Enhanced sampling in generalized ensemble with large gap of sampling parameter: Case study in temperature space random walk

Cheng Zhang¹ and Jianpeng Ma^{1,2,a)}

¹Department of Bioengineering, Rice University, Houston, Texas 77005, USA

²Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030, USA

(Received 12 March 2009; accepted 29 April 2009; published online 20 May 2009)

We present an efficient sampling method for computing a partition function and accelerating configuration sampling. The method performs a random walk in the λ space, with λ being any thermodynamic variable that characterizes a canonical ensemble such as the reciprocal temperature β or any variable that the Hamiltonian depends on. The partition function is determined by minimizing the difference of the thermal conjugates of λ (the energy in the case of $\lambda = \beta$), defined as the difference between the value from the dynamically updated derivatives of the partition function and the value directly measured from simulation. Higher-order derivatives of the partition function are included to enhance the Brownian motion in the λ space. The method is much less sensitive to the system size, and to the size of λ window than other methods. On the two dimensional Ising model, it is shown that the method asymptotically converges the partition function, and the error of the logarithm of the partition function is much smaller than the algorithm using the Wang–Landau recursive scheme. The method is also applied to off-lattice model proteins, the AB models, in which cases many low energy states are found in different models.

© 2009 American Institute of Physics. [DOI: 10.1063/1.3139192]

I. INTRODUCTION

In a canonical ensemble, the partition function is defined as a sum over configurations (denoted by X),

$$Z(\lambda) = \sum_X \exp[-\mathcal{H}_\lambda(X)], \quad (1)$$

where $\mathcal{H}_\lambda(X)$ is the reduced Hamiltonian of the system with a dependence on a variable λ . In the case that the λ dependence only exists in the energy function $E_\lambda(X)$, $\mathcal{H}_\lambda(X)$ can be written as $\beta E_\lambda(X)$, where $\beta = 1/T$ is the reciprocal temperature. However, λ can be the temperature β , leaving the energy function $E(X)$ independent of β .

A quantity of particular interest is the ratio of the partition function $Z(\lambda_1)/Z(\lambda_0)$ at two given λ s, λ_1 and λ_0 (if λ does not explicitly involve the temperature, the ratio can be translated to the free energy difference as $\Delta F = -T \ln[Z(\lambda_1)/Z(\lambda_0)]$). In our previous study,¹ the partition function is computed in an expanded ensemble, where a regular simulation is coupled with random transitions among different λ s, e.g., temperatures β s or volumes V s. This approach requires two neighboring distributions of the corresponding macroscopic quantities, such as the energy for the temperature or the virial for the volume, to overlap sufficiently. Accordingly the spacing $\Delta\lambda$ is proportional to $1/\sqrt{N}$, and the number of λ sampling points increases as $\sqrt{N}$, as the system size N grows. Many other methods, such as replica exchange,² simulated tempering,³ and others,^{4,5} have similar issues.

To overcome this problem of increasingly large number of sampling points, we define λ as a continuous variable instead of a discrete one. The partition function $Z(\lambda)$ is characterized as a continuous function by a few adjustable parameters (e.g., derivatives with respect to λ) for a large λ window. In this way, the number of λ windows can be significantly reduced, and one can handle a much larger system. Moreover, the ratio of the partition function at end points (i.e., boundaries of a λ window), $Z(\lambda_1)/Z(\lambda_0)$, can be asymptotically determined through a simulation.

In this paper, we first present the theoretical background of the method in a general framework. Next the simulation protocol is exemplified in a special case where λ is the reciprocal temperature β . We then numerically test the performance of the method on the two-dimensional Ising model, and apply the method to folding model proteins. At the end, we conclude the method with discussions.

II. METHOD

A. General theory

The method aims at determining the ratio of the partition function $Z(\lambda_1)/Z(\lambda_0)$. To do so, we perform a simulation in a generalized ensemble of a wide λ range and use trajectory averages to obtain runtime estimates of derivatives of the partition function. As simulation progresses, one can simultaneously reach an asymptotically converged partition function $Z(\lambda)$ as well as an efficient random walk in the λ space.

