Effective sampling in the configurational space of a small peptide by the multicanonical-multioverlap algorithm

Satoru G. Itoh* and Yuko Okamoto†

Department of Physics, School of Science, Nagoya University, Nagoya, Aichi 464-8602, Japan
(Received 10 October 2006; published 22 August 2007)

We propose a generalized-ensemble algorithm, which we refer to as the multicanonical-multioverlap algorithm. By utilizing a non-Boltzmann weight factor, this method realizes a random walk in the multidimensional, energy-overlap space and explores widely in the configurational space including specific configurations, where the overlap of a configuration with respect to a reference state is a measure for structural similarity. We apply the multicanonical-multioverlap molecular dynamics method to a penta peptide, Met-enkephalin, in vacuum as a test system. We also apply the multicanonical and multioverlap molecular dynamics methods to this system for the purpose of comparisons. We see that the multicanonical-multioverlap molecular dynamics method realizes effective sampling in the configurational space including specific configurations more than the other two methods. Furthermore, from the results of the multicanonical-multioverlap molecular dynamics simulation, we obtain a local-minimum state of the Met-enkephalin system.

I. INTRODUCTION

In order to understand the protein folding mechanisms, it is essential that we investigate free-energy landscapes of protein systems. In experiments, however, to obtain the free-energy landscapes of protein systems is very difficult. Therefore, computer simulations are now widely used for problems of the protein folding mechanisms. In computer simulations, however, it is still difficult to obtain accurate free-energy landscapes of protein systems. Accordingly, many efforts are devoted to develop efficient simulation algorithms.

In complex systems such as proteins, we must realize effective samplings in the configurational space. In usual canonical-ensemble simulations [1–6], however, it is difficult to achieve this. This is because the usual canonical-ensemble simulations tend to become trapped in a few of many local-minimum states. To overcome these difficulties, the generalized-ensemble algorithms have been proposed (for a review, see, for instance, Ref. [7]).

The multicanonical algorithm [8–11] is one of the most well-known methods among the generalized-ensemble algorithms. In the multicanonical ensemble, we have a flat probability distribution of the potential energy. Therefore, multicanonical-ensemble simulations realize free random walks in the potential-energy space and have effective samplings in the configurational space. The random walk in the potential-energy space allows one to calculate various thermodynamic quantities as functions of temperature for a wide temperature range from the results of a single simulation run by the single-histogram reweighting techniques [12,13]. This method is suitable to widely sample the configurational space, but not to have samplings that focus on a specific configuration because of the very nature of the algorithm. Consequently, it is difficult to obtain an accurate free-energy

landscape around specific configurations in multicanonical simulations, while we can obtain the free-energy landscape over wide areas.

The multioverlap algorithm was recently proposed for the Monte Carlo (MC) method [14] and the molecular dynamics (MD) method [15,16] in order to investigate the stability of specific configurations and transition states among specific configurations. In the multioverlap ensemble, we have a flat probability distribution in the overlap space. The method aims at achieving effective samplings that focus on specific configurations. Accordingly, we can obtain an accurate free-energy landscape around the specific configurations and estimate correctly transition states among the specific configurations. The multioverlap method is especially useful for polymeric molecules such as proteins, and other similar methods [17–19] are developed for research of the phase transition of the crystalline system and applied to various systems [20,21]. In the multioverlap method, however, we do not have a random walk in the potential-energy space, and hence, thermodynamic quantities can be obtained at temperatures only near the simulation temperatures.

In this paper, we propose a simulation method, which we refer to as the multicanonical-multioverlap method, to sample widely the configurational space and effectively the vicinity of specific conformations of a protein. We apply the multicanonical-multioverlap MD method to Met-enkephalin in vacuum and test the effectiveness of the method by comparing the results with those of the multicanonical MD method and the multioverlap MD method.

In Sec. II we summarize the formulation of the multicanonical-multioverlap algorithms. We present the details of the three simulations that we performed and their results in Sec. III. Section IV is devoted to conclusions.

II. MULTICANONICAL-MULTIOVERLAP ALGORITHMS

A. Definition of dihedral-angle distance

We introduce a dihedral-angle distance, which is a complementary quantity of the overlap, as a reaction coordi-

*itoh@tb.phys.nagoya-u.ac.jp
†okamoto@phys.nagoya-u.ac.jp

nate. The overlap O with respect to a reference configuration is defined as follows [14,22]:

$$O = 1 - d, \qquad (1)$$

where d is the dihedral-angle distance given by

$$d = \frac{1}{n\pi} \sum_{i=1}^{n} d_a(v_i, v_i^0). \qquad (2)$$

Here, n is the total number of dihedral angles, v_i is the dihedral angle i, and v_i^0 is the dihedral angle i of the reference configuration. The distance $d_a(v_i, v_i^0)$ between two dihedral angles is given by

$$d_a(v_i, v_i^0) = \min(|v_i - v_i^0|, 2\pi - |v_i - v_i^0|). \qquad (3)$$

If $d=0$, from Eq. (2), all dihedral angles are coincident with those of the reference configuration, and the two structures are identical. The dihedral-angle distance is thus an indicator of how similar the conformation is to the reference conformation.

Sampling the vicinity of the reference configuration is equivalent to sampling around $d=0$. In order to realize a sampling that focuses on the reference configuration, therefore, we just have to sample the neighborhood of $d=0$. We then would like to realize an effective sampling, which covers widely the configurational space and focuses on the reference configuration. In other words, we want to sample effectively and widely the configurational space, including near $d=0$.

B. Effective sampling in the energy-overlap space

In the case of canonical ensemble at a constant temperature T_0, the probability distribution P_c of potential energy E is represented by the product of the density of states $n(E)$ and the Boltzmann weight factor $W_c(E; T_0)$:

$$P_c(E; T_0) = n(E) W_c(E; T_0) = n(E) e^{-\beta_0 E}, \qquad (4)$$

where β_0 is given by $\beta_0 = 1/k_B T_0$ (k_B is the Boltzmann constant). In the multicanonical ensemble at a constant temperature T_0, by employing the non-Boltzmann weight factor $W_{muca}(E)$, which we refer to as the multicanonical weight factor, a uniform probability distribution of potential energy is obtained:

$$P_{muca}(E) = n(E) W_{muca}(E) = n(E) e^{-\beta_0 E_{muca}(E)} \equiv \text{const}, \qquad (5)$$

where $E_{muca}(E)$ is the multicanonical potential energy. The multicanonical weight factor, or the multicanonical potential energy, is not *a priori* known and has to be determined by short preliminary simulations. There exist many methods for the determination of the weight factor (see, e.g., Refs. [23–31]). Equation (5) implies that multicanonical simulations realize a free random walk in the potential-energy space and are able to effectively sample the configurational space.

