JCTC Journal of Chemical Theory and Computation

Article

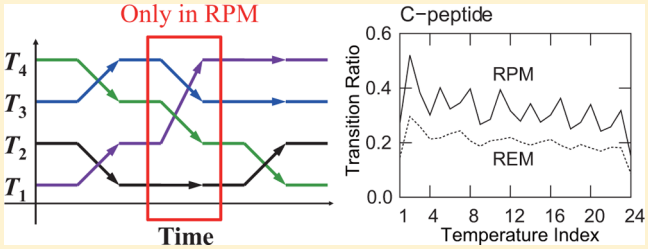
pubs.acs.org/JCTC

Replica-Permutation Method with the Suwa–Todo Algorithm beyond the Replica-Exchange Method

Satoru G. Itoh* and Hisashi Okumura*

Department of Theoretical and Computational Molecular Science, Institute for Molecular Science, Okazaki, Aichi 444-8585, Japan
Department of Structural Molecular Science, The Graduate University for Advanced Studies, Okazaki, Aichi 444-8585, Japan

ABSTRACT: We propose a new method for molecular dynamics and Monte Carlo simulations, which is referred to as the replica-permutation method (RPM), to realize more efficient sampling than the replica-exchange method (REM). In RPM, not only exchanges between two replicas but also permutations among more than two replicas are performed. Furthermore, instead of the Metropolis algorithm, the Suwa–Todo algorithm is employed for replica-permutation trials to minimize its rejection ratio. We applied RPM to particles in a double-well potential energy, Met-enkephalin in a vacuum, and a C-peptide analog of ribonuclease A in explicit water. For comparison purposes, replica-exchange molecular dynamics simulations were also performed. As a result, RPM sampled not only the temperature space but also the conformational space more efficiently than REM for all systems. From our simulations of C-peptide, we obtained the α-helix structure with salt bridges between Gly2 and Arg10, which is known in experiments. Calculating its free-energy landscape, the folding pathway was revealed from an extended structure to the α-helix structure with the salt bridges. We found that the folding pathway consists of the two steps: The first step is the "salt-bridge formation step," and the second step is the "α-helix formation step."



1. INTRODUCTION

In recent years, generalized-ensemble algorithms have been frequently employed to study biomolecular systems (for reviews, see, e.g., refs 1 and 2). This is because it is difficult to obtain sufficient sampling in the conformational space of these systems by conventional canonical-ensemble simulations.[3−9] The canonical simulations tend to get trapped in a local-minimum free-energy state of the biomolecular systems.

The replica-exchange method (REM)[10,11] is one of the most well-known methods among the generalized-ensemble algorithms. Noninteracting copies (replicas) of a target system are employed in REM. Different temperatures are assigned to the replicas. By exchanging the temperatures between the replicas, random walks of the replicas in the temperature space are realized. Accordingly, the simulation can escape from local-minimum states. It is easier to perform replica-exchange molecular dynamics (MD) or Monte Carlo (MC) simulations than to perform multicanonical MD[12,13] or MC simulations,[14,15] although the multicanonical algorithm is also one of the most well-known generalized-ensemble algorithms. In the multicanonical and similar algorithms[16−25] non-Boltzmann weight factors are used as the weight factors. These non-Boltzmann weight factors are not *a priori* known and have to be determined by iterative procedures. In REM, however, the usual Boltzmann weight factor is employed for each replica. Therefore, there is no necessity to perform the procedures of obtaining the weight factor.

In REM, the Metropolis algorithm[3] is utilized to exchange the temperatures between the replicas. The Metropolis algorithm is a Markov chain Monte Carlo (MCMC) method and widely employed to obtain a required ensemble. The Metropolis algorithm is designed so as to satisfy the detailed balance condition, which is a sufficient condition to perform state transitions based on MCMC. Trials of such state transitions are accepted or rejected stochastically, and their rejection ratio increases when there is a large difference between the probability distributions. In order to minimize the rejection ratio, a new algorithm was proposed recently by Suwa and Todo.[26]

In the Suwa–Todo algorithm, the detailed balance condition is not imposed. They introduced a graphical procedure called *weight allocation* instead of solving the detailed balance condition algebraically. By minimizing the rejection ratio with the weight allocation, this algorithm realizes efficient sampling of states. For a system which has only two states for each particle like the Ising model, however, this algorithm is exactly the same as the Metropolis algorithm.[26]

To take advantage of the Suwa–Todo algorithm, one might consider exchanging temperatures in REM using this algorithm. For example, let us consider the following exchange of the temperatures T_1 and T_2 between the replicas, replica 1 and replica 2: (replica 1 at T_1; replica 2 at T_2) → (replica 1 at T_2; replica 2 at T_1). In this case, the number of all combinations of the replicas and temperatures is only two, (replica 1 at T_1; replica 2 at T_2) and (replica 1 at T_2; replica 2 at T_1). For this

Received: September 11, 2012
Published: December 6, 2012

dx.doi.org/10.1021/ct3007919 | *J. Chem. Theory Comput.* 2013, 9, 570–581

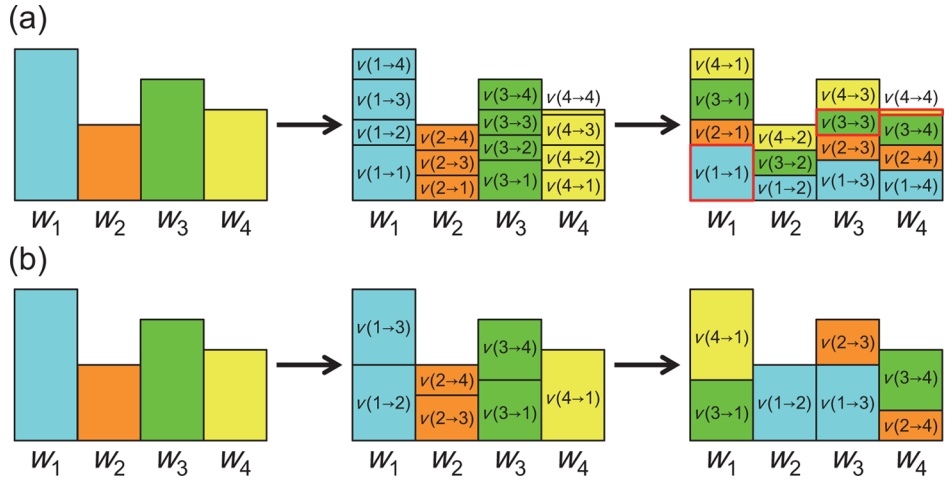


Figure 1. Schematic figures of the weight allocation of the (a) Metropolis and (b) Suwa−Todo algorithms. Red frame blocks represent rejected flows $\nu(i \to i)$ $(i = 1,...,4)$.

exchange, therefore, the Suwa−Todo algorithm is exactly the same as the Metropolis algorithm. In general, because temperatures are exchanged between two replicas as in this example, REM is not able to take advantage of the Suwa−Todo algorithm.

In order to solve this difficulty, we propose a new generalized-ensemble algorithm, which we refer to as the *replica-permutation method* (RPM). In this method, temperatures are permuted among more than two replicas using the Suwa−Todo algorithm. We can utilize the Suwa−Todo algorithm because the number of all combinations of the replicas and temperatures is larger than two.

We first apply this new method to particles in a double-well potential energy. For comparison purposes, RPM with the Metropolis algorithm and REM are also performed. As the second application of RPM, we employ Met-enkephalin in a vacuum. This penta-peptide is often used as a test system to see the usefulness of new algorithms.[27−30] Furthermore, RPM is applied to a C-peptide analog of ribonuclease A in explicit water, which is known to form an α-helix structure,[31−35] to see its sampling efficiency for a larger biomolecular system. The results of the second and third applications are compared with those of REM.

It is considered that the α-helix structure of the C-peptide analog is stabilized by salt bridges (SBs) between Gly2 and Arg10.[34] We discuss the role of the SBs for the α-helix structure. Furthermore, a folding pathway based on a free-energy landscape is presented from our simulations.

In section II, we describe the formulation of RPM. The Suwa−Todo algorithm and the graphical procedure, weight allocation, for MCMC are also introduced in this section. We present the details of our simulations in section III. The results are shown in section IV. Section V is devoted to conclusions.

II. METHODS

A. Markov Chain Monte Carlo Method with the Weight Allocation. We first describe the usual MCMC method with the Metropolis and Suwa−Todo algorithms. Let us consider that a system has n states and that the weight of state i is expressed by w_i $(i = 1,...,n)$. In MCMC, the weight w_i is given by

$$w_i = \sum_{j=1}^{n} w_j P(j \to i) \tag{1}$$

where $P(j \to i)$ is the transition probability from state j to state i. By defining the amount of stochastic flow from state j to state i as

$$\nu(j \to i) \equiv w_j P(j \to i) \tag{2}$$

Equation 1 can be rewritten as

$$w_i = \sum_{j=1}^{n} \nu(j \to i) \tag{3}$$

This amount of stochastic flow $\nu(j \to i)$ also satisfies

$$w_j = \sum_{i=1}^{n} \nu(j \to i) \tag{4}$$

because

$$\sum_{i=1}^{n} P(j \to i) = 1 \tag{5}$$

From eqs 3 and 4, the global balance equation is derived:

$$\sum_{i=1}^{n} \nu(i \to j) = \sum_{i=1}^{n} \nu(j \to i) \tag{6}$$

By performing state transitions with $\nu(j \to i)$ satisfying this equation, a required ensemble is obtained.

