Efficient Conformational Sampling of Proteins Using Fragment-Based Replica-Exchange Method with Adaptive Parameter Tuning

Masaaki Suzuki

***Abstract*—The native structure of a protein corresponds to the global minimum of free energy. The replica-exchange method (REM) has recently been used to search for the energy minimum in a wide range of protein conformations. For large systems, however, applying REM can be costly because the number of replicas required for conformational sampling increases. In the study reported here, a variant of multidimensional REM with automatic parameter tuning was developed. The proposed REM enables us to reduce the number of replicas needed for the simulation, using the fragment-based expression for the Hamiltonian. In our fragment-based REM, we defined the residue fragments and introduced a replica-exchange process based on the potential energy of the "target fragment". The number of replicas needed in the fragment-based REM depends solely on the number of degrees of freedom in the target fragment rather than on the size of the whole system, so the shortcoming of the conventional REM is overcome. Although there is a one-to-one correspondence between replicas and temperatures in the conventional REM, the fragment-based REM treats the fragment size as a second parameter characterizing the replicas. Replica groups with a small fragment size accelerate the local structure refinement, and those with a large fragment size accelerate the global structure refinement. In this study, an automatic tuning scheme, which provides appropriate fragment size adaptively, has also been developed. Computing equilibrium distributions for peptides, we found that the proposed REM successfully provides appropriate fragment length and thus good conformational sampling performance.**

***Index Terms*—protein structure prediction, conformation sampling, replica-exchange method, protein fragment, adaptive parameter tuning.**

I. Introduction

PROTEINS are biopolymers in which between a few dozen and several thousand amino acids are linked together in a chain. A protein folds into its unique native conformation under physiological conditions. Its biological functions are closely related to its conformation; hence, the analysis of protein tertiary structure is very important for post-genomic research.

It is widely believed that molecular simulation will become a powerful tool for elucidating the molecular mechanisms behind protein folding. Many researchers have recently attempted to predict protein structures using computer simulations along with X-ray structure analysis and other experiments. Computational protein structure prediction is based on the following hypotheses, known as Anfinsen's dogma [1]:

- Protein tertiary structure is determined solely by the amino acid sequence information.
- Native protein structure corresponds to the global minimum of free energy.

Therefore, assuming that the energy function of the system is given accurately, the native conformation of proteins can be calculated using methods similar to optimization algorithms, the Monte Carlo (MC), or molecular dynamics (MD).

Predicting the structure of a protein solely from its amino acid sequence remains a difficult challenge. One of the reasons for this difficulty is the need for an efficient sampling method that enables searching for the energy minimum among a huge range of conformations. Because the number of possible conformations for each protein is immense, each protein can exist in many states corresponding to the local energy minima. Because the thermal fluctuations at low temperatures are small, conventional MD simulations in the canonical ensemble would be trapped in one of these states. One way to tackle this multiple-minima problem is to perform a simulation based on the non-Boltzmann probability weight factors so that a random walk in the energy space can be realized. Compared with the conventional methods, the random walk allows the simulation to pass any energy barrier and to sample a much wider phase space. Of these methods, the replica-exchange method (REM) is well suited to large-scale parallel computing.

REM was originally developed by Hukushima and Nemoto [2] for spin glass simulation in the framework of MC simulations. Sugita and Okamoto [3] applied REM to protein systems and combined REM with MD (see [4] for a detailed description of the algorithm). REM considers a number of non-interacting copies (or replicas) of the original system in the canonical ensemble at different temperatures. An REM simulation is realized by alternately performing the following two steps: (i) each replica is simulated simultaneously and independently; and (ii) the pairs of replicas at neighboring temperatures are exchanged based on Metropolis-type criteria. In this way, simulations can escape from metastable states, but, as explained in Section II.B, applying REM to large systems becomes costly because the number of replicas required for conformational sampling increases.

To overcome this difficulty, we have developed a variant of REM called fragment-based REM, which is based on the idea that there are correlations between the local amino acid sequence and the local structure [5]. The number of replicas needed in the fragment-based REM depends solely on the number of degrees of freedom in the target residue fragment focused on, rather than on the size of the whole system, and so the shortage of the conventional REM is overcome.

Manuscript received January 8, 2018.

M. Suzuki is with the Department of Industrial Administration, Tokyo University of Science, Noda, Chiba 278-8510, Japan. e-mail: m-suzuki@rs.tus.ac.jp.

