Analysis of Helix-Helix Interactions of Bacteriorhodopsin by Replica-Exchange Simulations

Hironori Kokubo[†] and Yuko Okamoto[‡§*]

[†]Department of Chemistry, University of Houston, Houston, Texas; [‡]Department of Physics, School of Science, Nagoya University, Nagoya, Japan; and [§]JST/BIRD, Nagoya, Japan

ABSTRACT We performed long-time replica-exchange Monte Carlo simulations of bacteriorhodopsin transmembrane helices, which made it possible that wide conformational space was sampled. Using only the helix-helix interactions and starting from random initial configurations, we obtained the nativelike helix arrangement successfully and predicted a part of the configurations (three helices out of seven) precisely. By the principal component analysis we classified low-energy structures into some clusters of similar structures, and we showed that the above nativelike three-helix configuration is reproduced properly in most clusters and that not only the van der Waals interactions but also the electrostatic interactions contributed to the stabilization of the native structures.

INTRODUCTION

Membrane proteins exist in membranes and help to transfer many kinds of molecules in or out of cells selectively. Thus, they are main targets for drug design and are known to play essential roles in many diseases and our sense of vision, smell, etc. Nevertheless, most membrane protein functions are unknown. The number of experimentally determined protein structures is increasing year after year. However, most of them are soluble proteins. Membrane proteins occupy as much as ~25% of the total proteins (1), but the number of experimentally determined structures is $<1\%$ of the total number because of technical difficulties in experiments. We consider that it is important not only to determine structures but also to understand how membrane proteins make their stable native structures for understanding the mechanism of their functions. Prediction of membrane protein structures by molecular simulations is a promising method to solve these two problems at the same time. If we can predict the structures by simulations, we will be able to analyze in a straightforward manner what interactions are important physically and how structures are made.

The transmembrane regions of membrane proteins often consist of only either α-helices or β-sheets. This suggests that the intrachain hydrogen bonds through the formations of α-helices or β-sheets have energetical advantage. The two-stage model (2) assumes that the structure formation of membrane proteins, which have transmembrane helices, can be divided into two different stages. In Stage 1, the helix structure of each transmembrane region is formed independently from one another, and Stage 2 is the helix-helix associations that drive the native tertiary-structure formations. Some experimental results support that this is true for bacteriorhodopsin (3), lactose permease (4), rhodopsin (5), and the red cell anion exchanger protein (6). In this article, we study Stage 2 only, assuming that each transmembrane helix structure is given.

Suwa et al. (7) and Hirokawa et al. (8) developed a method to predict the native bacteriorhodopsin helix positioning by treating transmembrane helices as continuum rods. Although a partial helix positioning could be successfully predicted, it was not an atomic-level structure prediction. Vaidehi et al. (9) developed a method to predict the structures of G protein-coupled receptors. However, their method needed the experimental electron density maps, and the structure search was limited to the rotation around the electron density map.

We recently proposed a prediction method (10–13) for helical membrane protein structures by the replica-exchange method (REM) (14,15), which is the atomic-level structure prediction method and samples a wide configurational space by shuffling helices and searches the global-minimum free energy state. REM can sample a wide configurational space without getting trapped in local-minimum free-energy states and we can find stable structures at low temperatures (for a review, see (16)). The first chemistry and biophysics applications of the method can be found in the literature (17–21). By employing the Monte Carlo (MC) version of REM, we could successfully predict the two-helix configuration of glycophorin A dimer (10–12).

In this article, we present the results of long REM simulations of seven transmembrane helices of bacteriorhodopsin. Preliminary results have been reported elsewhere (13). Bacteriorhodopsin consists of seven transmembrane helices and is one of the G protein-coupled receptors. We show that some helices of bacteriorhodopsin have the ability of assembling to the nativelike structure precisely by themselves from a random initial configuration. We also discuss the importance of the electrostatic interactions as well as the van der Waals (vdW) interactions for the native helix assembly.

Submitted January 8, 2008, and accepted for publication August 6, 2008.

*Correspondence: okamoto@phys.nagoya-u.ac.jp

Editor: Jose Onuchic.

0006-3495/09/02/0765/12 $2.00

doi: 10.1529/biophysj.108.129015

In Methods, the details of the REM simulation for bacteriorhodopsin are explained. We next present Results and Discussion, and the final section gives our Conclusions.

METHODS

Simulation setup

We review essential parts of our methods here briefly (the complete methodological details are given in (13)).

We used only transmembrane helices in our simulations. Our simulations neglect the loop regions of membrane proteins, lipid, and water. Generally speaking, water around membrane surfaces, lipid, and loop regions of bacteriorhodopsin may contribute to the structure stability, but it is not clear how much they play roles in making the native structure. Popot et al. (3) experimentally showed that the seven bacteriorhodopsin transmembrane helices can assemble by themselves into functional proteins.

The coordinates of the seven helices of the transmembrane regions were extracted from the PDB structure of bacteriorhodopsin (PDB code: 1C3W), which were named as A, B, C, D, E, F, and G from the N-terminus. Their sequences are EWIWLALGTALMGLGTLYFLVKG; KFYAITTLVPAIAFT MYLSMLL; IYWARYADWLFTTPLLLLDLALL; QGTILALVGADGI MIGTGLVGAL; RFVWWAISTAAMLYILYVLFFGF; TFKVLRNVTV VLWSAYPVVWLIGSE; and LNIETLLFMVLDVSAKVGFGLILL. The N- and C-termini of each helix were blocked with acetyl and *N*-methyl groups, respectively. We used the CHARMM param19 parameter set (polar hydrogen model) (22,23) for the potential energy of the system. No cutoff was introduced to the nonbonded energy terms.

Each helix structure was first minimized subject to harmonic restraints on all the heavy atoms. We treated the backbone of the α-helices as rigid body and only side-chain structures were made flexible. Each helix also has the freedom of translation and rotation. This is consistent with the two-stage model, in which each individual helix is stable as a domain and the native configurations are built by the interactions among stable 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. We update configurations with rigid translations and rigid rotations of each α-helix and torsion rotations of side chains. There are $2N_{\mathrm{H}} + N_{\mathrm{D}}$ kinds of MC moves, where N_{H} is the total number of transmembrane helices in the protein, and 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 step in this article is defined to be an update of one of these degrees of freedom, which is accepted or rejected according to the Metropolis criterion.

We add the following three simple harmonic constraints to the original potential energy to make conformational sampling efficiency better and mimic the effects of membrane boundary surfaces that are defined by the lipid molecules and surrounding water molecules.

The first harmonic potential constrains pairs of adjacent helices along the amino-acid chain not to be apart from each other too much. The C atom of the C-terminus of the i^{th} helix and the C_α atom of the N-terminus of the $(i+1)^{\mathrm{th}}$ helix are constrained not to be apart from each other by >20 Å. There is no constraint potential imposed within this distance.

The second harmonic potential constrains each helix N-terminus and C-terminus to be located near membrane boundary planes. N-terminus or C-terminus of each helix near the fixed lower membrane boundary and the upper membrane boundary are constrained to remain near the membrane boundary planes. There is no constraint imposed within 2 Å from the planes. If each helix is apart from the planes >2 Å, it is pulled back by this harmonic potential. This constraint is considered to be a simple implicit membrane model, which holds the transmembrane helices in the lipid region.

The third harmonic potential constrains all C_α atoms within the sphere of radius 100 Å. Helices move around in this large sphere freely during simulations. The radius of the sphere is large enough for a wide configurational space sampling of this system. This is introduced to avoid the helices from flying too far away from the origin.

These three harmonic constraints limit sampling space and improve the sampling efficiency. The conformational space, which is appropriate for transmembrane helices, is sampled by these constraints.

Replica-exchange method

These simulations were performed with the dielectric constant $\varepsilon = 1.0$ as in the literature (10–12). We showed in those references that $\varepsilon = 1.0$ is more appropriate to use in the structure prediction. This is because almost no lipid molecules can exist among helices and the value $\varepsilon = 4.0$, which is close to the lipid environment underestimates the electrostatic effects.

The MC simulations were performed with the CHARMM program package (24), and REM was implemented in it (see the review in (16) for details of REM). In Kokubo and Okamoto (13), we performed a REM MC simulation of 168,000,000 MC steps, starting from random helix configurations, and we used 32 replicas and the following 32 temperatures: 200, 218, 238, 260, 284, 310, 338, 369, 410, 455, 505, 561, 623, 691, 768, 853, 947, 1052, 1125, 1202, 1285, 1374, 1469, 1642, 1835, 2051, 2293, 2679, 3132, 3660, 4278, and 5000 K. Although these simulations predicted the structures similar to the native one successfully, the sampling in the previous work (13) was not enough to enable us to discuss the statistical properties.

During that simulation we observed that when the temperature went under 500 K, the helix arrangement changed little, although the energy went down further. This may be because we fixed the backbone structure in our simulations and higher temperatures than usual are necessary for sampling.

