Generalized-Ensemble Algorithms for Molecular Simulations of Biopolymers

Ayori Mitsutake,[a,b] Yuji Sugita,[a,b] and Yuko Okamoto[a,b,1]

[a] *Department of Theoretical Studies*
Institute for Molecular Science
Okazaki, Aichi 444-8585, Japan
and
[b] *Department of Functional Molecular Science*
The Graduate University for Advanced Studies
Okazaki, Aichi 444-8585, Japan

cond-mat/0012021, revised; *Biopolymers (Peptide Science)* **60**, 96–123 (2001).

Keywords: *protein folding; generalized-ensemble algorithm; multicanonical algorithm; simulated tempering; replica-exchange method; parallel tempering*

ABSTRACT

In complex systems with many degrees of freedom such as peptides and proteins there exist a huge number of local-minimum-energy states. Conventional simulations in the canonical ensemble are of little use, because they tend to get trapped in states of these energy local minima. A simulation in generalized ensemble performs a random walk in potential energy space and can overcome this difficulty. From only one simulation run, one can obtain canonical-ensemble averages of physical quantities as functions of temperature by the single-histogram and/or multiple-histogram reweighting techniques. In this article we review uses of the generalized-ensemble algorithms in biomolecular systems. Three well-known methods, namely, multicanonical algorithm, simulated tempering, and replica-exchange method, are described first. Both Monte Carlo and molecular dynamics versions of the algorithms are given. We then present three new generalized-ensemble algorithms which combine the merits of the above methods. The effectiveness of the methods for molecular simulations in the protein folding problem is tested with short peptide systems.

[1] *Correspondence to:* Y. Okamoto, Department of Theoretical Studies, Institute for Molecular Science, Okazaki, Aichi 444-8585, Japan e-mail: okamotoy@ims.ac.jp

1 INTRODUCTION

Despite the great advancement of computer technology in the past decades, simulations of complex systems such as spin glasses and biopolymers are still greatly hampered by the multiple-minima problem. It is very difficult to obtain accurate canonical distributions at low temperatures by conventional Monte Carlo (MC) and molecular dynamics (MD) methods. This is because simulations at low temperatures tend to get trapped in one of huge number of local-minimum-energy states. The results thus will depend strongly on the initial conditions. One way to overcome this multiple-minima problem is to perform a simulation in a *generalized ensemble* where each state is weighted by a non-Boltzmann probability weight factor so that a random walk in potential energy space may be realized. The random walk allows the simulation to escape from any energy barrier and to sample much wider phase space than by conventional methods. Monitoring the energy in a single simulation run, one can obtain not only the global-minimum-energy state but also canonical ensemble averages as functions of temperature by the single-histogram [1] and/or multiple-histogram [2, 3] reweighting techniques (an extension of the multiple-histogram method is also referred to as *weighted histogram analysis method* (WHAM) [3]).

One of the most well-known generalized-ensemble methods is perhaps *multicanonical algorithm* (MUCA) [4, 5] (for a recent review, see Ref. [6]). (The method is also referred to as *entropic sampling* [7] and *adaptive umbrella sampling* [8] *of the potential energy* [9]. The mathematical equivalence of multicanonical algorithm and entropic sampling has been given in Ref. [10].) MUCA and its generalizations have been applied to spin glass systems (see, e.g., Refs. [11]–[14]). MUCA was also introduced to the molecular simulation field [15] (for previous reviews of generalized-ensemble approach in the protein folding problem, see, e.g., Refs. [16]–[18]). Since then MUCA has been extensively used in many applications in protein and related systems [19]–[45]. Molecular dynamics version of MUCA has also been developed [27, 28, 9] (see also Refs. [46, 27] for Langevin dynamics version). Moreover, multidimensional (or multicomponent) extensions of MUCA can be found in Refs. [26, 30, 34].

While a simulation in multicanonical ensemble performs a free 1D random walk in potential energy space, that in *simulated tempering* (ST) [47, 48] (the method is also referred to as the *method of expanded ensemble* [47]) performs a free random walk in temperature space (for a review, see, e.g., Ref. [49]). This random walk, in turn, induces a random walk in potential energy space and allows the simulation to escape from states of energy local minima. ST has also been applied to protein folding problem [50]–[53].

A third generalized-ensemble algorithm that is related to MUCA is *1/k-sampling* [54]. A simulation in 1/k-sampling performs a free random walk in entropy space, which, in turn, induces a random walk in potential energy space. The relation among the above three generalized-ensemble algorithms was discussed and the effectiveness of the three methods in protein folding problem was compared [52].

The generalized-ensemble method is powerful, but in the above three methods the probability weight factors are not *a priori* known and have to be determined by iterations of short trial simulations. This process can be non-trivial and very tedius for complex systems with many local-minimum-energy states. Therefore, there have been attempts to accelerate the convergence of the iterative process for MUCA [11, 26, 55, 56, 57, 9] (see also Ref. [6]).

A new generalized-ensemble algorithm that is based on the weight factor of Tsallis sta-

tistical mechanics [58] was recently developed with the hope of overcoming this difficulty [59, 60], and the method was applied to a peptide folding problem [61, 62]. A similar but slightly different formulation is given in Ref. [63]. See also Ref. [64] for a combination of Tsallis statistics with simulated tempering. (Optimization problems were also addressed by simulated annealing algorithms [65] based on the Tsallis weight in Refs. [66]–[68]. For reviews of molecular simulations based on Tsallis statistics, see, e.g., Refs. [69]–[71].) In this generalized ensemble the weight factor is known, once the value of the global-minimum energy is given [59]. The advantage of this ensemble is that it greatly simplifies the determination of the weight factor. However, the estimation of the global-minimum energy can still be very difficult.

In the *replica-exchange method* (REM) [72]–[74], the difficulty of weight factor determination is greatly alleviated. (Closely related methods were independently developed in Refs. [75]–[77]. REM is also referred to as *multiple Markov chain method* [78] and *parallel tempering* [49]. Details of literature about REM and related algorithms can be found in a recent review [79].) In this method, a number of non-interacting copies (or replicas) of the original system at different temperatures are simulated independently and simultaneously by the conventional MC or MD method. Every few steps, pairs of replicas are exchanged with a specified transition probability. The weight factor is just the product of Boltzmann factors, and so it is essentially known.

REM has already been used in many applications in protein systems [80, 81, 53, 82, 83, 84, 85]. Systems of Lennard-Jones particles have also been studied by this method in various ensembles [86]–[89]. Moreover, REM was applied to cluster studies in quantum chemistry field [90]. The details of molecular dynamics algorithm have been worked out for REM [82] (see also Refs. [80, 91]). We then developed a multidimensional REM which is particularly useful in free energy calculations [84] (see also Refs. [92, 86, 93]).

However, REM also has a computational difficulty: As the number of degrees of freedom of the system increases, the required number of replicas also greatly increases, whereas only a single replica is simulated in MUCA or ST. This demands a lot of computer power for complex systems. Our solution to this problem is: Use REM for the weight factor determinations of MUCA or ST, which is much simpler than previous iterative methods of weight determinations, and then perform a long MUCA or ST production run. The first example is the *replica-exchange multicanonical algorithm* (REMUCA) [94]. In REMUCA, a short replica-exchange simulation is performed, and the multicanonical weight factor is determined by the multiple-histogram reweighting techniques [2, 3]. Another example of such a combination is the *replica-exchange simulated tempering* (REST) [95]. In REST, a short replica-exchange simulation is performed, and the simulated tempering weight factor is determined by the multiple-histogram reweighting techniques [2, 3].

We have introduced a further extension of REMUCA, which we refer to as *multicanonical replica-exchange method* (MUCAREM) [94]. In MUCAREM, the multicanonical weight factor is first determined as in REMUCA, and then a replica-exchange multicanonical production simulation is performed with a small number of replicas.

In this article, we describe the six generalized-ensemble algorithms mentioned above. Namely, we first review three familiar methods: MUCA, ST, and REM. We then present the three new algorithms: REMUCA, REST, and MUCAREM. The effectiveness of these methods is tested with short peptide systems.

2 GENERALIZED-ENSEMBLE ALGORITHMS

2.1 Multicanonical Algorithm and Simulated Tempering

Let us consider a system of N atoms of mass m_k $(k = 1, \cdots, N)$ with their coordinate vectors and momentum vectors denoted by $q \equiv \{\boldsymbol{q}_1, \cdots, \boldsymbol{q}_N\}$ and $p \equiv \{\boldsymbol{p}_1, \cdots, \boldsymbol{p}_N\}$, respectively. The Hamiltonian $H(q, p)$ of the system is the sum of the kinetic energy $K(p)$ and the potential energy $E(q)$:

$$H(q, p) = K(p) + E(q) \ , \tag{1}$$

where

$$K(p) = \sum_{k=1}^{N} \frac{\boldsymbol{p}_k{}^2}{2m_k} \ . \tag{2}$$

In the canonical ensemble at temperature T each state $x \equiv (q, p)$ with the Hamiltonian $H(q, p)$ is weighted by the Boltzmann factor:

$$W_{\rm B}(x; T) = e^{-\beta H(q,p)} \ , \tag{3}$$

where the inverse temperature β is defined by $\beta = 1/k_{\rm B}T$ ($k_{\rm B}$ is the Boltzmann constant). The average kinetic energy at temperature T is then given by

$$\langle \ K(p) \ \rangle_T = \left\langle \sum_{k=1}^{N} \frac{\boldsymbol{p}_k{}^2}{2m_k} \right\rangle_T = \frac{3}{2} N k_{\rm B} T \ . \tag{4}$$

Because the coordinates q and momenta p are decoupled in Eq. (1), we can suppress the kinetic energy part and can write the Boltzmann factor as

$$W_{\rm B}(x; T) = W_{\rm B}(E; T) = e^{-\beta E} \ . \tag{5}$$

The canonical probability distribution of potential energy $P_{\rm B}(E; T)$ is then given by the product of the density of states $n(E)$ and the Boltzmann weight factor $W_{\rm B}(E; T)$:

$$P_{\rm B}(E; T) \propto n(E) W_{\rm B}(E; T) \ . \tag{6}$$

Since $n(E)$ is a rapidly increasing function and the Boltzmann factor decreases exponentially, the canonical ensemble yields a bell-shaped distribution which has a maximum around the average energy at temperature T. The conventional MC or MD simulations at constant temperature are expected to yield $P_{\rm B}(E; T)$, but, in practice, it is very difficult to obtain accurate canonical distributions of complex systems at low temperatures by conventional simulation methods. This is because simulations at low temperatures tend to get trapped in one or a few of local-minimum-energy states.

In the multicanonical ensemble (MUCA) [4, 5], on the other hand, each state is weighted by a non-Boltzmann weight factor $W_{\rm mu}(E)$ (which we refer to as the *multicanonical weight factor*) so that a uniform energy distribution $P_{\rm mu}(E)$ is obtained:

$$P_{\rm mu}(E) \propto n(E) W_{\rm mu}(E) \equiv \text{constant} \ . \tag{7}$$

The flat distribution implies that a free random walk in the potential energy space is realized in this ensemble. This allows the simulation to escape from any local minimum-energy

states and to sample the configurational space much more widely than the conventional canonical MC or MD methods.

From the definition in Eq. (7) the multicanonical weight factor is inversely proportional to the density of states, and we can write it as follows:

$$W_{\rm mu}(E) \equiv e^{-\beta_0 E_{\rm mu}(E;T_0)} = \frac{1}{n(E)} \ , \tag{8}$$

where we have chosen an arbitrary reference temperature, $T_0 = 1/k_{\rm B}\beta_0$, and the "*multicanonical potential energy*" is defined by

$$E_{\rm mu}(E;T_0) = k_{\rm B}T_0 \ln n(E) = T_0 S(E) \ . \tag{9}$$

Here, $S(E)$ is the entropy in the microcanonical ensemble. Since the density of states of the system is usually unknown, the multicanonical weight factor has to be determined numerically by iterations of short preliminary runs [4, 5] as described in detail below.

A multicanonical Monte Carlo simulation is performed, for instance, with the usual Metropolis criterion [96]: The transition probability of state x with potential energy E to state x' with potential energy E' is given by

$$w(x \to x') = \begin{cases} 1 \ , & \text{for } \Delta E_{\rm mu} \le 0 \ , \\ \exp\left(-\beta_0 \Delta E_{\rm mu}\right) \ , & \text{for } \Delta E_{\rm mu} > 0 \ , \end{cases} \tag{10}$$

where

$$\Delta E_{\rm mu} \equiv E_{\rm mu}(E';T_0) - E_{\rm mu}(E;T_0) \ . \tag{11}$$

The molecular dynamics algorithm in multicanonical ensemble also naturally follows from Eq. (8), in which the regular constant temperature molecular dynamics simulation (with $T = T_0$) is performed by solving the following modified Newton equation: [27, 28, 9]

$$\dot{\boldsymbol{p}}_k \ = \ -\frac{\partial E_{\rm mu}(E;T_0)}{\partial \boldsymbol{q}_k} \ = \ \frac{\partial E_{\rm mu}(E;T_0)}{\partial E} \ \boldsymbol{f}_k \ , \tag{12}$$

where $\boldsymbol{f}_k$ is the usual force acting on the k-th atom ($k = 1, \cdots, N$). From Eq. (9) this equation can be rewritten as

$$\dot{\boldsymbol{p}}_k \ = \ \frac{T_0}{T(E)} \ \boldsymbol{f}_k \ , \tag{13}$$

where the following thermodynamic relation gives the definition of the "effective temperature" $T(E)$:

$$\left.\frac{\partial S(E)}{\partial E}\right|_{E=E_a} \ = \ \frac{1}{T(E_a)} \ , \tag{14}$$

with

$$E_a \ = \ < E >_{T(E_a)} \ . \tag{15}$$

The multicanonical weight factor is usually determined by iterations of short trial simulations. The details of this process are described, for instance, in Refs. [11, 22]. For the first run, a canonical simulation at a sufficiently high temperature T_0 is performed, i.e., we set

$$\begin{cases} E^{(1)}_{\rm mu}(E;T_0) \ = \ E \ , \\ W^{(1)}_{\rm mu}(E;T_0) \ = \ W_{\rm B}(E;T_0) \ = \ \exp\left(-\beta_0 E\right) \ . \end{cases} \tag{16}$$

We define the maximum energy value $E_{\rm max}$ under which we want to have a flat energy distribution by the average potential energy at temperature T_0:

$$E_{\rm max} =< E >_{T_0} \ . \tag{17}$$

Above $E_{\rm max}$ we have the canonical distribution at $T = T_0$. In the ℓ-th iteration a simulation with the weight $W_{\rm mu}^{(\ell)}(E;T_0) = \exp\left(-\beta_0 E_{\rm mu}^{(\ell)}(E;T_0)\right)$ is performed, and the histogram $N^{(\ell)}(E)$ of the potential energy distribution $P_{\rm mu}^{(\ell)}(E)$ is obtained. Let $E_{\rm min}^{(\ell)}$ be the lowest-energy value that was obtained throughout the preceding iterations including the present simulation. The multicanonical potential energy for the $(\ell+1)$-th iteration is then given by

$$E_{\rm mu}^{(\ell+1)}(E;T_0) = \begin{cases} E \ , & \text{for } E \geq E_{\rm max}, \\ E_{\rm mu}^{(\ell)}(E;T_0) + k_{\rm B}T_0 \ln N^{(\ell)}(E) - c^{(\ell)} \ , & \text{for } E_{\rm min}^{(\ell)} \leq E < E_{\rm max}, \\ \left.\dfrac{\partial E_{\rm mu}^{(\ell+1)}(E;T_0)}{\partial E}\right|_{E=E_{\rm min}^{(\ell)}} \left(E - E_{\rm min}^{(\ell)}\right) + E_{\rm mu}^{(\ell+1)}(E_{\rm min}^{(\ell)};T_0), & \text{for } E < E_{\rm min}^{(\ell)}, \end{cases} \tag{18}$$

where the constant $c^{(\ell)}$ is introduced to ensure the continuity at $E = E_{\rm max}$ and we have

$$c^{(\ell)} = k_{\rm B}T_0 \ln N^{(\ell)}(E_{\rm max}) \ . \tag{19}$$

We iterate this process until the obtained energy distribution becomes reasonably flat, say, of the same order of magnitude, for $E < E_{\rm max}$. When the convergence is reached, we should have that $E_{\rm min}^{(\ell)}$ is equal to the global-minimum potential energy value.

It is also common especially when working in MD algorithm to use polynomials and other smooth functions to fit the histograms during the iterations [23, 28, 9]. We have shown that the cubic spline functions work well [94].

However, the iterative process can be non-trivial and very tedius for complex systems, and there have been attempts to accelerate the convergence of the iterative process [11, 26, 55, 56, 57, 9].

After the optimal multicanonical weight factor is determined, one performs a long multicanonical simulation once. By monitoring the potential energy throughout the simulation, one can find the global-minimum-energy state. Moreover, by using the obtained histogram $N_{\rm mu}(E)$ of the potential energy distribution $P_{\rm mu}(E)$ the expectation value of a physical quantity A at any temperature $T = 1/k_{\rm B}\beta$ is calculated from

$$< A >_T = \frac{\sum\limits_E A(E)\ n(E)\ e^{-\beta E}}{\sum\limits_E n(E)\ e^{-\beta E}} \ , \tag{20}$$

where the best estimate of the density of states is given by the single-histogram reweighting techniques (see Eq. (7)) [1]:

$$n(E) = \frac{N_{\rm mu}(E)}{W_{\rm mu}(E)} \ . \tag{21}$$

In the numerical work, we want to avoid round-off errors (and overflows and underflows) as much as possible. It is usually better to combine exponentials as follows (see Eq. (8)):

$$< A >_T = \frac{\sum\limits_E A(E)\ N_{\rm mu}(E)\ e^{\beta_0 E_{\rm mu}(E;T_0)-\beta E}}{\sum\limits_E N_{\rm mu}(E)\ e^{\beta_0 E_{\rm mu}(E;T_0)-\beta E}} \ . \tag{22}$$

We now briefly review the original *simulated tempering* (ST) method [47, 48]. In this method temperature itself becomes a dynamical variable, and both the configuration and the temperature are updated during the simulation with a weight:

$$W_{\rm ST}(E;T) = e^{-\beta E + a(T)} \ , \tag{23}$$

where the function $a(T)$ is chosen so that the probability distribution of temperature is flat:

$$P_{\rm ST}(T) = \int dE \ n(E) \ W_{\rm ST}(E;T) = \int dE \ n(E) \ e^{-\beta E + a(T)} = \text{constant} \ . \tag{24}$$

Hence, in simulated tempering the *temperature* is sampled uniformly. A free random walk in temperature space is realized, which in turn induces a random walk in potential energy space and allows the simulation to escape from states of energy local minima.