In the multioverlap ensemble at a constant temperature T_0, the probability distribution of dihedral-angle distances is defined by

$$P_{muov}(d_1, \ldots, d_N)$$

$$= \int dE \, n(E; d_1, \ldots, d_N) W_{muov}(E; d_1, \ldots, d_N)$$

$$= \int dE \, n(E; d_1, \ldots, d_N) e^{-\beta_0 E + f(d_1, \ldots, d_N)} \equiv \text{const}, \quad (6)$$

where d_i is the dihedral-angle distance with respect to reference configuration i ($i = 1, \ldots, N$), $n(E; d_1, \ldots, d_N)$ is the density of states, $W_{muov}(E; d_1, \ldots, d_N)$ is the multioverlap weight factor, and $f(d_1, \ldots, d_N)$ is the "dimensionless free energy." The dimensionless free energy is not *a priori* known and has to be determined by short preliminary simulations. This method performs a random walk in the N-dimensional dihedral-angle-distance space, in which the simulation visits the N reference configurations often. We can thus obtain accurate information about transition states among these N states.

We want simulations to have effective and wide samplings including near $d=0$ in the configurational space. Therefore, we consider to carry out a simulation that performs a random walk both in the potential-energy space and in the dihedral-angle-distance space (energy-overlap space). In other words, the simulation needs to have a constant probability distribution in the energy-overlap space. In analogy with the multicanonical ensemble in Eq. (5) or the multioverlap ensemble in Eq. (6), by employing the non-Boltzmann weight factor $W_{mco}(E, d_1, \ldots, d_N)$, which we refer to as the multicanonical-multioverlap weight factor, a uniform probability distribution with respect to the potential energy and dihedral-angle distances is obtained:

$$P_{mco}(E, d_1, \ldots, d_N) = n(E, d_1, \ldots, d_N) W_{mco}(E, d_1, \ldots, d_N)$$

$$\equiv \text{const.} \qquad (7)$$

The multicanonical-multioverlap weight factor is not *a priori* known and again has to be determined by short preliminary simulations. The multicanonical-multioverlap weight factor $W_{mco}(E, d_1, \ldots, d_N)$ at a constant temperature T_0 can be written as

$$W_{mco}(E, d_1, \ldots, d_N) = e^{-\beta_0 E_{mco}(E, d_1, \ldots, d_N)}, \qquad (8)$$

where $E_{mco}(E, d_1, \ldots, d_N)$ is the multicanonical-multioverlap potential energy. We remark that by definition the multicanonical weight factor in Eq. (5) and the multicanonical-multioverlap weight factor in Eq. (7) are independent of temperature, whereas the multioverlap weight factor in Eq. (6) depends on temperature.

C. Monte Carlo methods in the multicanonical-multioverlap ensemble

Canonical MC simulations are performed with the Metropolis criterion [1]. In the Metropolis criterion, the transition probability from state x with potential energy E to state x' with potential energy E' is given by

$$w(x \to x') = \begin{cases} 1, & \text{for } \Delta E \leq 0, \\ \exp(-\beta_0 \Delta E), & \text{for } \Delta E > 0, \end{cases} \quad (9)$$

where

$$\Delta E \equiv E' - E. \quad (10)$$

In multicanonical-multioverlap MC simulations, the transition probability from state x with potential energy E and dihedral-angle distances $d_1, \ldots, d_N$ to state x' with potential energy E' and dihedral-angle distances $d'_1, \ldots, d'_N$ is correspondingly given by

$$w(x \to x') = \begin{cases} 1, & \text{for } \Delta E_{mco} \leq 0, \\ \exp(-\beta_0 \Delta E_{mco}), & \text{for } \Delta E_{mco} > 0, \end{cases}$$
$$(11)$$

where E_{mco} is the multicanonical-multioverlap potential energy in Eq. (8) and

$$\Delta E_{mco} \equiv E'_{mco}(E', d'_1, \ldots, d'_N) - E_{mco}(E, d_1, \ldots, d_N). \quad (12)$$

We remark that another MC method, which realizes a random walk in two-dimensional energy-order-parameter space, was proposed by Wang and Landau [32].

D. Equations of motion in the multicanonical-multioverlap ensemble

Solving regular Newton's equations of motion leads to the microcanonical ensemble. There are several methods to realize the canonical ensemble by the MD simulation (see, for example, Refs. [2–6]). Here, we just consider one of these methods, namely the Gaussian constraint method [2,3] (the realization by other methods is also straightforward). In the Gaussian constraint method, the following equations of motion with Gaussian thermostat are solved:

$$\dot{\boldsymbol{q}}_i = \frac{d\boldsymbol{q}_i}{dt} = \frac{\boldsymbol{p}_i}{m_i},$$

$$\dot{\boldsymbol{p}}_i = \boldsymbol{F}_i - \zeta_c \boldsymbol{p}_i, \quad (13)$$

where m_i, $\boldsymbol{q}_i$, and $\boldsymbol{p}_i$ are the mass, coordinate vector, and momentum vector of atom i. The force $\boldsymbol{F}_i$ acting on atom i is given by

$$\boldsymbol{F}_i = -\frac{\partial E}{\partial \boldsymbol{q}_i}, \quad (14)$$

where E is the potential energy. The coefficient ζ_c is chosen so as to guarantee that the total kinetic energy is constant:

$$\zeta_c = \frac{\sum_i \boldsymbol{F}_i \cdot \dot{\boldsymbol{q}}_i}{2\sum_i \frac{\boldsymbol{p}_i^2}{2m_i}}. \quad (15)$$

Correspondingly, the molecular dynamics algorithm in the multicanonical-multioverlap ensemble naturally follows from Eq. (8) (see Refs. [10,11] for the case of multicanonical