We first construct a generalized ensemble composed of canonical ensembles of a continuous λ range. The probability of λ being in the interval $(\lambda, \lambda + d\lambda)$ is given by

^{a)}Electronic mail: jpmma@bcm.tmc.edu.

$$\begin{aligned}
 w(\lambda)d\lambda &= \tilde{w}(\lambda) \frac{Z(\lambda)}{\tilde{Z}(\lambda)} d\lambda \\
 &= \sum_X \tilde{w}(\lambda) \exp[-\mathcal{H}_\lambda(X) - \ln \tilde{Z}(\lambda)] d\lambda, \quad (2)
 \end{aligned}$$

where we have introduced an approximate partition function $\tilde{Z}(\lambda)$, as well as a predefined weight function $\tilde{w}(\lambda)$. In the second line, the partition function is expanded using Eq. (1). If, in a special case, $\tilde{w}(\lambda)$ is a constant, the generalized ensemble corresponds to a flat λ histogram.^{1,2,4,5}

An efficient sampling has three requirements. First, the method should yield the correct partition function ratio at the end points $Z(\lambda_1)/Z(\lambda_0)$, or the free energy difference. Second, the difference between the actual weight $w(\lambda)$ should be close to the desired one $\tilde{w}(\lambda)$, or equivalently $\tilde{Z}(\lambda)$ should be close to $Z(\lambda)$. Last, an efficient scheme for the λ -space random walk is required.

The first requirement can be rephrased to a condition on the average properties of the ensemble,

$$\begin{aligned}
 0 &= \frac{Z(\lambda_1)}{\tilde{Z}(\lambda_1)} - \frac{Z(\lambda_0)}{\tilde{Z}(\lambda_0)} \\
 &= \int_{\lambda_0}^{\lambda_1} \frac{\partial}{\partial \lambda} \left(\frac{Z}{\tilde{Z}} \right) d\lambda \\
 &= \int_{\lambda_0}^{\lambda_1} \left[-\frac{\partial \ln \tilde{Z}}{\partial \lambda} - \left\langle \frac{\partial \mathcal{H}_\lambda}{\partial \lambda} \right\rangle_\lambda \right] \left(\frac{Z}{\tilde{Z}} \right) d\lambda \\
 &= \langle \tilde{P}(\lambda) \rangle_{\tilde{w}} - \langle \langle P_\lambda(X) \rangle_\lambda \rangle_{\tilde{w}}. \quad (3)
 \end{aligned}$$

Here, $\langle \dots \rangle_\lambda$ denotes a configuration average at a fixed λ ; $\tilde{P}(\lambda) \equiv -\partial \ln \tilde{Z} / \partial \lambda$ is the derivative of the estimated partition function; $P_\lambda(X) \equiv \partial \mathcal{H}_\lambda(X) / \partial \lambda$ is the thermal conjugate of λ ; $\langle A \rangle_{\tilde{w}}$ denotes a weighted average in the generalized ensemble for quantity A ,

$$\langle A \rangle_{\tilde{w}} \equiv \int_{\lambda_0}^{\lambda_1} A(\lambda) \frac{w(\lambda)}{\tilde{w}(\lambda)} d\lambda = \int_{\lambda_0}^{\lambda_1} A(\lambda) \times \frac{Z(\lambda)}{\tilde{Z}(\lambda)} d\lambda.$$

It is evident that as long as the two averages on the right hand side of the last line of Eq. (3) are equal, the requirement on the correct ratio of the partition function is satisfied. In simulation, $\langle \tilde{P}(\lambda) \rangle_\lambda$ from the estimated partition function is dynamically adjusted to be equal to the measured average $\langle \langle P_\lambda(X) \rangle_\lambda \rangle_{\tilde{w}}$ (practically, the twofold average is translated to the average in the generalized ensemble, which is then replaced by a trajectory average measured along simulation). As a result, one can eventually obtain the correct ratio of the partition function.

The second requirement is satisfied by minimizing the following quantity:

$$S = \langle [\tilde{P}(\lambda) - \langle P_\lambda(X) \rangle_\lambda]^2 \rangle_{\tilde{w}}. \quad (4)$$

Since $\tilde{P}(\lambda)$ and $\langle P_\lambda(X) \rangle_\lambda$ are the derivatives of $-\ln \tilde{Z}(\lambda)$ and $-\ln Z(\lambda)$, respectively, S is minimized when the two are equal at any λ . However, a perfect match between $\tilde{P}(\lambda)$ and

