Eur Biophys J (2000) 28: 663–673

ARTICLE

Hyeon S. Son · Ian D. Kerr · Mark S.P. Sansom

Simulation studies on bacteriorhodopsin bundle of transmembrane α segments

Received: 19 April 1999 / Revised version: 6 September 1999 / Accepted: 17 September 1999

Abstract Bacteriorhodopsin (BR) is a membrane protein which pumps protons through the plasma membrane. Seven transmembrane BR helical segments are subjected to simulation studies in order to investigate the packing process of transmembrane helices. A Monte Carlo simulated annealing protocol is employed to optimize the helix bundle system. Helix packing is optimized according to a semi-empirical potential mainly composed of six components: a bilayer potential, a crossing angle potential, a helix dipole potential, a helix-helix distance potential, a helix orientation potential and a helix-helix distance restraint potential (a loop potential). Necessary parameters are derived from theoretical studies and statistical analysis of experimentally determined protein structures. The structures from the simulations are compared with the experimentally determined structures in terms of geometry. The structures generated show similar shapes to the experimentally suggested structure even without the helix-helix distance restraint potential. However, the relative locations of individual helices were reproduced only when the helix-helix distance restraint potential was used with restraint conditions. Our results suggest that transmembrane helix bundles resembling those observed experimentally may be generated by simulations using simple potentials.

H.S. Son (✉)
National Creative Research Initiative Center
for Superfunctional Materials, Department of Chemistry,
Pohang University of Science and Technology,
San 31, Hyojadong, Namgu, Pohang 790-784,
Republic of Korea
e-mail: hyeon@chem.postech.ac.kr

H.S. Son · M.S.P. Sansom
Laboratory of Molecular Biophysics,
The Rex Richards Building, University of Oxford,
South Parks Road, Oxford OX1 3QU, UK

I.D. Kerr
Nuffield Department of Clinical Biochemistry,
University of Oxford, Level 4, John Radcliffe Hospital,
Oxford OX3 9DU, UK

Key words Simulated annealing · Monte Carlo · Simulation · Bilayer potential · Bacteriorhodopsin

Introduction

Halobacterium halobium synthesizes a particular region of the membrane that converts energy from visual light into stored energy by pumping protons through the plasma membrane. This region is the purple membrane, of which bacteriorhodopsin (BR) is the major protein. This protein has a prosthetic group, retinal, which is responsible for the proton pump (Stoeckenius and Rawen 1967; Oesterhelt and Stoecknius 1973). The seven-transmembrane domain structure of BR was first determined at 7.0 Å resolution by electron microscopy (Henderson and Unwin 1975). This structure has been extended to near atomic resolution, 3.0 Å (horizontal) and 10.0 Å (vertical), providing invaluable information about transmembrane helix packing (Henderson et al. 1990). The vertical resolution of the structure has been improved to 4.3 Å. The structure was refined using tight geometric restraints to improve the model (Grigorieff et al. 1996). BR forms an extremely regular two-dimensional array (Blaurock and Stoeckenius 1971). BR has seven transmembrane helices which are roughly perpendicular to the plane of the lipid bilayer and the helices are linked by the extracellular and the cytoplasmic loops. Helices A to G are arranged in anticlockwise order when they are viewed from outside the bilayer. All neighbouring helices are anti-parallel except for the GA pair which is formed by the first helix (A) and the last helix (G) which close the bundle. Helices B, C and F contain proline residues which are believed to induce the helix kinks (Barlow and Thornton 1988; Sankararamakrishnan and Vishveshwara 1992).

Simulated annealing via molecular dynamics (SA/MD; Kirkpatrick et al. 1983) has been used successfully to refine structures from X-ray diffraction and from NMR spectroscopy (Brünger et al. 1990; Karplus and

Pesko 1990; Brünger 1992). Similar SA/MD methods may be able to predict structures provided that the initial structure is reasonably close to the native conformation. This is to set the system within the region where the global minimum could be found without falling into one of the local minima. To set the initial structure, information from experiments and theoretical studies could be used. For example, membrane spanning regions could be predicted and the alignment of the transmembrane segments could be determined by labelling experiments (Hubbell and Altenbach 1994). If a complex problem can be broken down into several smaller ones, the problem may be solved hierarchically. The proposed "two-stage folding" model by Popot and Engelman (1990) suggests that this may be possible for membrane proteins. According to this proposal, the process of folding for transmembrane proteins can be divided into two distinctive thermodynamic stages: first, stable helices are formed, and secondly, these transmembrane helices pack together to form the native structure of the integral membrane protein. BR has proven to be an excellent testing system to verify this hypothesis. A number of experiments on isolated BR segments have been carried out and it has been successfully shown that reconstituting BR (in vitro) is possible (Arseniev et al. 1988; Barsukov et al. 1992; Kahn and Engelman 1992; Pervushin and Arseniev 1992; Orekhov et al. 1994). This "two-stage folding" model implies that transmembrane helices can be treated as pre-existing folding elements, i.e., if the helix domains are predicted, the helix packing simulation does not have to start from a random or a linear peptide conformation. The lipid bilayer environment plays a crucial role in transmembrane helix packing. Therefore, it is desirable to include the bilayer environment in the simulation system in order to characterize the behaviour of transmembrane helices. Membrane lipids are amphipathic molecules and so a bilayer has two hydrophilic surfaces with a hydrophobic region in between. Using this property, a simple method mimicking the lipid bilayer environment was employed (Son and Sansom 1999). An exponential type of function suggested by Edholm and Jähnig (1988; Jähnig and Edholm 1992) was used to represent the hydrophobic potential differences experienced by constituent amino acids of helices.

