Free-energy landscapes of proteins in solution by generalized-ensemble simulations

Yuji Sugita[1,2]

[1]Advanced Science Institute, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan, [2]CREST & BIRD, Japan Science and Technology Agency, 4-1-8 Honcho Kawaguchi, Saitama 332-0012, Japan

TABLE OF CONTENTS

1. ABSTRACT

Free-energy landscapes of proteins in solution are essential for understanding molecular mechanism of protein folding, stability, and dynamics. Because of the multiple-minima problem (or quasi-ergodicity problem), the conventional molecular dynamics or Monte Carlo methods cannot provide the landscapes accurately at low temperatures. By contrast, the simulations based on the generalized-ensemble algorithms can sample wider conformational spaces than the conventional approaches, thereby providing better free-energy landscapes of proteins at low temperatures. In this article, we review two well-known generalized-ensemble algorithms, namely, multicanonical algorithm and replica-exchange method, and then introduce further extensions of the above two methods, which are applicable to larger systems with rugged energy landscapes. These simulation methods have been applied to the protein folding simulations of the C-peptide in ribonuclease A with explicit solvent. We also demonstrate how the methods and the free-energy landscapes of proteins are useful for the biological research, by showing the simulation results on the phospholamban, a reversible regulator of sarco(end)plasmic reticulum Ca^{2+}-pump.

2. INTRODUCTION

Understanding molecular mechanisms of protein folding and stability is one of the central issues in molecular biology. Folding pathways of proteins have been explored with the experimentally detected intermediate states between the native and denatured states. Recent advances in experimental studies allow us to characterize the structures of these intermediate states, including the transition states in protein-folding reactions (1-3). In theoretical studies, a different view on protein folding has appeared recently (4, 5). In this view, free-energy landscapes of proteins play essential roles for understanding the folding kinetics: rapid folding of proteins occurs in smooth free-energy landscapes, whereas proteins with rough free-energy landscapes fold more slowly, causing kinetic traps during the folding processes. This view has been derived from the free-energy landscapes of simple lattice models for proteins and summarized to the principle of minimum frustration in protein structures by Wolynes and co-workers (4). Note that this principle is essentially the same as the consistency principle in proteins proposed by Go in which he claimed that maximum consistency among various interactions is realized in the native structure (6).

The free-energy landscapes of proteins from the simulations based on the all-atom representation of proteins and solvent molecules may be the most accurate ones, because all the interactions acting on the systems are taken into account in the simulations (7). However, due to the ruggedness of the landscapes at low temperatures, it is not straightforward to obtain them accurately by conventional molecular dynamics (MD) or Monte Carlo (MC) simulations. Unless the computational time is sufficiently long, simulations of proteins at low temperatures tend to get trapped at one of the local-energy-minimum states and can not sample the rest of the local-energy-minimum states. In most cases, it is difficult to obtain the global-energy-minimum state, which corresponds to the native structures of proteins, by the conventional simulation methods. To avoid the difficulty, we require much more powerful conformational sampling algorithms for proteins or other systems with rugged energy landscapes.

Generalized-ensemble algorithms have been developed for spin systems with strong frustration (8-11) and have been used for biological systems recently (12). The algorithms have two advantages over the conventional simulation methods. First, generalized-ensemble simulations can avoid trapping at one of the local-energy-minimum states by performing a random walk in the potential-energy space, implying that the simulations can sample wider conformational spaces than the conventional simulations. The random walk in the potential-energy space can be achieved by non-Boltzmann weight factor used in the algorithms. Secondly, we can provide canonical-ensemble averages at any temperatures precisely just from one long production run of the generalized-ensemble simulation, by combining the histogram-reweighting techniques (13-15). Over the last 15 years, various new algorithms that can be categorized in the generalized-ensemble algorithms have been developed by different researchers (12). Of these, multicanonical algorithm (MUCA) (8, 9) and replica-exchange method (REM) (10) may be two of the most widely used algorithms.

MUCA and REM have been shown to be very powerful in the simulations of small proteins or polypeptides in gas phase or in implicit solvent (16-20). However, the energy ranges that should be covered in these methods become very wide in larger biological systems, implying that a random walk in the potential-energy space may be difficult to achieve during the simulations. To solve the problem, we combined the merits of MUCA and REM, and developed new methods, namely, replica-exchange multicanonical algorithm (REMUCA) and multicanonical replica-exchange method (MUCAREM) (21-23). In this article, we first review the algorithms of MUCA and REM, and then, introduce the methods of REMUCA and MUCAREM. Finally, we demonstrate two recent applications of protein folding and stability by the generalized-ensemble simulations. In the first application, we discuss about the folding mechanism of the C-peptide in ribonuclease A in explicit solvent by using the free-energy landscape of the peptide (24). In the second application, we show the effect of phosphorylation on phospholamban

(PLN), a reversible regulator of sarco(end)plasmic reticulum Ca^{2+}-pump (SERCA), by comparing the free-energy landscapes of unphosphorylated (PLN) and phosphorylated PLN (pPLN) at room temperatures (25). These two applications indicate that the free-energy landscapes of proteins in explicit solvent by the generalized-ensemble simulations should have significant contributions for the research on biological sciences.

3. GENERALIZED-ENSEMBLE ALGORITHMS

3.1. Multicanonical algorithm

Multicanonical algorithm (MUCA) has been first developed by Berg et al. for the simulations of spin glasses (8, 9). In the algorithm, the multicanonical ensemble is based on the non-Boltzmann weight factor (multicanonical weight factor), $W_{mu}(E)$, which is the inverse of the density of states, $n(E)$:

$$W_{mu}(E) = \exp\left[-\beta_0 E_{mu}(E;T_o)\right] = \frac{1}{n(E)}$$ (1.1)

where $E_{mu}(E;T_0)$ is the so-called multicanonical potential energy function at arbitrary reference temperature, $T_0 = 1/k_B\beta_0$ (k_B is the Boltzmann's constant). The potential-energy distribution in the multicanonical ensemble, $P_{mu}(E)$, becomes constant, because it is given by the product of $W_{mu}(E)$ and $n(E)$:

$$P_{mu}(E) \propto n(E)W_{mu}(E) = \text{const.}$$ (1.2)

The flat distribution implies that a free random walk in the potential-energy space is realized in the ensemble, so that the simulations based on MUCA can escape from any local-minimum-energy states. Thus, the simulations based on the algorithm can sample wider conformational spaces than the conventional simulations.

MC or MD simulations based on MUCA can be performed just by replacing the total potential energy, E, with the multicanonical potential energy, $E_{mu}(E;T_0)$. In multicanonical MC simulations, the Metropolis criterion is modified as follows:

$$w(x \rightarrow x') = \begin{cases} 1 & \text{if } \Delta E_{mu} \leq 0 \\ \exp(-\beta_0 \Delta E_{mu}) & \text{if } \Delta E_{mu} > 0 \end{cases}$$ (1.3)

where

$$\Delta E_{mu} = E_{mu}(E';T_0) - E_{mu}(E;T_0)$$ (1.4)

In MD simulations, constant-temperature MD is performed by solving the modified Newton equation (26, 27):

$$\frac{d\mathbf{p}_k}{dt} = -\frac{\partial E_{mu}(E;T_0)}{\partial q_k} = \frac{\partial E_{mu}(E;T_0)}{\partial E}\mathbf{f}_k$$ (1.5)

where $\mathbf{f}_k$ is the usual force acting on the k-th atom.