$\langle P_\lambda(X) \rangle_\lambda$ is practically impossible to reach. We thus adopt a variational approach, where $\tilde{P}(\lambda)$ is approximated as a linear combination of a few trial functions $\phi_k(\lambda)$ s as

$$\tilde{P}(\lambda) = \sum_k a_k \phi_k(\lambda).$$

An example of the expansion is a power series $\tilde{P}(\lambda) = a_0 + a_1\lambda + a_2\lambda^2 + \dots$, where $\phi_k(\lambda) = \lambda^k$ and a_k corresponds to the k th order derivative of $\tilde{P}(\lambda)$, $a_k = (1/k!) d^{(k)}\tilde{P}/d\lambda^{(k)}$. Now one can minimize S with respect to the coefficients a_k s as $\partial S / \partial a_k = 0$. This determines a_k s from the following set of equations:

$$\sum_k a_k \langle \phi_j(\lambda) \phi_k(\lambda) \rangle_{\tilde{w}} = \langle \phi_j(\lambda) \langle P_\lambda(X) \rangle_\lambda \rangle_{\tilde{w}}. \quad (5)$$

Similar to the case of Eq. (3), the twofold average $\langle \phi_j(\lambda) \times \langle P_\lambda(X) \rangle_\lambda \rangle_{\tilde{w}}$ is equivalent to the average of $\phi_j(\lambda) P_\lambda(X)$ in the generalized ensemble, and can be evaluated from a simulation trajectory. The parameters a_k s are regularly updated in simulation accordingly to Eq. (5) to enforce the minimization of S . Once all a_k s are obtained, the ratio of the estimated partition function can be calculated as

$$\ln \tilde{Z}(\lambda_1) - \ln \tilde{Z}(\lambda_0) = - \int_{\lambda_0}^{\lambda_1} \tilde{P}(\lambda) d\lambda = - \sum_k a_k \Delta \Phi_k, \quad (6)$$

where $\Delta \Phi_k \equiv \int_{\lambda_0}^{\lambda_1} \phi_k(\lambda') d\lambda'$. We now show that Eq. (5) is compatible with the first requirement of Eq. (3); assuming $\phi_0 = 1$, the first equation of Eq. (5), i.e., the $j=0$ case, becomes $\langle \tilde{P}(\lambda) \rangle_{\tilde{w}} = \langle \sum_k a_k \phi_k \rangle_{\tilde{w}} = \langle \langle P_\lambda(X) \rangle_\lambda \rangle_{\tilde{w}}$, which is identical to Eq. (3).

Last, sampling in the generalized ensemble can be implemented by a regular configurational sampling at a fixed λ , as well as a random walk in the λ space. Any constant temperature algorithm can be used to generate configurational moves at a fixed λ . For the λ -space sampling, a convenient choice is to follow a Langevin equation,

$$\frac{d\lambda}{dt} = -\frac{1}{\tilde{w}} [P_\lambda(X) - \tilde{P}(\lambda)] + \frac{\xi}{\sqrt{\tilde{w}}}, \quad (7)$$

where ξ is a Gaussian white noise that satisfies $\langle \xi(t) \xi(t') \rangle = 2\delta(t-t')$, with t being the simulation time. The equation is derived by treating $V_\lambda(X) = \mathcal{H}_\lambda(X) + \ln \tilde{Z}(\lambda)$ in Eq. (2) as the “potential,” and its derivative $-\partial V_\lambda(X) / \partial \lambda = -[P_\lambda(X) - \tilde{P}(\lambda)]$ as the “force” of the λ -space random walk. To show that Eq. (7) yields the correct λ distribution, we examine the time evolution of the λ distribution $\rho(\lambda)$, described by the corresponding Fokker–Planck equation, $\partial \rho / \partial t = (\partial / \partial \lambda) \times [(\partial V_\lambda / \partial \lambda)(\rho / \tilde{w})] + (\partial^2 / \partial \lambda^2)(\rho / \tilde{w})$, whose stationary solution ($\partial \rho / \partial t = 0$) indeed gives the desired distribution $\rho \sim \tilde{w} \exp[-V_\lambda(X)]$.

B. A case study: Temperature space sampling

In the following discussion, λ is assumed to be the reciprocal temperature β . Thus, $P_\lambda(X) = \partial[\beta E(X)] / \partial \beta$ in Eq. (3) is the energy E , and $\tilde{P}(\lambda)$ can be interpreted as the estimated average energy $\tilde{E}(\beta) = -\partial \ln \tilde{Z} / \partial \beta$.