Simulated annealing via Monte Carlo (SA/MC) simulation (TMH simulation) studies, described in our previous paper (Son and Sansom 1999), have shown that helix bundles resembling those observed experimentally may be predicted by simulations using simple potentials. BR-like, δ-endotoxin-like and cytochrome *c* oxidase-like structures have been observed as low energy structures (the BR-like and the δ-endotoxin-like structures are from a 7-Ala_{20} simulation and the cytochrome *c* oxidase-like structure is from a 12-Ala_{20} simulation). Both BR-like and δ-endotoxin-like structures co-exist as low-energy structures in simulations using native BR sequence. We know from the results of experimental studies (Henderson et al. 1990; Grigorieff et al. 1996) that the real BR sequence does not adopt the δ-endotoxin-like structure, but it was still observed as the lowest energy conformation in the simulation. Therefore, deciding on representative structures among low-energy configurations was not straightforward. TMH simulation has so far employed a semi-empirical potential composed of four components: a helix-helix distance potential, a crossing angle potential, a helix dipole potential and a bilayer potential. These components could be categorized into two different types of potentials: "helix-helix interaction potentials" and "helix-environment interaction potentials". Another potential, a "helix orientation potential", is an addition to the "helix-environment interaction potentials" and is intended to solve the problem of obtaining representative structures among low-energy configurations. This potential was included in the TMH potential function. The "helix orientation potential" is intended to orient certain helix faces to be in contact with the surrounding lipids. TMH simulation including this new potential produced an ensemble of structures among which the BR-like structures were the lowest energy structures. This means that the "helix orientation potential" mediated the helix orientation in the simulation to generate the native-like structure as the lowest energy structure. However, the positions of constituent helices generated were different from the energy minimized (EM) determined structure. This was an expected result, because there were no restraints in the TMH simulation to arrange the helices in any specified positions. Further simulations were conducted with sets of restraints [by employing a "helix-helix distance restraint potential (loop potential)"], and the results showed that the helices were arranged in a preferred order. These results suggested that the lowest energy structure from the simulation could be used as the initial structure for the further refinement using a SA/MD simulation method. A SA/MD simulation was conducted in order to refine the TMH produced representative structure. In this study, other atoms were introduced to allow internal flexibility of the helices, i.e., inter-atomic interaction could be simulated, and helical kink/bend was also allowed.

This paper describes the implementation of the potentials, the "helix orientation potential" and the "helix-helix distance restraint potential", and discusses the structures resulting from the BR TMH simulations that implemented the potentials. The SA/MD simulation conducted with the representative BR structure from TMH simulation is described and the results are also discussed. Structures at the end of each simulation are analysed and the results are presented in the relevant sections.

Materials and methods

BR sequence

Only the seven transmembrane helical segments were used in our simulations. The residues are labelled such

that the residue numbers restart from the first residue of each segment. Helix segments are defined as follows:

Helix A: E_1WIWL_5ALGTA$_{10}$LMGLG$_{15}$TLYFL$_{20}$VK-GM
Helix B: P_1DAKK$_5$FYAIT$_{10}$TLVPA$_{15}$IAFTM$_{20}$YLS-ML$_{25}$L
Helix C: Y_1WARY$_5$ADWLF$_{10}$TTPLL$_{15}$LLDLA$_{20}$LLV
Helix D: T_1ILAL$_5$VGADG$_{10}$IMIGT$_{15}$GLVGA$_{20}$L
Helix E: F_1VWWA$_5$ISTAA$_{10}$MLYIL$_{15}$YVLFF$_{20}$GFT
Helix F: E_1VAST$_5$FKVLR$_{10}$NVTVV$_{15}$LWSAY$_{20}$PVV-WL$_{25}$IGS
Helix G: N_1IETL$_5$LFMVL$_{10}$DVSAK$_{15}$VGFGL$_{20}$ILL-RS$_{25}$RA