In the numerical work we discretize the temperature in M different values, T_m ($m = 1, \cdots, M$). Without loss of generality we can order the temperature so that $T_1 < T_2 < \cdots < T_M$. The lowest temperature T_1 should be sufficiently low so that the simulation can explore the global-minimum-energy region, and the highest temperature T_M should be sufficiently high so that no trapping in an energy-local-minimum state occurs. The probability weight factor in Eq. (23) is now written as

$$W_{\rm ST}(E;T_m) = e^{-\beta_m E + a_m} \ , \tag{25}$$

where $a_m = a(T_m)$ ($m = 1, \cdots, M$). The parameters a_m are not known *a priori* and have to be determined by iterations of short simulations. This process can be non-trivial and very difficult for complex systems. Note that from Eqs. (24) and (25) we have

$$e^{-a_m} \propto \int dE \ n(E) \ e^{-\beta_m E} \ . \tag{26}$$

The parameters a_m are therefore "dimensionless" Helmholtz free energy at temperature T_m (i.e., the inverse temperature β_m multiplied by the Helmholtz free energy).

Once the parameters a_m are determined and the initial configuration and the initial temperature T_m are chosen, a simulated tempering simulation is then realized by alternately performing the following two steps [47, 48]:

1. A canonical MC or MD simulation at the fixed temperature T_m is carried out for a certain MC or MD steps.

2. The temperature T_m is updated to the neighboring values $T_{m\pm1}$ with the configuration fixed. The transition probability of this temperature-updating process is given by the Metropolis criterion (see Eq. (25)):

$$w(T_m \to T_{m\pm1}) = \begin{cases} 1 \ , & \text{for } \Delta \le 0 \ , \\ \exp(-\Delta) \ , & \text{for } \Delta > 0 \ , \end{cases} \tag{27}$$

   where

$$\Delta = (\beta_{m\pm1} - \beta_m) \, E - (a_{m\pm1} - a_m) \ . \tag{28}$$

Note that in Step 2 we exchange only pairs of neighboring temperatures in order to secure sufficiently large acceptance ratio of temperature updates.

As in multicanonical algorithm, the simulated tempering parameters $a_m = a(T_m)$ $(m = 1, \cdots, M)$ are also determined by iterations of short trial simulations (see, e.g., Refs. [49, 50, 52] for details). Here, we give the one in Ref. [52].

During the trial simulations we keep track of the temperature distribution as a histogram $N_m = N(T_m)$ $(m = 1, \cdots, M)$.

1. Start with a short canonical simulation (i.e., $a_m = 0$) updating only configurations at temperature $T_m = T_M$ (we initially set the temperature label m to M) and calculate the average potential energy $< E >_{T_M}$. Here, the histogram N_n will have non-zero entry only for $n = m = M$.

2. Calculate new parameters a_n according to

$$a_n = \begin{cases} a_n - \ln N_n \ , & \text{for} \ \ m \le n \le M \ , \\ a_n - < E >_{T_m} (\beta_{m-1} - \beta_m) \ , & \text{for} \ \ n = m - 1 \ , \\ -\infty \ , & \text{for} \ \ n < m - 1 \ . \end{cases} \tag{29}$$

   This weight implies that the temperature will range between T_{m-1} and T_M.

3. Start a new simulation, now updating both configurations and temperatures, with weight $W_{\rm ST}(E; T_n) = e^{-\beta_n E + a_n}$ and sample the distribution of temperatures T_n in the histogram $N_n = N(T_n)$. For $T = T_{m-1}$ calculate the average potential energy $< E >_{T_{m-1}}$.

4. If the histogram N_n is approximately flat in the temperature range $T_{m-1} \le T_n \le T_M$, set $m = m - 1$. Otherwise, leave m unchanged.

5. Iterate the last three steps until the obtained temperature distribution N_n becomes flat over the whole temperature range $[T_1, T_M]$.

After the optimal simulated tempering weight factor is determined, one performs a long simulated tempering run once. From the results of this production run, one can obtain the canonical ensemble average of a physical quantity A as a function of temperature from Eq. (20), where the density of states is given by the multiple-histogram reweighting techniques [2, 3] as follows. Let $N_m(E)$ and n_m be respectively the potential-energy histogram and the total number of samples obtained at temperature $T_m = 1/k_{\rm B}\beta_m$ $(m = 1, \cdots, M)$. The best estimate of the density of states is then given by [2, 3]

$$n(E) = \frac{\displaystyle\sum_{m=1}^{M} g_m^{-1} \ N_m(E)}{\displaystyle\sum_{m=1}^{M} g_m^{-1} \ n_m \ e^{f_m - \beta_m E}} \ , \tag{30}$$

where

$$e^{-f_m} = \sum_E \ n(E) \ e^{-\beta_m E} \ . \tag{31}$$

Here, $g_m = 1 + 2\tau_m$, and τ_m is the integrated autocorrelation time at temperature T_m. Note that Eqs. (30) and (31) are solved self-consistently by iteration [2, 3] to obtain the dimensionless Helmholtz free energy f_m (and the density of states $n(E)$). We remark that

in the numeraical work, it is often more stable to use the following equations instead of Eqs. (30) and (31):

$$P_{\rm B}(E;T) = n(E)e^{-\beta E} = \frac{\displaystyle\sum_{m=1}^{M} g_m^{-1}\ N_m(E)}{\displaystyle\sum_{m=1}^{M} g_m^{-1}\ n_m\ e^{f_m-(\beta_m-\beta)E}}\ , \tag{32}$$

where

$$e^{-f_m} = \sum_E\ P_{\rm B}(E;T_m)\ . \tag{33}$$

The equations are solved iteratively as follows. We can set all the f_m ($m = 1, \cdots, M$) to, e.g., zero initially. We then use Eq. (32) to obtain $P_{\rm B}(E;T_m)$ ($m = 1, \cdots, M$), which are substituted into Eq. (33) to obtain next values of f_m, and so on.

2.2 Replica-Exchange Method

The *replica-exchange method* (REM) [72]–[75] was developed as an extension of simulated tempering [72] (thus it is also referred to as *parallel tempering* [49]) (see, e.g., Ref. [82] for a detailed description of the algorithm). The system for REM consists of M *non-interacting* copies (or, replicas) of the original system in the canonical ensemble at M different temperatures T_m ($m = 1, \cdots, M$). We arrange the replicas so that there is always exactly one replica at each temperature. Then there is a one-to-one correspondence between replicas and temperatures; the label i ($i = 1, \cdots, M$) for replicas is a permutation of the label m ($m = 1, \cdots, M$) for temperatures, and vice versa:

$$\left\{ \begin{array}{rcl} i &=& i(m) \ \equiv\ f(m)\ , \\ m &=& m(i) \ \equiv\ f^{-1}(i)\ , \end{array} \right. \tag{34}$$

where $f(m)$ is a permutation function of m and $f^{-1}(i)$ is its inverse.

Let $X = \left\{x_1^{[i(1)]}, \cdots, x_M^{[i(M)]}\right\} = \left\{x_{m(1)}^{[1]}, \cdots, x_{m(M)}^{[M]}\right\}$ stand for a "state" in this generalized ensemble. The state X is specified by the M sets of coordinates $q^{[i]}$ and momenta $p^{[i]}$ of N atoms in replica i at temperature T_m:

$$x_m^{[i]} \equiv \left(q^{[i]}, p^{[i]}\right)_m\ . \tag{35}$$

Because the replicas are non-interacting, the weight factor for the state X in this generalized ensemble is given by the product of Boltzmann factors for each replica (or at each temperature):

$$W_{\rm REM}(X) = \exp\left\{-\sum_{i=1}^{M} \beta_{m(i)} H\left(q^{[i]}, p^{[i]}\right)\right\} = \exp\left\{-\sum_{m=1}^{M} \beta_m H\left(q^{[i(m)]}, p^{[i(m)]}\right)\right\}\ , \tag{36}$$

where $i(m)$ and $m(i)$ are the permutation functions in Eq. (34).

We now consider exchanging a pair of replicas in the generalized ensemble. Suppose we exchange replicas i and j which are at temperatures T_m and T_n, respectively:

$$X = \left\{\cdots, x_m^{[i]}, \cdots, x_n^{[j]}, \cdots\right\} \longrightarrow\ X' = \left\{\cdots, x_m^{[j]\prime}, \cdots, x_n^{[i]\prime}, \cdots\right\}\ . \tag{37}$$

Here, i, j, m, and n are related by the permutation functions in Eq. (34), and the exchange of replicas introduces a new permutation function f':

$$\left\{ \begin{array}{ll} i & = f(m) \longrightarrow j = f'(m) \ , \\ j & = f(n) \longrightarrow i = f'(n) \ . \end{array} \right. \tag{38}$$

The exchange of replicas can be written in more detail as

$$\left\{ \begin{array}{lcl} x_m^{[i]} \equiv \left(q^{[i]}, p^{[i]} \right)_m & \longrightarrow & x_m^{[j]\prime} \equiv \left(q^{[j]}, p^{[j]\prime} \right)_m \ , \\ x_n^{[j]} \equiv \left(q^{[j]}, p^{[j]} \right)_n & \longrightarrow & x_n^{[i]\prime} \equiv \left(q^{[i]}, p^{[i]\prime} \right)_n \ , \end{array} \right. \tag{39}$$

where the definitions for $p^{[i]\prime}$ and $p^{[j]\prime}$ will be given below. We remark that this process is equivalent to exchanging a pair of temperatures T_m and T_n for the corresponding replicas i and j as follows:

$$\left\{ \begin{array}{lcl} x_m^{[i]} \equiv \left(q^{[i]}, p^{[i]} \right)_m & \longrightarrow & x_n^{[i]\prime} \equiv \left(q^{[i]}, p^{[i]\prime} \right)_n \ , \\ x_n^{[j]} \equiv \left(q^{[j]}, p^{[j]} \right)_n & \longrightarrow & x_m^{[j]\prime} \equiv \left(q^{[j]}, p^{[j]\prime} \right)_m \ . \end{array} \right. \tag{40}$$

In the original implementation of the *replica-exchange method* (REM) [72]–[75], Monte Carlo algorithm was used, and only the coordinates q (and the potential energy function $E(q)$) had to be taken into account. In molecular dynamics algorithm, on the other hand, we also have to deal with the momenta p. We proposed the following momentum assignment in Eq. (39) (and in Eq. (40)) [82]:

$$\left\{ \begin{array}{ll} p^{[i]\prime} & \equiv \sqrt{\dfrac{T_n}{T_m}} \ p^{[i]} \ , \\ p^{[j]\prime} & \equiv \sqrt{\dfrac{T_m}{T_n}} \ p^{[j]} \ , \end{array} \right. \tag{41}$$

which we believe is the simplest and the most natural. This assignment means that we just rescale uniformly the velocities of all the atoms in the replicas by the square root of the ratio of the two temperatures so that the temperature condition in Eq. (4) may be satisfied.

In order for this exchange process to converge towards an equilibrium distribution, it is sufficient to impose the detailed balance condition on the transition probability $w(X \to X')$:

$$W_{\rm REM}(X) \ w(X \to X') = W_{\rm REM}(X') \ w(X' \to X) \ . \tag{42}$$

From Eqs. (1), (2), (36), (41), and (42), we have

$$\begin{aligned} \frac{w(X \to X')}{w(X' \to X)} & = \exp \Big\{ -\beta_m \left[K\left(p^{[j]\prime} \right) + E\left(q^{[j]} \right) \right] - \beta_n \left[K\left(p^{[i]\prime} \right) + E\left(q^{[i]} \right) \right] \\ & \qquad + \beta_m \left[K\left(p^{[i]} \right) + E\left(q^{[i]} \right) \right] + \beta_n \left[K\left(p^{[j]} \right) + E\left(q^{[j]} \right) \right] \Big\} \ , \\ & = \exp \Big\{ -\beta_m \frac{T_m}{T_n} K\left(p^{[j]} \right) - \beta_n \frac{T_n}{T_m} K\left(p^{[i]} \right) + \beta_m K\left(p^{[i]} \right) + \beta_n K\left(p^{[j]} \right) \\ & \qquad - \beta_m \left[E\left(q^{[j]} \right) - E\left(q^{[i]} \right) \right] - \beta_n \left[E\left(q^{[i]} \right) - E\left(q^{[j]} \right) \right] \Big\} \ , \\ & = \exp\left(-\Delta \right) \ , \end{aligned} \tag{43}$$

where

$$\Delta \equiv (\beta_n - \beta_m) \left(E\left(q^{[i]}\right) - E\left(q^{[j]}\right) \right) \ , \tag{44}$$

and i, j, m, and n are related by the permutation functions (in Eq. (34)) before the exchange:

$$\begin{cases} i &= f(m) \ , \\ j &= f(n) \ . \end{cases} \tag{45}$$

This can be satisfied, for instance, by the usual Metropolis criterion [96]:

$$w(X \to X') \equiv w\left(x_m^{[i]} \ \middle| \ x_n^{[j]} \right) = \begin{cases} 1 \ , & \text{for } \Delta \le 0 \ , \\ \exp\left(-\Delta\right) \ , & \text{for } \Delta > 0 \ , \end{cases} \tag{46}$$

where in the second expression (i.e., $w(x_m^{[i]}|x_n^{[j]})$) we explicitly wrote the pair of replicas (and temperatures) to be exchanged. Note that this is exactly the same criterion that was originally derived for Monte Carlo algorithm [72]–[75].

Without loss of generality we can again assume $T_1 < T_2 < \cdots < T_M$. A simulation of the *replica-exchange method* (REM) [72]–[75] is then realized by alternately performing the following two steps:

1. Each replica in canonical ensemble of the fixed temperature is simulated *simultaneously* and *independently* for a certain MC or MD steps.

2. A pair of replicas at neighboring temperatures, say $x_m^{[i]}$ and $x_{m+1}^{[j]}$, are exchanged with the probability $w\left(x_m^{[i]} \ \middle| \ x_{m+1}^{[j]} \right)$ in Eq. (46).

Note that in Step 2 we exchange only pairs of replicas corresponding to neighboring temperatures, because the acceptance ratio of the exchange process decreases exponentially with the difference of the two β's (see Eqs. (44) and (46)). Note also that whenever a replica exchange is accepted in Step 2, the permutation functions in Eq. (34) are updated.

The REM simulation is particularly suitable for parallel computers. Because one can minimize the amount of information exchanged among nodes, it is best to assign each replica to each node (exchanging pairs of temperature values among nodes is much faster than exchanging coordinates and momenta). This means that we keep track of the permutation function $m(i;t) = f^{-1}(i;t)$ in Eq. (34) as a function of MC or MD step t during the simulation. After parallel canonical MC or MD simulations for a certain steps (Step 1), $M/2$ pairs of replicas corresponding to neighboring temperatures are simulateneously exchanged (Step 2), and the pairing is alternated between the two possible choices, i.e., (T_1, T_2), (T_3, T_4), $\cdots$ and (T_2, T_3), (T_4, T_5), $\cdots$.

The major advantage of REM over other generalized-ensemble methods such as multicanonical algorithm [4, 5] and simulated tempering [47, 48] lies in the fact that the weight factor is *a priori* known (see Eq. (36)), while in the latter algorithms the determination of the weight factors can be very tedius and time-consuming. A random walk in "temperature space" is realized for each replica, which in turn induces a random walk in potential energy space. This alleviates the problem of getting trapped in states of energy local minima. In REM, however, the number of required replicas increases as the system size N increases (according to $\sqrt{N}$) [72]. This demands a lot of computer power for complex systems.

2.3 Replica-Exchange Multicanonical Algorithm and Replica-Exchange Simulated Tempering

The *replica-exchange multicanonical algorithm* (REMUCA) [94] overcomes both the difficulties of MUCA (the multicanonical weight factor determination is non-trivial) and REM (a lot of replicas, or computation time, is required). In REMUCA we first perform a short REM simulation (with M replicas) to determine the multicanonical weight factor and then perform with this weight factor a regular multicanonical simulation with high statistics. The first step is accomplished by the multiple-histogram reweighting techniques [2, 3]. Let $N_m(E)$ and n_m be respectively the potential-energy histogram and the total number of samples obtained at temperature $T_m = 1/k_{\rm B}\beta_m$ of the REM run. The density of states $n(E)$ is then given by solving Eqs. (30) and (31) self-consistently by iteration [2, 3].

Once the estimate of the density of states is obtained, the multicanonical weight factor can be directly determined from Eq. (8) (see also Eq. (9)). Actually, the multicanonical potential energy, $E_{\rm mu}(E;T_0)$, thus determined is only reliable in the following range:

$$E_1 \le E \le E_M \ , \tag{47}$$

where

$$\left\{ \begin{array}{rcl} E_1 & = & < E >_{T_1} \ , \\ E_M & = & < E >_{T_M} \ , \end{array} \right. \tag{48}$$

and T_1 and T_M are respectively the lowest and the highest temperatures used in the REM run. Outside this range we extrapolate the multicanonical potential energy linearly:

$$\mathcal{E}_{\rm mu}^{\{0\}}(E) \equiv \left\{ \begin{array}{ll} \left. \dfrac{\partial E_{\rm mu}(E;T_0)}{\partial E} \right|_{E=E_1} (E - E_1) + E_{\rm mu}(E_1;T_0) \ , & \text{for } E < E_1, \\ E_{\rm mu}(E;T_0) \ , & \text{for } E_1 \le E \le E_M, \\ \left. \dfrac{\partial E_{\rm mu}(E;T_0)}{\partial E} \right|_{E=E_M} (E - E_M) + E_{\rm mu}(E_M;T_0) \ , & \text{for } E > E_M. \end{array} \right. \tag{49}$$

The multicanonical MC and MD runs are then performed with the Metropolis criterion of Eq. (10) and with the Newton equation in Eq. (12), respectively, in which $\mathcal{E}_{\rm mu}^{\{0\}}(E)$ in Eq. (49) is substituted into $E_{\rm mu}(E;T_0)$. We expect to obtain a flat potential energy distribution in the range of Eq. (47). Finally, the results are analyzed by the single-histogram reweighting techniques as described in Eq. (21) (and Eq. (20)).

Some remarks are now in order. From Eqs. (9), (14), (15), and (48), Eq. (49) becomes

$$\mathcal{E}_{\rm mu}^{\{0\}}(E) = \left\{ \begin{array}{ll} \dfrac{T_0}{T_1}(E - E_1) + T_0 S(E_1) = \dfrac{T_0}{T_1} E + \text{constant} \ , & \text{for } E < E_1 \equiv < E >_{T_1}, \\ T_0 S(E) \ , & \text{for } E_1 \le E \le E_M, \\ \dfrac{T_0}{T_M}(E - E_M) + T_0 S(E_M) = \dfrac{T_0}{T_M} E + \text{constant} \ , & \text{for } E > E_M \equiv < E >_{T_M}. \end{array} \right. \tag{50}$$

The Newton equation in Eq. (12) is then written as (see Eqs. (13), (14), and (15))

$$\dot{\boldsymbol{p}}_k = \left\{ \begin{array}{ll} \dfrac{T_0}{T_1} \ \boldsymbol{f}_k \ , & \text{for } E < E_1, \\ \dfrac{T_0}{T(E)} \ \boldsymbol{f}_k \ , & \text{for } E_1 \le E \le E_M, \\ \dfrac{T_0}{T_M} \ \boldsymbol{f}_k \ , & \text{for } E > E_M. \end{array} \right. \tag{51}$$

Because only the product of inverse temperature β and potential energy E enters in the Boltzmann factor (see Eq. (5)), a rescaling of the potential energy (or force) by a constant, say α, can be considered as the rescaling of the temperature by α^{-1} [27, 91]. Hence, our choice of $\mathcal{E}_{\rm mu}^{\{0\}}(E)$ in Eq. (49) results in a canonical simulation at $T = T_1$ for $E < E_1$, a multicanonical simulation for $E_1 \le E \le E_M$, and a canonical simulation at $T = T_M$ for $E > E_M$. Note also that the above arguments are independent of the value of T_0, and we will get the same results, regardless of its value.