We therefore changed the temperature distribution to improve sampling efficiency to: 500, 539, 580, 625, 673, 724, 781, 841, 906, 976, 1051, 1132, 1219, 1313, 1414, 1523, 1641, 1767, 1904, 2051, 2209, 2379, 2562, 2760, 2973, 3202, 3449, 3715, 4001, 4310, 4642, and 5000 K. We used the 32 final structures of the previous simulation in Kokubo and Okamoto (13) as the initial conformations and performed four independent REM simulations with this initial configuration by using four different random number seeds. This time we performed 900,000,000 MC steps in total.

The above temperature distribution was chosen so that all the acceptance ratios of replica exchange are almost uniform and sufficiently large (>10%) for computational efficiency (18,16). The highest temperature was chosen sufficiently high so that no trapping in local-minimum-energy states occurs. Replica exchange was attempted once at every 50 MC steps.

Principal component analysis

We next analyzed these simulation data by the principal component analysis (PCA) (25–29). At first 8800 conformations were chosen at each temperature from the replica-exchange simulations. The structures were taken from the trajectories at even intervals. Each structure is superimposed on the arbitrary reference structure. In this work we chose the structure of PDB structure (1C3W) as the reference structure. We then calculated the average structure of $N = 8800$ structures and superimposed every structure on this average structure at each temperature. We define the variance-covariance matrix as

$$C_{ij} = \left\langle \left(q_i - \langle q_i \rangle\right)\left(q_j - \langle q_j \rangle\right)\right\rangle, \tag{1}$$

where $\vec{q} = (q_1, q_2, q_3, q_4, q_5, q_6, \ldots, q_{3n-2}, q_{3n-1}, q_{3n}) = (x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_n, y_n, z_n)$ and $\langle \vec{q} \rangle = \Sigma_{k=1}^{N} \vec{q}(k)/N$, x_i,y_i,z_i are Cartesian coordinates of the i^{th} atom, and n is the total number of atoms. This symmetric $3n \times 3n$ matrix was diagonalized and the eigenvectors and eigenvalues were obtained. The first superposition is performed to remove large eigenvalues from the translations and rotations of the system because we want to analyze the internal differences of structures. Therefore, the six eigenvalues from the smallest one are very close to zero within the limit of arithmetic precision ($\sim 1.0 \times 10^{-12}$). We order the eigenvalues in the decreasing order of magnitude. The first, second, and third principal component axes are thus defined as the eigenvectors corresponding to the largest, second-largest, and third-

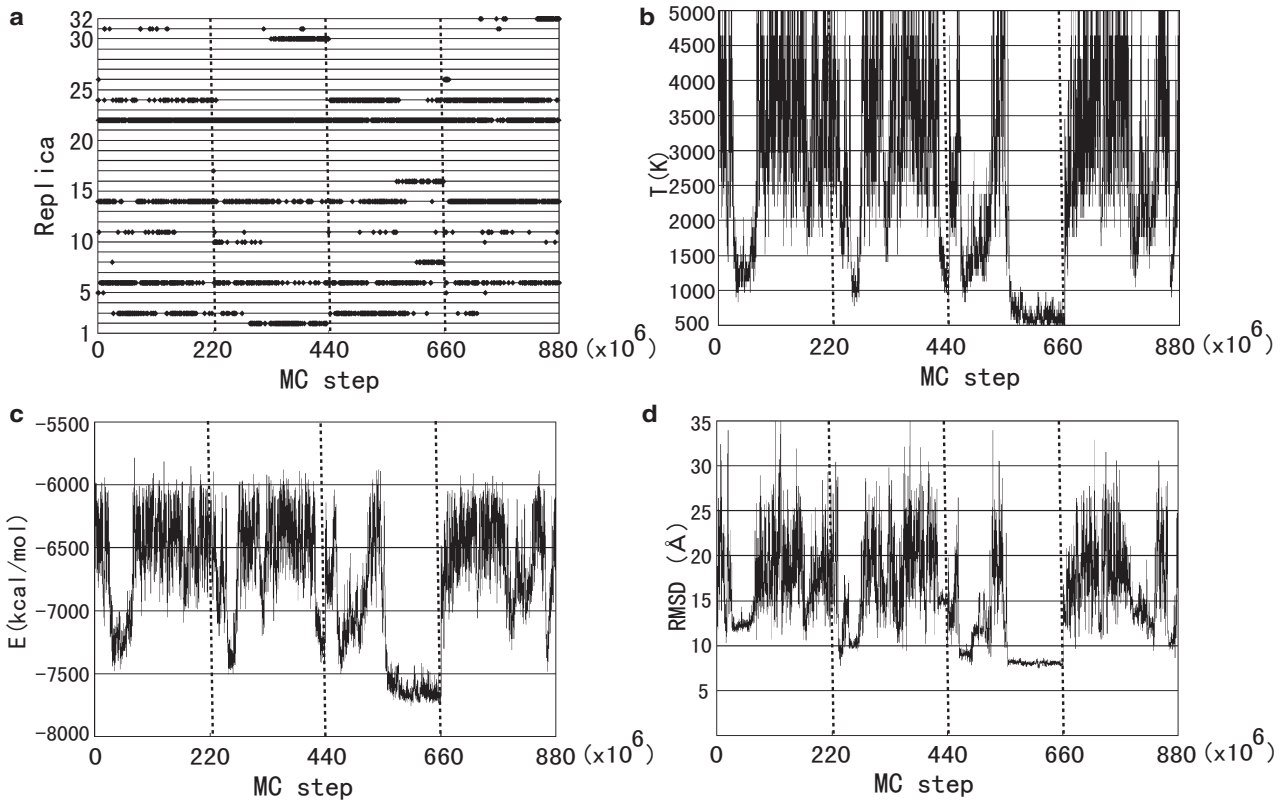

FIGURE 1 Time series of various quantities. (*a*) Replica exchange at $T = 500$ K, (*b*) temperature exchange for one of the replicas (Replica 16), (*c*) the total potential energy for Replica 16, and (*d*) the RMS deviation (in Å) of backbone atoms from the PDB structure for Replica 16. We added three vertical dotted lines to show that these simulations consist of four independent runs. In panel *a*, it may appear that more than one point occupy each time step because many points are plotted. However, only one point (one replica) occupies 500 K at each time step.

largest eigenvalues, respectively. The i^{th} principal component of each sampled structure $\vec{q}$ is defined by the inner product

$$\mu_i = \vec{v}_i \cdot (\vec{q} - \langle \vec{q} \rangle), \ (i = 1, 2, \ldots), \tag{2}$$

where $\vec{v}_i$ is the (normalized) i^{th} eigenvector.

RESULTS AND DISCUSSION

Time series properties of REM simulations

We first examine how the replica-exchange simulations performed. In Fig. 1, we show the time series of these REM simulations. Starting from the same initial states, four independent REM MC simulations were performed with four different random number seeds. Each simulation consisted of 220×10^6 MC steps. Each figure of Fig. 1 is plotted by connecting the four independent runs. In Fig. 1 *a*, the time series of replica exchange at the lowest temperature ($T = 500$ K) is shown. We see that many replicas take the lowest temperature many times, and we observe a random walk in the replica space. These simulations were performed by using 512 CPUs on a Hitachi Super Technical Server SR11000 at the Institute for Molecular Science for a month.

The complementary picture is the temperature exchange for each replica. The result for one of the replicas (Replica 16) is shown in Fig. 1 *b*. We observe random walks in the temperature space between the lowest and highest temperatures. Other replicas perform random walks similarly. In Fig. 1 *c* the corresponding time series of the total potential energy is shown. We see that a random walk in the potential energy space between low and high energy regions is also realized. Note that there is a strong correlation between the behaviors in Fig. 1, *b* and *c*. In Table 1, we list the acceptance

TABLE 1 Acceptance ratios of replica exchange corresponding to pairs of neighboring temperatures

Pairs of T	Acceptance ratio	Pairs of T	Acceptance ratio
500 ↔ 539 K	0.34	1641 ↔ 1767 K	0.43
539 ↔ 580 K	0.36	1767 ↔ 1904 K	0.43
580 ↔ 625 K	0.36	1904 ↔ 2051 K	0.45
625 ↔ 673 K	0.36	2051 ↔ 2209 K	0.46
673 ↔ 724 K	0.35	2209 ↔ 2379 K	0.47
724 ↔ 781 K	0.33	2379 ↔ 2562 K	0.49
781 ↔ 841 K	0.35	2562 ↔ 2760 K	0.51
841 ↔ 906 K	0.35	2760 ↔ 2973 K	0.53
906 ↔ 976 K	0.35	2973 ↔ 3202 K	0.55
976 ↔ 1051 K	0.35	3202 ↔ 3449 K	0.57
1051 ↔ 1132 K	0.35	3449 ↔ 3715 K	0.58
1132 ↔ 1219 K	0.35	3715 ↔ 4001 K	0.59
1219 ↔ 1313 K	0.36	4001 ↔ 4310 K	0.60
1313 ↔ 1414 K	0.38	4310 ↔ 4642 K	0.60
1414 ↔ 1523 K	0.39	4642 ↔ 5000 K	0.61
1523 ↔ 1641 K	0.41		

