Replica-exchange multicanonical and multicanonical replica-exchange Monte Carlo simulations of peptides. II. Application to a more complex system

Ayori Mitsutake, Yuji Sugita, and Yuko Okamoto

Citation: J. Chem. Phys. **118**, 6676 (2003); doi: 10.1063/1.1555849
View online: https://doi.org/10.1063/1.1555849
View Table of Contents: http://aip.scitation.org/toc/jcp/118/14
Published by the American Institute of Physics

Replica-exchange multicanonical and multicanonical replica-exchange Monte Carlo simulations of peptides. II. Application to a more complex system

Ayori Mitsutake[a)]
Department of Physics, Faculty of Science and Technology, Keio University, Yokohama, Kanagawa 223-8522, Japan

Yuji Sugita[b)] and Yuko Okamoto[c)]
Department of Theoretical Studies, Institute for Molecular Science, Okazaki, Aichi 444-8585, Japan and Department of Functional Molecular Science, The Graduate University for Advanced Studies, Okazaki, Aichi 444-8585, Japan

(Received 22 November 2002; accepted 6 January 2003)

In Paper I of this series the formulations of the replica-exchange multicanonical algorithm and the multicanonical replica-exchange method for Monte Carlo versions have been presented. The effectiveness of these algorithms were then tested with the system of a penta peptide, Met-enkephalin, in the gas phase. In this article the detailed comparisons of performances of these algorithms together with the regular replica-exchange method are made, taking a more complex system of a 17-residue helical peptide. It is shown that these two new algorithms are more efficient than the regular replica-exchange method. © *2003 American Institute of Physics.*
[DOI: 10.1063/1.1555849]

I. INTRODUCTION

In protein folding simulations, powerful simulation algorithms are indispensable because conventional constant temperature simulations at low temperatures will get trapped in states of energy local minima. Generalized-ensemble algorithms such as the multicanonical algorithm (MUCA)[1,2] and the replica-exchange method (REM)[3–5] are now commonly used to overcome the multiple-minima problem (for recent reviews, see Refs. 6–8). Recently, two new generalized-ensemble algorithms have been developed for molecular dynamics simulations, which combines the merits of MUCA and REM.[9] In the first method that we refer to as the replica-exchange multicanonical algorithm (REMUCA), the multicanonical weight factor is determined from a short REM simulation with the multiple-histogram reweighting techniques (WHAM).[10,11] A long multicanonical production run with high statistics is then performed with this weight factor. The process of determining the multicanonical weight factor in this method is faster and simpler than that in the usual iterative determination. In the second method that we refer to as the multicanonical replica-exchange method (MUCAREM), a REM simulation with replicas corresponding to multicanonical ensembles with different energy ranges is performed.

In the preceding paper (which we refer to as Paper I), the formulations of the REMUCA and MUCAREM algorithms for the Monte Carlo version were explained in detail. These algorithms were tested with the benchmark system of a penta peptide, Met-enkephalin, in the gas phase.

In this article we present detailed comparisons of performance of these new algorithms as well as regular REM. For this purpose we have performed the simulations of a more complex system of a 17-residue helical peptide.

The multicanonical weight factors that are determined in REMUCA and MUCAREM simulations for more complex systems sometimes realize random walk in energy space but are not good enough to realize the flat probability distributions of energy to obtain accurate thermodynamic quantities. We need to refine these weight factors to estimate the correct thermodynamic quantities. We explain how to obtain the refined multicanonical weight factors by iterative applications of REMUCA and/or MUCAREM simulations.

II. METHODS

The details of the methods, MUCA, REM, REMUCA, and MUCAREM, have been presented in Paper I.

The multicanonical ensemble is based on the following non-Boltzmann weight factor $W_{mu}(E)$ so that the probability distribution of potential energy $P_{mu}(E)$ is flat and a one-dimensional random walk in potential energy space is realized:

$$P_{mu}(E) \propto n(E) W_{mu}(E) = \text{const}, \tag{1}$$

where $n(E)$ is the density of states. We then have

$$W_{mu}(E) \propto n^{-1}(E). \tag{2}$$

In REMUCA the multicanonical weight factor $W_{mu}(E)$ is first determined from the results of a short REM simulation by the multiple-histogram reweighting techniques [see Eqs. (23), (24), (29), and (30) in Paper I]. A multicanonical production run is then performed with this weight factor.

a)Electronic mail: ayori@rk.phys.keio.ac.jp
b)Present address: Department of Structural Biology, Institute of Molecular and Cellular Biosciences, University of Tokyo, Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan; Electronic mail: sugita@iam.u-tokyo.ac.jp
c)Electronic mail: okamotoy@ims.ac.jp

When no other estimate is available, the same multicanonical weight factor is also used in the MUCAREM simulations. In MUCAREM a replica-exchange production run with replicas corresponding to multicanonical ensembles with a different energy range is then performed.

The final results by both REMUCA and MUCAREM simulations therefore depend on the accuracy of the estimate of the multicanonical weight factor, or the density of states, that is obtained by the short REM simulation. For a simple benchmark system in Paper I (a penta peptide in the gas phase), an accurate estimate of the multicanonical weight factor was obtained by only one short REM simulation. For more complex systems, however, a single REM simulation will not be sufficient for the satisfactory multicanonical weight factor determination unless we make the REM simulation very long. In such cases we can iterate MUCA and/or MUCAREM simulations in which the estimate of the multicanonical weight factor is updated by the single- and/or multiple-histogram reweighting techniques, respectively. The MUCA production run (in REMUCA) or MUCAREM production run with the refined multicanonical weight factor will make a more frequent random walk in energy space and can yield a flatter probability distribution of potential energy.

To be more specific, this iterative process can be described as follows: The REMUCA production run corresponds to a MUCA simulation with the weight factor $W_{mu}(E)$ [see Eq. (25) in Paper I]. The new estimate of the density of states can be obtained by the single-histogram reweighting techniques[12] as follows [see Eq. (5) in Paper I]:

$$n(E)=\frac{N_{mu}(E)}{W_{mu}(E)}, \tag{3}$$

where $N_{mu}(E)$ is the energy histogram obtained from the MUCA production run in REMUCA.

On the other hand, from the MUCAREM production run, the improved density of states can be obtained by the multiple-histogram reweighting techniques as follows [see Eqs. (41) and (42) in Paper I]:[10,11]

$$n(E)=\frac{\sum_{m=1}^{\mathcal{M}} N_{MR}^{\{m\}}(E)}{\sum_{m=1}^{\mathcal{M}} n_m \exp(f_m) W_{MR}^{\{m\}}(E)}, \tag{4}$$

and

$$\exp(-f_m)\equiv\sum_E n(E) W_{MR}^{\{m\}}(E). \tag{5}$$

Here, $N_{MR}^{\{m\}}(E)$ and n_m are the energy histogram and the total number of samples obtained for the mth multicanonical ensemble (or energy range) in the MUCAREM production run, respectively, and $W_{MR}^{\{m\}}(E)$ is the corresponding multicanonical weight factor [see Eq. (35) in Paper I].

The improved density of states thus obtained [either from Eq. (3) or from Eq. (4)] leads to a new multicanonical weight factor [see Eq. (2)]. The next iteration can be either a MUCA production run (as in REMUCA) or a MUCAREM production run with the weight factors obtained from Eqs. (26), (29), and (30) or Eqs. (34), (35), (37), and (38) in Paper I, respectively. The results of this production run may yield an optimal multicanonical weight factor that yields a sufficiently flat energy distribution. If not, we can repeat the above process by obtaining the third estimate of the multicanonical weight factor either from Eq. (3) or from Eq. (4), and so on.

We remark that as the estimate of the multicanonical weight factor becomes more accurate, one is required to have a fewer number of replicas for a successful MUCAREM simulation, because each replica will have a flat energy distribution for a wider energy range. Hence, for a large, complex system, it is often more efficient to first try MUCAREM and iteratively reduce the number of replicas so that eventually one needs only one or a few replicas (instead of trying REMUCA directly from the beginning and iterating MUCA simulations).

III. RESULTS AND DISCUSSION

A. Computational details

We performed the REM, REMUCA, and MUCAREM simulations with the system of a 17-residue fragment of ribonuclease T1. It is known by experiments that this peptide fragment tends to form α-helical conformations.[13] The amino acid sequence is SSDVSTAQIAAYKLHED, which is the part from residue Ser-13 through Asp-29 with the mutations A21I and G23A (these are underlined above) from the native sequence.[13] We first tried to obtain an optimal multicanonical weight factor by the usual iterative process.[14,15] However, we failed. Thus, we think that this is a good system to test the effectiveness of REMUCA and MUCAREM.