The aim is to compute the ratio of the partition function $Z(\beta_1)/Z(\beta_0)$. The ratio is to be calculated from the estimated average energy $\tilde{E}(\beta)$. For this reason, we look for a best fit between estimated average energy $\tilde{E}(\beta)$ and the actual one $\langle E \rangle_\beta$. If one expands the estimated average energy as $\tilde{E}(\beta) = a_0 + a_1\beta + a_2\beta^2 + \dots$, the task is to determine the best fitting coefficients a_k s. As in the general formalism, the trial functions are $\phi_0=1, \phi_1=\beta, \phi_2=\beta^2, \dots$, and the coefficients correspond to derivatives of the estimated partition function, e.g., $a_0 = \tilde{E}(0) = -\partial \ln Z / \partial \beta|_{\beta=0}$ and $a_1 = \partial \tilde{E} / \partial \beta|_{\beta=0} = -\partial^2 \ln Z / \partial \beta^2|_{\beta=0} \dots$

The simulation procedure is described as follows. To be more specific, we assume that a third order expansion of $\tilde{E}(\beta) = a_0 + a_1\beta + a_2\beta^2$ is used. The method has two components: A regular configurational sampling at a given temperature β and a random walk in the temperature space. For generating configurational moves at a temperature β , the Metropolis algorithm or a constant temperature molecular dynamics method can be used. After a configurational step is finished, we accumulate averages for $\langle \beta \rangle$, $\langle \beta^2 \rangle$, $\langle \beta^3 \rangle$, $\langle \beta^4 \rangle$, $\langle E \rangle$, $\langle \beta E \rangle$, and $\langle \beta^2 E \rangle$, where the first four values correspond to $\langle \phi_j(\lambda) \phi_i(\lambda) \rangle_{\tilde{w}}$ and the rest to $\langle \phi_j(\lambda) \langle P_\lambda(X) \rangle_\lambda \rangle_{\tilde{w}}$ in Eq. (5). Note, the symbol $\langle \dots \rangle$ here is a shorthand notation for a trajectory average; it corresponds to an ensemble average in the generalized ensemble, where the averaging over both configuration and β is implied. For simplicity, we have also assumed $\tilde{w}(\beta)$ to be a constant and thus dropped the $\tilde{w}$ subscript.

The temperature space random walk is realized by assuming β as a continuous variable within the temperature range of interest (β_0, β_1) . Although one can divide the whole temperature range into several subwindows, we assume only one window in this example for the sake of simplicity. The current temperature β is updated regularly, i.e., every few configurational sampling steps. Before a temperature update, the coefficients a_0 , a_1 , and a_2 are determined by solving Eq. (5), or explicitly

$$\begin{pmatrix} 1 & \langle \beta \rangle & \langle \beta^2 \rangle \\ \langle \beta \rangle & \langle \beta^2 \rangle & \langle \beta^3 \rangle \\ \langle \beta^2 \rangle & \langle \beta^3 \rangle & \langle \beta^4 \rangle \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \langle E \rangle \\ \langle \beta E \rangle \\ \langle \beta^2 E \rangle \end{pmatrix}.$$

Then the temperature β is updated according to the Langevin equation [Eq. (7)], or explicitly

$$d\beta/dt = (a_0 + a_1\beta + a_2\beta^2) - E + \xi,$$

where E is the current energy; ξ is a Gaussian white noise that satisfies $\langle \xi(t) \xi(t') \rangle = 2\delta(t-t')$, which can be conveniently generated using a random number generator. If the Langevin equation drives the current temperature out of the entire temperature range, the update is rejected and the old temperature is preserved. At the end, the ratio of the estimated partition function can be calculated as,

$$\begin{aligned} \ln \tilde{Z}(\beta_0) - \ln \tilde{Z}(\beta_1) \\ = a_0(\beta_1 - \beta_0) + a_1(\beta_1^2 - \beta_0^2)/2 + a_2(\beta_1^3 - \beta_0^3)/3. \end{aligned}$$

As the simulation progresses, the coefficients a_0 , a_1 , and a_2

gradually converge to fixed values, and the ratio of the estimated partition function asymptotically approaches the ratio of the correct one $\tilde{Z}(\beta_0)/\tilde{Z}(\beta_1) \rightarrow Z(\beta_0)/Z(\beta_1)$.

