Journal of Structural Biology 165 (2009) 37–46

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

# A dual-scale approach toward structure prediction of retinal proteins

C.-C. Chen, C.-M. Chen *

*Department of Physics, National Taiwan Normal University, 88 sec. 4 Ting-Chou Rd., Taipei 116, Taiwan*

ARTICLE INFO

*Article history:*
Received 19 February 2008
Received in revised form 8 October 2008
Accepted 20 October 2008
Available online 26 October 2008

*Keywords:*
Retinal proteins
Coarse-grained protein model
Parallel tempering
Monte-Carlo simulations
All-atom molecular dynamics simulations
Structure prediction

ABSTRACT

We propose a dual-scale approach to predict the native structures of retinal proteins (RPs) by combining coarse-grained (CG) Monte-Carlo simulations and all-atom (AA) molecular dynamics simulations to pack their transmembrane helices correctly. This approach has been applied to obtain the structures of five RPs, including bacteriorhodopsin (BR), halorhodopsin (HR), sensory rhodopsin I (SRI), sensory rhodopsin II (SRII), and (bovine) rhodopsin. The proposed CG model predicts a reasonably good structure of RPs in days using a desktop computer, which also gives clear physical picture for the packing, tilting, and orientation of transmembrane helices. A high-resolution protein structure is obtained from the AA molecular dynamics simulations by refining the predicted CG structure. The root mean square deviation in coordinates of backbone atoms from the X-ray structure is 1.89 Å for HR, 1.92 Å for SRII, 2.64 Å for BR, and 5.54 Å for rhodopsin. Reasonable predictions of HR structure can be obtained by this approach in the case of using predicted secondary structures with certain alignment error. Since the crystal structure of SRI is not available in the protein data bank, the predicted structure of SRI from our dual-scale approach is compared to that obtained from homology modeling.



## 1. Introduction

Membrane proteins (MPs) play key roles in living cells, such as ion channels, drug receptors, and information transfers (White and Wimley, 1999). Functionally normal MPs are vital to survival and their defects lead to many known diseases. The clinical importance of MPs is demonstrated by the fact that more than 50% of known drugs are targeting on MPs (Moreau and Huber, 1999), which are also responsible for the uptake, metabolism, and clearance of these pharmacologically active substances. Although analyses show that more than a quarter of proteins coded in genomes are MPs (Gerstein, 1998; Wallin and von Heijne, 1998; Krogh et al., 2001), due to difficulties in crystallizing MPs, less than 200 unique structures have been derived (data obtained from http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html). Therefore, there exist great incentives for computational and theoretical studies of MPs (Milik and Skolnick, 1992; Chen, 2000; Floriano et al., 2000; Chen and Chen, 2003; Dobbs et al., 2002; Kokubo and Okamoto, 2004; Ou et al., 2007).

Retinal proteins (RPs) include MPs found in the purple membrane of Halobacterium salinarium (Zheng and Herzfeld, 1992), each with different functions: bacteriorhodopsin (BR) is a proton pump (Pebay-Peyroula et al., 1997), halorhodopsin (HR) is a chloride pump (Kolbe et al., 2000), while sensory rhodopsin I (SRI) and sensory rhodopsin II (SRII) (Royant et al., 2001) are photosensoric proteins. The two ion pumps, BR and HR, convert light energy for the bacteria to synthesize ATPs. The two photosensors, SRI and SRII, direct the bacteria toward optimal light conditions and to avoid exposure to photooxidative conditions. RPs are the focus of much interest and have become a paradigm for MPs in general and transporters in particular (Oesterhelt and Stoeckenius, 1973; Oesterhelt, 1976; Henderson, 1977). Their structure and function have been analyzed in great detail using a variety of experimental techniques. Structurally, they have a topology of seven transmembrane (TM) helices arranged in two arcs, an inner one containing helices B, C, and D and an outer one comprising helices E, F, G, and A. Between helices B, C, F, and G, there is a TM pore, which accommodates a retinal to separate the extracellular half channel from the cytoplasmic half channel. Since the general structure of RPs has a relatively simple topology and has been studied extensively using various experimental techniques, they serves as excellent model systems for constructing a physical model to predict the structure and thermodynamics of MP folding, particularly for seven transmembrane (7TM) receptors.

The construction of a general model for MP structure prediction remains to be one of the great challenges (Bowie, 2005). A traditional approach for predicting MP structures is the homology modeling, which relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence. Many investigators have used the structure of BR or bovine rhodopsin as a template to build models of 7TM receptors (Pardo et al., 1992; Davies et al., 1996; Baldwin, 1998; Herzyk and Hubbard, 1998). However, due to the low sequence identity (less than 30%) between most 7TM receptors and rhodopsin (or BR), it should be noted

* Corresponding author.
*E-mail address:* cchen@phy.ntnu.edu.tw (C.-M. Chen).

1047-8477/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2008.10.001

that their arrangement of TM helices could be very different, which leads to an incorrect prediction of their structures. In fact, the average sequence identity of 99% of human 7TM receptors to bovine rhodopsin is lower than 20% (Archer et al., 2003). Another template based method for structure prediction is the threading method (Zhang et al., 2006; Yarov-Yarovoy et al., 2006), whose success depends on the completeness of the library of solved structures in the protein library. An alternative method for structure prediction of MPs without using homology has been developed by Goddard III and coworkers (Trabanino et al., 2004; Kalani et al., 2004). The disadvantage of this method is that an experimental electron density map of MPs is required as an initial input. The *ab initio* approach is based on the global minimization of a physical potential energy function, which thus far has had limited success for small proteins (Simons et al., 1997; Liwo et al., 1999; Zhang et al., 2003). We note that most previous studies of MP structures (such as homology modeling and threading) show little interest on the thermodynamic hypothesis of protein folding, which is the main focus of this article. In addition, most previous methods in studying the structure of large proteins often require extensive computation. Our study intends to obtain a reasonably good structure within limited computational time by a dual-scale approach.

It has been suggested that folding of many integral MPs can be understood based on the two-stage model: Independently stable helices are formed in lipid bilayers in the first stage, and the helices interact with others to form a functional MP in the second stage (Popot and Engelman, 1990, 2000; Booth and Curran, 1999; Pappu et al.,1999). In this article, based on the two-stage model, we demonstrate the feasibility of predicting the native structure of RPs by a dual-scale approach of computer simulations. Five RPs, including BR, HR, SRI, SRII, and (bovine) rhodopsin are tested without using their 3D structure information in the protein data bank (PDB) (Berman et al., 2000). Their secondary structures are considered to be known here for simplicity, and the derivation of such information is demonstrated in the Supplementary information (supplementary information, 2008).[1] Using possible secondary structures of HR from our secondary structure prediction algorithm, predicted 3D structures of HR are compared to its crystal structure to further validate this dual-scale approach toward a more general structure prediction algorithm. Since most MPs contain large number of amino acids (the average length of MPs of known structures is about 400 amino acids), it is extremely difficult to study their folded structures and folding dynamics by computer simulations at atomic resolution. Therefore, we first construct a coarse-grained (CG) protein model for helical bundle MPs (HBMPs), which includes most dominant physical interactions of the system. A detailed description of our CG model is given in Section 2. The lowest-energy state structure of HBMPs can be identified using the parallel tempering (PT) algorithm (Hansmann, 1997), as described in Section 3. PT is an efficient algorithm in finding the ground state, but the dynamic information of MP folding is missing. On the other hand, Monte-Carlo (MC) simulations are not so efficient to find the ground state, but can be used to obtain thermodynamic information of MP folding. In Section 4, we delineate our dual-scale simulation methods. At a low resolution scale, CG MC simulations are performed to find the ground state structure identified by PT simulations and to obtain folding dynamics from a random initial structure. At a high resolution scale, an all atom (AA) representation of our lowest energy CG structure of RPs is refined to give the predicted 3D structures of RPs using AA molecular dynamics (MD) simulations. In Section 5, we discuss results from our computer simulations of four RPs. The predicted packing, tilting, and orientation of helices are found to be consistent with experimental data by comparing the native structures of RPs with our predicted structures. The root mean square deviation (RMSD) in the tilting angles is 4.8 degrees for HR and is 3.2 degrees for SRII, while RMSD in the orientation angles is 6.8 degrees for HR and 23 degrees for SRII. Section 6 gives our conclusion.

[1] Supplementary information can be found at the Web page http://www.phy.ntnu.edu.tw/~cchen/paper/mp.htm.