TMH simulation potential

The SA/MC method was employed to minimize the energy of the system. The Metropolis prescription (Metropolis et al. 1953) was used to determine whether or not to accept each step of the selection. Helix packing was optimized by the SA/MC method (TMH method) employing a semi-empirical potential composed of six terms:

$$V_T = w_d V_d + w_b V_b + w_\Omega V_\Omega + w_\mu V_\mu + w_\omega V_\omega + w_{loop} V_{loop} \quad (1)$$

where V_T is the total potential of the system, V_d is a potential function dependent on the inter-helical separation distance, V_b represents the sequence dependent helix/bilayer interaction energy, V_Ω is a crossing angle term allowing for "ridges into grooves" (Chothia et al. 1981) packing of adjacent helices, V_μ is the dipole potential, V_ω is the helix orientation potential and V_{loop} is the helix-helix distance restraint potential (loop potential). $w_d, w_b, w_\Omega, w_\mu, w_\omega$ and w_{loop} are the weights applied to each energy term. The same potential function with only first four terms [in Eq. (1)] was used to investigate the packing process of transmembrane helices in our previous study. The definition and the description of the terms are fully given in our previous report (Son and Sansom 1999).

Helix orientation potential

Previous TMH simulation studies (Son and Sansom 1999) suggested that more term(s) should be added to the potential function to generate a lowest energy structure as a representative structure. This problem was noticed when BR- and δ-endotoxin-like structures were generated when simulated with seven transmembrane helix bundles (7-Ala$_{20}$ and simulations with real BR sequence). The δ-endotoxin-like structure is a compact bundle of six helices surrounding a central helix, while the BR-like structure forms an elongated elliptically shaped structure. A clear difference between them is the central helix, i.e., the BR-like structure lacks a central helix which has no contact with the surrounding lipid environment. Furthermore, previous TMH methods were not able to predict the rotational orientation of individual helices with respect to the lipid environment. A "helix orientation" potential was thus implemented into the existing TMH potential function in order to solve these problems. The potential is intended to keep the hydrophobic helix faces in contact with the lipid environment and consequently to make the δ-endotoxin-like structure unfavourable (if the central helix in the δ-endotoxin-like structure is strongly lipid accessible, then this helix-buried structure is not favourable). Analysis of helix face orientation with respect to the lipid environment shows that helices have a tendency of orienting specific faces pointing to the lipid environment (PERSCAN analysis); analysis on the PS/RC structure from *Rhodobacter sphaeroides* showed that the lipid-exposed residues were more hydrophobic than buried interior residues (Rees et al. 1989). Also, from the analysis of PS/RC sequences (Yeates et al. 1987) from closely related species, it has been found that the lipid-exposed residues of transmembrane helices were poorly conserved (instead, those buried inside the bundles were observed to be more highly conserved). Donnelly et al. (1993) used Fourier analysis of sequence variability between 33 PS/RC sequences in order to tabulate a substitution table for lipid accessible residues of transmembrane helices.

A "helix orientation angle (ω)" of a helix is defined by the angle between the helix orientation vector and the vector joining the centre of the bundle and the centre of the helix (Fig. 1). The following equation is used to form a "helix orientation" potential:

$$E_{\omega l} = \cos(\omega) \quad (2)$$

where $E_{\omega l}$ is the "helix orientation" potential of helix l, and ω is the angle between most outward-facing vector of helix l and the vector joining the centre of the bundle and the centre of the helix:

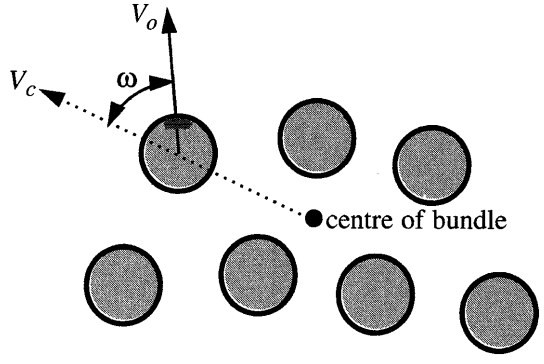

Fig. 1 Definition of helix orientation angle (ω). Transmembrane helices are simplified as cylinders and they are viewed perpendicular to the lipid plane from outside of the lipid bilayer. V_c and V_o are direction vectors

$$V_\omega = \sum_{l=1}^{N} E_{\omega l} \tag{3}$$

where N is the number of helices in the system. This potential was included in the total potential [Eq. (1)] of the TMH simulation system for Metropolis Monte Carlo selection.