Note, in this method, although a_k s correspond to different order derivatives of the partition function, the determination of these parameters by Eq. (5) does not involve averages of high-order moments of $P_\lambda(X)$ such as $\langle [P_\lambda(X)]^2 \rangle_{\tilde{w}}$ or $\langle [P_\lambda(X)]^3 \rangle_{\tilde{w}}$, or averages of high-order derivatives of $\mathcal{H}_\lambda(X)$ such as $\langle \partial^2 \mathcal{H}_\lambda(X) / \partial \lambda^2 \rangle_{\tilde{w}} = -\langle \partial P_\lambda(X) / \partial \lambda \rangle_{\tilde{w}}$ and $\langle \partial^3 \mathcal{H}_\lambda(X) / \partial \lambda^3 \rangle_{\tilde{w}} = -\langle \partial^2 P_\lambda(X) / \partial \lambda^2 \rangle_{\tilde{w}}$. This is a desirable feature because these high-order quantities are usually difficult to compute or may not be well-defined. We naturally avoid these quantities by using moments of the variable λ instead. In the above example, a_0 , a_1 , and a_2 are determined by averages, such as $\langle \beta^3 \rangle$ and $\langle \beta^2 E \rangle$, but not high-order moments of E , such as $\langle \beta E^3 \rangle$ and $\langle \beta E^2 \rangle$. This feature makes the updating more robust and the method more applicable for a general λ .

III. NUMERICAL RESULTS

A. Two-dimensional Ising model

We first perform a test on the 32×32 Ising model using the first, second, and third order series expansion of $\tilde{E}(\beta)$. The Metropolis algorithm is used to generate configuration changes. The range of β is (0,0.25), the corresponding T range is (4, $+\infty$). The time step for integrating the Langevin equation [Eq. (7)] is 5×10^{-5} . The results are shown in Fig. 1. First let us examine the β histogram, which corresponds to the β distribution $w(\beta)$ defined in Eq. (2). According to Eq. (3), the values of the β histogram at the end points should be equal, i.e., $w(\beta_0) = w(\beta_1)$ [$\tilde{w}(\beta)$ is constant here and $Z(\beta_0)/\tilde{Z}(\beta_0) = Z(\beta_1)/\tilde{Z}(\beta_1)$ according to Eq. (3)]. Figure 1(a) agrees with the expectation in all the three cases. In addition, we expect the approximate partition function to be sufficiently close to the exact one. The difference of the two can be examined from the difference between the actual weight $w(\beta)$ and the desired weight $\tilde{w}(\beta)$. Since $\tilde{w}(\beta)$ is a constant in this case, the β histogram that represents $w(\beta)$ should be sufficiently flat. It can be seen from Fig. 1(a), the first-order algorithm yields a β distribution peaked at both the boundaries $\beta=0$ and $\beta=0.25$, while the third order algorithm yields an almost flat histogram. For the energy histograms [Fig. 1(b)], since the corresponding energy distributions are very different at $\beta=0$ and $\beta=0.25$, $\tilde{E}(\beta)$ cannot be represented by a constant value, and thus the first order algorithm becomes ineffective (no overlap between two energy distributions at the given β gap). On the other hand, using a higher order version, $\tilde{E}(\beta)$ can more effectively approximate the average energy as a function of β . As a result, we can achieve a flatter β histogram as well as a broader energy histogram. The example shows that higher order algorithms can handle a much larger temperature gap than the first order one.

We also compare the current method with the method from a previous study¹ which uses the Wang-Landau (WL) updating scheme⁶ to converge the partition function. In both cases, we update the temperature after a sweep of configura-

FIG. 1. (a) The β histogram using different orders of $\tilde{E}(\beta)$ expansion. The right axis is for the first order (the solid line). The left axis is for the rest two (the dashed and dotted line). (b) The corresponding energy histograms using different orders expansions. For comparison, the constant temperature energy histograms (using the Metropolis algorithm) for $\beta=0$ and $\beta=0.25$ are shown in the shaded area.

tion sampling. For the current method, the third order expansion with a single temperature window is used. For the previous method, 25 sampling temperatures are evenly distributed in the temperature range with $\Delta\beta=0.01$. Such a fine temperature interval ensures that the previous method targets the flat β -histogram ensemble and it has a good transition rate between neighboring temperatures. The parameters of the WL updating scheme are the following. The initial value for the modification factor $\ln f=1.0$ and it is shrunk by a factor of 2 at the end of each stage. The criterion for terminating a stage depends on the flatness of the temperature histogram. Three different choices of the flatness thresholds 20%, 50%, and 99% are used (in the last case, a stage is terminated when each sampling temperature is visited at least once). We also use a recipe⁷ of improving the convergence of the WL updating scheme in final stages, where the modification factor $\ln f$ is specified as $1/t_n$ in final stages regardless of the histogram flatness. Here t_n is defined as the number of Monte Carlo steps divided by the number of temperatures.