ISBN: 978-988-14047-8-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

In this study, an automatic tuning scheme, which provides appropriate fragment size adaptively, has been developed because the fragment size is the dominant factor for efficient conformational sampling. Using peptide simulations, we compare the sampling performance among conventional REM, fragment-based REM, and fragment-based REM with adaptive fragment-length.

II. Replica-Exchange Method

A. Replica-Exchange Molecular Dynamics Method [3][4]

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

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

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

$$W_B(\mathbf{x}) = \exp\{-\beta H(\mathbf{q}, \mathbf{p})\} \tag{2}$$

The system for REM consists of M non-interacting copies of the original system in the canonical ensemble at M different temperatures T_m $(m = 1, 2, \ldots, M)$. Let $\mathbf{X} = (\mathbf{x}^{[1]}_{m(1)}, \mathbf{x}^{[2]}_{m(2)}, \ldots, \mathbf{x}^{[M]}_{m(M)})$ represent a state in this generalized ensemble. The state $\mathbf{X}$ is specified by the M sets of coordinates $\mathbf{q}^{[i]}$ and momenta $\mathbf{p}^{[i]}$ of N atoms in replica i $(i = 1, 2, \ldots, M)$ at temperature T_m:

$$\mathbf{x}^{[i]}_m \equiv (\mathbf{q}^{[i]}, \mathbf{p}^{[i]})_m \tag{3}$$

Because the replicas are non-interacting, the weighting factor for the state X in this generalized ensemble is given by the product of Boltzmann factors for each replica:

$$W_{\mathrm{REM}}(\mathbf{X}) = \exp\left\{-\sum_{i=1}^{M} \beta_{m(i)} H(\mathbf{q}^{[i]}, \mathbf{p}^{[i]})\right\} \tag{4}$$

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

$$\mathbf{X} = (\ldots, \mathbf{x}^{[i]}_m, \ldots, \mathbf{x}^{[j]}_n, \ldots) \rightarrow \mathbf{X}' = (\ldots, \mathbf{x}^{[j]'}_m, \ldots, \mathbf{x}^{[i]'}_n, \ldots) \tag{5}$$

For this exchange process to converge toward the equilibrium distribution, it is sufficient to impose the detailed balance condition on the transition probability $w(\mathbf{X} \rightarrow \mathbf{X}')$ as:

$$W_{\mathrm{REM}}(\mathbf{X})\ w(\mathbf{X} \rightarrow \mathbf{X}') = W_{\mathrm{REM}}(\mathbf{X}')\ w(\mathbf{X}' \rightarrow \mathbf{X}) \tag{6}$$

From Equations (1), (4) and (6), we obtain

$$\frac{w(\mathbf{X} \rightarrow \mathbf{X}')}{w(\mathbf{X}' \rightarrow \mathbf{X})} = \exp(-\Delta) \tag{7}$$

where

$$\Delta \equiv [\beta_n - \beta_m]\left(E(\mathbf{q}^{[i]}) - E(\mathbf{q}^{[j]})\right) \tag{8}$$

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

$$w(\mathbf{X} \rightarrow \mathbf{X}') \equiv w(\mathbf{x}^{[i]}_m | \mathbf{x}^{[j]}_n) = \begin{cases} 1 & (\Delta \le 0) \\ \exp(-\Delta) & (\Delta > 0) \end{cases} \tag{9}$$

An REM simulation is realized by alternately performing the following two steps:

1. Computation of each replica is performed simultaneously and independently for a certain number of MD steps.
2. Pairs of replicas at neighboring temperatures are exchanged with the transition probability given by Equation (9).

B. Issues for Large-Scale REM

Exchanges between adjacent replicas should be frequently accepted for efficient conformational sampling. The relationship between the number of replicas and the degrees of freedom in the simulated system has previously been estimated [6]. Considering the potential energy fluctuations of two replicas sampling at each target temperature T_n and $T_n - 1$ (Figure 1), their instantaneous energy fluctuations δE_n and δE_{n-1} scale as $\sqrt{f}\ T_n$ and $\sqrt{f}\ T_{n-1}$, respectively. The average energy gap ΔE between the two neighboring replicas is proportional to $f\Delta T$. Here, f is the number of degrees of freedom and $\Delta T = T_n - T_{n-1}$. To obtain a reasonable acceptance ratio, the replica energy gap, ΔE, needs to be similar to the energy fluctuations δE_n and δE_{n-1}. Thus, $\Delta E/\delta E$ should be near unity. Because $\Delta E/\delta E$ is proportional to $\Delta T\sqrt{f}/T$ and hence $\Delta T \approx 1/\sqrt{f}$, the acceptable temperature gap between the neighboring replicas becomes narrower as the size of the system increases, and the number of replicas needed increases in proportion to $f^{1/2}$. This implies that, for larger systems, more and more replicas are needed and thus REM becomes impractical for very large systems such as proteins.

