Prediction of membrane protein structures by replica-exchange Monte Carlo simulations: Case of two helices

Hironori Kokubo, and Yuko Okamoto

Citation: J. Chem. Phys. **120**, 10837 (2004); doi: 10.1063/1.1712942
View online: https://doi.org/10.1063/1.1712942
View Table of Contents: http://aip.scitation.org/toc/jcp/120/22
Published by the American Institute of Physics

Prediction of membrane protein structures by replica-exchange Monte Carlo simulations: Case of two helices

Hironori Kokubo[a)]
Department of Functional Molecular Science, The Graduate University for Advanced Studies, Okazaki, Aichi 444-8585, Japan

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

(Received 13 January 2004; accepted 1 March 2004)

We test our prediction method of membrane protein structures with glycophorin A transmembrane dimer and analyze the predicted structures in detail. Our method consists of two parts. In the first part, we obtain the amino-acid sequences of the transmembrane helix regions from one of existing WWW servers and use them as an input for the second part of our method. In the second part, we perform a replica-exchange Monte Carlo simulation of these transmembrane helices with some constraints that indirectly represent surrounding lipid and water effects and identify the predicted structure as the global-minimum-energy state. The structure obtained in the case for the dielectric constant $\epsilon=1.0$ is very close to that from the nuclear magnetic resonance experiments, while that for $\epsilon=4.0$ is more packed than the native one. Our results imply that the helix–helix interaction is the main driving force for the native structure formation and that the stability of the native structure is determined by the balance of the electrostatic term, van der Waals term, and torsion term, and the contribution of electrostatic energy is indeed important for correct predictions. The inclusion of atomistic details of side chains is essential for estimating this balance accurately because helices are tightly packed. © *2004 American Institute of Physics.* [DOI: 10.1063/1.1712942]

I. INTRODUCTION

It is one of the most important problems in the structural genomics era to predict protein tertiary structures from the amino-acid sequence information. We can obtain various information about the function and stability by the knowledge of protein structures. Therefore many efforts are devoted to structural determination of proteins.

It is estimated that 20%–30% of all genes in most genomes encode membrane proteins.[1,2] However, only a small number of detailed structures have been obtained for membrane proteins because of technical difficulties in experiments such as high quality crystal growth. About 23 000 protein structures are currently registered on the Protein Data Bank (PDB),[3] but most of them are structures of soluble proteins, and the number of membrane protein structures are less than 100. The database analysis based on bioinformatics such as homology search are thus unreliable due to lack of enough samples. Therefore, it is desirable to develop a method for predicting membrane protein structures by computer simulations (for previous attempts see, for instance, Refs. 4–9).

Although the number of known membrane protein structures is small in PDB, we can still extract several features of their structures. Transmembrane regions of most membrane proteins in inner membrane are composed of helices, and those in outer membrane are composed of a β-sheet. In other words, membrane proteins have only one type of secondary structures in these regions, and in this sense their structures are simpler than those of soluble proteins. Another feature is that membrane protein structures are known to be more tightly packed than soluble protein structures. These features should be taken into account and utilized when we consider membrane protein structure predictions.

The two-stage model was proposed for the structure formation of membrane proteins which are composed of several transmembrane helices in Ref. 10. In the two-stage model, individual helices of a membrane protein are postulated to be stable separately as domains in a lipid bilayer and then side-to-side helix association is driven, resulting in a functional protein. In fact, some experimental evidence indicates that the formation of α-helices and the positioning of transmembrane helices are independent: Separated fragments of bacteriorhodopsin formed independently α-helical conformations in the membrane, and the native structure could be recovered by mixing the fragments.[11,12] Therefore, it is reasonable to assume that processes of helix formation and positioning can be predicted separately.

Considering the difficulties in experiments and homology-based predictions, it is particularly desirable to develop effective prediction methods of membrane protein structures by molecular simulations. Molecular simulations allow us to understand the physical mechanism of the stability and functions of membrane proteins and help us to construct a unified view of their structures and functions.

a)Electronic mail: kukubo@ims.ac.jp
b)Electronic mail: okamotoy@ims.ac.jp

In this article, we target membrane proteins which are composed of transmembrane helices and propose the prediction method of transmembrane helices (we do not consider β-sheet membrane proteins in the present work). Our prediction method consists of two parts. In the first part, amino-acid sequences of the transmembrane helix regions are obtained from database analyses.[1,13–19] In the second part, we perform a molecular simulation of these transmembrane helices with some constraints and identify the predicted structure as the global-minimum-energy state.

However, it is difficult to obtain a global-minimum state in potential energy surface by conventional molecular dynamics (MD) or Monte Carlo (MC) simulations. This is because there exist a huge number of local-minimum-energy states, and the simulations tend to get trapped in one of the local-minimum states. One popular way to overcome this multiple-minima problem is to perform a generalized-ensemble simulation (for reviews, see Refs. 20 and 21), which is based on non-Boltzmann probability weight factors so that a random walk in potential energy space may be realized. The random walk allows the simulation to go over any energy barrier and sample much wider configurational space than by conventional methods. One of well-known generalized-ensemble algorithms is the replica-exchange method (REM) (Refs. 22–24) (the method is also referred to as parallel tempering[25]). We apply this method to the structure prediction of membrane proteins. We can obtain not only the global-minimum-energy state but also canonical-ensemble averages of physical quantities as a function of temperature from only one REM simulation run by using the multiple-histogram reweighting techniques.[26,27]

The dimeric transmembrane domain of glycophorin A is often used as a model system of helix–helix interaction of membrane proteins.[6,7] In this article, we test the second part of our prediction method using this membrane protein. Namely, given the amino-acid sequences of the transmembrane helices, we performed a replica-exchange Monte Carlo simulation to predict the helix configurations. Preliminary results have been already reported elsewhere.[28] Here, we give the details of the present approach and results.

Section II summarizes the details of our method for predicting transmembrane helix configurations. In Sec. III the details of the condition of the REM simulations for the dimeric transmembrane domain of glycophorin A are explained. In Sec. IV the results of the application to the structure prediction of the dimeric transmembrane domain of glycophorin A are given. Section V is devoted to conclusions.

II. METHODS

Our method consists of two parts. In the first part, amino-acid sequences of the transmembrane helix regions of the target protein are identified. It is already established that the transmembrane helical segments can be predicted by analyzing mainly the hydrophobicity of amino-acid sequences, without having any information about the higher order structures. There exist many WWW servers such as TMHMM,[1] MEMSAT,[16] SOSUI,[18] and HMMTOP,[19] in which given the amino-acid sequence of a protein they judge whether the protein is a membrane protein or not and (if yes) predict the regions in the amino-acid sequence that correspond to the transmembrane helices. Hence, in the first part of our method, we obtain the amino-acid sequences of the transmembrane helices from one of these WWW servers and use them as an input for the second part of our method.

In the second part, given the amino-acid sequences of transmembrane parts which were identified in the first part, we assume transmembrane parts as helices and construct ideal canonical α-helices (3.6 residues per turn) of these sequences. For our simulations we introduce the following rather drastic approximations: (1) We treat the backbone of the α-helices as rigid body and fix the backbone structures of helices. Only side-chain structures are made flexible. Each helix also has the freedom of translation and rotation. (2) We use only the transmembrane parts in the simulation and neglect the rest of the amino acids of the membrane protein (such as loop regions). (3) We neglect surrounding molecules such as lipids.

Approximation (1) above is introduced following the two-stage model, in which each helix is stable as a domain and the native configurations are built mainly by the interactions between helices. We believe that the flexibility of side chains is also important because membrane proteins are very tightly packed and the packed structures are searched by varying side-chain structures.