Finally, although we did not find any difficulty in the case of protein systems that we studied, a single REM run in general may not be able to give an accurate estimate of the density of states (like in the case of a strong first-order phase transition [72]). In such a case we can still greatly simplify the process of the multicanonical weight factor determination by combining the present method with the previous iterative methods [11, 22, 26, 55, 56, 57, 9].

We finally present the new method which we refer to as the *replica-exchange simulated tempering* (REST) [95]. In this method, just as in REMUCA, we first perform a short REM simulation (with M replicas) to determine the simulated tempering weight factor and then perform with this weight factor a regular ST simulation with high statistics. The first step is accomplished by the multiple-histogram reweighting techniques [2, 3], which give the dimensionless Helmholtz free energy f_m (see Eqs. (30) and (31)).

Once the estimate of the dimensionless Helmholtz free energy f_m are obtained, the simulated tempering weight factor can be directly determined by using Eq. (25) where we set $a_m = f_m$ (compare Eq. (26) with Eq. (31)). A long simulated tempering run is then performed with this weight factor. Let $N_m(E)$ and n_m be respectively the potential-energy histogram and the total number of samples obtained at temperature $T_m = 1/k_{\rm B}\beta_m$ from this simulated tempering run. The multiple-histogram reweighting techniques of Eqs. (30) and (31) can be used again to obtain the best estimate of the density of states $n(E)$. The expectation value of a physical quantity A at any temperature T $(= 1/k_{\rm B}\beta)$ is then calculated from Eq. (20).

The formulations of REMUCA and REST are simple and straightforward, but the numerical improvement is great, because the weight factor determination for MUCA and ST becomes very difficult by the usual iterative processes for complex systems.

2.4 Multicanonical Replica-Exchange Method

In the previous subsection we presented a new generalized-ensemble algorithm, REMUCA, that combines the merits of replica-exchange method and multicanonical algorithm. In REMUCA a short REM simulation with M replicas are first performed and the results are used to determine the multicanonical weight factor, and then a regular multicanonical production run with this weight is performed. The number of replicas, M, that is required in the first step should be set minimally as long as a random walk between the lowest-energy region and the high-energy region is realized. This number can still be very large for complex systems. This is why the (multicanonical) production run in REMUCA is performed with a "single replica." While multicanonical simulatoins are usually based on local updates, a replica-exchange process can be considered to be a global update, and global updates enhance the sampling further. Here, we present a further modification of REMUCA and refer to the new method as *multicanonical replica-exchange method* (MUCAREM) [94]. In MUCAREM the final production run is not a regular multicanonical

simulation but a replica-exchange simulation with a few replicas, say $\mathcal{M}$ replicas, in the multicanonical ensemble. (We remark that replica-exchange simulations based on the generalized ensemble with Tsallis weights were introduced in Ref. [80].) Because multicanonical simulations cover much wider energy ranges than regular canonical simulations, the number of required replicas for the production run of MUCAREM is much less than that for the regular REM ($\mathcal{M} \ll M$), and we can keep the merits of REMUCA (and improve the sampling further).

The details of MUCAREM are as follows. As in REMUCA, we first perform a short REM simulation with M replicas with M different temperatures (we order them as $T_1 < T_2 < \cdots < T_M$) and obtain the best estimate of the density of states $n(E)$ in the whole energy range of interest (see Eq. (47)) by the multiple-histogram reweighting techniques of Eqs. (30) and (31). We then choose a number $\mathcal{M}$ ($\mathcal{M} \ll M$) and assign $\mathcal{M}$ pairs of temperatures $(T_{\rm L}^{\{m\}}, T_{\rm H}^{\{m\}})$ $(m = 1, \cdots, \mathcal{M})$. Here, we assume that $T_{\rm L}^{\{m\}} < T_{\rm H}^{\{m\}}$ and arrange the temperatures so that the neighboring regions covered by the pairs have sufficient overlaps. In particular, we set $T_{\rm L}^{\{1\}} = T_1$ and $T_{\rm H}^{\{\mathcal{M}\}} = T_M$. We then define the following quantities:

$$\left\{ \begin{array}{lcl} E_{\rm L}^{\{m\}} & = & < E >_{T_{\rm L}^{\{m\}}} \ , \\ E_{\rm H}^{\{m\}} & = & < E >_{T_{\rm H}^{\{m\}}} \ , \ (m = 1, \cdots, \mathcal{M}) \ . \end{array} \right. \tag{52}$$

We also choose $\mathcal{M}$ (arbitrary) temperatures T_m $(m = 1, \cdots, \mathcal{M})$ and assign the following multicanonical potential energies:

$$\mathcal{E}_{\rm mu}^{\{m\}}(E) = \left\{ \begin{array}{ll} \left. \dfrac{\partial E_{\rm mu}(E;T_m)}{\partial E} \right|_{E=E_{\rm L}^{\{m\}}} (E - E_{\rm L}^{\{m\}}) + E_{\rm mu}(E_{\rm L}^{\{m\}};T_m) \ , & \text{for } E < E_{\rm L}^{\{m\}}, \\ E_{\rm mu}(E;T_m) \ , & \text{for } E_{\rm L}^{\{m\}} \le E \le E_{\rm H}^{\{m\}}, \\ \left. \dfrac{\partial E_{\rm mu}(E;T_m)}{\partial E} \right|_{E=E_{\rm H}^{\{m\}}} (E - E_{\rm H}^{\{m\}}) + E_{\rm mu}(E_{\rm H}^{\{m\}};T_m) \ , & \text{for } E > E_{\rm H}^{\{m\}}, \end{array} \right. \tag{53}$$

where $E_{\rm mu}(E;T)$ is the multicanonical potential energy that was determined for the whole energy range of Eq. (47). As remarked around Eq. (50), our choice of $\mathcal{E}_{\rm mu}^{\{m\}}(E)$ in Eq. (53) results in a canonical simulation at $T = T_{\rm L}^{\{m\}}$ for $E < E_{\rm L}^{\{m\}}$, a multicanonical simulation for $E_{\rm L}^{\{m\}} \le E \le E_{\rm H}^{\{m\}}$, and a canonical simulation at $T = T_{\rm H}^{\{m\}}$ for $E > E_{\rm H}^{\{m\}}$.

The production run of MUCAREM is a replica-exchange simulation with $\mathcal{M}$ replicas with $\mathcal{M}$ different temperatures T_m and multicanonical potential energies $\mathcal{E}_{\rm mu}^{\{m\}}(E)$. By following the same derivation that led to the original REM, we have the following transition probability of replica exchange of neighboring temperatures (see Eqs. (44) and (46)):

$$w\left(x_m^{[i]} \ \middle| \ x_{m+1}^{[j]}\right) = \left\{ \begin{array}{ll} 1 \ , & \text{for } \Delta \le 0 \ , \\ \exp\left(-\Delta\right) \ , & \text{for } \Delta > 0 \ , \end{array} \right. \tag{54}$$

where

$$\Delta = \beta_{m+1} \left\{ \mathcal{E}_{\rm mu}^{\{m+1\}}\left(E\left(q^{[i]}\right)\right) - \mathcal{E}_{\rm mu}^{\{m+1\}}\left(E\left(q^{[j]}\right)\right) \right\} - \beta_m \left\{ \mathcal{E}_{\rm mu}^{\{m\}}\left(E\left(q^{[i]}\right)\right) - \mathcal{E}_{\rm mu}^{\{m\}}\left(E\left(q^{[j]}\right)\right) \right\} \ . \tag{55}$$

Note that we need to newly evaluate the multicanonical potential energy, $\mathcal{E}_{\rm mu}^{\{m\}}(E(q^{[j]}))$ and $\mathcal{E}_{\rm mu}^{\{m+1\}}(E(q^{[i]}))$, because $\mathcal{E}_{\rm mu}^{\{m\}}(E)$ and $\mathcal{E}_{\rm mu}^{\{n\}}(E)$ are, in general, different functions for

$m \neq n$. We remark that the same additional evaluation of the potential energy is necessary for the multidimensional replica-exchange method [84].

For obtaining the canonical distributions, the multiple-histogram reweighting techniques [2, 3] are again used. Let $N_m(E)$ and n_m be respectively the potential-energy histogram and the total number of samples obtained at T_m with the multicanonical potential energy $\mathcal{E}_{\rm mu}^{\{m\}}(E)$ $(m = 1, \cdots, \mathcal{M})$. The expectation value of a physical quantity A at any temperature $T = 1/k_{\rm B}\beta$ is then obtained from Eq. (20), where the best estimate of the density of states is given by solving the multiple-histogram reweighting equations, which now read

$$n(E) = \frac{\sum_{m=1}^{\mathcal{M}} g_m^{-1} \; N_m(E)}{\sum_{m=1}^{\mathcal{M}} g_m^{-1} \; n_m \; e^{f_m - \beta_m \mathcal{E}_{\rm mu}^{\{m\}}(E)}} \; , \tag{56}$$

and

$$e^{-f_m} = \sum_E \; n(E) \; e^{-\beta_m \mathcal{E}_{\rm mu}^{\{m\}}(E)} \; . \tag{57}$$

3 EXAMPLES OF SIMULATION RESULTS

We now present some examples of the simulation results by the algorithms described in the previous section. A few short peptide systems were considered.

For Monte Carlo simulations, the potential energy parameters were taken from ECEPP/2 [97]–[99]. The generalized-ensemble algorithms were implemented in the computer code KONF90 [100, 101] for the actual simulations. Besides gas phase simulations, various solvation models have been incorporated. The simplest one is the sigmoidal, distance-dependent dielectric function [102, 103]. The explicit form of the function we used is given in Ref. [104], which is a slight modification of the one in Ref. [105]. A second (and more accurate) model that represents solvent contributions is the term proportional to the solvent-accessible surface area of solute molecule. The parameters we used are those of Ref. [106]. For the calculation of solvent-accessible surface area, we used the computer code NSOL [107], which is based on the code NSC [108]. The third (and most rigorous) method that represents solvent effects is based on the reference interaction site model (RISM) [109]–[111]. The model of water molecule that we adopted is the SPC/E model [112]. A robust and fast algorithm for solving RISM equations was recently developed [113, 114], which we employed in our calculations [115, 116, 45].

For molecular dynamics simulations, the force-field parameters were taken from the all-atom versions of AMBER [117]–[119]. The computer code developed in Refs. [120, 121], which is based on PRESTO [122], was used. The unit time step was set to 0.5 fs. The temperature during the canonical MD simulations was controlled by the constraint method [123, 124]. Besides gas phase simulations, we have also performed MD simulations with explicit water molecules of TIP3P model [125].

As described in detail in the previous section, in generalized-ensemble simulations and subsequent analyses of the data, potential energy distributions have to be taken as histograms. For the bin size of these histograms, we used the values ranging from 0.5 to 2 kcal/mol, depending on the system studied.

We first illustrate how effectively generalized-ensemble simulations can sample the configurational space compared to the conventional simulations in the canonical ensemble. It is known by experiments that the system of a 17-residue peptide fragment from ribonuclease T1 tends to form α-helical conformations [126]. We have performed both a canonical MC simulation of this peptide at a low temperature ($T = 200$ K) and a multicanonical MC simulation [127]. In Figure 1 we show the time series of potential energy from these simulations.

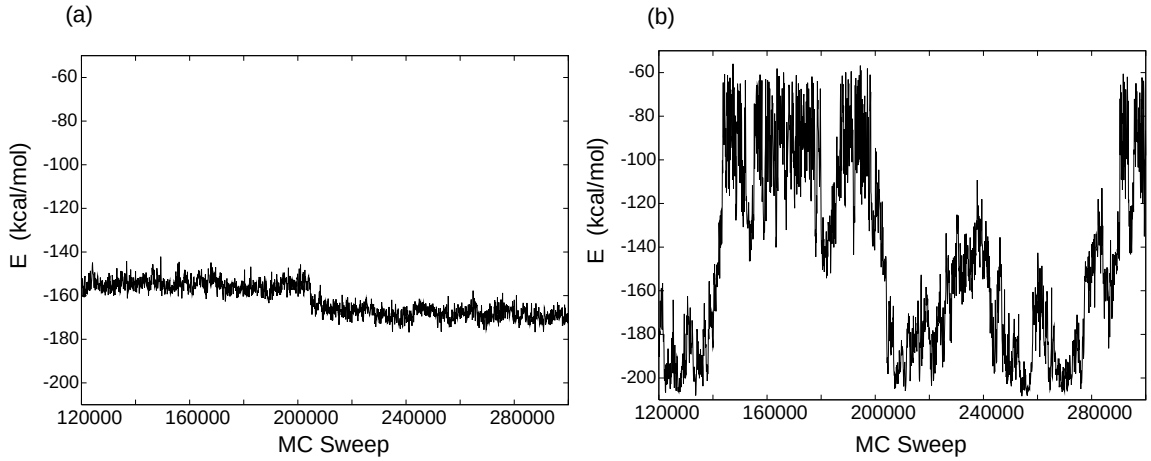

Figure 1: Time series (from 120,000 MC sweeps to 300,000 MC sweeps) of potential energy of the peptide fragment of ribonuclease T1 from (a) a conventional canonical MC simulation at $T = 200$ K and (b) a multicanonical MC simulation.

We see that the canonical simulation thermalizes very slowly. On the other hand, the MUCA simulation indeed performed a random walk in potential energy space covering a very wide energy range. Four conformations chosen during this period (from 120,000 MC sweeps to 300,000 MC sweeps) are shown in Figure 2 for the canonical simulation and in Figure 3 for the MUCA simulation. We see that the MUCA simulation samples much wider conformational space than the conventional canonical simulation.

The next examples of the systems that we studied by multicanonical MC simulations are homo-oligomer systems. We studied the helix-forming tendencies of three amino-acid homo-oligomers of length 10 in gas phase [21, 22] and in aqueous solution (the solvent effects are represented by the term that is proportional to solvent-accessible surface area) [43]. Three characteristic amino acids, alanine (helix former), valine (helix indifferent), and glycine (helix breaker) were considered. In Figure 4 the lowest-energy conformations obtained both in gas phase and in aqueous solution by MUCA simulations are shown [43]. The lowest-energy conformations of $(\mathrm{Ala})_{10}$ (Figures 4(a) and 4(b)) have six intrachain backbone hydrogen bonds that characterize the α-helix and are indeed completely helical. Those of $(\mathrm{Val})_{10}$ (Figures 4(c) and 4(d)) are also in almost ideal helix state (from residue 2 to residue 9 in gas phase and from residue 2 to residue 8 in aqueous solution). On the other hand, those of $(\mathrm{Gly})_{10}$ (Figures 4(e) and 4(f)) are not helical and rather round.

We calculated the average values of the total potential energy and its component terms of $(\mathrm{Ala})_{10}$ as a function of temperature both in gas phase and in aqueous solution [43]. The results are shown in Figure 5. For homo-alanine in gas phase, all the conformational energy

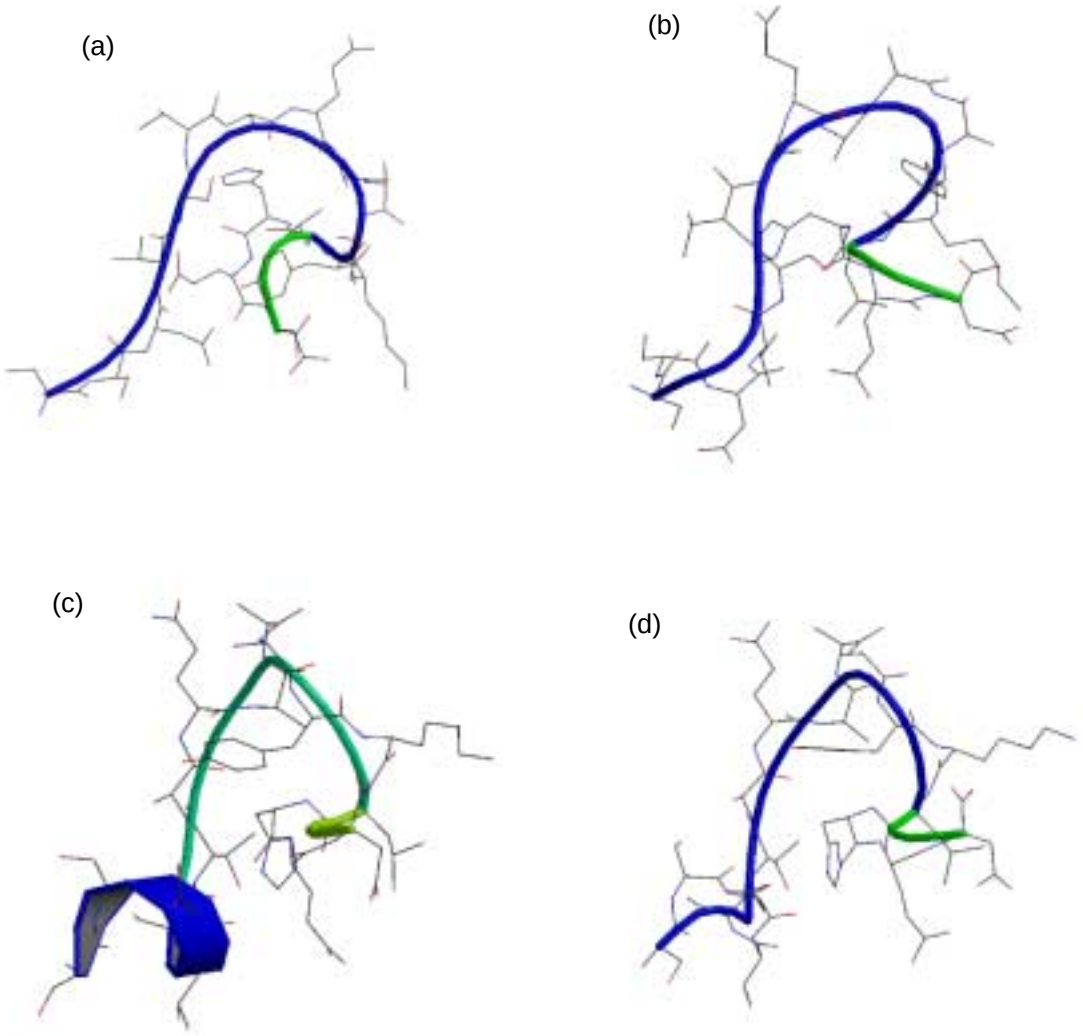

Figure 2: Typical snapshots from the canonical MC simulation of Figure 1(a). The figures were created with Molscript [128] and Raster3D [129, 130].