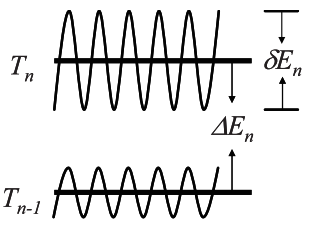

Fig. 1. Schematic Diagram Illustrating the Energy Fluctuations of Simulations at Two Temperatures for Neighboring Replicas

III. Development of Fragment-Based REM with Adaptive Fragment Length

A. Fragment-Based REM [5]

As previously mentioned, a complete amino acid sequence has a one-to-one correspondence with its unique native structure. In addition, the existence of a local correlation

ISBN: 978-988-14047-8-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

between the sequence and structure has been implied by studies on knowledge-based protein structure prediction [7].

Motivated by our desire to enhance the sampling of a fraction of the residue fragments of a protein, we proposed a replica-exchange process based on the potential energy of the "target fragment", F. In this study, we defined the residue fragments as depicted in Figure 2. Instead of preparing replicas with different temperatures, in our proposed method, fragment REM (FREM), we introduced replicas with different Hamiltonians. In FREM, the m-th replica is defined by the Hamiltonian $E_{m(i)}(\mathbf{q}^i)$:

$$E_{m(i)}(\mathbf{q}^i) \equiv S_{m(i)} \sum_{k \in F} E^k_{AA}(\mathbf{q}^i) + \sum_{k \notin F} E^k_{AA}(\mathbf{q}^i) \quad (10)$$

where $E^k_{AA}(\mathbf{q}^i)$ is the potential energy of the k-th amino acid $(k = 1, 2, \ldots, N_{AA})$ in replica i, and $S_{m(i)}$ is a scaling factor. This is followed by a procedure similar to that based on the original replica-exchange criterion, resulting in the following transition probability for replica-exchange:

$$w_{\mathrm{FREM}}(\mathbf{x}^i|\mathbf{x}^j) = \begin{cases} 1 & \text{for } \Delta_{\mathrm{FREM}} \leq 0 \\ \exp(-\Delta_{\mathrm{FREM}}) & \text{for } \Delta_{\mathrm{FREM}} > 0 \end{cases} \quad (11)$$

where

$$\Delta_{\mathrm{FREM}} \equiv \beta(S_{n(j)} - S_{m(i)}) \left\{ \sum_{k \in F} E^k_{AA}(\mathbf{q}^i) - \sum_{k \notin F} E^k_{AA}(\mathbf{q}^j) \right\} \quad (12)$$

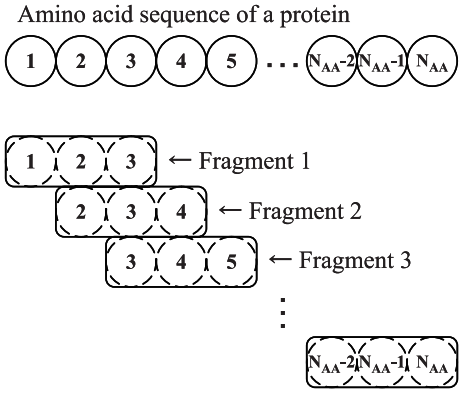

Fig. 2. Definition of Three-residue Length Fragments. Each Circle Represents an Amino Acid

Note that, when the target fragment size equals the total number of amino acids, and so the whole Hamiltonian is scaled (i.e., $E_m(\mathbf{q}) = S_m E(\mathbf{q})$), it precisely corresponds to the conventional REM with the inverse temperature $\beta_m = \beta S_m$; both give the same Boltzmann factor $\exp\{-\beta S_m E(\mathbf{q})\}$. In FREM, instead of flattening the energy landscape by scaling the whole Hamiltonian, only part of the Hamiltonian specific to the degrees of freedom in the target fragment is weakened by scaling. By doing this, we can change the "effective temperature" of the partial system. The number of replicas needed in FREM thus depends solely on the number of degrees of freedom in the target fragment rather than on the size of the whole system, and so the shortcoming of conventional REM that we mentioned in the previous chapter is overcome.