MD). The multicanonical-multioverlap MD simulation is carried out by solving the following modified equations of motion with Gaussian thermostat:

$$\dot{\boldsymbol{q}}_i = \frac{d\boldsymbol{q}_i}{dt} = \frac{\boldsymbol{p}_i}{m_i},$$

$$\dot{\boldsymbol{p}}_i = \boldsymbol{F}_i^{mco} - \zeta_{mco} \boldsymbol{p}_i. \quad (16)$$

The "force" $\boldsymbol{F}_i^{mco}$ acting on atom i is calculated from

$$\boldsymbol{F}_i^{mco} = -\frac{\partial E_{mco}}{\partial \boldsymbol{q}_i}, \quad (17)$$

where E_{mco} is the multicanonical-multioverlap potential energy in Eq. (8). The coefficient ζ_{mco} is defined by

$$\zeta_{mco} = \frac{\sum_i \boldsymbol{F}_i^{mco} \cdot \dot{\boldsymbol{q}}_i}{2\sum_i \frac{\boldsymbol{p}_i^2}{2m_i}}. \quad (18)$$

E. Reweighting techniques

The results of the multicanonical-multioverlap production run can be analyzed by the reweighting techniques. Suppose that we have determined the multicanonical-multioverlap potential energy E_{mco} at a constant temperature T_0 and that we have made a production run at this temperature. The expectation value of a physical quantity A at any temperature T is calculated from

$$\langle A \rangle_T = \frac{\sum_{E, d_1, \ldots, d_N} A(E, d_1, \ldots, d_N) n(E, d_1, \ldots, d_N) e^{-\beta E}}{\sum_{E, d_1, \ldots, d_N} n(E, d_1, \ldots, d_N) e^{-\beta E}}, \quad (19)$$

where the best estimate of the density of states is given by the single-histogram reweighting techniques [12,13] [see Eq. (7)]:

$$n(E, d_1, \ldots, d_N) = \frac{N_{mco}(E, d_1, \ldots, d_N)}{W_{mco}(E, d_1, \ldots, d_N)}, \quad (20)$$

and $N_{mco}(E, d_1, \ldots, d_N)$ is the histogram of the probability distribution that was obtained by the multicanonical-multioverlap production run. By substituting Eqs. (8) and (20) into Eq. (19), we have

$$\langle A \rangle_T = \frac{\displaystyle\sum_{E,d_1,\ldots,d_N} A(E,d_1,\ldots,d_N) N_{mco}(E,d_1,\ldots,d_N) e^{\beta_0 E_{mco}(E,d_1,\ldots,d_N) - \beta E}}{\displaystyle\sum_{E,d_1,\ldots,d_N} N_{mco}(E,d_1,\ldots,d_N) e^{\beta_0 E_{mco}(E,d_1,\ldots,d_N) - \beta E}}. \tag{21}$$

We can also calculate the free energy (or, the potential of mean force) with appropriate reaction coordinates. For example, the free energy $F(\xi_1,\xi_2;T)$ with reaction coordinates ξ_1 and ξ_2 at temperature T is given by

$$F(\xi_1,\xi_2;T) = -k_B T \ln P_c(\xi_1,\xi_2;T), \tag{22}$$

where $P_c(\xi_1,\xi_2;T)$ is the reweighted canonical probability distribution of ξ_1 and ξ_2 and given by [see Eq. (21)]

$$P_c(\xi_1,\xi_2;T) = \frac{\displaystyle\sum_{E,d_1,\ldots,d_N} N_{mco}(\xi_1,\xi_2;E,d_1,\ldots,d_N) e^{\beta_0 E_{mco}(E,d_1,\ldots,d_N) - \beta E}}{\displaystyle\sum_{\xi_1,\xi_2,E,d_1,\ldots,d_N} N_{mco}(\xi_1,\xi_2;E,d_1,\ldots,d_N) e^{\beta_0 E_{mco}(E,d_1,\ldots,d_N) - \beta E}}, \tag{23}$$

and $N_{mco}(\xi_1,\xi_2;E,d_1,\ldots,d_N)$ is the histogram of the probability distribution that was obtained from the multicanonical-multioverlap production run.

These reweighting techniques are valid only in the range of variable where we have a sufficient number of nonzero entries in the histograms. Beyond this range, the density of states such as the one in Eq. (20) is not accurate and the average values thus calculated are not reliable. The larger the range that are covered by the simulation is the larger the applicability range of the reweighting techniques becomes. This gives a quantitative measure for the improvement of simulation algorithms.

III. APPLICATION TO MET-ENKEPHALIN IN GAS PHASE

A. Computational details

In order to demonstrate the effectiveness of the multicanonical-multioverlap MD method, we compare a multicanonical-multioverlap MD simulation with multicanonical and multioverlap MD simulations. We apply the three simulations to the system of Met-enkephalin in vacuum. Met-enkephalin is one of the simplest peptides and has the amino-acid sequence Tyr-Gly-Gly-Phe-Met. This peptide is often adopted as a test system in biomolecular simulations. In our simulations the N-terminus and the C-terminus were blocked with the acetyl group and the N-methyl group, respectively. This is because we wanted the total charge of the Met-enkephalin system to be neutral. The force field that we adopted is the CHARMM (Chemistry at Harvard Molecular Mechanics) param 22 parameter set [33]. Leap-frog algorithm [34] was employed for the numerical integration and the time step was taken to be 0.2 fs. The reason for using such a small time step is to perform simulations with high accuracy. In leap-frog algorithm, statistical errors are proportional to the square of a step size. Therefore, we can realize high-accuracy simulations if a step size is small.