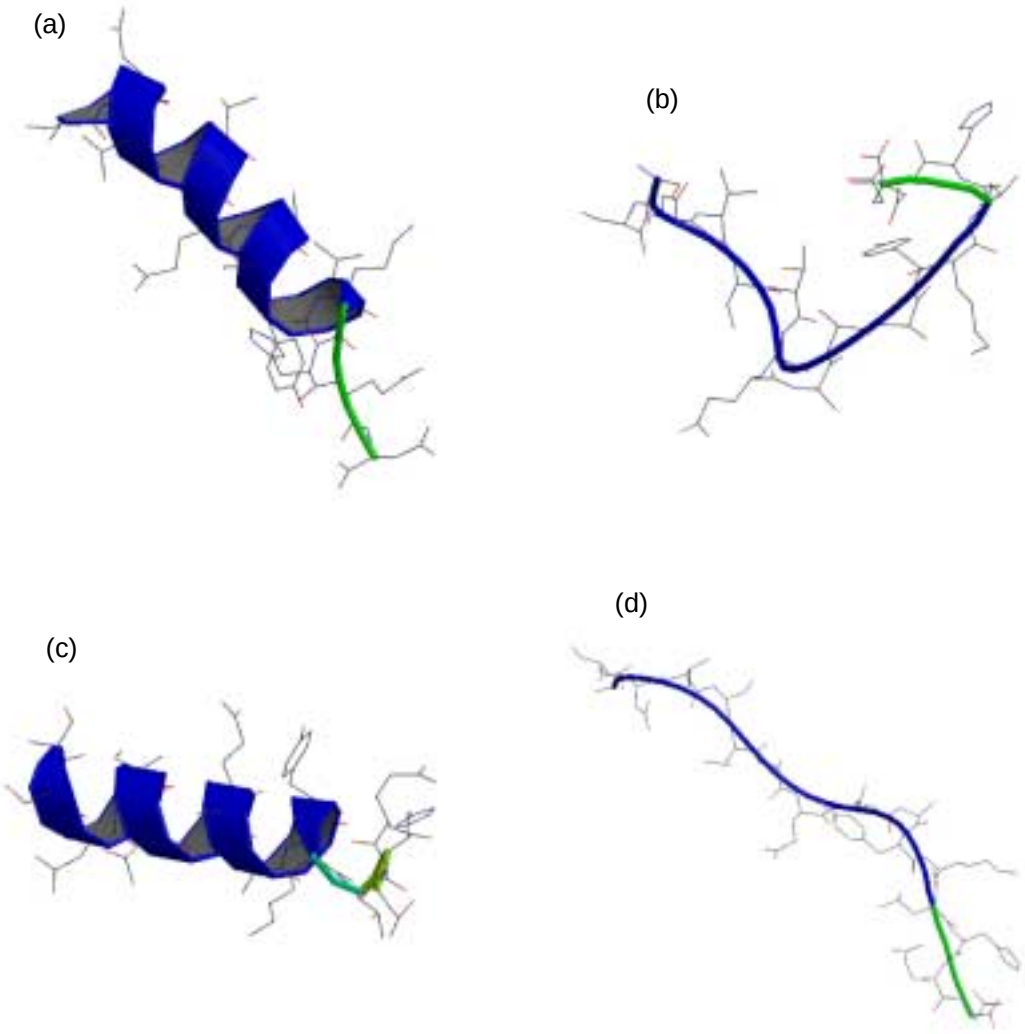

Figure 3: Typical snapshots from the multicanonical MC simulation of Figure 1(b). The figures were created with Molscript [128] and Raster3D [129, 130].

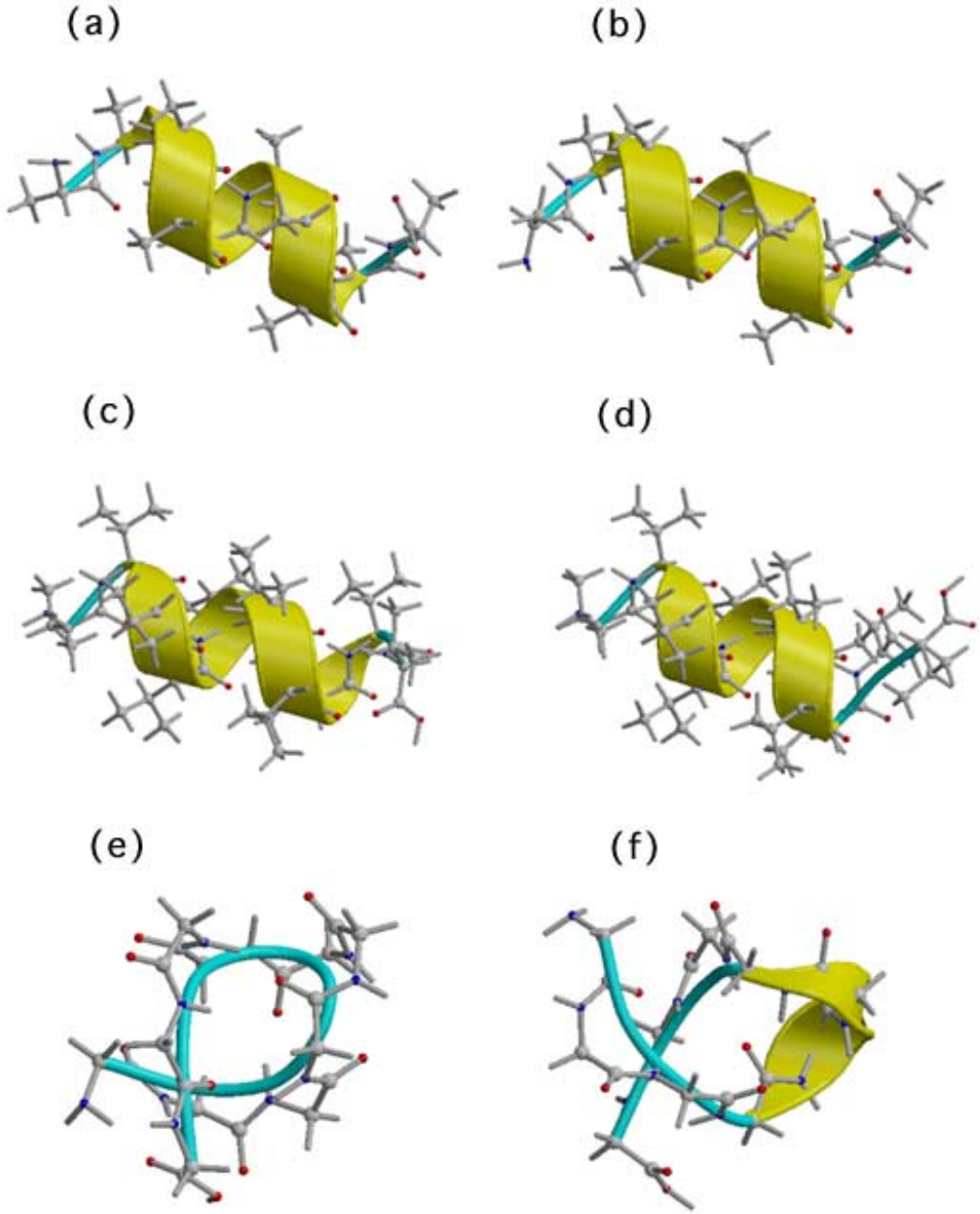

Figure 4: The lowest-energy conformations of $(Ala)_{10}$ ((a) and (b)), $(Val)_{10}$ ((c) and (d)), and $(Gly)_{10}$ ((e) and (f)) obtained from the multicanonical MC simulations in gas phase and in aqueous solution, respectively. The figures were created with Molscript [128] and Raster3D [129, 130].

terms increase monotonically as temperature increases. The changes of each component terms are very small except for the Lennard-Jones term $E_{\rm v}$, indicating that $E_{\rm v}$ plays an important role in the folding of homo-alanine [22].

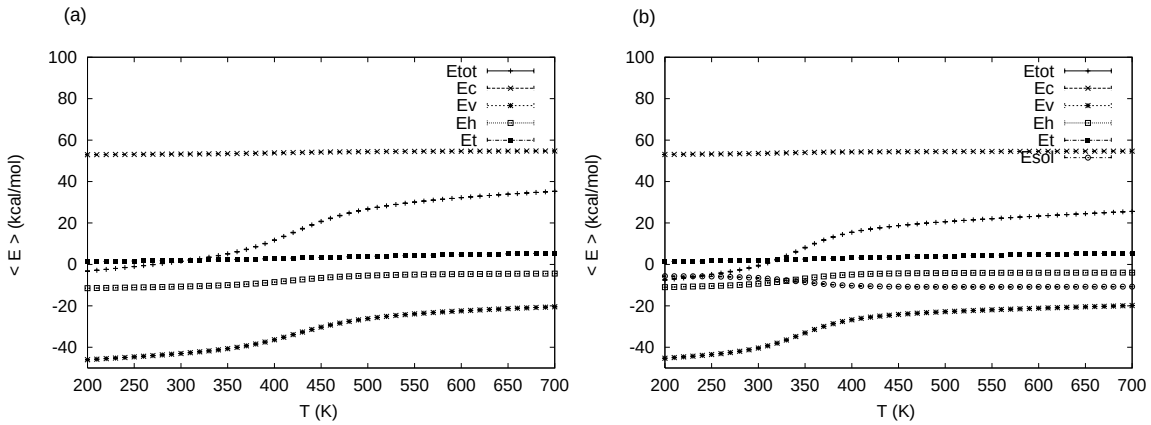

Figure 5: Average of the total potential energy $E_{\rm tot}$ and averages of its component terms, electrostatic energy $E_{\rm c}$, hydrogen-bond energy $E_{\rm h}$, Lennard-Jones energy $E_{\rm v}$, torsion energy $E_{\rm t}$, and solvation free energy $E_{\rm sol}$ (only for the case in aqueous solution) for homo-alanine as a function of temperature T (a) in gas phase and (b) in aqueous solution. The values for each case were calculated from one multicanonical production run of 1,000,000 MC sweeps by the single-histogram reweighting techniques.

In aqueous solution the overall behaviors of the conformational energy terms are very similar to those in gas phase. The solvation term, on the other hand, decreases monotonically as temperature increases. These results imply that the solvation term favors random-coil conformations, while the conformational terms favor helical conformations.

The rapid changes (decrease for the solvation term and increase for the rest of the terms) of all the average values occur at the same temperature (around at 420 K in gas phase and 340 K in solvent). We thus calculated the specific heat for $(\mathrm{Ala})_{10}$ as a function of temperature. The specific heat here is defined by the following equation:

$$C(T) = \beta^2 \, \frac{< E_{\rm tot}^2 >_T - < E_{\rm tot} >_T{}^2}{N}, \tag{58}$$

where N (= 10) is the number of residues in the oligomer. In Figure 6 we show the results. We observe sharp peaks in the specific heat for both environment. The temperatures at the peak, helix-coil transition temperatures, are $T_c \approx 420$ K and 340 K in gas phase and in aqueous solution, respectively.

We calculated the average number of helical residues $< n >_T$ in a conformation as a function of temperature. In Figure 7 we show this quantity as a function of temperature for the three homo-oligomers in aqueous solution. The average helicity tends to decrease monotonically as the temperature increases because of the increased thermal fluctuations.

At $T = 200$ K, $< n >_T$ for homo-alanine is 8. If we neglect the terminal residues, in which α-helix tends to be frayed, $n = 8$ corresponds to the maximal helicity, and the conformation can be considered completely helical. The homo-alanine is thus in an ideal helical structure at $T = 200$ K. Around the room temperature, the homo-alanine

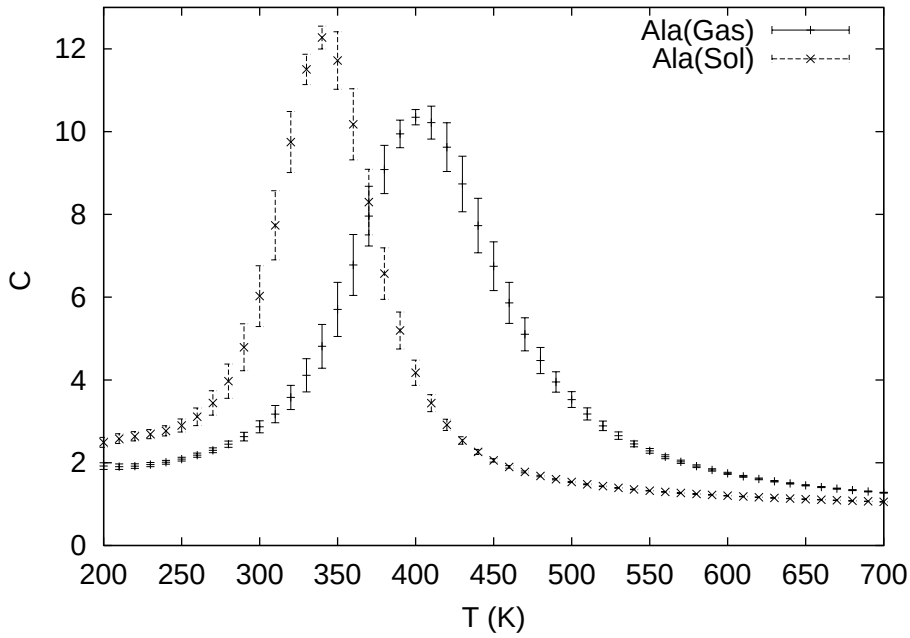

Figure 6: Specific heat C as a function of temperature T for $(\mathrm{Ala})_{10}$ in gas phase and in aqueous solution. The values for each case were calculated from one multicanonical production run of 1,000,000 MC sweeps by the single-histogram reweighting techniques.

is still substantially helical ($\approx$ 70 % helicity). This is consistent with the experimental fact that alanine is a strong helix former. We observe that $< n >_T$ is 5 (50 % helicity) at the transition temperature obtained from the peak in specific heat (around 340 K). This implies that the peak in specific heat indeed implies a helix-coil transition between an ideal helix and a random coil.

The next example is a penta peptide, Met-enkephalin, whose amino-acid sequence is: Tyr-Gly-Gly-Phe-Met. Since this is one of the simplest peptides with biological functions, it served as a bench mark system for many simulations.

Here, we present the latest results of a multicanonical MC simulation of Met-enkephalin in gas phase [39]. The conformations were classified into six groups of similar structures according to their intrachain hydrogen bonds. In Figure 8 we show the lowest-energy conformations in each group identified by the MUCA simulation. The lowest-energy conformation of group C25 (Figure 8(a)) has two hydrogen bonds, connecting residues 2 and 5, and forms a type II′ β-turn. The ECEPP/2 energy of the conformation is -12.2 kcal/mol, and this conformation corresponds to the global-minimum-energy state of Met-enkephalin in gas phase. The conformation is essentially identical with those found by others [132, 133]. The lowest-energy conformation of group C14 (Figure 8(b)) has two hydrogen bonds, connecting residues 1 and 4, and forms a type II β-turn. The energy is -11.1 kcal/mol, and this conformation corresponds to the second-lowest-energy state. Other groups correspond to high-energy states.

We now study the distributions of conformations in these groups as a function of temperature. The results are shown in Figure 9. As can be seen in the Figure, group C25 is dominant at low temperatures. Conformations of group C14 start to appear from $T \approx 100$ K. At $T \approx 300$ K, the distributions of these two groups, C25 and C14, balance

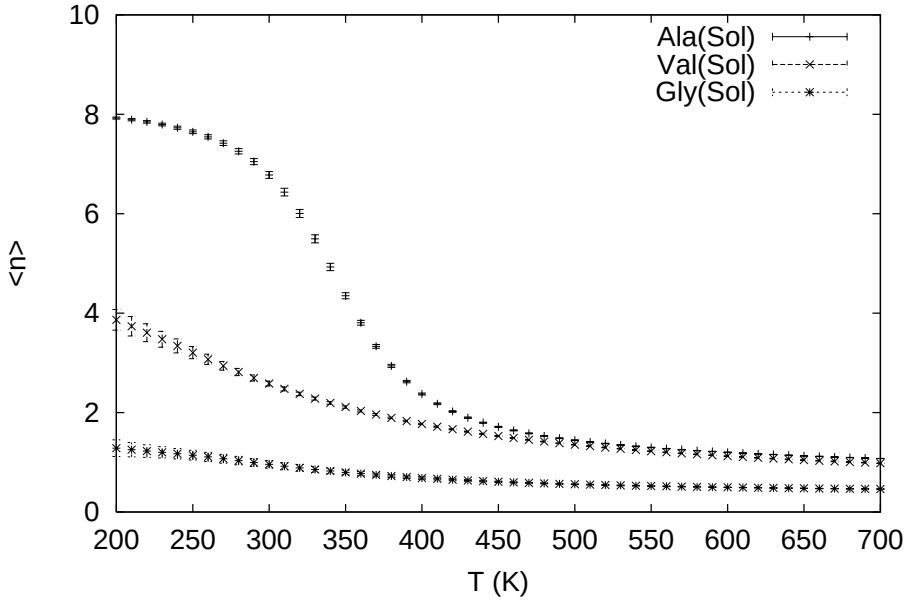

Figure 7: Average helicity $< n >_T$ as a function of temperature T for $(\mathrm{Ala})_{10}$, $(\mathrm{Val})_{10}$, and $(\mathrm{Gly})_{10}$ in aqueous solution. The values for each case were calculated from one multicanonical production run of 1,000,000 MC sweeps by the single-histogram reweighting techniques.

(≈ 25 % each) and constitute the main groups. Above $T \approx 300$ K, the contributions of other groups become non-negligible (those of group C24 and group C13 are about 10 % and 8 %, respectively, at $T = 400$ K). Note that the distribution of conformations that do not belong to any of the six groups monotonically increases as the temperature is raised. This is because random-coil conformations without any intrachain hydrogen bonds are favored at high temperatures.

The same peptide in gas phase was studied by the replica-exchange MD simulation [82]. We made an MD simulation of 2×10^6 time steps (or, 1.0 ns) for each replica, starting from an extended conformation. We used the following eight temperatures: 700, 585, 489, 409, 342, 286, 239, and 200 K, which are distributed exponentially, following the annealing schedule of simulated annealing simulations [101]. As is shown below, this choice already gave an optimal temperature distribution. The replica exchange was tried every 10 fs, and the data were stored just before the replica exchange for later analyses.

As for expectation values of physical quantities at various temperatures, we used the multiple-histogram reweighting techniques of Eqs. (30) and (31). We remark that for biomolecular systems the integrated autocorrelation times τ_m in the reweighting formulae (see Eq. (30)) can safely be set to be a constant [3], and we do so throughout the analyses in this section.

For an optimal performance of REM simulations the acceptance ratios of replica exchange should be sufficiently uniform and large (say, > 10 %). In Table 1 we list these quantities. The values are indeed uniform (all about 15 % of acceptance probability) and large enough (more than 10 %).