Approximation (2) is based on the following idea: The environment inside the membrane is very hydrophobic and that outside the membrane is very hydrophilic. Loop regions of membrane proteins are often outside the membrane and the structure outside the membrane are often flexible. Therefore we assume that the regions inside the membrane form stable structures by themselves and approximation (2) follows.

We consider that helix–helix interactions are the main driving force of the structure formation. If we perform the simulations using explicit lipids, large computational time will be necessary and it is difficult to search the wide conformational space. Therefore we introduce approximation (3) and do not treat lipids explicitly in the simulations in this article. This approximation is justified by the two-stage model again, which implies that helix–helix interactions (and not helix–lipid interactions) are the main driving force of the structure formation. In the future we plan to treat the effects of lipids more accurately. We remark that a generalized Born theory of lipids has been recently introduced.[29]

In principle, we can also use molecular dynamics method, but we employ Monte Carlo algorithm here. We update configurations with rigid translations and rigid rotations of each α-helix and torsion rotations of side chains.

We use a standard force field such as CHARMM (Refs. 30 and 31) for the potential energy of the system and add the following simple harmonic constraints to the original force-field potential energy:

$$E_{\text{constr}} = \sum_{i=1}^{N_H-1} k_1\theta(r_{i,i+1}-d_{i,i+1})[r_{i,i+1}-d_{i,i+1}]^2 + \sum_{i=1}^{N_H} \{k_2\theta(|z_i^L-z_0^L|-d_i^L)[|z_i^L-z_0^L|-d_i^L]^2 + k_2\theta(|z_i^U-z_0^U|-d_i^U)[|z_i^U-z_0^U|-d_i^U]^2\} + \sum_{C_\alpha} k_3\theta(r_{C_\alpha}-d_{C_\alpha})[r_{C_\alpha}-d_{C_\alpha}]^2, \quad (1)$$

where N_H is the total number of transmembrane helices in the protein and $\theta(x)$ is the step function

$$\theta(x) = \begin{cases} 1, & \text{for } x \geqslant 0, \\ 0, & \text{otherwise}, \end{cases} \quad (2)$$

and k_1, k_2, and k_3 are the force constants of the harmonic constraints, $r_{i,i+1}$ is the distance between the C atom of the C-terminus of the ith helix and the N atom of the N-terminus of the $(i+1)$th helix, z_i^L and z_i^U are the z-coordinate values of the C_α (or C) atom of the N-terminus (or C-terminus) of the ith helix near the fixed lower boundary value z_0^L and the upper boundary value z_0^U of the membrane, respectively, r_{C_α} are the distance of C_α atoms from the origin, and $d_{i,i+1}$, d_i^L, d_i^U, and d_{C_α} are the corresponding central value constants of the harmonic constraints.

The first term in Eq. (1) is the energy that constrains pairs of adjacent helices along the amino-acid chain not to be apart from each other too much (loop constraints). This term has a non-zero value only when the distance $r_{i,i+1}$ becomes longer than $d_{i,i+1}$. Only the structures in which the distance between helices is short are searched because of this constraint term. Our purpose is to find the optimal packed configurations of helices, therefore it is reasonable to set this constraint energy.

The second term in Eq. (1) is the energy that constrains helix N-terminus and C-terminus to be located near membrane boundary planes. This term has a nonzero value only when the C atom of each helix C-terminus and C_α atom of each helix N-terminus are apart more than d_i^L (or d_i^U). Based on the knowledge that most membrane proteins are placed in parallel, this constraint energy is included so that the helix ends are not too much apart from the membrane boundary planes.

The third term in Eq. (1) is the energy that constrains all C_α atoms within the sphere (centered at the origin) of radius d_{C_α}. This term has a nonzero value only when C_α atoms go out of this sphere. The term is introduced so that the center of mass of the molecule stays near the origin. The radius of the sphere is set to a large value in order to guarantee that a wide configurational space is sampled.

These three terms do not impose any restraints on the possible structures as membrane proteins if the constraint constants of d_i^L, d_i^U, and d_{C_α} are set large enough as we can understand from the fact that the step function is used.

We now briefly review the replica-exchange method (REM) (Refs. 22–24) (see Refs. 21 and 33 for details). The system for REM consists of M noninteracting copies (or replicas) of the original system in the canonical ensemble at M different temperatures T_m $(m=1,...,M)$. Let $X=(\ldots,x_m^{[i]},\ldots)$ stand for a state in this ensemble. Here, the superscript i and the subscript m in $x_m^{[i]}$ label the replica and the temperature, respectively. The state X is specified by M sets of $x_m^{[i]}$, which in turn is specified by the coordinates $q^{[i]}$ of all the atoms in replica i. A REM simulation is then realized by alternately performing the following two steps. Step 1: Each replica in canonical ensemble of the fixed temperature is simulated simultaneously and independently for a certain MC or MD steps. Step 2: A pair of replicas, say i and j, which are at neighboring temperatures T_m and T_n, respectively are exchanged: $X=(\ldots,x_m^{[i]},\ldots,x_n^{[j]},\ldots)\rightarrow X'=(\ldots,x_m^{[j]},\ldots,x_n^{[i]},\ldots)$. The transition probability of this replica exchange is given by the Metropolis criterion,

$$w(X\rightarrow X') \equiv w(x_m^{[i]}|x_n^{[j]}) = \begin{cases} 1, & \text{for } \Delta \leqslant 0, \\ \exp(-\Delta), & \text{otherwise}, \end{cases} \quad (3)$$

where

$$\Delta = (\beta_m-\beta_n)(E(q^{[j]})-E(q^{[i]})). \quad (4)$$

Here, $E(q^{[i]})$ and $E(q^{[j]})$ are the potential energy of the ith replica and the jth replica, respectively. In the present work, we employ Monte Carlo algorithm for Step 1. There are $2N_H+N_D$ kinds of Monte Carlo moves, where N_D is the total number of dihedral angles in the side chains of N_H helices. The first term corresponds to the rigid translation and rigid rotation of the helices and the second to the dihedral-angle rotations in the side chains. One MC sweep is defined to consist of $2N_H+N_D$ updates that are randomly chosen from these MC moves with the Metropolis evaluation for each update. We predict the native structure of membrane spanning regions as the global-minimum-energy state obtained by the REM simulations.

From only one simulation run, one can obtain not only the global-minimum structure but also canonical-ensemble averages of physical quantities as functions of temperature by using the multiple-histogram reweighting techniques[26,27] (see also Ref. 32) as follows. Suppose we have made M independent simulation runs at M different temperatures. Let $N_m(E)$ and n_m be the energy histogram and the total number of samples obtained in the mth run, respectively. The expectation value of a physical quantity A at any intermediate temperature T is given by

$$\langle A\rangle_T = \frac{\sum_E A(E)n(E)\exp(-\beta E)}{\sum_E n(E)\exp(-\beta E)}, \quad (5)$$

where the density of states $n(E)$ is obtained by solving the following WHAM equations:

$$n(E) = \frac{\sum_{m=1}^{M} g_m^{-1}N_m(E)}{\sum_{m=1}^{M} g_m^{-1}n_m \exp(f_m-\beta_m E)}, \quad (6)$$

and

$$\exp(-f_m) \equiv \sum_E n(E)\exp(-\beta_m E). \tag{7}$$

Here, $g_m = 1 + 2\tau_m$, and τ_m is the integrated autocorrelation time at temperature T_m. For biomolecular systems the quantity g_m can safely be set to be a constant in the reweighting formulas,[27] and so we set $g_m = 1$ throughout the analyses in the present work. Note that $n(E)$ and f_m are solved self-consistently by iteration. Moreover, ensemble averages of any physical quantity A (including those that cannot be expressed as functions of potential energy) can now be obtained from the "trajectory" of configurations of the production run. Namely, we first obtain f_m $(m = 1,...,M)$ by solving Eqs. (6) and (7) self-consistently, and then we have[32]

$$\langle A \rangle_T = \frac{\sum_{m=1}^{M} \sum_{x_m} A(x_m) \dfrac{g_m^{-1}}{\sum_{\ell=1}^{M} g_\ell^{-1} n_\ell \exp(f_\ell - \beta_\ell E(x_m))} \exp(-\beta E(x_m))}{\sum_{m=1}^{M} \sum_{x_m} \dfrac{g_m^{-1}}{\sum_{\ell=1}^{M} g_\ell^{-1} n_\ell \exp(f_\ell - \beta_\ell E(x_m))} \exp(-\beta E(x_m))}, \tag{8}$$

where x_m are the configurations at temperature T_m. Here, the trajectories x_m are taken for each temperature T_m separately.