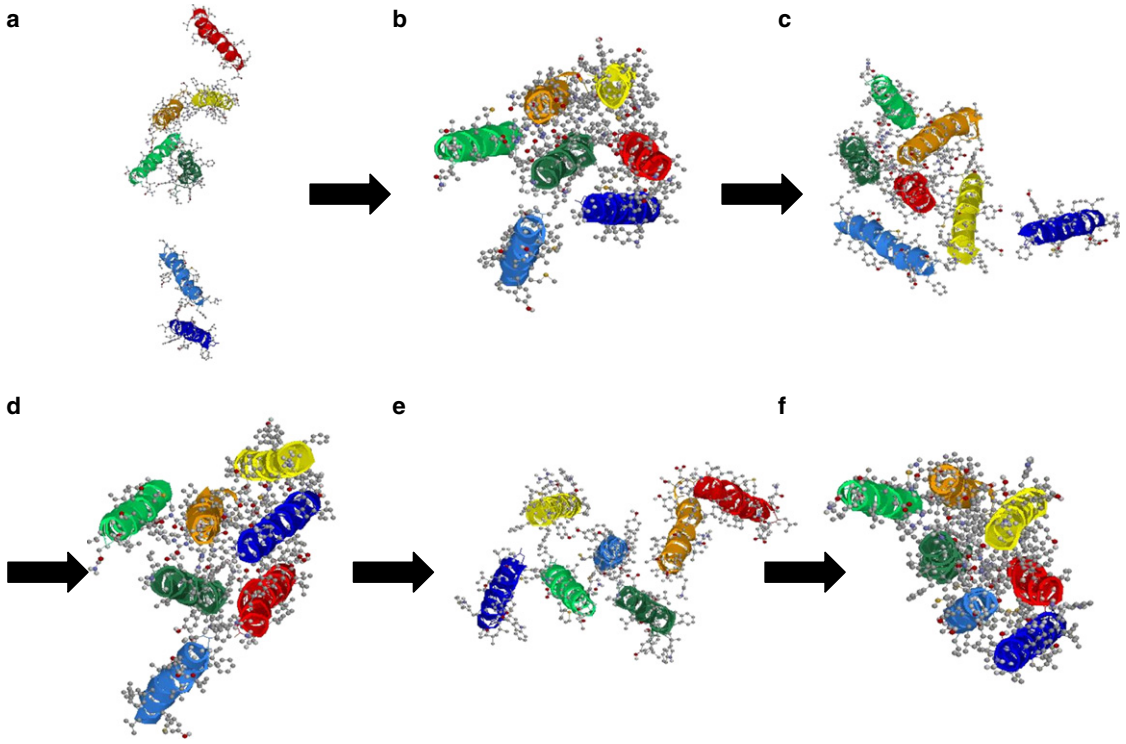

FIGURE 2 Typical snapshots from the REM simulations. Panels *a–f* are structures at 460, 480, 500, 520, 540, and 560 ($\times 10^6$) MC steps in Fig. 1, *b–d* (Replica16), respectively. The corresponding temperatures are 4001 K (*a*), 1414 K (*b*), 1767 K (*c*), 1904 K (*d*), 5000 K (*e*), and 724 K (*f*). The color of the helices from the N-terminus is as follows: Helix A (*blue*), Helix B (*aqua*), Helix C (*green*), Helix D (*yellow-green*), Helix E (*yellow*), Helix F (*orange*), and Helix G (*red*). The figures were created with RasMol (32).

ratios of replica exchange between all pairs of neighboring temperatures. We find from Table 1 that the acceptance ratios are indeed uniform and large.

We also calculated the number of tunneling events (30,31), which gives a measure for the quality of the random walk in the potential energy space. One tunneling event is defined by a trajectory that goes from E_H to E_L and back, where E_H and E_L are the values near the highest energy and the lowest energy, respectively, which the random walk can reach. If E_H is sufficiently high, the trajectory gets completely uncorrelated when it reaches E_H. On the other hand, when the trajectory reaches near E_L, it tends to get trapped in local-minimum

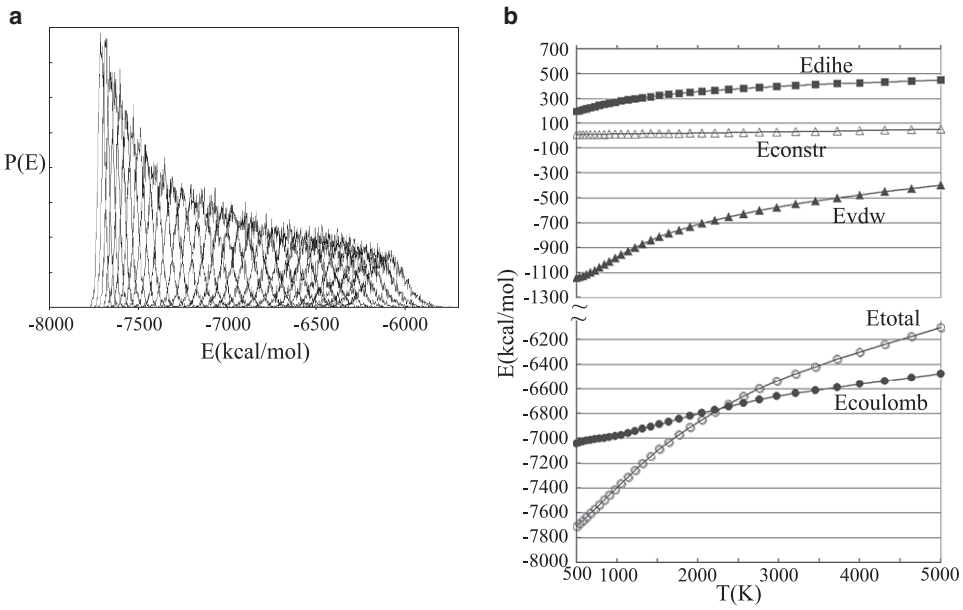

FIGURE 3 (*a*) The canonical probability distributions of the total potential energy obtained from the replica-exchange MC simulation at the 32 temperatures. The distributions correspond to the following temperatures (from *left* to *right*): 500, 539, 580, 625, 673, 724, 781, 841, 906, 976, 1051, 1132, 1219, 1313, 1414, 1523, 1641, 1767, 1904, 2051, 2209, 2379, 2562, 2760, 2973, 3202, 3449, 3715, 4001, 4310, 4642, and 5000 K. (*b*) The averages of the total potential energy E_{total} and its component terms: electrostatic energy $E_{coulomb}$, van der Waals energy E_{vdw}, dihedral energy E_{dihe}, and constraint energy E_{constr} as functions of temperature.

states. We thus consider that the more tunneling events we observe during a fixed number of MC steps, the more efficient the method is as a generalized-ensemble algorithm (or, the average quantities obtained by the reweighting techniques are more reliable). Here, we took E_H as the average potential energy at the highest temperature (5000 K) and E_L as the average potential energy at the lowest temperature (500 K). The total number of the tunneling events was nine, which is good enough for this purpose.

All these results confirm that these REM simulations have properly performed. We next examine how widely the configurational space was sampled during these simulations. We plot the time series of the root mean-square deviation (RMSD) of all the C_α atoms from the experimental structure (PDB code: 1C3W) in Fig. 1 *d*. When the temperature becomes high, the RMSD takes large values, and when the temperature becomes low, the RMSD takes small values. By comparing Fig. 1, *c* and *d*, we see that there is a strong correlation between the total potential energy and the RMSD values. The fact that RMSD at high temperatures is large implies that our simulations did not get trapped in local-minimum potential-energy states.

In Fig. 2, typical snapshots of Replica 16 from the REM simulations are shown. In Fig. 2 *a*, the helix configuration is quite different from the native one (see Fig. 7 below). As the simulation proceeds, the temperature becomes higher and then drops to low values through the replica-exchange process, and the compact structure is obtained in Fig. 2 *f*. Although the conformation of Helices E (*yellow*) and F (*orange*) is different from the native one, the rest of the helix configuration is similar to the native one (compare Fig. 2 *f* with Fig. 7 below). These figures confirm that our simulations indeed sampled a wide configurational space. We see that the REM simulation performed random walks not only in energy space but also in conformational space and that they did not get trapped in one of a huge number of local-minimum-energy states.