The results in Table 1 imply that one should observe a free random walk in temperature space. The results for one of the replicas are shown in Figure 10(a). We do observe a

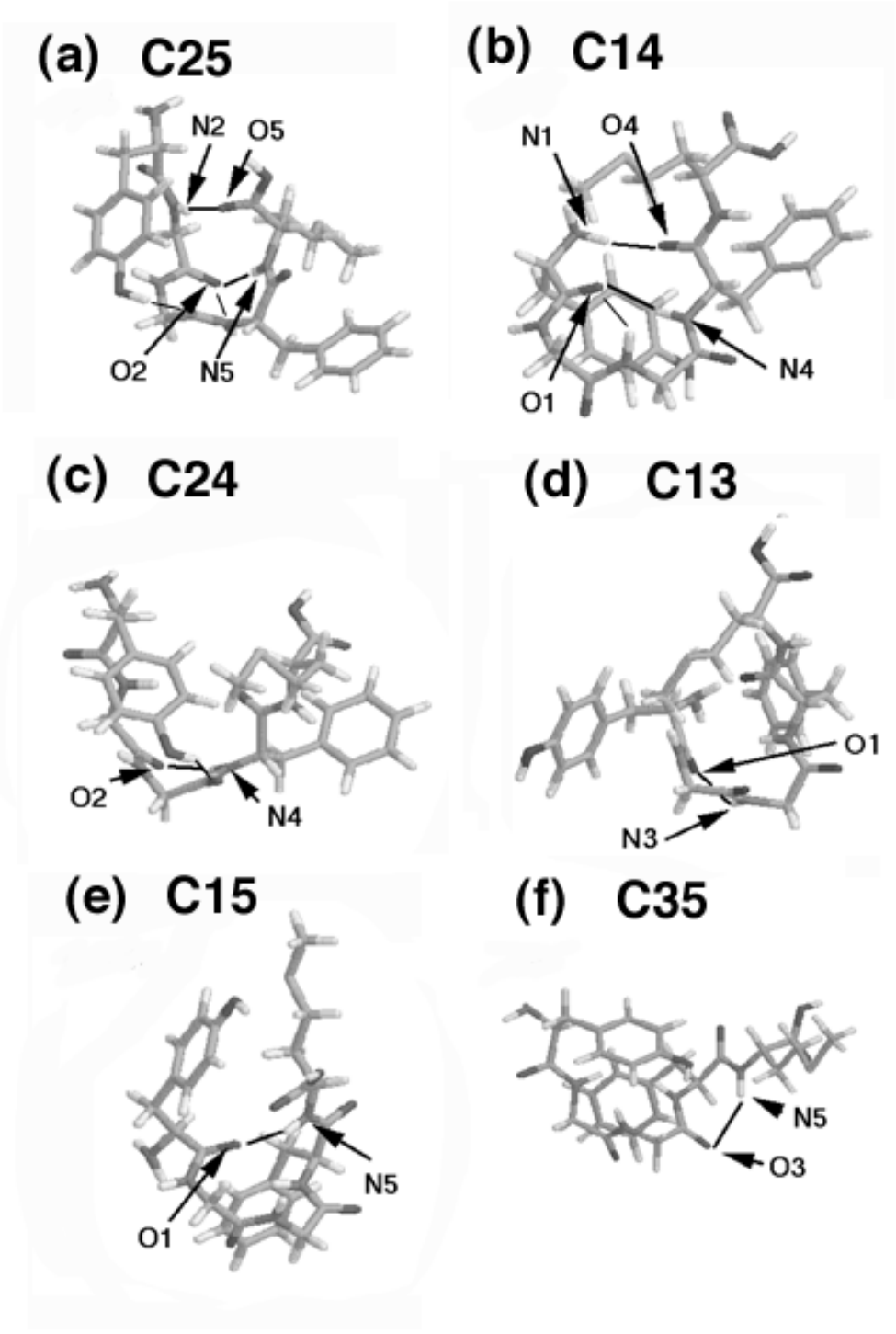

Figure 8: The lowest-energy conformations of Met-enkephalin in each group obtained by the multicanonical MC simulation of 1,000,000 MC sweeps. These conformations correspond to groups (a) C25, (b) C14, (c) C24, (d) C13, (e) C15, and (f) C35. The figures were created with RasMol [131].

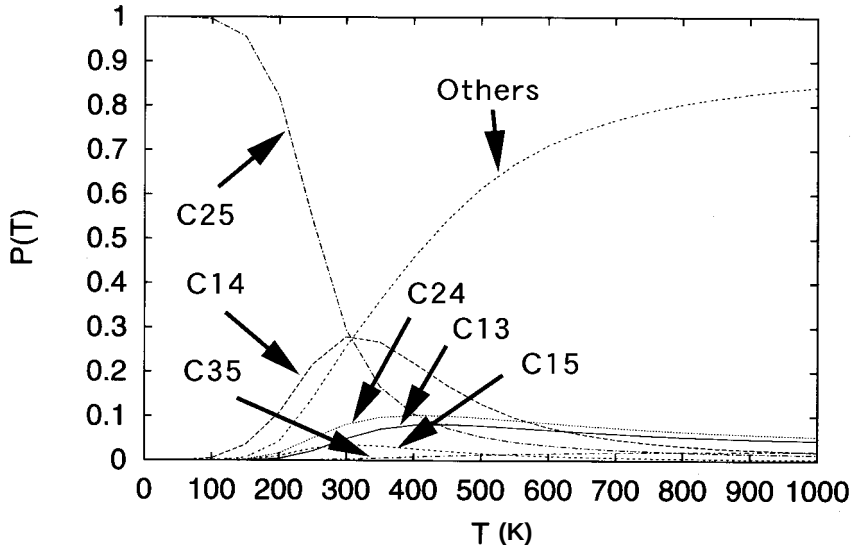

Figure 9: The distributions of each group of similar structures of Met-enkephalin in gas phase as a function of temperature.

Table 1: Acceptance Ratios of Replica Exchange Corresponding to Pairs of Neighboring Temperatures

Pair of Temperatures (K)	Acceptance Ratio
200 $\longleftrightarrow$ 239	0.160
239 $\longleftrightarrow$ 286	0.149
286 $\longleftrightarrow$ 342	0.143
342 $\longleftrightarrow$ 409	0.139
409 $\longleftrightarrow$ 489	0.142
489 $\longleftrightarrow$ 585	0.146
585 $\longleftrightarrow$ 700	0.146

random walk in temperature space between the lowest and highest temperatures. In Figure 10(b) the corresponding time series of the total potential energy is shown. We see that a random walk in potential energy space between low and high energies is realized. We remark that the potential energy here is that of AMBER in Ref. [117]. Note that there is a strong correlation between the behaviors in Figures 10(a) and 10(b).

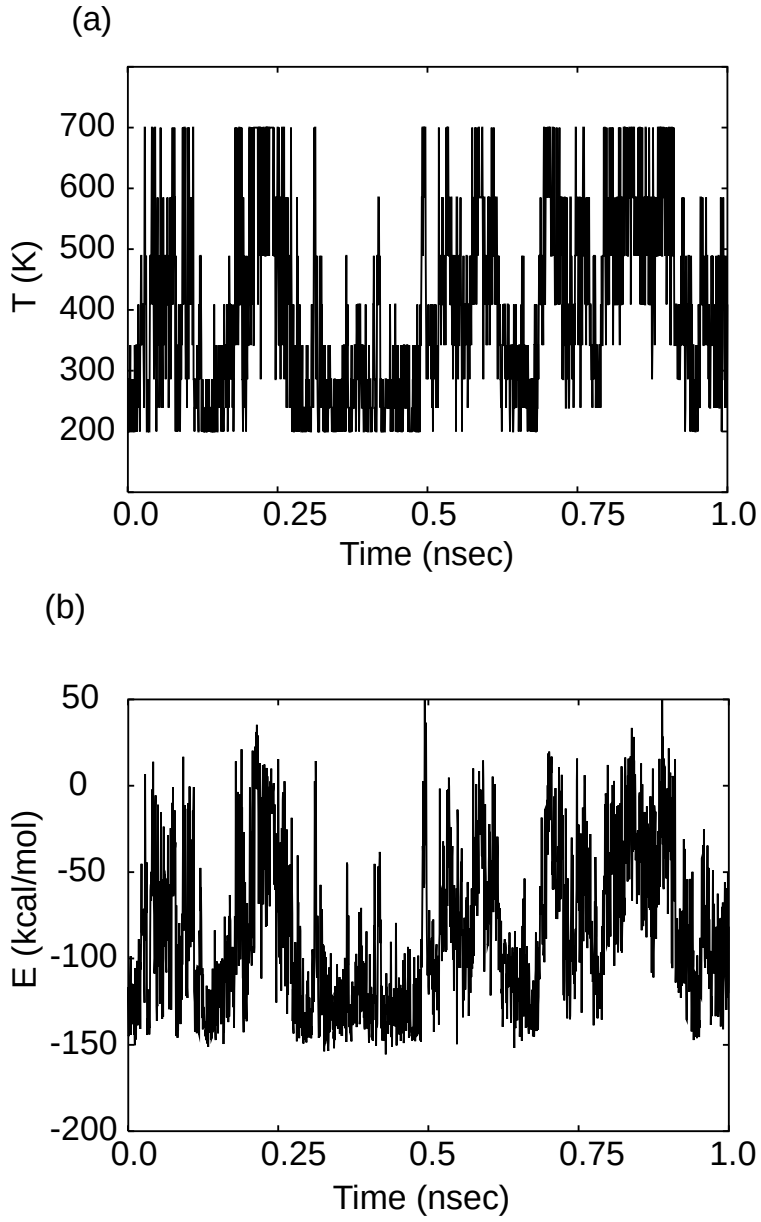

Figure 10: Time series of (a) temperature exchange and (b) the total potential energy for one of the replicas from a replica-exchange MD simulation of Met-enkephalin in gas phase.

In Figure 11 the canonical probability distributions obtained at the chosen eight temperatures from the replica-exchange simulation are shown. We see that there are enough overlaps between all pairs of distributions, indicating that there will be sufficient numbers of replica exchanges between pairs of replicas (see Table 1).

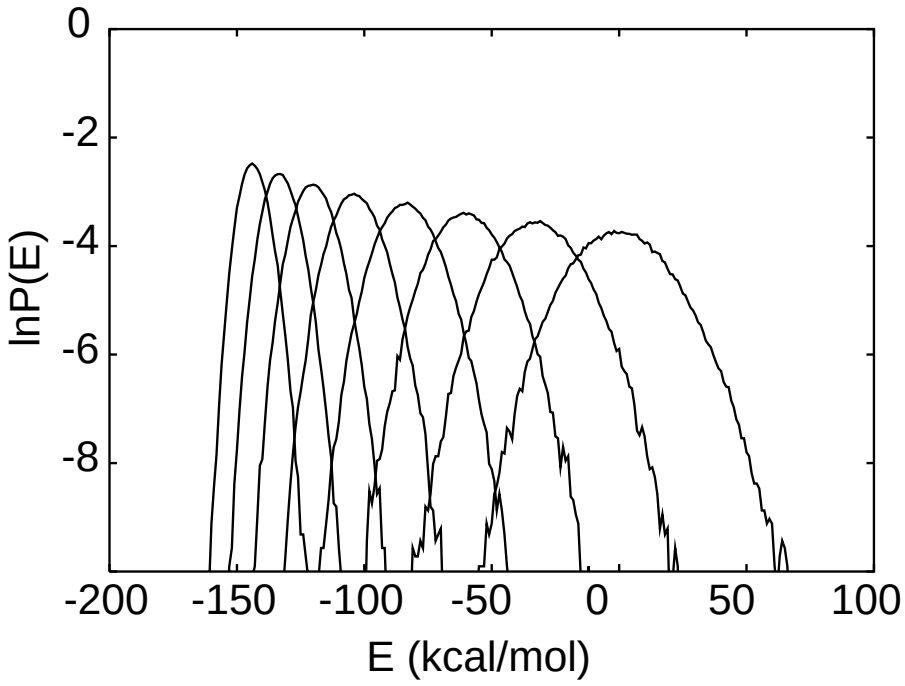

Figure 11: The canonical probability distributions of the total potential energy of Met-enkephalin in gas phase obtained from the replica-exchange MD simulation at the eight temperatures. The distributions correspond to the following temperatures (from left to right): 200, 239, 286, 342, 409, 489, 585, and 700 K.

We further compare the results of the replica-exchange simulation with those of a single canonical MD simulation (of 1 ns) at the corresponding temperatures. In Figure 12 we compare the distributions of a pair of dihedral angles (ϕ, ψ) of Gly-2 at two extreme temperatures ($T = 200$ K and 700 K). While the results at $T = 200$ K from the regular canonical simulation are localized with only one dominant peak, those from the replica-exchange simulation have several peaks (compare Figures 12(a) and 12(b)). Hence, the replica-exchange run samples much broader configurational space than the conventional canonical run at low temperatures. The results at $T = 700$ K (Figures 12(c) and 12(d)), on the other hand, are similar, implying that a regular canonical simulation can give accurate thermodynamic quantities at high temperatures.

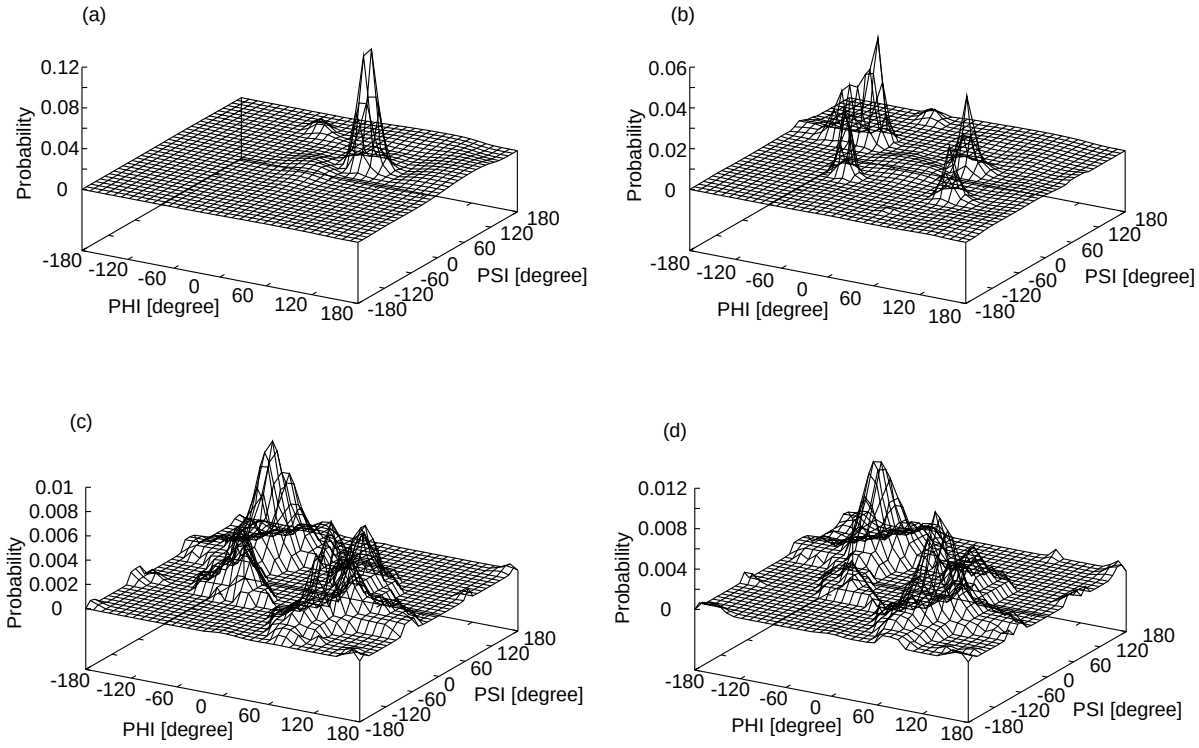

Figure 12: Distributions of a pair of dihedral angles (ϕ, ψ) of Gly-2 for: (a) $T = 200$ K from a regular canonical MD simulation, (b) $T = 200$ K from the replica-exchange MD simulation, (c) $T = 700$ K from a regular canonical MD simulation, and (d) $T = 700$ K from the replica-exchange MD simulation.

In Figure 13 we show the average total potential energy as a function of temperature. As expected from the results of Figure 12, we observe that the canonical simulations at low temperatures got trapped in states of energy local minima, resulting in the discrepancies in average values between the results from the canonical simulations and those from the replica-exchange simulation.

We now present the results of MD simulations based on replica-exchange multicanonical algorithm and multicanonical replica-exchange method [94]. The Met-enkephalin in gas phase was studied again. The potential energy is, however, that of AMBER in

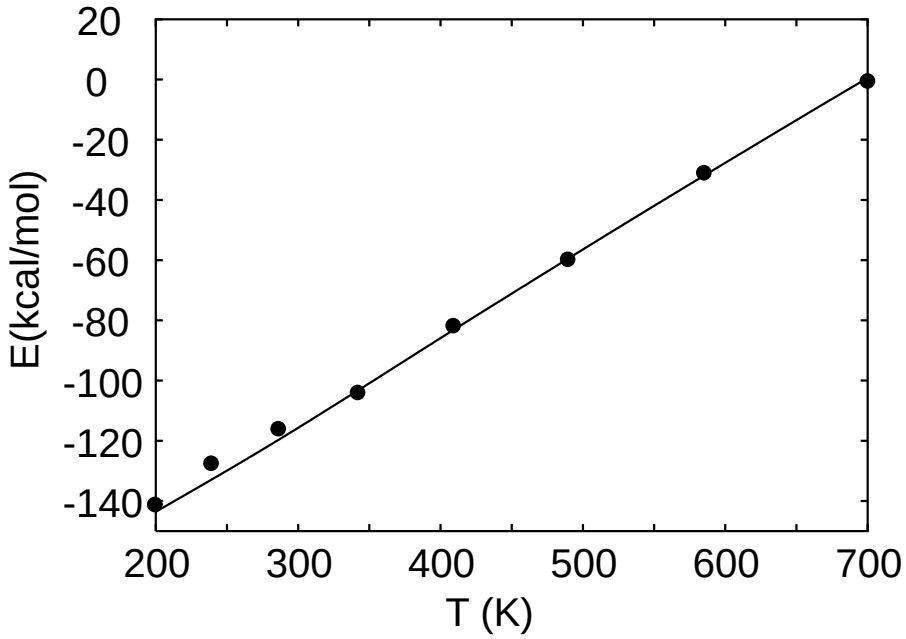

Figure 13: Average total potential energy of Met-enkephalin in gas phase as a function of temperature. The solid curve is the result from the replica-exchange MD simulation and the dots are those of regular canonical MD simulations.

Ref. [118] instead of Ref. [117]. In Table 2 we summarize the parameters of the simulations that were performed. As discussed in the previous section, REMUCA consists of two simulations: a short REM simulation (from which the density of states of the system, or the multicanonical weight factor, is determined) and a subsequent production run of MUCA simulation. The former simulation is referred to as REM1 and the latter as MUCA1 in Table 2. A production run of MUCAREM simulation is referred to as MUCAREM1 in Table 2, and it uses the same density of states that was obtained from REM1. Finally, a production run of the original REM simulation was also performed for comparison and it is referred to as REM2 in Table 2. The total simulation time for the three production runs (REM2, MUCA1, and MUCAREM1) was all set equal (i.e., 5 ns).