III. COMPUTATIONAL DETAILS

Our method consists of two parts. In the first part, we obtain amino-acid sequences of the transmembrane helix regions from existing WWW servers such as those in Refs. 1, 16, 18, and 19. However, the precision of these programs in the WWW servers is about 85% and needs improvement. We thus focus our attention on the effectiveness of the second part of our method, leaving this improvement to the developers of the WWW servers. Namely, we use the experimentally known amino-acid sequence of transmembrane regions (without relying on the WWW servers) and try to predict their conformations, following the prescription of the second part of our method described in the previous section. Here, we chose one of the simplest systems: the transmembrane dimer of glycophorin A (PDB code: 1AFO). The number of amino acids for each helix is 18 and the sequence is TLIIFGVMAGVIGTILLI.

We first constructed the ideal canonical α-helix (3.6 residues per turn) of this sequence. The N and C termini of this helix were blocked with acetyl and N-methyl groups, respectively. The force field that we used is the CHARMM PARAM19 parameter set (polar hydrogen model).[30,31] No cutoff was introduced to the nonbonded energy terms, and the dielectric constant ϵ was set equal to 1.0. We have also studied the case of $\epsilon = 4.0$, because it is the value close to that for the lipid environment. The computer code based on the CHARMM macromolecular mechanics program[34] was used and the replica-exchange method was implemented in it. This helix structure was minimized subject to harmonic restraints on all the heavy atoms. The initial configuration for the REM simulation was that two α-helices of identical sequence and structure thus prepared were placed in parallel at a distance of 20 Å. These helices are quite apart from each other and the starting configuration is indeed very different from the native one. Note that the only information derived from the NMR experiments[35] is the amino-acid sequence of the individual helices.

The values of the constants for the constraints in Eq. (1) were set as follows: $N_H = 2$, $k_1 = k_2 = 0.5$ kcal/(mol Å^2), $k_3 = 0.05$ kcal/(mol Å^2), $d_{i,i+1} = 20$ Å, $z_0^L = -13.35$ Å, $z_0^U = +13.35$ Å, $d_i^L = d_i^U = 1.0$ Å, and $d_{C_\alpha} = 50$ Å. The values for z_0^L and z_0^U were taken from the z-coordinates of the initial configuration [in Fig. 3(a) below; the z axis is placed vertically in the figure]. In the present example of glycophorin A dimer, the first term in Eq. (1) was imposed on both terminal ends (i.e., two kinds of $r_{i,i+1}$ were prepared: one is the distance between a pair of N atoms at the N-terminus of the two helices and the other is the distance between a pair of C atoms at the C-terminus of the two helices). As explained in Sec. II, the constraint terms do not impose any constraints on the possible structures as membrane proteins if the constraint constants are set properly as we can understand from the fact that the step function is used.

We performed two REM MC simulations of 1 000 000 MC sweeps, starting from the parallel configuration of Fig. 3(a) below: one with the dielectric constant $\epsilon = 1.0$ and the other with $\epsilon = 4.0$. We used the following 13 temperatures: 200, 239, 286, 342, 404, 489, 585, 700, 853, 1041, 1270, 1548, and 1888 K, which are distributed almost exponentially. The highest temperature was chosen sufficiently high so that no trapping in local-minimum-energy states occurs. This temperature distribution was chosen so that all the acceptance ratios are almost uniform and sufficiently large (>10%) for computational efficiency. Replica exchange was attempted once at each MC sweep.

IV. RESULTS AND DISCUSSION

We first examine whether the present REM simulations performed properly. The acceptance ratios of replica exchange are listed in Table I for both cases of dielectric constants. We see that the acceptance ratios of replica exchange between all pairs of neighboring temperatures are uniform and large enough (>10%) for computational efficiency. The results in Table I imply that one should observe a free random walk in the replica space and temperature space.

In Figs. 1 and 2 we show the "time series" of the present

TABLE I. Acceptance ratios of replica exchange corresponding to pairs of neighboring temperatures with the dielectric constant $\epsilon=1.0$ and $\epsilon=4.0$.

Pairs of temperatures	Acceptance ratio ($\epsilon=1.0$)	Acceptance ratio ($\epsilon=4.0$)
200↔239 K	0.41	0.37
239↔286 K	0.40	0.37
286↔342 K	0.39	0.37
342↔404 K	0.40	0.41
404↔489 K	0.32	0.34
489↔585 K	0.34	0.36
585↔700 K	0.33	0.36
700↔853 K	0.28	0.30
853↔1041 K	0.29	0.31
1041↔1270 K	0.36	0.36
1270↔1548 K	0.42	0.42
1548↔1888 K	0.46	0.47

REM simulations with the dielectric constant $\epsilon=1.0$ and $\epsilon=4.0$, respectively. In Figs. 1(a) and 2(a) the "time series" of replica exchange at the lowest temperature ($T=200$ K) are shown. We see that every replica takes the lowest temperature many times, and we indeed observe a random walk in the replica space. The complementary picture to this is the temperature exchange for each replica. The results for one of the replicas (Replica 6) are shown in Figs. 1(b) and 2(b). We again observe random walks in the temperature space between the lowest and highest temperatures. Other replicas perform random walks similarly. In Figs. 1(c) and 2(c) the corresponding time series of the total potential energy are shown. We see that random walks in the potential energy space between low and high energy regions are also realized. Note that there is a strong correlation between the behaviors in Figs. 1(b) and 1(c) as there should. The same is true for Figs. 2(b) and 2(c). All these results confirm that the present REM simulations have been properly performed.

We now study how widely the configurational space was sampled during the present simulations. We first examine the case for $\epsilon=1.0$. We plot the time series of the root-mean-square (RMS) deviation of the backbone atoms from the

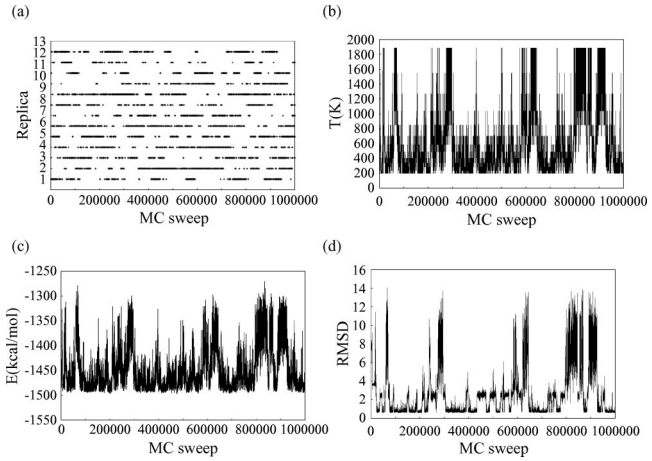

FIG. 1. Time series of replica exchange at $T=200$ K (a), temperature exchange for one of the replicas (Replica 6) (b), the total potential energy for Replica 6 (c), and the RMS deviation (in Å) of backbone atoms from the NMR structure for Replica 6 (d) with the dielectric constant $\epsilon=1.0$.