Average energies as functions of temperatures

In Fig. 3 *a*, the canonical probability distributions of the total potential energy obtained at the chosen 32 temperatures from the REM simulations are shown. We see that there are enough overlaps between all neighboring pairs of distribution. This ensures that the number of replicas is sufficient. In Fig. 3 *b*, the average of the total potential energy E_{total} and averages of its component terms, namely, the electrostatic energy $E_{coulomb}$, van der Waals energy E_{vdw}, torsion energy E_{dihe} (these three terms are from the CHARMM force field (22,23)), and constraint energy E_{constr} as functions of temperature T are shown. At high temperatures the helices are generally far apart from each other because the entropy effects are dominant. At low temperatures, on the other hand, we observe the side-chain packing among helices. We see that as the temperature becomes lower, E_{vdw}, E_{dihe},

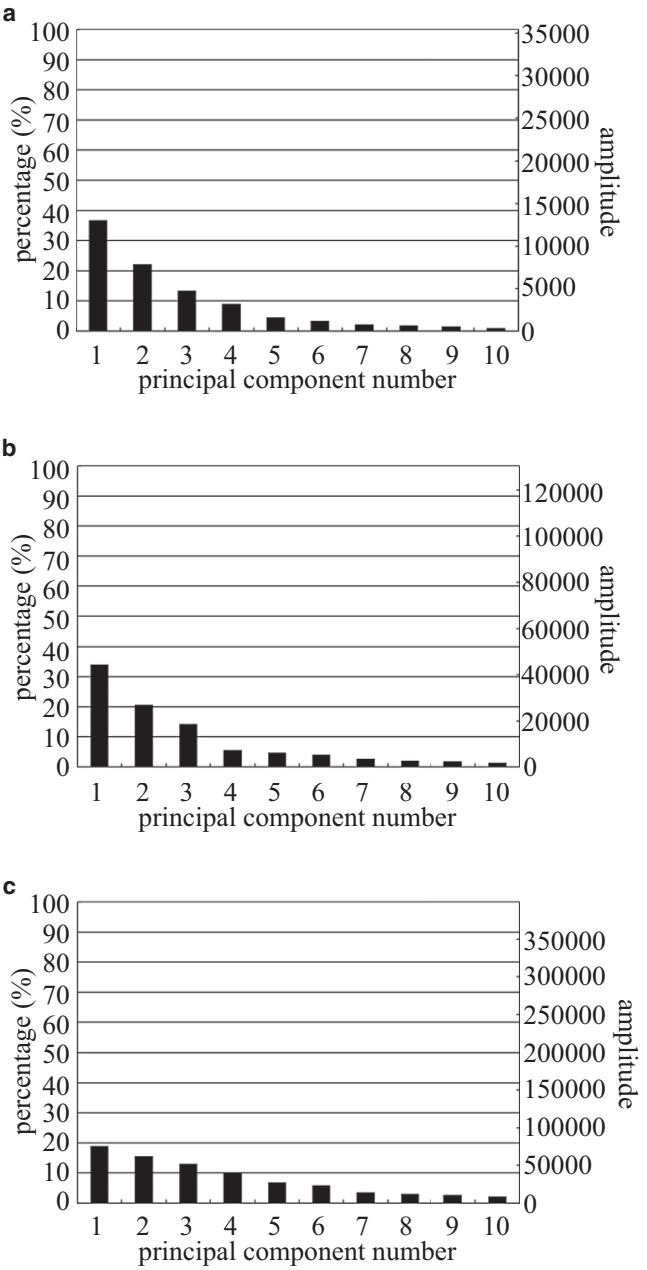

FIGURE 4 The percentage (on the *left ordinate*) and the amplitude (on the *right ordinate*) of the principal components from the structures in the replica-exchange simulations at the following temperatures: 500 K (*a*), 976 K (*b*), and 5000 K (*c*).

and $E_{coulomb}$ decrease almost linearly up to ~1500 K, and as a result E_{total} is also almost linearly decreasing up to ~1500 K. On the other hand, when the temperature becomes <1500 K, E_{vdw} and E_{dihe} contribute more to the decrease of E_{total}. This is reasonable because E_{vdw} and E_{dihe} decrease as a result of side-chain packing and the stability of the conformation increases. The contribution from E_{constr} is always small at every temperature.

We used only transmembrane regions in these simulations. Transmembrane helices are generally considered to

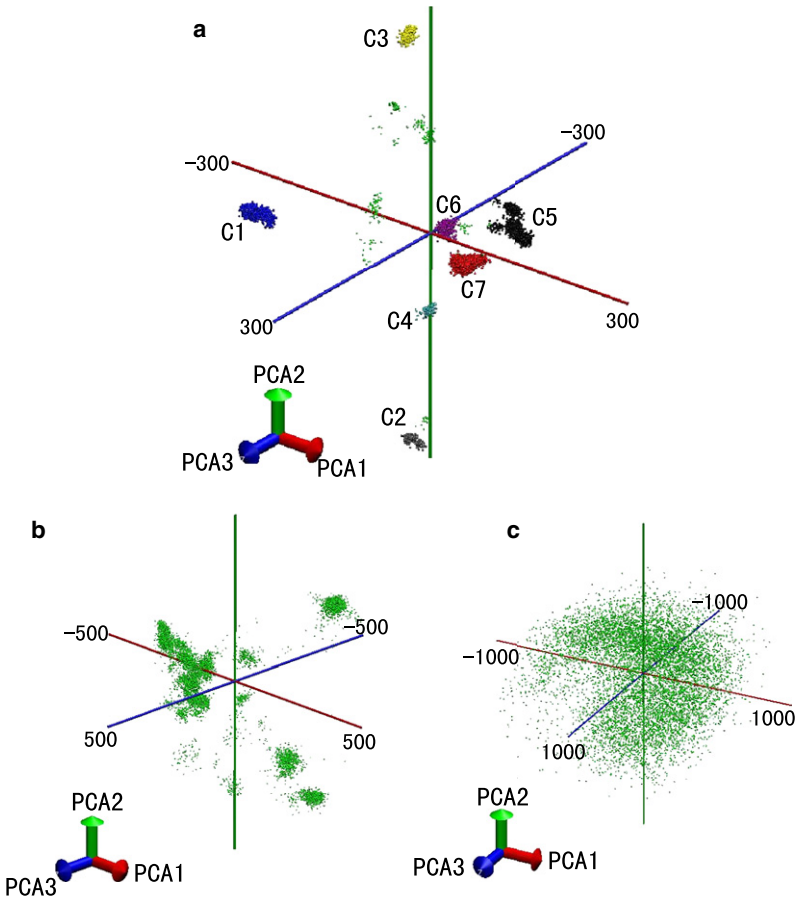

FIGURE 5 The projection of the sampled structures on the first, second, and third principal axes from the replica-exchange simulations at 500 K (*a*), 976 K (*b*), and 5000 K (*c*). Only structures at 500 K (*a*) are classified into clusters of similar structures and used for the detail analyses. C1, C2, ⋯, C7 in panel *a* stand for Clusters 1, 2, ⋯, 7, respectively: C1 (*blue*), C2 (*gray*), C3 (*yellow*), C4 (*aqua*), C5 (*black*), C6 (*purple*), and C7 (*red*). Here, in panel *a*, a small portion of structures which do not belong to any of the seven clusters are shown in green and were excluded from the analyses. The structures in panels *b* and *c* were not used in the cluster analyses and they are also plotted in green. PCA1, PCA2, and PCA3 represent the directions of the principal axes 1, 2, and 3, respectively.

be hydrophobic, and helix-helix association is sometimes considered only by vdW packing (lock-and-key model). However, Fig. 3 *b* shows that E_{coulomb} also changes much as a function of temperature. This implies that electrostatic effects among helices can never be neglected. They also contribute to the formation of the native helix configuration. We discuss this point further below.

Classification of the sampled structures by PCA

We next employed the principal component analysis (PCA) (25–29) to classify the sampled structures into clusters of similar structures. In Fig. 4 we show the percentage (the left ordinate) and amplitude (the right ordinate) of the first 10 principal components at the chosen three temperatures of 500 (lowest), 976, and 5000 K (highest). The amplitudes correspond to the eigenvalues of the variance-covariance in Eq. 1. We see from the percentage values in Fig. 4 that as the temperature becomes higher, more principal component axes are needed to represent the fluctuations of the system, as it is expected. We observe from the amplitude values in Fig. 4 that the amplitude becomes larger as the temperature becomes higher. This is reasonable because as the temperature becomes higher, the fluctuations of the system become larger and the simulations sample a wider conformational space. In Fig. 4 *a*, we see that >70% of the total amplitude is expressed by the first three principal components. Although we can express the system more precisely as we use more principal axes, we here classify and analyze the sampled structures at the lowest temperature by the first three principal components. The fact that most of the amplitudes of fluctuations in this protein system is represented only by a small number of principal components, is consistent with that protein folding dynamics can be expressed as the diffusion over a low-dimensional free energy surface as is elucidated in the energy landscape theory (33). Fig. 4 *c* shows that many principal component axes are needed to express the sampled structures properly at the highest temperature. The sampled structures are sometimes analyzed by other reaction coordinates such as native contact, RMSD, and radius of gyration. These are suitable in some cases but may not be appropriate in others. We do not know how many reaction coordinates we need to be able to identify important local-minimum free energy states in the free energy landscape. The principal component analysis is one of the methods that naturally provide us with the information as to how many reaction coordinates we need for such investigations.

