mathrm{m} - \beta)]}, \qquad (24)$$

where the values of the three parameters were determined as $P_0 = 0.61$, $\Delta E = 38.96$ kJ mol^{-1}, and $\beta_\mathrm{m} = 0.350$ kJ^{-1} mol by regression, see Fig. 7(b). Here, ΔE is the energy difference between the folded state and the unfolded state; β_m is the melting temperature. In the simplest two-state model, the folding fraction increases monotonically to unity as the temperature decreases to zero. Here, the maximal fraction was changed from 1.0 to an adjustable parameter P_0, which helped fitting the calculated $P(\beta)$ to the two-state model. Physically, such a modification implies the existence of many configurations with energy similar to that the native one but with different structures in our simulation trajectory.

If the volume change during folding is ignored, the enthalpy change between the folded state and the unfolded state is roughly equal to the energy change $\Delta H \approx 39.0$ kJ mol^{-1}. The entropy difference of the folded and unfolded states is $\Delta S = \beta_\mathrm{m} \Delta E = 113$ J mol^{-1}. These values are relatively small compared with the experimental values $\Delta H^\mathrm{exp} = 70.2$ kJ mol^{-1}, $\Delta S^\mathrm{exp} = 203.3$ J mol^{-1}.[17] However, the estimated folding temperature from our calculation 344 K is close to the experimental result 345 K.[17] From trajectory 3, which yields the highest folding fraction at 300 K among four trajectories, the fraction of folded states at 300 K is roughly 55%, which still differs significantly from the ex-

perimental value 91%.[17] For the other three trajectories, the calculated fractions are even smaller. The difference between our calculation and experiments were likely due to insufficient sampling and/or force field inaccuracy.

Figure 8 shows the heat capacity versus temperature from Eq. (19). The difference between three independent trajectories was small, suggesting thermodynamic properties of the entire system, protein and water, reaching convergence. It is interesting that in our simulation, the heat capacity is not a constant, but inversely proportional to the temperature as $C_V \approx 4.2 \times 10^3/T$ kJ mol^{-1} K^{-1}.

Figure 9(a) shows the joint distribution of the radius of gyration R_g (calculated from C_α atoms) and the C_α-RMSD. Generally the two measures are positively correlated. However, the smallest radius of gyration does not occur at the native structure for the hairpin, whose R_g is roughly around 5.8 Å. A non-native structure, on the other hand, had a smaller R_g around 4.9 Å, but a larger RMSD around 3 Å. Figure 9(b) shows the joint distribution of temperature and energy. A typical temperature fluctuation is only around 5 K, which is roughly the magnitude of temperature gap in other tempering methods based on a discrete temperature, such as replica exchange. On the other hand, the temperature window size used in our simulation was much larger. This suggests that our method could more efficiently use statistics to facilitate the temperature-space random walk.

B. Trp-cage

The second application is a 20 amino acid alpha helical protein, tryptophan cage (trp-cage).[23] The PDB code is
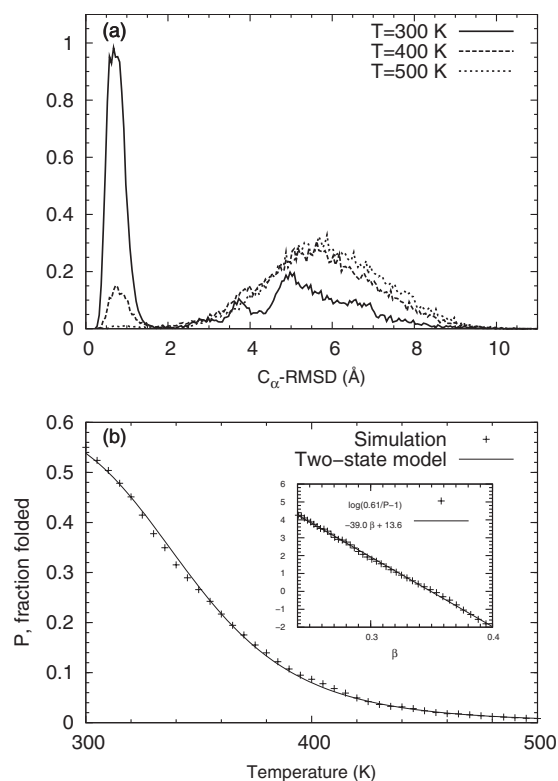
FIG. 7. Trpzip2: (a) distribution along the RMSD from the native structure at three temperatures 300, 400, and 500 K, calculated from trajectory 3 and (b) fraction P of the folded state vs the temperature. Inset: linear fitting of $\log(P_0/P-1)$ vs β according to the two-state model.

1L2Y, and the amino acid sequence is NLYIQWLKDG-GPSSGRPPPS. The system was extensively studied by experiments[23,24] as well as by various sampling techniques, either in explicit solvent[25,26] or in implicit solvent.[20,27] We again used AMBER99SB as the force field.[22]

We simulated the system in a cubic $46 \times 46 \times 46$ Å^3 box, filled with 3161 water molecules and two Cl$^-$ ions. The grid spacing for Fourier transform was 1.19 Å, and the alpha parameter was 0.3123 Å^{-1}. The initial structure was an open chain see Fig. 10(a), which was constructed by bending a fully extended chain to fit into the box.

We report three independent 1 μs simulation trajectories. The temperature range for trajectory 1 was β
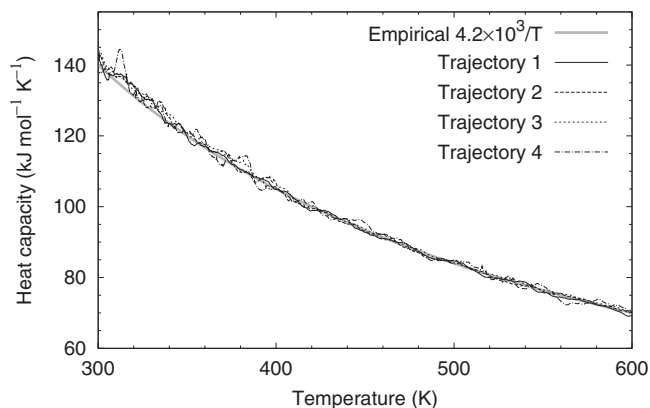