The energy function E_{TOT} that we used is the sum of the conformational energy term of the solute E_P and the solvation free energy term E_{SOL} for the interaction of the peptide with the surrounding solvent: $E_{\mathrm{TOT}}=E_P+E_{\mathrm{SOL}}$ (we remark that in Paper I we only used the conformational term E_P as the total energy). Here, the solvation term E_{SOL} is the sum of terms that are proportional to the solvent-accessible surface area of the atomic groups of the solute. The parameters in the conformational energy as well as the molecular geometry were taken from ECEPP/2.[16–18] The parameters of the solvent term were adopted from Ref. 19. For the total energy histograms we took the bin size of $\epsilon=1$ kcal/mol. The computer code KONF 90[20,21] was used, and MC simulations based on the REM, REMUCA, and MUCAREM were performed. For the calculation of a solvent-accessible surface area, we used the computer code NSOL,[22] which is a modification of the code NSC.[23]

The dihedral angles ϕ and ψ in the main chain and χ in the side chains constituted the variables to be updated in the simulations as in Paper I. The number of degrees of freedom for the peptide is 80 (it was 19 for Met-enkephalin in Paper I). One MC sweep consists of updating all these angles once with Metropolis evaluation for each update. The simulations were started from randomly generated conformations. The CPU time on Origin3800 and NEC SX-5 was, respectively, about 320 and 60 min to run 10 000 MC sweeps for each replica.

B. REMUCA and MUCAREM simulations

In Table I we summarize the parameters of the simulations that were performed in the present work. As discussed in detail in Paper I, 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 I. A production run of the MUCAREM simulation is referred to as MUCAREM1 and it uses the same multicanonical weight factor that was obtained from REM1. Moreover, a long production run of the original REM simulation was also performed for comparison and it is referred to as REM2 in Table I. As explained in the previous section, the results of REMUCA and MUCAREM simulations can be used to obtain better estimates of the multicanonical weight factor than that used in these REMUCA and MUCAREM simulations [see Eqs. (3) and (4)]. The production MUCA runs with these weights are referred to as MUCA2 and MUCA3 in Table I, respectively. The total number of MC sweeps for the five production runs (REM2, MUCA1, MUCAREM1, MUCA2, and MUCA3) was set essentially equal (i.e., about 4 000 000 MC sweeps). Before taking the data, we made equilibration simulations of 10 000 MC sweeps (per replica) for all the runs. Replica exchange was tried every 25 MC sweeps for REM1 and REM2 and every 50 MC sweeps for MUCAREM1.

We first check whether the REM simulations indeed performed properly. In REM1 and REM2 there exist 14 replicas with 14 different temperatures, ranging from 200 to 700 K, as listed in Table I (i.e., $T_1=200$ K and $T_M=T_{14}=700$ K). The temperatures are distributed exponentially between T_1 and T_M, following the optimal distribution found in Ref. 24. In Fig. 1 the canonical probability distributions obtained at the chosen 14 temperatures from REM1 are shown. We see that there are enough overlaps between all pairs of neighboring distributions, indicating that there will be sufficient numbers of replica exchange between pairs of replicas. For an optimal performance of REM simulations, the acceptance ratios of replica exchange should be sufficiently uniform and large (say, $>10\%$). In Table II 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 14% and 30%).

TABLE I. Summary of parameters in REM, REMUCA, and MUCAREM simulations.

Run	No. of replicas, M	Temperature, T_m (K) ($m=1,...,M$)	MC sweeps
REM1	14	200, 220, 243, 267, 294, 324, 357, 393, 432, 476, 524, 577, 636, 700	60 000[a]
REM2	14	200, 220, 243, 267, 294, 324, 357, 393, 432, 476, 524, 577, 636, 700	290 000[b]
MUCA1	1		4 000 000
MUCAREM1	4		1 000 000[c]
MUCA2	1		4 000 000
MUCA3	1		4 000 000

[a]The total number of MC sweeps is 840 000.
[b]The total number of MC sweeps is 4 060 000.
[c]The total number of MC sweeps is 4 000 000.

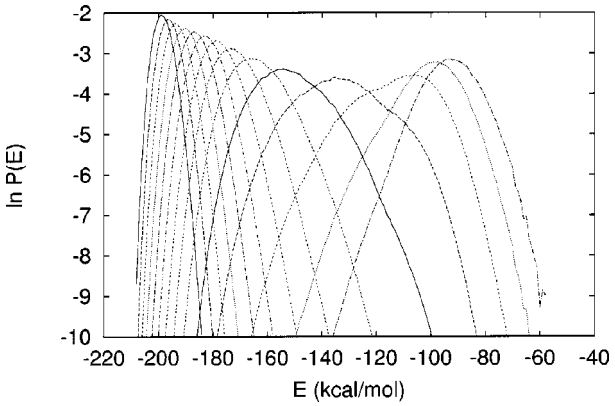

FIG. 1. Probability distribution of total energy obtained from REM1 (see Table I for the parameters of the simulation).

In Fig. 2 the time series of total energy for each replica in REM2 are shown (the results for REM1 essentially correspond to the first 60 000 MC sweeps of these figures). We see random walks in energy space, as expected. The behaviors are, however, quite different among the replicas; some random walks are more conspicuous than others. The trajectories as a whole cover a wide energy space between low-temperature regions and high-temperature ones. We remark that the multicanonical weight factor obtained from a short replica-exchange simulation is stable without much oscillations, which are common in the usual iterative process, because we can use the information for a wide temperature range. We next compare the performances of the usual canonical MC simulation and REM simulation. The time series of total energy from the canonical MC simulations (left-hand side) and REM2 (right-hand side) at temperatures 294 K (a), 524 K (b), and 700 K (c) are shown in Fig. 3. The results for REM2 are from different replicas when they are at the corresponding temperatures. At the low temperature (294 K), the energy in the usual canonical MC simulation fluctuates around -165 kcal/mol, while that in the REM simulation at 294 K fluctuates around lower energy (-185 kcal/mol). This suggests that the conventional canonical MC simulation got trapped in a state of energy local minimum. We also see that the results are similar at the intermediate temperature (524 K) and almost identical at the high temperature (700 K).

TABLE II. Acceptance ratios of a replica exchange in REM1.

Pair of temperature (K)	Acceptance ratio
200↔220	0.29
220↔243	0.30
243↔267	0.29
267↔294	0.29
294↔324	0.29
324↔357	0.29
357↔393	0.28
393↔432	0.25
432↔476	0.21
476↔524	0.19
524↔577	0.14
577↔636	0.25
636↔700	0.30