During the FREM simulation, as shown in Figure 3, only the randomly chosen target fragment is simulated over a range of "effective temperatures" (in other words, over a range of $S_m\beta$) with periodic exchanges performed according to Equations (11) and (12), while the remainder of the system is maintained at the same temperature for all replicas. After several hundred exchange attempts, the next target fragment is chosen randomly.

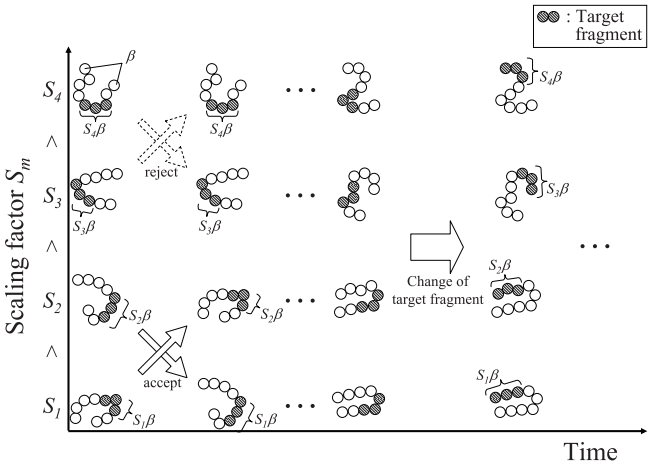

Fig. 3. Schematic Illustration of the FREM, with Four Replicas

Sugita et al. [8] proposed a multidimensional extension of REM in which the Hamiltonian of the system depends on a parameter, with different parameter values for different replicas. Fukunishi et al. [9] developed an alternative REM called Hamiltonian REM, which is a specific case of multidimensional REM. Although we employ a common approach, this is the first time that the fragment-based expression for the Hamiltonian (Equations (11) and (12)) has been used in this way.

Non-local interaction in the amino acid sequence is the dominant factor for the formation of β-sheet structures, whereas local hydrogen bonding is crucial for α-helices. Thus, it is anticipated that the fragmentation will not work well for a β-sheet structures.

The multiple fragment-size replica-exchange method (MFREM) is a two-dimensional extension of FREM for β-sheet-rich proteins. We treat the "fragment size" $N_\zeta(\zeta = 1, 2, \ldots, N^L_{\mathrm{rep}})$ as a second parameter characterizing the replicas and write the Hamiltonian for the i-th replica as:

$$E_{m,\zeta}(\mathbf{q}^i) \equiv S_m \sum_{k \in F}^{N_\zeta} E^k_{AA}(\mathbf{q}^i) + \sum_{k \notin F}^{N_{AA}-N_\zeta} E^k_{AA}(\mathbf{q}^i) \quad (13)$$

where $m = 1, 2, \ldots, N^T_{\mathrm{rep}}$ and the total number of replicas $N_{\mathrm{replica}} = N^T_{\mathrm{rep}} \times N^L_{\mathrm{rep}}$. Although replica i and the scaling factor S_m are in one-to-one correspondence in FREM, replica i and the parameter set (S_m, N_ζ) are in one-to-one correspondence in MFREM (see Figure 4).

In MFREM, the following replica-exchange processes are performed alternately:

ISBN: 978-988-14047-8-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

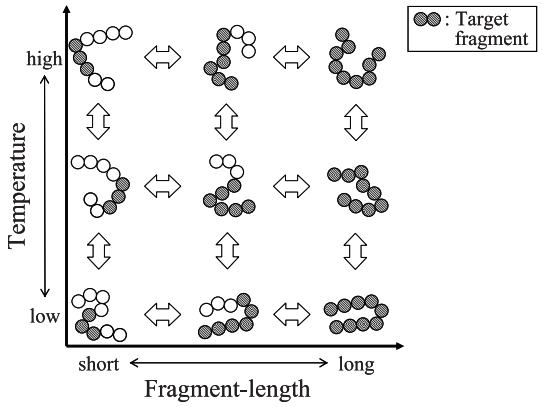

Fig. 4. Schematic Illustration of the MFREM, with Nine Replicas

1. Pairs of replicas corresponding to neighboring temperatures, (S_m, N_ζ) and (S_{m+1}, N_ζ) are exchanged. This is called "T-exchange".
2. Pairs of replicas corresponding to "neighboring" fragment size, (S_m, N_ζ) and $(S_m, N_{\zeta+1})$ are exchanged. This is called "L-exchange".