Table 2: Summary of Parameters in REM, REMUCA, and MUCAREM Simulations

Run	No. of Replicas, M	Temperature, T_m (K) $(m = 1, \cdots, M)$	MD Steps
REM1	10	200, 239, 286, 342, 409, 489, 585, 700, 836, 1000	2×10^5
REM2	10	200, 239, 286, 342, 409, 489, 585, 700, 836, 1000	1×10^6
MUCA1	1	1000	1×10^7
MUCAREM1	4	375, 525, 725, 1000	2.5×10^6

After the simulation of REM1 is finished, we obtained the density of states, $n(E)$,

by the multiple-histogram reweighting techniques of Eqs. (30) and (31). The density of states will give the average values of the potential energy from Eq. (20), and we found

$$\begin{cases} E_1 &= < E >_{T_1} = -30 \text{ kcal/mol} , \\ E_M &= < E >_{T_M} = 195 \text{ kcal/mol} . \end{cases} \tag{59}$$

Then our estimate of the density of states is reliable in the range $E_1 \leq E \leq E_M$. The multicanonical potential energy $\mathcal{E}_{mu}^{\{0\}}(E)$ was thus determined for the three energy regions ($E < E_1$, $E_1 \leq E \leq E_M$, and $E > E_M$) from Eq. (49). Namely, the multicanonical potential energy, $E_{mu}(E;T_0)$, in Eq. (9) and its derivative, $\frac{\partial E_{mu}(E;T_0)}{\partial E}$, were determined by fitting $\ln n(E)$ by cubic spline functions in the energy region of ($-30 \leq E \leq 195$ kcal/mol) [94]. Here, we have set the arbitrary reference temperature to be $T_0 = 1000$ K. Outside this energy region, $E_{mu}(E;T_0)$ was linearly extrapolated as in Eq. (49).

After determining the multicanonical weight factor, we carried out a multicanonical MD simulation of 1×10^7 steps (or 5 ns) for data collection (MUCA1 in Table 2). In Figure 14 the probability distribution obtained by MUCA1 is plotted. It can be seen that a good flat distribution is obtained in the energy region $E_1 \leq E \leq E_M$. In Figure 14 the canonical probability distributions that were obtained by the reweighting techniques at $T = T_1 = 200$ K and $T = T_M = 1000$ K are also shown (these results are essentially identical to one another among MUCA1, MUCAREM1, and REM2, as discussed below). Comparing these curves with those of MUCA1 in the energy regions $E < E_1$ and $E > E_M$ in Figure 14, we confirm our claim in the previous section that MUCA1 gives canonical distributions at $T = T_1$ for $E < E_1$ and at $T = T_M$ for $E > E_M$, whereas it gives a multicanonical distribution for $E_1 \leq E \leq E_M$.

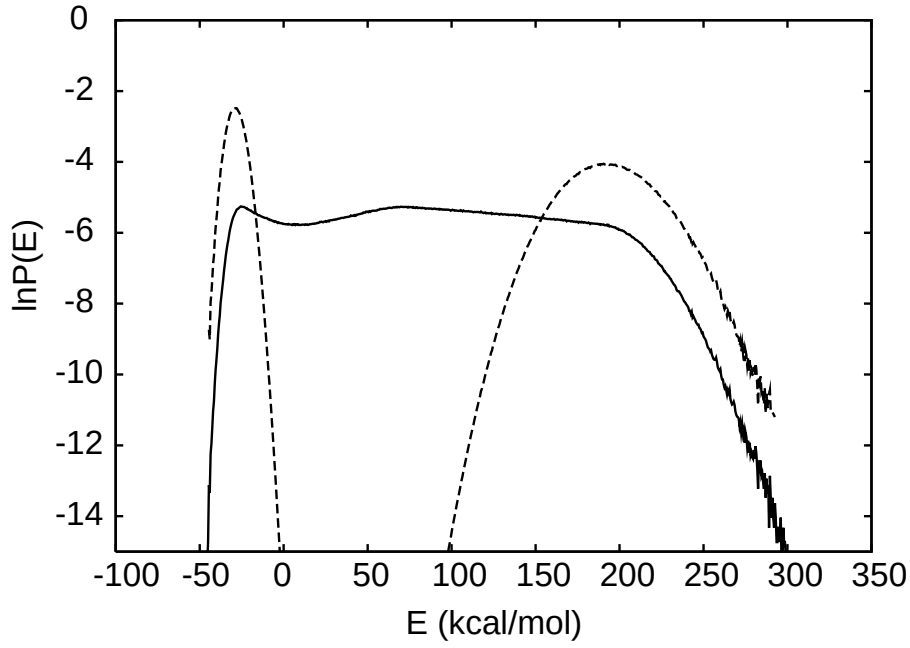

Figure 14: Probability distribution of potential energy of Met-enkephalin in gas phase that was obtained from MUCA1 (see Table 2). The dotted curves are the probability distributions of the reweighted canonical ensemble at $T = 200$ K (left) and 1000 K (right).

In the previous works of multicanonical simulations of Met-enkephalin in gas phase (see, for instance, Refs. [15, 39]), at least several iterations of trial simulations were required for the multicanonical weight determination. We emphasize that in the present case of REMUCA (REM1), only one simulation was necessary to determine the optimal multicanonical weight factor that can cover the energy region corresponding to temperatures between 200 K and 1000 K.

From the density of states obtained by REMUCA (i.e., REM1), we prepared the multicanonical weight factors (or the multicanonical potential energies) for the MUCAREM simulation (see Eq. (53)). The parameters of MUCAREM1, such as energy bounds $E_L^{\{m\}}$ and $E_H^{\{m\}}$ $(m = 1, \cdots, \mathcal{M})$ are listed in Table 3. The choices of $T_L^{\{m\}}$ and $T_H^{\{m\}}$ are, in general, arbitrary, but significant overlaps between the probability distributions of adjacent replicas are necessary. The replica-exchange process in MUCAREM1 was tried every 200 time steps (or 100 fs). It is less frequent than in REM1 (or REM2). This is because we wanted to ensure a sufficient time for system relaxation.

Table 3: Summary of Parameters in MUCAREM1

m	$T_L^{\{m\}}$ (K)	$T_H^{\{m\}}$ (K)	T_m (K)	$E_L^{\{m\}}$ (kcal/mol)	$E_H^{\{m\}}$ (kcal/mol)
1	200	375	375	-30	20
2	300	525	525	-5	65
3	375	725	725	20	120
4	525	1000	1000	65	195

In Figure 15 the probability distributions of potential energy obtained by MUCAREM1 are shown. As expected, we observe that the probability distributions corresponding to the temperature T_m are essentially flat for the energy region $E_L^{\{m\}} \leq E \leq E_H^{\{m\}}$, are of the canonical simulation at $T = T_L^{\{m\}}$ for $E < E_L^{\{m\}}$, and are of the canonical simulation at $T = T_H^{\{m\}}$ for $E > E_H^{\{m\}}$ $(m = 1, \cdots, \mathcal{M})$. As a result, each distribution in MUCAREM is much broader than those in the conventional REM and a much smaller number of replicas are required in MUCAREM than in REM ($\mathcal{M} = 4$ in MUCAREM versus $M = 10$ in REM).

In Figure 16 the time series of potential energy for the first 500 ps of REM2 (a), MUCA1 (b), and MUCAREM1 (c) are plotted. They all exhibit a random walk in potential energy space, implying that they all perfomed properly as generalized-ensemble algorithms. To check the validity of the canonical-ensemble expectation values calculated by the new algorithms, we compare the average potential energy as a function of temperature in Figure 17. In REM2 and MUCAREM1 we used the multiple-histogram techniques [2, 3], whereas the single-histogram method [1] was used in MUCA1. We can see a perfect coincidence of these quantities among REM2, MUCA1, and MUCAREM1 in Figure 17.

We now present the results of a replica-exchange simulated tempering MC simulation of Met-enkephalin in gas phase [95]. The potential energy is again that of ECEPP/2 [97]–[99]. In Table 4 we summarize the parameters of the simulations that were performed. As described in the previous section, REST consists of two simulations: a short REM simulation (from which the dimensionless Helmholtz free energy, or the simulated tempering weight factor, is determined) and a subsequent ST production run. The former simulation

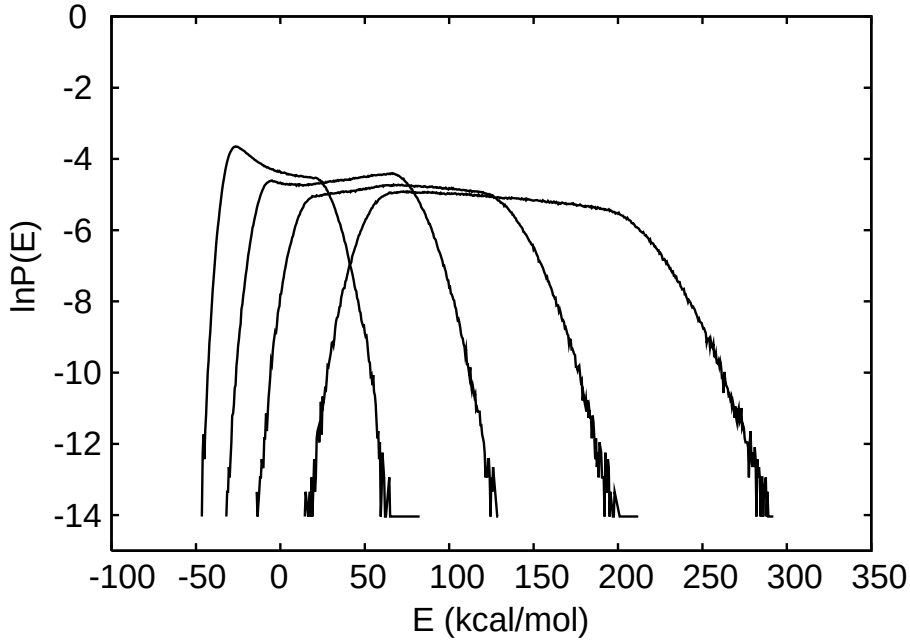

Figure 15: Probability distributions of potential energy obtained from MUCAREM1 (see Tables 2 and 3).

is referred to as REM1 and the latter as ST1 in Table 4. In REM1 there exist 8 replicas with 8 different temperatures ($M = 8$), ranging from 50 K to 1000 K as listed in Table 4 (i.e., $T_1 = 50$ K and $T_M = T_8 = 1000$ K). The same set of temperatures were also used in ST1. The temperatures were distributed exponentially between T_1 and T_M, following the optimal distribution found in the previous simulated annealing schedule [101], simulated tempering run [52], and replica-exchange simulation [82]. After estimating the weight factor, we made a ST production run of 10^6 MC sweeps (ST1). In REM1 and ST1, a replica exchange and a temperature update, respectively, were tried every 10 MC sweeps.

Table 4: Summary of Parameters in REST Simulations

Run	No. of Replicas, M	Temperature, T_m (K) ($m = 1, \cdots, M$)	MC Sweeps
REM1	8	50, 77, 118, 181, 277, 425, 652, 1000	5×10^4
ST1	1	50, 77, 118, 181, 277, 425, 652, 1000	1×10^6

We first check whether the replica-exchange simulation of REM1 indeed performed properly. For an optimal performance of REM the acceptance ratios of replica exchange should be sufficiently uniform and large (say, > 10 %). In Table 5 we list these quantities. It is clear that both points are met in the sense that they are of the same order (the values vary between 10 % and 40 %).

After determining the simulated tempering weight factor, we carried out a long ST simulation for data collection (ST1 in Table 4). In Figure 18 the time series of temperature and potential energy from ST1 are plotted. In Figure 18(a) we observe a random walk in temperature space between the lowest and highest temperatures. In Figure 18(b) the

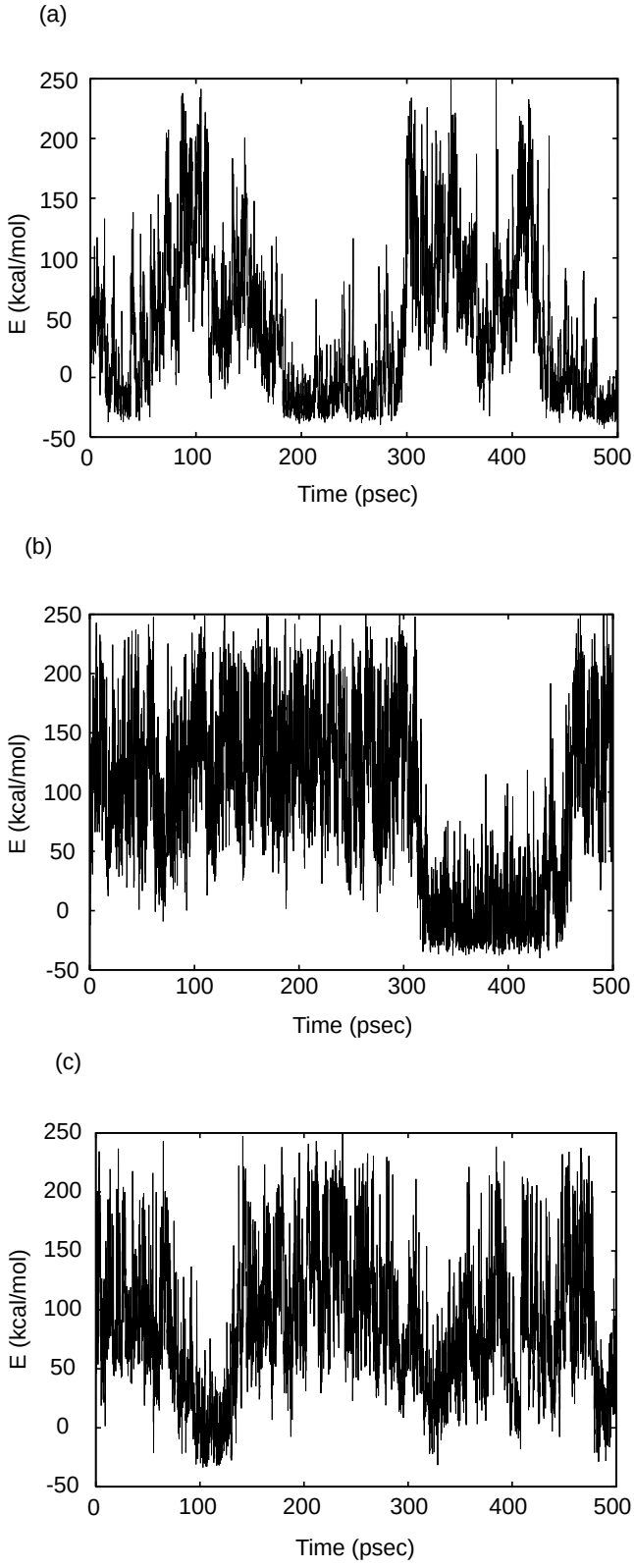

Figure 16: Time series of potential energy of Met-enkephalin in gas phase for one of the replicas in (a) REM2, (b) MUCA1, and (c) MUCAREM1 (see Tables 2 and 3 for the parameters of the simulations).

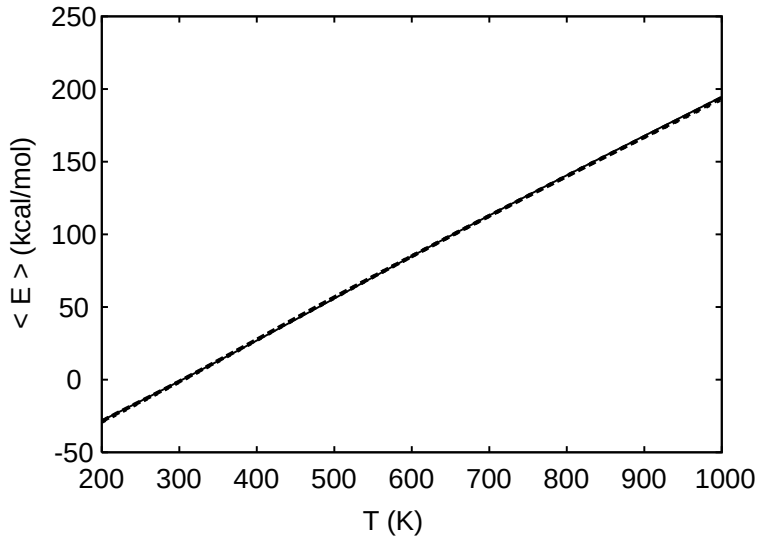

Figure 17: The average potential energy of Met-enkephalin in gas phase as a function of temperature. The solid, dotted, and dashed curves are obtained from REM2, MUCA1, and MUCAREM1, respectively (see Tables 2 and 3 for the parameters of the simulations).

Table 5: Acceptance Ratios of Replica Exchange in REM1 of Table 4

Pair of Temperatures (K)	Acceptance Ratio
50 $\longleftrightarrow$ 77	0.30
77 $\longleftrightarrow$ 118	0.27
118 $\longleftrightarrow$ 181	0.22
181 $\longleftrightarrow$ 277	0.17
277 $\longleftrightarrow$ 425	0.10
425 $\longleftrightarrow$ 652	0.27
652 $\longleftrightarrow$ 1000	0.40

corresponding random walk of the total potential energy between low and high energies is observed. Note that there is a strong correlation between the behaviors in Figures 18(a) and 18(b), as there should. It is known from our previous works that the global-minimum-energy conformation for Met-enkephalin in gas phase has the ECEPP/2 energy value of -12.2 kcal/mol [19, 39]. Hence, the random walk in Figure 18(b) indeed visited the global-minimum region many times. It also visited high energy regions, judging from the fact that the average potential energy is around 15 kcal/mol at $T = 1000$ K [15, 39] (see also Figure 19 below).