The results of the comparison are shown in Fig. 2. Since the exact partition function for the Ising model is available,⁸ the error of the logarithmic partition function, defined as $\epsilon=|\Delta \ln \tilde{Z}-\Delta \ln Z|$, can be measured. Here we show the averaged error over 1000 independent runs. Note, since we

FIG. 2. The logarithmic error of the ratio of the partition function $\epsilon(\ln[Z(\lambda_1)/\ln Z(\lambda_0)])$ vs the simulation time t . For the WL updating scheme, three thresholds of the histogram flatness 20%, 50%, and 99% are used. The result from the $\ln f=1/t_n$ correction (where t_n is the number of MC steps divided by the number of temperatures) is shown as the dotted line. For the current method (solid line), the error ϵ scales with the simulation time t as $\epsilon \sim t^{-0.52}$. By using the parallel version (four copies, dashed line), the error is further reduced. All results are averaged over 1000 independent runs.

averaged the absolute value of the error, both the random error and systematic bias (if any) are included. For the current method, the error reduces steadily as the simulation progresses. The error ϵ as a function simulation sweeps (Monte Carlo (MC) steps per site) t can be fitted by regression as $\epsilon=11.6t^{-0.52}$. It is clear that the original WL recursive scheme suffers from the problem of saturation at a long simulation time. Although the $1/t_n$ recipe eases the problem, it is still less efficient than the method introduced here. The error of the current method at the end of 10^5 sweeps is 0.0297, while for the WL recursion with $1/t_n$ recipe, the error is 0.121. The same precision $\epsilon=0.121$ can be achieved by the current method at the end of 6000 sweeps. This shows that the current method is one to two orders of magnitude more efficient than the WL recursion in terms of simulation time in this case.

The efficiency of the method can be further improved when parallel computers are available. We briefly describe a parallel extension here. In the parallel version, multiple copies of simulations run simultaneously using a same set of a_k s. All copies contribute to the trajectory averages, such as $\langle\beta^2\rangle$ and $\langle\beta E\rangle$. The parameters a_k s shared by all copies are calculated from the averages from multiple trajectories and therefore are more accurate. In Fig. 2, we also show that the result from the parallel version (dashed line) using four copies. The error at the end of 10^5 sweeps is 0.0156, which is about half of the single copy version. According to the $t^{-1/2}$ scaling relation, the convergence rate is about four times as fast as the single copy one, as we expected. By contrast, the WL type updating does not have a convenient parallel counterpart.

B. Atomistic model protein

The method is also applied to locate low energy states of an AB model protein,⁹ which has been extensively studied in

literature.^{1,10,11} It has two types of residues, A: hydrophobic, B: hydrophilic. Particularly, we used the second set of molecular force fields,⁹ which produces more globular low energy structures.

The major challenge of this system is that it contains many different low-energy wells separated by high barriers. Due to its rugged low-energy landscape, the AB model serves as a stringent testing case for the ability of configurational sampling of the algorithm. Although in principle thermal properties are determined by averages from all low-energy wells, only the one with the lowest energy has a dominant contribution at a low temperature. For example, if two energy wells have a $\Delta E=2.0$ difference in their energy (which is common for a system with 55 or 89 residues), the contribution from the higher energy well is only $\exp(-\Delta E/T) \approx 2 \times 10^{-9}$ times that of the lower energy well at temperature $T=0.1$. Since lower energy states are gradually discovered as the simulation proceeds, average properties estimated in an earlier time must be promptly corrected according to the newly found low energy states.