In the Metropolis algorithm, the detailed balance condition, which is a sufficient condition for eq 6, is employed to obtain the required ensemble. Here, the detailed balance condition is described in terms of the amount of stochastic flow as follows:

$$\nu(i \to j) = \nu(j \to i) \tag{7}$$

$\nu(j \to i)$ is given by the following equation so as to satisfy eq 7:

$$\nu(j \to i) = \frac{1}{n-1} \min[w_j, w_i], \quad j \neq i \tag{8}$$

where the coefficient $1/(n-1)$ comes from a random selection of state i from $n-1$ candidates except state j.

When $\nu(j \to i)$ satisfies eqs 3 and 4, eq 6 is automatically fulfilled. Therefore, we focus on eqs 3 and 4 after this. These

dx.doi.org/10.1021/ct3007919 | J. Chem. Theory Comput. 2013, 9, 570−581

two equations can be understood visually by the *weight allocation* as follows (see Figure 1a): eq 3 is regarded as filling a box whose size is w_i by blocks whose sizes are $v(j \to i)$ $(j = 1,...,n)$ without any vacant space. Equation 4 is regarded as dividing a block whose size is w_j into smaller blocks whose sizes are $v(j \to i)$ $(i = 1,...,n)$. In order to satisfy both equations simultaneously for any i and j, therefore, we prepare blocks w_j and boxes w_i $(i,j = 1,...,n)$. By using blocks $v(j \to i)$ divided from w_j, all w_i are filled without any space. Figure 1a shows this weight allocation of the Metropolis algorithm for a system which has four states $(n = 4)$. The block size of $v(j \to i)$ is calculated from eq 8. Red frame blocks represent rejected flow $v(i \to i)$ $(i = 1,...,4)$, and their sizes are associated directly with the average rejection ratio $\sum_i v(i \to i)/\sum_i w_i$.

A new algorithm was proposed recently by Suwa and Todo through the weight allocation to minimize the average rejection ratio for state transitions in MCMC. We refer to this algorithm as the Suwa−Todo algorithm. In this algorithm, $v(j \to i)$ satisfies eqs 3 and 4 without imposing the detailed balance condition in eq 7. In the Suwa−Todo weight allocation, w_j is divided and w_i is filled as follows (see Figure 1b): (i) Choose the state which has the maximum weight. If two or more states have the maximum weight, one of them is chosen. Here, we assume that w_1 is the maximum weight without a loss of generality. (ii) Box w_2 is filled by block w_1 $(v(1 \to 2))$. If block w_1 still remains after filling box w_2, try to fill the next box w_3 $(v(1 \to 3))$. This process is continued until the block size of w_1 becomes 0 $(v(1 \to 4),...,v(1 \to k)))$. (iii) By using block w_2, fill boxes in turns from the last partially filled box at step ii $(v(2 \to k),...,v(2 \to l))$. This procedure is repeated for the blocks $w_3,...,w_n$ $(v(3 \to l),...)$. (iv) Once all the boxes except w_1 are saturated, box w_1 $(...,v(n \to 1))$ is filled.

Figure 1b shows the weight allocation for $n = 4$. In this figure, the average rejection ratio $\sum_i v(i \to i)/\sum_i w_i$ is 0. In this Suwa−Todo algorithm, $v(j \to i)$ is given by

$$v(j \to i) = \max[0, \min[\Delta_{ji}, w_j + w_i - \Delta_{ji}, w_j, w_i]] \tag{9}$$

where

$$\Delta_{ji} \equiv S_j - S_{i-1} + w_1 \tag{10}$$

and

$$S_i \equiv \sum_{j=1}^{i} w_j,\ S_0 \equiv S_n \tag{11}$$

By performing state transitions based on the amount of stochastic flow $v(j \to i)$ or the transition probability $P(j \to i) = v(j \to i)/w_j$, the required ensemble is obtained.

Regarding the rejection ratio, from eq 9,

$$v(i \to i) = \begin{cases} \max[0, 2w_1 - S_n], & i = 1 \\ 0, & i \geq 2 \end{cases} \tag{12}$$

Therefore, the rejection ratio becomes 0, if

$$w_1 \leq \frac{S_n}{2} \tag{13}$$

It means a reject-free MC simulation is realized if w_1 is less than or equal to half of the total weight $S_n = \sum_{j=1}^{n} w_j$. This condition can be fulfilled in most cases.

B. Replica-Exchange Method with the Metropolis Algorithm. We consider a system of N atoms with their coordinate vectors and momentum vectors denoted by $q \equiv \{\boldsymbol{q}_1,...,\boldsymbol{q}_N\}$ and $p \equiv \{\boldsymbol{p}_1,...,\boldsymbol{p}_N\}$, respectively. The Hamiltonian H in state $x \equiv (q,p)$ is given by the sum of the kinetic energy K and potential energy V:

$$H(x) = K(p) + V(q) \tag{14}$$

In the canonical ensemble at temperature T, each state x is weighted by the Boltzmann factor:

$$W_B(x) = e^{-\beta H(x)} \tag{15}$$

where $\beta = 1/k_B T$ (k_B is the Boltzmann constant).

Let us suppose that there are M *noninteracting* copies (or replicas) of the original system in the canonical ensemble at M different temperatures T_m $(m = 1,...,M)$. In REM, the replicas are arranged so that there would be always exactly one replica at each temperature. In other words, there is a one-to-one correspondence between the replicas and temperatures. Therefore, the label i $(i = 1,...,M)$ for the replicas is a permutation of the label m $(m = 1,...,M)$ for the temperatures, and vice versa:

$$\begin{cases} i = i(m) \equiv f(m) \\ m = m(i) \equiv f^{-1}(i) \end{cases} \tag{16}$$

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

Let $X_\alpha = \{x_1^{[i(1)]},...,x_M^{[i(M)]}\} = \{x_{m(1)}^{[1]},...,x_{m(M)}^{[M]}\}$ stand for a "state" in REM. Here, the superscript i and the subscript m in $x_m^{[i]}$ are labels of the replicas and temperatures, respectively. All possible combinations between the replicas and temperatures are labeled by the subscript α. The state X_α is specified by the M sets of coordinates $q^{[i]}$ and momenta $p^{[i]}$ of N atoms in replica i at temperature T_m:

$$x_m^{[i]} \equiv (q^{[i]}, p^{[i]})_m \tag{17}$$

Because the replicas are noninteracting, the weight factor $w_R(X_\alpha)$ for the state X_α is given by the product of Boltzmann factors for each replica i (or at temperature T_m):

$$\begin{aligned} w_R(X_\alpha) &= \prod_{i=1}^{M} \exp\{-\beta_{m(i)} H(x_{m(i)}^{[i]})\} \\ &= \prod_{m=1}^{M} \exp\{-\beta_m H(x_m^{[i(m)]})\} \end{aligned} \tag{18}$$

where $i(m)$ and $m(i)$ are the permutation functions in eq 16.

We now consider exchanging a pair of replicas in REM. Suppose we exchange replicas j and k which are at temperatures T_m and T_n, respectively $(j = i(m),\ k = i(n))$:

$$X_\alpha = \{..., x_m^{[j]}, ..., x_n^{[k]}, ...\} \to X_\beta = \{..., x_m^{[k]}, ..., x_n^{[j]}, ...\} \tag{19}$$

From eq 8 the amount of stochastic flow $v(X_\alpha \to X_\beta)$ for this replica exchange is given by

$$v(X_\alpha \to X_\beta) = C \min[w_R(X_\alpha), w_R(X_\beta)] \tag{20}$$

and the transition probability $P(X_\alpha \to X_\beta)$ is expressed by

$$P(X_\alpha \to X_\beta) = C \min\left[1, \frac{w_R(X_\beta)}{w_R(X_\alpha)}\right] \tag{21}$$

where $C = 1/({}_M C_2)$ if replicas j and k were selected randomly among M replicas. Here, ${}_M C_2$ is the number of 2-combinations from M elements: ${}_M C_2 = M!/\{(M-2)!2!\}$. In the usual REM,

dx.doi.org/10.1021/ct3007919 | J. Chem. Theory Comput. 2013, 9, 570−581

Table 1. Example of an Assignment of the Labels in RPM with Three Replicas

temperature/label α of X_α	1	2	3	4	5	6
T_1	replica 1	replica 1	replica 2	replica 2	replica 3	replica 3
T_2	replica 2	replica 3	replica 1	replica 3	replica 1	replica 2
T_3	replica 3	replica 2	replica 3	replica 1	replica 2	replica 1

replica exchanges are tried only between two neighboring temperatures. In this case, $C = 1$ because $M = 2$.