In Fig. 5, the structures obtained from the replica-exchange simulation are projected on the first, second, and third principal component axes at the chosen three temperatures.

TABLE 2 Various average properties of the structures classified by the principal component analysis, obtained at the temperature of 500 K by the REM simulations

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	500 K
Number	1420	274	739	176	1385	368	4197	8800
$\langle E_{total} \rangle$	−7707.7 ± 17.7	−7715.8 ± 17.9	−7697.7 ± 16.0	−7712.1 ± 18.1	−7705.1 ± 16.7	−7717.1 ± 17.8	−7715.1 ± 17.9	−7710.5 ± 18.6
$\langle E_{vdw} \rangle$	−1144.2 ± 14.2	−1152.9 ± 14.1	−1149.1 ± 14.3	−1158.9 ± 14.3	−1136.7 ± 15.7	−1147.8 ± 15.4	−1142.4 ± 14.7	−1142.9 ± 15.8
$\langle E_{coulomb} \rangle$	−7038.4 ± 13.1	−7045.6 ± 12.7	−7026.9 ± 11.9	−7029.8 ± 12.8	−7045.3 ± 12.8	−7047.4 ± 13.9	−7042.0 ± 13.3	−7041.1 ± 14.8
$\langle E_{dihe} \rangle$	195.3 ± 10.5	200.9 ± 10.8	203.4 ± 10.5	203.2 ± 10.9	196.3 ± 11.2	199.2 ± 10.7	193.8 ± 10.3	196.2 ± 11.1
$\langle E_{constr} \rangle$	8.5 ± 3.1	10.7 ± 2.7	3.8 ± 1.6	2.3 ± 1.1	9.4 ± 4.1	7.7 ± 2.4	4.4 ± 1.7	6.2 ± 3.5
⟨RMSD⟩	4.91 ± 0.21	11.86 ± 0.14	8.85 ± 0.07	10.17 ± 0.08	8.00 ± 0.35	7.63 ± 0.24	7.14 ± 0.11	7.37 ± 1.56

"Number" stands for the numbers of conformations that belong to each cluster. The following abbreviations are used: the total potential energy, E_{total}; van der Waals energy, E_{vdw}; electrostatic energy, $E_{coulomb}$; dihedral energy, E_{dihe}; constraint energy, E_{constr} (all in kcal/mol); and root-mean-square deviation of all C_α atoms RMSD (in Å). The values after ± are the standard deviations. The entries under "500 K" stand for the results of the average over all the 8800 structures used in the cluster analysis at $T = 500$ K.

As the temperature becomes higher, these clusters become less distinguishable, and a wider conformational space is sampled without getting trapped in local-minimum free energy states. At the highest temperature there are no distinct clusters.

If we perform constant temperature simulations at the lowest temperature, the simulations will get trapped in one of the clusters in Fig. 5 *a*, depending on the initial configurations of the simulations. However, each replica of the replica-exchange simulations will not get trapped in one of the local-minimum free energy states, by going through high-temperature regions. Every replica can climb over energy barriers in Fig. 5 *c* by temperature exchange during simulations. This is the reason why we adopted replica-exchange simulations. We classified most of the sampled structures at the lowest temperature into seven distinct clusters in Fig. 5 *a*. They lie in the ranges (−263 ~ −222; −59 ~ −27; 0 ~ 42); (165 ~ 194; −130 ~ −111; 249 ~ 267); (−18 ~ 2; 262 ~ 282; 17 ~ 52); (95 ~ 119; −30 ~ −8; 121 ~ 154); (0 ~ 98, −43 ~ 1, −133 ~ −61); (0 ~ 38, −11 ~ 28, −26 ~ 14); and (23 ~ 80; −45 ~ −10; −31 ~19), which we refer to as Cluster 1, Cluster 2, …, and Cluster 7, respectively. A total of 8559 structures from the total 8800 structures belong to one of these clusters. Note that the reason why we classified the structures in this way is that the largest part of the total amplitude is represented by the first three principal components, as was shown in Fig. 4 *a*.

Average properties of clusters

Table 2 lists various average properties of the seven clusters of similar structures. The column "500 K" represents the various average values over all 8800 structures corresponding to the temperature 500 K used for the cluster analysis. These 8800 structures were extracted from the trajectories at even intervals. The columns of Cluster 1, Cluster 2,…, and Cluster 7 represent the various average values over the structures that belong to each cluster. We see from the value in the "Number" column in Table 2 that Cluster 7 was sampled most frequently. Therefore, Cluster 7 is the global-minimum free energy state in our simulations. Other clusters are considered to be local-minimum free energy states. Unfortunately, judging from the RMSD values, Cluster 7 is not close to the native structure, and it is instead one of the local-minimum free energy states. Cluster 1 instead is close to the native structure. Note that the standard deviation of RMSD of each cluster became much smaller than that before clustering. This implies that the structures sampled are properly divided into seven clusters. The structures included

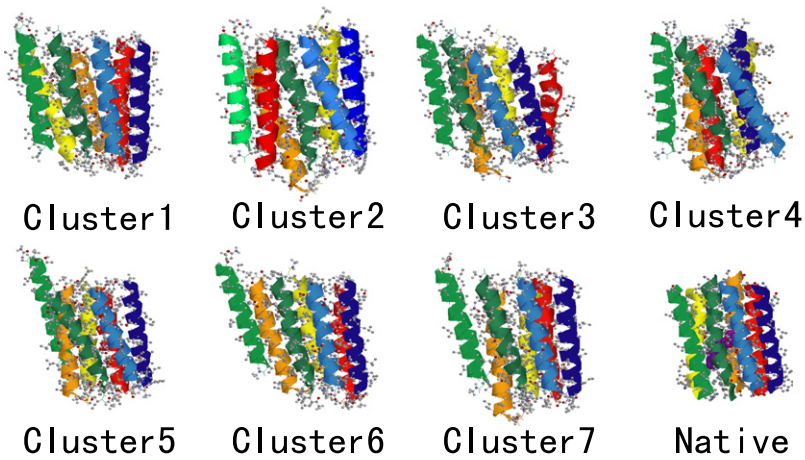

FIGURE 6 Typical structures of clusters obtained from the REM simulations (side view). The RMSD from the native configuration is 4.85 Å, 7.19 Å, 8.86 Å, 8.51 Å, 13.67 Å, 10.14 Å, 8.40 Å, 7.55 Å, 8.72 Å, 7.77 Å, and 7.10 Å with respect to all C_α atoms. See also the legend of Fig. 2.

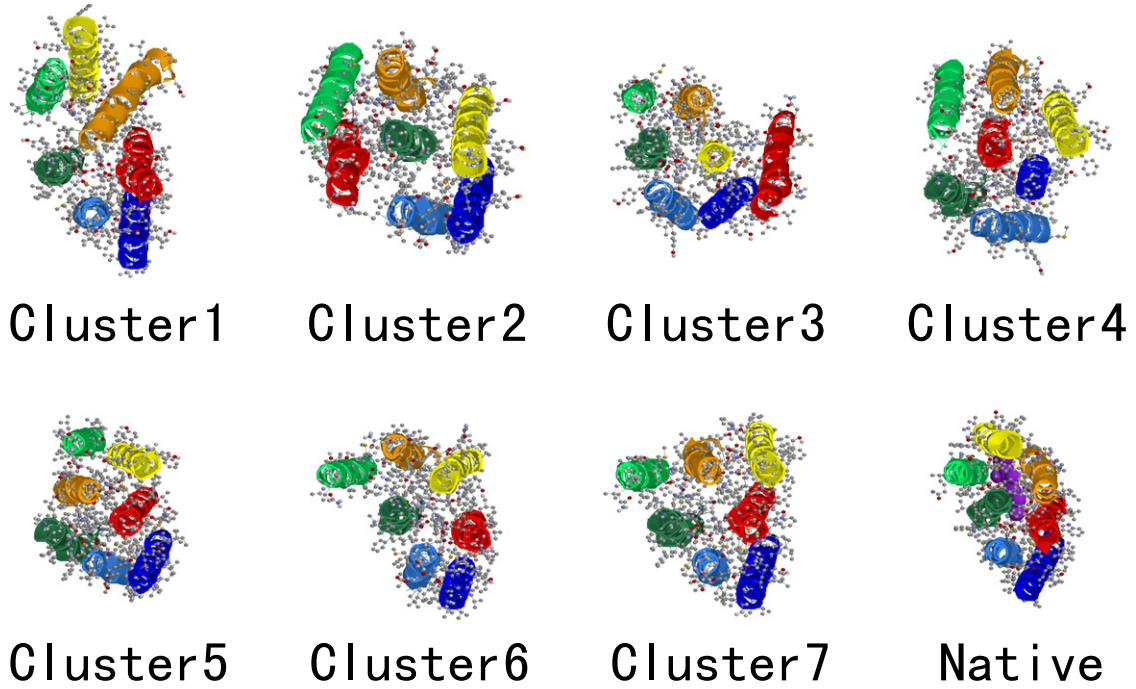

FIGURE 7 Typical structures of clusters obtained from the REM simulations (*top view*). See also the legends of Figs. 2 and 6.

in each cluster have essentially the same backbone structures because of this nature.