Due to this reason, we use a more aggressive averaging scheme that favors recent statistics. In this scheme, we introduce a memory factor $\gamma < 1$ to gradually shrink the weight of previous statistics. For example, the average energy $\langle E \rangle$ is computed as $\mathcal{E}/\mathcal{N}$, where $\mathcal{E}$ is the total energy and $\mathcal{N}$ is the total weight (which are both accumulated from the beginning of simulation). The total energy and total weight are updated as $\mathcal{E} \rightarrow \gamma\mathcal{E} + E$ and $\mathcal{N} \rightarrow \gamma\mathcal{N} + 1$ after each sampling step, where $\gamma \leq 1$. If $\gamma=1$, the averaging scheme is reduced to a regular average. However, when a factor $\gamma < 1$ is used, a recently visited state would have a larger statistical weight than a state visited much earlier (e.g., the weight of a state sampled five steps before is reduced to γ^5 while the weight of the current state is unity). In this way the average energy can now be readily adjusted according to recent statistics. Since the deterministic part of the Langevin equation [Eq. (7)] for the temperature-space random walk is controlled by the difference between the current energy and the average energy, the new scheme can effectively avoid the temperature being trapped to a local value. In this way, the random walk in the temperature space can be significantly enhanced. The modified averaging scheme is most useful for improving low temperature sampling in a rugged energy landscape in which different energy wells can give rise to a significant trapping effect due to the difference in their local-minimum energy. Practically, γ should still be close to 1.0 (or gradually reduced to a value close to 1.0) to maintain a good sampling accuracy. In this example, $\gamma=1.0-10^{-7}$ is used.

In implementation, Brownian dynamics is used for constant temperature simulation. The equation of motion is $d\vec{x}/dt = \vec{F} + \vec{\eta}$, where $\vec{F}$ is the force derived from a molecular potential; $\vec{\eta}$ is a vector of Gaussian white noise specified by $\langle \eta_i(t) \eta_j(t') \rangle = 2T\delta(t-t')\delta_{ij}$; T is the current temperature. The time step is 3×10^{-3} . For polymer of 34 and 55 residues, the temperature range is (0.1,0.7). The method can easily locate the known lowest energy configurations with $E=-98.3571$ and -178.1339 , respectively.¹ No other lower energy state is discovered.

The polymer of 89 residues is more challenging and its lowest energy configuration has not been reported in literature to authors' best knowledge. In this case, the temperature range is $T=(0.15,0.4)$ with five windows, separated at $T=0.20, 0.23, 0.26$, and 0.3 . Within each window, a second order expansion in term of T , $\tilde{E}=a_0+a_1T$ is used. This expansion is more suitable than the expansion in terms of β at a low temperature because a_0 is roughly the ground state energy while a_1 is the average heat capacity. The weighting factor $\tilde{w}(T)dT \sim 1/\{1+[(T-T_c)/\Delta]^2\}dT$ concentrated on $T_c=0.25$ with a width $\Delta=0.1$ is used to accelerate the temperature-space random walk. In addition, since the goal is to find the ground state instead of calculating the free energy, we replaced the weighted averages by regular averages in Eq. (5) and remove the constraint Eq. (3) to make the averaging and updating process more stable. The lowest energy state found in this study is $E=-311.6134$. The corresponding configuration is shown in Fig. 3(a). Here, another state with a very close energy $E=-311.5391$ but with a very different configuration is also shown in Fig. 3(b). The result indicates an extreme ruggedness of the low energy landscape.

IV. CONCLUDING DISCUSSIONS

In summary, we demonstrated an enhanced sampling method using a generalized ensemble. The method computes the partition function by minimizing the difference between the derivative of the estimated partition function and that of the actual one. One advantage of the method is that it allows a large gap of the macroscopic variable λ of the partition function. For example, when $\lambda=\beta$, it can afford a much larger temperature gap than other tempering methods¹⁻³ that rely on overlap between distributions. This feature makes the method more suitable for handling larger systems with much narrower distributions of thermodynamic quantities. The method also delivers asymptotic convergence of the partition function, which makes it superior to other methods based on the WL recursive scheme.⁶ The efficiency of the method is demonstrated on the two dimensional Ising model and the off-lattice protein models.

One of the most important features of our method is its scalability to large systems. The method performs a random walk in the λ space, e.g., the reciprocal temperature β as in the examples, and the ratio of partition functions between two end points of a λ window is calculated by minimizing the difference between the estimated and measured (from simulation trajectory) values of the thermal conjugates of λ (the energy in the case of $\lambda=\beta$). By including parameters that correspond to higher-order derivatives of the partition function in the Langevin equation controlling the λ -space random walk, the profile of the partition function within the λ window is more accurately approximated and the Brownian motion in the λ space is augmented. Thus, as long as the thermal conjugates varies smoothly within a λ window, the method can bring an efficient sampling of the entire λ window. Such a feature makes the method much less sensitive to the size of system, as well as the size of λ window.