FIG. 8. Trpzip2: the heat capacity, computed from four independent trajectories. Bold solid line: empirical formula $C_V \approx 4.2 \times 10^3/T$.
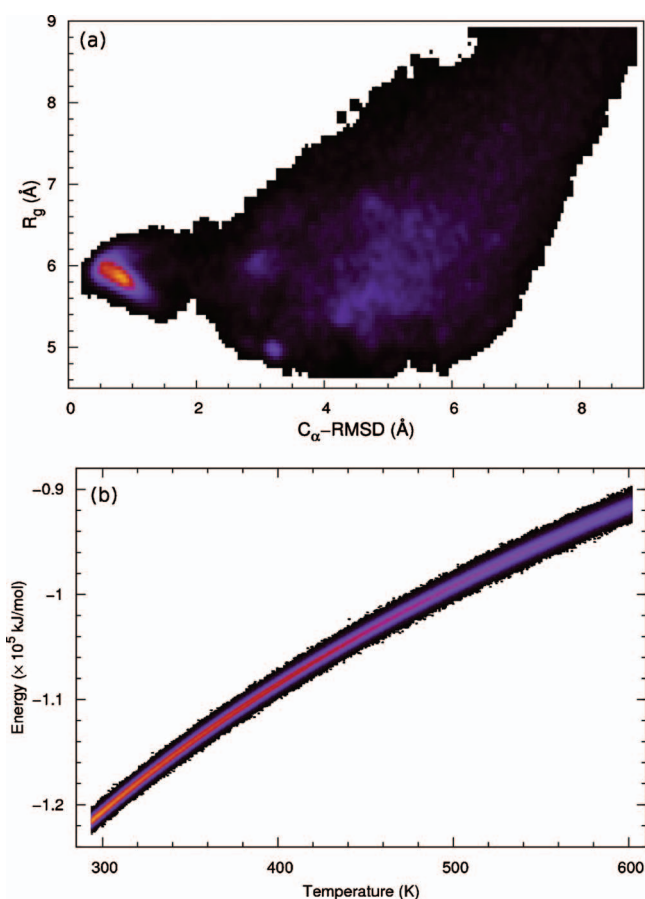

FIG. 9. Trpzip2: joint distributions of (a) the radius of gyration vs C$_\alpha$-RMSD and (b) the potential energy vs temperature. A brighter color represents a higher population density.

$\in (0.24, 0.41)$, or $T \in (293.6 \text{ K}, 501.5 \text{ K})$. In trajectories 2 and 3, we used a larger temperature range $\beta \in (0.20, 0.41)$, which covered a 293.6 K $\sim$ 601.9 K temperature range. The temperature bin size $\delta\beta$ was 0.0005, 0.0005, and 0.0002 respectively. In applying integral estimators, the temperature windows size was 10%, 10%, and 8% of temperature value for trajectories 1, 2, and 3, respectively. In all cases, we used the alternative estimator introduced in Appendix B.

All three simulations independently reached atomically accurate native configurations. A typical folded structure in trajectory 1 is shown in Fig. 10(b). The C$_\alpha$-RMSDs from the three trajectories were 0.43, 0.48, and 0.44 Å, respectively. The average C$_\alpha$-RMSD for the native structure was around 0.8 Å. The lowest RMSDs for all heavy atoms were 1.34, 1.47, and 1.46 Å, respectively.

The C$_\alpha$-RMSD, radius of gyration, instantaneous temperature, and potential energy along trajectory are shown in panels (a)–(d), respectively, of Fig. 11, for the three independent simulations. In each trajectory, there were two folding events reaching an atomic accuracy, e.g., for trajectory 2, the native structure was reached at 350 and 560 ns. As in the trpzip2 case, the system stayed around a low-temperature and low-energy state for 30–100 ns upon reaching the native state. For the folding speed, it appears that simulations 2 and 3, which used a higher roof temperature T_{max}=600 K, tended to reach the native state sooner than simulation 1, whose roof temperature T_{max}=500 K was lower. However,
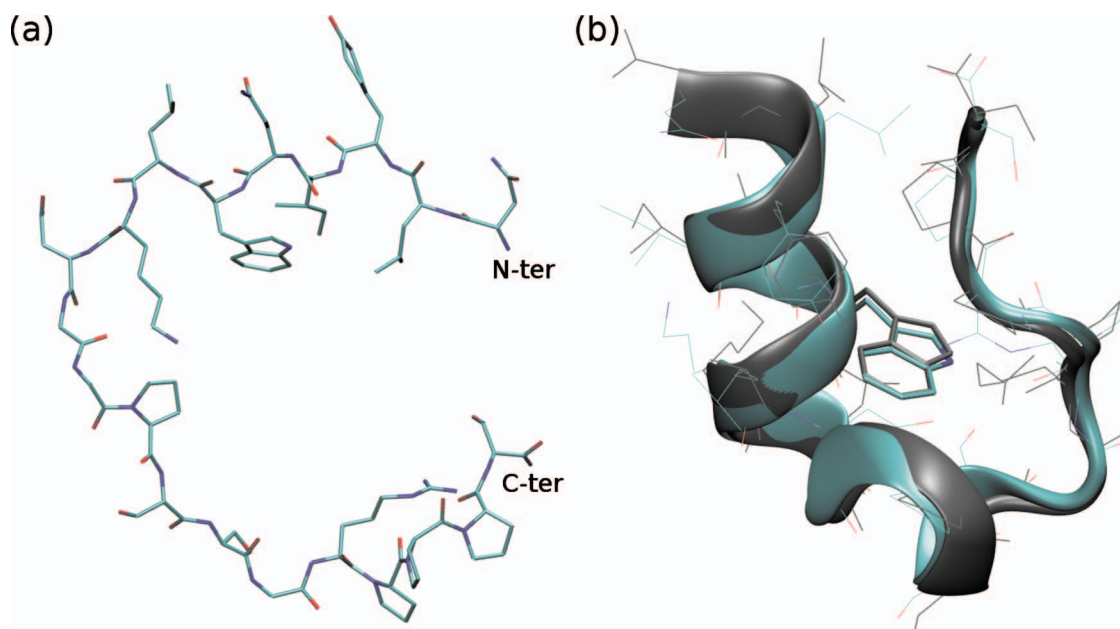
FIG. 10. Trp-cage: (a) the initial fully extended structure, (b) a typical folded structure. The C_α-RMSD and the all heavy atom RMSD are 0.44 and 1.54 Å, respectively. Gray: reference structure (PDB ID: 1L2Y).

in simulation 1, the system was able to stay in the native states longer and performed a more detailed sampling at the low temperature end.