Helix-helix distance restraint potential (loop potential)

Inter-helical distance restraints can be set by this potential term. This potential was originally designed to mimic the existence of loops between transmembrane helical segments. If the maximum length that the loop region can stretch is known, this potential will maintain two helices within the "loop length distance". However, the potential is flexible enough that the end-to-end distance of any pair of helices can be restrained; solid state NMR could give information about helix-helix distances and other chemical methods such as fluorescence energy spectroscopy, and cysteine scanning mutagenesis (Kono et al. 1996) experiments could also restrain the helix-helix distances. The loop potential is defined as follows:

$$E_{\text{loop}} = \begin{cases} E_{\min} & \text{if } d_e \le D_{\text{cut}} \\ 0 & \text{if } d_e > D_{\text{cut}} \end{cases} \tag{4}$$

where d_e is the end-to-end distance of a pair of helices and D_{cut} is the maximum cutoff distance which two helices can separate. The total loop potential V_{loop} is defined as:

$$V_{\text{loop}} = \sum_{l=1}^{N-1} \sum_{m=n+1}^{N} E_{\text{loop}} \tag{5}$$

where l and m are helix indices and N is the total number of helices. The summation is over every pair of helices to which the loop potential is applied. This total loop potential term, V_{loop}, was included in the total potential function [Eq. (1)].

PERSCAN analysis: parameterization of orientation potential

The rotational orientation of helices can be characterized by the helix orientation angles. The PERSCAN analysis is designed to predict the lipid-inaccessible faces for the BR helices. Initially, 34 sequences for members of the bacteriorhodopsin family were retrieved from the PIR (protein information resource) database version 47.0. The program AMPS (Barton 1990) was used to calculate the pairwise amino acid identities and these values were used to reduce the initial sequence set for each family to a representative subset. A total of 21 sequences were excluded and the remaining 13 sequences [the PIR accession codes and classification of the remaining 13 sequences are as follows: A34178 (archaerhodopsin precursor from *Halobacterium* spp.), S14731 (archaerhodopsin 2 from *Halobacterium* spp.), S35730 (bacteriorhodopsin from *Halobacterium* spp.), S51206 (cruxrhodopsin from *Halourcula* spp.), S76743 (cruxrhodopsin-2), S55297 (sensory rhodopsin II apoprotein from *Haloarcula vallismortis*), S55300 (sensory rhodopsin II apoprotein from *Natronobacterium pharaonis*), S29988 (halorhodopsin from *Halobacterium* spp.), D43766 (halorhodopsin from *Halobacterium salinarium*), A35002 (halorhodopsin from *Natronobacterium pharaonis*), A26161 (halorhodopsin precursor from *Halobacterium* spp.), S29989 (sensory rhodopsin I from *Halobacterium* spp.), S09277 (sensory rhodopsin I from *Halobacterium halobium*)] were then subjected to the multiple sequence alignment analysis using the program AMPS. Transmembrane helical segments were predicted using the program TMS-PRED (Persson and Argos 1994). Periodicity analysis of putative transmembrane segments was then performed by the program PERSCAN (Donnelly et al. 1993). PERSCAN employs Fourier transform analysis to identify periodicity, in characteristics such as lipid inaccessibility, hydrophobicity and residue conservation, along the sequence blocks from multiple sequence alignments. From the results, the lipid-inaccessible face for each helix was identified. Figure 2 presents the results in the form of helical wheel.

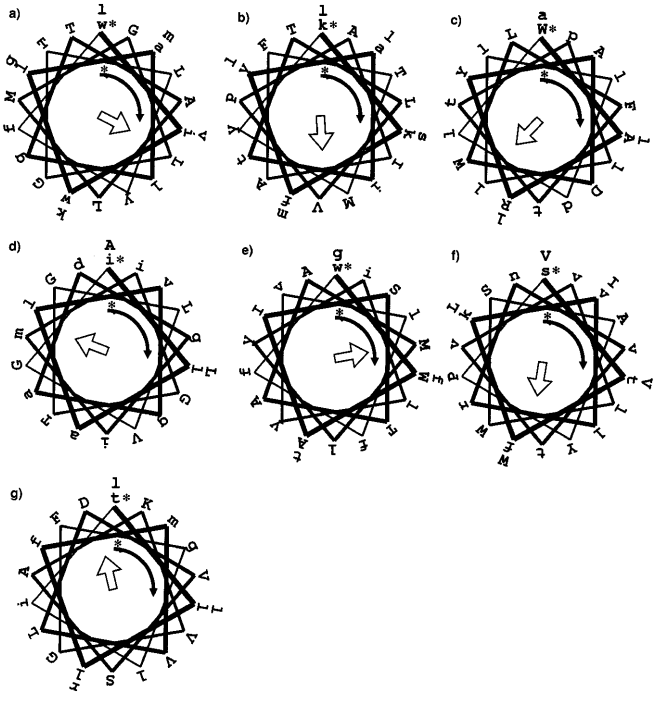

Fig. 2a–g Results of PERSCAN analysis for seven transmembrane segments of BR. Each helical wheel contains the BR transmembrane segment and periodicity analysis. The lipid-inaccessible faces are directed by *open arrows*, the symbol * indicates the first residue of the segment and the *solid arrows* indicate the directions of the helix wheel