In each of these two processes, pairs of replicas are simultaneously exchanged, and the pairing is further alternated between the two possibilities. The acceptance criterion for these replica exchanges is given by equation (9), where equation (8) now becomes

$$\Delta \equiv \beta \left\{ \left[E_{m(i),\zeta}(\mathbf{q}^j) + E_{n(j),\zeta}(\mathbf{q}^i) \right] - \left[E_{m(i),\zeta}(\mathbf{q}^i) + E_{n(j),\zeta}(\mathbf{q}^j) \right] \right\} \quad (14)$$

for T-exchange, and

$$\Delta \equiv \beta \left\{ \left[E_{m,\zeta(I)}(\mathbf{q}^J) + E_{m,\eta(J)}(\mathbf{q}^I) \right] - \left[E_{m,\zeta(I)}(\mathbf{q}^I) + E_{m,\eta(J)}(\mathbf{q}^J) \right] \right\} \quad (15)$$

for L-exchange. In MFREM, replica groups with a small fragment size N_ζ accelerate the local structure refinement. On the other hand, replica groups with a large N_ζ accelerate the global structure refinement.

B. Influence of Fragment Length on Performance of Fragment-Based REM

In MFREM, fragment sizes are expected to have a strong influence on conformational sampling performance. Here, the sensitivity of MFREM's performance to the selection of fragment size was first evaluated using computation for a 16-residue poly-alanine chain as an example. In the numerical experiments, 16 replicas were simulated at eight different temperatures with two different fragment sizes (there are $N_{\mathrm{rep}}^L = 2$ fragment sizes at $N_{\mathrm{rep}}^T = 8$ temperatures, so that the total number of replicas is given by $N_{\mathrm{replica}} = N_{\mathrm{rep}}^T \times N_{\mathrm{rep}}^L = 16$) as depicted in Figure 5. We compared potential energy histograms for various combinations of fragment lengths. The parameters characterizing the replicas are summarized in Table 1. For all calculations, a distance-dependent dielectric, which mimics the presence of a solvent, was used. Electrostatic and Van del Waals interactions were calculated without using a distance-related cut-off, using the direct summation method. The MD time step was set to 0.5 fs. Starting from the fully extended conformation, we performed 250-ps equilibration MD, followed by a 1-ns production MD run. The AMBER ff94 forcefield [10] was used.

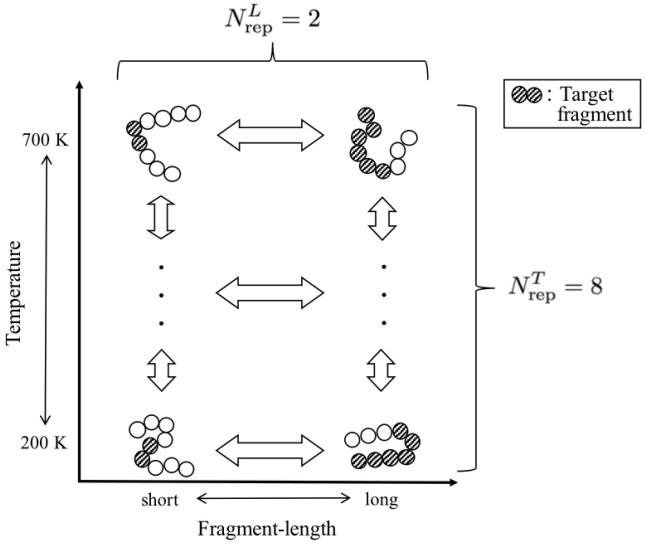

Fig. 5. Configuration of Replicas Used for Numerical Experiments

Median values of the total potential energy of poly-alanine at 200 K for each combination of the fragment size obtained by MFREM are shown in Table 2. Figure 6 provides histograms of the total potential energy of poly-alanine at 200 K obtained by conventional REM and MFREM. Here, for MFREM, the results with combination of the fragment size with the best (minimum) median value and with the worst (maximum) median value are shown. When appropriate fragment size was provided in MFREM, we succeeded in efficiently sampling low-energy structure with fewer replicas than conventional REM. On the other hand, it can be seen that the performance can be degraded compared with that of conventional REM if inappropriate fragment size is provided.