## 2. Model

Previous studies using lattice MC simulations have shown the feasibility in predicting the number and location of TM α-helices of HBMPs (as also demonstrated in the Supplementary information). Insertion of TM helices into the membrane in vivo occurs either spontaneously or, more probably, via a translocon. In the latter case, our computer simulations suggest that the formation of TM helices is much faster than the packing of TM segments since only local interactions are involved in helix formation (Chen and Chen, 2003). As proposed by the two-stage model, we assume that the initial structure of RPs contains seven rigid cylinders, which randomly reside in the membrane and are constrained by flexible inter-helix loops. Each cylinder consists of 4 or 6 segments (5 or 7 monomers) and length of each segment varies from 5.66 Å to 8.16 Å depending on the number of residues in each helix. As shown in Supplementary Fig. S2 (supplementary information, 2008), we illustrate the CG representation of four RPs from their secondary structures, in which each monomer roughly corresponds to a helical turn of TM helices. These TM helices can be identified on the basis of hydrophobicity as described in the Supplementary information and, as an example, the secondary structure of BR adopted here is (WIWL) (ALGT) (ALMGL) (GTLY) (FLVK) (helix A: 10–30), (PDAKK) (FYAIT) (TLVPA) (IAFTM) (YLSML) (helix B: 37–61), (RYAD) (WLFT) (TPL) (LLLD) (LALL) (helix C: 82–100), (QGTIL) (ALVG) (ADGI) (MIGT) (GLVGA) (helix D: 105–126), (RFVW) (WAIS) (TAA) (MLYI) (LYVL) (helix E: 134–152), (EVAST) (FKVLR) (NVTV) (VLWSA) (YPVVW) (helix F: 166–189), and (NIET) (LLFM) (VLD) (VSAK) (VGFG) (helix G: 202–220). Here amino acids within the same group are considered to belong to the same monomer in our CG model. The loop constraint is modeled by limiting the head-to-tail distance between two consecutive helices, which is proportional to the number of residues in the loops (each residue is about 3.8 Å). If the head-to-tail distance ($r_{ht}$) between two consecutive helices is smaller than the length of the corresponding loop, they can move without feeling the loop constraint. However, it is not permitted for $r_{ht}$ to be larger than the maximal length. In other words, the potential for the loop constraint is a step function of $r_{ht}$, which is zero if $r_{ht}$ is less than its maximal distance and is infinity if it is larger than the maximal distance. These helices are allowed to diffuse in the membrane, as well as to tilt ($\Theta$) and rotate ($\Phi$) along the membrane normal direction, as shown in Fig. 1. The value of $\Omega$ is irrelevant in the CG model due to our simplification of TM helices, and will be determined by the hydrophobicity of helix surfaces in the AA model. The membrane core is modeled as a hydrophobic layer of thickness ($L$) 26 Å sandwiched by two water regions (White and Wimley, 1999).

Among various physical interactions, evidences show that van der Waals (vdW) interaction and side-chain packing among TM helices mostly determine the tertiary structure of MPs (White and Wimley, 1999; Popot and Engelman, 2000). Although inter-helical hydrogen bonding, ion pairs, and disulfide bonds have been considered as alternative sources of stability, there are only few cases demonstrating the importance of these alternative interactions. In our model, the vdW interaction between helices is expressed as

$$E_{\mathrm{vdw}} = e_1 \sum_{\substack{i,j \\ i \neq j}} \sum_{m,n} \left\{ \left[ \frac{r_0}{r(m_i, n_j)} \right]^{12} - \left[ \frac{r_0}{r(m_i, n_j)} \right]^{6} \right\}, \quad (1)$$

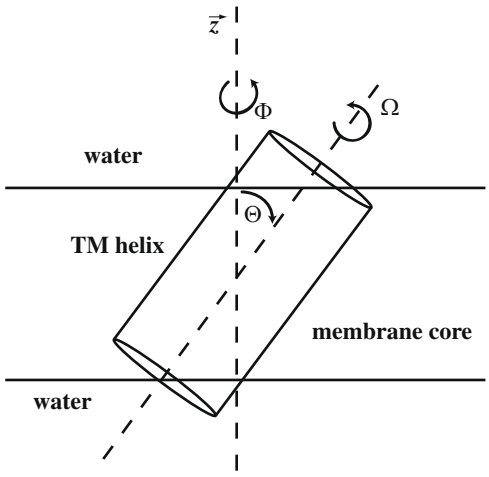


**Fig. 1.** A schematic representation of a TM helix showing its tilting ($\Theta$), orientation ($\Phi$), and rotational ($\Omega$) angles.

where $e_1$ is the strength of the vdW interaction and $r_0$ determines the minimum of $E_{\mathrm{vdw}}$. In this representation, each helix is considered to be a rigid polymer cylinder and a monomer represents 3–5 residues. The distance between $m$-th monomer in helix $i$ and $n$-th monomer in helix $j$ is denoted by $r(m_i,n_j)$ $(i \neq j)$. The packing interaction in Eq. (1) is a sum of all vdW energy between monomers, which is approximated without sequence dependence in our CG model. A mean parameter $r_0$ is employed instead of considering sequence heterogeneity explicitly, which is an average value evaluated using AA model of the helices. To estimate its value, based on the secondary structure of RPs, we first construct those seven helices of RPs individually as standard helices with the $\varphi$ and $\psi$ torsional angles of residues equal to −60 and −40 degree. Each standard helix is subject to an energy minimization using AMBER7 (Case et al., 2002). The lowest vdW energy of each pair of neighboring helices is calculated as a function of their separation distance by varying their tilting ($\Theta$), orientation ($\Phi$), and rotational ($\Omega$) angles at each distance, as shown in Fig. 2. This packing parameter $r_0$ is calculated to be 8.3 Å for HR (experimental measurement is 8.24 Å), 8.4 Å for SRII (experimental measurement is 8.5 Å), and 7.8 Å for BR (experimental measurement is 7.9 Å) by averaging its values predicted from these curves (Berman et al., 2000). We note that the value of $r_0$ is in general sequence dependent since different proteins would have different packing size. We further note that the determination of $r_0$ relies on AA calculations for pairs of neighboring helices, which requires knowledge on the topology of helix packing. Since a slight change in the value of $r_0$ only changes the packing size but not the packing topology, a rough estimation of $r_0$ can be used to obtain the packing topology first. This packing topology will then allow a more accurate estimation of $r_0$.

The helix–water interaction $E_{\mathrm{hw}}$ can be modeled by using a rescaled Kyte–Doolittle hydropathy (HP) index (Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val) = (0.4, −1, −0.78, −0.78, 0.56, −0.78, −0.78, −0.09, −0.71, 1, 0.84, −0.87, 0.42, 0.62, −0.36, −0.18, −0.16, −0.2, −0.29, 0.93) with strength $e_2$, which is mainly determined by the Gibbs free energy change for transferring amino acids from the water phase into their compressed gas phase (Kyte and Doolittle, 1982). Here the HP index of residues is rescaled to be values between −1 and 1. In our CG model, the hydropathy index of each monomer corresponds to its cumulative hydropathy (CHP) index of associated residues, i.e., $\mathrm{CHP}_j = \begin{cases} \sum_i \mathrm{HP}_i, \text{if monomer } j \text{ is in water} \\ 0, \text{if monomer } j \text{ is in membrane} \end{cases}$, where $i$ is the residue index in monomer $j$ of a TM helix. Thus one can express

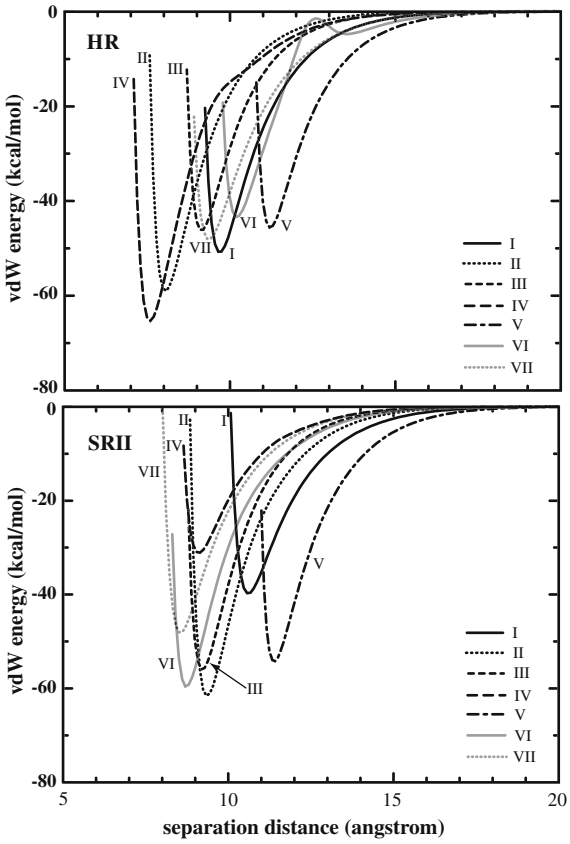


**Fig. 2.** Energy curves of vdW interaction for each pair of neighboring helices in HR and SRII as a function of their separation distance. The vdW energy curves are of the lowest energy at each separation distance by varying tilting, orientation, and rotational angles. Curve I is for the pair of helices A and B, curve II is for the pair of helices B and C, curve III is for the pair of helices C and D, curve IV is for the pair of helices D and E, curve V is for the pair of helices E and F, curve VI is for the pair of helices F and G, and curve VII is for the pair of helices G and A.