C. Replica-Permutation Method with the Suwa−Todo Algorithm. We consider performing a replica permutation among all M replicas as a generalization of a replica exchange in eq 19:

$$X_\alpha = \{x_1^{[i(1)]}, ..., x_M^{[i(M)]}\} \rightarrow X_\beta = \{x_1^{[j(1)]}, ..., x_M^{[j(M)]}\} \tag{22}$$

where both i and j are permutation functions and $i \neq j$. Not only an exchange between two replicas but also a permutation among more than two replicas are allowed in this method. Note that the number of all possible combinations between the replicas and temperatures is $M!$ Therefore, the index α of X_α takes a value between 1 and $M!$

In the Metropolis algorithm, the transition probability $P(X_\alpha \rightarrow X_\beta)$ for this replica permutation is also given by eq 21. The replicas are allowed to transit to non-neighboring temperatures, but $P(X_\alpha \rightarrow X_\beta)$ takes a quite small value for such replica permutation. Accordingly, most of such replica permutations are rejected. The number of replica permutations in which any replica does not transit to non-neighboring temperatures is given by

$$\sum_{n=1}^{[M/2]} {}_{M-n}C_n \tag{23}$$

where

$$\left[\frac{M}{2}\right] = \begin{cases} \dfrac{M}{2}, & \text{for even number of } M \\ \dfrac{M-1}{2}, & \text{for odd number of } M \end{cases} \tag{24}$$

On the other hand, the number of all replica-permutation candidates is $M! - 1$ because the current combination between the replicas and temperatures is not included. Therefore, the probability of trying replica permutations which do not include non-neighboring transitions is $(\sum_{n-1}^{[M/2]} {}_{M-n}C_n)/(M! - 1)$. Thus, most of the replica-permutation trials are rejected for large M, because

$$\sum_{n=1}^{[M/2]} {}_{M-n}C_n < \sum_{n=0}^{M} {}_{M}C_n = 2^M \ll M! \tag{25}$$

To avoid this rejection problem, we apply the Suwa−Todo algorithm to the replica permutations. As in section II.A, we assume that $w_R(X_1)$ is the maximum weight without a loss of generality. The amount of stochastic flow $v(X_\alpha \rightarrow X_\beta)$ is determined by the weight allocation in the same way as in section II.A only by replacing the weight w_i with $w_R(X_\alpha)$. From eqs 9, 10, and 11, $v(X_\alpha \rightarrow X_\beta)$ is given by

$$v(X_\alpha \rightarrow X_\beta) = \max[0, \min[\Delta_{\alpha\beta}, w_R(X_\alpha) + w_R(X_\beta) - \Delta_{\alpha\beta}, w_R(X_\alpha), w_R(X_\beta)]] \tag{26}$$

where

$$\Delta_{\alpha\beta} \equiv S_\alpha - S_{\beta-1} + w_R(X_1) \tag{27}$$

and

$$S_\alpha \equiv \sum_{\beta=1}^{\alpha} w_R(X_\beta),\ S_0 \equiv S_{M!} \tag{28}$$

If $w_R(X_\gamma)$ $(\gamma \neq 1)$ is the maximum weight more generally, eqs 27 and 28 are modified as follows:

$$\Delta_{\alpha\beta} \equiv S_\alpha - S_{\beta-1} + w_R(X_\gamma) \tag{29}$$

and

$$S_\alpha \equiv \begin{cases} \displaystyle\sum_{\beta=\gamma}^{\alpha} w_R(X_\beta), \text{ for } \alpha \geq \gamma \\ \displaystyle\sum_{\beta=\gamma}^{M!} w_R(X_\beta) + \sum_{\beta=1}^{\alpha} w_R(X_\beta), \text{ for } \alpha < \gamma \end{cases}$$

$$S_0 \equiv S_{M!} \tag{30}$$

A replica-permutation simulation with the Suwa−Todo algorithm is performed as follows:

Step 1: The label α $(\alpha = 1,...,M!)$ of X_α is assigned to all combinations between the replicas and temperatures. Table 1 shows an example for $M = 3$ (three replicas).

Step 2: For each replica, a canonical MD or MC simulation at the assigned temperature is carried out simultaneously and independently for certain steps.

Step 3: A replica-permutation trial is performed as follows: First, each weight is obtained with eq 18, and the maximum weight $w_R(X_\gamma)$ is determined. Next, we calculate the amount of stochastic flow $v(X_\alpha \rightarrow X_\beta)$ in eq 26 and the transition probability $P(X_\alpha \rightarrow X_\beta) = v(X_\alpha \rightarrow X_\beta)/w_R(X_\alpha)$ for $\beta = 1,...,M!$. Finally, transition from state X_α to state X_β is accepted stochastically with the probability $P(X_\alpha \rightarrow X_\beta)$.

Repeating steps 2 and 3, we can carry out the replica permutation MD or MC simulation.

Figure 2 shows an example of time series of temperatures in RPM. This method realizes not only minimization of the rejection ratio but also transitions of the replicas to non-neighboring temperatures.

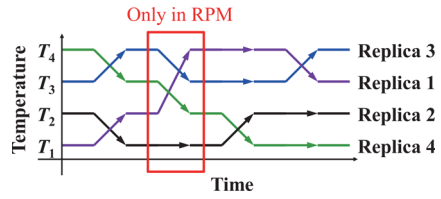


Figure 2. An example of time series of temperatures in RPM. The transitions of replicas in the red square frame are not realized in REM.

The number of combinations between the replicas and temperatures increases in proportion to $M!$. For a large number of replicas, we can divide all replicas and temperatures into

dx.doi.org/10.1021/ct3007919 | J. Chem. Theory Comput. 2013, 9, 570−581

subsets to decrease the number of combinations. Although such a division is not necessary, three to eight replicas are appropriate in each subset. As an example, let us consider that the total number of the replicas is eight and that they are divided into two subsets which have four replicas. In this case, the replica-permutation simulation is performed as follows:

Step 1: Let us suppose that the temperatures are assigned to the replicas at an initial state as

$$\begin{pmatrix} \text{replica 1 at } T_1 \\ \text{replica 2 at } T_2 \\ \vdots \\ \text{replica 8 at } T_8 \end{pmatrix}$$

They are divided into two subsets:

$$\begin{pmatrix} \text{replica 1 at } T_1 \\ \text{replica 2 at } T_2 \\ \text{replica 3 at } T_3 \\ \text{replica 4 at } T_4 \end{pmatrix} \text{ and } \begin{pmatrix} \text{replica 5 at } T_5 \\ \text{replica 6 at } T_6 \\ \text{replica 7 at } T_7 \\ \text{replica 8 at } T_8 \end{pmatrix}$$

All combinations between the replicas and temperatures for the former subset and those for the latter subset are labeled X_α^1 and X_α^3 (α = 1,...,4!), respectively. Moreover, two more subsets are prepared so that components would differ from those of the previous subsets:

$$\begin{pmatrix} \text{replica 3 at } T_3 \\ \text{replica 4 at } T_4 \\ \text{replica 5 at } T_5 \\ \text{replica 6 at } T_6 \end{pmatrix} \text{ and } \begin{pmatrix} \text{replica 7 at } T_7 \\ \text{replica 8 at } T_8 \\ \text{replica 1 at } T_1 \\ \text{replica 2 at } T_2 \end{pmatrix}$$

The labels for the former subset and the latter subset are X_α^2 and X_α^4, respectively.

Step 2: For each replica, a canonical MD or MC simulation at the assigned temperature is carried out simultaneously and independently for a certain steps.

Step 3: At an odd number of trail time, replica permutations for X_α^1 and X_α^3 are carried out. That is, four replicas are permuted among four corresponding temperatures in each subset. At an even number of trial time, replica permutations for X_α^2 and X_α^4 are performed.

Step 4: As a result of the replica permutation for X_α^1, let us suppose that the combination of the replicas and temperatures is changed as

$$\begin{pmatrix} \text{replica 1 at } T_1 \\ \text{replica 2 at } T_2 \\ \text{replica 3 at } T_3 \\ \text{replica 4 at } T_4 \end{pmatrix} \rightarrow \begin{pmatrix} \text{replica 1 at } T_1 \\ \text{replica 3 at } T_2 \\ \text{replica 2 at } T_3 \\ \text{replica 4 at } T_4 \end{pmatrix} \quad (31)$$

Due to this permutation, the combination in X_α^2 is automatically changed without a replica permutation for this subset:

$$\begin{pmatrix} \text{replica 3 at } T_3 \\ \text{replica 4 at } T_4 \\ \text{replica 5 at } T_5 \\ \text{replica 6 at } T_6 \end{pmatrix} \xrightarrow{\text{by eq 31}} \begin{pmatrix} \text{replica 2 at } T_3 \\ \text{replica 4 at } T_4 \\ \text{replica 5 at } T_5 \\ \text{replica 6 at } T_6 \end{pmatrix} \quad (32)$$

To avoid relabeling for X_α^2, we perform the next permutation regarding replica 2 as replica 3. Although the combination of the temperatures and replicas may be changed by the last replica permutation, we do not need to relabel the combinations as in Table 1 by replacing the replicas in this way.