In the multicanonical-multioverlap MD simulation, we must have a reference conformation. Therefore, we adopted

the conformation in Fig. 1 as the reference conformation and set $N=1$ in Eqs. (7) and (8). This conformation was obtained by the simulated annealing MD method [35] as follows: During the simulated annealing run, the temperature was decreased linearly from 1000 K to 100 K with an increment of 50 K, and the canonical MD simulations were performed for 200 ps at each temperature. This simulated annealing MD run was repeated 10 times with different initial random numbers. The obtained final conformations were further minimized by the conjugate gradient method, and we finally obtained two conformations from the results of these minimizations. One of the two conformations was the refer-
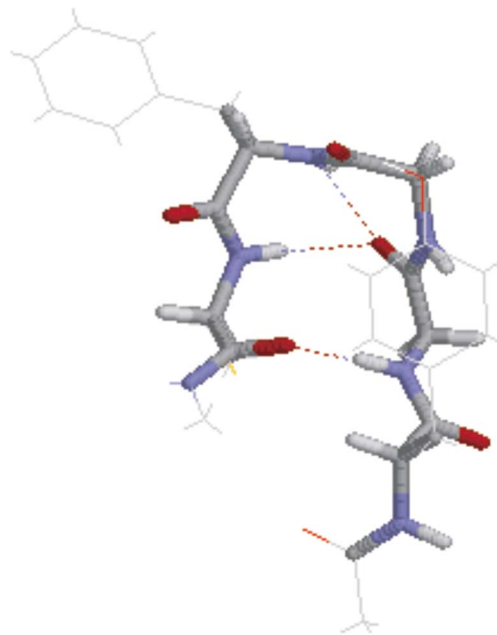

FIG. 1. (Color online) Reference configuration that was used in the multicanonical-multioverlap MD simulation. The dotted lines denote the hydrogen bonds. The N-terminus and the C-terminus are on the right-hand side and on the left-hand side, respectively. The figure was created with the computer program RASMOL [40].

ence conformation in Fig. 1. From our previous results of the multioverlap MD simulations [15,16], we see that this reference conformation is one of the conformations in a local-minimum state and another conformation obtained by the results of these minimizations is the global-minimum state with the CHARMM param 22 parameter set.

The backbone dihedral angles are of three types: the rotation angle around the N-C$^{\alpha}$ bond of the backbone (ϕ), that around the C$^{\alpha}$-C bond (ψ), and that around the peptide bond C-N (ω). Our multicanonical-multioverlap MD simulation was performed using the all-atom model, but we used only ϕ and ψ angles in the definition of the dihedral-angle distances in Eq. (2). This is because the dihedral angles of the backbone ω have almost the fixed value of 180° for the peptide bond C-N. Furthermore, by using only the backbone dihedral angles (and not side-chain dihedral angles) as the elements of the dihedral-angle distances, we focused on the backbone structures of Met-enkephalin. In Eq. (2), consequently, the number n of the elements of the dihedral-angle distances is ten because Met-enkephalin has five pairs of ϕ and ψ.

For the purpose of comparisons, we also performed a multicanonical MD simulation and a multioverlap MD simulation for 9 ns at $T_0 = 300$ K. We determined the multicanonical weight factor in Eq. (5) as follows. We carried out canonical MD simulations at eight temperatures between 300 K and 1000 K with equal increments of 100 K and obtained ensemble averages of the potential energy at each temperature. From the ensemble averages of the potential energy, we calculated the derivative of the multicanonical potential energy [29–31]:

$$\left.\frac{\partial E_{muca}(E)}{\partial E}\right|_{E=E_{ave}} = \frac{T_0}{T(E_{ave})}, \qquad (24)$$

with

$$E_{ave} = \langle E \rangle_{T(E_{ave})}. \qquad (25)$$

We integrated the derivative of the multicanonical potential energy and obtained the multicanonical weight factor. We adopted a random-coil conformation for the initial conformation of the multicanonical MD simulation production run.

In the multioverlap MD simulation, we employed the two-dimensional version of this method. In other words, we used two reference conformations in the multioverlap MD simulation and $N=2$ in Eq. (6). One of the two reference conformations was the conformation in Fig. 1, and the other one is shown in Fig. 2. This reference conformation in Fig. 2 is one of the conformations in the global-minimum state with the CHARMM param 22 parameter set (see Ref. [15,16]). We determined the multioverlap weight factor in Eq. (6) by the following process [23,24]. Suppose that we have the dimensionless free energy $f=f^{(l)}$ in the lth iteration of the short multioverlap MD simulation. In the $l+1$th iteration of the short multioverlap MD simulation, $f^{(l+1)}$ is calculated from

$$f^{(l+1)}(d_1,d_2) = f^{(l)}(d_1,d_2) - \ln N^{(l)}(d_1,d_2), \qquad (26)$$

where d_1 and d_2 are the dihedral-angle distance for reference conformation 1 (RC1) in Fig. 1 and reference conformation 2 (RC2) in Fig. 2, respectively. $N^{(l)}(d_1,d_2)$ in Eq. (26) is the
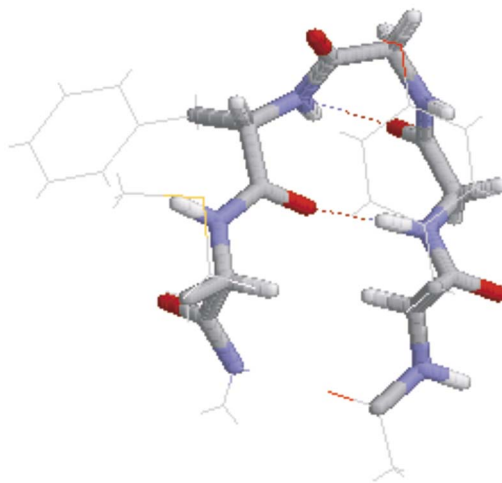

FIG. 2. (Color online) The other reference configuration that was used in the multioverlap MD simulation. See also the caption of Fig. 1.

histogram obtained from the results of the lth iteration. For this calculation, the dihedral-angle distances were discretized with a bin size of 0.01. Moreover, we interpolated the dimensionless free energy by a polynomial, following the techniques that were introduced in Ref. [36] [see Eq. (94) there]. The initial value was set as follows:

$$f^{(1)}(d_1,d_2) = 0. \qquad (27)$$

In the first iteration, therefore, we performed a short usual canonical MD simulation. We stopped the iterations after seven short multioverlap MD simulations each for 1.8 ns and we obtained the multioverlap weight factor. The initial conformation for the multioverlap production run was a conformation equilibrated at 300 K by the canonical simulation. Since the multioverlap weight factor has temperature dependence, it is appropriate that we employ this initial conformation.