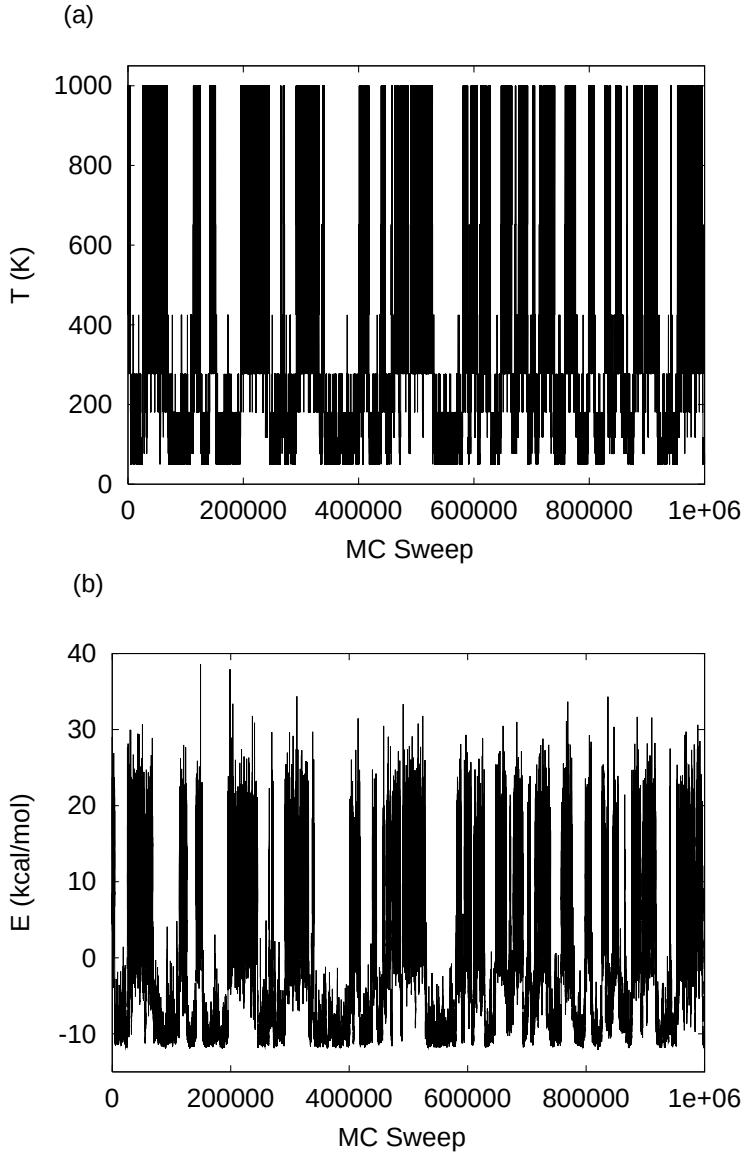

Figure 18: Time series of (a) temperature and (b) potential energy in ST1 (see Table 4 for the parameters of the simulation).

For an optimal performance of ST, the acceptance ratios of temperature update should be sufficiently uniform and large. In Table 6 we list these quantities. It is clear that both points are met (the values vary between 26 % and 57 %); we find that the present ST run (ST1) indeed properly performed. We remark that the acceptance ratios in Table 6 are

significantly larger and more uniform than those in Table 5, suggesting that ST runs can sample the configurational space more effectively than REM runs, provided the optimal weight factor is obtained.

Table 6: Acceptance Ratios of Temperature Update in ST1

Pair of Temperatures (K)	Acceptance Ratio
50 $\longrightarrow$ 77	0.47
77 $\longrightarrow$ 50	0.47
77 $\longrightarrow$ 118	0.43
118 $\longrightarrow$ 77	0.43
118 $\longrightarrow$ 181	0.37
181 $\longrightarrow$ 118	0.42
181 $\longrightarrow$ 277	0.29
277 $\longrightarrow$ 181	0.29
277 $\longrightarrow$ 425	0.30
425 $\longrightarrow$ 277	0.26
425 $\longrightarrow$ 652	0.43
652 $\longrightarrow$ 425	0.42
652 $\longrightarrow$ 1000	0.57
1000 $\longrightarrow$ 652	0.56

We remark that the details of Monte Carlo versions of REMUCA and MUCAREM have also been worked out and tested with Met-enkephalin in gas phase [134]. Here in Figure 19, we just show the average ECEPP/2 potential energy as a function of temperature that was calculated from the four generalized-ensemble algorithms, MUCA, REMUCA, MUCAREM, and REST [134]. The results are in good agreement.

We have so far presented the results of generalized-ensemble simulations of Met-enkephalin in gas phase. However, peptides and proteins are usually in aqueous solution. We therefore want to incorporate rigorous solvation effects in our simulations in order to compare with experiments.

Our first example with rigorous solvent effects is a multicanonical MC simulation, where the solvation term was included by the RISM theory [45]. While low-energy conformations of Met-enkephalin in gas phase are compact and form β-turn structures [39], it turned out that those in aqueous solution are extended. In Figure 20 we show the lowest-energy conformations of Met-enkephalin obtained during the multicanonical MC simulation with RISM theory incorporated [45]. They exhibit characteristics of almost fully extended backbone structure with large side-chain fluctuations. The results are in accord with the observations in NMR experiments, which also suggest extended conformations [135].

We also calculated an average of the end-to-end distance of Met-enkephalin as a function of temperature. The results in aqueous solution (the present simulation) and in the gas phase (a previous simulation [39]) are compared in Figure 21. The end-to-end distance in aqueous solution at all temperatures varies little (around 12 Å); the conformations are extended in the entire temperature range. On the other hand, in the gas phase, the end-to-end distance is small at low temperatures due to intrachain hydrogen bonds, while

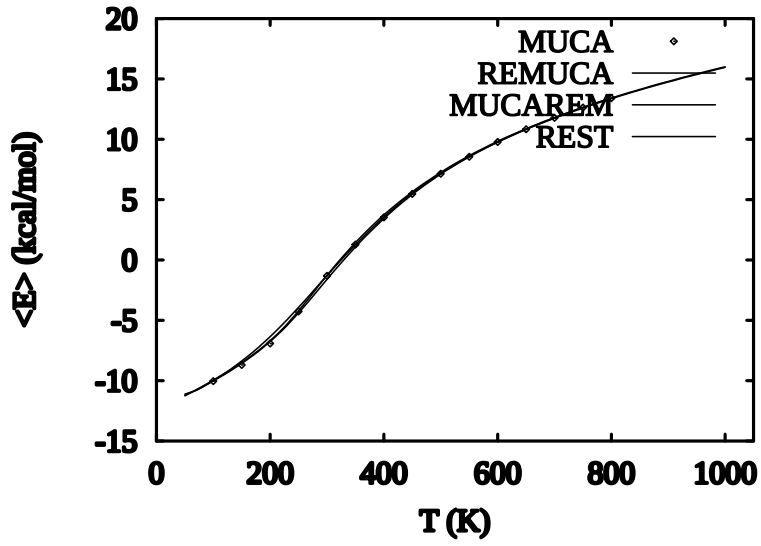

Figure 19: The average potential energy of Met-enkephalin in gas phase as a function of temperature. The results from the four generalized-ensemble algorithms, MUCA, REMUCA, MUCAREM, and REST, are superimposed.

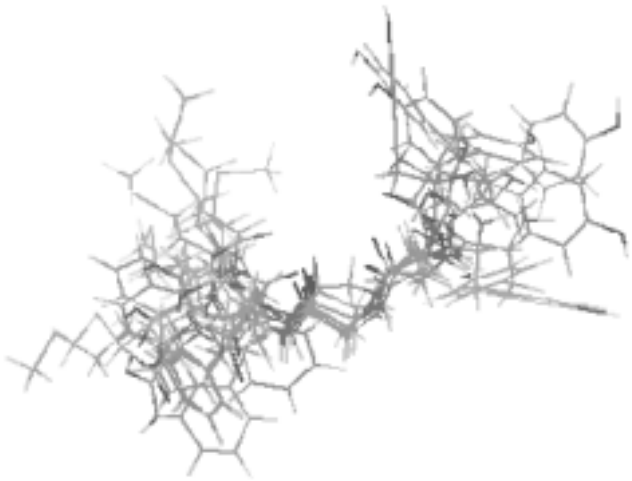

Figure 20: Superposition of eight representative low-energy conformations of Met-enkephalin obtained by the multicanonical MC simulation in aqueous solution based on RISM. The figure was created with RasMol [131].

the distance is large at high temperatures, because these intrachain hydrogen bonds are broken.

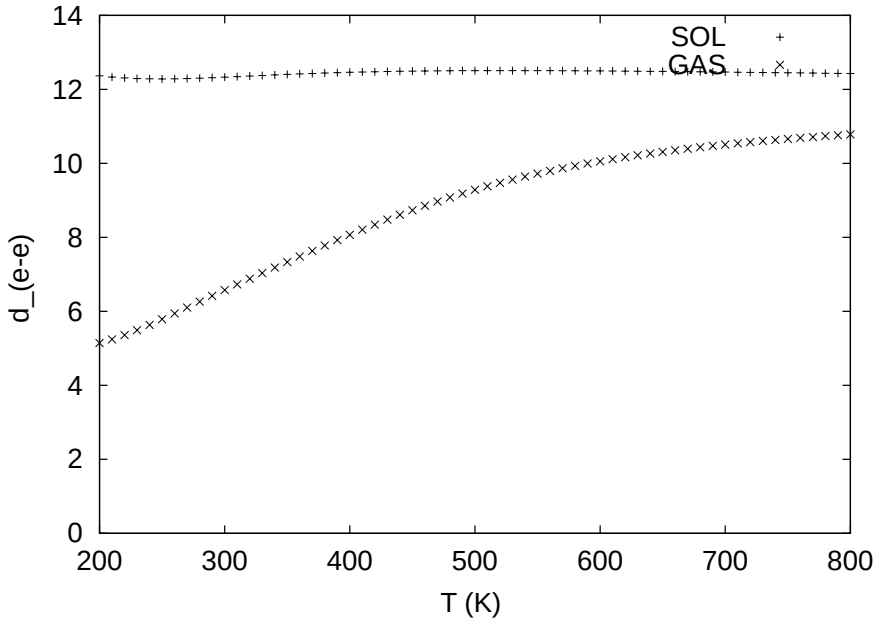

Figure 21: Average end-to-end distance of Met-enkephalin in aqueous solution (SOL) and in gas phase (GAS) as a function of temperature. Here, the end-to-end distance is defined as the distance between the nitrogen atom at the N terminus and the oxygen atom at the C terminus.

The same peptide was also studied by MD simulations of replica-exchange and other generalized-ensemble simulations in aqueous solution based on TIP3P water model [136]. Two AMBER force fields [118, 119] were used. The number of water molecules was 526 and they were placed in a sphere of radius of 16 Å. The initial configuration is shown in Figure 22.

In Figure 23 the canonical probability distributions obtained at the 24 temperatures from the replica-exchange simulation are shown. We see that there are enough overlaps between all pairs of distributions, indicating that there will be sufficient numbers of replica exchanges between pairs of replicas. The corresponding time series of the total potential energy for one of the replicas is shown in Figure 24. We do observe a random walk in potential energy space, which covers an energy range of as much as 2,000 kcal/mol.

Finally, the average end-to-end distance as a function of temperature was calculated by the multiple-histogram reweighting techniques of Eqs. (30) and (31). The results both in gas phase and in aqueous solution are shown in Figure 25. The results are in good agreement with those of ECEPP/2 energy plus RISM solvation theory [45] in the sense that Met-enkephalin is compact at low temperatures and extended at high temperatures in gas phase and extended in the entire temperature range in aqueous solution (compare Figures 21 and 25).

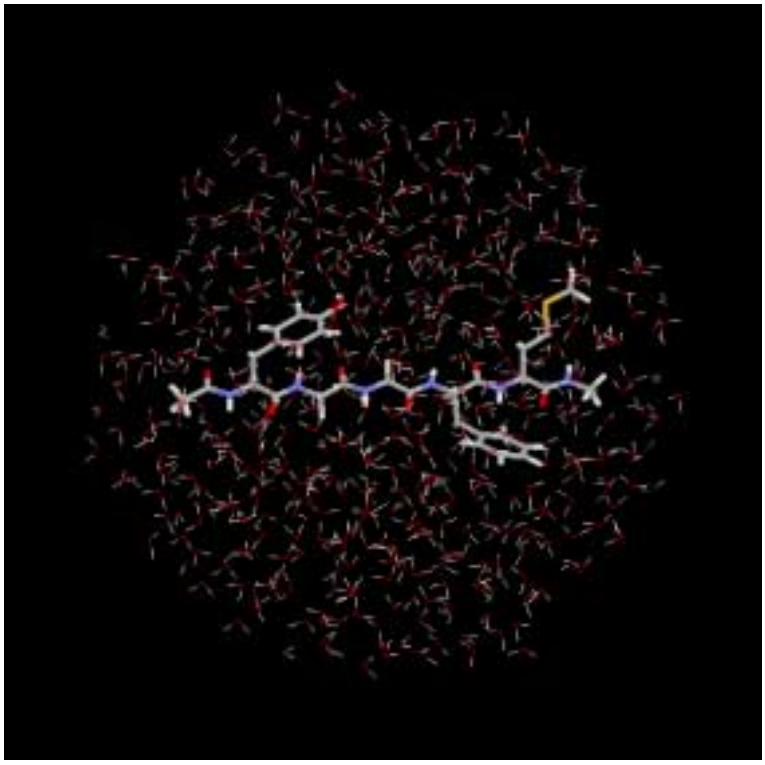

Figure 22: Initial configuration of replica-exchange MD simulations of Met-enkephalin in aqueous solution with 526 TIP3P water molecules.

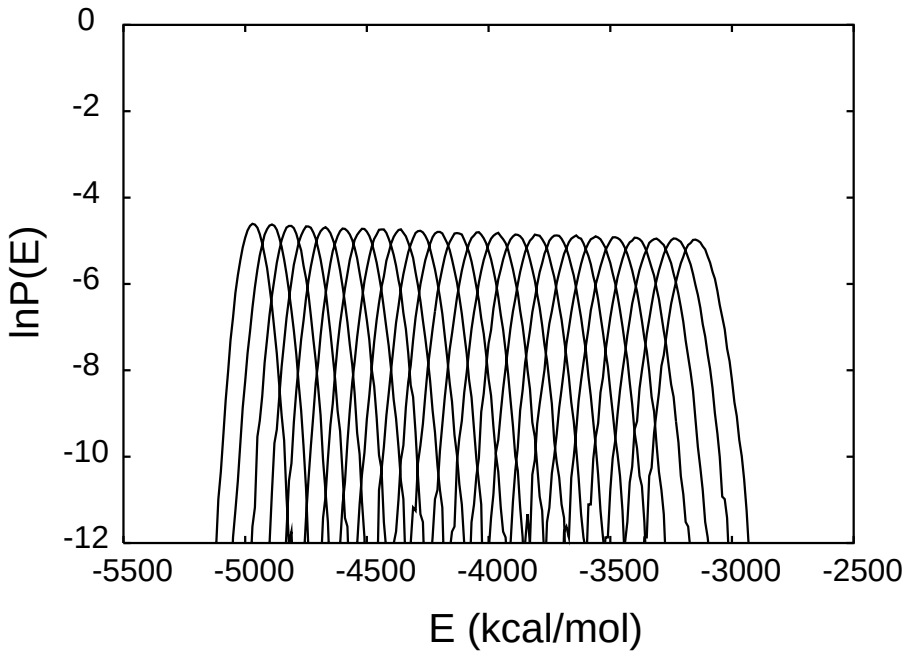

Figure 23: The canonical probability distributions of the total potential energy of Met-enkephalin in aqueous solution obtained from the replica-exchange MD simulation at the 24 temperatures.

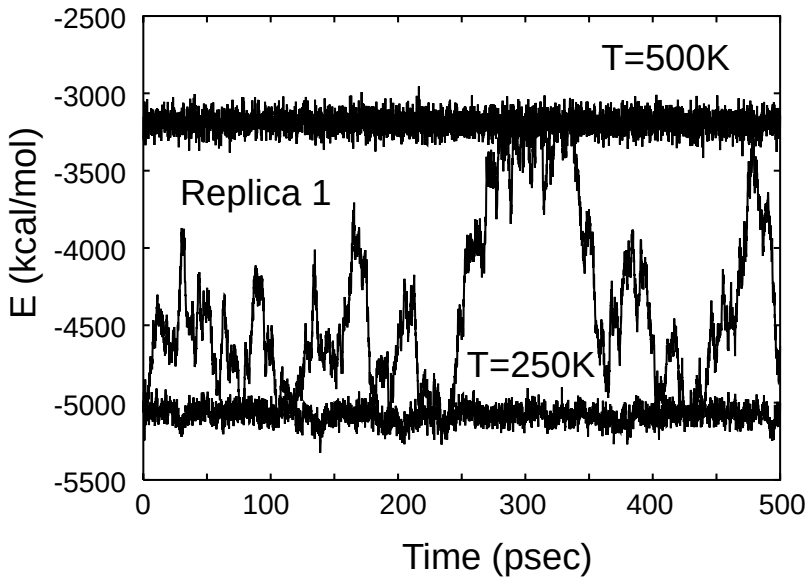

Figure 24: Time series of the total potential energy of Met-enkephalin in aqueous solution obtained for one of the replicas from the replica-exchange MD simulation. Corresponding times series in the canonical ensemble at temperatures 250 K and 500 K are also shown.

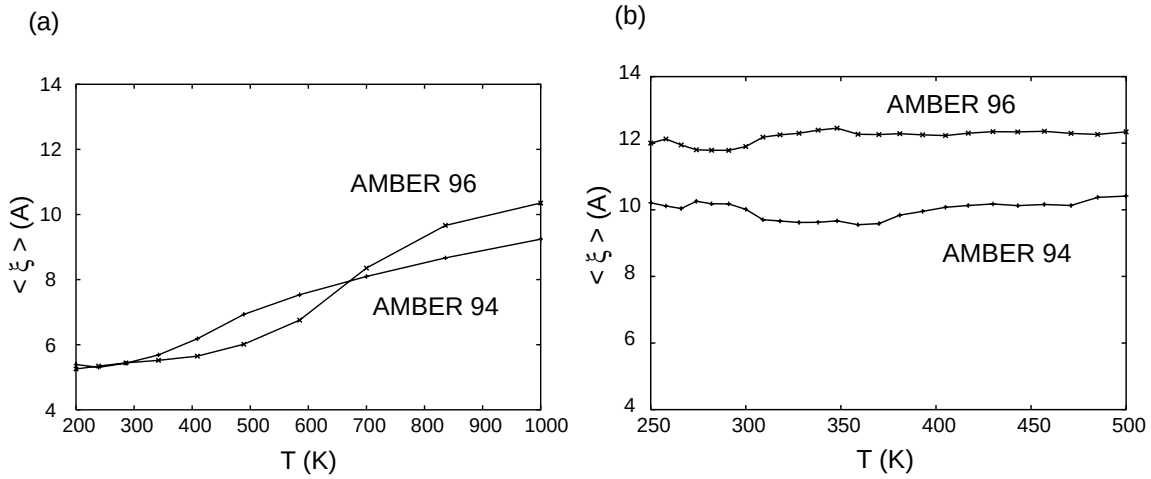

Figure 25: Average end-to-end distance of Met-enkephalin (a) in gas phase and (b) in aqueous solution as a function of temperature.

4 CONCLUSIONS

In this article we have reviewed uses of generalized-ensemble algorithms in molecular simulations of biomolecules. A simulation in generalized ensemble realizes a random walk in potential energy space, alleviating the multiple-minima problem that is a common difficulty in simulations of complex systems with many degrees of freedom.

Detailed formulations of the three well-known generalized-ensemble algorithms, namely, multicaonical algorithm (MUCA), simulated tempering (ST), and replica-exchange method (REM), were given. We then introduced three new generalized-ensemble algorithms that combine the merits of the above three methods, which we refer to as replica-exchange multicanonical algorithm (REMUCA), replica-exchange simulated tempering (REST), and multicanonical replica-exchange method (MUCAREM).