Repeating steps 2−4, we continue the replica permutation MD or MC simulation.

D. Reweighting Techniques. The results obtained from RPM can be analyzed with the reweighting techniques as in REM.[36,37] Let us suppose that we have carried out a RPM simulation with M replicas and M different temperatures T_m (m = 1,...,M).

For appropriate reaction coordinates ξ_1 and ξ_2, the canonical probability distribution $P_T(\xi_1,\xi_2)$ at any temperature T can be calculated from

$$P_T(\xi_1, \xi_2) = \sum_E \frac{\sum_{m=1}^M (g_m)^{-1} N_m(E; \xi_1, \xi_2) e^{-\beta E}}{\sum_{m=1}^M (g_m)^{-1} n_m e^{f_m - \beta_m E}} \quad (33)$$

and

$$e^{-f_m} = \sum_{\xi_1,\xi_2} P_{T_m}(\xi_1, \xi_2) \quad (34)$$

Here, $g_m = 1 + 2\tau_m$, τ_m is the integrated autocorrelation time at temperature T_m, $N_m(E;\xi_1,\xi_2)$ is the histogram of the potential energy and the reaction coordinates ξ_1 and ξ_2 at T_m, and n_m is the total number of samples obtained at T_m. Note that this probability distribution is not normalized. Equations 33 and 34 are solved self-consistently by iteration. For biomolecular systems, the quantity g_m can safely be set to be a constant in the reweighting formulas,[37] and so we set g_m = 1 throughout the analyses in the present work. These equations can be easily generalized to any numbers of reaction coordinates (ξ_1,ξ_2,...).

From the canonical probability distribution $P_T(\xi_1,\xi_2)$ in eq 33, the expectation value of a physical quantity A at any temperature T is given by

$$\langle A \rangle_T = \frac{\sum_{\xi_1,\xi_2} A(\xi_1, \xi_2) P_T(\xi_1, \xi_2)}{\sum_{\xi_1,\xi_2} P_T(\xi_1, \xi_2)} \quad (35)$$

We can also calculate the free energy (or the potential of mean force) as a function of the reaction coordinates ξ_1 and ξ_2 at any temperature T from

$$F_T(\xi_1, \xi_2) = -k_B T \ln P_T(\xi_1, \xi_2) \quad (36)$$

III. COMPUTATIONAL DETAILS

A. Asymmetric Double-Well Potential Energy. In order to verify that RPM gives correct ensembles, we first applied this method to a simple system. The system has 100 noninteracting particles with a one-dimensional asymmetric double-well potential energy. The potential energy $V(q)$ at a coordinate q is defined by

$$V(q) = ((q + 1)^2 - 1)((q - 1)^2 - 0.9) \text{ kcal/(mol·Å}^4) \quad (37)$$

dx.doi.org/10.1021/ct3007919 | J. Chem. Theory Comput. 2013, 9, 570−581

To see the usefulness of the Suwa−Todo algorithm for the replica-permutation method, we performed replica-permutation MD simulations with both Metropolis and Suwa−Todo algorithms. From now on, we will call the replica-permutation method with the Suwa−Todo algorithm RPM and that with the Metropolis algorithm M-RPM. The MD versions of the former and latter are called RPMD and M-RPMD, respectively. Conventional replica-exchange MD (REMD) simulations were also carried out for comparison purposes. We prepared 40 different initial conditions (ICs) for each method. The MD simulation from each IC was performed for 10.0 ns including an equilibration run for 1.0 ns. The time step was taken to be 1.0 fs. The mass of each particle was 1.0 u. Six replicas were used, and temperatures were set at 200, 235, 275, 325, 380, and 450 K. The temperatures were controlled by the Gaussian constraint method[4,5] to avoid the problem of nonergodicity in the Nosé-Hoover thermostat[6−8] for the noninteracting particle system. The trajectory data were stored every 10.0 fs. Replica permutations or exchanges were tried every 1.0 ps.

B. Met-Enkephalin in a Vacuum. To see the usefulness of RPM for biomolecular systems, we next employed a Met-enkephalin molecule in a vacuum as a test system. The results were compared with those obtained from a REMD simulation. The AMBER parm99SB force field[38,39] was used. The N-terminus and the C-terminus of Met-enkephalin were blocked by the acetyl group and the N-methyl group, respectively. Therefore, the amino acid sequence is Ace-YGGFM-Nme. The SHAKE algorithm[40] was employed to constrain bond lengths with the hydrogen atoms during our simulations. The temperature was controlled by the Nosé-Hoover thermostat.[6−8] The time step was taken to be 1.0 fs. The initial conformation was an extended structure.

The number of replicas was 12, and temperatures were 200, 230, 265, 300, 340, 385, 435, 490, 555, 635, 720, and 820 K. For PRM, we divided the 12 replicas into two subsets, which had six replicas and six temperatures. The RPMD and REMD simulations were performed for 50.0 ns per replica including an equilibration run for 1.0 ns. The trajectory data were stored every 100 fs. Replica permutations or exchanges were tried every 1.0 ps.

C. C-Peptide in Explicit Water. In order to demonstrate the sampling efficiency of RPM for a lager system, we performed a RPMD simulation of a C-peptide analog in explicit water solvent.[32] This peptide is known to form an α-helix structure at a lower temperature than 318 K at pH 5.2.[32,35] Because the charges at the peptide termini affect helix stability,[31] we blocked the termini of C-peptide with neutral Nme and Ace. Therefore, the amino acid sequence is Ace-AETAAAKFLRAHA-Nme. The histidine residue was protonated to conform our simulation conditions to the experimental pH. A REMD simulation was also carried out. The number of water molecules was 1800, and two chlorine ions were added as counterions. The AMBER parm99SB[38,39] was used. The model for the water molecules was the TIP3P rigid-body model.[41] The temperature was controlled by the Nosé-Hoover thermostat.[6−8] The SHAKE algorithm[40] was employed to constrain bond lengths with the hydrogen atoms of C-peptide and to fix the water molecule structures during our simulations. This system was put in a cubic unit cell with a side length of 38.6 Å with the periodic boundary conditions. The cutoff distance for the Lennard-Jones potential energy was 12.0 Å. The electrostatic potential energy was calculated using the Ewald method.[42] The multiple-time-step method[43] was employed in our MD simulations. For interactions between the C-peptide atoms and those between the C-peptide atoms and the water atoms, the time step was taken to be 1.0 fs. The time step of 4.0 fs was used for interactions between the water atoms. The initial conformation was an extended structure.

The numbers of replicas in the RPMD and REMD simulations were 24, and temperatures were 281, 285, 289, 294, 299, 304, 309, 314, 320, 326, 332, 338, 344, 351, 358, 365, 372, 380, 388, 396, 405, 414, 423, and 433 K. These replicas were divided into four subsets in RPM. Each subset had six replicas and six temperatures. The RPMD and REMD simulations were performed for 40.0 ns per replica including equilibration run for 4.0 ns. Namely, the production run of each simulation was carried out for 864.0 ns in total. The trajectory data were stored every 400 fs. Trials of replica permutations and exchanges were performed every 4.0 ps.

IV. RESULTS AND DISCUSSION

A. Asymmetric Double-Well Potential Energy. We first show that RPM (and M-RPM) gives correct probability distributions for the asymmetric double-well potential energy. The probability distributions $P(q)$ in the RPMD, M-RPMD, and REMD simulations are presented in Figure 3. Here, $P(q)$

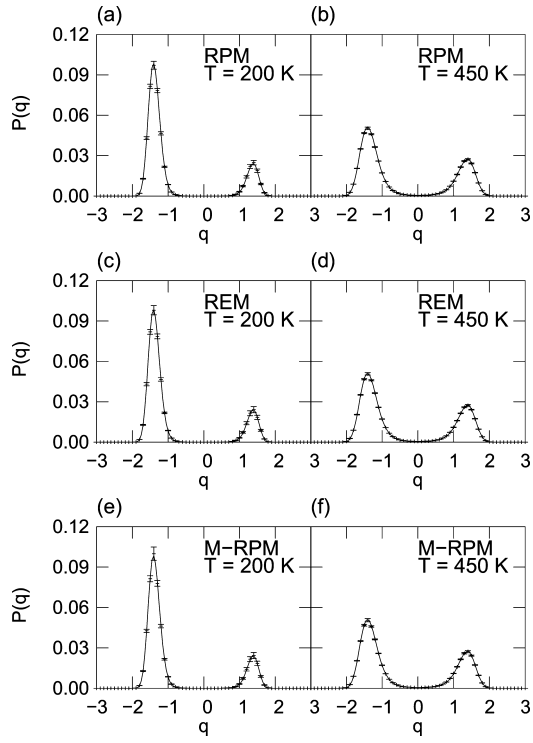


Figure 3. The probability distributions $P(q)$ of the coordinates q at (a, c, e) T = 200 K and (b, d, f) T = 450 K. These results were obtained from the (a, b) RPMD simulations, (c, d) REMD simulations, and (e, f) M-RPMD simulations. The solid lines are the probability distributions in eq 38.

for each method was obtained from the 40 simulations starting from the different ICs. The bin size Δq for $P(q)$ was taken to be 0.05 Å. The errors were estimated as the standard deviations of the 40 simulations. To see the accuracy of the simulation results, exact probability distributions $P_{\text{exact}}(q)$ are also illustrated as the solid lines in the figure. $P_{\text{exact}}(q)$ at a temperature T_0 was calculated numerically by

dx.doi.org/10.1021/ct3007919 | J. Chem. Theory Comput. 2013, 9, 570−581

$$P_{\text{exact}}(q) = C_{\text{DW}} \int_{q-\Delta q/2}^{q+\Delta q/2} \mathrm{d}q' \exp(-\beta_0 V(q')) \tag{38}$$

where C_{DW} is the normalization constant: $C_{\text{DW}} = (\int_{-\infty}^{\infty} \mathrm{d}q\, P_{\text{exact}}(q))^{-1}$. The integration in this equation was computed using Simpson's method. As shown in Figure 3, all $P(q)$ show good agreement with $P_{\text{exact}}(q)$ at all temperatures. Correct probability distributions are obtained by not only REM but also RPM (and M-RPM).