The multicanonical-multioverlap MD simulation was carried out at $T_0 = 300$ K. We first have to determine the multicanonical-multioverlap weight factor $W_{mco}(E,d_1)$ in Eq. (8) to obtain a flat probability distribution in the energy-overlap space (E,d_1). For this purpose we used a similar procedure to that in Eq. (26). Namely, suppose that we have $E_{mco}=E_{mco}^{(l)}$ in the lth iteration of the short multicanonical-multioverlap MD simulation. In the $l+1$th iteration of the short multicanonical-multioverlap MD simulation, $E_{mco}^{(l+1)}$ is calculated from

$$E_{mco}^{(l+1)}(E,d_1) = E_{mco}^{(l)}(E,d_1) + k_B T_0 \ln N^{(l)}(E,d_1), \qquad (28)$$

where $N^{(l)}(E,d_1)$ is the histogram obtained from the results of the lth iteration. For this calculation, the potential energy and the dihedral-angle distance were discretized with a bin size of 1.0 kcal/mol and a bin size of 0.01, respectively. We also interpolated the multicanonical-multioverlap potential energy by the polynomial. The initial value was set as follows:
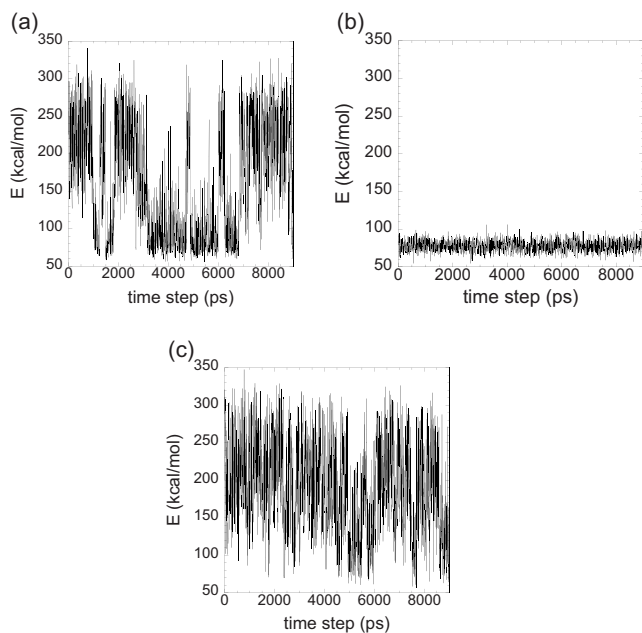
FIG. 3. The time series of the potential energy E. (a) Results from the multicanonical MD simulation, (b) the multioverlap MD simulation at $T_0 = 300$ K, and (c) the multicanonical-multioverlap MD simulation.

$$E_{mco}(E, d_1) = E_{muca}(E), \qquad (29)$$

where $E_{muca}(E)$ is the multicanonical weight factor that was determined as above. We then performed three iterations of the multicanonical-multioverlap MD simulations in Eq. (16) for 3 ns. The multicanonical-multioverlap weight factor $E_{mco}(E, d_1)$ was updated by Eq. (28) after each multicanonical-multioverlap MD simulation. Finally, the multicanonical-multioverlap MD production run was then performed with this weight factor for 9 ns after equilibration of 1 ns. For the initial conformation of the multicanonical-multioverlap MD simulation production run, we also adopted a random-coil conformation. We remark that the CPU costs are not much different among the three production runs. The CPU time for each of the three production runs was about 6 hours on a Xeon 3.20 GHz machine.

B. Comparisons of the three simulations

We first compare the time series of the potential energy and the dihedral-angle distance obtained from the multicanonical, multioverlap, and multicanonical-multioverlap MD simulations. Figure 3 shows the time series of the potential energy of the three simulations. The multicanonical and multicanonical-multioverlap MD simulations cover widely the potential-energy space, as we can see in Figs. 3(a) and 3(c). In other words, the two simulations realized free-random walks in the potential-energy space and sampled widely the conformational space. In the multioverlap MD simulation, however, we can sample only a narrow region in the potential-energy space as in Fig. 3(b). Therefore, in contrast with the other two simulations, the multioverlap MD
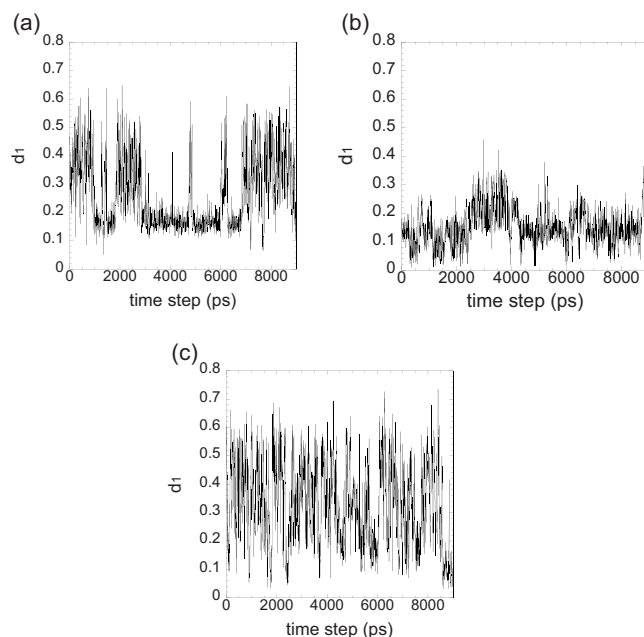

FIG. 4. The time series of the dihedral-angle distance d_1. (a) Results from the multicanonical MD simulation, (b) the multioverlap MD simulation at $T_0 = 300$ K, and (c) the multicanonical-multioverlap MD simulation.

simulations are not suitable to sample widely the conformational space.

In Fig. 4 we show the time series of the dihedral-angle distance d_1 with respect to conformation RC1 in Fig. 1. When $d_1 = 0$, the values of the backbone dihedral angles are completely coincident with those of RC1. In the multioverlap and multicanonical-multioverlap MD simulations, we see from Figs. 4(b) and 4(c) that the efficient samplings were realized in the neighborhood of $d_1 = 0$. In other words, the multioverlap and multicanonical-multioverlap MD simulations could sample efficiently the vicinity of RC1. The multicanonical MD simulation sampled infrequently the neighborhood of $d_1 = 0$, as we can see in Fig. 4(a). Thus, it is difficult to sample specific conformations in multicanonical MD simulations.

From Fig. 3(c) and Fig. 4(c) we see that the multicanonical-multioverlap MD simulation sampled both widely the conformational space and efficiently the vicinity of the reference conformation. Therefore, the multicanonical-multioverlap MD method has the advantages of both the multicanonical MD method and the multioverlap MD method.