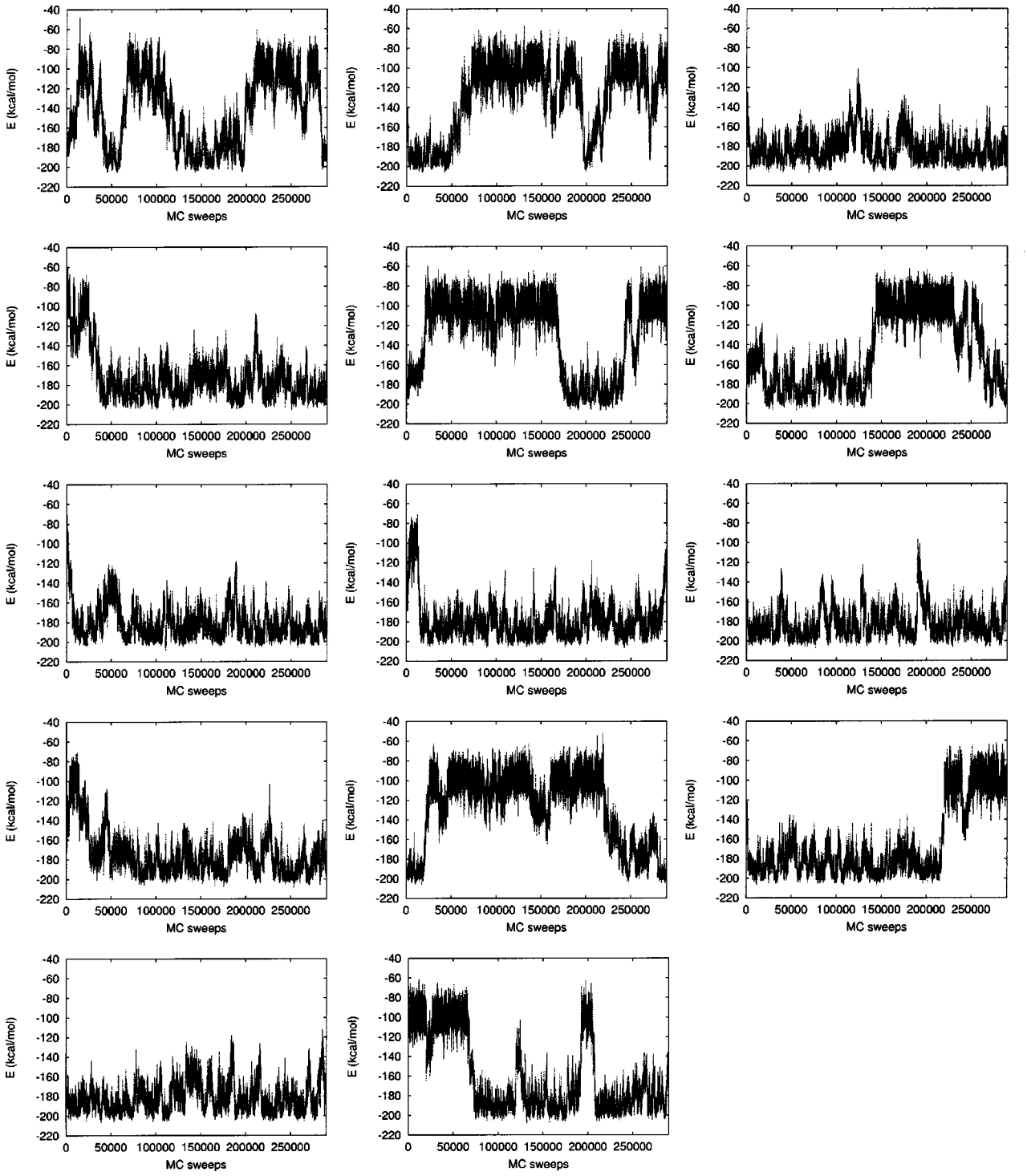

FIG. 2. Time series of total energy for all the replicas in REM2 (see Table I for the parameters of the simulation).

Typical snapshots from these canonical MC simulations and REM2 at the corresponding temperatures are shown in Fig. 4. At the low temperature (294 K), an ideal helix structure (global-minimum-energy state) is dominant for REM2, whereas partially unfolded helical structures corresponding to local-minimum-energy states are obtained for the canonical simulation. At the intermediate temperature (524 K), the helical conformations and random-coil ones coexist in both simulations (it turns out that this temperature corresponds to the helix–coil transition temperature, as discussed below). Finally, at a high temperature (700 K), the conformations of the two simulations are random coils, as expected.

We now present the results of our REMUCA simulations. Before the REMUCA simulation, we have also tried to determine the multicanonical weight factor of this peptide system by the regular iterative process.[15] We made 40 iterations of short trial multicanonical simulations of 50 000 MC sweeps (the total number of MC sweeps is 2 000 000). How-

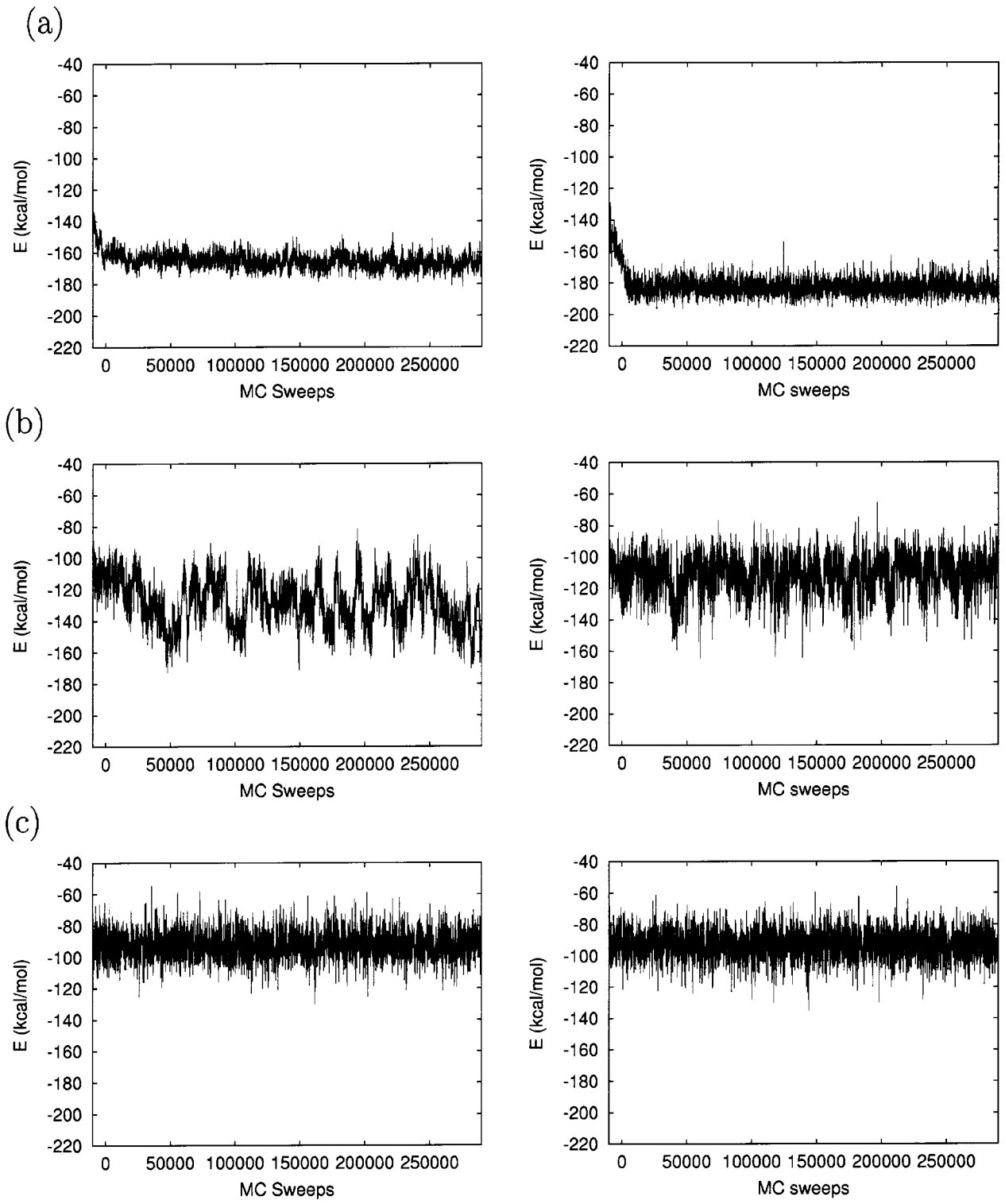

FIG. 3. Time series of total energy from the canonical MC simulations (the left-hand side) and REM2 (the right-hand side) at temperatures 294 (a), 524 (b), and 700 K (c). The negative values of MC sweeps correspond to the part of equilibration.

ever, we failed in obtaining a satisfactory multicanonical weight factor in the sense that we could not get a random walk in a wide energy range. In Fig. 5(a) we show the time series of a production run with this incomplete multicanonical weight factor. We can see that the simulation got trapped in a state of energy local minima and did not realize a random walk in a wide energy range.

The REMUCA simulations consist of a short REM simulation (REM1 in Table I) and a MUCA production run (MUCA1 in Table I). After REM1, we obtained the density of states, $n(E)$, by the multiple-histogram reweighting techniques [see Eqs. (23) and (24) in Paper I]. We also estimated E_1 and E_M that are the canonical expectation values of the total potential energy at temperatures T_1 and T_M, respec-

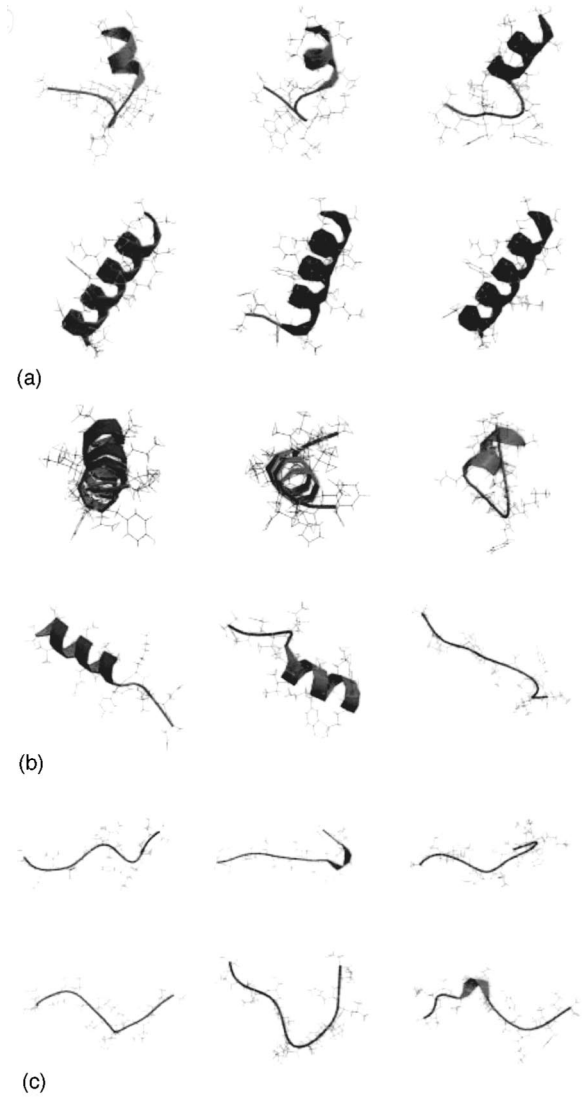

FIG. 4. Typical snapshots from the canonical MC simulations (above) and REM2 (below) at temperatures 294 K (a), 524 K (b), and 700 K (c). The snapshots for each simulation on the left-hand side, middle, and right-hand side are the conformations at 50 000, 150 000, and 250 000 MC sweeps, respectively. These figures were created with MOLSCRIPT (Ref. 25) and RASTER3D (Refs. 26 and 27).