Trajectory 1 yielded the largest fraction of the folded state at 300 K, 19%, which is still much lower than experiment value, 70%.[23,28] By fitting the folding fraction, computed from the fraction of states with C_α-RMSD less than 2.2 Å cutoff, to the two-state formula Eq. (24), the parameters are $P_0 = 0.50$, $\Delta E = 21.1$ kJ mol^{-1}, and $\beta_m = 0.367$ kJ^{-1} mol. The enthalpy change is thus $\Delta H \approx 21$ kJ mol^{-1}, which is relatively small compared with the experimental values

$\Delta H^{\text{exp}} = 56.2$ kJ mol^{-1}.[28] The estimated folding temperature from our calculation is 328 K, which is slightly higher than the experimental result 315 K.[28]

C. Villin headpiece

Our last application was the villin headpiece, a 36 residue alpha-helical protein HP36. The PDB ID is 1VII, and the amino acid sequence is **M**LSDEDFKAVFGMTRSAFAN-LPLWKQQ**N**LKKEKGLF. This system was the first protein partially folded in explicit solvent.[29] Recently, a high reso-
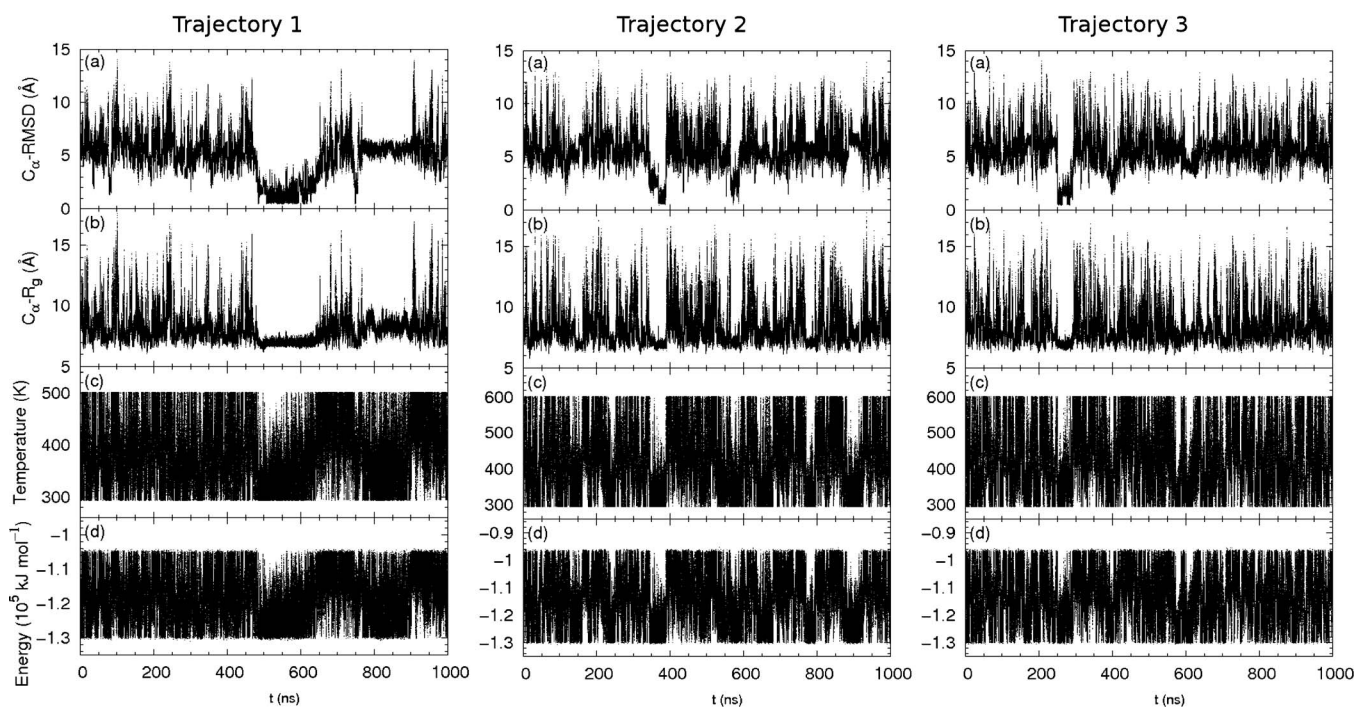

FIG. 11. Trp-cage: three independent trajectories (a) C_α-RMSD from native structure, (b) C_α radius of gyration, (c) temperature, and (d) potential energy.

lution x-ray structure of slightly modified protein HP35, PDB ID 1YRF, was published.[30] The sequence of HP35 is LSDEDFKAVFGMTRSAFANLPLWKQQ**H**LKKEKGLF, in which the N-terminal methionine (MET) of the original sequence was chopped off and the 28th residue, an asparagine, was replaced by a hisidine. Both sequences were studied in literature, both by simulations,[31–34] and in experiments.[35,36]

However, there is a significant difference between the NMR structure of HP36 and the x-ray structure of HP35, about 1.62 Å difference in terms of C_α-RMSD. Further, while molecular dynamics simulations on HP35 reached an atomic accuracy in implicit solvent[32] as well as in explicit solvent,[34] simulations on HP36, especially in explicit solvent, yielded relatively poorer results. The difference between the two structures could be due to (1) the intrinsic difference between HP35 and HP36 or (2) differences in the two experimental techniques, NMR versus x ray.[30]

In this simulation, we used AMBER03 (Ref. 37) as the force field, which was previously used to fold HP35 to an atomic accuracy in implicit solvent.[32] Another reason of choosing AMBER03 instead of AMBER99SB was that the latter force field might slightly disfavor helical conformations.[38] To reduce simulation size, we used a dodecahedron simulation box with edge length being 24.1 Å to accommodate the protein as well as 3343 water molecules, and two Cl$^-$ ions. The volume of the box was $53.3 \times 53.3 \times 37.6$ Å$^3 = 1.069 \times 10^5$ Å^3. The initial conformation of the protein was a fully extended chain, which was bent from a linear chain to fit into the box, see Fig. 12(a). The grid spacing for Fourier transform was 1.19 Å and the alpha parameter was 0.3123 Å^{-1}.