Can helices self-assemble only by helix-helix interactions?

Typical structures of seven clusters in Table 2 and the experimental structure (PDB code: 1C3W) are shown in Figs. 6 and 7 by the side and top views, respectively. The purple molecule in the center of bacteriorhodopsin of the native structure is a retinal molecule. Remember that Cluster 1 has the smallest RMSD. In fact the overall helix configuration (or, topology) is identical with that of the native one. Nevertheless, it is not clear which part really resembles the native one yet.

Therefore, we list in Table 3 the RMSD from the native structure with respect to C_α atoms of neighboring pairs of helices to examine which parts resemble the native ones closely. It is now clear that the RMSD of helix pairs A-B, A-G, and B-G in Clusters 1, 5, 6, and 7 are very small (they vary from 0.24 Å to 0.64 Å). This implies that these parts greatly resemble the native structure.

One of the possible reasons why our predictions are poorer for the partial structures of Helices C, D, E, and F is because we did not include the retinal molecule in our simulations. The native experimental structures in Fig. 7 suggest that a retinal molecule exists among these helices, and the retinal seems to be necessary for predicting the perfect bacteriorhodopsin structure of these four helices. The RMSD values of Helices B-C and C-G of Clusters 1, 5, 6, and 7 are also small ($<$2.0 Å), which is smaller relatively, but it is not clear how close they really are only by the RMSD values. We next check how much these parts resemble the native ones.

In Fig. 8, we compared a typical Cluster 7 structure with the experimental one to see how much they resemble each

TABLE 3 The RMSD (in Å) of pairs of helices from those of the corresponding PDB structure with respect to all C_α atoms

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	500 K
⟨RMSD-AB⟩	0.27 ± 0.12	0.33 ± 0.15	4.26 ± 0.08	4.45 ± 0.09	0.24 ± 0.07	0.27 ± 0.14	0.27 ± 0.08	0.80 ± 1.43
⟨RMSD-BC⟩	1.84 ± 0.20	4.19 ± 0.11	3.40 ± 0.06	3.39 ± 0.05	1.24 ± 0.22	1.47 ± 0.38	1.56 ± 0.20	1.88 ± 0.83
⟨RMSD-CD⟩	4.21 ± 0.57	5.41 ± 0.10	2.50 ± 0.11	5.36 ± 0.19	7.02 ± 0.90	4.95 ± 0.27	3.71 ± 0.58	4.39 ± 1.44
⟨RMSD-DE⟩	3.63 ± 0.29	8.41 ± 0.42	7.08 ± 0.09	7.42 ± 0.19	7.22 ± 0.91	8.51 ± 0.30	7.44 ± 0.26	6.81 ± 1.52
⟨RMSD-EF⟩	5.38 ± 0.23	4.75 ± 0.23	4.15 ± 0.06	4.61 ± 0.07	4.72 ± 0.51	4.59 ± 0.68	4.81 ± 0.34	4.81 ± 0.47
⟨RMSD-FG⟩	4.11 ± 0.13	6.91 ± 0.15	6.10 ± 0.23	4.44 ± 0.10	3.26 ± 0.58	4.17 ± 1.11	3.27 ± 0.29	3.85 ± 1.10
⟨RMSD-AG⟩	0.48 ± 0.21	6.99 ± 0.20	4.17 ± 0.13	2.05 ± 0.06	0.36 ± 0.06	0.41 ± 0.11	0.39 ± 0.07	1.06 ± 1.78
⟨RMSD-BG⟩	0.64 ± 0.36	4.94 ± 0.11	7.10 ± 0.16	3.96 ± 0.10	0.41 ± 0.09	0.55 ± 0.12	0.50 ± 0.09	1.42 ± 2.16
⟨RMSD-CF⟩	4.85 ± 0.16	4.47 ± 0.21	2.78 ± 0.11	5.62 ± 0.18	2.83 ± 0.24	1.88 ± 0.48	3.36 ± 0.31	3.54 ± 0.94
⟨RMSD-CG⟩	2.00 ± 0.18	4.95 ± 0.11	7.28 ± 0.17	1.34 ± 0.15	1.42 ± 0.30	1.76 ± 0.41	1.72 ± 0.20	2.41 ± 1.81

Transmembrane helices are named as A, B, C, D, E, F, and G from the N-terminus. The values after ± are the standard deviations. The entries under "500 K" stand for the results of the average over all the structures used in the cluster analysis at $T = 500$ K.

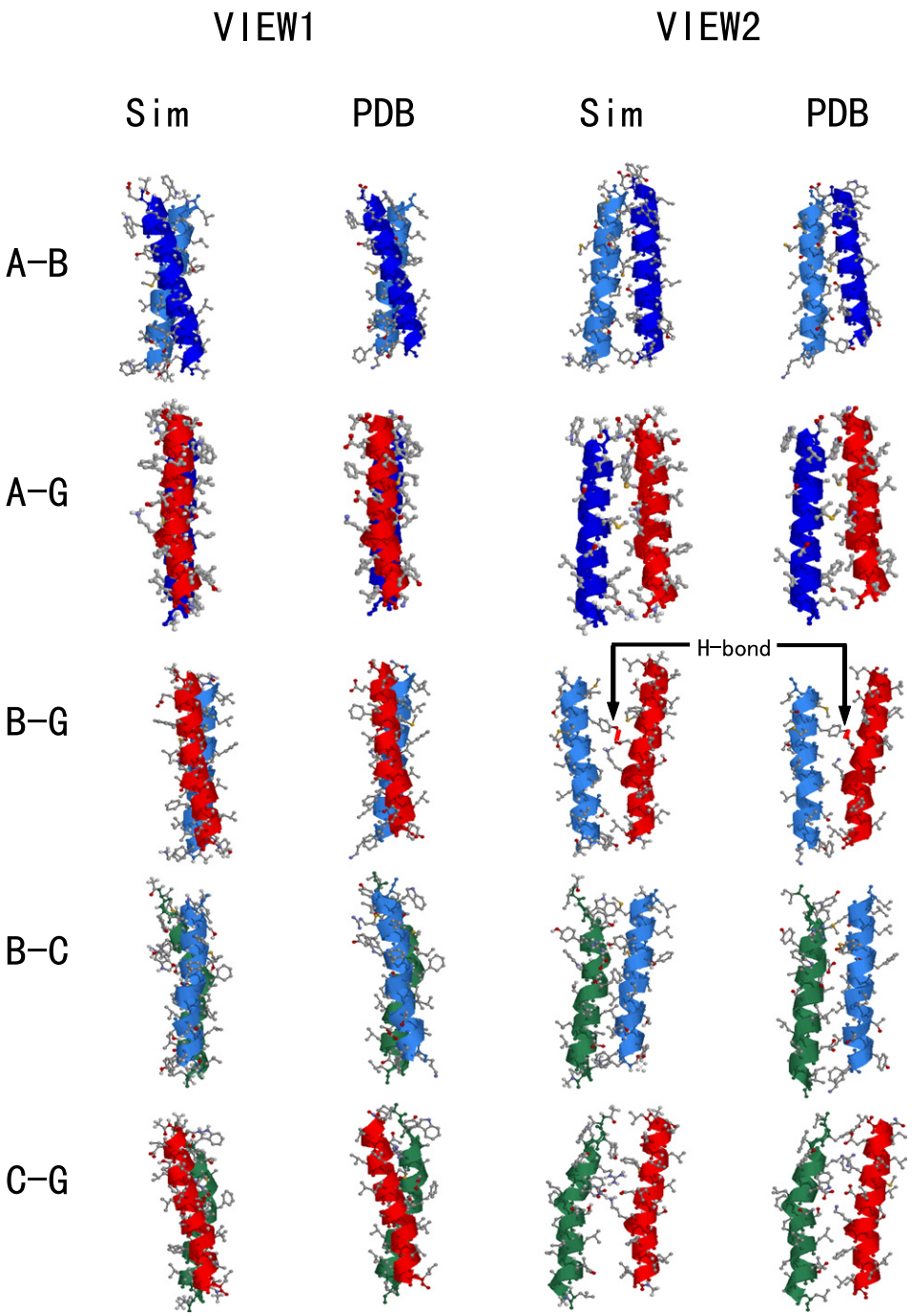

FIGURE 8 The structures of some pairs of helices of Cluster 7 and Native in Fig. 6. Sim and PDB express the partial structures of Cluster 7 and PDB (1C3W), respectively. Helix colors are the same as in Fig. 2. For each entry, two figures of the same structure viewed from different angles (VIEW1 and VIEW2) are shown. The figures were created with RasMol (32).

other. The five pairs of helices (A-B, A-G, B-G, B-C, and C-G) that have low RMSD values in Table 3 are depicted and compared with the corresponding native helix pairs. These five pairs of helices are shown in ascending order of RMSD. We confirm from this figure that the structures of the first three pairs of helices (A-B, A-G, and B-G) obtained by the simulations are in remarkable agreement with those of the PDB structure including some side-chain packing. It means that Clusters 1, 5, 6, and 7 (these occupy 83.75% of the total conformations) have partial structures similar to the native ones about Helices A, B, and G. On the other hand, we see that side-chain packings are somewhat different from the experimental ones for the conformations of B-C and C-G. The red marks between Helices B and G (*right side*) in Fig. 8 represent the interhelix hydrogen bond, which is between Tyr OH of Helix B and Asn O of Helix G. This hydrogen bond was successfully predicted by our simulations. This is the only interhelix hydrogen bond that exists in Helices A, B, and G of the native structure.