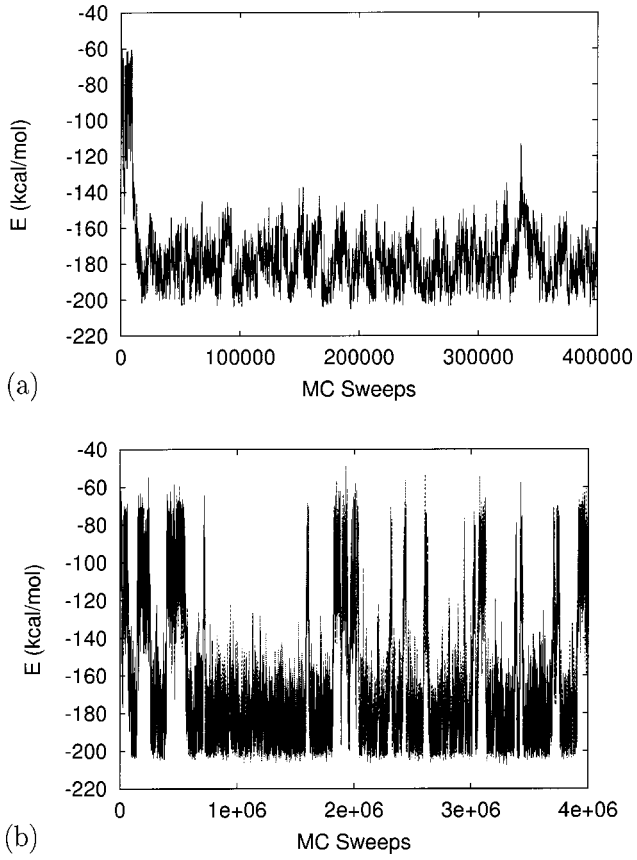

FIG. 5. The time series of total energy from the MUCA production runs with the multicanonical weight factors that were determined by the usual iterative process (a) and REMUCA (b). The total numbers of MC sweeps in (a) and (b) are 400 000 and 4 000 000, respectively.

tively ($T_1=200$ K and $T_M=700$ K are the lowest and the highest temperatures that were used in REM1). The results of REM1 gave

$$E_1=-199 \text{ kcal/mol}, \quad E_M=-93 \text{ kcal/mol}. \tag{6}$$

As explained in Paper I, the obtained density of states is reliable only in the energy range $E_1 \leqslant E \leqslant E_M$. The multicanonical weight factor was then determined for the three energy regions ($E<E_1$, $E_1 \leqslant E \leqslant E_M$, and $E>E_M$) from Eqs. (26), (29), and (30) in Paper I. We carried out a MUCA production run of 4 000 000 MC sweeps with this multicanonical weight factor (MUCA1 in Table I). In Fig. 5(b) we show the time series of MUCA1. We now observe a random walk in energy space. The simulation still has a tendency to stay in the low-energy region, but we see an improvement in behavior from the previous method [compare Figs. 5(a) and 5(b)]. Note that the required MC sweeps for the multicanonical weight factor determination for REMUCA was 860 000 (see Table I), while that for the previous iterative method was 2 000 000. Four snapshots of conformations during MUCA1 are shown in Fig. 6. We observe a large conformational changes during the simulation. The low-energy conformations are in almost ideal helix state [see Figs. 6(a) and 6(c)], and the high-energy conformations are random coils [see Figs. 6(b) and 6(d)]. Thus, a random walk in conformational space is also realized. In Fig. 7 the probability distribution obtained by MUCA1 is plotted. It can be seen that a flat distribution is obtained in the energy region $E_1 \leqslant E \leqslant E_M$ (although there is some weight toward the low-energy region). In Fig. 7, the canonical probability distributions that were obtained by the reweighting techniques at $T=T_1=200$ K, $T=T_{11}=524$ K, and $T=T_M=700$ K are also shown. We find 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 $E_1 \leqslant E \leqslant E_M$ [see also Eq. (6)].

We now present the results of our MUCAREM simulation (MUCAREM1 in Table I). From the density of states obtained by REM1, we prepared the multicanonical weight factors for the MUCAREM simulation [see Eqs. (34)–(38) in Paper I]. We used four replicas to do this simulation. The parameters of MUCAREM1 such as energy bounds $E_L^{\{m\}}$ and $E_H^{\{m\}}$ ($m=1,...,4$) are listed in Table III. The choices of $E_L^{\{m\}}$ and $E_H^{\{m\}}$ are, in general, arbitrary, but significant overlaps

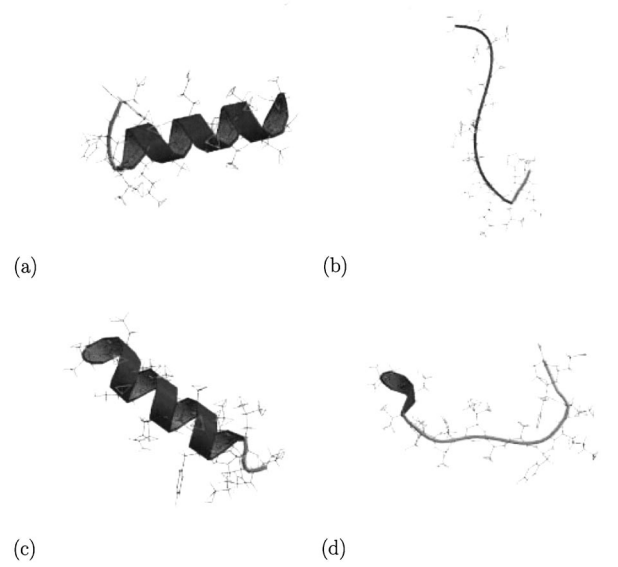

FIG. 6. Typical snapshots from MUCA1 (see Table I for the parameters of the simulation). The snapshots of (a), (b), (c), and (d) correspond to the conformations at 100 000, 200 000, 300 000, and 400 000 MC sweep, respectively. The conformations of (a) and (c) are low-energy conformations. On the other hand, the conformations of (b) and (d) are high-energy structures [see Fig. 5(b)]. The figures were created with MOLSCRIPT (Ref. 25) and RASTER3D (Refs. 26 and 27).

between the probability distributions of adjacent replicas are necessary. In Fig. 8(a) we show the time series of total energy in MUCAREM1 for each replica. We observe random walks in energy space covering the energy range between $E_L^{\{1\}}$ and $E_H^{\{4\}}$ (see Table III). In Fig. 8(b) the probability distributions of total energy obtained by this simulation are shown. We observe that the probability distributions corresponding to the mth multicanonical ensemble is essentially flat for the energy region $E_L^{\{m\}} \leqslant E \leqslant E_H^{\{m\}}$, is of the canonical simulations at $T = T_L^{\{m\}}$ for $E < E_L^{\{m\}}$, and is of the canonical simulation at $T = T_H^{\{m\}}$ for $E > E_H^{\{m\}}$ $(m = 1,...,4)$. Note that each distribution in MUCAREM1 is much broader than that in REM1 [compare Figs. 1 and 8(b)]. The number of required replicas is thus smaller than in regular REM simulations (4 in MUCAREM1 versus 14 in REM1 and REM2).

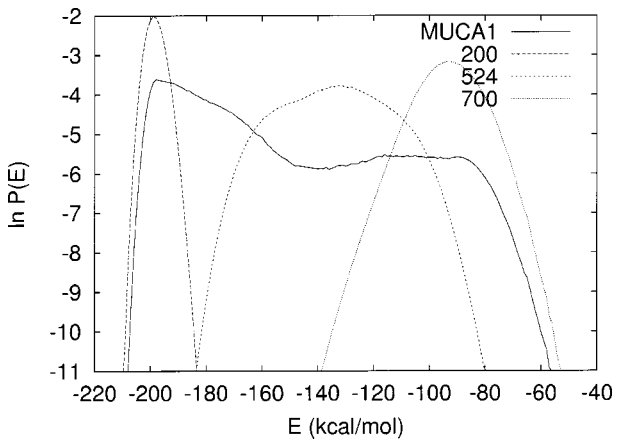

FIG. 7. Probability distributions of total energy obtained from MUCA1. Also added are the probability distributions (dotted curves) of the reweighted canonical ensemble at $T = 200$ K (left), 524 K (middle), and 700 K (right).