The transition ratios of the replicas from a temperature to another temperature in each method are shown in Figure 4a.

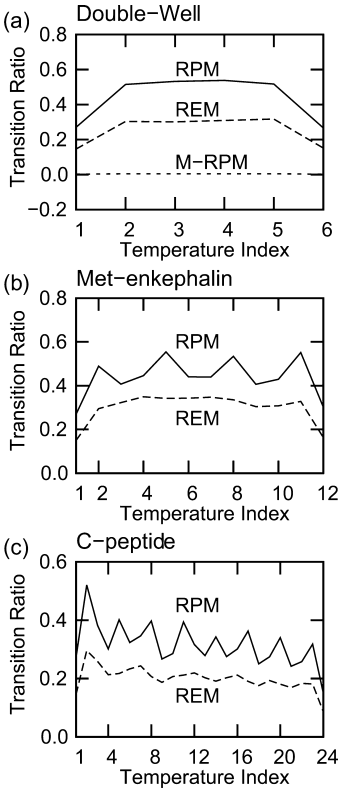


Figure 4. Transition ratios of the replicas for (a) particles in the double-well potential, (b) Met-enkephalin in a vacuum, and (c) C-peptide in explicit water. Temperatures are represented as the temperature indices. The smallest and the highest indices correspond to the lowest and the highest temperatures, respectively.

The transition ratio of the replicas is defined here as a probability with which a replica at the temperature is transferred to another temperature. RPM has the largest transition ratios at all temperatures among all methods. On the other hand, those of M-RPM are extremely small (the values are 0.003−0.004). This is because most of the replica-permutation trials were rejected. To realize efficient replica permutations, therefore, it is essential to employ the Suwa−Todo algorithm. The total numbers of tunneling times of all replicas during the simulations are listed in Table 2. Here, when a replica makes a round trip between the lowest and highest temperatures, it is counted as one tunneling. The number of the tunneling times is useful information to see how efficiently the simulation samples the temperature space. The value for each method in the table was obtained by taking an average of the 40 simulations' results. The number of the tunneling times in RPM was 1.7 times larger than that in REM. It means that RPM realizes efficient sampling in the temperature space. On the other hand, M-RPM did not have such a large number of tunneling times. This is because its transition ratios were very small.

Table 2. The Total Numbers of Tunneling Times during the Simulations

	double-well[a]	enkephalin	C-peptide
RPM	424 ± 64	459 ± 21	58 ± 6
REM	254 ± 39	357 ± 14	27 ± 4
M-RPM	14 ± 3		

[a]The values were obtained by taking an average of the 40 simulations' results.

To estimate the convergence speed to the exact probability distributions, we examined the time series of the average deviations of probability distributions $P^{i,j}_{\text{particle}}(q;t)$ from $P_{\text{exact}}(q)$, where $P^{i,j}_{\text{particle}}(q;t)$ is a probability distribution obtained from a single MD simulation of particle i from IC j accumulated until time t. The average deviation $\Delta_q(t)$ at time t is defined by

$$\Delta_q(t) = \frac{1}{N_{\text{bin}} N_{\text{IC}} N_{\text{particle}}} \times \sum_{k=1}^{N_{\text{bin}}} \sum_{j=1}^{N_{\text{IC}}} \sum_{i=1}^{N_{\text{particle}}} |P^{i,j}_{\text{particle}}(q_k; t) - P_{\text{exact}}(q_k)| \tag{39}$$

where N_{bin} is the number of the bins, N_{IC} is the number of ICs, N_{particle} is the number of the particles in a single simulation, and q_k is the coordinate at bin k. A total of 120 bins ($N_{\text{bin}} = 120$) ranging between $q_k = -3$ and $q_k = 3$ were taken into account. Figure 5 shows $\Delta_q(t)$ calculated from each method.

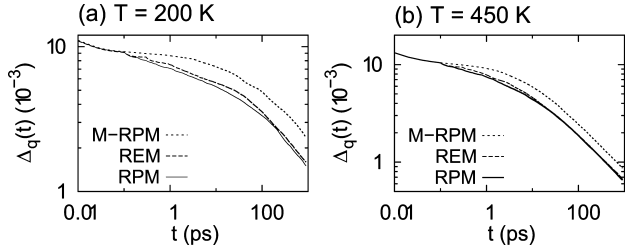


Figure 5. Average deviation $\Delta_q(t)$ in eq 39 at (a) T = 200 K and (b) T = 450 K. The solid line, the dashed line, and the dotted line show $\Delta_q(t)$ obtained from the RPMD simulations, the REMD simulations, and the M-RPMD simulations, respectively.

Convergence of M-RPM is the slowest due to the low transition ratios of the replicas. RPM shows slightly faster convergence than REM at the lowest temperature (T = 200 K) although convergence at the highest temperature (T = 450 K) is almost the same. It is important to increase the sampling efficiency at a low temperature because the conformational sampling at a high temperature is originally easier than at a low temperature. Therefore, RPM realizes efficient sampling not only in the temperature space but also in the coordinate space at a low temperature. It is again because the transition ratios of the replicas by RPM are higher than those by REM.

B. Met-Enkephalin in a Vacuum. The replica-transition ratios and the total numbers of tunneling times for Met-enkephalin are shown in Figure 4b and Table 2, respectively. The transition ratios in RPM are larger than those of REM at all temperatures. RPM also has larger tunneling times than REM. Thus, efficient sampling in the temperature space was realized by RPM for this biomolecular system, too.

dx.doi.org/10.1021/ct3007919 | J. Chem. Theory Comput. 2013, 9, 570−581

Its local-minimum free-energy structures in a vacuum have been reported for various force fields, although those for the AMBER parm99SB force field[39] have yet to be reported. For example, the global-minimum and local-minimum structures for the CHARMM22 force field[44] are shown in Figure 6. To obtain

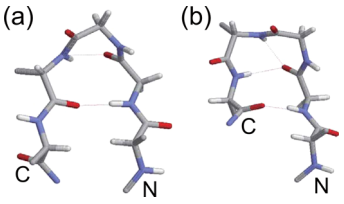


Figure 6. The (a) global-minimum and (b) local-minimum structures of Met-enkephalin in a vacuum for the CHARMM22 force field.

global-minimum and local-minimum structures in the AMBER parm99SB force field and to see the sampling efficiency of RPM, we illustrate free-energy landscapes at 200 K in Figure 7. The abscissa and ordinate are the root-mean-square deviation (RMSD) with respect to the structure in Figure 6a ($RMSD_1$) and that with respect to the structure in Figure 6b ($RMSD_2$), respectively. Here, the RMSD is defined by

$$\mathrm{RMSD} = \min\left[\sqrt{\frac{1}{n}\sum_j (q_j - q_j^0)^2}\right] \tag{40}$$

where n is the number of the backbone atoms in Met-enkephalin, $\{q_j^0\}$ are the coordinates in the reference conformation, and the minimization is over the rigid translations and rigid rotations for the coordinates of the conformation $\{q_j\}$ with respect to the center of geometry. The free-energy landscapes in Figure 7a and b were calculated from eq 36 with the reweighting techniques. These landscapes show good agreement with each other and have five local-minimum free-energy states. The five local-minimum states are labeled A to E, as shown in the figures. The free-energy landscapes in Figure 7c and d were obtained from the raw histogram without the reweighting techniques. In the landscape of REM in Figure 7d, states D and E are not observed, although these states are observed in RPM.

In order to discuss sampling efficiency at the lowest temperature more quantitatively, we counted the numbers of visiting times in each state, as listed in Table 3. Here, when a replica visited a local-minimum state at the lowest temperature after the replica had visited another state at the lowest temperature, it is counted as one visit. The errors were estimated using the jackknife method[45−47] in which the production run was divided into 20 segments. We regarded the regions presented in Table 4 as those for the local-minimum states. As shown in Table 3, REM did not visit state D and state E at the lowest temperature. This is the reason why these states were not observed in Figure 7d. On the other hand, RPM visited all states, and the numbers of visiting times in RPM are larger than REM for all states. RPM thus samples the conformational space more efficiently than REM at the lowest temperature.