TMH simulation protocol

The protocol used for the simulation presented in this section is similar to the one stated in our previous paper (Son and Sansom 1999). Ideal helices of Cα atoms were constructed with the BR sequence. The bundle system was randomized by heating the system to a temperature of 90,000 K. Subsequent abrupt cooling of the system to 7000 K fixed the randomized structure and then the system was cooled slowly down to 300 K to adopt the simulated annealing process. A step temperature scale of 50 K was used to decrement the temperature from 7000 K to 300 K, and for each step temperature, 600 times of sampling was done (300 times of translational sampling and 300 times of rotational). An ensemble of 50 structures was generated at the end of the simulation.

An inter-helical distance of 8.8 Å and crossing angles of 14° for parallel and −173° for anti-parallel pairs of helices (Sansom et al. 1995) were set as the parameters of the helix distance potential and crossing angle potential. The bilayer potential parameter was set up using the GES scale (Engelmen et al. 1986). The dielectric constant ε was set to be 2.0. The weights were set up so that the minimum energy values of constituent potentials would be −10.0 kcal/mol (−20 kcal/mol for the loop potential) in all the simulations unless otherwise stated. Inter-helical restraints were imposed by the loop potential (E_{loop}), and various simulations were conducted with different restraint conditions. Table 1 lists the definition of the simulation models. In the BR-0 simulation, there was no inter-helical restraint imposed upon the simulation potential, while in the BR-11 simulation 11, restraints were imposed to hold neighbouring helices together, i.e., for a neighbouring helix pair $D_{\text{cut}} = 11$ Å was employed.

Simulations were performed on a DEC Alpha 2100 4/275. For each simulation, an ensemble of 50 structures was generated. For the generation of one structure, CPU time of approximately 120 min was required. Analyses of the structures were done on Silicon Graphics R3000 Indigos. Helix bundle geometry was analysed by ACTIVE (H.S. Son and M.S.P. Sansom, unpublished) and graphical analysis was done by ACTIVE and RasMol (Sayle 1996).

SA/MD simulation protocol

The lowest energy structure was selected from the ensemble of structures generated from the TMH simulation BR-11. The selected structure itself is similar to the EM-determined BR structure in terms of its shape, helix orientation and positions of the helices in the bundle. The initial structure for further simulation was thus provided by the TMH simulation. The subsequent SA/MD simulation is to refine the structure in order to generate more stable structures. The helices in the initial structure are rigid rods, i.e., no internal flexibility was allowed in the process of the TMH simulation. The SA/MD simulation introduced the flexibility by a full atom represented simulation procedure.

From the TMH BR-11 simulation, the lowest energy structure was selected to be the initial structure. Only Cα atoms were specified in this structure initially. The remaining atoms were then added onto the corresponding Cα atom positions (the TMH simulation procedure replaced the initial stage of the existing SA/MD protocol). These backbone and sidechain atoms were then released to be "grown out" from the fixed Cα atoms. Simulation started at 1000 K and at this temperature geometrical interactions and later a repulsive van der Waals term were slowly introduced. The system was then cooled down slowly to 300 K. The Cα atoms were fixed at their original locations. At the end of this stage of the simulation, five structures were generated and they were then fed into the MD stage (this is a molecular dynamics stage as a module of the SA/MD protocol) to generate ensembles of 25 structures. At this MD stage the Cα atoms were gradually allowed to move from their initial locations. A temperature of 500 K was initially assigned to the system. An electrostatic term was introduced by scaling polar sidechain partial atomic charges as the temperature was reduced from 500 K to 300 K. Partial sidechain charges were gradually scaled from 0 to 50% of their standard values. For the treatment of non-bonded interactions, a constant dielectric condition was employed ($\varepsilon = 2.0$) (Brooks et al. 1983). On reaching 300 K, a 5 ps burst of constant temperature dynamics was performed, followed by 1000 steps of conjugate gradient energy minimization. These ensembles were then subjected to the analysis which is described below. Xplor V3.0 (Brünger 1992) was used with the CHARMm PARAM19 (Brooks et al. 1983) parameter set to perform the SA/MD simulation throughout. Simulations were performed on a DEC Alpha 2100 4/275 workstation; for the generation of one structure, CPU time of approximately 130 min was required. The geometrical analyses were performed using ACTIVE and graphical analysis was done by ACTIVE and RasMol on Silicon Graphics R3000 Indigos.