$E_{hw} = e_2 \sum_j \mathrm{CHP}_j$, where $j$ runs over all monomers in TM helices. As a first order approximation, each monomer is either in the lipid phase or in the water phase in the present calculation. This simplification will be relieved to include the size effect of each monomer in the future.

In addition, detailed studies of model hydrophobic helices in phospholipid bilayers have shown that lipids in the immediate neighborhood of a helix are perturbed due to the helix–lipid interaction (Huschilt et al., 1985; Subczynski et al., 1998). We thus model the helix–lipid interaction by a tilting energy $E_{hl} = e_3 \sum_i (1 - \cos\Theta_i)$ of the helices in the membrane, where $\Theta_i$ is the tilting angle of the $i$-th helix. The tilting energy increases if a helix is tilted from the membrane normal, due to the increase in the contact between lipids and helix. In principle, the competition of the helix–water and helix–lipid interactions would determine the tilting angles of helices. A proper estimation of the value of $e_3/e_2$ would allow us to predict the tilting angles of helices correctly. By performing MC simulations for various values of $e_3/e_2$, we calculate the tilting angles of helices (the lowest energy $\Theta_i$ of $i$-th helix) as a function of $e_3/e_2$ for 12 HBMPs in the PDB, including four 7TM receptors, three transporters, two ion channels, two aquaporins, and one intramembrane protease. The predicted tilting angles of helices are then compared with their values acquired from the PDB. In this article, we define the RMSD of the helix tilting angles to be $\Theta_{\mathrm{rmsd}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\Theta_i - \Theta_i^0)^2}$, where $\Theta_i$ and $\Theta_i^0$ are the predicted and acquired tilting angles of $i$-th helix. As shown in Fig. 3, when the membrane thickness is taken to be 26 Å, the best value of $e_3/e_2$ is estimated to be 0.7 by minimizing $\Theta_{\mathrm{rmsd}}$.

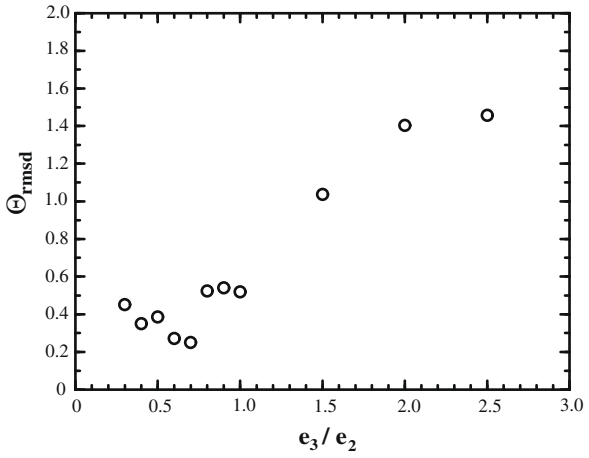


**Fig. 3.** RMSD (from the PDB structure) of the helix tilting angles ($\Theta_{rmsd} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\Theta_i - \Theta_i^0)^2}$) as a function of $e_3/e_2$. The thickness of membrane core is set to be 26 Å.

Finally we discuss the effect of the retinal in stabilizing the channel state over a hexagonal packing state. The hexagonal packing state, in which one helix of RPs is surrounded by the other six helices, is expected to have the lowest packing energy since there are more contacts between helices (Kokubo and Okamoto, 2004). However, for such a hexagonal packing state, the retinal will be outside the helix bundle and has unfavorable contacts with lipids. Conversely, the retinal can form hydrogen bonding with water molecules and have favored contacts with helices, if it is in the channel state (Nina et al., 1995; Baudry et al., 1999). In our CG model, the retinal is represented by a rod of length 12 Å and radius 1.6 Å, which is covalently bound to Lys-216 of the G-helix and allowed to move in the membrane (Tajkhorshid et al., 1999). We thus model the non-covalent interaction between the retinal and its environment by a contact energy between retinal and helices, $E_{contact} = e_4\sum_{i=1}^{7}\varepsilon(\Delta r_i)$, where $\Delta r_i$ is the shortest distance between the axes of retinal and $i$-th helix, and $\varepsilon(\Delta r_i)$ is 1 if $\Delta r_i$ is between 6 Å and 9 Å and 0 otherwise. We note that $E_{contact}$ is residue independent since it is used to model mainly the energy difference between retinal-water interaction and retinal-lipid interaction.

According to the thermodynamic hypothesis of protein folding, the native state of proteins is the global minimum of free energy (Anfinsen, 1973). To find the ground state structure of RPs, the relevant physical quantity to be minimized in our model is the total energy $E_{total} = E_{packing} + E_{hw} + E_{hl} + E_{contact}$ (Chen et al.,). We note that our model is a minimum model for MP folding, which has only been tested for five RPs. A more comprehensive model and a more realistic representation of proteins are still under our investigation, in which case all residues of TM helices are explicitly considered. In this CG model, typical values of the parameters are $e_1 = 0.25$, $e_2 = 1$, $e_3 = 0.7$, $e_4 = -0.5$, and $kT = 0.1$. The choice of these parameters is not unique. For example, the RMSD (in coordinates of helix backbone atoms) of the ground state structure of BR is unchanged if we slightly change the value of $e_1$ or $e_4$. However, for $e_1 > 0.4$ ($e_4 = -0.5$) or $e_4 > -0.3$ ($e_1 = 0.25$), the structure of the ground state has a RMSD (in coordinates of helix backbone atoms) of 5.67 Å from the native structure of BR. The estimated value of $e_3/e_2$ generally depends on the thickness of membrane core, $L$. Thermal energy has no effect on the predicted structure as long as enough simulation time is used.

## 3. Search the ground state structure

Parallel tempering (PT), also known as the replica exchange method, has been shown to have good search properties in protein studies. The basic idea of PT is a conditional temperature swapping of two configurations of the protein, each in a regular canonical simulation at different temperatures. This approach effectively enhances the probability of the protein for getting out of local energy minima. Given two configurations, each with energies and temperatures $E_1$, $T_1$ and $E_2$, $T_2$, respectively, the probability of swapping their simulation temperatures is given by $P = \min\{1, \exp[-(E_2 - E_1)(\frac{1}{kT_1} - \frac{1}{kT_2})]\}$. Our simulation data are gathered as follows. For a protein sequence, MC simulations of 10 replicas of a protein with random initial states are executed independently at different temperatures ($T = 0.08$, 0.1, 0.13, 0.17, 0.21, 0.28, 0.36, 0.46, 0.59, and 0.77). In our distributed computing system, MC simulation of each replica is carried out at a client computer (using a single core 1.8 GHz AMD Opteron CPU). After a run time of $10^3$ MC steps, all information of replicas will be collected by a server computer. Here, 1 MC step is defined to be the time period within which each helix attempts to move 50 times. The server computer will then attempt to swap temperature between replicas based on the above swapping probability. We display in Supplementary Fig. S3 (supplementary information, 2008) for a typical replica of BR the time series of temperatures that are visited in the course of a simulation of $10^7$ MC steps. Due to the successive temperature swapping, the replica moves randomly between low and high temperatures. The time series of energies of the replica during the simulation is shown in Fig. 4, in which inset (a) is the initial conformation and inset (b) is the ground state conformation on the lipid mid-plane (LMP). The energy of the ground state obtained from PT is −8.7 and its RMSD in coordinates of helix backbone atoms is 4.08 Å from its PDB structure.

PT is an efficient algorithm in finding the ground state, but the dynamic information of MP folding is missing. On the other hand, regular MC simulations are not so efficient to find the ground state, but can be used to obtain thermodynamic information of MP folding. In the case of a rugged folding landscape, the searching time for the ground state structure by regular MC simulations is often too long such that the lowest energy state found by regular MC simulations is not guaranteed to be the ground state. In our case, the ground state of our model energy $E_{total}$ is first identified by the PT simulations. An independent search of the configuration space of MP folding for the ground state is then carried out by regular MC simulations, which would help us to judge the roughness of our model energy $E_{total}$ and to obtain thermodynamic information of MP folding. Indeed, we find that the landscape of our model

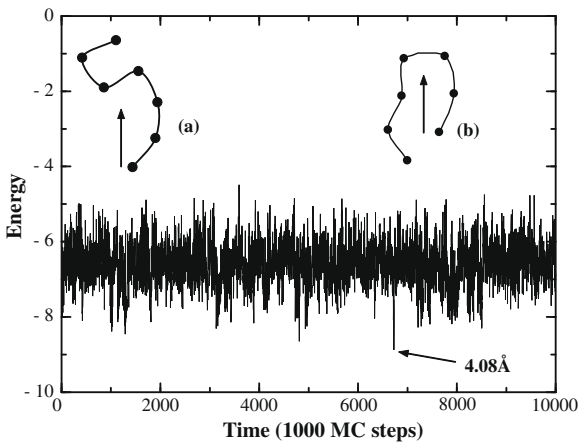


**Fig. 4.** Time series of energy for one of the 10 replicas over the $10^7$ MC steps in the parallel tempering simulation. The lowest energy state is observed at around $7 \times 10^6$ MC steps. The insets are the initial conformation (a) and the ground state conformation (b) of the replica. Here, each filled circle represents the position of one helix on LMP and the arrow represents the projection of the retinal on LMP.

energy $E_{total}$ is rather smooth such that the ground state structures of the four RPs can be found within 100 million MC steps by regular MC simulations, as demonstrated in the next section. The RMSD in coordinates of helix backbone atoms between the AA representation of BR for these two ground state structures (from MC and from PT) is only 1.8 Å. A detailed comparison of packing on LMP, tilting, and orientation of helices between these two ground state structures is shown in Supplementary Fig. S4 (supplementary information, 2008).