We report two independent simulation trajectories, each of 2 μs. The temperature range for trajectory 1 was $\beta \in (0.18, 0.41)$, or $T \in (293.6 \text{ K}, 668.7 \text{ K})$; that for trajectory 2 was $\beta \in (0.20, 0.41)$, or $T \in (293.6 \text{ K}, 601.9 \text{ K})$. The temperature bin size was 0.0002 in both cases. The integral estimator introduced in the Sec. II was used, and the temperature windows size was 8% of current temperature value.

The alpha-carbon root mean square deviation can be calculated from the NMR structure as well as the x-ray structure. In both cases, the N-terminal MET and leucine (LEU) as well as the C-terminal phenylalanine (PHE) are not included in our calculation.[32] Due to the flexibility of the C-terminal and the N-terminal helix, we also calculate the C_α-RMSD for residues 9–32 of HP36 as in the literature,[29,31] denoted as RMSD$_{\text{core}}$ here. Note, in terms of C_α-RMSD$_{\text{core}}$, the NMR reference structure differs from the x-ray reference structure by 0.87 Å.

The lowest C_α-RMSDs reached in the two simulation trajectories are listed in Table II. The lowest C_α-RMSD$_{\text{core}}$ from the NMR structure are 0.72 and 0.73 Å, for the two trajectories, respectively. These figures are smaller than those from a previous study, in which the lowest RMSD$_{\text{core}}$ was around 1.5 Å.[31] In both trajectories, the best folded structures are more similar to the x-ray structure than to the NMR structure. The average C_α-RMSD and C_α-RMSD$_{\text{core}}$ from the NMR structure are 1.30 and 1.95 Å, respectively. In comparison, the two figures drop to 0.90 and 1.10 Å if the reference is switched to the x-ray structure.
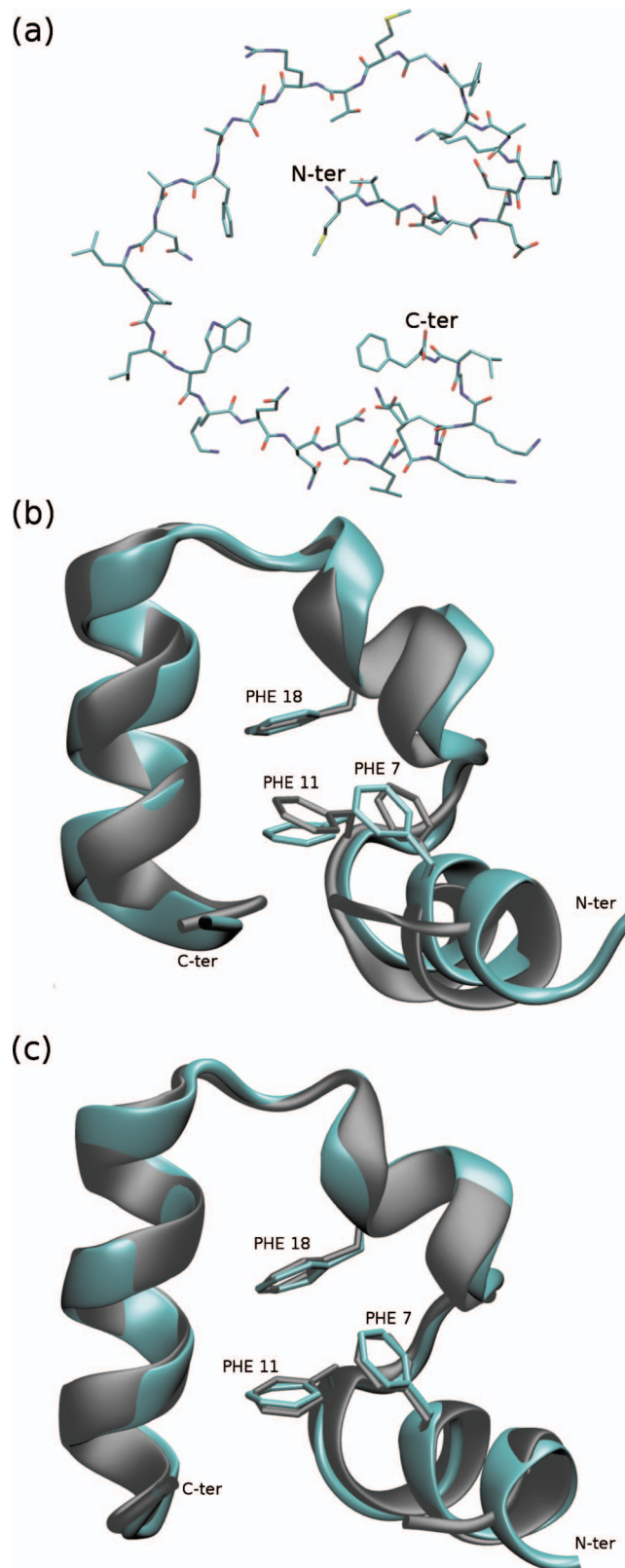

FIG. 12. Villin headpiece: (a) the initial fully extended structure, (b) a typical folded structure compared with an NMR reference structure (gray, PDB ID: 1VII, C_α-RMSD: 1.15 Å), (c) a typical folded structure compared with an x-ray structure (gray, PDB ID: 1YRF, C_α-RMSD: 0.47 Å), and the N-terminal is not shown due to the sequence difference.

TABLE II. Lowest RMSDs in angstrom reached in two folding trajectories of villin headpiece. Snapshots of reaching the lowest C_α-RMSDs can differ slightly from those of reaching the lowest heavy atom RMSDs. The reference structures (either the NMR or x-ray structure) are denoted in parentheses.