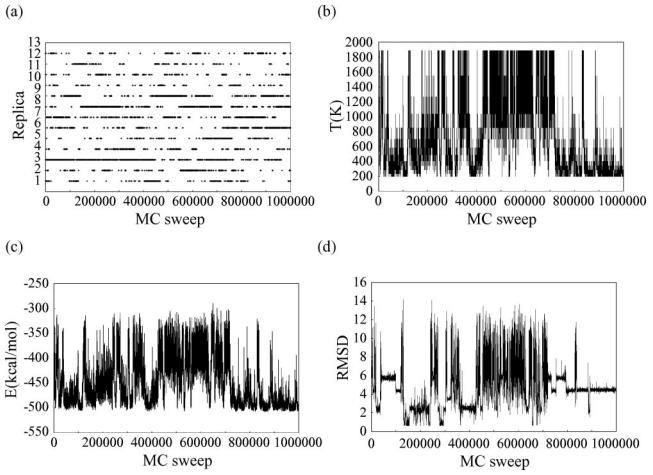

FIG. 2. Time series of replica exchange at $T=200$ K (a), temperature exchange for one of the replicas (Replica 6) (b), the total potential energy for Replica 6 (c), and the RMS deviation (in Å) of backbone atoms from the NMR structure for Replica 6 (d) with the dielectric constant $\epsilon=4.0$.

NMR structure[35] in Fig. 1(d). When the temperature becomes high, the RMS deviation takes a large value [the largest value in Fig. 1(d) is 14.1 Å, and the maximum value among all the replicas is 15.7 Å], and when the temperature becomes low, the RMS deviation takes a small value [the smallest value in Fig. 1(d) is 0.55 Å, and the minimum value among all the replicas is also 0.55 Å]. By comparing Fig. 1(c) and Fig. 1(d), we see that there is a strong correlation between the total potential energy and the RMS deviation values. In particular, it is remarkable that when the energy is the lowest (around -1490 kcal/mol), most of the RMS values are as small as about 0.5 Å. This implies that the global-minimum-energy state is indeed very close to the native structure.

We now examine the case with $\epsilon=4.0$. We plot the time series of the RMS deviation of the backbone atoms from the NMR structure[35] in Fig. 2(d). When the temperature becomes high, the RMS deviation takes a large value [the largest value in Fig. 2(d) is 14.2 Å, and the maximum value among all the replicas is 14.8 Å]. The RMS deviation sometimes takes a small value [the smallest value in Fig. 2(d) is 0.58 Å, and the minimum value among all the replicas is 0.56 Å], and when this occurs, the temperature is low and the potential energy takes a small value. When the temperature becomes low and the potential energy is low, however, the RMS deviation is not always small and takes several values (around 0.7, 2.5, 3.5, 4.4, or 5.7 Å) contrary to the case with $\epsilon=1.0$. This implies that there are several stable structures at low temperatures. We will discuss this matter more in detail below.

In Fig. 3 typical snapshots from the REM simulations of Figs. 1 and 2 are shown. Figure 3(a) is the initial configuration of our simulations, in which the two helices are placed in parallel. Figures 3(b1)–3(b4) and Figs. 3(c1)–3(c4) are typical snapshots with the dielectric constant $\epsilon=1.0$ and $\epsilon=4.0$, respectively. These figures confirm that our simulations indeed sampled wide configurational space. We see that the REM simulations perform random walks not only in en-

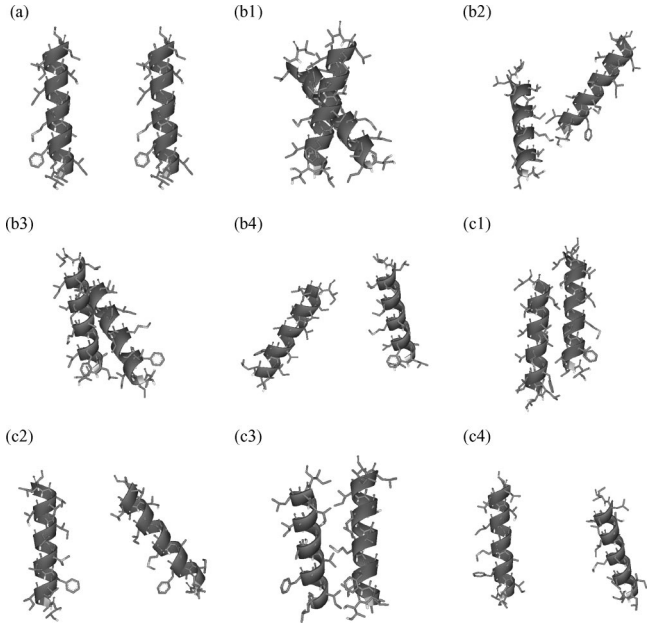

FIG. 3. Typical snapshots from the REM simulation. The initial configuration (a), the configurations with the dielectric constant $\epsilon=1.0$ (b1)–(b4) and with the dielectric constant $\epsilon=4.0$ (c1)–(c4).

ergy space but also in conformational space and that they do not get trapped in one of a huge number of local-minimum-energy states.

In Figs. 4(a) and 5(a) the canonical probability distributions of the total potential energy obtained at the chosen 13 temperatures from the REM simulation with the dielectric constant $\epsilon=1.0$ and $\epsilon=4.0$ are shown, respectively. We see that there are enough overlaps between all neighboring pairs of distributions, indicating that there will be sufficient numbers of replica exchange between pairs of replicas. In Fig. 4(b) (the case for $\epsilon=1.0$) and Fig. 5(b) (the case for $\epsilon=4.0$), the average of the total potential energy E_{tot} and averages of its component terms, namely, the electrostatic energy E_c, van der Waals energy E_v, torsion energy E_t (these

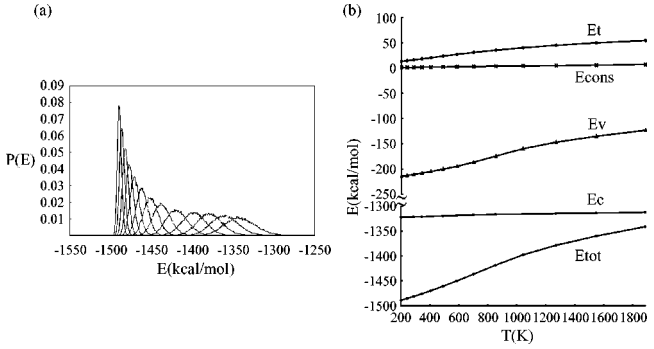

FIG. 4. (a) The canonical probability distributions of the total potential energy obtained from the replica-exchange MC simulation at the 13 temperatures with the dielectric constant $\epsilon=1.0$. The distributions correspond to the following temperatures (from left to right): 200, 239, 286, 342, 404, 489, 585, 700, 853, 1041, 1270, 1548, and 1888 K. (b) The averages of the total potential energy E_{tot} and its component terms: electrostatic energy E_c, van der Waals energy E_v, dihedral energy E_t, and constraint energy E_{cons} as functions of temperature T with the dielectric constant $\epsilon=1.0$. The values were calculated by the multiple-histogram reweighting techniques.