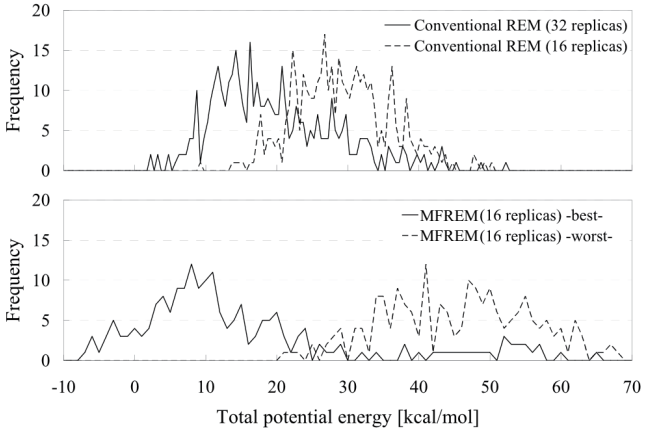

Fig. 6. Histogram of the Total Potential Energy of Poly-alanine at 200 K Obtained by Conventional REM (Upper) and by MFREM (Lower)

C. Adaptive Control of Fragment Length

Although the necessity of selecting an appropriate fragment length in MFREM has been indicated in the previous

ISBN: 978-988-14047-8-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE I
SUMMARY OF THE REPLICA PARAMETERS FOR THE α-HELIX PEPTIDE SIMULATIONS

	Conventional REM	MFREM
No. of replicas	16, 32	16 ($N^T_{\rm rep} = 8, N^L_{\rm rep} = 2$)
Temperatures[a] (K)	(200, 217, 236, 257, 279, 303, 330, 359, 390, 424, 461, 501, 544, 592, 643, 700) for 16 replicas (200, 208, 217, 226, 235, 245, 255, 265, 276, 288, 299, 312, 325, 338, 352, 366, 382, 397, 414, 431, 448, 467, 486, 506, 527, 549, 571, 595, 620, 645, 672, 700) for 32 replicas	(200, 239, 286, 342, 409, 489, 585, 700)
Replica-exchange interval (fs)	10	10
Fragment size (residue)	—	($N_{\xi=1}, N_{\xi=2}$) = (2,6), (2,10), (2,14), (2,18), (6,10), (6,14), (6,18), (10,14) (10,18), (14,18)

[a] For MFREM, "effective temperatures" of the target fragment. The remainder of the system is maintained at 200 K.

TABLE II
MEDIAN OF THE TOTAL POTENTIAL ENERGY (KCAL/MOL) OF POLY-ALANINE AT 200 K OBTAINED BY MFREM FOR EACH COMBINATION OF FRAGMENT SIZE

		Length of longer fragment (residues)			
		6	10	14	18
Length of shorter fragment (residues)	2	44.7	43.4	29.6	11.5
	6		43.7	28.3	10.0
	10			45.1	21.0
	14				45.5

section, in general, there are few precise and universal guidelines for configuring such hyper-parameters that govern the performance: their configuration is problem-specific. Additionally, the tuning of these parameters is not simple and generally requires many preliminary calculations. Furthermore, which fragment sizes are suitable may vary at every stage of the solution search. In this research, we implemented a function to adaptively select an appropriate fragment length in the process of structure search in MFREM. During the simulation, the update frequency of the best (minimum) total potential energy value is compared for each constant period between replica groups having different fragment lengths. In accordance with the comparison result, the fragment lengths of each replica group of the next period are determined as follows:

$$\begin{aligned} L_l(T+1) &= L_l(T) \\ L_s(T+1) &= L_s(T) + \Delta L \end{aligned} \quad (N_l(T) > c \cdot N_s(T)) \tag{16}$$

$$\begin{aligned} L_l(T+1) &= L_l(T) - \Delta L \\ L_s(T+1) &= L_s(T) \end{aligned} \quad (N_s(T) > c \cdot N_l(T)) \tag{17}$$

$$\begin{aligned} L_l(T+1) &= L_l(T) - \Delta L \\ L_s(T+1) &= L_s(T) + \Delta L \end{aligned} \quad (otherwise) \tag{18}$$

where $N_l(T)$ and $N_s(T)$ are the numbers of updates of the best (minimum) potential energy value in T phase in the replica group with a long fragment length and that in the replica group with a short fragment length, while $L_l(T)$ and $L_s(T)$ are the fragment lengths of the replica group having the long fragment length and that of the replica group having the short fragment length within the T phase, respectively. ΔL is the amount of increase or decrease of the fragment length, and $c \geq 1$ is a constant. For example, if the update frequency of the best (minimum) potential energy value in the replica group having a long fragment length is significantly greater than that of the replica group having a short fragment length, the fragment length of the replica group having the short fragment length is expanded. When the update frequencies of both replica groups are about the same, each fragment length is made longer or shorter.