Traj. ID	$RMSD_{core}$ (NMR)	RMSD (NMR)	$RMSD_{core}$ (x ray)	RMSD (x ray)	First time of reaching atomic accuracy of the native state (ns)
1	0.72	1.32	0.30	0.42	310
2	0.73	0.99	0.30	0.40	26

Table II also lists the first time for reaching the native structure. In both trajectories, the time (310 ns for trajectory 1 and 26 ns for trajectory 2) is significantly shorter than those in a very recent study (where the folding occurs in $5-6$ μs).[34]

In Figs. 12(b) and 12(c), we show the superposition of typical folded structures to the NMR reference structure and the x-ray reference structure, respectively, from simulation 1. Although generally deviations from the native structures are small, the position of the N-terminal helix differs appreciably from the NMR native structure [notice the difference in position of PHE 7 and PHE 11 in Fig. 12(b)], whereas the difference is much smaller compared with the x-ray structure [PHE 7 and PHE 11 of the two structures are superimposable in Fig. 12(c)].

Figures 13(a) and 13(b) show the C_α-$RMSD_{core}$ (the stable region residues 9–32) versus the C_α-RMSD (entire chain) for the NMR and x-ray reference structures, respectively. It is clear that the best folded structures have larger deviations from the NMR structure than from the x-ray structure. Moreover, the C_α-$RMSD_{core}$ is less consistent with the C_α-RMSD with respect to the NMR structure than to the x-ray structure, which can be a result of the extremely flexible N-terminal helix. In general, we found that the folded structures in our simulation are closer to the x-ray structure. In both case, we observe many metastable states around the native state, demonstrating an extremely rugged energy landscape.

The folding fraction, computed according to C_α-RMSD from the x-ray structure and using 3.0 Å as cutoff, was fitted against Eq. (24), and yielded $P_0=0.21$, $\Delta E=20.9$ kJ mol^{-1}, and $\beta_m=0.349$ kJ^{-1} mol. Assuming that the volume change during folding is negligible, we estimate the enthalpy change $\Delta H=21$ kJ mol^{-1}, and the folding temperature $T_m=345$ K. Although the folding temperature agrees well with the experimental value 342 K, the folding enthalpy is relatively small compared to $\Delta H^{exp}=113$ kJ mol^{-1}.[35]

V. CONCLUDING DISCUSSIONS

In conclusion, we presented a single-copy enhanced sampling method for studying large complex biological systems. The method was validated in an Ising model as well as in a Lennard-Jones fluid, and was successfully applied to folding of three small proteins, trpzip2, trp-cage, and villin headpiece, in explicit solvent. In all three protein cases, we reversibly reached atomic accuracy of the native structures within a microsecond. Since our method is based on a single trajectory, it is computationally less demanding to reach a

long time scale than tempering methods based on multiple copies, such as the replica exchange method.

In terms of sampling efficiency, in addition to the advantage from using a continuous temperature generalize ensemble, which eases the issue of low temperature-transition rate in large complex systems,[2] a major improvement in this work is the employment of integral identity that efficiently estimates the average energy $\tilde{E}(\beta)$ for the temperature space random walk. The integral identity draws estimates from a large temperature window instead of from a single bin to improve statistics. This strategy makes the method more robust and applicable to large and complex systems.

Additionally, the adaptive averaging scheme used for refreshing statistics can effectively boost the temperature space random walk in early stages, when the damping magnitude γ
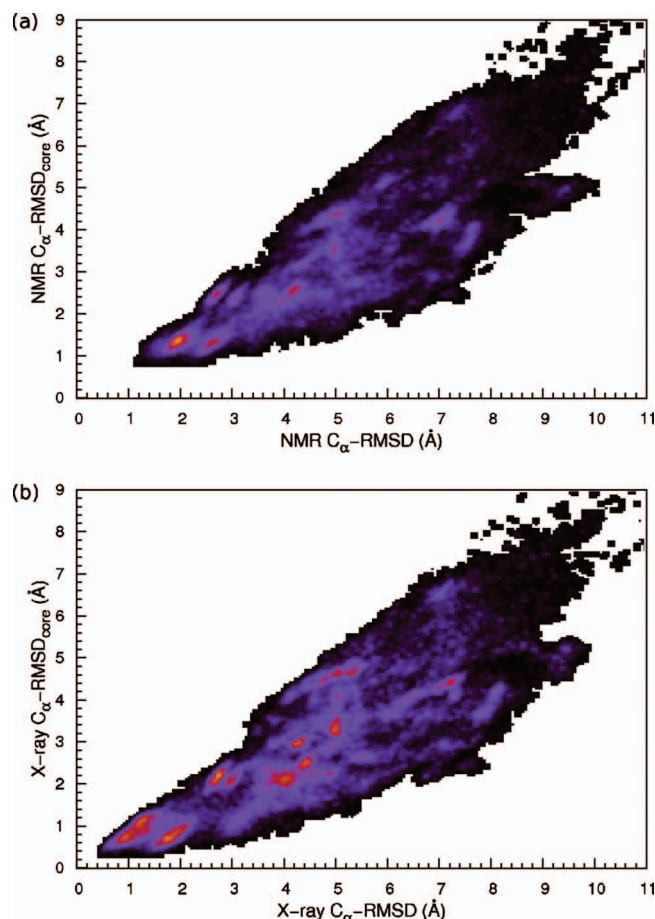

FIG. 13. Villin headpiece: joint distributions of (a) C_α-RMSD and C_α-$RMSD_{core}$ (from residues 9~32) using the NMR structure as the reference and (b) C_α-RMSD and C_α-$RMSD_{core}$, using the x-ray structure as the reference. Statistics from the two trajectories were combined.

still differs significantly from unit. It could account for several fast folding (<50 ns) events we observed. However, in later stages of simulations, the boosting effect gradually weakens as the damping magnitude γ is switched toward 1.0 [see Eq. (23)] to achieve an asymptotic convergence.

In folding applications, the calculated values for the folding enthalpy and folding temperature still showed discrepancies from experimental results. This was possibly due to errors in the force fields[26,34] as well as limited simulation time. These factors might lead to overpopulation of non-native conformations at room temperature. This observation is in line with a very recent long time replica exchange simulation on trp-cage with an aggregated 40×1 μs, which also revealed discrepancy between experimental and simulation results.[26] Even in the presence of enhanced sampling, one would desire to perform much longer simulations to allow much more folding events to atomic accuracy. In this way, one can more accurately examine population ratios of the native conformation to various folding intermediate and non-native conformations.

ACKNOWLEDGMENTS

APPENDIX A: FOKKER–PLANCK EQUATION AND STATIONARY DISTRIBUTION

Though the temperature distribution of the generalized ensemble is uniquely specified by Eq. (8), various Langevin equations exist for sampling the distribution, e.g., both Eq. (1) and the one used in the previous study[2] correctly populate the distribution.