## 4. Simulation methods

The dynamic CG simulation of HBMP folding is performed in a simulation box, which is divided into three regions: a membrane core of thickness around 26 Å sandwiched by two water regions. The protein chain consists of seven rigid cylinders (each cylinder represents a TM helix) located in the membrane core and the loop constraint is imposed on these cylinders. One end of the retinal rod is permanently linked to the G-helix and the other end is allowed to move in the membrane. The presence of this retinal molecule in the structure formation of RPs will block helices from entering the pore region of the helix-bundle. The folding of RPs is simulated by the Metropolis MC algorithm in a continuum space at a constant temperature $T$ (Chen and Higgs, 1998). In our simulations, seven rigid helices and the retinal molecule are all allowed to move in the simulation box by changing their $\Theta$ and $\Phi$ angles, as well as the position of their center of mass. Each helix is treated as a straight and rigid cylinder as schematically illustrated in Fig. 1. At each instant, a cylinder is picked up at random and attempts to diffuse, tilt, or rotate along the $z$-axis. The transformation matrix of residue positions by imposing an angular change to a helix from $(\theta,\ \varphi)$ to $(\theta+\Delta\theta,\ \varphi+\Delta\varphi)$ can be easily derived as

$$J=\begin{pmatrix}\cos\phi' & \sin\phi' & 0\\ -\sin\phi' & \cos\phi' & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\cos\Delta\theta & 0 & -\sin\Delta\theta\\ 0 & 1 & 0\\ \sin\Delta\theta & 0 & \cos\Delta\theta\end{pmatrix}\begin{pmatrix}\cos\phi & \sin\phi & 0\\ -\sin\phi & \cos\phi & 0\\ 0 & 0 & 1\end{pmatrix},$$

where $\varphi' = -(\varphi+\Delta\varphi)$. The value of $\Omega$ is irrelevant in our CG model due to the simplification of TM helices. If any attempted move of cylinders satisfies the constraints of excluded volume and inter-helical loops, the move is accepted with probability $w = \min[1, \exp(-\Delta E/kT)]$, where $\Delta E$ is the energy change of the system.

In addition to examining folded structure of RPs using a CG model, an AA calculation is desired to see if one can obtain folded structures of RPs at the atomic level. To construct the AA representation of the lowest energy structure of RPs from our CG model, the seven energy minimized helices (as constructed for calculating $r_0$ in Section 2) are used to replace those rigid cylinders of RPs in our CG model by fitting the center of mass and the axis of helices. The rotation angle along the long axis ($\Omega$) of each helix is chosen to align the most hydrophobic surface of helices to face the membrane core (Trabanino et al., 2004). However, this policy for the rotational angles of helices based on hydrophobicity may not be universal for all MPs. Other interactions might also contribute to the rotational angles of helices. Inter–helix loops are added to connect consecutive helices using MC simulations with a spring potential between corresponding ends of loops and helices, in which case helices of HBMPs are frozen as a template to be added with loops. The retinal is covalently bound to Lys-216 of the G-helix and allowed to move in the simulation. The atomic charges of the retinal and Lys-216 are taken from a previous literature (Tajkhorshid et al., 1999). This structure is then refined by an energy minimization, which is proceeded with 5000 steps of steep descent method and 10000 steps of conjugate gradient method. Here, the hydrophobic core of membrane is treated as a dielectric medium of dielectric constant $\kappa = 2.5$ [its value is between 2 and 4 (Tsong, 1990)]. As a first order approximation, we treat the environment of MPs as a uniform dielectric, which screens out charges by a factor $1/\kappa$. More sophisticated model of the environment of MPs can be adopted to improve the predicted structures of MPs. Other values of $\kappa$ (2.0 and 3.0) are also used, but no substantial differences in the folded structure are observed. Starting from the energy-minimized structure, we carry out restrained MD simulations to further refine the folded structure by allowing both helices and loop segments to move. The restraints include the torsional angles ($\varphi$ and $\psi$) and the distance between N and O atoms of hydrogen bonds in the helices, as well as the residue position on LMP. The time step is 2 fs. The bonds associated with hydrogen atoms were fixed at their equilibrium bond lengths. The cutoff distance for non-bonded interactions is 100 Å to include all atom-atom non-bonded interactions. The temperature coupling parameter for a constant temperature simulation is set to be 5 ps. We note that typical values of temperature coupling parameter are between 0.5 ps and 5 ps for protein simulations, and too small values of this parameter might cause unrealistic fluctuations. For the purpose of structure refinement, any value in this range should make no significant difference.

As a summary, we have outlined the procedure of our dual-scale approach by the following steps:

Step 1: Acquiring secondary structure of RPs

The secondary structure of RPs is used to determine the exact sequence and length of each TM helix. A CG model of these helices is represented by straight and rigid cylinders, which contain 5 or 7 monomers depending on their lengths. Here, each monomer roughly represents a turn in the helix. Taking account of the hydrogen bonding along the backbone of helices, bending or twisting of cylinders is not allowed for simplicity. The length of inter-helix loops is calculated based on the number of amino acids in each loop, which is used to be the head-to-tail constraint between two consecutive helices. A deviation in the secondary structure of RPs affects the quality of our structure prediction, which will be discussed in Section 5.

Step 2: Generating a random initial CG configuration of RPs

A random initial structure of seven rigid cylinders with satisfied loop constraints is placed into the lipid phase in a simulation box [as shown in inset (A) of Figs. 4 and 5]. The simulation box contains a lipid phase of dielectric constant $\kappa = 2.5$, sandwiched by two water phases of dielectric constant $\kappa = 80$. The retinal is represented by a rod of length 12 Å and radius 1.6 Å, and one of its ends is permanently linked to the G-helix.

Step 3: Identifying the ground state of the CG model by PT simulations

The total energy $E_{total}$ of the CG model requires the input of two parameters, $r_0$ and $e_3/e_2$. The value of $r_0$ is estimated from the vdW energy between nearest-neighbor pairs of helices in an AA model. The value of $e_3/e_2$ is estimated from fitting the experimental tilting angles of 12 HBMPs in the PDB. In general, the energy landscape of protein folding is rough and the searching of the ground state is nontrivial. Since PT algorithm largely enhances the probability of the protein for getting out of local energy minima, we use PT simulations to identify the ground state of total energy $E$. The drawback of PT simulations is that it is not possible to get the kinetic information of protein folding from an initial structure to the native structure.

Step 4: Obtaining the ground state structure and folding kinetics by MC simulations

Using the ground state structure identified by PT algorithm as the target structure, MC simulations are used to get kinetic information of MP folding from a random structure. All helices of a MP are represented by cylinders, which are straight and rigid. These cylinders can move by diffusion of their center of mass and rotation of their long axis ($\Theta$ and $\Phi$) under the loop constraints. One end of the retinal molecule is permanently linked to the G-helix and the other end is allowed to move in the membrane.

Step 5: Generating an AA representation of the CG ground state structure

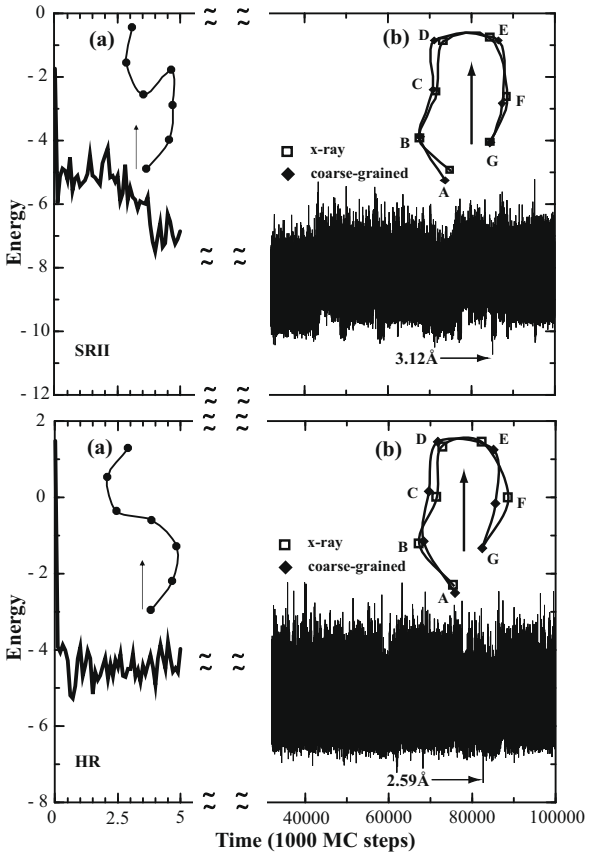


**Fig. 5.** Energy curves of SRII and HR as a function of simulation time calculated from MC simulations. The energy of the chain is plotted for every 100 MC steps initially and for every 10,000 steps otherwise. Both the highest and the lowest energies during the time window of 10,000 steps are plotted. Starting from a random initial configuration as shown in inset (a), the lowest energy states of these energy curves are also noted with their RMSD from the PDB structure (in coordinates of helix backbone atoms), whose structures (filled diamonds) are compared to their crystal ones (open squares), as shown in the inset (b). The retinal molecule is represented by an arrow.