IV. NUMERICAL EXPERIMENTS

THE sampling efficiency of the proposed methods is discussed using computations for an α-helical peptide (16-residue poly-alanine chain) and a β-hairpin peptide (10-residue Chignolin) as examples. For the β-hairpin peptide simulations, the AMBER ff96 forcefield [11] was used. Other settings are the same as in Section III.B.

The 16-residue poly-alanine peptide predominantly adopts a helical structure below ∼300 K. We computed the equilibrium distribution at 200 K using MFREM with and without adaptive control of fragment size, and compared the conformational sampling performance by examining the histogram of the total potential energy (Figure 7). Here, for MFREM without adaptive fragment size, the results with the combination of fragment size with the best (minimum) median value and with the worst (maximum) median value are shown. In the MFREM with adaptive fragment length (MFREM/AF) proposed in this research, the calculation result equivalent to the best case in the conventional MFREM is obtained without any preliminary calculations for parameter tuning. Time series of the lowest total potential energy of poly-alanine obtained by conventional MFREM and by MFREM/AF are shown in Figure 8.

We next performed a benchmark test on Chignolin. We performed a comparative study similar to that for poly-alanine. Figure 9 shows the histogram of the total potential energy of Chignolin at 200 K obtained by conventional MFREM and by MFREM/AF. As with experiments on an α-helix, MFREM/AF performs better because a superior quality of equilibrium distribution can be achieved without any preliminary runs. Figure 10 provides time series of the lowest total potential energy of Chignolin obtained by conventional MFREM with various combinations of fragment size ($N_{\xi=1}, N_{\xi=2}$) and by MFREM/AF. These analyses revealed that the proposed MFREM/AF successfully provides an appropriate fragment length and thus good conformational sampling performance.

ISBN: 978-988-14047-8-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

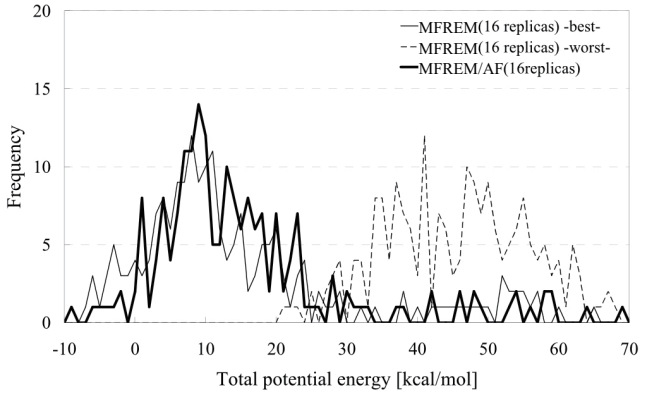

Fig. 7. Histogram of the Total Potential Energy of Poly-alanine at 200 K Obtained by Conventional MFREM and by MFREM/AF with 16 Replicas

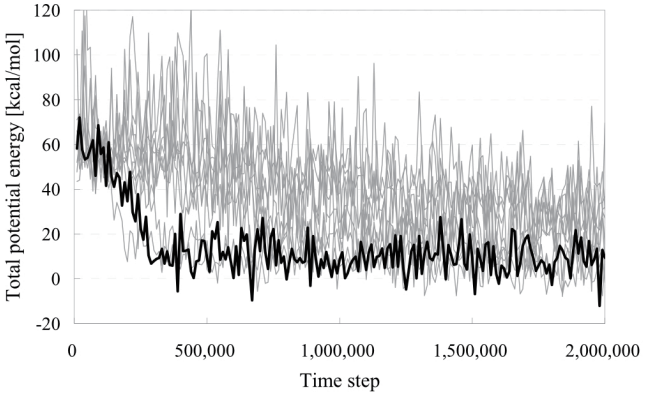

Fig. 8. Time Series of the Lowest Total Potential Energy of Poly-alanine Obtained by Conventional MFREM with Various Combinations of Fragment Size $(N_{\xi=1}, N_{\xi=2})$ (10 Gray Lines) and by MFREM/AF (Black Line) with 16 Replicas