Before applying a Langevin equation, one needs to find a proper "potential" $V(\beta)$ for the temperature to guide its diffusion. The potential is the negative logarithm of the temperature distribution

$$V(\beta) \equiv -\ln p(\beta, X).$$

The negative derivative of the potential naturally serves as the force of driving the temperature space random walk. According to Eq. (8), we have

$$\frac{\partial V(\beta)}{\partial \beta} = \frac{\partial[-\ln p(\beta, X)]}{\partial \beta} = E + \frac{\partial \ln \widetilde{Z}(\beta)}{\partial \beta} - \frac{\partial \ln w(\beta)}{\partial \beta}.$$

Equation (1) is now simplified as

$$\frac{d(1/\beta)}{dt} = \frac{dV(\beta)}{d\beta} + \frac{\sqrt{2}}{\beta}\xi. \tag{A1}$$

To show that Eq. (A1) is correct, we introduce a new variable $\tau = 1/\beta = k_B T$ and write down the Fokker–Planck equation that governs the distribution of τ or $\rho(\tau)$

$$\frac{\partial \rho(\tau)}{\partial t} = \frac{\partial}{\partial \tau}\left[-\frac{\partial V}{\partial \beta}\rho(\tau)\right] + \frac{\partial^2}{\partial \tau^2}[\tau^2 \rho(\tau)]$$

$$= \frac{\partial}{\partial \tau}\left[\rho(\tau)\tau^2 \frac{\partial V}{\partial \tau}\right] + \frac{\partial^2}{\partial \tau^2}[\tau^2 \rho(\tau)], \tag{A2}$$

where we have used $-\tau^2(\partial/\partial \tau) = \partial/\partial \beta$. A stationary solution of Eq. (A2) can be found by solving the following equation:

$$-[\tau^2 \rho(\tau)]\frac{\partial V}{\partial \tau} = \frac{\partial}{\partial \tau}[\tau^2 \rho(\tau)],$$

whose solution is readily obtained,

$$\tau^2 \rho(\tau) \sim \exp(-V) = p(\beta, X).$$

Finally, we translate the distribution of τ back to that of β according to the probability invariance $\rho(\beta)d\beta = \rho(\tau)d\tau$,

$$\rho(\beta)d\beta = \rho(\tau)d\tau = \rho(\tau)d\beta/\beta^2 = \rho(\tau)\tau^2 d\beta.$$

We thus conclude that $\rho(\beta) = \tau^2 \rho(\tau) = p(\beta, X)$, which proves that Eq. (8) is the stationary solution of the Fokker–Planck equation.

APPENDIX B: ALTERNATIVE FITTING BASED ESTIMATOR

Here we introduce an alternative estimator for $\widetilde{E}(\beta)$ based on linear extrapolation. First, we generalize Eq. (15) to the following. For any function

$$f(\beta) = \langle E\rangle_\beta + k(\bar{\beta}_i - \beta), \tag{B1}$$

where $\bar{\beta}_i = (\beta_i + \beta_{i+1})/2$ and k is an arbitrary constant, we have

$$\widetilde{E}_i = \frac{1}{\Delta\beta_i}\int_{\beta_i}^{\beta_{i+1}} d\beta\langle E\rangle_\beta = \frac{1}{\Delta\beta_i}\int_{\beta_i}^{\beta_{i+1}} d\beta f(\beta)$$

$$= \int_{\beta_-}^{\beta_+} d\beta\phi'_s(\beta)f(\beta) + \int_{\beta_-}^{\beta_+} d\beta\phi(\beta)f'(\beta), \tag{B2}$$

where on the second line, the linear term in Eq. (B1) vanishes after the integration. Equation (15) can be considered as a special case for $k=0$.

Equations (B1) and (B2) allow a properly extrapolation for the average energy $\langle E\rangle_\beta$ from β to $\bar{\beta}_i$ before applying the integral identity, or the temperature averaging. If we assume that $\langle E\rangle_\beta$ is roughly linear function of β in the window (β_-, β_+), the slope k of $\langle E\rangle_\beta$ versus β can be obtained from linear regression as

$$k = \langle\Delta\beta\Delta E\rangle_+/\langle\Delta\beta^2\rangle_+, \tag{B3}$$

where the angular bracket $\langle\cdots\rangle_+$ denotes a temperature average over (β_-, β_+).

The advantage of using Eq. (B2) is that it reduces the magnitude of the second integral and thus makes the estima-

tor more robust. If the relation of $\langle E \rangle_\beta$ and β is perfectly linear, $f'(\beta) = \partial \langle E \rangle_\beta / \partial \beta - k$ vanishes everywhere, and thus the second integral yields zero.

In practice, we use a $\phi(\beta)$ parameterized with $a_+ = a_- = 1/2$, and restrain the magnitude of the second integral within the average energy fluctuation $\sqrt{\langle \Delta E^2 \rangle_+}$ in the window to ensure stability of the estimator.

The estimator introduced here can be thought as a first-order generalization of Eq. (15), and thus allows a large temperature window or application to larger systems. Besides, one can also apply additional constraints to the derivative to improve the accuracy and stability of the estimator. For example, in our case of estimating the average energy, the derivative k corresponds to the negative energy fluctuation as $\partial \langle E \rangle_\beta / \partial \beta = -\langle \Delta E^2 \rangle_\beta = -k_B T^2 C_V$ (where C_V is the heat capacity), and its value can be restrained within a certain range. In the systems tested in this work, however, it did not produce significantly different estimates from the one introduced in the Sec. II D.