It is still possible that configurations of Helices A, B, and G were not disordered enough to achieve the random helix shuffling at high temperatures during the simulations even

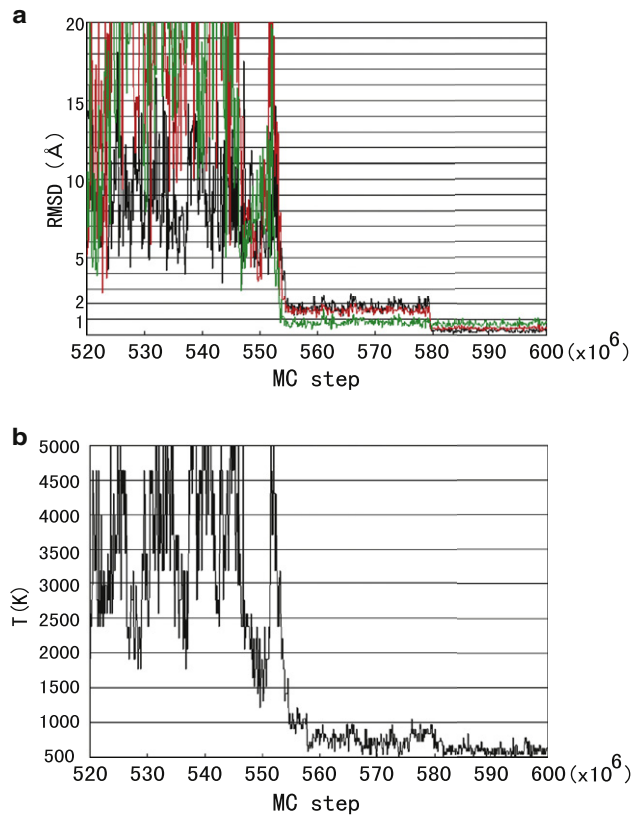

FIGURE 9 (*a*) Time series of the RMSD (in Å) of backbone atoms of helix pairs from the PDB structure for Replica 16. Black, red, and green curves correspond to helix pairs A and B, A and G, and B and G, respectively. (*b*) The corresponding time series of temperature for Replica 16.

if other helices moved much. To confirm that this partial structure was really recovered from random conformations, we show in Fig. 9 *a*, the time series of RMSD of the helix pairs (A-B, A-G, and B-G) from the experimental ones in the case of Replica 16. The corresponding time series of temperature is shown in Fig. 9 *b*. We see random walks in RMSD value and this partial structure was actually recovered during the simulations from very different configurations. We confirmed that some other replicas also behaved like this. Remember that we considered only helix-helix interactions in our model. Therefore, this implies that these helices themselves can assemble only by the helix-helix interactions.

It is now clear that we could predict the structures of Helices A, B, and G precisely without the use of the native structure information of the helix arrangement of bacteriorhodopsin. Here, we focus our discussion only on Helices A, B, and G, which were successfully predicted by our simulations. Helices A and B in Clusters 1, 2, 5, 6, and 7 have the structures close to the native ones. We refer to these clusters as native clusters for Helices A and B. On the other hand, those in Clusters 3 and 4 have structures different from the native ones. We refer to these as nonnative clusters for Helices A and B. Similarly, Clusters 1, 5, 6, and 7 are native clusters, and Clusters 2, 3, and 4 are nonnative clusters for Helices A and G. Clusters 1, 5, 6, and 7 are native clusters, and Clusters 2, 3, and 4 are nonnative clusters for Helices B and G. What makes the structures of these helix pair structures in native clusters more stable than those of nonnative ones? Are the packing interactions by vdW energy dominant? In Tables 4–6, we list various average potential energies of the structures of Helices A and B, A and G, and B and G, respectively. We see in Table 4 that the electrostatic energy of the nonnative clusters (Clusters 3 and 4) for Helices A and G is less than that of native clusters (Clusters 1, 2, 5, 6, and 7), although the vdW energy is larger. Therefore, the native clusters are stabilized by the vdW interactions instead of the electrostatic interactions. This implies that the native structure of Helices A and B is the best packing structure. The native clusters also have the least total potential energy, and it is expected that the native structure of Helices A and B can self-assemble only by these two helices. We confirmed this by performing replica-exchange simulations of only Helices A and B (data not shown).

On the other hand, in Table 5 we see that the native clusters (Clusters 1, 5, 6, and 7) for Helices A and G have less electrostatic energy than other nonnative clusters. The vdW energy of the native clusters is not necessarily less than that of the nonnative clusters. Therefore, the contribution from the electrostatic energy instead of the vdW energy stabilizes the native structure of Helices A and G. In this case, the native structure of Helices A and G is not necessarily the best packing structure. The native clusters also have the lowest total potential energy, so it is expected that the native structure of Helices A and G can self-assemble only by themselves. We again confirmed it by performing replica-exchange simulations of only Helices A and G (data not shown).

Table 6 lists the case of the helix pair B and G. Native clusters for Helices B and G are Clusters 1, 5, 6, and 7, and nonnative clusters are Clusters 2, 3, and 4. The vdW energy of the native clusters is less than that of nonnative clusters, but the electrostatic energy of the native clusters is larger than that

TABLE 4 Various average energies (in kcal/mol) of the partial structures of the helix pair A and B classified by the principal component analysis, which were obtained at the temperature of 500 K by the REM simulations

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
$\langle E_{\text{total}} \rangle$	−1874.9 ± 9.6	−1864.1 ± 9.5	−1854.6 ± 10.0	−1852.4 ± 11.2	−1881.7 ± 9.0	−1876.4 ± 9.1	−1878.9 ± 8.7
$\langle E_{\text{vdw}} \rangle$	−248.0 ± 6.8	−258.3 ± 6.9	−227.7 ± 5.4	−225.3 ± 5.7	−258.0 ± 6.2	−253.9 ± 6.8	−255.5 ± 6.0
$\langle E_{\text{coulomb}} \rangle$	−1750.1 ± 7.5	−1735.3 ± 6.3	−1755.0 ± 7.5	−1752.6 ± 7.0	−1747.6 ± 6.0	−1747.0 ± 6.5	−1744.9 ± 5.8
$\langle E_{\text{dihe}} \rangle$	50.8 ± 6.0	57.0 ± 5.6	55.8 ± 5.8	52.1 ± 5.3	50.4 ± 5.4	52.1 ± 5.8	48.8 ± 5.4
$\langle E_{\text{constr}} \rangle$	0.3 ± 0.6	0.3 ± 0.4	0.2 ± 0.4	1.2 ± 0.9	1.3 ± 1.4	0.2 ± 0.3	0.5 ± 0.6

TABLE 5 Various average energies (in kcal/mol) of the partial structures of the helix pair A and G classified by the principal component analysis, which were obtained at the temperature of 500 K by the REM simulations

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
$\langle E_{\text{total}} \rangle$	−2204.6 ± 10.1	−2103.8 ± 9.8	−2202.4 ± 9.5	−2154.3 ± 12.7	−2225.0 ± 11.6	−2220.9 ± 11.1	−2223.4 ± 9.7
$\langle E_{\text{vdw}} \rangle$	−238.3 ± 7.5	−206.3 ± 6.1	−255.7 ± 7.1	−249.3 ± 7.6	−259.0 ± 7.3	−259.4 ± 6.7	−256.8 ± 6.9
$\langle E_{\text{coulomb}} \rangle$	−2087.3 ± 7.2	−2022.4 ± 7.6	−2071.3 ± 8.0	−2029.3 ± 9.3	−2084.9 ± 7.7	−2083.5 ± 7.7	−2087.4 ± 6.8
$\langle E_{\text{dihe}} \rangle$	51.7 ± 5.8	55.5 ± 5.2	54.8 ± 5.8	54.7 ± 5.7	49.6 ± 5.6	52.9 ± 5.9	51.3 ± 5.6
$\langle E_{\text{constr}} \rangle$	0.2 ± 0.5	0.3 ± 0.4	0.7 ± 0.9	0.6 ± 0.6	0.3 ± 0.6	0.1 ± 0.2	0.5 ± 0.7

of the nonnative clusters. This implies that the native structure is stabilized by the vdW energy, although one interhelix hydrogen bond exists between Helices B and G.