The representative conformations at the local-minimum free-energy states are also shown in Figure 7. The structural features are as follows: (State A) The structure in state A is the global-minimum free-energy structure for the AMBER parm99SB force field and similar to that for the CHARMM22 force field in

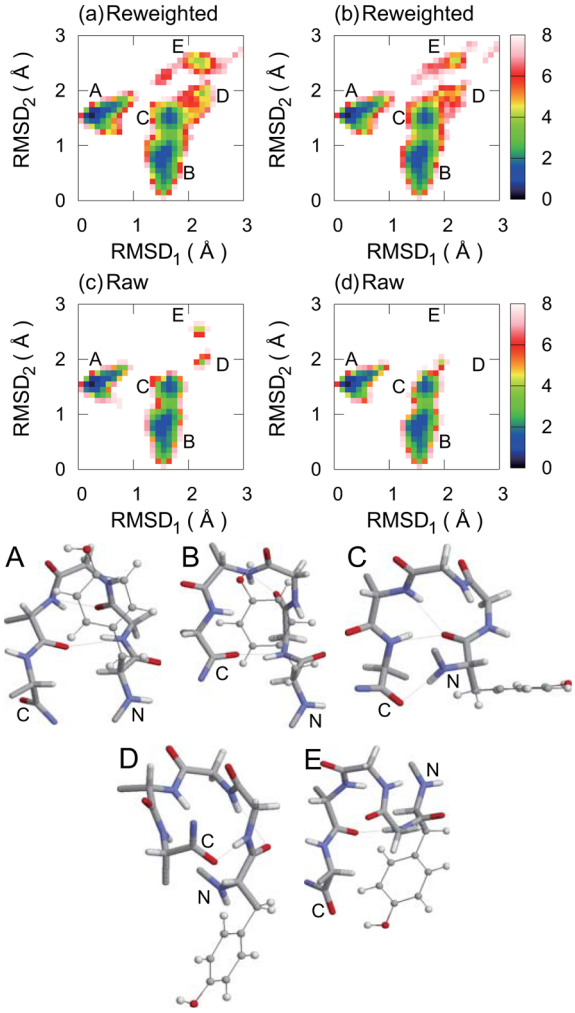


Figure 7. Free-energy landscapes at T = 200 K obtained using the reweighting techniques from the (a) RPMD and (b) REMD simulations and those calculated from the raw histograms obtained by the (c) RPMD and (d) REMD simulations. The abscissa is the RMSD with respect to the structure in Figure 6a. The ordinate is the RMSD with respect to the structure in Figure 6b. The unit of the free-energy landscape is kcal/mol. The labels A to E show the global-minimum and local-minimum free-energy states. The representative conformations at these states are also presented. The dotted lines denote hydrogen bonds. The figures were drawn by RasMol.[53]

Table 3. The Number of Visiting Times in Each State of Met-Enkephalin in a Vacuum

method	A	B	C	D	E
RPM	97 ± 9	98 ± 6	88 ± 7	1 ± 1	2 ± 1
REM	86 ± 5	86 ± 6	65 ± 5	0 ± 0	0 ± 0

Table 4. RMSD Ranges for Each Local-Minimum Free-Energy State

state	$RMSD_1$ (Å)	$RMSD_2$ (Å)
A	0.0−0.6	1.3−1.8
B	1.3−1.7	0.4−1.0
C	1.5−1.8	1.3−1.7
D	2.1−2.4	1.8−2.1
E	2.0−2.4	2.3−2.8

dx.doi.org/10.1021/ct3007919 | J. Chem. Theory Comput. 2013, 9, 570−581

Figure 6a. This structure has two hydrogen bonds between NH of Gly2 and CO of Phe4 and between CO of Gly2 and NH of Phe4. The hydroxy group of the Tyr1 side chain is close to CO of Gly3. (State B) The structure in state B is almost the same as the structure in Figure 6b. Two hydrogen bonds are formed between NH of Gly2 and CO of Met5 and between CO of Gly2 and NH of Phe4. However, this structure does not have the hydrogen bond between CO of Gly2 and NH of Met5 which exists in Figure 6b. The distance between the hydroxy group of Tyr1 and CO of Gly3 is small, as in state A. (State C) There are two hydrogen bonds between CO of Tyr1 and NH of Phe4 and between CO of Tyr1 and NH of Met5 in state C. (State D) Two hydrogen bonds between CO of Tyr1 and NH of Gly3 and between NH of Gly2 and CO of Met5 are formed in state D. As for the structures in states C and D, the Tyr1 hydroxy group is not close to any backbone CO. (State E) The structure in state E has a hydrogen bond between NH of Gly2 and CO of Phe4. This structure also has a small distance between the Tyr1 hydroxy group and CO of Met5.

C. C-peptide in Explicit Water. The transition ratios of the replicas and the total numbers of tunneling times for C-peptide in explicit water are shown in Figure 4c and Table 2, respectively. The transition ratios in RPM are larger than those in REM at all temperatures, again. The tunneling time of RPM was about 2.1 times larger than that of REM. The time series of the temperatures of replica 1, replica 9, and replica 17 are shown in Figure 8. Figure 8 actually shows more frequent tunneling in RPM than in REM. In REM, six replicas had never made a round trip between the lowest and highest temperatures during the simulation. In contrast, all replicas had at least one tunneling in RPM. Most of them had more than two tunnelings. Therefore, RPM samples the temperature space more efficiently than REM.

It was reported in the experiments that C-peptide has a helix structure with salt bridges (SBs) between Glu2 and Arg10 at a low temperature.[33,34] We obtained helix structures which had such SBs in our RPMD and REMD simulations, as in these reports. The lowest potential-energy conformation among these helix structures for each simulation is presented in Figure 9. Here, we employed the DSSP (define secondary structure of proteins) criteria[48] for hydrogen bonds between the side chains of Glu2 and Arg10 and for secondary structures of C-peptide. The structure in Figure 9a obtained from the RPMD simulation has two hydrogen bonds between O_ε of Glu2 and H_η of Arg10 and between O_ε of Glu2 and H_ε of Arg10. The residues from Ala4 to Ala11 form the α-helix structure. As for the structure from the REMD simulation in Figure 9b, the two O_ε atoms of Glu2 form the hydrogen bonds with the two H_η atoms of Arg10. The α-helix structure is formed between Ala4 and Leu9. Although these structures are slightly different from each other, both RPM and REM sampled conformations near the lowest potential-energy helix structure in the other method.

To see the effects of the SBs on the α-helix structures, we calculated probabilities of the α-helix structures with the SBs as well as without them. These probabilities at 281 K for each residue are shown in Figure 10. The probabilities without the SBs in RPM agree well with those in REM. The probabilities for residues 4 to 7 in both methods are high regardless of the existence of the SBs. This is because their amino acid sequence is AAAK, and this sequence is known for having α-helix structures.[49] In RPM, the probabilities with the SBs are especially higher than those without the SBs while both

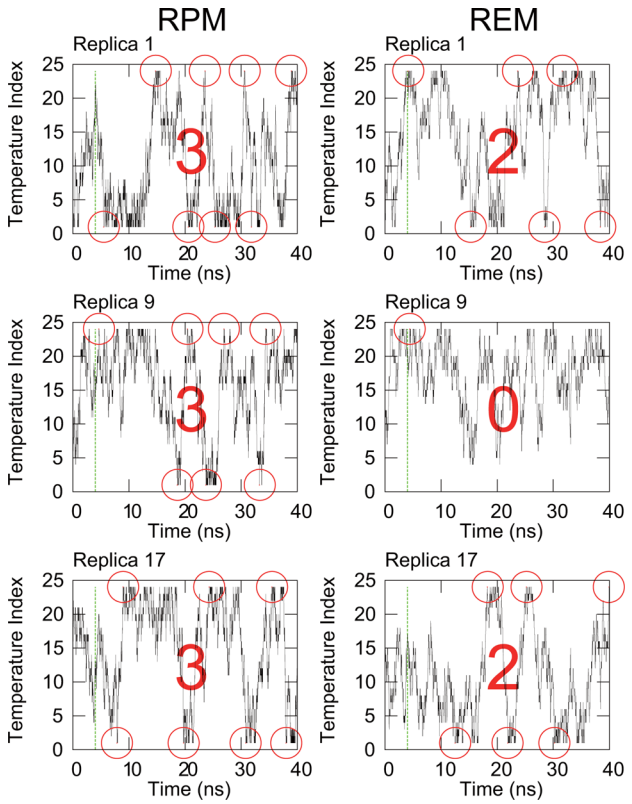


Figure 8. Time series of the temperatures of replica 1, replica 9, and replica 17. The left-hand figures and the right-hand figures are obtained from the RPMD simulation and the REMD simulation, respectively. The temperatures are represented as the temperature indices. Indices 1 and 24 correspond the lowest and highest temperatures, respectively. The production runs started from the green dashed lines. Red circles indicate the steps at which the replicas reached the highest (lowest) temperature after they had visited the lowest (highest) temperature. The red numbers present the tunneling times of the replicas.