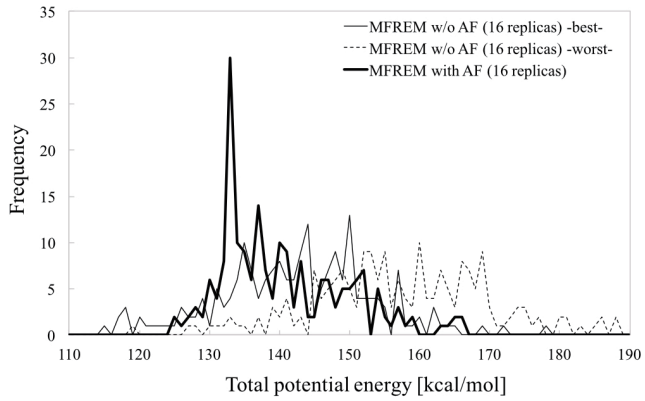

Fig. 9. Histogram of the Total Potential Energy of Chignolin at 200 K Obtained by Conventional MFREM and by MFREM/AF with 16 Replicas

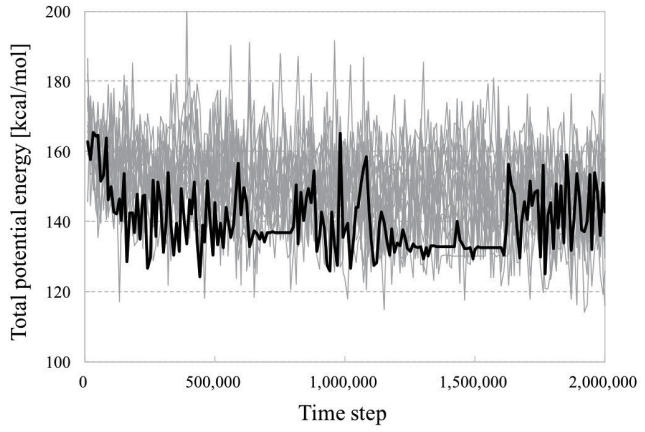

Fig. 10. Time Series of the Lowest Total Potential Energy of Chignolin Obtained by Conventional MFREM with Various Combinations of Fragment Size $(N_{\xi=1}, N_{\xi=2})$ (15 Gray Lines) and by MFREM/AF (Black Line) with 16 Replicas

V. Conclusion

A VARIANT of multidimensional REM with automatic parameter tuning has been developed. We have developed the fragment-based replica-exchange method, which is based on the existence of a correlation between the local amino-acid sequence and the local structure. We found that the proposed REM reduces the number of replicas needed for the simulation. Fragment size is the dominant factor for efficient conformation sampling. An automatic tuning scheme, which provides appropriate fragment size adaptively, has also been developed. By performing peptide folding simulations, we have found that the modified REM successfully provides appropriate fragment length and thus good conformational sampling performance.

References

[1] C. B. Anfinsen, "Principles that govern the folding of protein chains," *Science*, 181 (1973), pp. 223–230.
[2] K. Hukushima and K. Nemoto, "Exchange Monte Carlo method and application to spin glass simulations," *J. Phys. Soc. Jpn*, 65 (1996), pp. 1604–1608.
[3] Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding," *Chem. Phys. Lett.*, 314 (1999), pp. 141–151.
[4] A. Mitsutake et al., "Generalized-ensemble algorithms for molecular simulations of biopolymers," *Biopolymers (Pept. Sci.)*, 60 (2001), pp. 96–123.
[5] M. Suzuki and H. Okuda, "Fragment replica-exchange method for efficient protein conformation sampling," *Mol. Sim.*, 34 (2008), pp. 267–275.
[6] X. Cheng et al., "Modified replica exchange simulation methods for local structure refinement," *J. Phys. Chem. B*, 109 (2005), pp. 8220–8230.
[7] K. T. Simons et al., "Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions," *J. Mol. Biol.*, 268 (1997), pp. 209–225.
[8] Y. Sugita et al., "Multidimensional replica-exchange method for free-energy calculations," *J. Chem. Phys.*, 113 (2000), pp. 6042–6051.
[9] H. Fukunishi et al., "On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: application to protein structure prediction," *J. Chem. Phys.*, 116 (2002), pp. 9058–9067.
[10] W. D. Cornell et al., "Second generation force field for the simulation of proteins, nucleic acids and organic molecules," *J. Am. Chem. Soc.*, 117 (1995), pp. 5179–5197.
[11] D. A. Pearlman et al., "AMBER a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules," *Comput. Phys. Commun.*, 91 (1995), pp. 1–41.

ISBN: 978-988-14047-8-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)