TABLE III. Summary of parameters in MUCAREM1.

m	$T_L^{\{m\}}$	$T_H^{\{m\}}$	$E_L^{\{m\}}$ (kcal/mol)	$E_H^{\{m\}}$ (kcal/mol)
1	200	323	−198.0	−182.0
2	294	432	−186.0	−164.0
3	392	577	−172.0	−113.0
4	524	700	−133.0	−90.0

C. Iterations of REMUCA and MUCAREM simulations

In the previous section we presented the results of REMUCA and MUCAREM simulations. These simulations were based on the multicanonical weight factor that was determined from the results of a single short REM simulation of 840 000 MC sweeps (namely, REM1 in Table I). We did observe random walks in energy space. However, the obtained probability distributions were not completely flat [see Figs. 7 and 8(b)]. This suggests that the above multicanonical weight factor needs some refinement. As explained in Sec. II, we can iterate the REMUCA and MUCAREM production runs to improve the multicanonical weight factor.

We recalculated the multicanonical weight factor by Eq. (3) from the results of the first 840 000 MC sweeps of the REMUCA production run (MUCA1) and performed a MUCA production run of 4 000 000 MC sweeps with this weight factor (this run is referred to as MUCA2 in Table I). Likewise, we recalculated the multicanonical weight factor by Eq. (4) from the results of the first 210 000 MC sweeps (per replica) of the MUCAREM production run (MUCAREM1) (the total number of MC sweeps is again 840 000). We then performed a MUCA production run of 4 000 000 MC sweeps with this weight factor (this run is referred to as MUCA3 in Table I).

In Fig. 9 the time series and probability distribution of the total energy from MUCA2 are shown. We do observe a great improvement in the sense that the results of MUCA2 give a more uniform random walk in energy space and a flatter energy distribution than those of MUCA1 [compare Fig. 9 with Figs. 5(b) and 7].

The time series and probability distribution of the total energy from MUCA3 are shown in Fig. 10. A similar improvement of the multicanonical weight factor from MUCA1 to MUCA3 is achieved [compare Fig. 10 with Figs. 5(b) and 7].

To further test the efficiency of these algorithms, we calculated the number of tunneling events during REM2, MUCA1, MUCAREM1, MUCA2, and MUCA3. One tunneling event is defined by a trajectory that goes from E_M to E_1 and back. If E_M is sufficiently high, the trajectory of a simulation gets completely uncorrelated when it reaches E_M. We thus consider that the more tunneling events we observe during a fixed number of MC sweeps, the more efficient the method is as a generalized-ensemble algorithm. In Table IV we summarize the number of tunneling events that were observed up to 2 000 000, 3 000 000, and 4 000 000 MC sweeps in REM2, MUCA1, MUCAREM1, MUCA2, and MUCA3.

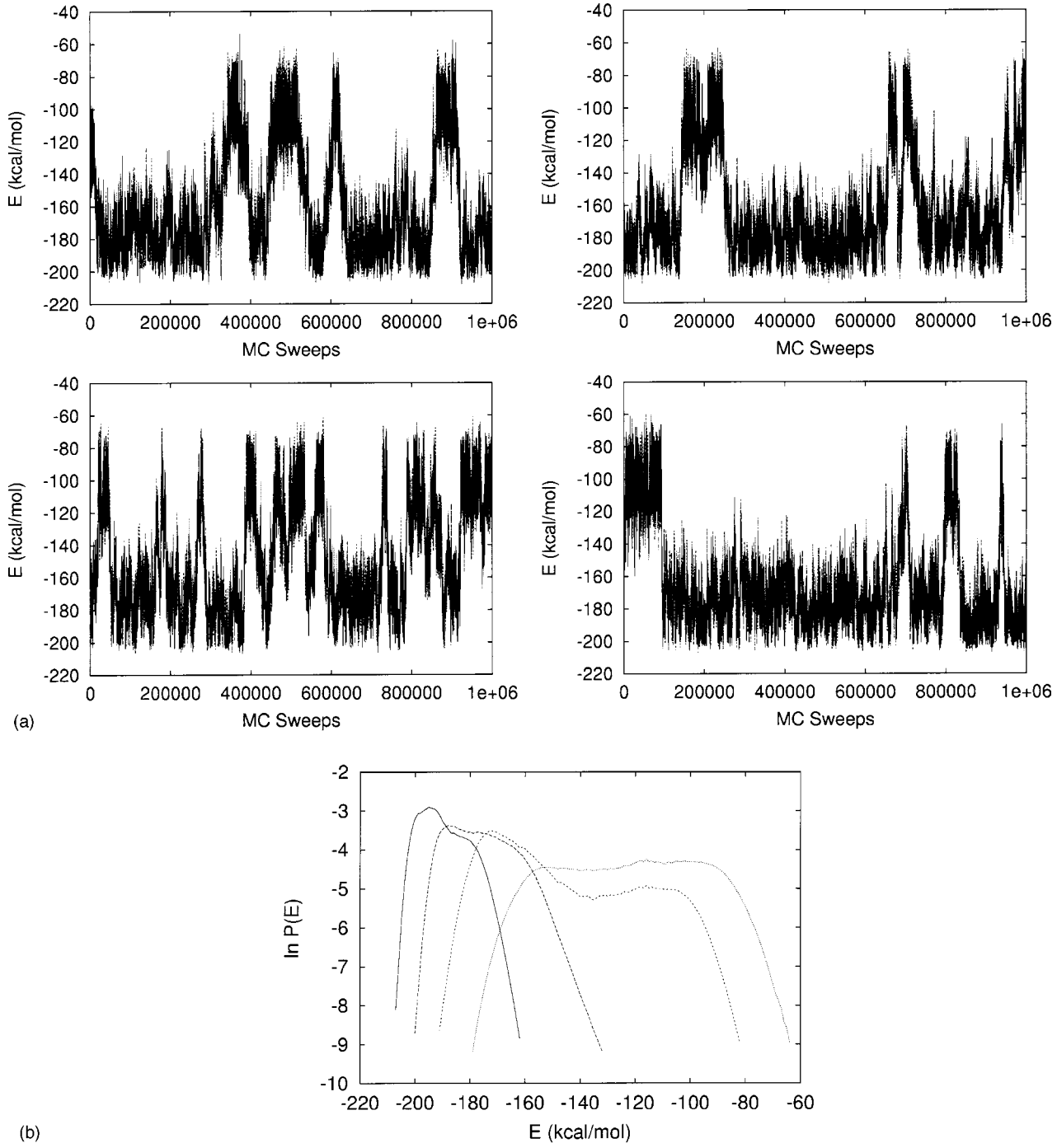

FIG. 8. Time series (a) and probability distributions (b) of total energy from MUCAREM1 (see Tables I and III for the parameters of the simulation).

In REM2 and MUCAREM1, we added the number of tunneling events of each replica to get the total number of tunneling events. The number of tunneling events of REM2 is the least among the five, suggesting that REMUCA and MUCAREM are more efficient in configurational sampling than REM. Among the remaining four production runs, the numbers of tunneling events of MUCA2 and MUCA3 are significantly more than those of MUCA1 and MUCAREM1. This means that the multicanonical weight factors that were used in MUCA2 and MUCA3 are much better than that used in MUCA1 (and MUCAREM1). Note that the multicanonical weight factor of MUCA1 was obtained from the results of a REM simulation of 840 000 MC sweeps in total (REM1), while those of MUCA2 and MUCA3 were obtained from the results of a MUCA simulation (MUCA1) and a MUCAREM simulation (MUCAREM1), respectively, of 840 000 MC sweeps in total. Simulations with more tunneling events will thus yield better estimates of the multicanonical weight factor. Finally, one can say that MUCAREM1 gave a slightly larger number of tunneling events than MUCA1. This means that when the estimate of the multicanonical weight factor is not good, the process of replica exchange enhances a random walk in energy space.