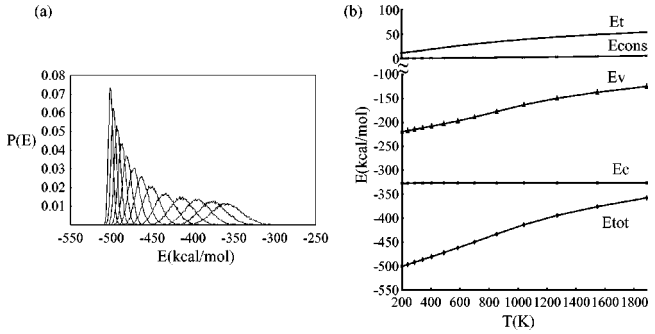

FIG. 5. (a) The canonical probability distributions of the total potential energy obtained from the replica-exchange MC simulation at the 13 temperatures with the dielectric constant $\epsilon=4.0$. The distributions correspond to the following temperatures (from left to right): 200, 239, 286, 342, 404, 489, 585, 700, 853, 1041, 1270, 1548, and 1888 K. (b) The averages of the total potential energy E_{tot} and its component terms: electrostatic energy E_c, van der Waals energy E_v, and dihedral energy E_t as functions of temperature T with the dielectric constant $\epsilon=4.0$. The values were calculated by the multiple-histogram reweighting techniques.

three terms are from the CHARMM force field), and constraint energy E_{cons} as a function of temperature T are shown. The multiple-histogram reweighting techniques in Eq. (8) were used for these calculations. We see that as the temperature becomes low, E_{tot} becomes low mainly because E_v and E_t become low. The changes of E_c and E_{cons} as the temperature is varied are small and the contribution of E_c and E_{cons} can be said to be smaller on the average than those of E_v and E_t. In the case for $\epsilon=4.0$, this temperature variation of E_c is particularly small and less than 1.0 kcal/mol.

In Fig. 6 the configuration obtained by the NMR experiments[35] and the global-minimum-energy configurations with the dielectric constant $\epsilon=1.0$ and $\epsilon=4.0$ obtained by the REM simulations are compared. The predicted structure with $\epsilon=1.0$ is in remarkable agreement with that from the NMR experiments. At first sight, it is rather surprising that the result with $\epsilon=1.0$ is much closer to the experimental result than that with $\epsilon=4.0$, because the dielectric constant for a lipid system is closer to 4.0 than to 1.0. However, on second thoughts we understand that the present results are reasonable because the pairs of helices in transmembrane proteins are tightly packed and almost no lipid molecules can exist between helices. This implies that helix–helix interactions are the main driving force in the final stage of the structure formation of membrane proteins. We remark that the good agreement of the prediction with the NMR structure was also obtained for glycophorin A in Ref. 7, although they used a very different potential energy function (a slightly modified united-atom OPLS force field without electrostatic interactions). This fact also supports the above claim that the side-chain packing between helices are very important in determining the transmembrane configurations.

For further understanding we compare some properties of the three structures from Fig. 6 in Table II. We see that the structure with $\epsilon=4.0$ is stabilized by E_v and E_t compared with the structure with $\epsilon=1.0$. The difference of E_{cons} between the structures of $\epsilon=1.0$ and $\epsilon=4.0$ is smaller than those of E_v and E_t (0.62 kcal/mol versus 5.4 and 2.0 kcal/mol). Thus the existence of the constraint terms is not the

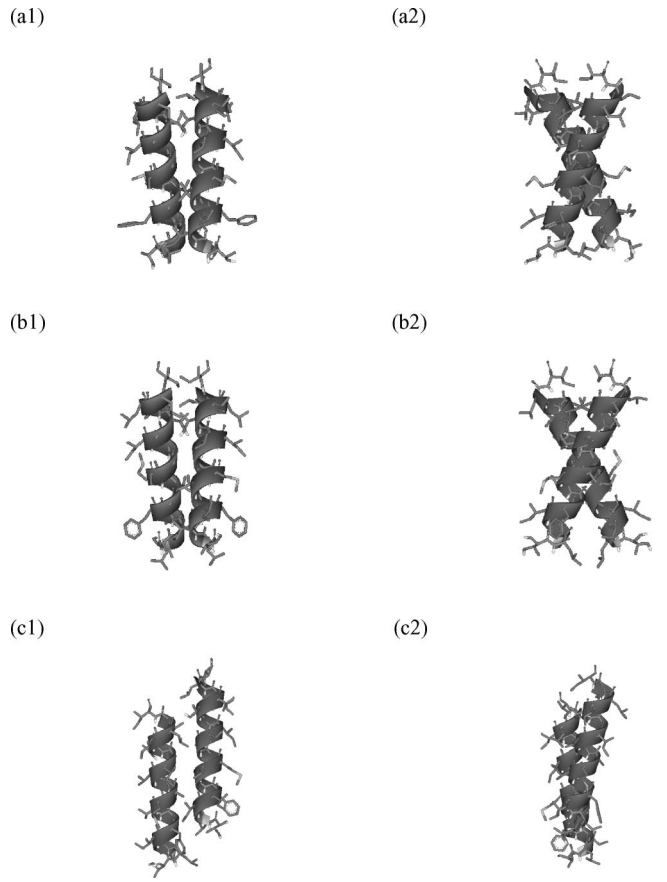

FIG. 6. The NMR configuration (PDB code 1AFO, MODEL 16) (a1) (a2), the global-minimum-energy configuration that was obtained by the REM simulation with the dielectric constant $\epsilon=1.0$ (b1) (b2), and the global-minimum-energy configurations that was obtained by the REM simulation with the dielectric constant $\epsilon=4.0$ (c1) (c2). The pair (a1) and (a2) correspond to the same structure viewed from different angles. Likewise, the pair (b1) and (b2) and the pair (c1) and (c2) correspond to the same structures viewed from different angles. The RMS deviation from the native configuration (a) is 0.64 Å (b) and 4.48 Å (c) with respect to all backbone atoms, and it is 1.31 Å (b) and 5.55 Å (c) with respect to all atoms.

major cause for the difference between $\epsilon=1.0$ and $\epsilon=4.0$. The predicted structure with $\epsilon=4.0$ had smaller solvent accessible surface area and was more packed than the native one. This means that although only four hydrophilic amino acids are included in 36 amino acids used in our simulations, the electrostatic energy term contributes to the stability of this membrane protein and forces the native structure to be a little less packed than the case with weakened electrostatic interactions. In other words, the stability of the native structure is determined by the balance of E_c, E_v, and E_t, and the contribution of electrostatic energy is also important. The interhelical crossing angle in this table is defined as the angle between the principal axes of moment of inertia for each helix. From the interhelical crossing angle, the structure with $\epsilon=4.0$ is near parallel. This can also be understood from Fig. 6(c). One of the helices in the structure with $\epsilon=4.0$ appears to be slightly off from the membrane boundary. We see that the structure with $\epsilon=1.0$ is indeed close to the native structure in every property.

In Table III the interhelical distances of the three structures in Fig. 6 are compared with those of the solid-state

TABLE II. Various properties of the native structure and the global-minimum-energy structure by the REM simulation with the dielectric constant $\epsilon=1.0$ and $\epsilon=4.0$. The following abbreviations are used: the total potential energy E_{tot}, van der Waals energy E_v, electrostatic energy E_c, dihedral energy E_t, constraint energy E_{cons}, RMS deviation RMSD, radius of gyration RGYR, interhelical crossing angle IHCA, and solvent accessible surface area SA. The energy is in kcal/mol, distance is in Å, angle is in degrees, and area is in Å^2.