In most cases, the density of states or multicanonical weight factor for the system is not *a priori* known, implying that we need the way to estimate the factor. In practice, $W_{mu}(E)$ is determined by iterations of short trial simulation runs. Once the optimal $W_{mu}(E)$ is determined, a long production simulation with the factor can be performed to obtain canonical-ensemble averages at any temperature T :

$$\langle A \rangle_T = \frac{\sum_E A(E)n(E)e^{-\beta E}}{\sum_E n(E)e^{-\beta E}} \tag{1.6}$$

where the best estimate of the density of states is given by the single histogram reweighting techniques (13):

$$n(E) = \frac{P_{mu}(E)}{W_{mu}(E)} \tag{1.7}$$

3.2. Replica-exchange method

Replica-exchange method (REM) has been developed for the simulations of spin systems by Hukushima et al. (10). Swendsen and co-workers independently developed a closely related method (28). Similar method, in which the same equation is used but emphasis is laid on optimization, has also been developed (29). Because of the history, REM is also referred to as multiple-Markov chain method (30) or parallel-tempering method (PT) (11).

In REM, a generalized-ensemble consists of M non-interacting copies of the original system (or replicas). Each replica has its own temperature, T_m ($m = 1, 2, …, M$). Then, the following two steps are alternately performed:

Step 1: each replica in canonical ensemble at the constant temperature is simulated for a certain MD or MC steps.
Step 2: a pair of replicas, for instance, i and j, which are at the neighboring temperatures T_m and T_n, are exchanged according to the Metropolis criteria:

$$w(x \rightarrow x') = \begin{cases} 1 & \text{if } \Delta \leq 0 \\ \exp(-\Delta) & \text{if } \Delta > 0 \end{cases} \tag{1.8}$$

where

$$\Delta = (\beta_m - \beta_n)(E_j - E_i) \tag{1.9}$$

Here, E_i and E_j are the potential energies of the i-th replica and j-th replica, respectively. Through the steps, each replica walks randomly in temperature space. In most cases, configurations at high temperatures have high-energy values, whereas those at low temperatures exist in low-energy regions. Therefore, a random walk in temperature space can achieve a random walk in the potential-energy space. Thus, the simulation based on REM

can escape from any energy barriers between local-energy-minimum states and sample wider conformational space like MUCA. Note that the transition probability, $w(x \rightarrow x')$, is satisfied with the detailed-balance conditions, because it is given by the Metropolis criterion like the conventional MC algorithm (10). It means that the probability distribution at each temperature should reach the canonical ensemble after long iterations of step 1 and step 2.

In MD version of REM (REMD), we also have to deal with the momenta of the systems. We have shown that rescaling the momenta uniformly by the square root of the ratio of T_m and T_n is able to cancel the kinetic energy terms in the Boltzmann factor (20). We can use the same transition probability in REMD as that in MC version of REM.

Like MUCA, the canonical-ensemble average at any temperatures can be obtained from the simulations based on REM by using the equation (1.6). Here, the best estimate of the density of states is given by the multiple-histogram reweighting techniques (14, 15):

$$n(E) = \frac{\sum_{m=1}^{M} g_m^{-1} N_m(E)}{\sum_{m=1}^{M} g_m^{-1} n_m e^{f_m - \beta_m E}} \text{ and } e^{-f_m} = \sum_E P(E;\beta_m) \tag{1.10}$$

where $P(E;\beta_m)$, $N_m(E)$, and n_m are the potential-energy distribution, the energy histogram, and the number of data for m-th replica, respectively. g_m is related to the integral autocorrelation time, τ_m, at temperature T_m as $g_m = 1 + 2\tau_m$. We usually set $g_m = 1$.

3.3. Extensions of MUCA and REM

These two algorithms work very well for small systems, such as small proteins or polypeptides in gas phase or in implicit solvent (16-20). However, both algorithms have their own weak points when they are applied to larger systems with rough free-energy landscapes. In MUCA, obtaining the optimal weight factor may be a very difficult and tedious task. In most cases, a large number of iterations are required to refine and update the weight factor before the production run. Wang-Landau method (31) is one of the efficient methods for determining the multicanonical weight factor. In contrast, no such effort is required in REM, because the weight factor is automatically determined in the algorithm. However, number of replicas that is required to keep sufficient probabilities of replica-exchanges increases rapidly as the system size becomes large. It implies that the simulations based on REM become very expensive for large systems, such as protein systems in explicit solvent.

To overcome the difficulties, we have developed two new methods, namely, replica-exchange multicanonical algorithm (REMUCA) and multicanonical replica-exchange method (MUCAREM) (22, 23). They are

hybrid approaches of MUCA and REM. In REMUCA, the multicanonical weight factor, $W_{mu}(E)$, or equivalently, the density of states, $n(E)$, is first estimated by using the potential-energy distributions of a short REM simulation as input data of the multiple-histogram reweighting techniques (14, 15). A production dynamics is performed after obtaining the optimized weight factor. In MUCAREM, a replica-exchange production run with replicas corresponding to multicanonical ensembles with different energy ranges is performed. The weight factors and exchange probabilities are given by,

$$E_{mu}^{\{m\}}(E) = \begin{cases} \left.\dfrac{\partial E_{mu}(E;T_m)}{\partial E}\right|_{E=E_L^{\{m\}}} (E - E_L^{\{m\}}) + E_{mu}(E_L^{\{m\}};T_m) & \text{for } E < E_L^{\{m\}} \\[2ex] E_{mu}(E;T_m) & \text{for } E_L^{\{m\}} < E < E_H^{\{m\}} \\[2ex] \left.\dfrac{\partial E_{mu}(E;T_m)}{\partial E}\right|_{E=E_H^{\{m\}}} (E - E_H^{\{m\}}) + E_{mu}(E_H^{\{m\}};T_m) & \text{for } E_H^{\{m\}} < E \end{cases}$$

(1.11)

and

$$w(x \rightarrow x') = \begin{cases} 1 & \text{if } \Delta \le 0 \\ \exp(-\Delta) & \text{if } \Delta > 0 \end{cases},$$

(1.12)

where

$$\Delta = \beta_m [E_{mu}^{\{m\}}(E_j) - E_{mu}^{\{m\}}(E_i)] - \beta_n [E_{mu}^{\{n\}}(E_j) - E_{mu}^{\{n\}}(E_i)].$$

(1.13)

Note that we need to evaluate the multicanonical potential energy, $E_{mu}^{\{m\}}(E_j)$ and $E_{mu}^{\{n\}}(E_i)$, because $E_{mu}^{\{m\}}(E)$ and $E_{mu}^{\{n\}}(E)$ are, in general, different functions for $m \ne n$. The same additional evaluation of the potential energy is necessary for the multidimensional replica-exchange method (MREM) (32) (or a special case of which is also referred to as Hamiltonian replica-exchange method (HREM) (33)).