Once the results of these production runs are obtained,

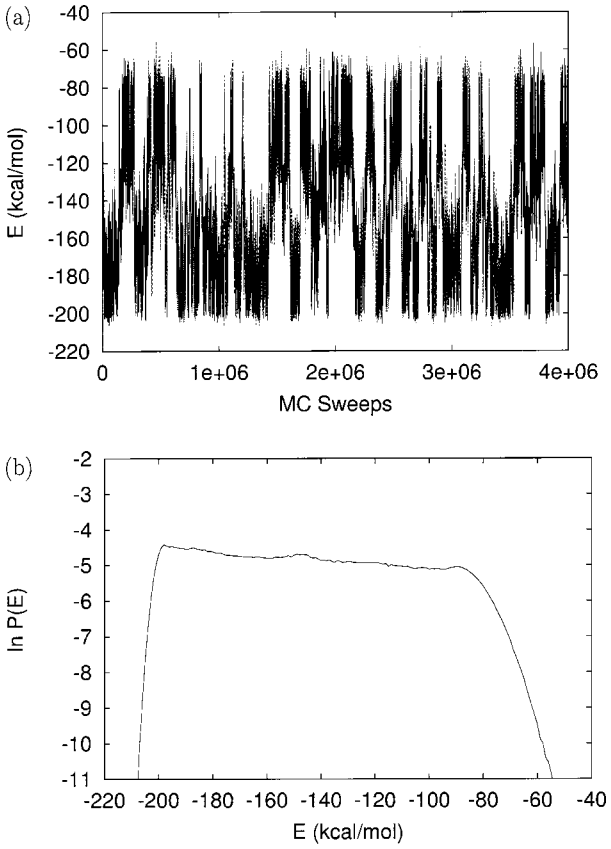

FIG. 9. Time series (a) and probability distribution (b) of total energy from MUCA2.

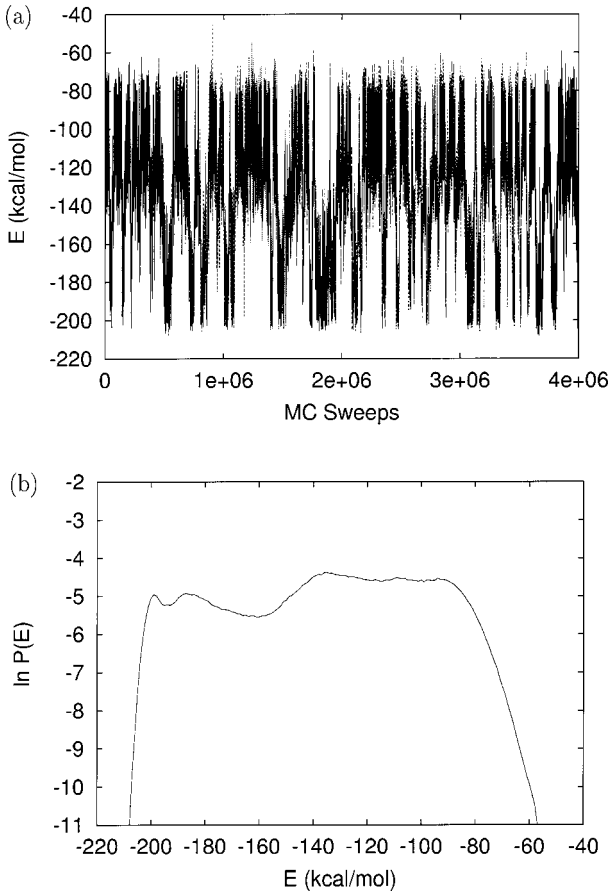

FIG. 10. Time series (a) and probability distribution (b) of total energy from MUCA3.

one can calculate ensemble averages of physical quantities as functions of temperature by the reweighting techniques [see Eqs. (3), (5), (23), and (41) in Paper I]. The quality of the estimate of the multicanonical weight factor can also be examined by calculating the ensemble averages of physical quantities.

The average total energy as a function of temperature that was obtained from the five production runs (namely, REM2, MUCA1, MUCAREM1, MUCA2, and MUCA3) are shown in Fig. 11. Each figure is a superposition of four results that correspond to the trajectories up to 1 000 000, 2 000 000, 3 000 000, 4 000 000 MC sweeps, which allows us to check the convergence of the results. We see that all the results are almost identical and converged rather quickly, except for those from REM2, which had the least number of tunneling events (see Table IV).

The average total energies increase monotonically as temperature increases. The drastic change around 500 K suggests the existence of a certain phase transition. In order to obtain the transition temperature, we calculated the specific heat as a function of temperature. The specific heat here is defined by the following equation:

$$C(T)=\beta^2\frac{\langle E^2\rangle_T-\langle E\rangle_T^2}{N}, \tag{7}$$

where N $(=17)$ is the number of residues in this peptide. The results from the five production runs are shown in Fig. 12. The peak of the specific heat gives the transition temperature T_c, and we find $T_c\approx 510$ K. We see some variations in the results among the production runs, which reflects the difficulty in the computation of this quantity. The height of the specific heat in REM2 up to 1 000 000 MC sweeps is much lower than those up to 2 000 000, 3 000 000, and 4 000 000 MC sweeps. This discrepancy can be understood because the number of tunneling events in REM2 is only one before 1 000 000 MC sweeps (data not shown).

The average number of helical residues, or helicity, as a function of temperature ($\langle n\rangle_T$) that were obtained from the five production runs are shown in Fig. 13. Here, we consider that a residue is in the α-helix configuration when the dihedral angles (ϕ, ψ) fall in the range $(-70\pm 30°, -37\pm 30°)$. The nature of the convergence of the results is similar to that of the average total energy (see Fig. 11). The average helicity tends to decrease monotonically as the temperature increases because of the increased thermal fluctua-

TABLE IV. Number of tunneling events in REM2, MUCA1, MUCAREM1, MUCA2, and MUCA3.

MC sweeps	REM2	MUCA1	MUCAREM1	MUCA2	MUCA3
2 000 000	2	7	9	16	18
3 000 000	5	12	16	23	29
4 000 000	9	17	22	32	38

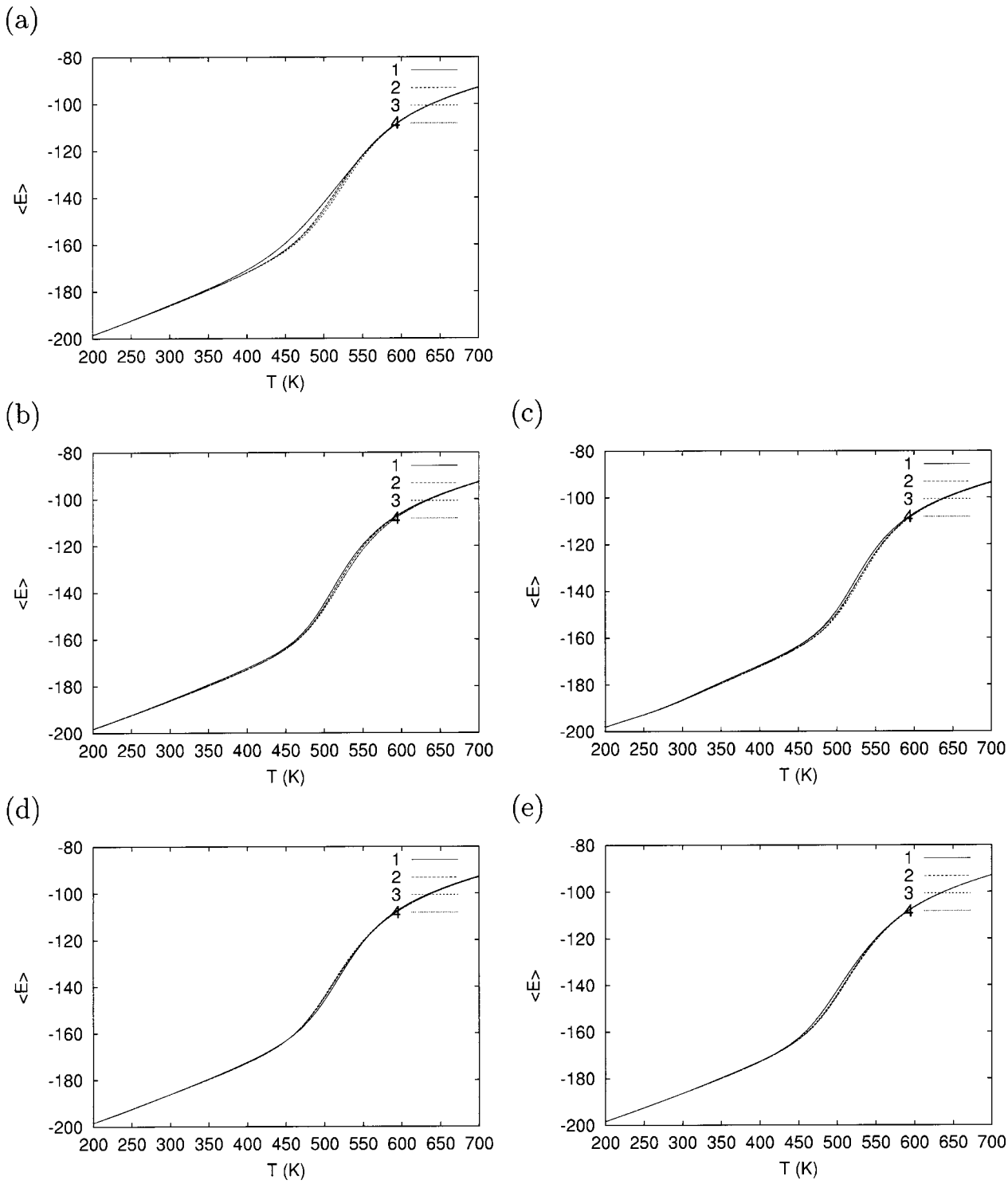

FIG. 11. Averages total energy as a function of temperature in REM2 (a), MUCA1 (b), MUCAREM1 (c), MUCA2 (d), and MUCA3 (e). The labels 1, 2, 3, and 4 in the legend correspond to the results that were calculated from the trajectories up to 1 000 000, 2 000 000, 3 000 000, and 4 000 000 MC sweeps, respectively.