In this case, the native clusters do not have the lowest total potential energy. Cluster 2, which is nonnative, has the lowest total potential energy. Therefore, it suggests that the native structure of Helices B and G cannot be recovered only by themselves and other helices are necessary to fold into the native structure. This means that although nonnative structures are more stable only for isolated Helices B and G, they are unstable in the entire structure of bacteriorhodopsin.

CONCLUSIONS

We examined the helix-helix association of bacteriorhodopsin transmembrane helices by the techniques of replica-exchange simulation and principal component analysis. This is a first attempt to clarify the global characteristics of helix association by sampling wide helix conformations including helix shuffling of bacteriorhodopsin by molecular simulation.

Most residues of transmembrane helices of bacteriorhodopsin are hydrophobic ones, but they also contain many polar and charged residues. By plotting the average energy components as functions of temperature, we clarified that the van der Waals energy, torsion energy, and electrostatic energy are all important for assembling helices, and none are negligible.

We classified low-energy structures into clusters of similar structures by the principal component analysis to find which helix pairs are stable at low temperatures. We found seven clusters of similar structures, one of which had a rather similar overall structure to the native one. We also showed that the partial structures of Helices A, B, and G were particularly well predicted in our simulations. These findings essentially confirm and support the results of our previous preliminary simulation run (13) with confidence.

Therefore, we focused our attention on how these helices assemble to the native structure. Our supplemental simulations suggested that Helices A and B can self-assemble by themselves mainly by the van der Waals packing interactions, Helices A and G can self-assemble by themselves mainly by the electrostatic interactions, and Helices B and G cannot self-assemble by themselves and need other helices for making the proper, native configuration.

It seems that, to understand the whole transmembrane structure including Helices C, D, E, and F, we need to take into consideration the retinal molecule explicitly, which exists among these helices.

The computations were performed on the computers at the Research Center for Computational Science, Institute for Molecular Science.

This work was supported, in part, by the Grants-in-Aid for the Next Generation Super Computing Project, Nanoscience Program and for Scientific Research in Priority Areas, "Molecular Theory for Real Systems", from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

REFERENCES

1. Krogh, A., B. Larsson, G. v. Heijne, and E. L. L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305:567–580.
2. Popot, J. L., and D. M. Engelman. 2000. Helical membrane protein folding, stability, and evolution. *Annu. Rev. Biochem.* 69:881–922.
3. Popot, J. L., S. E. Getchman, and D. M. Engelman. 1987. Refolding of bacteriorhodopsin in lipid bilayers: a thermodynamically controlled two-stage process. *J. Mol. Biol.* 198:665–676.
4. Bibi, E., and H. R. Kaback. 1990. In vivo expression of the LacY gene in two segments leads to functional Lac permease. *Proc. Natl. Acad. Sci. USA.* 87:4325–4329.
5. Ridge, K. D., S. S. J. Lee, and L. L. Yao. 1995. In vivo assembly of rhodopsin from expressed polypeptide fragments. *Proc. Natl. Acad. Sci. USA.* 92:3204–3208.
6. Groves, J. D., and M. J. A. Tanner. 1995. Co-expressed complementary fragments of the human red cell anion exchanger (Band 3, AE1)

TABLE 6 Various average energies (in kcal/mol) of the partial structures of the helix pair B and G classified by the principal component analysis, which were obtained at the temperature of 500 K by the REM simulations

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
$\langle E_{\text{total}} \rangle$	−1982.9 ± 9.5	−2058.0 ± 9.3	−1992.2 ± 8.8	−1979.3 ± 10.6	−1995.7 ± 11.1	−1991.1 ± 9.7	−1992.3 ± 9.8
$\langle E_{\text{vdw}} \rangle$	−213.3 ± 6.4	−212.8 ± 6.9	−210.7 ± 6.0	−209.4 ± 5.5	−228.0 ± 5.6	−224.9 ± 6.4	−223.5 ± 6.5
$\langle E_{\text{coulomb}} \rangle$	−1876.0 ± 7.5	−1953.9 ± 7.2	−1890.7 ± 5.8	−1882.9 ± 8.4	−1876.0 ± 8.5	−1873.3 ± 7.5	−1876.5 ± 7.0
$\langle E_{\text{dihe}} \rangle$	54.8 ± 5.6	57.1 ± 5.3	57.0 ± 5.9	61.1 ± 5.7	55.8 ± 5.5	55.5 ± 5.1	56.1 ± 5.4
$\langle E_{\text{constr}} \rangle$	0.2 ± 0.4	0.2 ± 0.4	0.8 ± 0.9	0.6 ± 0.8	1.1 ± 1.0	0.2 ± 0.3	0.3 ± 0.4

generate stilbene disulfonate-sensitive anion transport. *J. Biol. Chem.* 270:9097–9105.

7. Suwa, M., T. Hirokawa, and S. Mitaku. 1995. A continuum theory for the prediction of lateral and rotational positioning of α-helices in membrane proteins: bacteriorhodopsin. *Proteins.* 22:363–377.
8. Hirokawa, T., J. Uechi, H. Sasamoto, M. Suwa, and S. Mitaku. 2000. A triangle lattice model that predicts transmembrane helix configuration using a polar jigsaw puzzle. *Protein Eng.* 13:771–778.
9. Vaidehi, N., W. B. Floriano, R. Trabanino, S. E. Hall, P. Freddolino, et al. 2000. Prediction of structure and function of G protein-coupled receptors. *Proc. Natl. Acad. Sci. USA.* 99:12622–12627.
10. Kokubo, H., and Y. Okamoto. 2004. Prediction of transmembrane helix configurations by replica-exchange simulations. *Chem. Phys. Lett.* 383:397–402.
11. Kokubo, H., and Y. Okamoto. 2004. Prediction of membrane protein structures by replica-exchange Monte Carlo simulations: case of two helices. *J. Chem. Phys.* 120:10837–10847.
12. Kokubo, H., and Y. Okamoto. 2004. Classification and prediction of low-energy membrane protein helix configurations by replica-exchange Monte Carlo method. *J. Phys. Soc. Jpn.* 73:2571–2585.
13. Kokubo, H., and Y. Okamoto. 2004. Self-assembly of transmembrane helices of bacteriorhodopsin by a replica-exchange Monte Carlo simulation. *Chem. Phys. Lett.* 392:168–175.
14. Geyer, C.J. 1991. Markov chain Monte Carlo maximum likelihood. *In* Computing Science and Statistics. *Proc. 23rd Symp. Interface.* 156–163.
15. Hukushima, K., and K. Nemoto. 1996. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.* 65:1604–1608.
16. Mitsutake, A., Y. Sugita, and Y. Okamoto. 2001. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers (Peptide Sci.).* 60:96–123.
17. Hansmann, U. H. E. 1997. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* 281:140–150.
18. Sugita, Y., and Y. Okamoto. 1999. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314:141–151.
19. Irback, A., and E. Sandelin. 1999. Monte Carlo study of the phase structure of compact polymer chains. *J. Chem. Phys.* 110:12256–12262.
20. Wu, M. G., and M. W. Deem. 1999. Analytical rebridging Monte Carlo: application to *cis/trans* isomerization in proline-containing, cyclic peptides. *J. Chem. Phys.* 111:6625–6632.
21. Yan, Q., and J. J. de Pablo. 1999. Hyper-parallel tempering Monte Carlo: application to the Lennard-Jones fluid and the restricted primitive model. *J. Chem. Phys.* 111:9509–9516.
22. Reiher III, W.E. 1985. Theoretical studies of hydrogen bonding. PhD thesis. Department of Chemistry, Harvard University, Cambridge, MA.
23. Neria, E., S. Fischer, and M. Karplus. 1996. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* 105:1902–1921.
24. Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, et al. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217.
25. Teeter, M. M., and D. A. Case. 1990. Harmonic and quasiharmonic descriptions of crambin. *J. Phys. Chem.* 94:8091–8097.
26. Kitao, A., F. Hirata, and N. Gō. 1991. 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.
27. Garcia, A. E. 1992. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.* 68:2696–2699.
28. Abagyan, R., and P. Argos. 1992. Optimal protocol and trajectory visualization for conformational searches of peptides and proteins. *J. Mol. Biol.* 225:519–532.
29. Amadei, A., A. B. M. Linssen, and H. J. C. Berendsen. 1993. Essential dynamics of proteins. *Proteins.* 17:412–425.
30. Berg, B. A., and T. Neuhaus. 1992. Multicanonical ensemble—a new approach to simulate 1st order phase transitions. *Phys. Rev. Lett.* 68:9–12.
31. Mitsutake, A., Y. Sugita, and Y. Okamoto. 2003. 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.
32. Sayle, R. A., and E. J. Milner-White. 1995. RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.* 20:374–376.
33. Onuchic, J. N., Z. Luthey-Schulten, and P. G. Wolynes. 1997. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* 48:545–600.