In MUCAREM, a random walk in the potential-energy space is realized not only by the non-Boltzmann weight factor but also by the replica-exchange processes. Because the latter is a global update for the structures between the neighboring replicas, sampling efficiency of MUCAREM may be much improved compared with the original MUCA and REM. We have used MUCAREM in the simulation of G-peptide, a 16-residue polypeptide of the C-terminal end of streptococcal protein G B1 domain, in explicit solvent (34-36). In the simulation, only 8 replicas were required to cover the energy range corresponding to the temperatures between 275 K and 689 K, whereas at least 64 replicas were necessary for the original REM in the same system.

4. FOLDING SIMULATIONS OF THE C-PEPTIDES IN SOLUTION

Here, we show the efficiency of the new generalized-ensemble algorithms in protein-folding studies. Molecular mechanism of folding and stability for the C-peptide in ribonuclease A has been studied both by experimental and theoretical researchers. Circular Dichroism (CD) (37, 38) and nuclear magnetic resonance (NMR) (39) experiments have shown that the C-peptide analogs have a partially distorted alpha-helix structure at room temperatures. In their studies, two side-chain interactions have been pointed out to be essential for the alpha-helical structure: one is a salt-bridge interaction between Glu2 and Arg10, the other is an aromatic-aromatic interaction between Phe8 and His12 (37, 40). At neutral pH, side chain of Glu is negatively charged, whereas that of Arg is positive, suggesting that Glu2 and Arg10 may form a stable salt bridge by electrostatic interactions. In theoretical studies, the alpha-helical structure has been obtained from the random-coil structure or fully extended structure by simulated annealing in gas phase (41), or the simulations based on the generalized-ensemble algorithms (19, 22). However, the roles of the side-chain interactions pointed out by the experiments have not been fully examined.

We performed all-atom MD simulations of the C-peptide in explicit solvents (24) and then, analyzed the free-energy landscapes of the peptide at various temperatures. The simulations were based on REMUCA (21). The peptide analog comprises 13 amino-acid residues: Ala-Glu-Thr-Ala-Ala-Ala-Lys-Phe-Leu-Arg-Ala-His-Ala, in which the N and C-termini were blocked with the acetyl group and the N-methyl group, respectively. The initial configuration of the peptide was a fully extended conformation and was simulated at 1000 K in gas phase. Then, one of the structures in the simulation was solvated with 1387 water molecules in a sphere of radius 22 Å. AMBER parm99 (42, 43) and TIP3P model (44) were used for the force-field parameters of proteins and water molecules.

We first performed a REMD simulation with 32 replicas (REMD1) for 100 ps per replica, covering the temperature range from 250 K to 700 K. During this simulation, replica-changes were attempted every 200 MD steps. The potential-energy distributions were used as input data to the multiple-histogram analysis (14, 15) and then, we obtained the first estimate of $W_{mu}(E)$. We divided the factor into four factors that covered different energy regions. We then carried out a MUCAREM simulation with 4 replicas (MUCAREM) for 1 ns per replica, in which replica-exchange was attempted every 1000 MD steps. We again updated the estimate of $W_{mu}(E)$ by using the potential-energy distributions obtained in the MUCAREM simulation. Finally, we performed a 15-ns multicanonical MD run with 1 replica (REMUCA) and analyzed the simulation trajectory in detail. In Figure 1, the potential-energy distributions obtained from REMD, MUCAREM, and REMUCA are shown. Note that in REMD and MUCAREM, sufficient overlaps of the potential energy distributions between neighboring replicas are observed. In REMUCA, a flat distribution is obtained. These suggest that good random walks are realized in REMD, MUCAREM, and REMUCA.
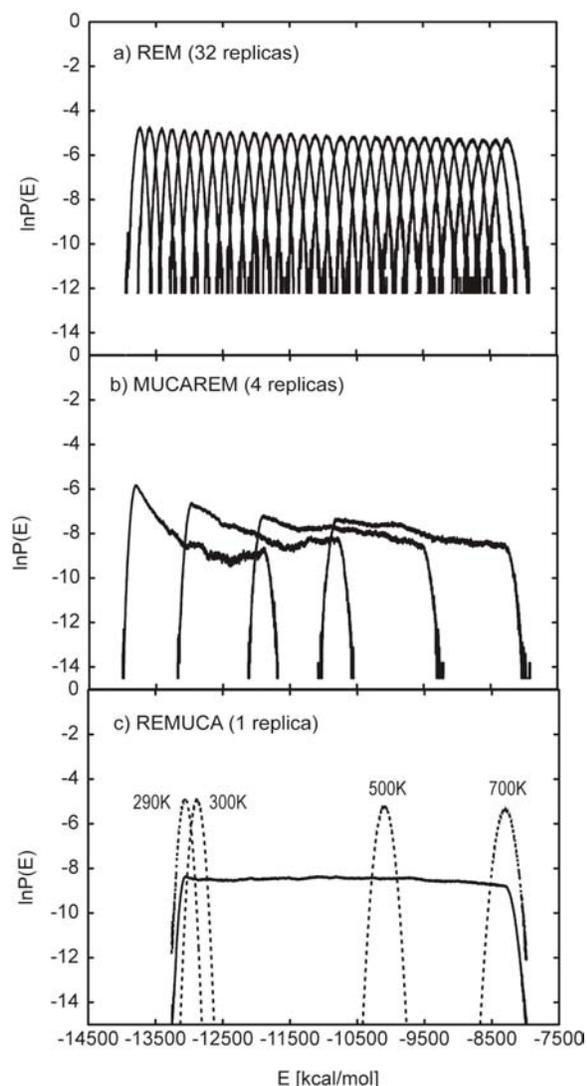
Figure 1. Probability distributions of potential energy of the C-peptide obtained from (a) REMD, (b) MUCAREM, and (c) REMUCA. Dashed curves in Figure 1(c) are the reweighted canonical distributions at 290, 300, 500 and 700 K, respectively (from left to right).

To construct the free-energy landscapes of the C-peptide at various temperatures, we performed principal component analysis (PCA), which is one of the standard methods for studying conformational dynamics or fluctuations of proteins (45-47). We used the first two principal component (PC) axes as reaction coordinates, because the first and second PCs contributed significantly to the total root-mean-square fluctuations of the proteins (22.8 and 17.0 %, respectively). To obtain the landscapes at different temperatures, we employed histogram-reweighting techniques and calculated the potentials of mean force along the first two PCs:

$$\Delta G(PC1, PC2; T) = -k_B T \ln P_B(PC1, PC2; T) , \qquad (1.14)$$

where PC1 and PC2 are the first two PC values. $P_B(PC1, PC2; T)$ is the "reweighted" canonical distribution at temperature T. In Figure 2, smooth free-energy landscapes were observed at 400 K and 500 K. In contrast, at 300 K, the landscape became much more rugged and consisted of three distinct local-minimum-energy states (LM1, LM2, and LM3), which were separated with high-energy barriers. Of these states, LM1 corresponded to the global free-energy minimum state (GM).