tions. At $T=200$ K, $\langle n\rangle_T$ was about 15. If we neglect the terminal residues, where the α helix tends to be frayed, $n=15$ corresponds to the maximal helicity, and the conformation can be considered completely helical. Then the peptide is in an ideal helix state at $T=200$ K [see Figs. 6(a) and 6(c)]. At the transition temperature (around 510 K), $\langle n\rangle_T$ is 8.5 (50% helicity). This implies that the phase transition observed above by the peak in specific heat is indeed a helix–coil transition between an ideal helix and a random coil.

Finally, the results of the above three thermodynamic quantities from the five production runs are superimposed and shown in Fig. 14. The results of the conventional canonical MC simulations at four temperatures ($T=200$, 294, 542, and 700 K) are also plotted. All the results from the five

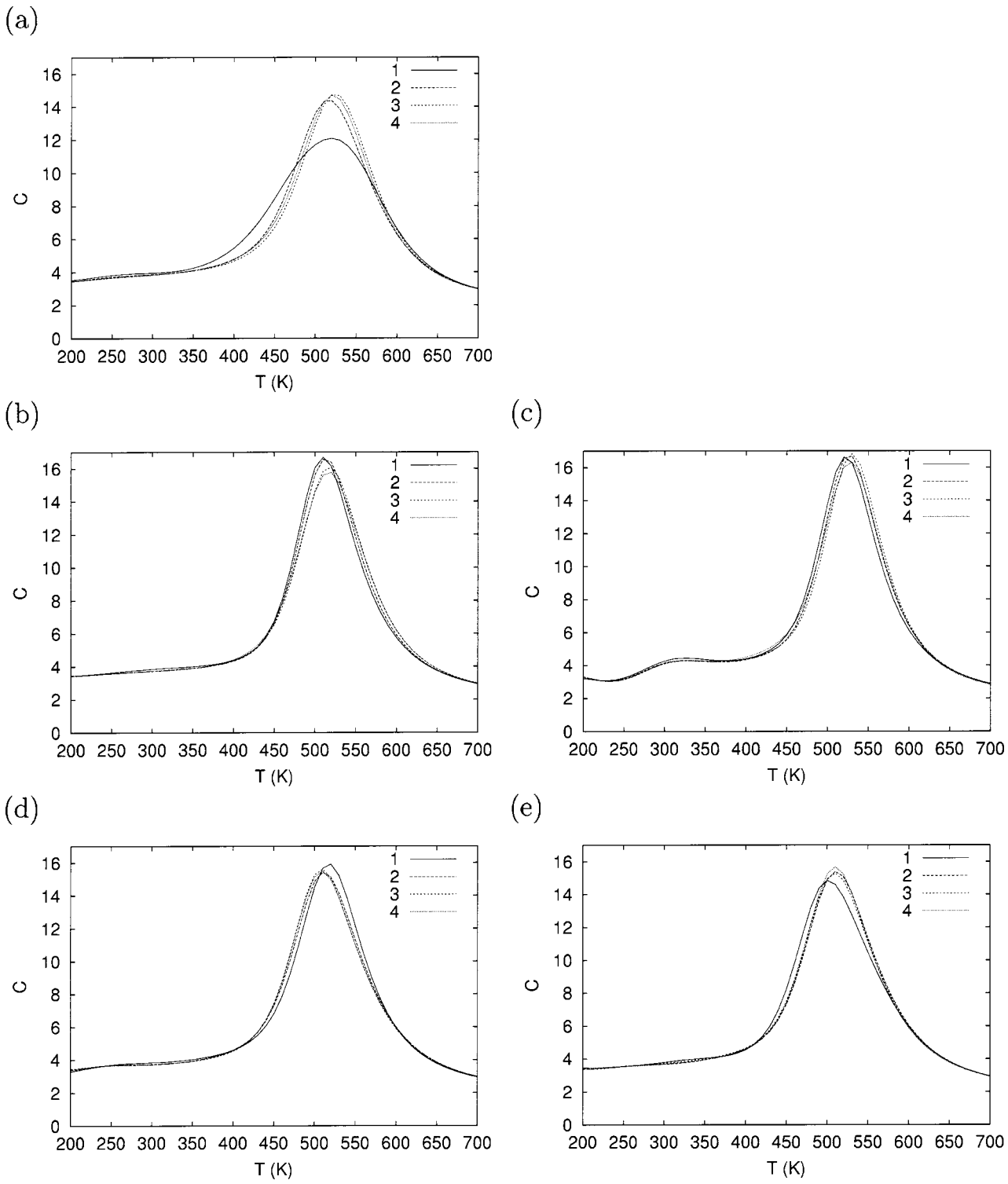

FIG. 12. Specific heat as a function of temperature in REM2 (a), MUCA1 (b), MUCAREM1 (c), MUCA2 (d), and MUCA3 (e). The labels 1, 2, 3, and 4 in the legend correspond to the results that were calculated from the trajectories up to 1 000 000, 2 000 000, 3 000 000, and 4 000 000 MC sweeps, respectively.

generalized-ensemble production runs are more or less similar, while those from the regular canonical simulations strongly deviate at lower temperatures (200 and 294 K). The canonical simulations apparently got trapped in states of energy local minima at these temperatures, as discussed above (see Figs. 3 and 4). Among the results of the five production runs, those of MUCA2 and MUCA3 are essentially identical (even for the specific heat). These two simulations had many tunneling events and gave the most accurate results.

IV. CONCLUSIONS

In the first article of the series of the present work, we have presented the details of the Monte Carlo version of the