Table 1 The types of TMH simulations. Simulation types are named according to the number of inter-helical restraints imposed by the loop potential

Simulation type	Inter-helical restraints (by E_{loop})
BR-0	None
BR-6	A-B, B-C, C-D, D-E, E-F, F-G
BR-7	A-B, B-C, C-D, D-E, E-F, F-G, C-F
BR-11	A-B, B-C, C-D, D-E, E-F, F-G, B-G, C-G, C-F, C-E, D-F

Results

TMH simulation

There are two sets of simulations: one is without any inter-helical distance restraints (simulation BR-0) and

the other is with the inter-helical restraints (simulations BR-6, -7 and -11). Different restraints were imposed to see if the restraints oriented the helices in the desired order. Figures 3–6 present the resultant structures from the simulations; the lowest energy structures are shown. The shapes of the structures are similar to the experimentally determined BR structure. However, the order of the helices in BR-0 is not that observed in the experimentally determined BR structure. The simulations BR-7 and BR-11 generated structures whose constituent helices are arranged as observed in BR. Therefore the helix orientation potential generated the BR-like structures and the loop potential sorted out the order of the helices. The structures are geometrically analysed and Table 2 shows the results of the analysis. The results all show that the simulation method has successfully optimized the system. The average crossing angles (over the ensemble of 50 structures) between neighbouring pairs of helices in the BR-11 simulation are −174° for AB, −173° for BC, −170° for CD, −172° for DE, −176° for

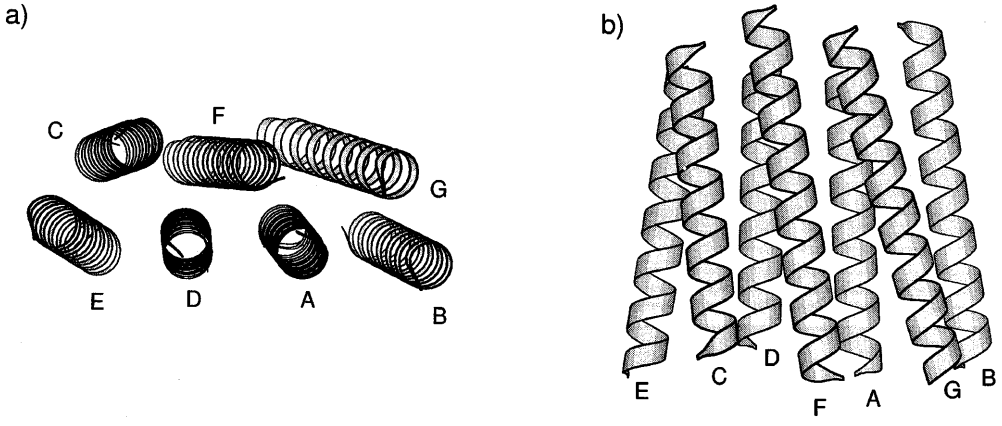

Fig. 3a, b Lowest energy structure of TMH simulation BR-0: **a** perpendicular view and **b** parallel view to the lipid bilayer plane

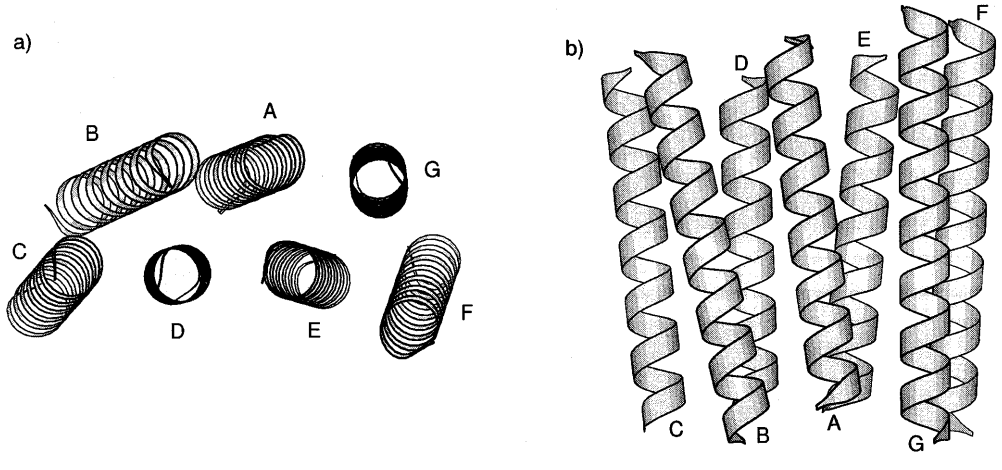

Fig. 4a, b Lowest energy structure of TMH simulation BR-6: **a** perpendicular view and **b** parallel view to the lipid bilayer plane