To characterize each LM, we focused on the conformations whose total potential energy was lower than -12,500 kcal/mol (including the contributions from water molecules). -12,500 kcal/mol was chosen, because it was the highest energy value that can be reached for the energy distribution at 300 K in Figure 1c. Figure 3 indicated the different characters of the LMs. In LM2, the peptide contained the salt-bridge interaction between Glu2 and Arg10, whereas two aromatic residues, Phe8 and His12, were separated. In contrast, in LM3, the two aromatic residues made contacts, whereas the salt-interaction was not observed. In LM1, which corresponds to the global free-energy-minimum state (GM), both interactions were observed. The structures in GM agreed with the X-ray structure of the C-peptide, which has a partially distorted alpha-helix with these two side-chain interactions.

The two side-chain interactions in the protein have quite different characteristics: the aromatic-aromatic interaction between Phe8 and His12 is a short-range interaction mainly caused by hydrophobic forces, whereas the salt interaction between Glu2 and Arg10 occurs due to the long-range electrostatic interactions. In LM2 and LM3, the C-peptide is likely trapped at collapsed local-energy-minimum states, where one of the major stabilizing factors in the native structure is lost. Thus, the simulation reveals a funnel-like shape of the free-energy landscape of the C-peptide, where all the important interactions consistently exist only in GM. This may be the mechanism how such a short peptide can fold into a stable conformation rather than a random-coil state in solution.

To understand the role of each side-chain interaction on the folding reaction of the C-peptide, the free-energy landscape is useful. In Figure 3, the green points (salt-bridge interaction) and the blue points (alpha-helix formation) do not overlap except for the region around GM. In contrast, the red points (aromatic-aromatic interaction) overlap significantly with the blue points, suggesting that the formation of alpha-helix is strongly correlated with the aromatic-aromatic interaction between Phe8 and His12. The salt interaction between Glu2 and Arg10 may contribute to prevent the structure from expanding to the N-terminus. Although the simulation trajectories of the generalized-ensemble simulations do not provide the kinetic information directly, careful analysis on the free-energy landscape can provide the information on how proteins fold from the random-coil or the denatured states.
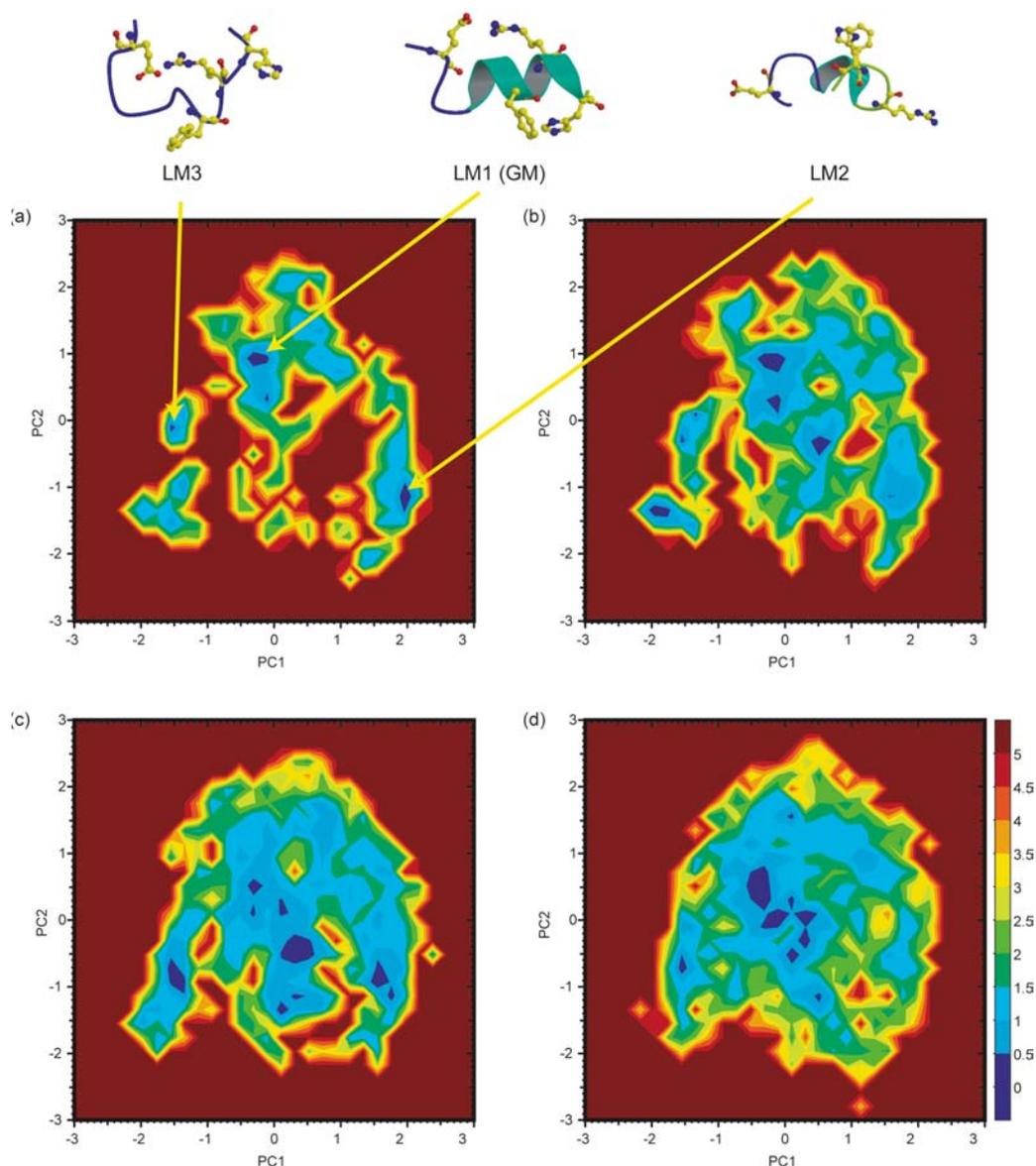
Figure 2. The potential of mean force of the C-peptide along the first two principal component axes at (a) 300, (b) 350, (c) 400 and (d) 500 K. LM1(GM), LM2 and LM3 represent three distinct local-energy-minimum states. Representative snapshot in LM1, LM2, and LM3 are also shown. The aromatic-aromatic interaction between Phe8 and His12 and the salt-bridge interaction between Glu2 and Arg10 are shown explicitly in ball-and-stick model.

5. CONFORMATIONAL CHANGES OF PHOSPHOLAMBAN BY PHOSPHORYLATION AT SER16

Phospholamban (PLN) is a 52-residue integral membrane protein that regulates the activity of the sarco(endo)plasmic reticulum calcium pump (SERCA) in cardiac muscle (48). Its inhibitory action is relieved when PLN is phosphorylated at Ser16 by cAMP-dependent protein kinase or at Thr17 by Ca^{2+}/calmodulin-dependent kinase (49, 50). However, molecular mechanism of how the phosphorylation can change the inhibitory action remained to be answered. To understand the mechanism, we studied the effect of phosphorylation on the structure of PLN. We

performed REMD simulations of the unphosphorylated form of PLN (PLN) and phosphorylated form of PLN (pPLN) in solution to explore the free-energy landscapes of the two forms (25).