We now consider the probability distributions of configurations from the three simulations. In Fig. 5 we show the raw data of the histograms with respect to the potential energy E and dihedral-angle-distance d_1. From Fig. 5(a) it is apparent that the multicanonical MD simulation had only a partial sampling in the vicinity of RC1 (low d_1 regions). Moreover, the multicanonical MD simulation did not sample widely in the conformational space at the low-energy region, although it had a wide sampling in the conformational space at the high-energy region. In the multioverlap MD simulation, on the other hand, the sampling that focuses on RC1 was real-
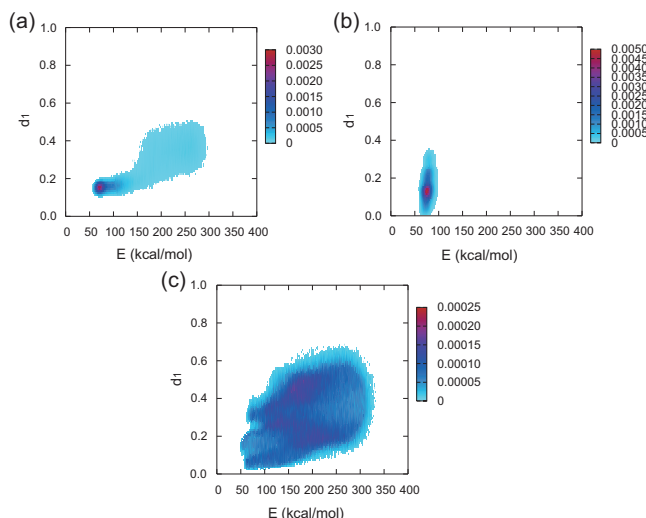
FIG. 5. (Color) The raw data of the probability distribution with respect to the potential energy E and dihedral-angle distance d_1. (a) Results from the multicanonical MD simulation, (b) the multioverlap MD simulation at $T_0 = 300$ K, and (c) the multicanonical-multioverlap MD simulation.

ized [see Fig. 5(b)]. At the high-energy region, however, we could not sample at all in the multioverlap MD simulation. From Fig. 5(c) we see that the multicanonical-multioverlap MD simulation performed the effective sampling in the conformational space in comparison with the other two simulations. In fact, we could sample widely the conformational space at the low-energy region as well as the high-energy region and the vicinity of RC1 in the multicanonical-multioverlap MD simulation.

C. Physical quantities calculated by the reweighting techniques

We present various physical quantities calculated from the results of the three simulations by the reweighting techniques. The reweighting techniques for the multicanonical-multioverlap MD method were explained in Sec. II E. The reweighting techniques for the other MD methods are accounted, for instance, in Refs. [7,15]. In Fig. 6 we show the average potential energy and specific heat calculated as functions of temperature by the reweighting techniques. Here, the error bars were calculated by the jackknife method [37–39]. The number of bins was taken to be eight. The specific heat here is defined by

$$C_v = \frac{1}{k_B}\frac{d\langle E\rangle_T}{dT} = \beta^2(\langle E^2\rangle_T - \langle E\rangle_T^2). \tag{30}$$

From Fig. 6 we see that the results from the multicanonical-multioverlap MD simulations coincide well with those from the multicanonical MD simulation. The results from the multioverlap MD simulation, however, are in agreement with those from the other two simulations only near $T = 300$ K. As mentioned in Sec. II E, the reweighting techniques are valid only in the range of variables where we have a sufficient number of nonzero entries in the histo-
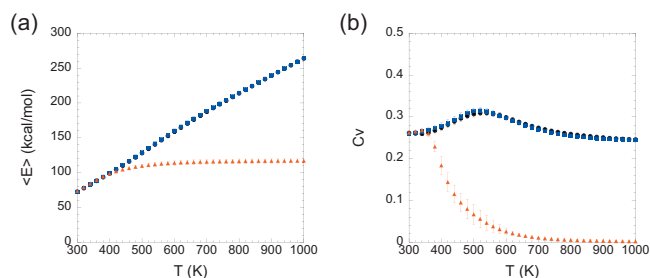

FIG. 6. (Color) (a) Average potential energy as a function of temperature, and (b) specific heat as a function of temperature. These results were calculated from the multicanonical MD simulation (blue square), the multioverlap MD simulation (red triangle), and the multicanonical-multioverlap MD simulation (black circle) by the reweighting techniques. The results from the multicanonical and multicanonical-multioverlap simulations are essentially identical.

grams. The multioverlap algorithm sampled only a narrow potential-energy range between about 60 kcal/mol and about 90 kcal/mol as shown in Fig. 3(b). Therefore, a reliable estimate of the density of states in Eq. (20) was obtained only in the narrow potential-energy range. Indeed, we see that the average potential energy was estimated correctly only in the corresponding range, namely, between about 60 kcal/mol and about 90 kcal/mol, from Fig. 6. As for the temperature range, the accurate average potential energy was obtained only between about 300 K and about 360 K. In Fig. 7 we show the reweighted canonical probability distributions of the potential energy at $T = 300$ K and $T = 400$ K. Reweighted canonical probability distributions of the potential energy $P_c(E)$ are given by

$$P_c(E) = \frac{\displaystyle\sum_{d_1,\ldots,d_N} n(E,d_1,\ldots,d_N)e^{-\beta E}}{\displaystyle\sum_{E,d_1,\ldots,d_N} n(E,d_1,\ldots,d_N)e^{-\beta E}}, \tag{31}$$

and $n(E,d_1,\ldots,d_N)$ is the density of states calculated from the reweighting techniques [see Eq. (20)]. From Fig. 7(a) the results from the multioverlap MD simulation well coincide
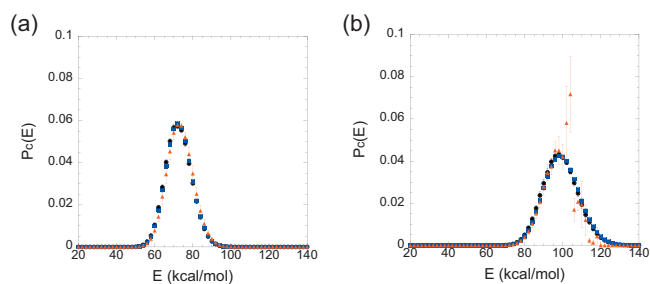