APPENDIX C: CHOICE OF THE TEMPERATURE DISTRIBUTION $\omega(\beta)$

Here, we discuss the optimal choice of $w(\beta)$. For molecular systems, we invariably use $w(\beta) \approx p(\beta) \sim 1/\beta$ basing on the following reason. The overall energy distribution $p(E)$ of the generalized ensemble is a superposition of energy distributions of canonical ensembles from different temperatures. To ensure a fixed degree of overlap, the height of $p(E)$ only needs to match that of a canonical distribution at β, whose average energy $\langle E(\beta) \rangle$ is roughly equal to E. The average height of a canonical ensemble is inversely proportional to its width $\sqrt{\langle \Delta E^2 \rangle_\beta}$, and thus,

$$p(E)dE \propto \frac{dE}{\sqrt{\langle \Delta E^2 \rangle_\beta}}.$$

To translate the energy distribution to the temperature space, we change variable from E to β. According to the probability invariance $p(E)dE = p(\beta)d\beta$, and $|dE/d\beta| = \langle \Delta E^2 \rangle_\beta$, we have

$$w(\beta) \propto \frac{|dE/d\beta|}{\sqrt{\langle \Delta E^2 \rangle_\beta}} \propto \sqrt{\langle \Delta E^2 \rangle_\beta} \propto \frac{\sqrt{C}}{\beta},$$

In the last step, we have used fact that $k_B \beta^2 \langle \Delta E^2 \rangle_\beta = C$, where C is the heat capacity.

For a Lennard-Jones-like system, the heat capacity C is roughly a constant, accordingly the optimal $w(\beta)$ is proportional to $1/\beta$. On the other hand, our protein simulations show that the heat capacity of the entire system, water and protein, roughly follows $C \propto \beta$, see Fig. 8. According to this observation, the optimal $w(\beta)$ should be proportional to $1/\sqrt{\beta}$. However, in our simulations, we still used $1/\beta$ as $w(\beta)$, this setup slightly biases the ensemble toward the high temperature end in order to encourage a faster motion at higher temperature and to help overcome broken ergodicity at lower temperature.

APPENDIX D: INTEGRAL IDENTITY FOR A NON-NEGATIVE QUANTITY

For a histogram-like quantity, regular integral identities such as Eq. (15) should be modified to ensure the output is non-negative. Here, we briefly sketch a technique for this purpose (a more detailed and general treatment is presented elsewhere[6]).

Suppose $h(\beta)$ is a non-negative quantity, and we are interested in estimating its value at a particular temperature β^*. Instead of applying an integral identity to $h(\beta)$ itself, we introduce a smooth modulating function $f(\beta)$, with $f(\beta^*) = 1$, and apply the identity to the product $h(\beta)f(\beta)$,

$$h(\beta^*) = h(\beta^*)f(\beta^*)$$
$$= \int_{\beta_-}^{\beta_+} [h(\beta)f(\beta)]\phi'(\beta)d\beta + \int_{\beta_-}^{\beta_+} [h(\beta)f(\beta)]' \phi(\beta)d\beta.$$

(D1)

If we choose

$$f(\beta) = \exp\left(-\int_{\beta^*}^{\beta} [\ln h(\beta')]' d\beta'\right),$$

(D2)

the second term on the right hand side of Eq. (D1) vanishes because $[h(\beta)f(\beta)]' = 0$ everywhere. Thus, the estimated value is always non-negative given that $\phi'(\beta) \geq 0$. $\phi'(\beta)$ satisfies the normalization condition $\int_{\beta_-}^{\beta_+} \phi'(\beta)d\beta = 1$, and should be proportional to $w(\beta)/f(\beta)$ to minimize the error. Thus, the general expression for $h(\beta^*)$ is

$$h(\beta^*) = \frac{\int_{\beta_-}^{\beta_+} w(\beta)h(\beta)d\beta}{\int_{\beta_-}^{\beta_+} w(\beta)\exp\left(\int_{\beta^*}^{\beta} [\ln h(\beta')]' d\beta'\right)d\beta}.$$

(D3)

In the case of energy histogram, $h(\beta)$ can defined as an integral of the canonical energy distribution at β over a small energy bin ΔE,

$$h(\beta) = \int_{E}^{E+\Delta E} \frac{g(E')\exp(-\beta E')}{Z(\beta)} dE'$$
$$= \left(\int_{E}^{E+\Delta E} p_\beta(E')dE'\right)/w(\beta),$$

where $g(E')$ is the density of states; on the second line, we have used the distribution function of the generalized ensemble $p_\beta(E') = w(\beta)g(E')\exp(-\beta E')/Z(\beta)$.

It can be straightforwardly verified that

$$[\ln h(\beta)]' = -[\ln Z(\beta)]' - \frac{\int_{E}^{E+\Delta E} E' g(E')\exp(-\beta E')dE'}{\int_{E}^{E+\Delta E} g(E')\exp(-\beta E')dE'}$$
$$\approx -[\ln Z(\beta)]' - (E + \Delta E/2),$$

where on the last line, we assumed that the energy bin size ΔE is sufficiently small such that a typical energy distribution does not vary drastically within a bin. Practically, it only requires the bin size ΔE to be much smaller than the fluctuation (or width) of a typical energy distribution. We finally reached

$$h(\beta^*) \approx \frac{\int_{\beta_-}^{\beta_+}\int_E^{E+\Delta E} p(\beta,E')dE'd\beta}{\int_{\beta_-}^{\beta_+} w(\beta)Z(\beta^*)/Z(\beta)\exp[-(\beta-\beta^*)(E+\Delta E/2)]d\beta}.$$

By substituting $\widetilde{Z}(\beta)$ for $Z(\beta)$, we recover Eq. (20).

[1] A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsovvelyaminov, J. Chem. Phys. **96**, 1776 (1992); E. Marinari and G. Parisi, Europhys. Lett. **19**, 451 (1992); C. Zhang and J. Ma, Phys. Rev. E **76**, 036708 (2007); H. Li, M. Fajer, and W. Yang, J. Chem. Phys. **126**, 024106 (2007); Y. Q. Gao, *ibid.* **128**, 064105 (2008); J. Kim, J. E. Straub, and T. Keyes, Phys. Rev. Lett. **97**, 050601 (2006).

[2] C. Zhang and J. Ma, J. Chem. Phys. **130**, 194112 (2009).

[3] C. J. Geyer, *Proceedings of the 23rd Symposium on the Interface* (American Statistical Association, New York, 1991); U. H. E. Hansmann, Chem. Phys. Lett. **281**, 140 (1997); K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996); R. H. Swendsen and J. S. Wang, Phys. Rev. Lett. **57**, 2607 (1986); J. Kim, J. E. Straub, and T. Keyes, J. Chem. Phys. **132**, 224107 (2010).

[4] E. Darve and A. Pohorille, J. Chem. Phys. **115**, 9169 (2001); M. Fasnacht, R. H. Swendsen, and J. M. Rosenberg, Phys. Rev. E **69**, 056704 (2004).