In the simulations, we did not include the transmembrane region of PLN or pPLN, because the phosphorylation site (Ser16) is far from the transmembrane region of PLN. Thus, we used the first 20 amino-acid residues of PLN in the simulations, whose amino-acid sequence is Met-Glu-Lys-Val-Gln-Tyr-Leu-Thr-Arg-Ser-Ala-Ile-Arg-Arg-Ala-Ser-Thr-Ile-Glu-Met. The C-termini of the protein was blocked with an N-methyl group. The starting structure of PLN was taken from the corresponding region of
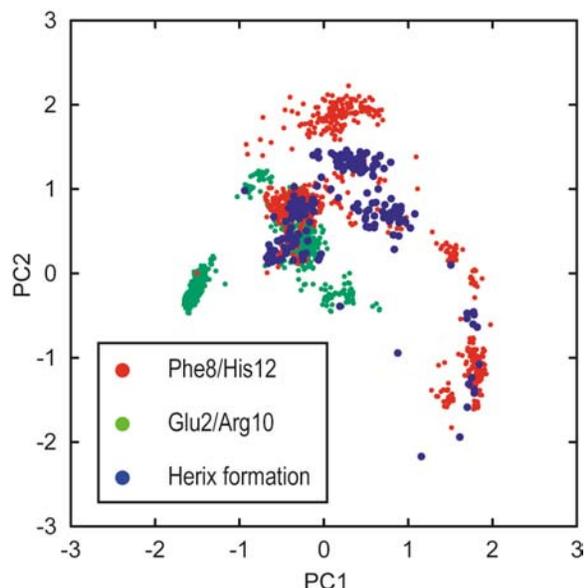
Figure 3. Interaction consistency in the free-energy landscape of the C-peptide system. Red, green, blue points represent conformations with the aromatic interaction between Phe8, those with the salt-bridge interaction between Glu2 and Arg10, and those with more than six helical residues defined by DSSP(62). Only the conformations whose total potential energy is less than -12,500 kcal/mol are considered.

the first conformer of the NMR structures reported by Zamoon et al. (51). For pPLN, the starting structure was prepared by simply changing the side chain of Ser16 to the phosphorylated form. Since the phosphate is negativelycharged at neutral pH, the phosphorylated Ser16 can form salt bridges with arginine residues in the peptide. PLN and pPLN were solvated in a sphere with a radius of 30 Å, in which 8 K^+, 11 Cl^- and 3405 water molecules, and 9 K^+, 10 Cl^- and 3399 water molecules were included, respectively. CHARMM27 CMAP potential function (52, 53) was used for proteins, and the TIP3P model (44) was used for water molecules. Other simulation details were written in the original paper (25).

We used sixty replicas with different temperatures from 290 K to 653 K. In the simulations of PLN and pPLN, we performed 8 ns and 10 ns of MD simulations for each replica, respectively, and attempted replica-exchanges every 100 steps. The total simulation lengths were 480 ns for PLN and 600 ns for pPLN. The coordinates of PLN or pPLN saved every 500 steps for the final 6 ns were used to construct the free-energy landscapes. To characterize the landscapes at 310 K, we used the first (PC1) and third (PC3) axes, because different local-energy-minimum states appeared more clearly in the PC1-PC3 plane than in the PC1-PC2 plane. Contributions to the mean square fluctuation from the PC2 and PC3 were similar (8.6 and 8.4 %, respectively).

In the free-energy landscape of PLN (Figure 4), three major clusters were found. In cluster I, PLN consisted of a long alpha-helix from Glu2 and Glu19, whereas the C-terminal part of the alpha-helix was unwound in clusters II

(between Thr17 and Glu19) and III (between Ile12 and Glu19), suggesting that the structural ensemble of PLN in solution contains the conformations similar to the atomic model of PLN bound to SERCA (54). In contrast, the free-energy landscape of pPLN exhibited a broader distribution comprising eight clusters, which were further classified into three groups. In the first group, pPLN consisted of a long alpha-helix from Lys3 to pSer16, and contained, at most, a single salt-bridge between pSer16 and Arg13. In the second group, two salt-bridges were formed between pSer16, and two arginines (Arg9 and Arg13). Due to the two salt bridges, the C-terminal part of the alpha-helix was unwound or divided into two short helices. The third group contained disordered conformations, in which a salt bridge was formed between pSer16 and Arg14. Thus, phosphorylation of PLN at Ser16 likely form salt-bridges with arginines and increased the number of conformational substates by distorting the alpha-helix.

The effect of phosphorylation at Ser16 of PLN has been extensively studied by experimentalists. However, nuclear magnetic resonance (NMR) (55) and fluorescence resonance electron transfer (FRET) (56, 57) did not provide a consistent picture of the effect of phosphorylation. NMR studies of a mutant of PLN (AFA-PLN) (55) showed that the phosphorylation unwinds the C-terminal part of the cytoplasmic alpha-helix and shifts the conformational equilibrium toward disordered states. The authors suggested that a stable salt bridge involving pSer16 is unlikely. In contrast, FRET studies suggested that pPLN takes more compact helical conformation stabilized by a salt bridge between pSer16 and Arg13 (56, 57). One of the major purposes of the simulation study is to explain this experimental inconsistency.

The simulation results on PLN and pPLN agree with the NMR measurements (55), because the phosphorylation increases the degree of conformational disorder as shown in the free-energy landscapes of PLN and pPLN (Figure 4). Average alpha-helicity of each residue in PLN and pPLN at 310 K in Figure 5 represents the degree of disorder more quantitatively. Although a direct comparison between FRET and the simulations is not possible, Figure 6 shows approximately 2 Å decrease in the Cα-distance between Tyr6 and Met20 upon the phosphorylation at Ser16 of PLN. It is quantitatively consistent with the results of FRET (56, 57), in which the distance between Tyr6 and a fluorophore, *N-(1-pyrenylmaleimide)*, attached to Cys24 decreased by 3 Å by the phosphorylation. The decrease was initially interpreted as an elongation of the cytoplasmic alpha-helix (56, 57). However, the free-energy landscapes of PLN and pPLN indicated that the decrease is due to the conformational disorder of alpha-helix upon the phosphorylation at Ser16. Thus, two experimental measurements and the simulation results are consistent with each other.