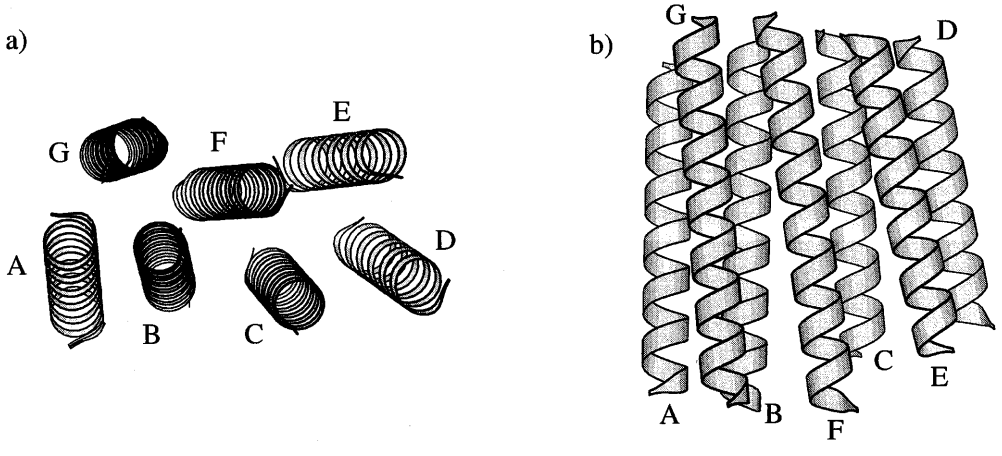

Fig. 5a, b Lowest energy structure of TMH simulation BR-7: **a** perpendicular view and **b** parallel view to the lipid bilayer plane

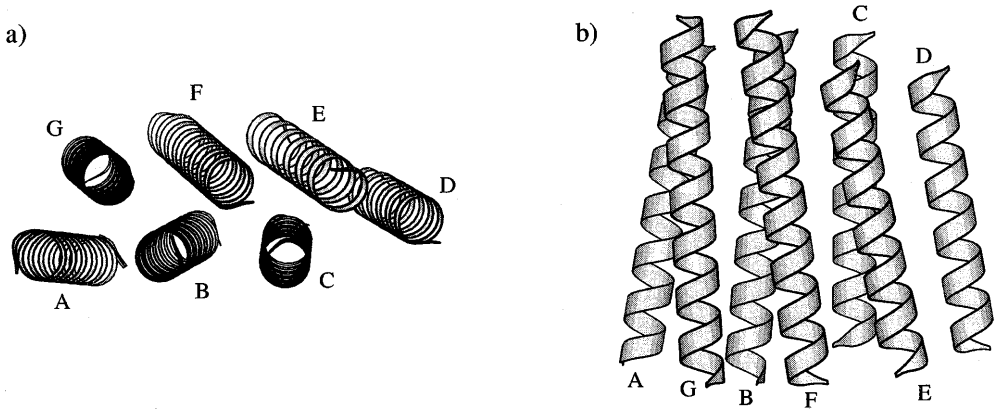

Fig. 6a, b Lowest energy structure of TMH simulation BR-11: **a** perpendicular view and **b** parallel view to the lipid bilayer plane

EF, −172° for FG, and 2.8° for GA. The EM-determined structure of BR has a positive crossing angle of 167° for the CD helix pair (Table 3a), i.e., the TMH simulation has not been able to predict this positive angle [it is too early to discuss this dissimilarity; the discussion is left for the later section (below) that describes the SA/MD results]. Figure 7 shows the energy trajectories for the lowest energy structures from the simulation BR-11. In order to investigate the effectiveness of the helix orientation potential, structure clustering was considered. Each structure was projected onto a two-dimensional plane (effectively onto the bilayer plane). A pair-wise root mean square deviation (RMSD) value was found by comparing two projected structures; one structure was fixed and the other structure was rotated through 360°. For each 1.0° rotation, the RMSD value between two structures was found by calculating the distances between corresponding helix pairs. Among these 359 RMSD values, the minimum value was taken as the pair-wise RMSD of the two structures. A pair-wise RMSD matrix thus was created. This matrix was then subjected to a cluster analysis using a C program "oc" (G.J. Barton, unpublished). The BR-0 simulation (and other simulations also) generated clusters of structures: δ-endotoxin-like, BR-like, rhodopsin-like, along with other randomly oriented structures. The occurrence of structural clustering was confirmed by visual inspection of the structures. The BR-like structures are lower energy structures than the others (in all the simulations with the helix orientation potential). This result is different from the simulation result that was presented in our previous report (Son and Sansom 1999). This indicates that the new "helix orientation potential" effectively influenced the simulation.