According to the sequence of each helix, we first represent each cylinder as a standard helix with the $\varphi$ and $\psi$ torsional angles of residues equal to $-60$ and $-40$ degree. This standard helix is subject to an energy minimization using AMBER7 to get its lowest energy conformation (which allows the kinks produced by prolines). The seven energy minimized helices obtained are used to replace those rigid cylinders of RPs in our CG model by fitting the center of mass and the axis of helices. The rotation angles along the long axis ($\Omega$) of each helix are chosen to align the most hydrophobic surface of helices to face the membrane core. Amino acids in the loops are added by MC simulations to connect consecutive helices. The retinal molecule is covalently bound to Lys-216 of the G-helix as in the predicted CG structure. RMSD (in coordinates of helix backbone atoms) of the predicted CG structure from the X-ray structure can then be calculated.

Step 6: Obtaining AA refined structure of RPs by AMBER7

The AA structure from Step 5 is first refined by an energy minimization using AMBER. Starting from the energy-minimized structure, we carry out restrained MD simulations to further refine the folded structure by allowing both helices and loop segments to move. The restraints include the torsion angles ($\varphi$ and $\psi$) and the distance between N and O atoms of hydrogen bonds in the helices, as well as the residue position on LMP. The values of these restraint variables can only vary within a small range, i.e. ±1 degree for torsion angles, ±0.1 Å for the distance between N and O atoms, and ±0 Å for residue positions on LMP. The AA structure with the lowest potential energy is chosen to be our prediction, whose RMSD (in coordinates of helix backbone atoms) from the X-ray structure is calculated.

## 5. Results and discussion

According to the thermodynamic hypothesis of protein folding, the native state of proteins is the global minimum of free energy (Anfinsen, 1973). London and coworkers have demonstrated the reversibility of denaturation and renaturation of BR under a wide variety of conditions (Huang et al., 1981; London and Khorana, 1982). It is also found that bound retinal is not necessary for maintenance of native secondary structure, but it does play a key role in tertiary structure formation. From AA calculations (Kokubo and Okamoto, 2004), the lowest energy state of BR in the absence of retinal is found to be a non-channel hexagonal structure without a pore, which is consistent with the result of our CG simulations. With the retinal, however, the interaction between a retinal protein and its environment would favor the native channel state, which has the lowest energy in our CG model. Typical energy curves of the folding of RPs are shown in Fig. 5. Starting from a random initial configuration as shown in insets (a), the energies of SRII and HR drop rapidly during the first few 1000 MC steps to avoid unfavorable contacts between helices. The lowest energy state ($-10.6$) of SRII has a RMSD (in coordinates of helix backbone atoms) of 3.12 Å from its PDB structure, which is observed at about 80 million MC steps. The lowest energy ($-7.3$) state of HR also occurs at about 80 million MC steps, whose RMSD (in coordinates of helix backbone atoms) is about 2.59 Å from its PDB structure. These ground states are consistent with the lowest energy states obtained using the PT algorithm, as described in Section 3. A comparison of our prediction and the crystal structure on LMP is shown in the insets (b) of Fig. 5 for SRII and HR. Without using experimental structural information of these tested MPs in PDB as an input, the resemblance between our prediction and the native structure of RPs implies the validity of our model energy $E_{\text{total}}$. For these predicted CG structures, the overall RMSD (in coordinates of helix backbone atoms) between their AA representations (as constructed in step 5) and their X-ray structures is 3.99 Å for BR, 3.12 Å for SRII, 2.59 Å for HR, and 5.56 Å for rhodopsin. The scattered diagrams of SRII and HR in Fig. 6 shows the relationship between energy of CG structures and RMSD of their energy-minimized AA representation (in coordinates of helix backbone atoms) from the PDB structure for 300 structures observed in our simulations, including 70 lowest energy structures and 230 randomly chosen structures. It is clear that these low energy structures resemble the native structure and their RMSD (in coordinates of helix backbone atoms) from the PDB structure is smaller than 4 Å. The calculated correlation coefficient between energy and RMSD is 0.73 for SRII and 0.71 for HR, implying a high consistency between our model structures and their energy. We have not observed structures of low energy but high RMSD (false positive prediction) or structures of high energy but low RMSD (false negative prediction). Approximately, starting from a random initial configuration of RPs, it takes a few days to find the ground state of our CG model for regular MC simulations using a single core 1.8 GHz AMD Opteron CPU.

By aligning the shortest helix of the PDB structure in the $z$-axis, the tilting and orientation angles of helices are listed in Supplementary Table T1 for SRII and HR. $\Theta_{\text{rmsd}}$ (from its PDB structure) calculated for our CG structures is 4.8 degrees for HR and 5.6 degrees for SRII, while it is 4.8 degrees for HR and 3.2 degrees for SRII from the refined AA structures. As expected, the predicted tilting angles are consistent with their values acquired from X-ray structures, demonstrating that the helix–water and helix–lipid terms are the

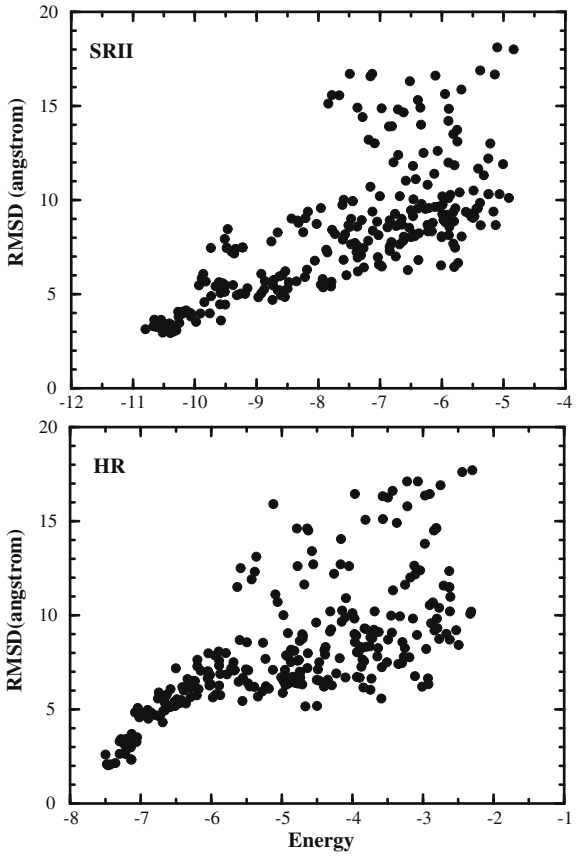


**Fig. 6.** Scatter plots of SRII and HR in the energy–RMSD (from the PDB structure) plane for 80 observed structures, including 10 lowest energy structures and 70 randomly chosen structures. The value of RMSD is calculated only for backbone atoms in transmembrane helices.

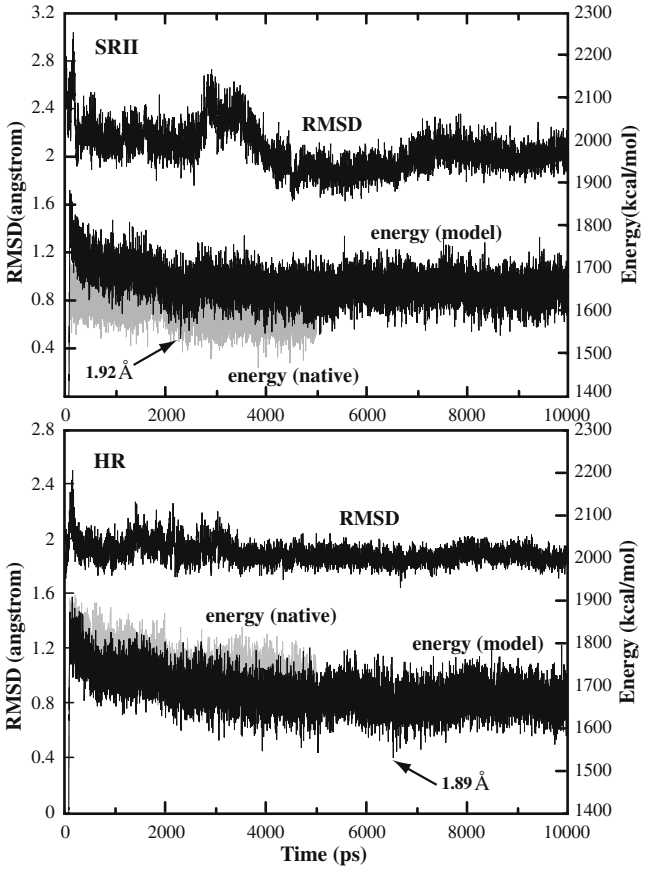


**Fig. 7.** RMSD (from the PDB structure) and potential energy of SRII and HR calculated for MD simulations. Curve 1 is the RMSD of backbone atoms of the seven helices in the MD trajectory, curve 2 is the potential energy curve obtained from the restrained MD simulation starting from MC predicted structure, and curve 3 is the potential energy curve of a restrained MD simulation starting from the X-ray structure.