	Solution NMR structure	Global-minimum structure ($\epsilon=1.0$)	Global-minimum structure ($\epsilon=4.0$)
E_{tot}	⋯	−1497.8	−509.6
E_v	⋯	−219.6	−225.0
E_c	⋯	−1322.1	−327.0
E_t	⋯	10.5	8.5
E_{cons}	⋯	0.12	0.74
RMSD	⋯	0.64	4.48
RGYR	10.00	10.12	10.83
IHCA	45.8	42.7	14.0
SA	3152.3	3133.5	3087.2

NMR experiments.[36] The structure obtained from the solid-state NMR experiments is considered to be closer to the native structure than those obtained from the solution NMR experiments because it is crystallized in lipid bilayer as in the native state. On the other hand, the solution NMR structures were determined in detergent micelles. In the case for $\epsilon=1.0$, three distances out of six are between the solid-state NMR values and the solution NMR values and the distance between Gly79 C and Val80 C is closer to the solid-state NMR experiments. This suggests that the predicted structure in Fig. 6(b) is closer to the solid-state NMR structure than to the solution NMR structure in Fig. 6(a). The distances of the structure with $\epsilon=4.0$ are totally different from those of the solid-state and solution NMR structures as we can expect from Fig. 6(c).

In Fig. 7 we show the average values of the RMS deviation, the radius of gyration, the interhelical crossing angle, and the solvent accessible surface area as functions of temperature T for the REM simulations with $\epsilon=1.0$. The multiple-histogram reweighting techniques of Eq. (8) were used again. In Fig. 7(a) we see that the average RMSD decreases monotonically as the temperature is lowered. This means that when the temperature is high the structures that are very distant from the native one are often sampled (RMSD is as large as 8.0 Å) and when the temperature is low the structures that are close to the native one are mainly sampled (RMSD is as small as 1.0 Å). Figure 7(b) implies that the packed conformations are searched when the temperature is low and disjointed structures are searched when the temperature is high. When the temperature is 200 K, the average value of the radius of gyration is close to the native one (10.0 Å; see Table II). In Fig. 7(c) the average interhelical crossing angle is about 50° at low temperatures and about 40° at high temperatures. The crossing angle at the low temperature (about 50°) is a little larger (about 5°) than that of the native structure (45.8°; see Table II). From Fig. 7(d) we see that the average solvent accessible surface area is large when the temperature is high and it is small when the temperature is low. This is reasonable because the packed

TABLE III. The interhelical distances (in Å) of the solid-state NMR (Ref. 36), the solution NMR (Ref. 35), and the global-minimum-energy structure obtained from the REM simulation with the dielectric constant $\epsilon=1.0$ and $\epsilon=4.0$.

	Solid-state NMR structure	Solution NMR structure	Global-minimum structure ($\epsilon=1.0$)	Global-minimum structure ($\epsilon=4.0$)
Gly79 C Gly79 CA	4.1	4.7	4.5	8.9
Gly79 CA Ile76 C	4.8	4.8	5.7	12.1
Gly83 C Gly 83 CA	4.3	5.1	4.7	9.8
Gly83 CA Val80 C	4.2	4.3	4.3	12.4
Gly79 C Val80 C	4.0	2.9	4.4	12.4
Gly83 C Val84 C	4.0	3.7	3.8	13.1

conformations with small surface area are searched at low temperatures and disjointed conformations are searched at high temperatures. When the temperature is 200 K, the value of the surface area is indeed close to the native one (3152 $Å^2$; see Table II).

Similarly in Fig. 8, we show the corresponding quantities of Fig. 7, which were calculated from the results for the case of $\epsilon=4.0$. In Fig. 8(a) we see that the average RMS deviation is always more than 3.0 Å and the simulation often samples different structures from the native one. The RMS deviation at 600 K is smaller than that at 200 K, suggesting that structures more similar to the native one are sampled. The similar nature is observed in Figs. 8(b), 8(c), and 8(d). Namely, the quantities around 600 K are closer to those of the native structure than at other temperatures. This point is discussed further below around Fig. 10. The values at high temperatures in Fig. 8 are similar to the corresponding ones in Fig. 7, because the simulations search various structures without getting trapped in one or a few of local-minimum-energy structures. At low temperatures the simulations sample one or a few of them. In Fig. 8(a) the average RMS deviation at 200 K is close to that of the structure in Fig. 6(c) (see Table II). The averages of the radius of gyration [Fig. 8(b)], the interhelical crossing angle [Fig. 8(c)], and solvent accessible surface area [Fig. 8(d)] also take values close to those of the structure in Fig. 6(c). This implies that at the lowest temperature of 200 K, only the structure in Fig. 6(c) is mostly sampled.

In Figs. 7 and 8 we saw average quantities as functions of temperature. Here, we study how many dominant structures contribute to the averages by examining the probability distribution of RMSD. In Fig. 9 the results with $\epsilon=1.0$ at four temperatures are shown. From Fig. 9(a) we see that the structures close to the native one (RMSD=0.7 Å) are mainly sampled at $T=200$ K. There also exists a small contribution around RMSD=2.5 Å. This cannot be understood from the average properties. Actually, we do observe this second structure in Fig. 1. Namely, we see that the structures around RMSD=2.5 Å are certainly sampled at low temperatures [compare Figs. 1(b) and 1(d)]. Similar behavior, although less conspicuous, is observed at $T=342$ K [Fig. 9(b)]. As the temperature becomes higher, the distributions become broader. The two peaks become almost equally important at $T=585$ K [Fig. 9(c)]. This suggests that structures are sampled around only two local-minimum-energy states below 585 K. At the highest temperature of 1888 K [Fig. 9(d)] we no longer see any peaks in the histogram and various structures are sampled, and the simulation does not get

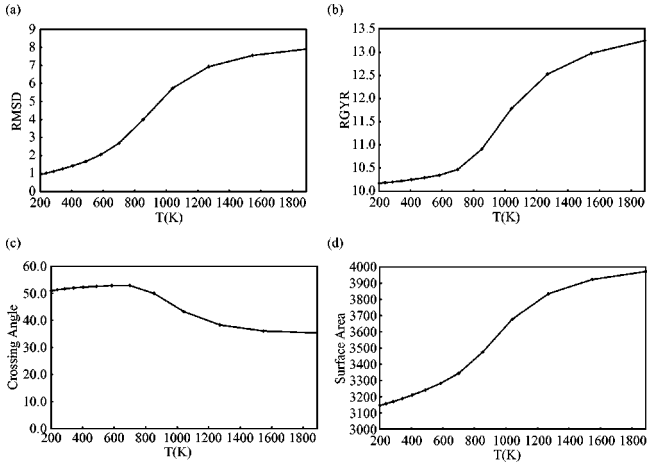

FIG. 7. The averages of the RMS deviation (a), the radius of gyration (RGYR) (b), the interhelical crossing angle (c), and the solvent accessible surface area (d) as functions of temperature T in the case of the dielectric constant $\epsilon=1.0$. The values were calculated by the multiple-histogram reweighting techniques. The distance is in Å, angle is in degrees, and area is in $Å^2$.