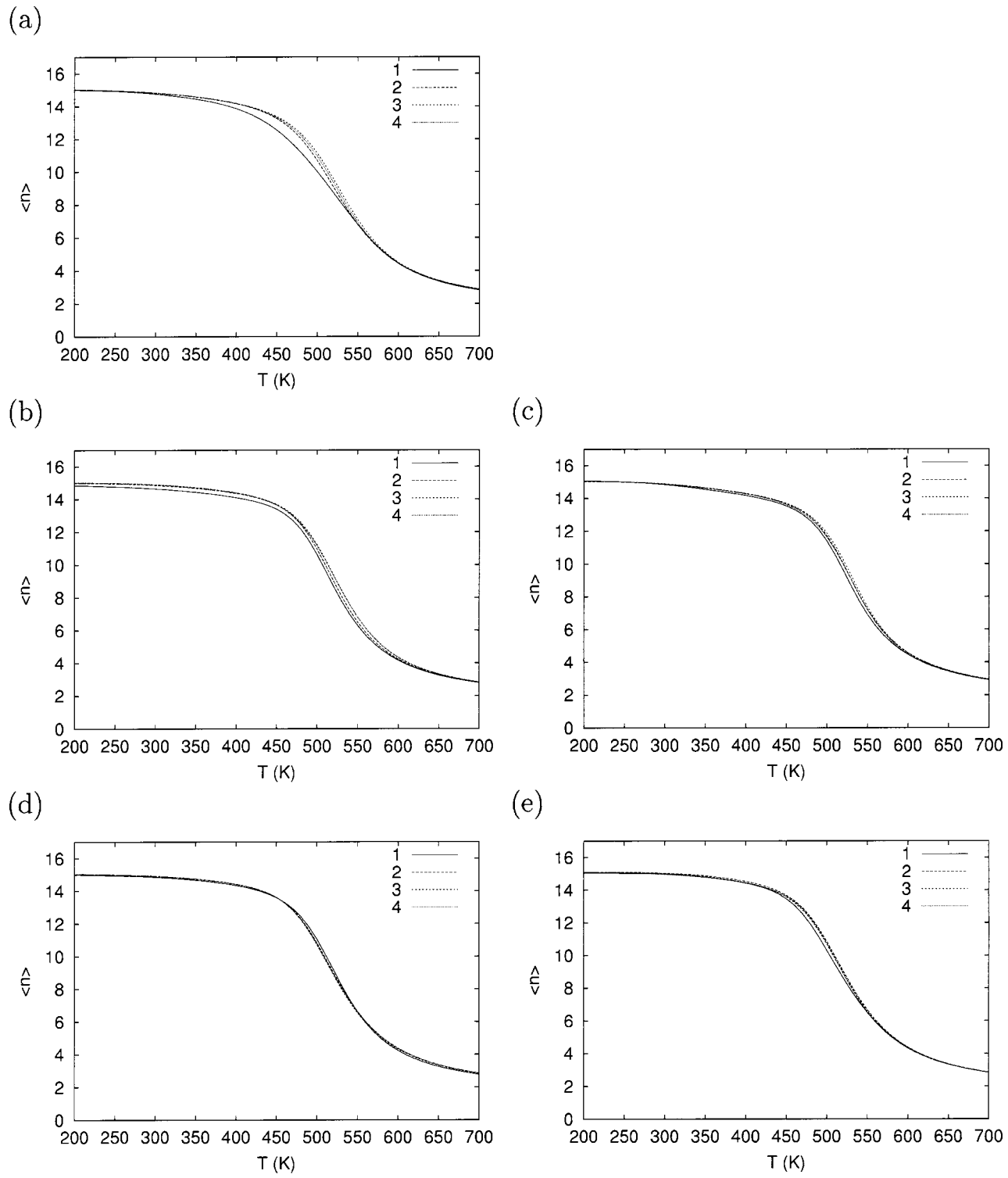

FIG. 13. Average helicity as a function of temperature in REM2 (a), MUCA1 (b), MUCAREM1 (c), MUCA2 (d), and MUCA3 (e). The labels 1, 2, 3, and 4 in the legend correspond to the results that were calculated from the trajectories up to 1 000 000, 2 000 000, 3 000 000, and 4 000 000 MC sweeps, respectively.

newly developed generalized-ensemble algorithms, replica-exchange multicanonical algorithm (REMUCA) and multicanonical replica-exchange method (MUCAREM). The effectiveness of these methods were tested with a penta peptide. In this article, which is the second of this series, we gave detailed comparisons of performances of these algorithms with a more complex system (a 17-residue helical peptide). It was shown that these new algorithms are much more efficient and effective than a well-known generalized-ensemble algorithm, namely a replica-exchange method. It turned out

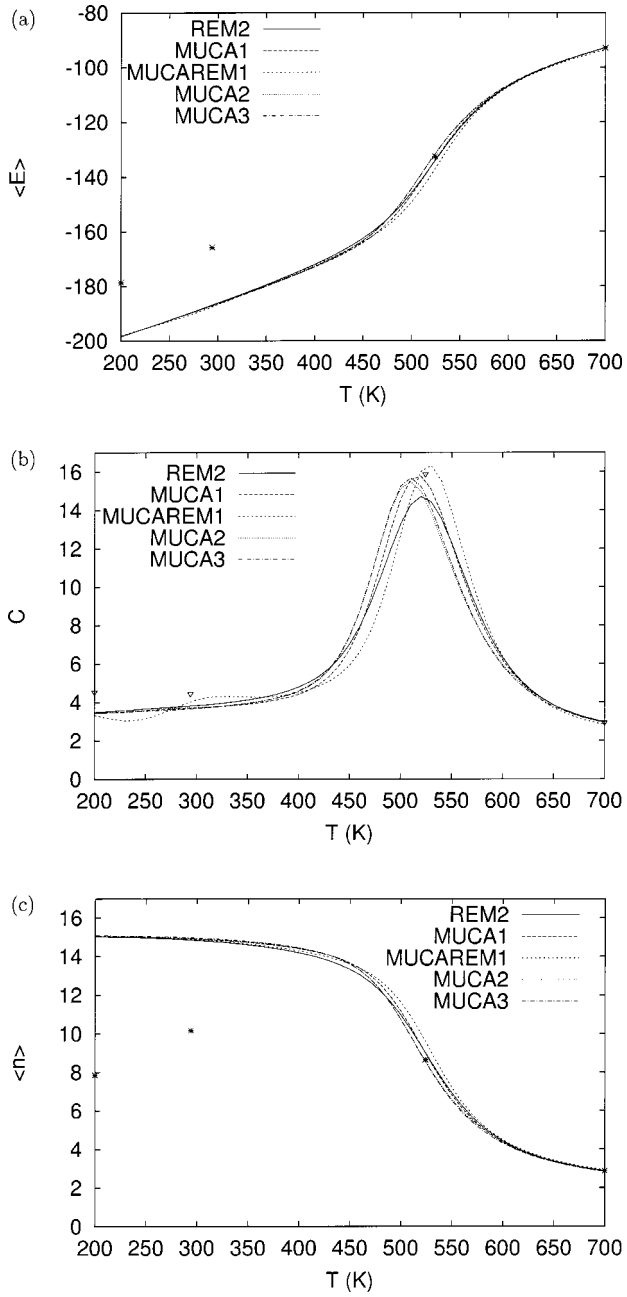

FIG. 14. Average total energy (a), specific heat (b), and average helicity as functions of the temperature in REM2, MUCA1, MUCAREM1, MUCA2, and MUCA3. The results were obtained from the simulations of 4 000 000 MC sweeps. The asterisks stand for the corresponding quantities obtained from the conventional canonical MC simulations of 300 000 MC sweeps at $T=200$, 294, 542, and 700 K.

that iterations of REMUCA and/or MUCAREM simulations further improve the performance. This iterative process presents one of the simplest and most efficient methods for the determination of the density of states of the system, or multicanonical weight factor.

ACKNOWLEDGMENTS

The simulations were performed on the HITACHI and other computers at the Research Center for Computational Science, Okazaki National Research Institutes. This work was supported, in part, by grants from the Research for the Future Program of the Japan Society for the Promotion of Science (JSPS-RFTF98P01101) and from the Japanese Ministry of Education, Culture, Sports, Science, and Technology.

[1] B. A. Berg and T. Neuhaus, Phys. Lett. B **267**, 249 (1991).
[2] B. A. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992).
[3] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).
[4] K. Hukushima, H. Takayama, and K. Nemoto, Int. J. Mod. Phys. C **7**, 337 (1996).
[5] C. J. Geyer, in Computing Science and Statistics, *Proceedings of the 23rd Symposium on the Interface*, edited by E. M. Keramidas (Interface Foundation, Fairfax Station, 1991), p. 156.
[6] U. H. E. Hansmann and Y. Okamoto, in *Annual Reviews of Computational Physics VI*, edited by D. Stauffer (World Scientific, Singapore, 1999), p. 129.
[7] A. Mitsutake, Y. Sugita, and Y. Okamoto, Biopolymers **60**, 96 (2001).
[8] Y. Sugita and Y. Okamoto, in *Lecture Notes in Computational Science and Engineering*, edited by T. Schlick and H. H. Gan (Springer-Verlag, Berlin, 2002), p. 304.
[9] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **329**, 261 (2000).
[10] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).
[11] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, J. Comput. Chem. **13**, 1011 (1992).
[12] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988).
[13] J. K. Myers, C. N. Pace, and J. M. Scholtz, Proc. Natl. Acad. Sci. U.S.A. **94**, 2833 (1997).
[14] B. A. Berg, Fields Inst. Commun. **26**, 1 (2000); also see cond-mat/9909236.
[15] B. A. Berg, Nucl. Phys. B, Proc. Suppl. **63A–C**, 982 (1998).
[16] F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, J. Phys. Chem. **79**, 2361 (1975).
[17] G. Némethy, M. S. Pottle, and H. A. Scheraga, J. Phys. Chem. **87**, 1883 (1983).
[18] M. J. Sippl, G. Némethy, and H. A. Scheraga, J. Phys. Chem. **88**, 6231 (1984).
[19] T. Ooi, M. Oobatake, G. Némethy, and H. A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **84**, 3086 (1987).
[20] H. Kawai, Y. Okamoto, M. Fukugita, T. Nakazawa, and T. Kikuchi, Chem. Lett. **1991**, 213.
[21] Y. Okamoto, M. Fukugita, T. Nakazawa, and H. Kawai, Protein Eng. **4**, 639 (1991).
[22] M. Masuya, manuscript in preparation; see also http://www.ics.kagoshima-u.ac.jp/~masatom
[23] F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander, and M. Scharf, J. Comput. Chem. **16**, 273 (1995).
[24] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).
[25] P. J. Kraulis, J. Appl. Crystallogr. **24**, 946 (1991).
[26] D. Bacon and W. F. Anderson, J. Mol. Graphics **6**, 219 (1988).
[27] E. A. Merritt and M. E. P. Murphys, Acta Crystallogr., Sect. D: Biol. Crystallogr. **50**, 869 (1994).