main interactions for helix tilting. On the other hand, the predicted orientation angles could deviate substantially from their experimental values, due to the lack of explicit sequence dependence in our vdW interaction. The calculated value of $\Phi_{\rm rmsd} = \sqrt{\frac{1}{7}\sum_{i=1}^{7}(\Theta_i - \Theta_i^0)^2}$ is only 18.9 degrees for HR, but is 97.3 degrees for SRII. In principle, there are two ways to improve our predictions. One is to include side chain information of each amino acid of TM helices in the CG model, which is still under our investigation. The other is to refine our CG prediction by AA models, as demonstrated in the following.

The predicted HBMP structures from our CG model have been refined using AMBER7. In Fig. 7, the 10 ns MD simulation gives a RMSD curve (in coordinates of helix backbone atoms and from the PDB structure) ranged 1.6–3.0 Å for SRII and 1.6–2.4 Å for HR. The potential energy (model) curve decreases systematically with time from 1800 kcal/mol to below 1600 kcal/mol for SRII and from 1900 kcal/mol to below 1600 kcal/mol for HR. These energy curves are very close to those simulation curves (native) starting from the X-ray structures. The correlation coefficient between the energy (model) curve and the RMSD curve is 0.17 for SRII and is 0.15 for HR. The lowest energy structure observed in our MD simulation has a RMSD in coordinates of helix backbone atoms of 1.92 Å from the PDB structure for SRII and 1.89 Å from the PDB structure for HR. Fig. 8 shows the comparison of the crystal structures (side view) of SRII (a), HR (b), and BR (c) with our MC prediction and MD refinement. Top view of these structures is available in Supplementary Fig. S5 (supplementary information, 2008). It is apparent that our CG prediction (green ribbons) fits well with the native structure of SRII (blue lines), as shown in Fig. 8(a). The CG prediction of helices A, B, E, F, and G is quite consistent with the X-ray structure, while deviation in the prediction of helices C and D is visible. As shown in Supplementary Table T1, there exist an 8 degree discrepancy in the tilting angle for C helix and a 184 degree discrepancy in the orientation angle for D helix. These discrepancies are considerably reduced in the MD refined structure (red ribbons). In Fig. 8(b and c), we compare the crystal structure (blue ribbons) of HR and BR with their structures of 10 lowest energy states found in MD simulations (red ribbons). Again, good structure prediction is made by our approach. Great improvement in the prediction of orientation angle has been seen by including the side chain information in the MD simulations, as shown in Supplementary Table T1. Our MD simulations reduce the value of $\Phi_{\rm rmsd}$ (from the PDB structure) from 97 degrees to 23 degrees for SRII and from 19 degrees to 6.8 degrees for HR.

In Fig. 9, we show the 2D helix alignments on LMP (a) as well as the 3D ground state structure (b) and a 3D local minimum structure (c) of rhodopsin. The predicted structure of rhodopsin is not as good as our predictions of other retinal proteins (RMSD is between 1.89 Å and 2.64 Å), due to the fact that rhodopsin helices are much longer and have a kink which bends the helix. In our coarse-grained model, as a first order approximation, all helices are assumed to be a rigid cylinder without kinks. This assumption works very well for BR, HR, SRI, and SRII, but not for rhodopsin. In this simple model, as shown in Fig. 9(a), the D helix of rhodopsin in the ground state (open squares) tends to align on a more regular loop around the retinal in order to maximize $E_{\rm vdw}$ and to reduce $E_{\rm hl}$. On the other hand, the D helix in the PDB structure (filled

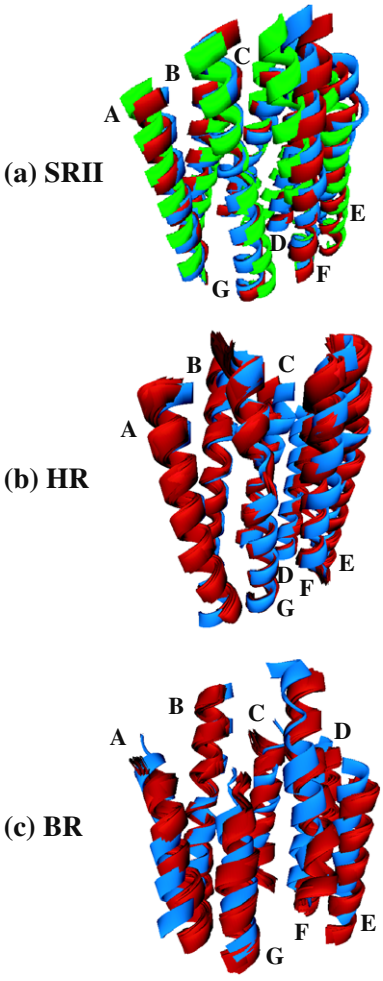


**Fig. 8.** Comparison of the crystal structures (side view) of SRII (a), HR (b), and BR (c) from our MC prediction and MD refinement. In (a), the crystal structure is represented by blue ribbons, our MC prediction is represented by green ribbons, and the MD refinement is represented by red ribbons. In (b) and (c), blue ribbons represent the crystal structure, while red ribbons represent the structure of 10 lowest energy states. (For interpretation of color mentioned in this figure the reader is referred to the web version of the article.)

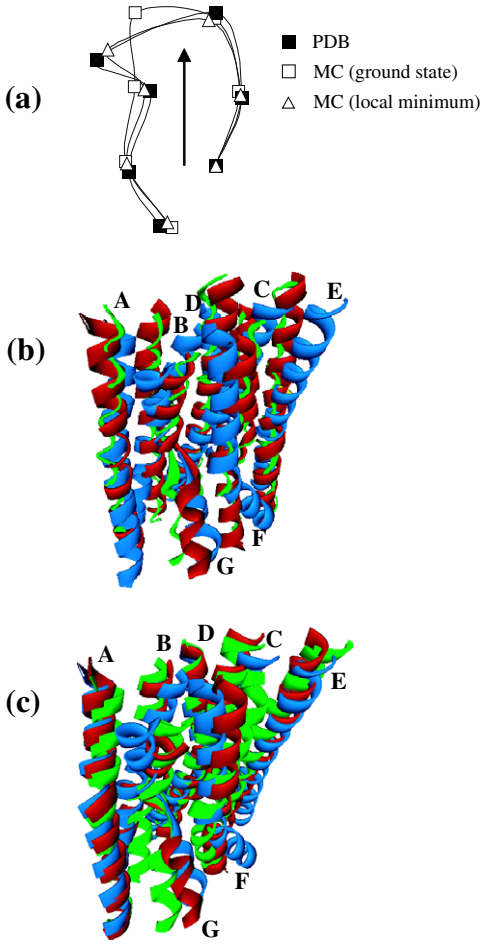


**Fig. 9.** Comparison of the PDB structure of bovine rhodopsin with our MC prediction and MD refinement: (a) helix alignments on LMP, (b) 3D side view of the ground state structure, and (c) 3D side view of a local minimum structure. In (B) the crystal structure is represented by blue ribbons, our MC prediction is represented by green ribbons, and the MD refinement is represented by red ribbons. (For interpretation of color mentioned in this figure the reader is referred to the web version of the article.)

squares) is squeezed out of the loop by helices C and E. The CG model energy of the ground state found in MC simulations is −11.8. However, we do find a slightly higher energy (−10.8) state whose helix alignment (open triangles) closely resembles that of the PDB structure. For the lowest energy alignment, the RMSD (in coordinates of helix backbone atoms) from the PDB structure of rhodopsin is 5.56 Å for our MC predicted structure, and is 5.54 Å for our MD refinement. For the second lowest energy alignment, the RMSD (in coordinates of helix backbone atoms) from the PDB structure of rhodopsin is 3.76 Å for the MC structure, and is 3.01 Å for its MD refinement. To improve our CG model, some modifications would be required in modeling long helices with kinks. Fig. 9(b) is a side view of rhodopsin (overlap of PDB structure, MC ground state structure, and MD refined structure), and the top view of rhodopsin is shown in Supplementary Fig. S6. In our prediction, the position of helices A, B, C, E, F, and G is roughly consistent with those in PDB. However, in the PDB structure of rhodopsin, helices B, D, F, and G have a visible kink and the tilting of helices A, B, C, and E is also different between our predicted structure and the PDB structure. Nonetheless, we note a considerable consistency between the PDB structure and the refined structure from the second lowest energy alignment, as shown in Fig. 9(c).