[5] A. B. Adib and C. Jarzynski, J. Chem. Phys. **122**, 014114 (2005).

[6] C. Zhang and J. Ma, arXiv:1005.0170 (2010).

[7] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988); **63**, 1195 (1989).

[8] A. E. Ferdinand and M. E. Fisher, Phys. Rev. **185**, 832 (1969).

[9] P. D. Beale, Phys. Rev. Lett. **76**, 78 (1996).

[10] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. Berendsen, J. Comput. Chem. **26**, 1701 (2005); B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, J. Chem. Theory Comput. **4**, 435 (2008); E. Lindahl, B. Hess, and D. van der Spoel, J. Mol. Model. **7**, 306 (2001); H. J. C. Berendsen and D. van der Spoel, Comput. Phys. Commun. **91**, 43 (1995).

[11] E. J. Sorin and V. S. Pande, Biophys. J. **88**, 2472 (2005).

[12] J. C. W. L. Jorgensen, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).

[13] U. Essman, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).

[14] G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**, 014101 (2007).

[15] S. Miyamoto and P. A. Kollman, J. Comput. Chem. **13**, 952 (1992).

[16] B. Hess, J. Chem. Theory Comput. **4**, 116 (2008).

[17] A. G. Cochran, N. J. Skelton, and M. A. Starovasnik, Proc. Natl. Acad. Sci. U.S.A. **98**, 5578 (2001).

[18] J. Zhang, M. Qin, and W. Wang, Proteins **62**, 672 (2006); J. W. Pitera, I. Haque, and W. C. Swope, J. Chem. Phys. **124**, 141102 (2006).

[19] L. Yang, Q. Shao, and Y. Q. Gao, J. Phys. Chem. B **113**, 803 (2009); C. D. Snow, L. Qiu, D. Du, F. Gai, S. J. Hagen, and V. S. Pande, Proc. Natl. Acad. Sci. U.S.A. **101**, 4077 (2004); A. E. Roitberg, A. Okur, and C.

Simmerling, J. Phys. Chem. B **111**, 2415 (2007); C. Chen and Y. Xiao, Bioinformatics **24**, 659 (2008); W. Y. Yang, J. W. Pitera, W. C. Swope, and M. Gruebele, J. Mol. Biol. **336**, 241 (2004).

[20] J. P. Ulmschneider, M. B. Ulmschneider, and A. Di Nola, J. Phys. Chem. B **110**, 16733 (2006).

[21] J. Wang, P. Cieplak, and P. A. Kollman, J. Comput. Chem. **21**, 1049 (2000).

[22] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Proteins **65**, 712 (2006).

[23] J. W. Neidigh, R. M. Fesinmeyer, and N. H. Andersen, Nat. Struct. Biol. **9**, 425 (2002).

[24] L. Qiu, S. A. Pabit, A. E. Roitberg, and S. J. Hagen, J. Am. Chem. Soc. **124**, 12952 (2002).

[25] R. Zhou, Proc. Natl. Acad. Sci. U.S.A. **100**, 13280 (2003); D. Paschek, S. Hempel, and A. E. Garcia, *ibid.* **105**, 17754 (2008); J. Juraszek and P. G. Bolhuis, *ibid.* **103**, 15859 (2006); S. Kannan and M. Zacharias, Proteins **76**, 448 (2008).

[26] R. Day, D. Paschek, and A. E. Garcia, Proteins **78**, 11 (2010).

[27] L. Yang, M. P. Grubb, and Y. Q. Gao, J. Chem. Phys. **126**, 125102 (2007); C. Simmerling, B. Strockbine, and A. E. Roitberg, J. Am. Chem. Soc. **124**, 11258 (2002); S. Chowdhury, M. C. Lee, G. Xiong, and Y. Duan, J. Mol. Biol. **327**, 711 (2003); C. D. Snow, B. Zagrovic, and V. S. Pande, J. Am. Chem. Soc. **124**, 14548 (2002); J. W. Pitera and W. Swope, Proc. Natl. Acad. Sci. U.S.A. **100**, 7587 (2003).

[28] W. W. Streicher and G. I. Makhatadze, Biochemistry **46**, 2876 (2007).

[29] Y. Duan and P. A. Kollman, Science **282**, 740 (1998).

[30] T. K. Chiu, J. Kubelka, R. Herbst-Irmer, W. A. Eaton, J. Hofrichter, and D. R. Davies, Proc. Natl. Acad. Sci. U.S.A. **102**, 7517 (2005).

[31] G. Jayachandran, V. Vishal, and V. S. Pande, J. Chem. Phys. **124**, 164902 (2006).

[32] H. Lei, C. Wu, H. Liu, and Y. Duan, Proc. Natl. Acad. Sci. U.S.A. **104**, 4925 (2007).

[33] D. L. Ensign, P. M. Kasson, and V. S. Pande, J. Mol. Biol. **374**, 806 (2007); C. D. Snow, H. Nguyen, V. S. Pande, and M. Gruebele, Nature (London) **420**, 102 (2002); M. Y. Shen and K. F. Freed, Proteins **49**, 439 (2002); A. Fernández, M. Y. Shen, A. Colubri, T. R. Sosnick, R. S. Berry, and K. F. Freed, Biochemistry **42**, 664 (2003); B. Zagrovic, C. D. Snow, M. R. Shirts, and V. S. Pande, J. Mol. Biol. **323**, 927 (2002).

[34] P. L. Freddolino and K. Schulten, Biophys. J. **97**, 2338 (2009).

[35] J. Kubelka, W. A. Eaton, and J. Hofrichter, J. Mol. Biol. **329**, 625 (2003).

[36] M. Wang, Y. Tang, S. Sato, L. Vugmeyster, C. J. McKnight, and D. P. Raleigh, J. Am. Chem. Soc. **125**, 6032 (2003).

[37] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman, J. Comput. Chem. **24**, 1999 (2003).

[38] R. B. Best, N. V. Buchete, and G. Hummer, Biophys J **95**, L07 (2008); R. B. Best and G. Hummer, J. Phys. Chem. B **113**, 9004 (2009); E. J. Thompson, A. J. DePaul, S. S. Patel, and E. J. Sorin, PLoS ONE **5**, e10056 (2010).

[39] W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graph. **14**, 33 (1996).