FIG. 7. (Color) Probability distributions of the potential energy at (a) $T = 300$ K and (b) $T = 400$ K. These results were calculated from the results of the multicanonical MD simulation (blue square), the multioverlap MD simulation (red triangle), and the multicanonical-multioverlap MD simulation (black circle) by the reweighting techniques.
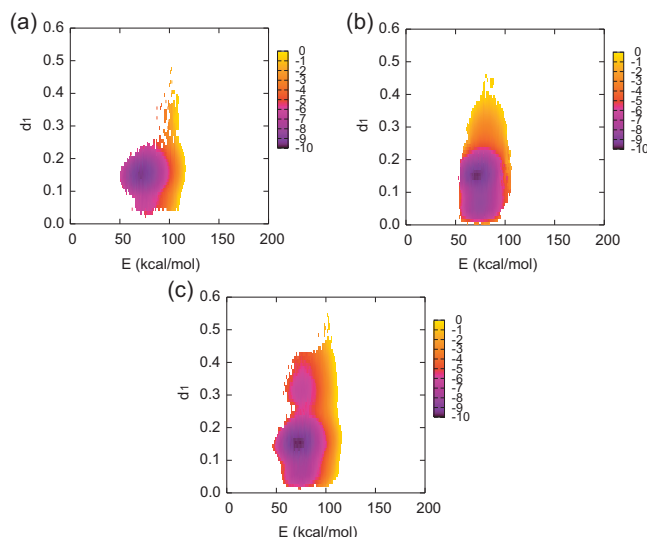
FIG. 8. (Color) The free-energy landscapes with respect to the potential energy E and dihedral-angle distance d_1 that were obtained from (a) the multicanonical MD simulation, (b) the multioverlap MD simulation at $T_0=300$ K, and (c) the multicanonical-multioverlap MD simulation.
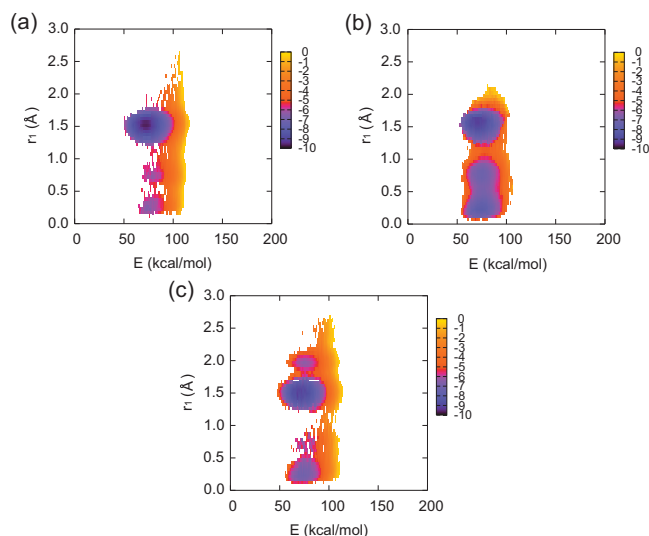

FIG. 9. (Color) The free-energy landscapes with respect to the potential energy E and RMSD r_1 that were obtained from (a) the multicanonical MD simulation, (b) the multioverlap MD simulation at $T_0=300$ K, and (c) the multicanonical-multioverlap MD simulation.

with those from the other two simulations at $T=300$ K. At $T=400$ K, however, the results from the multioverlap MD simulation were in disagreement with those from the other two simulations as shown in Fig. 7(b). In the high-energy region above 120 kcal/mol, especially, the reweighted canonical probability distribution of the potential energy obtained from the multioverlap MD simulation was underestimated including the error bars. The density of states was estimated wrongly as zero by the reweighting techniques because the multioverlap MD simulation had zero entries in the histograms in the high-energy region above 120 kcal/mol [see Eq. (20)]. Namely, the reason for the underestimate is not statistical errors but rather systematic errors in the reweighting techniques. The disagreement of the results from the multioverlap MD simulation in Fig. 6 also arose by the systematic errors in the reweighting techniques. Accordingly, we can obtain various physical quantities at any temperature in the multicanonical-multioverlap MD method as in the multicanonical MD method, although it is difficult to get such quantities at any temperature in the multioverlap MD method.

Subsequently, we present the free-energy landscape with respect to various reaction coordinates. The free-energy landscape was calculated from Eq. (22) with appropriate reaction coordinates by the reweighting techniques. In Fig. 8 we show the free-energy landscape at $T=300$ K obtained from the three simulations with respect to the potential energy E and dihedral-angle distance d_1. Since the multicanonical MD simulation did not sample the vicinity of RC1 and widely the conformational space at the low-energy region, the free-energy landscape at $T=300$ K was obtained only in a narrow region away from RC1 as shown in Fig. 8(a). In the multioverlap MD simulation, on the other hand, the free-energy landscape in Fig. 8(b) including the neighborhood of RC1 was calculated. From Fig. 8(c), the free-energy landscape

obtained from the multicanonical-multioverlap MD simulation was described over a wide range and near RC1. Moreover, we could identify a local-minimum state, which is located around $(E,d_1)=(75.0,0.32)$ in Fig. 8(c), although we did not find it by the other two methods. This is because the multicanonical-multioverlap MD simulation samples widely the conformational space at the low-energy region and effectively in the vicinity of RC1.

In Fig. 9 we also show the free-energy landscape at $T=300$ K obtained from the three simulations with respect to the potential energy and root-mean-square distance (RMSD) r_1. Here, the RMSD r_1 with respect to RC1 is defined by

$$r_1 = \min\left[\sqrt{\frac{1}{N}\sum_i (q_i - q_i^0)^2}\right], \qquad (32)$$

where N is the number of atoms, $\{q_i^0\}$ are the coordinates of RC1, and the minimization is taken over the rigid translations and rigid rotations of the coordinates of the configuration $\{q_i\}$. In this paper, we took into account only the backbone coordinates of Met-enkephalin in Eq. (32). In Fig. 9 we can also see a local-minimum state, which is located around $(E,r_1)=(75.0,2.0)$, only from the results of the multicanonical-multioverlap MD simulation.