FIG. 3. (a) The lowest-energy configuration of 89 residue model protein, $E = -311.6137$. (b) A different configuration with a similar local minimal energy $E = -311.5391$ (black, A; white, B).

In this study, we demonstrated the efficiency of the method in the reciprocal temperature space ($\lambda = \beta$). However, λ can be other variables. In our previous study,¹ we used the volume. Besides, it can be other λ parameter commonly used in free energy simulation.^{4,5,12,13}

Strictly speaking, the current method does not satisfy detailed balance due to the use of runtime averages, as in other algorithms before convergence.^{1,5-7} However, as simulation progresses the correction to the existing averages continuously decreases, and the deviation from detailed balance is negligible in the asymptotic limit. On the other hand, the runtime averaging process is essential to continuously improve the estimate of the partition function.

ACKNOWLEDGMENTS

The authors acknowledge support of grants from the National Institutes of Health (Grant No. R01-GM067801), the National Science Foundation (Grant No. MCB-0818353), and the Welch Foundation (Grant No. Q-1512).

¹C. Zhang and J. Ma, *Phys. Rev. E* **76**, 036708 (2007); **79**, 016703 (2009); *J. Chem. Phys.* **129**, 134112 (2008).

²R. H. Swendsen and J. S. Wang, *Phys. Rev. Lett.* **57**, 2607 (1986); C. J. Geyer, *Proceedings of the 23rd Symposium on the Interface* (American Statistical Association, New York, 1991); K. Hukushima and K. Nemoto, *J. Phys. Soc. Jpn.* **65**, 1604 (1996); U. H. E. Hansmann, *Chem. Phys. Lett.* **281**, 140 (1997).

³A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, *J. Chem. Phys.* **96**, 1776 (1992); E. Marinari and G. Parisi, *Europhys. Lett.* **19**, 451 (1999).

⁴H. Li, M. Fajer, and W. Yang, *J. Chem. Phys.* **126**, 024106 (2007); **129**, 034105 (2008).

⁵E. 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).

⁶F. Wang and D. P. Landau, *Phys. Rev. Lett.* **86**, 2050 (2001); *Phys. Rev. E* **64**, 056101 (2001).

⁷R. E. Belardinelli and V. D. Pereyra, *Phys. Rev. E* **75**, 046701 (2007).

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

⁹F. H. Stillinger, T. Head-Gordon, and C. L. Hirshfeld, *Phys. Rev. E* **48**, 1469 (1993); A. Irbäck, C. Peterson, F. Potthast, and O. Sommelius, *J. Chem. Phys.* **107**, 273 (1997).

¹⁰S. Y. Kim, S. B. Lee, and J. Lee, *Phys. Rev. E* **72**, 011916 (2005); J. Lee, K. Joo, S. Y. Kim, and J. Lee, *J. Comput. Chem.* **29**, 2479 (2008).

¹¹J. G. Kim, J. E. Straub, and T. Keyes, *Phys. Rev. Lett.* **97**, 050601 (2006); *Phys. Rev. E* **76**, 011913 (2007).

¹²J. G. Kirkwood, *J. Chem. Phys.* **3**, 300 (1935); E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral, *Chem. Phys. Lett.* **156**, 472 (1989).

¹³G. M. Torrie and J. P. Valleau, *J. Comput. Phys.* **23**, 187 (1977); C. Bartels and M. Karplus, *J. Comput. Chem.* **18**, 1450 (1997); P. A. Bash, U. C. Singh, R. Langridge, and P. A. Kollman, *Science* **236**, 564 (1987); R. W. Zwanzig, *J. Chem. Phys.* **22**, 1420 (1954); D. J. Tobias and C. L. Brooks III, *Chem. Phys. Lett.* **142**, 472 (1987); X. J. Kong and C. L. Brooks, *J. Chem. Phys.* **105**, 2414 (1996); A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12562 (2002); L. Zheng, M. Chen, and W. Yang, *ibid.* **105**, 20227 (2008); T. P. Straatsma and J. A. McCammon, *Annu. Rev. Phys. Chem.* **43**, 407 (1992).