Then, how does the phosphorylation relieve the inhibitory effect of PLN on SERCA in cardiac muscle? The current simulation indicated that the phosphorylation of PLN at Ser16 unwinds the cytoplasmic helix by forming salt
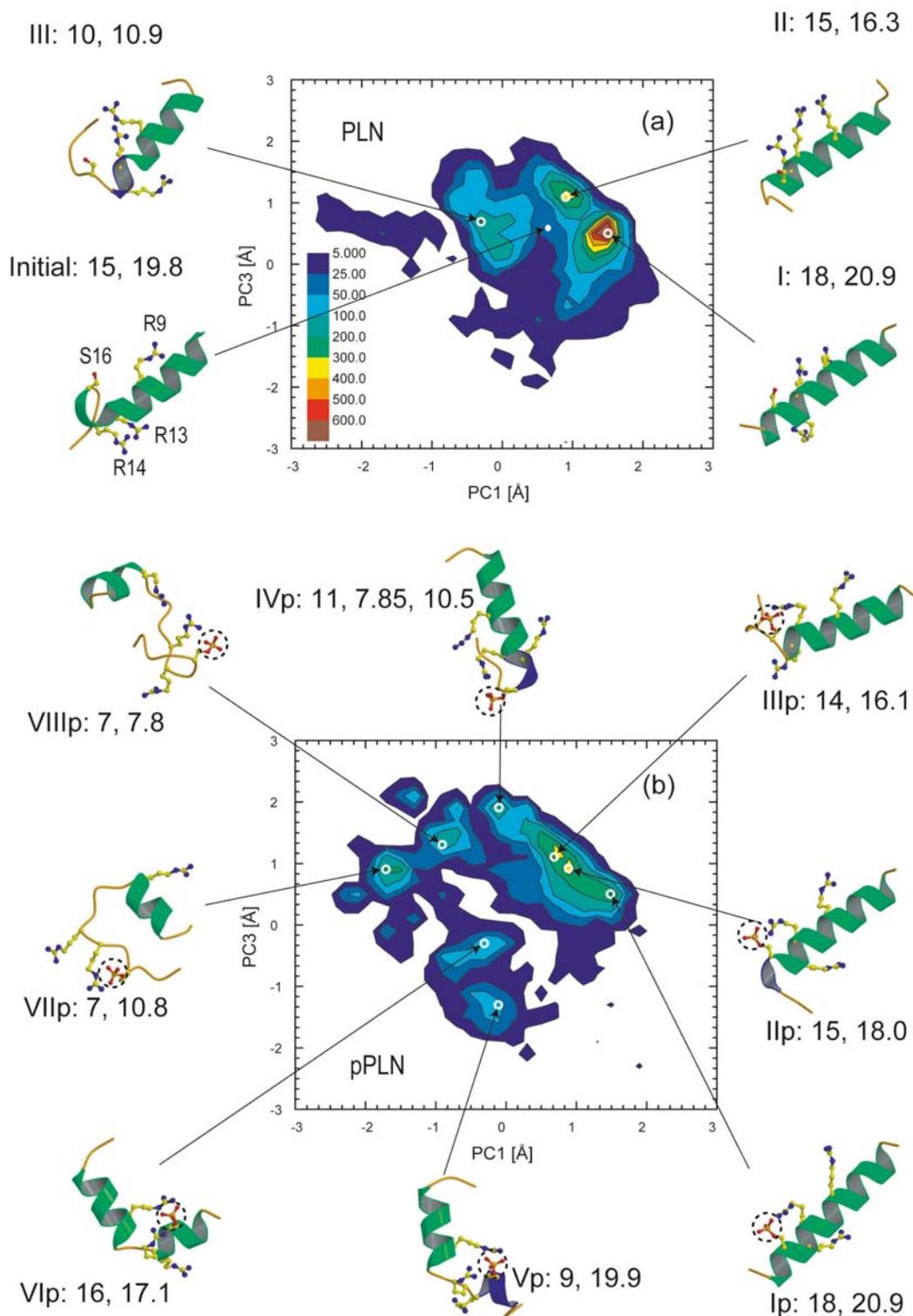
Figure 4. Free-energy landscapes of PLN (a) and pPLN (b) at 310 K in the REMD simulations. The starting conformation of PLN (white circle) and the representative stable conformations at local-energy-minimum (open circles) are also shown. The values are the number of alpha-helical residues and the Cα distance between Tyr6 and Met20 of the representative structure in each cluster. The side chains of Arg9, Arg13, Arg14, and Ser16 are represented in ball-and-stick form. The phosphate in pPLN was circled in Figure 4(b).
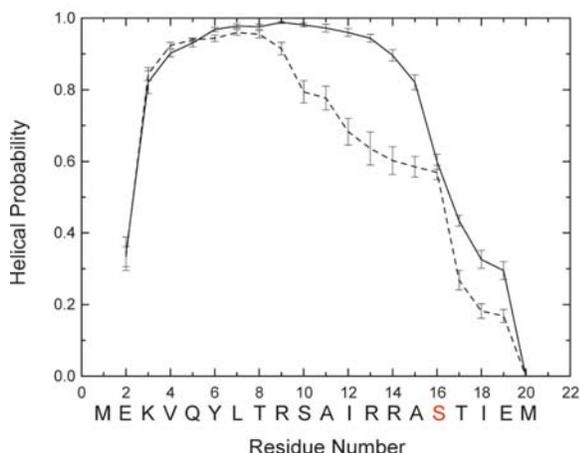
Figure 5. Average alpha-helicity of each residue in PLN (straight) and pPLN (dashed) at 310 K. An alpha-helical residue was defined by the same criteria that were used in DSSP(62). The error bars show the standard error of block averages that were calculated for every 1ns.
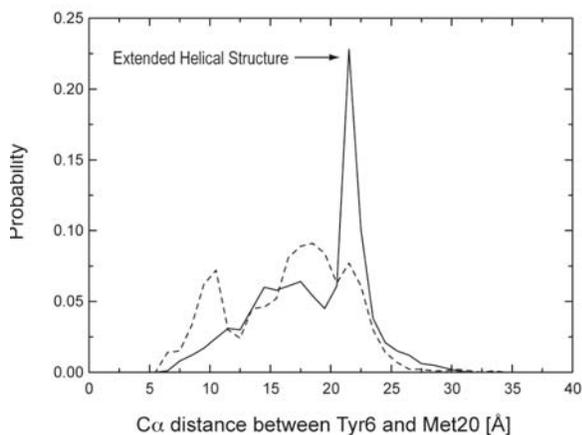
Figure 6. Distribution of the Cα distance between Tyr6 and Met20 at 310 K calculated in PLN (straight) and pPLN (dashed) in the REMD simulations.

bridges between pSer16 and three arginines. Based on the structural model (54), the crosslink between Lys3 in PLN and Lys397 or Lys400 in SERCA is possible only if the cytoplasmic domain takes an alpha-helical structure. These suggest that the conformational disorder upon the phosphorylation of PLN at Ser16 inhibits a stable binding of PLN to SERCA. To see the validity of the model, we should study the effect of phosphorylation of PLN bound to SERCA.

6. PERSPECTIVE

Recently, folding simulations of small polypeptides or proteins have been performed extensively by using the generalized-ensemble algorithms. In particular, REMD simulations with implicit solvent models

have been the most popular approach, because their computational cost is much smaller than the simulations with explicit solvent. Although implicit-solvent simulations have provided new insight on protein folding, the detailed description of the side-chain interactions during protein-folding processes may be difficult with the model. For instance, the implicit solvent model cannot describe the side-chain interactions that are bridged by a water molecule. In contrast, the simulations with explicit solvent can provide atomically detailed information on the side-chain interactions, as shown here.