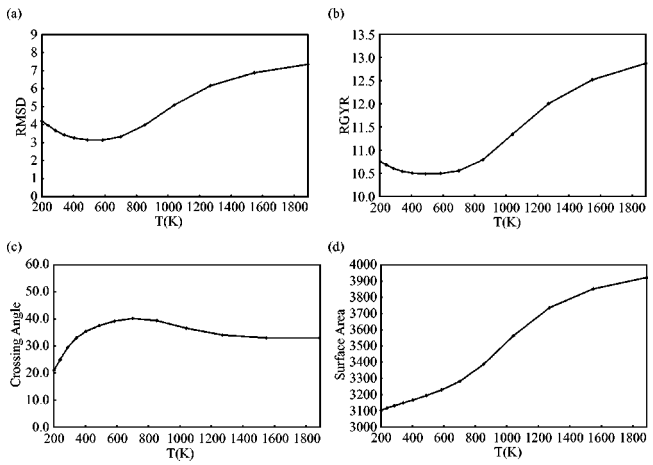

FIG. 8. The averages of the RMS deviation (a), the radius of gyration (RGYR) (b), the interhelical crossing angle (c), and the solvent accessible surface area (d) as functions of temperature T in the case of the dielectric constant $\epsilon=4.0$. The values were calculated by the multiple-histogram reweighting techniques. The distance is in Å, angle is in degrees, and area is in $Å^2$.

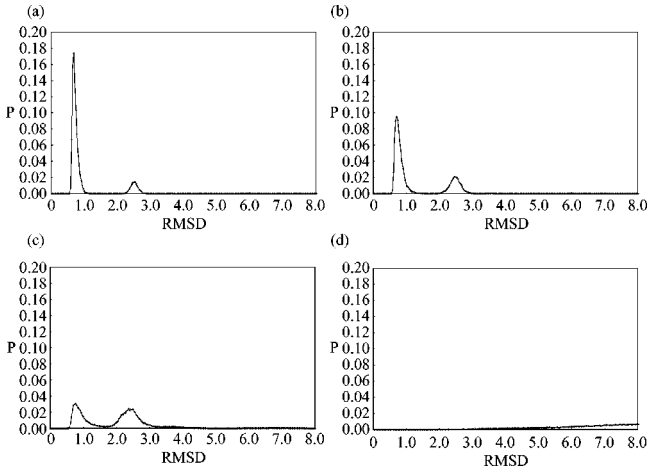

FIG. 9. The probability distributions of the RMSD obtained from the replica-exchange MC simulation with the dielectric constant $\epsilon=1.0$ at the chosen four temperatures. The distributions correspond to the following temperatures: 200 K (a), 342 K (b), 585 K (c), and 1888 K (d).

trapped in local-minimum-energy structures at this temperature.

The corresponding probability distribution for the case of $\epsilon=4.0$ are shown in Fig. 10. From Fig. 10(a) the situation is more complicated than the case of $\epsilon=1.0$. We see that the structures around RMSD$=4.5$ Å are mainly sampled at 200 K. However in this case as many as four other contributions also exist (RMSD$\sim$0.7, 2.5, 3.5, and 5.7 Å). As the temperature becomes high [Fig. 10(b) or Fig. 10(c)] from 200 K [Fig. 10(a)], the peak around RMSD$=4.5$ Å becomes small, and the four peaks (RMSD$\sim$0.7, 2.5, 4.5, and 5.7 Å) become almost equally important. This is the reason why the average values around 600 K are closer to those of the native structure than those around 200 K in Fig. 8. These results show that the simulation with $\epsilon=4.0$ also samples the native structure as one of the local-minimum-energy states (RMSD$=0.7$ Å), but it is not the dominant contributions at the lowest temperature.

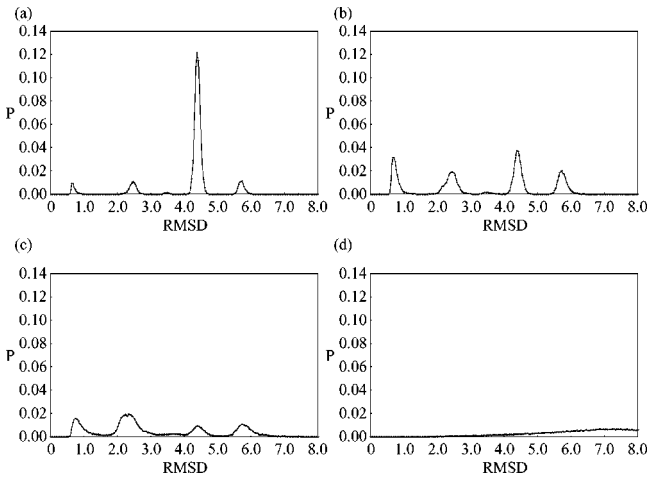

FIG. 10. The probability distributions of the RMSD obtained from the replica-exchange MC simulation with the dielectric constant $\epsilon=4.0$ at the chosen four temperatures. The distributions correspond to the following temperatures: 200 K (a), 342 K (b), 585 K (c), and 1888 K (d).

We list the average properties of the structures that correspond to the peaks of the histograms in Figs. 9(a) and 10(a) in Table IV. The global-minimum-energy structure for $\epsilon=1.0$ in Fig. 6(b) belongs to peak1 for $\epsilon=1.0$ in Table IV, and that for $\epsilon=4.0$ in Fig. 6(c) belongs to peak4 for $\epsilon=4.0$ in Table IV. We see that not only the RMS deviation but also radius of gyration, interhelical crossing angle, and solvent accessible surface area of peak1 in both cases of $\epsilon=1.0$ and $\epsilon=4.0$ are similar to one another. Hence, we conclude that the structures of peak1 for $\epsilon=1.0$ are essentially identical with those of peak1 for $\epsilon=4.0$ (they correspond to the native structure, see Table II). Likewise, the structures of peak2 for $\epsilon=1.0$ are the same as those of peak2 for $\epsilon=4.0$. Therefore, we can say that the low-energy structures that were sampled in the simulation with $\epsilon=1.0$ are subsets of structures of those that were sampled with $\epsilon=4.0$. However, only the case with $\epsilon=1.0$ gives the native structures as the global-minimum-energy state.

V. CONCLUSIONS

In this article we predicted the native structure of glycophorin A transmembrane domain as a test case of our method for the structure predictions of membrane proteins and analyzed the detailed properties of the predicted structures. Our method consists of two parts. In the first part, we obtain the amino-acid sequences of the transmembrane helix regions of the target protein from one of existing WWW servers. The precision of these programs in the WWW servers is at present about 85%, but it is expected to be further improved. In the second part of our method, we perform a generalized-ensemble simulation of these transmembrane helices with atomistic details to obtain the global-minimum-energy state, which we identify as the predicted structure. In order to save computation time, we introduced rather bold approximations in the second part: Backbones are treated as rigid body (only side-chain structures are made flexible) and the rest of the protein such as loop regions and the surrounding lipids and water are neglected.

With these assumptions, however, the structure obtained from the prediction in the case for the dielectric constant $\epsilon=1.0$ was very close to that from the NMR experiments. The fact that we can predict the native structure by the simulation without lipids implies that the helix–helix interaction is the main driving force of the native structure formation. It was found from the analysis of the average physical quantities as functions of temperature that the temperature variation of the electrostatic energy term was much smaller than the van der Waals and torsion energy terms and the contribution of the electrostatic energy to the stability was small on the average. However, only the case with $\epsilon=1.0$ gave the native structure as the global-minimum-energy structure, although structures close to the native one were also sampled with $\epsilon=4.0$ as one of the local-minimum-energy states. We saw that the predicted structure with $\epsilon=4.0$ has smaller solvent accessible surface area than the native one and is more stabilized by van der Waals energy than in the case for $\epsilon=1.0$. This finding suggests that although only four hydrophilic amino acids are included in 36 amino acids used in our simulations, the electrostatic energy term contributes to the stability of this mem-

TABLE IV. Various properties averaged over structures that correspond to each peak in histograms in Fig. 9(a) and Fig. 10(a). The abbreviations are the same as in Table II. The energy is in kcal/mol, distance is in Å, angle is in degrees, and area is in Å^2.