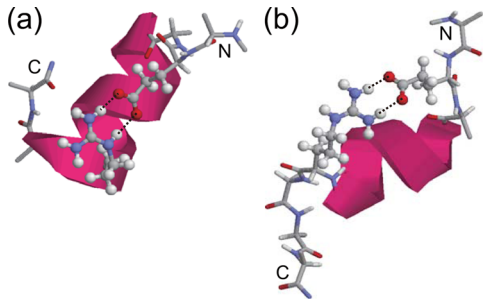


Figure 9. The lowest potential-energy conformations of C-peptide, which had α-helix structures with SBs between Glu2 and Arg10, obtained from the (a) RPMD and (b) REMD simulations. The dotted lines denote the hydrogen bonds between the side chains. The figures were created with RasMol.[53]

probabilities are almost the same in REM. We discuss the origin of this difference between RPM and REM later.

It is considered that the SBs between Glu2 and Arg10 stabilize the α-helix structure of C-peptide.[34] In order to investigate the relation between the SBs and the stability of the α-helix structure, we calculated a free-energy landscape at 281 K in each method from eq 36 with the reweighting techniques. These free-energy landscapes are shown in Figure 11. The

dx.doi.org/10.1021/ct3007919 | J. Chem. Theory Comput. 2013, 9, 570−581

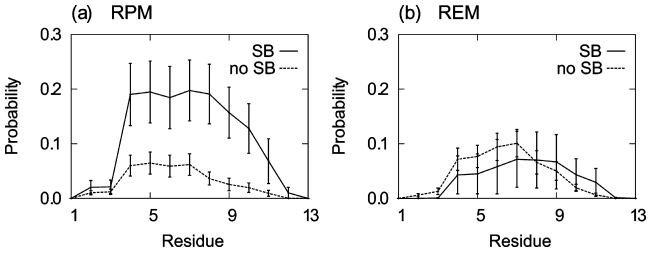


Figure 10. Probabilities of α-helix structures at 281 K for each residue in the (a) RPMD and (b) REMD simulations. Solid line and dashed line show the results with and without SBs between Glu2 and Arg10, respectively.

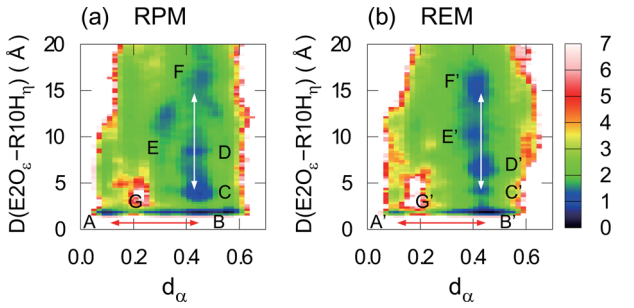


Figure 11. Free-energy landscapes at 281 K obtained from the (a) RPMD and (b) REMD simulations. The abscissa is the dihedral-angle distance with respect to the reference α-helix structure. The ordinate is the distance between O_ε of Glu2 and H_η of Arg10. The unit of the free-energy landscapes is kcal/mol. The labels A to F and A′ to F′ show the local-minimum free-energy states. The labels G and G′ are the transition states between states A and B and between states A′ and B′, respectively.

abscissa is the dihedral-angle distance d_α with respect to a reference α-helix structure. Here, a dihedral-angle distance d_α is defined by

$$d_\alpha = \frac{1}{n\pi}\sum_{i=1}^{n}\delta(v_i, v_i^0) \tag{41}$$

where n is the total number of dihedral angles, v_i is the dihedral angle i, and v_i^0 is the dihedral angle i of the reference conformation. The distance $\delta(v_i,v_i^0)$ between two dihedral angles is given by

$$\delta(v_i, v_i^0) = \min(|v_i - v_i^0|, 2\pi - |v_i - v_i^0|) \tag{42}$$

For d_α, only the backbone-dihedral angles of the residues 4−10 were employed as the elements in eq 41. We set the value of v_i^0 to $(\varphi,\psi) = (-\pi/3,-\pi/3)$. When the value of d_α is close to 0, therefore, C-peptide has an α-helix structure. The ordinate is the distance between O_ε of Glu2 and H_η of Arg10: D(E2O$_\varepsilon$–R10H$_\eta$). Here, this distance is defined to be the smallest among the four sets of the distances between two O_ε atoms of Glu2 and two H_η atoms of Arg10. Six local-minimum free-energy states are observed in both RPM and REM as shown in Figure 11a and b. We label these local-minimum states as A to F for RPM and A′ to F′ for REM. The transition states between states A and B and between states A′ and B′, are labeled as G and G′, respectively. The α-helix structure with the SBs as in Figure 9 corresponds to a structure in state A or A′. This fact means that the α-helix structure with the SBs is a stable structure. On the other hand, the α-helix structure without the SBs is not a stable structure because there are no local-minimum states for this structure. Therefore, the SBs play an important role in stabilizing the α-helix structure.

The global-minimum state at 281 K in RPM is state A, while that in REM is state B′. To see the reason for this difference, we counted the number of visiting times in state A for RPM and in state A′ for REM at the lowest temperature during the simulations. The region from 0.00 to 0.17 for d_α and from 1.0 Å to 2.2 Å for D(E2O$_\varepsilon$–R10H$_\eta$) was assigned to state A and state A′ here. The number of visiting times for each method is listed in Table 5. Here, we employed two criteria to count the number of visiting times. In criterion 1, when a replica visited in state A (or A′) at the lowest temperature after the replica had sampled d_α larger than 0.25 or D(E2O$_\varepsilon$–R10H$_\eta$) larger than 3.2 Å at the lowest temperature, it is counted as one visit. In criterion 2, sampling D(E2O$_\varepsilon$–R10H$_\eta$) larger than 3.2 Å was not taken into account. Therefore, the number of visiting times increases only when a conformation is changed from a nonhelical structure to a helix structure with the SBs in criterion 2. In addition to this, breaking and forming of the SBs also increases the number of visiting times in criterion 1. As shown in Table 5, the numbers of visiting times in RPM are much larger than those in REM in both criteria. In criterion 2, especially, the number of visiting times is 1 in REM. Because of such insufficient sampling in state A′, this state was underestimated in REM. This underestimation caused the small probabilities of the α-helix structures with the SBs in Figure 10b.

Table 5. The Numbers of Visiting Times in State A for RPM and in State A′ for REM

method	criterion 1	criterion 2
RPM	14 ± 8	5 ± 1
REM	4 ± 2	1 ± 0

The representative conformations for each state in Figure 11a are shown in Figure 12. The representative conformation for the transition state G is also presented. From these conformations and the free-energy landscape, we clarify a folding pathway from an extended structure in state F to the α-helix structure with the SBs in state A. Because the pathways in RPM and REM are almost the same, we will discuss that only in RPM: (State F to C via states E and D) The extended structure of C-peptide in state F changes to a globular structure as the side chains of Gly2 and Arg10 get close together. Turn structures are formed between Gly2 and Arg10 by this conformational change, as in states D and E. In state C the antiparallel β-bridge structure is occasionally created between Gly2 and Arg10. (State C to B) The SBs are formed between Gly2 and Arg10 by coming close together. Antiparallel β-bridges between Thr3 and Lys7 or between Leu9 and His12 are occasionally observed in state B. (State B to A via state G) In state G, a short α-helix or 3_{10}-helix structure is formed around residues 7−10, while maintaining the SBs. By growing this helix structure, the longer α-helix structure with the SBs is formed between residues 4 and 11 as in state A.

This folding pathway can be divided into two steps. The first step is the "salt-bridge formation step," and the second step is the "α-helix formation step." The first step and second step correspond to the transition from state F to B and that from B to A, respectively. These steps are drawn with the white and red arrows in Figures 11 and 12. We can also see in Figure 11 that

dx.doi.org/10.1021/ct3007919 | J. Chem. Theory Comput. 2013, 9, 570−581

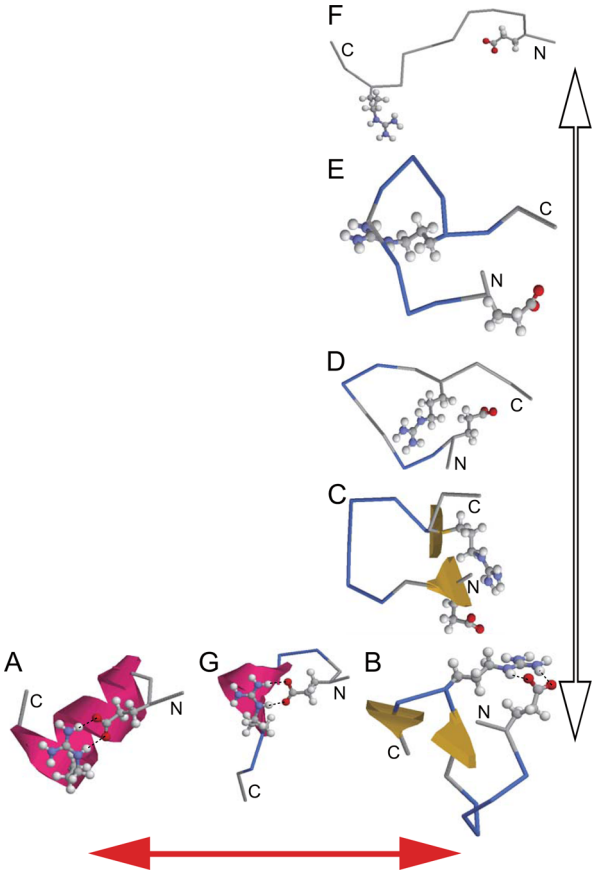


Figure 12. The representative conformations of the local-minimum free-energy states A–F and the transition state G obtained by the RPMD simulation. The dotted lines denote the SBs between Glu2 and Arg10. Backbone blue color shows turn structures. The figures were drawn with RasMol.[53]

C-peptide rarely takes a folding pathway in which the salt-bridge is formed after the α-helix structure formation.