Another important issue on protein folding and stability is the effect of salt ions on the conformational stability of proteins. In the simulations of PLN and pPLN, we added K^+ and Cl^- in the solvent sphere to mimic the experimental environment. However, due to the small size of the simulation system, we cannot study the details of the salt effect on the protein. In the near future, we may extend the size of the solvent sphere and study the salt effect on the protein stability in an atomic resolution. In such simulations, MUCAREM or REMUCA should be useful, because the number of replicas for a larger simulation system is still very small compared with the original REMD. The multi-dimensional replica-exchange method (MREM) (32), in which pre-defined reaction coordinates as well as temperatures are attempted to exchange during the simulations, (it is also referred to as Hamiltonian REM (33)) may be another candidate that are applicable to the large simulation systems. In MREM, a careful choice of the reaction coordinate is required before the production dynamics.

Toward better description of the free-energy landscapes of proteins, accuracy of the force-field parameters should also be essential. Yoda et al, have compared the structural tendency of six commonly used force-field parameters, by performing the generalized-ensemble simulations of C-peptide and G-peptide in explicit solvent (34, 36). The results indicated that alpha-helix is favored for AMBER94 (42) and AMBER99 (42, 43) and that beta-hairpin is favored for GROMOS96 (58), whereas CHARMM22 (52), AMBER96 (59), and OPLSAA/L (60, 61) have intermediate tendency. This result indicates that we should use different force fields depending on the folds (alpha, beta, alpha/beta, etc.) of the target proteins in their folding simulations from the first principles (from a random initial conformation). Therefore, further improvements on the force fields are necessary for studying the free-energy landscapes of protein folding or large-scale conformational changes in proteins.

7. ACKNOWLEDGEMENTS

8. REFERENCES

1. Dobson, C. M., A. Sali and M. Karplus: Protein folding: A perspective from theory and experiment. *Angewandte Chemie-International Edition* 37, 868-893 (1998)

2. Klein-Seetharaman, J., M. Oikawa, S. B. Grimshaw, J. Wirmer, E. Duchardt, T. Ueda, T. Imoto, L. J. Smith, C. M. Dobson and H. Schwalbe: Long-range interactions within a nonnative protein. *Science* 295, 1719-1722 (2002)

3. Fersht, A.: Structure and Mechanism in Protein Science. Freeman, New York (1999)

4. Onuchic, J. N., Z. LutheySchulten and P. G. Wolynes: Theory of protein folding: The energy landscape perspective. *Annu Rev Phys Chem* 48, 545-600 (1997)

5. Dill, K. A. and H. S. Chan: From Levinthal to pathways to funnels. *Nature Struct Biol* 4, 10-19 (1997)

6. Go, N.: Theoretical-Studies of Protein Folding. *Annu Rev Biophys Bioeng* 12, 183-210 (1983)

7. Boczko, E. M. and C. L. Brooks, 3rd: First-principles calculation of the folding free energy of a three-helix bundle protein. *Science* 269, 393-396 (1995)

8. Berg, B. A. and T. Neuhaus: Multicanonical Algorithms for 1st Order Phase-Transitions. *Phys Lett B* 267, 249-253 (1991)

9. Berg, B. A. and T. Neuhaus: Multicanonical Ensemble - a New Approach to Simulate 1st-Order Phase-Transitions. *Phys Rev Lett* 68, 9-12 (1992)

10. Hukushima, K. and K. Nemoto: Exchange Monte Carlo method and application to spin glass simulations. *J Phys Soc Jpn* 65, 1604-1608 (1996)

11. Marinari, E., G. Parisi and J. J. Ruiz-Lorenzo: Numerical simulations of spin glass systems. *Spin Glasses and Random Fields*, Young, A.P., Ed. 59-98 (1998)

12. Mitsutake, A., Y. Sugita and Y. Okamoto: Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers* 60, 96-123 (2001)

13. Ferrenberg, A. M. and R. H. Swendsen: New Monte-Carlo Technique for Studying Phase-Transitions. *Phys Rev Lett* 61, 2635-2638 (1988)

14. Ferrenberg, A. M. and R. H. Swendsen: Optimized Monte-Carlo Data-Analysis. *Phys Rev Lett* 63, 1195-1198 (1989)

15. Kumar, S., D. Bouzida, R. H. Swendsen, P. A. Kollman and J. M. Rosenberg: The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. *J Comp Chem* 13, 1011-1021 (1992)

16. Hansmann, U. H. E. and Y. Okamoto: Prediction of Peptide Conformation by Multicanonical Algorithm: New Approach to the Multiple-Minima Problem. *J Comp Chem* 14, 1333-1338 (1993)

17. Okamoto, Y. and U. H. Hansmann: Thermodynamics of Helix-Coil Transitions Studied by Multicanonical Algorithms. *J Phys Chem* 99, 11276-11287 (1995)

18. Hansmann, U. H. and Y. Okamoto: Effects of Side-Chain Charges on Alpha-Helix Stability in C-Peptide of Ribonuclease A Studied by Multicanonical Algorithm. *J Phys Chem B* 103, 1595-1604 (1999)

19. Ohkubo, Y. Z. and C. L. Brooks: Exploring Flory's isolated-pair hypothesis: Statistical mechanics of helix-coil transitions in polyalanine and the C-peptide from RNase A. *Proc Natl Acad Sci USA* 100, 13916-13921 (2003)

20. Sugita, Y. and Y. Okamoto: Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 314, 141-151 (1999)

21. Sugita, Y. and Y. Okamoto: Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape. *Chem Phys Lett* 329, 261-270 (2000)

22. Mitsutake, A., Y. Sugita and Y. Okamoto: Replica-exchange multicanonical and multicanonical replica-exchange Monte Carlo simulations of peptides. I. Formulation and benchmark test. *J Chem Phys* 118, 6664-6675 (2003)

23. Mitsutake, A., Y. Sugita and Y. Okamoto: Replica-exchange multicanonical and multicanonical replica-exchange Monte Carlo simulations of peptides. II. Application to a more complex system. *J Chem Phys* 118, 6676-6688 (2003)

24. Sugita, Y. and Y. Okamoto: Molecular mechanism for stabilizing a short helical peptide studied by generalized-ensemble simulations with explicit solvent. *Biophys J* 88, 3180-3190 (2005)

25. Sugita, Y., N. Miyashita, T. Yoda, M. Ikeguchi and C. Toyoshima: Structural changes in the cytoplasmic domain of phospholamban by phosphorylation at Ser16: a molecular dynamics study. *Biochemistry* 45, 11752-11761 (2006)

26. Hansmann, U. H., Y. Okamoto and F. Eisenmenger: Molecular dynamics, Langevin and hybrid Monte Carlo simulations in a multicanonical ensemble. *Chem Phys Lett* 259, 321-330 (1996)

27. Nakajima, N., H. Nakamura and A. Kidera: Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides. *J Phys Chem B* 101, 817-824 (1997)