	$\epsilon=1.0$	
	peak1	peak2
E_{tot}	−1489.1	−1488.2
E_v	−214.2	−215.6
E_c	−1322.0	−1320.8
E_t	13.2	13.0
E_{cons}	0.66	2.02
RMSD	0.73	2.52
RGYR	10.13	10.47
IHCA	50.7	52.1
SA	3137.9	3214.9

	$\epsilon=4.0$				
	peak1	peak2	peak3	peak4	peak5
E_{tot}	−498.2	−499.0	−498.6	−500.8	−499.1
E_v	−216.2	−217.8	−217.6	−219.9	−219.8
E_c	−328.0	−328.0	−327.4	−326.9	−326.0
E_t	12.1	11.7	12.6	11.9	12.8
E_{cons}	0.68	1.90	0.53	0.94	0.78
RMSD	0.70	2.47	3.49	4.41	5.74
RGYR	10.13	10.44	10.28	10.84	10.60
IHCA	50.3	51.2	38.2	16.1	21.0
SA	3130.6	3202.0	3077.7	3097.8	3049.9

brane protein and forces the native structure to be a little less packed than the case with weakened electrostatic interactions ($\epsilon=4.0$). In other words, the stability of the native structure is determined by the balance of the electrostatic term, van der Waals term, and torsion term, and the contribution of electrostatic energy is indeed important for correct predictions. We believe that the inclusion of atomistic details of side chains is important to estimate this balance accurately because transmembrane helices are usually tightly packed. This fact also justifies the validity of our assumptions in the sense that almost no lipid molecules can exist between helices.

In this article we presented the results of membrane protein structure predictions of only two transmembrane helices. The effectivess of the method should be further tested with proteins that have more transmembrane helices. We are now working with bacteriorhodopsin (seven helices) and our results show that we do get a very similar structure to the native one by the replica-exchange MC simulation starting from a random initial configuration.[37]

Our method is useful for the structure prediction of membrane proteins which is not known yet. In the future we have to make our approximations better. For example, we should introduce some flexibility in the helix backbone structures because membrane proteins are not necessarily composed of ideal α-helices. The electrostatic interactions, in which we used the dielectric constant value of 1.0 or 4.0, can also be made more accurate so that some environmental effects including lipids may be taken into account. Our results support the two stage model for the structure formation of membrane proteins. The applicability of our method to membrane proteins in which loop regions exist in transmembrane region and interact with transmembrane helices remains to be established.

ACKNOWLEDGMENTS

The computations were performed on the computers at the Research Center for Computational Science, Okazaki National Research Institutes and at ITBL, Japan Atomic Energy Research Institute. This work was supported, in part, by the Grants-in-Aid for the NAREGI Nanoscience Project and for Scientific Research in Priority Areas, ''Water and Biomolecules,'' from the Ministry of Education, Culture, Sports, Science, and Technology, Japan.

[1] A. Krogh, B. Larsson, G. v. Heijne, and E. L. L. Sonnhammer, J. Mol. Biol. **305**, 567 (2001).
[2] S. Mitaku, Biophysics (Engl. Transl.) **42**, 104 (2002) (in Japanese).
[3] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, Nucleic Acids Res. **28**, 235 (2000).
[4] W. R. Taylor, D. T. Jones, and N. M. Green, Proteins **18**, 281 (1994).
[5] M. Suwa, T. Hirokawa, and S. Mitaku, Proteins **22**, 363 (1995).
[6] P. D. Adams, D. M. Engelman, and A. T. Brünger, Proteins **26**, 257 (1996).
[7] R. V. Pappu, G. R. Marshall, and J. W. Ponder, Nat. Struct. Biol. **6**, 50 (1999).
[8] T. Hirokawa, J. Uechi, H. Sasamoto, M. Suwa, and S. Mitaku, Protein Eng. **13**, 771 (2000).
[9] N. Vaidehi, W. B. Floriano, R. Trabanino, S. E. Hall, P. Freddolino, E. J. Choi, G. Zamanakos, and W. A. Goddard III, Proc. Natl. Acad. Sci. U.S.A. **99**, 12622 (2002).
[10] J. L. Popot and D. M. Engelman, Annu. Rev. Biochem. **69**, 881 (2000).
[11] M. J. Liao, E. London, and H. G. Khorana, J. Biol. Chem. **258**, 9949 (1983).
[12] D. M. Engelman, B. D. Adair, J. F. Hunt, T. W. Kahn, and J. L. Popot, Curr. Topics Membr. Transport **36**, 71 (1990).
[13] P. Argos, J. K. Rao, and P. A. Hargrave, Eur. J. Biochem. **128**, 565 (1982).

[14] D. Eisenberg, R. M. Weiss, T. C. Terwilliger, and W. Wilcox, Faraday Symp. Chem. Soc. **17**, 109 (1982).
[15] K. Nakai and M. Kanehisa, Genomics **14**, 897 (1992).
[16] D. T. Jones, W. R. Taylor, and J. M. Thornton, Biochemistry **33**, 3038 (1994).
[17] B. Rost, R. Casadio, P. Fariselli, and C. Sander, Protein Sci. **4**, 521 (1995).
[18] T. Hirokawa, S. Boon-Chieng, and S. Mitaku, Bioinformatics **14**, 378 (1998).
[19] G. E. Tusnady and I. Simon, J. Mol. Biol. **283**, 489 (1998).
[20] U. H. E. Hansmann and Y. Okamoto, in *Annual Reviews of Computational Physics VI*, edited by D. Stauffer (World Scientific, Singapore, 1999), p. 129.
[21] A. Mitsutake, Y. Sugita, and Y. Okamoto, Biopolymers (Pept. Sci.) **60**, 96 (2001).
[22] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).
[23] K. Hukushima, H. Takayama, and K. Nemoto, Int. J. Mod. Phys. C **7**, 337 (1996).
[24] C. J. Geyer, in *Proceedings of the 23rd Symposium on the Interface*, edited by E. Keramidas (Interface Foundation, Fairfax Station, 1991), p. 156.
[25] E. Marinari, G. Parisi, and J. J. Ruiz-Lorenzo, in *Spin Glasses and Random Fields*, edited by A. P. Young (World Scientific, Singapore, 1998), p. 59.
[26] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).
[27] S. Kumar, D. Bouzida, R. H. Swendesen, P. A. Kollman, and J. M. Rosenberg, J. Comput. Chem. **13**, 1011 (1992).
[28] H. Kokubo and Y. Okamoto, Chem. Phys. Lett. **383**, 397 (2004).
[29] W. Im, M. Feig, and C. L. Brooks III, Biophys. J. **85**, 2900 (2003).
[30] W. E. Reiher III, ''Theoretical studies of hydrogen bonding,'' Ph.D. thesis, Department of Chemistry, Harvard University Cambridge, MA, 1985.
[31] E. Neria, S. Fischer, and M. Karplus, J. Chem. Phys. **105**, 1902 (1996).
[32] A. Mitsutake, Y. Sugita, and Y. Okamoto, J. Chem. Phys. **118**, 6664 (2003).
[33] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).
[34] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
[35] K. R. MacKenzie, J. H. Prestegard, and D. M. Engelman, Science **276**, 131 (1997).
[36] S. O. Smith, D. Song, S. Shekar, M. Groesbeek, M. Zilioz, and S. Aimoto, Biochemistry **40**, 6553 (2001).
[37] H. Kokubo and Y. Okamoto, Chem. Phys. Lett. (in press).