V. CONCLUSIONS

We proposed the replica-permutation method (RPM), in which the replicas are allowed to transit not only neighboring temperatures but also non-neighboring temperatures. For replica-permutation trials in this method, the Suwa–Todo algorithm was employed instead of the Metropolis algorithm. This is because most of the permutation trials are rejected in the Metropolis algorithm. The Suwa–Todo algorithm had been proposed originally to minimize average rejection ratios for state transitions in MCMC. We applied RPM and M-RPM to the particles with the double-well potential energy to clarify the usefulness of the Suwa–Todo algorithm for the replica permutations. For comparison purposes, REMD simulations were also performed. As a result, RPM realized the most efficient sampling in the temperature space, while replica permutations were hardly accepted in M-RPM.

We also applied RPM and REM to Met-enkephalin in a vacuum. RPM sampled the temperature space more efficiently than REM even in the biomolecular system. The five local-minimum free-energy states were obtained at 200 K in both methods by using the reweighting techniques. In the free-energy landscape estimated from the raw histogram in REM, however, two of the five local-minimum states were not observed. This is because REM did not sample these two states at the lowest temperature. On the other hand, RPM sampled all states even at the lowest temperature. It indicates that RPM realized efficient sampling not only in the temperature space but also in the conformational space.

Furthermore, the RPMD and REMD simulations were performed for the C-peptide analog in explicit water to see the usability of RPM for a larger biomolecular system. RPM showed higher sampling efficiency in the temperature space, again.

It is reported in experiments that C-peptide has the α-helix structure with the SBs between Gly2 and Arg10. We observed the α-helix structures in both simulations. We also showed that the SBs play an important role in stabilizing the α-helix structure. From the free-energy landscape, furthermore, the folding pathway from the extended structure to the α-helix structure with the SBs was clarified. This folding pathway consists of the two steps. The first step is the "salt-bridge formation step." In this step, the SBs are formed by changing its conformation from the extended structure to the globular structures. The second step is the "α-helix formation step." The α-helix structure is formed while maintaining the SBs.

We thus revealed that RPM realizes more efficient sampling in the conformational space at the low temperature than REM. Furthermore, because the transition ratios of the replicas in RPM were larger than those in REM at all temperatures for all systems, larger temperature intervals can be taken in RPM. Therefore, the number of replicas can be reduced.

Although only the results of the MD simulations of RPM were shown in this article, this method can be readily applied to the MC method. It is also straightforward to introduce this method to the multidimensional REM[50] (also called Hamiltonian REM[51]) and related methods.[52] We can enhance the sampling efficiency of these methods by replacing REM with RPM.

AUTHOR INFORMATION

Corresponding Author

*E-mail: itoh@ims.ac.jp, hokumura@ims.ac.jp.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The computations were performed on the computers at the Research Center for Computational Science, Okazaki Research Facilities, National Institutes of Natural Sciences.

REFERENCES

(1) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Biopolymers* **2001**, *60*, 96–123.
(2) Itoh, S. G.; Okumura, H.; Okamoto, Y. *Mol. Simul.* **2007**, *33*, 47–56.
(3) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
(4) Hoover, W. G.; Ladd, A. J. C.; Moran, B. *Phys. Rev. Lett.* **1982**, *48*, 1818–1820.
(5) Evans, D. J. *J. Chem. Phys.* **1983**, *78*, 3297–3302.
(6) Nosé, S. *Mol. Phys.* **1984**, *52*, 255–268.
(7) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.
(8) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
(9) Okumura, H.; Itoh, S. G.; Okamoto, Y. *J. Chem. Phys.* **2007**, *126*, 084103.
(10) Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604–1608.
(11) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(12) Hansmann, U. H. E.; Okamoto, Y.; Eisenmenger, F. *Chem. Phys. Lett.* **1996**, *259*, 321−330.
(13) Nakajima, N.; Nakamura, H.; Kidera, A. *J. Phys. Chem. B* **1997**, *101*, 817−824.
(14) Berg, B. A.; Neuhaus, T. *Phys. Lett. B* **1991**, *267*, 249−253.
(15) Berg, B. A.; Neuhaus, T. *Phys. Rev. Lett.* **1992**, *68*, 9−12.
(16) Okumura, H.; Okamoto, Y. *Chem. Phys. Lett.* **2004**, *383*, 391−396.
(17) Okumura, H.; Okamoto, Y. *Phys. Rev. E* **2004**, *70*, 026702.
(18) Okumura, H.; Okamoto, Y. *J. Phys. Soc. Jpn.* **2004**, *73*, 3304−3311.
(19) Okumura, H.; Okamoto, Y. *Chem. Phys. Lett.* **2004**, *391*, 248−253.
(20) Okumura, H.; Okamoto, Y. *J. Comput. Chem.* **2006**, *27*, 379−395.
(21) Berg, B. A.; Noguchi, H.; Okamoto, Y. *Phys. Rev. E* **2003**, *68*, 036126.
(22) Itoh, S. G.; Okamoto, Y. *Mol. Simul.* **2007**, *33*, 83−89.
(23) Okumura, H. *J. Chem. Phys.* **2008**, *129*, 124116.
(24) Okumura, H. *Phys. Chem. Chem. Phys.* **2011**, *13*, 114−126.
(25) Okumura, H. *Proteins* **2012**, *80*, 2397−2416.
(26) Suwa, H.; Todo, S. *Phys. Rev. Lett.* **2010**, *105*, 120603.
(27) Mitsutake, A.; Hansmann, U. H. E.; Okamoto, Y. *J. Mol. Graphics Modell.* **1998**, *16*, 226−238.
(28) Itoh, S. G.; Okamoto, Y. *Chem. Phys. Lett.* **2004**, *400*, 308−313.
(29) Itoh, S. G.; Okamoto, Y. *J. Chem. Phys.* **2006**, *124*, 104103.
(30) Itoh, S. G.; Okamoto, Y. *Phys. Rev. E* **2007**, *76*, 026705.
(31) Shoemaker, K. R.; Kim, P. S.; Brems, D. N.; Marqusee, S.; York, E. J.; Chaiken, I. M.; Stewart, J. M.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82*, 2349−2353.
(32) Shoemaker, K. R.; Kim, P. S.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Nature* **1987**, *326*, 563−567.
(33) Osterhout, J. J., Jr.; Baldwin, R. L.; York, E. J.; Stewart, J. M.; Dyson, H. J.; Wright, P. E. *Biochemistry* **1989**, *28*, 7059−7064.
(34) Fairman, R.; Shoemaker, K. R.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biophys. Chem.* **1990**, *37*, 107−119.
(35) Lim, D.; Moye-Sherman, D.; Ham, I.; Jin, S.; Scholtz, J. M.; Burgess, K. *Chem. Commun.* **1998**, 2375−2376.
(36) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63*, 1195−1198.
(37) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(38) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(39) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712−725.
(40) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311−1327.
(41) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(42) Ewald, P. *Ann. Phys.* **1921**, *369*, 253−287.
(43) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, U. K., 1987.
(44) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(45) Quenouille, M. H. *Biometrika* **1956**, *43*, 353−360.
(46) Miller, R. G. *Biometrika* **1974**, *61*, 1−15.
(47) Berg, B. A. *Markov Chain Monte Carlo Simulations and Their Statistical Analysis*; World Scientific: Singapore, 2004.
(48) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577−2637.
(49) Acevedo, O. E.; Lareo, L. R. *OMICS* **2005**, *9*, 391−399.
(50) Sugita, Y.; Kitao, A.; Okamoto, Y. *J. Chem. Phys.* **2000**, *113*, 6042−6051.
(51) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058−9067.
(52) Itoh, S. G.; Okumura, H.; Okamoto, Y. *J. Chem. Phys.* **2010**, *132*, 134105.
(53) Sayle, R. A.; Milner-White, E. J. *Trends Biochem. Sci.* **1995**, *20*, 374−376.