28. Swendsen, R. H. and J. S. Wang: Replica Monte Carlo simulation of spin glasses. *Phys Rev Lett* 57, 2607-2609 (1986)

29. Kimura, K. and K. Taki: *Proc. 13th IMACS World Cong. on Computation and Appl. Math (ICMACS '91)*, Vichnevetsky, R. & Miller, J.J.H., Eds. 827-828 (1991)

30. Geyer, C. J.: *Computing Science and Statistics: Proc. 23rd Sympo. on the Interface*, Keramidas, E.M., Ed. 156-163 (1991)

31. Wang, F. and D. P. Landau: Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys Rev Lett* 86, 2050-2053 (2001)

32. Sugita, Y., A. Kitao and Y. Okamoto: Multidimensional replica-exchange method for free-energy calculations. *J Chem Phys* 113, 6042-6051 (2000)

33. Fukunishi, H., O. Watanabe and S. Takada: On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J Chem Phys* 116, 9058-9067 (2002)

34. Yoda, T., Y. Sugita and Y. Okamoto: Comparisons of force fields for proteins by generalized-ensemble simulations. *Chem Phys Lett* 386, 460-467 (2004)

35. Yoda, T., Y. Sugita and Y. Okamoto: Cooperative folding mechanism of a beta-hairpin peptide studied by a multicanonical replica-exchange molecular dynamics simulation. *Proteins* 66, 846-859 (2007)

36. Yoda, T., Y. Sugita and Y. Okamoto: Secondary-structure preferences of force fields for proteins evaluated by generalized-ensemble simulations. *Chem Phys* 307, 269-283 (2004)

37. Shoemaker, K. R., R. Fairman, D. A. Schultz, A. D. Robertson, E. J. York, J. M. Stewart and R. L. Baldwin: Side-Chain Interactions in the C-Peptide Helix-Phe8-His12+. *Biopolymers* 29, 1-11 (1990)

38. Shoemaker, K. R., P. S. Kim, E. J. York, J. M. Stewart and R. L. Baldwin: Tests of the Helix Dipole Model for Stabilization of Alpha-Helices. *Nature* 326, 563-567 (1987)

39. Bruschweiler, R., D. Morikis and P. E. Wright: Hydration of the Partially Folded Peptide Rn-24 Studied by Multidimensional Nmr. *J Biomole NMR* 5, 353-356(1995)

40. Osterhout, J. J., R. L. Baldwin, E. J. York, J. M. Stewart, H. J. Dyson and P. E. Wright: H-1-Nmr Studies of the Solution Conformations of an Analog of the C-Peptide of Ribonuclease-A. *Biochemistry* 28, 7059-7064 (1989)

41. Okamoto, Y., M. Fukugita, T. Nakazawa and H. Kawai: Alpha-Helix Folding by Monte-Carlo Simulated Annealing in Isolated C-Peptide of Ribonuclease-A. *Protein Eng* 4, 639-647 (1991)

42. Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman: A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J Am Chem Soc* 117, 5179 - 5197 (1995)

43. Wang, J., P. Cieplak and P. A. Kollman: How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules. *J Comp Chem* 21, 1049-1074 (2000)

44. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein: Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* 79, 926-935 (1983)

45. Kitao, A., F. Hirata and N. Go: The Effects of Solvent on the Conformation and the Collective Motions of Protein - Normal Mode Analysis and Molecular-Dynamics Simulations of Melittin in Water and in Vacuum. *Chem Phys* 158, 447-472 (1991)

46. Garcia, A. E.: Large-Amplitude Nonlinear Motions in Proteins. *Phys Rev Lett* 68, 2696-2699 (1992)

47. Amadei, A., A. B. M. Linssen and H. J. C. Berendsen: Essential Dynamics of Proteins. *Proteins* 17, 412-425 (1993)

48. MacLennan, D. H. and E. G. Kranias: Phospholamban: a crucial regulator of cardiac contractility. *Nat Rev Mol Cell Biol* 4, 566-577 (2003)

49. Tada, M., M. Inui, M. Yamada, M. Kadoma, T. Kuzuya, H. Abe and S. Kakiuchi: Effects of phospholamban phosphorylation catalyzed by adenosine 3':5'-monophosphate- and calmodulin-dependent protein kinases on calcium transport ATPase of cardiac sarcoplasmic reticulum. *J Mol Cell Cardiol* 15, 335-346 (1983)

50. Tada, M., M. A. Kirchberger and A. M. Katz: Regulation of calcium transport in cardiac sarcoplasmic reticulum by cyclic AMP-dependent protein kinase. *Recent Adv Stud Cardiac Struct Metab* 9, 225-239 (1976)

51. Zamoon, J., A. Mascioni, D. D. Thomas and G. Veglia: NMR solution structure and topological orientation of monomeric phospholamban in dodecylphosphocholine micelles. *Biophys J* 85, 2589-2598 (2003)

52. MacKerell, A. D., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao,

H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus: All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102, 3586-3616 (1998)

53. Mackerell, A. D., Jr., M. Feig and C. L. Brooks, III: Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comp Chem* 25, 1400-1415 (2004)

54. Toyoshima, C., M. Asahi, Y. Sugita, R. Khanna, T. Tsuda and D. H. MacLennan: Modeling of the inhibitory interaction of phospholamban with the Ca^{2+} ATPase. *Proc Natl Acad Sci USA* 100, 467-472 (2003)

55. Metcalfe, E. E., N. J. Traaseth and G. Veglia: Serine 16 phosphorylation induces an order-to-disorder transition in monomeric phospholamban. *Biochemistry* 44, 4386-4396 (2005)

56. Li, J., D. J. Bigelow and T. C. Squier: Phosphorylation by cAMP-dependent protein kinase modulates the structural coupling between the transmembrane and cytosolic domains of phospholamban. *Biochemistry* 42, 10674-10682 (2003)

57. Li, J., D. J. Bigelow and T. C. Squier: Conformational changes within the cytosolic portion of phospholamban upon release of Ca-ATPase inhibition. *Biochemistry* 43, 3870-3879 (2004)

58. van Gunsteren, W. F., S. R. Billeter, A. A. Eising, P. H. Hunenberger, P. Kruger, A. E. Mark, W. R. P. Scott and I. G. Tironi: Biomolecular simulation: the GROMOS96 manual and user guide., Zurich (1996)

59. Kollman, P. A., R. Dixon, W. Cornell, T. Fox, C. Chipot and A. Pohorille: Computer Simulation of Biomolecular System. Elsevier, 83-96 (1997)

60. Jorgensen, W. L., D. S. Maxwell and J. Tirado-Rives: Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* 118, 11225-11236 (1996)

61. Kaminski, G. A., R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen: Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* 105, 6474-6487 (2001)

62. Kabsch, W. and C. Sander: Dictionary of Protein Secondary Structure - Pattern-Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* 22 2577-2637 (1983)

Send correspondence to: Yuji Sugita, Advanced Science Institute, RIKEN, 2-1 Hirosawa, Wako, Saitama, 351-0198, Japan, Tel: 81-48-462-1407, Fax: 81-48-467-4532, E-mail: sugita@riken.jp