By analyzing the conformations in the two local minima found in Figs. 8(c) and 9(c), we believe that they correspond to the same state. This is because common conformations were found in both states. In Fig. 10 we show a representative conformation in this local-minimum state. This conformation has a backbone hydrogen bond between hydrogen bond donor NH of Gly-2 and hydrogen bond acceptor CO of Met-5. The conformation also has a backbone hydrogen bond between CO of Try-1 and NH of Gly-3, but this hydrogen bond was frequently broken.
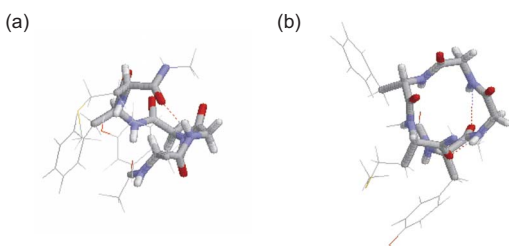
FIG. 10. (Color online) (a) A typical structure in the local-minimum state found in Figs. 8(c) and 9(c). (a) and (b) correspond to the same conformation viewed from different angles. See also the caption of Fig. 1.

IV. CONCLUSIONS

In this paper, we have proposed the multicanonical-multioverlap MD algorithm, which is useful to sample the conformational space widely and the vicinity of a reference conformation effectively. We applied this method to a penta-peptide system of Met-enkephalin in vacuum and compared the performance with those of multicanonical and multiover-lap MD methods. The multicanonical MD simulation sampled widely the conformational space at the high-energy region, but not at the low-energy region, and did not have the sampling around the reference conformation. The multiover-lap MD simulation could sample effectively the vicinity of the reference conformation. In the multioverlap MD simula-tion, however, we were not able to have the sampling in the high-energy region. On the other hand, the multicanonical-multioverlap MD simulation realized a free-random walk in the energy-overlap space and sampled the conformational space widely and the neighborhood of the reference confor-mation. Accordingly, we could obtain accurate free-energy landscape in the wide reaction-coordinate space including the vicinity of the reference conformation and discover a local-minimum state.

In the protein folding problem, the multicanonical-multioverlap method can be applied to deduce folding path-ways in which the protein system has an intermediate state such as a molten-globule state. This is because we can obtain free-energy landscapes, which include random-coil states, the native state, and the molten-globule state, from the re-sults of multicanonical-multioverlap simulations. Further-more, we can estimate transition states accurately between the native state (or denatured state) and the molten-globule state by employing the molten globule state as the reference conformation in multicanonical-multioverlap simulations.

[1] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[2] W. G. Hoover, A. J. C. Ladd, and B. Moran, Phys. Rev. Lett. **48**, 1818 (1982).
[3] D. J. Evans, J. Chem. Phys. **78**, 3297 (1983).
[4] S. Nosé, Mol. Phys. **52**, 255 (1984).
[5] S. Nosé, J. Chem. Phys. **81**, 511 (1984).
[6] W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).
[7] A. Mitsutake, Y. Sugita, and Y. Okamoto, Biopolymers **60**, 96 (2001).
[8] B. A. Berg and T. Neuhaus, Phys. Lett. B **267**, 249 (1991).
[9] B. A. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992).
[10] U. H. E. Hansmann, Y. Okamoto, and F. Eisenmenger, Chem. Phys. Lett. **259**, 321 (1996).
[11] N. Nakajima, H. Nakamura, and A. Kidera, J. Phys. Chem. B **101**, 817 (1997).
[12] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988).
[13] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1658 (1989).
[14] B. A. Berg, H. Noguchi, and Y. Okamoto, Phys. Rev. E **68**, 036126 (2003).
[15] S. G. Itoh and Y. Okamoto, Chem. Phys. Lett. **400**, 308 (2004).
[16] S. G. Itoh and Y. Okamoto, J. Chem. Phys. **124**, 104103 (2006).
[17] J. S. van Duijneveldt and D. Frenkel, J. Chem. Phys. **96**, 4655 (1992).
[18] E. B. Kim, R. Faller, Q. Yan, N. L. Abbott, and J. J. de Pablo, J. Chem. Phys. **117**, 7781 (2002).
[19] M. Chopra, M. Müller, and J. J. de Pablo, J. Chem. Phys. **124**, 134102 (2006).
[20] R. Radhakrishnan and K. E. Gubbins, Mol. Phys. **96**, 1249 (1999).
[21] R. Radhakrishnan and B. L. Trout, J. Chem. Phys. **117**, 1786 (2002).
[22] U. H. E. Hansmann, M. Masuya, and Y. Okamoto, Proc. Natl. Acad. Sci. U.S.A. **94**, 10652 (1997).
[23] B. A. Berg and T. Celik, Phys. Rev. Lett. **69**, 2292 (1992).
[24] Y. Okamoto and U. H. E. Hansmann, J. Phys. Chem. **99**, 11276 (1995).
[25] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).
[26] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **329**, 261 (2000).
[27] A. Mitsutake, Y. Sugita, and Y. Okamoto, J. Chem. Phys. **118**, 6664 (2003).
[28] A. Mitsutake, Y. Sugita, and Y. Okamoto, J. Chem. Phys. **118**, 6676 (2003).
[29] U. H. E. Hansmann, Phys. Rev. E **56**, 6200 (1997).
[30] Y. Sugita and Y. Okamoto, in *Lecture Notes in Computatinal Science and Engineering*, edited by T. Schlick and H. H. Gan

(Springer-Verlag, Berlin, 2002), pp. 304–332; see also e-print arXiv:cond-mat/0102296.

[31] T. Terada, Y. Matsuo, and A. Kidera, J. Chem. Phys. **118**, 4306 (2003).

[32] F. Wang and D. P. Landau, Phys. Rev. E **64**, 056101 (2001).

[33] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, J. Phys. Chem. B **102**, 3586 (1998).

[34] D. Brown and J. H. R. Clarke, Mol. Phys. **51**, 1243 (1984).

[35] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, Science **220**, 671 (1983).

[36] H. Okumura and Y. Okamoto, J. Comput. Chem. **27**, 379 (2006).

[37] M. H. Quenouille, Biometrika **43**, 353 (1956).

[38] R. G. Miller, Biometrika **61**, 1 (1974).

[39] B. A. Berg, *Markov Chain Monte Carlo Simulations and Their Statistical Analysis* (World Scientific, Singapore, 2004).

[40] R. A. Sayle and E. J. Milner-White, Trends Biochem. Sci. **20**, 374 (1995).