In previous discussions, we have obtained reasonably good predictions for several RPs using our dual-scale approach based on their known secondary structures. It is critical to assess the validity of this approach by starting from the amino acid sequence of MPs without knowing their exact secondary structures. In the Supplementary information, we have presented a model for secondary structure prediction and predicted two possible secondary structures (the two lowest energy states) of HR. Since our protein model is a CG model, it is usually a good practice to consider a few low energy states (including the ground state) as our starting secondary structure in order to obtain a good structure prediction. The secondary structure of HR in the PDB is denoted by s1 (5–30, helix A; 37–60, helix B; 85–106, helix C; 110–133, helix D; 137–169, helix E; 172–195, helix F; 206–234, helix G). The two predicted secondary structures are denoted as s2 (5–30, helix A; 40–69, helix B; 81–106, helix C; 109–132, helix D; 137–166, helix E; 174–202, helix F; 205–233, helix G) and s3 (5–30, helix A; 38–65, helix B; 81–106, helix C; 109–132, helix D; 137–166, helix E; 170–197, helix F; 205–233, helix G), respectively. The average length of helices is 27 amino acids for structure s1, 27.7 amino acids for structure s2, and 27.3 amino acids for structure s3. The mismatch is 12.6% between s1 and s2, and is 8.3% between s1 and s3. As shown in Supplementary Fig. S7(A) (supplementary information, 2008), reasonably good predictions of 3D

structure of HR have been obtained based on sequences s1, s2, and s3, whose packing patterns of helices are found to be similar to that of HR crystal structure. The predicted tilting ($\Theta$) and orientation ($\Phi$) angles of HR helices using AA model are displayed in Supplementary Fig. S7(B) for all three secondary structures (s1, s2, and s3). For the predicted 3D structure of HR using its secondary structure in the PDB (s1), its value (from the PDB structure) is 4.8 degrees in $\Theta_{\mathrm{rmsd}}$ and 6.8 degrees in $\Phi_{\mathrm{rmsd}}$. The RMSD values of $\Theta$ and $\Phi$ of the predicted 3D structures of HR from its PDB structure are 4.6 and 37.9 degrees for s2, as well as 3.7 and 35.6 degrees for structure s3. It is clearly seen that the prediction in $\Phi$ is improved if a more correct secondary structure is used for structure prediction. Larger deviation in the $\Phi$ prediction occurs for helices B, F, and G when the secondary structure s2 is assumed, and for helices B and G when the secondary structure s3 is used. In general, prediction error of $\Phi$ is consistent with the mismatch between the predicted secondary structures and the PDB secondary structure. The only exception is the G-helix, which has a large deviation in the $\Phi$ prediction but a small mismatch. It is also found that the predicted $\Theta$ values for these three secondary structures have similar level of accuracy since the predicted lengths of HR helices are about the same.

In predicting the structure of MPs, a typical approach is to build structural model of certain MP on the template of a homologous MP, although there is no physical principle to justify its applicability. It would be quite interesting to compare the predicted structure of MPs from our dual-scale approach with their model structure constructed by the homology modeling. For example, the sequence identity is 26% between SRI and BR, 23% between SRI and HR, and 32% between BR and HR (London and Khorana, 1982). The structure model of SRI has been thus built on the template of the homologous BR (Lin and Yan, 1997). One potential risk of the homology method is the packing of helices. The calculated value of the packing parameter $r_0$ is 8.3 Å for SRI but is only 7.9 Å for BR. The constructed model structure of SRI on the BR template would be over packed, which might vary the orientation angle of helices since both the packing size and helix orientation are dominated by those interactions to pack helices. As shown in Supplementary Fig. S8 (supplementary information, 2008), we have compared the 3D structures of SRI obtained by the homology method, MC simulation, and MD refinement (A), as well as their predicted tilting and orientation angles of seven helices (B). It is clear that, although overall packing of helices looks similar, discrepancy in the prediction of the orientation angle is observed for most helices. However, the prediction of the tilting angle is roughly consistent (except for the E-helix), since helix tilting is mostly dominated by the competition between helix–water and helix–lipid interactions. So far the crystal structure of SRI is not available in PDB, it is still difficult to resolve the above discrepancy in orientation angles between our dual-scale approach and the homology modeling. However, due to the fact that both SRI and HR have similar packing size, we believe that it is more sensible to predict the structure of SRI by using HR as the template in the homology approach. We note that an interesting structural similarity between human $\beta_2$ adrenergic receptor ($\beta_2$AR) and bovine rhodopsin has been found recently (Rasmussen et al., 2007). The RMSD for the alpha carbon backbone of the transmembrane segments of these two MPs is only 1.56 Å, in spite of a small sequence identity between them (21.1% for the whole sequence and 23.2% for transmembrane segments). In this case, the packing parameters of these two MPs are also found to be similar to each other (8.5 Å for bovine rhodopsin and 8.7 Å for $\beta_2$AR), and are much larger than that of BR (7.8 Å).

## 6. Conclusions

In summary, we have shown that the proposed CG model of HBMP folding can efficiently predict the structure of RPs with a RMSD from the PDB structure (in coordinates of helix backbone atoms) 2.59 Å for HR, 3.12 Å for SRII, 3.99 Å for BR, and 5.56 Å for bovine rhodopsin. The packing position and tilting angle of each helix of these RPs can be accurately predicted and well understood in the CG model. It is found that the vdW interaction among helices determines their packing position in membrane, while the competition between the helix–water and helix–lipid interactions determines the tilting angle of helices. Refinement of our CG structures by AA models reduces the value of RMSD from the PDB structure (in coordinates of helix backbone atoms) to 1.89 Å for HR, 1.92 Å for SRII, 2.64 Å for BR, and 5.54 Å for bovine rhodopsin. Side-chain packing is found to be responsible for the orientation of helices. It is clear that our dual-scale approach is effective in the structure prediction of RPs without using their crystal structure information in PDB. Furthermore, in the case of using predicted secondary structures with certain alignment error, reasonable prediction of HR structure has also suggested the validity of this approach. Although we believe our proposed interactions are generic for RPs, further investigation is required to test this dual-scale approach for other MPs. It is interesting to note that the sequence dependence of the vdW interaction term in our CG model is implicit on $r_0$ which can be calculated using AA models. We note that, for simplicity, $r_0$ is calculated using AMBER in the absence of lipids in this study, and more accurate calculation of $r_0$ is possible when lipids are included. In fact, both $r_0$ and $E_{\mathrm{hw}}$ depend on the exact sequence of the protein chain. The former determines the geometric size of packing and the latter determines the secondary structure. Although there is no crystal structure of SRI in PDB, we have made an interesting comparison of the predicted structures of SRI from our dual-scale approach and the homology method. Discrepancy in the prediction of the orientation angle of helices is observed for these two approaches. This comparison suggests that, in addition to sequence identity, the packing distance between helices also plays an important role in predicting the structure of HBMPs when the homology modeling is adopted.

In addition to good structure predicting properties of our approach, as demonstrated in the Supplementary information, we also note the possibility of studying thermodynamics of MP folding using MC simulations. Starting from a random initial state, it takes only a few days to find the native state of our CG model energy of RPs by MC simulations on a regular desktop PC. This implies that our model energy landscape of tested RPs is considerably smooth. This observation may also reflect the simple topology of the structures of tested RPs, whose loops are relatively short mostly and neighboring helices are effectively constrained. It should be noted that identifying ground state structure of proteins by regular MC simulations is usually not guaranteed. Therefore, the identification of ground state structures of HBMPs in this study is carried out by PT simulations, which is efficient in finding the ground state but lacks dynamic information of MP folding. We note that the proposed physical model in this article is a minimum model and has only been tested for specific membrane proteins. To go beyond our current study, a more general CG model of HBMP folding is under construction, which includes side-chain information of all residues. In this case, it would take much longer computational time to search the ground state structure of HBMP folding from a random initial structure by regular MC simulations. Considerable improvement should be made in computer speed and our simulation algorithm of HBMP folding in order to predict accurate folded structures of HBMPs as well as to study thermodynamics of HBMP folding at a higher resolution.

## Acknowledgments

This work is supported, in part, by the National Science Council of Taiwan under Grant of No. 96-2112-M-003-002.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jsb.2008.10.001.

## References


Anfinsen, C., 1973. Principles that govern the folding of protein chains. Science 181, 223–230.

Archer, E., Maigret, B., Escrieut, C., Pradayrol, L., Fourmy, D., 2003. Trends Pharmacol. Sci. 24, 36–40.

Baldwin, J.M., 1998. The probable arrangement of the helices in G protein-coupled receptors. EMBO J. 12, 1693–1703.