With these new methods available, we believe that we now have working simulation algorithms which we can use for conformational predictions of peptides and proteins from the first principles, using the information of their amino-acid sequence only. It is now high time that we addressed the question of the validity of the standard potential energy functions such as AMBER, CHARMM, GROMOS, ECEPP, etc. For this purpose, conventional simulations in the canonical ensemble are of little use because they will necessarily get trapped in states of local-minmum-energy states. It is therefore essential to use generalized-ensemble algorithms in order to test and develop accurate potential energy functions for biomolecular systems. Some preliminary results of comparisons among ECEPP/2 and different versions of AMBER force fields were given in the present article. We remark that more detailed analyses that compare different versions of AMBER by multicanonical MD simulations already exist [137]. Likewise, the validity of solvation theories should also be tested. For this, RISM theory [109]–[111] can be very useful. For instance, we have successfully given a molecular mechanism of secondary structural transitions in peptides due to addition of alcohol to solvent [138], which is very difficult to attain by regular molecular simulations.

Acknowledgements:
Our simulations were performed on the Hitachi and other computers at the Research Center for Computational Science, Okazaki National Research Institutes. This work is supported, in part, by a grant from the Research for the Future Program of the Japan Society for the Promotion of Science (JSPS-RFTF98P01101).

References

[1] Ferrenberg, A.M. & Swendsen, R.H. (1988) *Phys. Rev. Lett.* **61**, 2635–2638; *ibid.* (1989) **63**, 1658.

[2] Ferrenberg, A.M. & Swendsen, R.H. (1989) *Phys. Rev. Lett.* **63**, 1195–1198.

[3] Kumar, S., Bouzida, D., Swendsen, R.H., Kollman, P.A. & Rosenberg, J.M. (1992) *J. Comput. Chem.* **13**, 1011–1021.

[4] Berg, B.A. & Neuhaus, T. (1991) *Phys. Lett.* **B267**, 249–253.

[5] Berg, B.A. & Neuhaus, T. (1992) *Phys. Rev. Lett.* **68**, 9–12.

[6] Berg, B.A. (2000) *Fields Institute Communications* **26**, 1–24; also see cond-mat/9909236.

[7] Lee, J. (1993) *Phys. Rev. Lett.* **71**, 211–214; *ibid.* **71**, 2353.

[8] Mezei, M. (1987) *J. Comput. Phys.* **68**, 237–248.

[9] Bartels, C. & Karplus, M. (1998) *J. Phys. Chem. B* **102**, 865–880.

[10] Berg, B.A., Hansmann, U.H.E. & Okamoto, Y. (1995) *J. Phys. Chem.* **99**, 2236–2237.

[11] Berg, B.A. & Celik, T. (1992) *Phys. Rev. Lett.* **69**, 2292–2295.

[12] Berg, B.A., Celik, T. & Hansmann, U.H.E. (1993) *Europhys. Lett.* **22**, 63–68.

[13] Berg, B.A. & Janke, W. (1998) *Phys. Rev. Lett.* **80**, 4771–4774.

[14] Hatano, N. & Gubernatis, J.E. (2000) *Prog. Theor. Phys. (Suppl.)* **138**, 442–447.

[15] Hansmann, U.H.E. & Okamoto, Y. (1993) *J. Comput. Chem.* **14**, 1333–1338.

[16] Okamoto, Y. (1998) *Recent Res. Devel. in Pure & Applied Chem.* **2**, 1–23.

[17] Hansmann, U.H.E. & Okamoto, Y. (1999) in *Annual Reviews of Computational Physics VI*, Stauffer, D., Ed., World Scientific, Singapore, pp. 129–157.

[18] Hansmann, U.H.E. & Okamoto, Y. (1999) *Curr. Opin. Struct. Biol.* **9**, 177–183.

[19] Hansmann, U.H.E. & Okamoto, Y. (1994) *Physica* **A212**, 415–437.

[20] Hao, M.H. & Scheraga, H.A. (1994) *J. Phys. Chem.* **98**, 4940–4948.

[21] Okamoto, Y., Hansmann, U.H.E. & Nakazawa, T. (1995) *Chem. Lett.* **1995**, 391–392.

[22] Okamoto, Y. & Hansmann, U.H.E. (1995) *J. Phys. Chem.* **99**, 11276–11287.

[23] Kidera, A. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 9886–9889.

[24] Kolinski, A., Galazka, W. & Skolnick, J. (1996) *Proteins* **26**, 271–287.

[25] Urakami, N. & Takasu, M. (1996) *J. Phys. Soc. Jpn.* **65**, 2694–2699.

[26] Kumar, S., Payne, P. & Vásquez, M. (1996) *J. Comput. Chem.* **17**, 1269–1275.

[27] Hansmann, U.H.E., Okamoto, Y. & Eisenmenger, F. (1996) *Chem. Phys. Lett.* **259**, 321–330.

[28] Nakajima, N., Nakamura, H. & Kidera, A. (1997) *J. Phys. Chem. B* **101**, 817–824.

[29] Eisenmenger, F. & Hansmann, U.H.E. (1997) *J. Phys. Chem. B* **101**, 3304–3310.

[30] Higo, J., Nakajima, N., Shirai, H., Kidera, A. & Nakamura, H. (1997) *J. Comput. Chem.* **18**, 2086–2092.

[31] Nakajima, N., Higo, Kidera, A. & Nakamura, H. (1997) *J. Chem. Phys.* **278**, 297–301.

[32] Noguchi, H. & Yoshikawa, K. (1997) *Chem. Phys. Lett.* **278**, 184–188.

[33] Kolinski, A., Galazka, W. & Skolnick, J. (1998) *J. Chem. Phys.* **108**, 2608–2617.

[34] Iba, Y., Chikenji, G. & Kikuchi, M. (1998) *J. Phys. Soc. Jpn.* **67**, 3327–3330.

[35] Nakajima, N. (1998) *Chem. Phys. Lett.* **288**, 319–326.

[36] Hao, M.H. & Scheraga, H.A. (1998) *J. Mol. Biol.* **277**, 973–983.

[37] Shirai, H., Nakajima, N., Higo, J., Kidera, A. & Nakamura, H. (1998) *J. Mol. Biol.* **278**, 481–496.

[38] Schaefer, M., Bartels, C. & Karplus, M. (1998) *J. Mol. Biol.* **284**, 835–848.

[39] Mitsutake, A., Hansmann, U.H.E. & Okamoto, Y. (1998) *J. Mol. Graphics Mod.* **16**, 226–238; 262–263.

[40] Hansmann, U.H.E. & Okamoto, Y. (1999) *J. Phys. Chem. B* **103**, 1595–1604.

[41] Shimizu, H., Uehara, K., Yamamoto, K. & Hiwatari, Y. (1999) *Mol. Sim.* **22**, 285–301.

[42] Ono, S., Nakajima, N., Higo, J. & Nakamura, H. (1999) *Chem. Phys. Lett.* **312**, 247–254.

[43] Mitsutake, A. & Okamoto, Y. (2000) *J. Chem. Phys.* **112**, 10638–10647.

[44] Yasar, F., Celik, T., Berg, B.A. & Meirovitch, H. (2000) *J. Comput. Chem.* **21**, 1251–1261.

[45] Mitsutake, A., Kinoshita, M., Okamoto, Y. & Hirata, F. (2000) *Chem. Phys. Lett.* **329**, 295–303.

[46] Munakata, T. & Oyama, S. (1996) *Phys. Rev. E* **54**, 4394–4398.

[47] Lyubartsev, A.P., Martinovski, A.A., Shevkunov, S.V. & Vorontsov-Velyaminov, P.N. (1992) *J. Chem. Phys.* **96**, 1776–1783.

[48] Marinari E. & Parisi, G. (1992) *Europhys. Lett.* **19**, 451–458.

[49] Marinari, E., Parisi, G. & Ruiz-Lorenzo, J.J. (1998) in *Spin Glasses and Random Fields*, Young, A.P., Ed., World Scientific, Singapore, pp. 59–98.

[50] Irbäck, A. & Potthast, F. (1995) *J. Chem. Phys.* **103**, 10298–10305.

[51] Hansmann, U.H.E. & Okamoto, Y. (1996) *Phys. Rev. E* **54**, 5863–5865.

[52] Hansmann, U.H.E. & Okamoto, Y. (1997) *J. Comput. Chem.* **18**, 920–933.

[53] Irbäck, A. & Sandelin, E. (1999) *J. Chem. Phys.* **110**, 12256–12262.

[54] Hesselbo, B. & Stinchcombe, R.B. (1995) *Phys. Rev. Lett.* **74**, 2151–2155.

[55] Smith, G.R. & Bruce, A.D. (1996) *Phys. Rev. E* **53**, 6530–6543.

[56] Hansmann, U.H.E. (1997) *Phys. Rev. E* **56**, 6201–6203.

[57] Berg, B.A. (1998) *Nucl. Phys. B* (Proc. Suppl.) **63A-C**, 982–984.

[58] Tsallis, C. (1988) *J. Stat. Phys.* **52**, 479–487.

[59] Hansmann, U.H.E. & Okamoto, Y. (1997) *Phys. Rev. E* **56**, 2228–2233.

[60] Hansmann, U.H.E., Eisenmenger, F. & Okamoto, Y. (1998) *Chem. Phys. Lett.* **297**, 374–382.

[61] Hansmann, U.H.E., Masuya, M. & Okamoto, Y. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10652–10656.

[62] Hansmann, U.H.E., Okamoto, Y. & Onuchic, J.N. (1999) *Proteins* **34**, 472–483.

[63] Andricioaei, I. & Straub, J.E. (1997) *J. Chem. Phys.* **107**, 9117–9124.

[64] Munakata, T. & Mitsuoka, S. (2000) *J. Phys. Soc. Jpn.* **69**, 92–96.

[65] Kirkpatrick, S., Gelatt, C.D. Jr. & Vecchi, M.P. (1983) *Science* **220**, 671–680.

[66] Tsallis, C. & Stariolo, D.A. (1996) *Physica* **A233**, 395–406.

[67] Andricioaei, I. & Straub, J.E. (1996) *Phys. Rev. E* **53**, R3055–R3058.

[68] Hansmann, U.H.E. (1998) *Physica A* **242**, 250–257.

[69] Berne, B.J. & Straub, J.E. (1997) *Curr. Opin. Struct. Biol.* **7**, 181–189.

[70] Straub, J.E. & Andricioaei, I. (1999) *Braz. J. Phys.* **29**, 179–186.

[71] Hansmann, U.H.E. & Okamoto, Y. (1999) *Braz. J. Phys.* **29**, 187–198.

[72] Hukushima, K. & Nemoto, K. (1996) *J. Phys. Soc. Jpn.* **65**, 1604–1608.

[73] Hukushima, K., Takayama, H. & Nemoto, K. (1996) *Int. J. Mod. Phys. C* **7**, 337–344.

[74] Geyer, C.J. (1991) in *Computing Science and Statistics: Proc. 23rd Symp. on the Interface*, Keramidas, E.M., Ed., Interface Foundation, Fairfax Station, pp. 156–163.

[75] Swendsen, R.H. & Wang, J.-S. (1986) *Phys. Rev. Lett.* **57**, 2607–2609.

[76] Kimura, K. & Taki, K. (1991) in *Proc. 13th IMACS World Cong. on Computation and Appl. Math. (IMACS '91)*, Vichnevetsky, R. & Miller, J.J.H., Eds., vol. 2, pp. 827–828.

[77] Frantz, D.D., Freeman, D.L. & Doll, J.D. (1990) *J. Chem. Phys.* **93**, 2769–2784.

[78] Tesi, M.C., van Rensburg, E.J.J., Orlandini, E. & Whittington, S.G. (1996) *J. Stat. Phys.* **82**, 155–181.

[79] Iba, Y. (2000) "Extended ensemble Monte Carlo," submitted for publication; cond-mat/0012323.

[80] Hansmann, U.H.E. (1997) *Chem. Phys. Lett.* **281**, 140–150.

[81] Wu, M.G. & Deem, M.W. (1999) *J. Chem. Phys.* **111**, 6625–6632.

[82] Sugita, Y. & Okamoto, Y. (1999) *Chem. Phys. Lett.* **314**, 141–151.

[83] Gront, D., Kolinski, A. & Skolnick, J. (2000) *J. Chem. Phys.* **113**, 5065–5071.

[84] Sugita, Y., Kitao, A. & Okamoto, Y. (2000) *J. Chem. Phys.* **113**, 6042–6051.

[85] Garcia, A.E. & Sanbonmatsu, K.Y. (2001) *Proteins* **42**, 345–354.

[86] Yan, Q. & de Pablo, J.J. (1999) *J. Chem. Phys.* **111**, 9509–9516.

[87] Nishikawa, T., Ohtsuka, H., Sugita, Y., Mikami, M. & Okamoto, Y. (2000) *Prog. Theor. Phys. (Suppl.)* **138**, 270–271.

[88] Calvo, F., Neirotti, J.P., Freeman, D.L. & Doll, J.D. (2000) *J. Chem. Phys.* **112**, 10350–10357.

[89] Okabe, T., Kawata, M., Okamoto, Y. & Mikami, M. (2001) *Chem. Phys. Lett.* **335**, 435–439.

[90] Ishikawa, Y., Sugita, Y., Nishikawa, T. & Okamoto, Y. (2001) *Chem. Phys. Lett.* **333**, 199–206.

[91] Yamamoto, R. & Kob, W. (2000) *Phys. Rev. E* **61**, 5473–5476.

[92] Hukushima, K. (1999) *Phys. Rev. E* **60**, 3606–3614.

[93] Bunker, A. & Dünweg, B. (2000) "Parallel excluded volume tempering for polymer melts," *Phys. Rev. E*, in press.

[94] Sugita, Y. & Okamoto, Y. (2000) *Chem. Phys. Lett.* **329**, 261–270.

[95] Mitsutake, A. & Okamoto, Y. (2000) *Chem. Phys. Lett.* **332**, 131–138.

[96] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. & Teller, E. (1953) *J. Chem. Phys.* **21**, 1087–1092.

[97] Momany, F.A., McGuire, R.F., Burgess, A.W. & Scheraga, H.A. (1975) *J. Phys. Chem.* **79**, 2361–2381.

[98] Némethy, G., Pottle, M.S. & Scheraga, H.A. (1983) *J. Phys. Chem.* **87**, 1883–1887.

[99] Sippl, M.J., Némethy, G. & Scheraga, H.A. (1984) *J. Phys. Chem.* **88**, 6231–6233.

[100] Kawai, H., Okamoto, Y., Fukugita, M., Nakazawa, T. & Kikuchi, T. (1991) *Chem. Lett.* **1991**, 213–216.

[101] Okamoto, Y., Fukugita, M., Nakazawa, T. & Kawai, H. (1991) *Protein Eng.* **4**, 639–647.

[102] Hingerty, B.E., Ritchie, R.H., Ferrell, T. & Turner, J.E. (1985) *Biopolymers* **24**, 427–439.

[103] Ramstein, J. & Lavery, R. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7231–7235.

[104] Okamoto, Y. (1994) *Biopolymers* **34**, 529–539.

[105] Daggett, V., Kollman, P.A. & Kuntz, I.D. (1991) *Biopolymers* **31**, 285–304.

[106] Ooi, T., Oobatake, M., Némethy, G. & Scheraga, H.A. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3086–3090.

[107] Masuya, M., in preparation.

[108] Eisenhaber, F., Lijnzaad, P., Argos, P., Sander, C. & Scharf, M. (1995) *J. Comput. Chem.* **16**, 273–284.

[109] Chandler, D. & Andersen, H.C. (1972) *J. Chem. Phys.* **57**, 1930–1937.

[110] Hirata, F. & Rossky, P.J. (1981) *Chem. Phys. Lett.* **83**, 329–334.

[111] Perkyns, J.S. & Pettitt, B.M. (1992) *J. Chem. Phys.* **97**, 7656–7666.

[112] Berendsen, H.J.C., Grigera, J.R. & Straatsma, T.P. (1987) *J. Phys. Chem.* **91**, 6269–6271.

[113] Kinoshita, M., Okamoto, Y. & Hirata, F. (1997) *J. Comput. Chem.* **18**, 1320–1326.

[114] Kinoshita, M., Okamoto, Y. & Hirata, F. (1998) *J. Comput. Chem.* **19**, 1724–1735.

[115] Kinoshita, M., Okamoto, Y. & Hirata, F. (1997) *J. Chem. Phys.* **107**, 1586–1599.

[116] Kinoshita, M., Okamoto, Y. & Hirata, F. (1998) *J. Am. Chem. Soc.* **120**, 1855–1863.

[117] Weiner, S.J., Kollman, P.A., Nguyen, D.T. & Case, D.A. (1986) *J. Comput. Chem.* **7**, 230–252.

[118] Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M. Jr., Ferguson, D.M., Splellmeyer, D.C., Fox, T., Caldwell, J.W. & Kollman, P.A. (1995) *J. Am. Chem. Soc.* **117**, 5179–5197.

[119] Kollman, P., Dixon, R., Cornell, W., Fox, T. Chipot, C. & Pohorille, A. (1997) in *Computer Simulation of Biomolecular Systems Vol. 3*, van Gunsteren, W.F., Weiner, P.K. & Wilkinson, A.J., Eds., KLUWER/ESCOM, Dordrecht, pp. 83–96.

[120] Sugita Y. & Kitao, A. (1998) *Proteins* **30**, 388–400.

[121] Kitao, A., Hayward, S. & Gō, N. (1998) *Proteins* **33**, 496–517.

[122] Morikami, K., Nakai, T., Kidera, A., Saito, M. & Nakamura, H. (1992) *Comput. Chem.* **16**, 243–248.

[123] Hoover, W.G., Ladd, A.J.C. & Moran, B. (1982) *Phys. Rev. Lett.* **48**, 1818–1820.

[124] Evans, D.J. & Morris, G.P. (1983) *Phys. Lett.* **A98**, 433–436.

[125] Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. & Klein, M.L. (1982) *J. Chem. Phys.* **79**, 926–935.

[126] Myers, J.K., Pace, C.N. & Scholtz, J.M. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 2833–2837.

[127] Mitsutake, A. & Okamoto, Y. (2000), in preparation.

[128] Kraulis, P. J. (1991) *J. Appl. Cryst.* **24**, 946–950.

[129] Bacon, D. & Anderson, W. F. (1988) *J. Mol. Graphics*, **6**, 219–220.

[130] Merritt, E. A. & Murphy, M. E. P. (1994) *Acta Cryst.* **D50**, 869–873.

[131] Sayle, R.A. & Milner-White, E.J. (1995) *Trends Biochem. Sci.* **20**, 374–376.

[132] Li, Z. & Scheraga, H.A. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6611–6615.

[133] Meirovitch, H., Meirovitch, E., Michel, A.G. & Vásquez, M. (1994) *J. Phys. Chem.* **98**, 6241–6243.

[134] Mitsutake, A., Sugita, Y. & Okamoto, Y. (2000), in preparation.

[135] Graham, W.H., Carter, S.E. II & Hickes, P.R. (1992) *Biopolymers* **32**, 1755–1764.

[136] Sugita, Y. & Okamoto, Y. (2000), in preparation.

[137] Ono, S., Nakajima, N., Higo, J. & Nakamura, H. (2000) *J. Comput. Chem.* **21**, 748–762.

[138] Kinoshita, M., Okamoto, Y. & Hirata, F. (2000) *J. Am. Chem. Soc.* **122**, 2773–2779.