Baudry, J., Crouzy, S., Roux, B., Smith, J.C., 1999. Simulation analysis of the retinal conformational equilibrium in dark-adapted bacteriorhodopsin. Biophys. J. 76, 1909–1917.

Berman, H.M. et al., 2000. The protein data bank. Nucleic Acids Res. 28, 235–242.

Booth, P.J., Curran, A.R., 1999. Membrane protein folding. Curr. Opin. Struct. Biol. 9, 115–121.

Bowie, J.U., 2005. Solving the membrane protein folding problem. Nature 438, 581–589.

Case, D.A. et al. 2002. AMBER7.

Chen, C.-M., 2000. Lattice model of transmembrane polypeptide folding. Phys. Rev. E. 63, 010901.

Chen, C.-M., Chen, C.-C., 2003. Computer simulations of membrane protein folding: structure and dynamics. Biophys. J. 84, 1902–1908.

Chen, C.-M., Higgs, P.G., 1998. Monte-Carlo simulations of polymer crystallisation in dilute solution. J. Chem. Phys. 108, 4305–4314.

Chen, C.-C., Wei, C.-C., Sun, Y.-C., Chen, C.-M., 2008. Packing of transmembrane helices in bacteriorhodopsin folding: Structure and thermodynamics. J. Struct. Biol. 162, 237–247.

Davies, A., Schertler, G.F., Gowen, B.E., Saibil, H.R., 1996. Projection structure of an invertebrate rhodopsin. J. Struct. Biol. 117, 36–44.

Dobbs, H., Orlandini, E., Bonaccini, R., Seno, F., 2002. Optimal potentials for predicting inter-helical packing in transmembrane proteins. Proteins 49, 342–349.

Floriano, W.B., Vaidehi, N., Goddard III, W.A., Singer, M.S., Shepherd, G.M., 2000. Molecular mechanisms underlying differential odor responses of a mouse olfactory receptor. Proc. Natl. Acad. Sci. USA 97, 10712–10716.

Gerstein, M., 1998. Patterns of protein-fold usage in eight microbial genomes: a comprehensive structural census. Proteins 33, 518–534.

Hansmann, U.H.E., 1997. Parallel tempering algorithm for conformational studies of biological molecules. Chem. Phys. Lett. 281, 140–150.

Henderson, R., 1977. The purple membrane from Halobacterium halobium. Annu. Rev. Biophys. Bioeng. 6, 87–109.

Herzyk, P., Hubbard, R.E., 1998. Combined biophysical and biochemical information confirms arrangement of transmembrane helices visible from the three-dimensional map of frog rhodopsin. J. Mol. Biol. 281, 741–754.

Huang, K.-S., Bayley, H., Liao, M.-J., London, E., Khorana, H.G., 1981. Refolding of an integral membrane protein. Denaturation, renaturation and reconstitution of intact bacteriorhodopsin and two proteolytic fragments. J. Biol. Chem. 256, 3802–3809.

Huschilt, J.C., Hodges, R.S., Davis, J.H., 1985. Phase equilibria in an amphiphilic peptide-phospholipid model membrane by deuterium nuclear magnetic resonance difference spectroscopy. Biochemistry 24, 1377–1386.

Kalani, M.Y., Vaidehi, N., Hall, S.E., Trabanino, R.J., Freddolino, P.L., et al., 2004. The predicted 3D structure of the human D2 dopamine receptor and the binding site and binding affinities for agonist and antagonists. Proc. Natl. Acad. Sci. USA 101, 3815–3820.

Kokubo, H., Okamoto, Y., 2004. Self-assembly of transmembrane helices of bacteriorhodopsin by a replica-exchange monte carlo simulation. Chem. Phys. Lett. 392, 168–175.

Kolbe, M., Besir, H., Essen, L.-O., Oesterhelt, D., 2000. Structure of the light-driven chloride pump halorhodopsin at 1.8 Å resolution. Science 288, 1390–1396.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580.

Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132.

Lin, S.L., Yan, B., 1997. Three-dimensional model of sensory rhodopsin I reveals important restraints between the protein and the chromophore. Protein Eng. 10, 197–206.

Liwo, A., Lee, J., Ripoll, D.R., Pillardy, J., Scheraga, H.A., 1999. Protein structure prediction by global optimization of a potential energy function. Proc. Natl. Acad. Sci. USA 96, 5482–5485.

London, E., Khorana, H.G., 1982. Denaturation and renaturation of bacteriorhodopsin in detergents and lipid-detergent mixtures. J. Biol. Chem. 257, 7003–7011.

Milik, M., Skolnick, J., 1992. Spontaneous insertion of polypeptide-chains into membranes – a Monte-Carlo model. Proc. Natl. Acad. Sci. USA 89, 9391–9395.

Moreau, J.L., Huber, G., 1999. Central adenosine $A_{2A}$ receptors: an overview. Brain Res. Rev. 31, 65–82.

Nina, M., Roux, B., Smith, J.C., 1995. Functional interactions in bacteriorhodopsin: a theoretical analysis of retinal hydrogen bonding with water. Biophys. J. 68, 25–39.

Oesterhelt, D., 1976. Bacteriorhodopsin as an example of a lightdriven proton pump. Angew. Chem. Int. Ed. Engl. 15, 17–24.

Oesterhelt, D., Stoeckenius, W., 1973. Functions of a new photoreceptor membrane. Proc. Natl. Acad. Sci. USA 70, 2853–2857.

Ou, D.-M., Chen, C.-C., Chen, C.-M., 2007. Contact-induced structure transformation in transmembrane prion propagation. Biophys. J. 92, 2704–2710.

Pardo, L., Ballesteros, J.A., Osman, R., Weinstein, H., 1992. On the use of the transmembrane domain of bacteriorhodopsin as a template for modeling the three-dimensional structure of guanine nucleotide-binding regulatory protein-coupled receptors. Proc. Natl. Acad. Sci. USA 89, 4009–4012.

Pebay-Peyroula, E., Rummel, G., Rosenbusch, J.P., Landau, E.M., 1997. X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases. Science 277, 1676–1681.

Pappu, R.V., Marshall, G.R., Ponder, J.W., 1999. A potential smoothing algorithm accurately predicts transmembrane helix packing. Nat. Struct. Biol. 6, 50–55.

Popot, J.-L., Engelman, D.M., 1990. Membrane protein folding and oligomerization: The two-stage model. Biochemistry 29, 4031–4037.

Popot, J.-L., Engelman, D.M., 2000. Helical membrane protein folding, stability, and evolution. Annu. Rev. Biochem. 69, 881–922.

Rasmussen, S.G.F., Choi, H.-J., Rosenbaum, D.M., Kobilka, T.S., Thian, F.S., et al., 2007. Crystal structure of the human b2 adrenergic G-protein-coupled receptor. Nature 450, 383–388.

Royant, A., Nollert, P., Edman, K., Neutze, R., Landau, E.M., Pebay-Peyroula, E., Navarro, J., 2001. X-ray structure of sensory rhodopsin II at 2.1-Å resolution. Proc. Natl. Acad. Sci. USA 98, 10131–10136.

Simons, K.T., Kooperberg, C., Huang, E., Baker, D., 1997. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J. Mol. Biol. 268, 209–225.

Subczynski, W.K., Lewis, R.N.A.H., McElhaney, R.N., Hodges, R.S., Hyde, J.S., Kusumi, A., 1998. Molecular organization and dynamics of 1-palmitoyl-2-oleoylphosphatidylcholine bilayers containing a transmembrane α-helical peptide. Biochemistry 37, 3156–3164.

Tajkhorshid, E., Paizs, B., Suhai, S., 1999. Role of isomerization barriers in the p $K_a$ control of the retinal Schiff base: a density functional study. J. Phys. Chem. B. 103, 4518–4527.

Trabanino, R.J. et al., 2004. First principles predictions of the structure and function of G-protein-coupled receptors: validation for bovine rhodopsin. Biophys. J. 86, 1904–1921.

Tsong, T.Y., 1990. Electrical modulation of membrane proteins. Annu. Rev. Biophys. Chem. 19, 83–106.

Wallin, E., von Heijne, G., 1998. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. Protein Sci. 7, 1029–1038.

White, S.H., Wimley, W.C., 1999. Membrane protein folding and stability: physical principles. Annu. Rev. Biophys. Biomol. Struct. 28, 319–365.

Yarov-Yarovoy, V., Schonbrun, J., Baker, D., 2006. Multipass membrane protein structure prediction using Rosetta. Proteins 62, 1010–1025.

Zhang, Y., DeVries, M.E., Skolnick, J., 2006. Structure modeling of all identified G protein-coupled receptors in the human genome. PLoS Comput. Biol. 2, 88–99.

Zhang, Y., Kolinski, A., Skolnick, J., 2003. TOUCHSTONE II: a new approach to ab initio protein structure prediction. Biophys. J. 85, 1145–1164.

Zheng, L., Herzfeld, J.J., 1992. NMR studies of retinal proteins. J. Bioenerg. Biomembr. 24, 139–146.