Table 2 Geometrical analysis of TMH simulated BR helix bundle structures; d_c is the closest distance between two helices

Simulation type	Ω (parallel) (°)	Ω (antiparallel) (°)	d_c (neighbour) (Å)
BR-0	−173.0 (0.9)	13.6 (0.6)	8.9 (0.3)
BR-6	−172.6 (0.8)	12.7 (1.1)	9.0 (0.3)
BR-7	−172.0 (2.6)	12.4 (0.6)	9.0 (0.5)
BR-11	−173.0 (2.7)	14.4 (0.7)	8.8 (0.3)

SA/MD simulation

Figures 8 and 9 show the result of the SA/MD simulation. The structure was chosen by the ensemble average method. The coordinates of Cα atoms of 25 structures from the ensemble were averaged and an average structure was found. This ensemble average structure was chosen to be the representative predicted structure (Fig. 9). The simulation certainly further minimized the TMH generated structure. However, the overall shape of the structure has been retained. The representative structure has been geometrically analysed and Table 3c shows the results of the analysis. The results show that the simulation method optimized the system. The geometrical analysis results agreed in reasonable accuracy with the analysis results from the experimentally determined structure (Table 3a). However, there are some differences in crossing angles: the SA/MD refined structure shows crossing angles of −173° and 179° for CD and EF pairs, but the EM structure shows 167° and −169° for the same pairs of helices. It has been noticed that the EM-determined BR structure has an "exceptional" helix-helix packing orientation for dimer CD, which shows a positive crossing angle against the expected negative crossing angle by the "ridges into grooves" theory (Chothia et al. 1981). However, in the SA/MD simulated structure the EF pair is observed to have this "exceptional" orientation. To investigate this observation, Table 3b is provided. Table 3b is the results of the geometrical analysis of the TMH-generated representative structure. As shown in Table 3b, the crossing angle of the EF pair before SA/MD refinement is negative, i.e., SA/MD refinement mediated the crossing angle in the EF pair. It should be noted that after the SA/MD refinement the orientation of the helix pair CF (anti-parallel helix pair) shows an "expected" negative crossing angle which was positive at the end of the TMH simulation (Table 3c). Therefore it might be that the "unusual" orientation of helices are due to the packing

Table 3 Geometrical analysis of (a) the experimentally determined BR helices (Henderson et al. 1990), (b) TMH simulated BR-11 and (c) SA/MD simulated BR helix bundle structures. The crossing angles (Ω) and the closest distances (d_c) are shown

BR dimer		AB	AC	AD	AE	AF	AG	BC	BD	BE	BF	BG	CD	CE	CF	CG	DE	DF	DG	EF	EG	FG
(a)																						
	Ω (°)	−157.9	28.7	−161.7	26.0	−163.4	8.0	−172.1	−5.5	−159.6	21.8	−161.5	166.6	18.6	−155.9	23.6	−156.6	22.1	−163.5	−168.8	18.1	−170.9
	d_c (Å)	8.1	14.2	21.8	27.9	19.6	9.5	8.8	18.4	21.1	15.3	9.8	9.5	11.4	11.0	11.6	8.5	12.1	16.2	10.7	19.1	8.1
(b)																						
	Ω (°)	−173.0	11.6	−158.1	25.0	−158.5	15.1	−175.3	−15.1	−161.8	14.7	−171.0	−169.5	5.2	13.5	−170	−176.9	−1.4	−172.5	−176.3	10.5	−172.3
	d_c (Å)	9.1	17.7	26.5	19.7	13.6	8.6	8.6	17.5	11.8	8.3	8.8	9.3	8.7	9.0	18.5	8.8	18.5	27.0	8.4	17.4	8.8
(c)																						
	Ω (°)	−171.6	19.4	−154.0	27.3	−152.3	19.3	−168.9	17.6	−161.0	19.5	−166.8	−173.4	9.1	−170.4	12.2	−175.8	4.7	−165.4	179.4	12.7	−167.1
	d_c (Å)	7.9	16.8	27.6	22.0	15.0	9.5	8.7	19.7	15.2	9.2	10.2	10.8	9.4	9.8	20.0	8.9	17.9	29.0	10.2	20.3	10.0

Fig. 7 Energy trajectory for the lowest energy structure from the simulation BR-11

competition among neighbouring helices. The kink angle analysis shows that the SA/MD-generated structure exhibits a kink ranging from 13.5° to 19.5° (13.5° for helix B, 19.5° for helix C and 16.6° for helix F). The kink angles calculated from the experimentally determined BR structures are 5.8° (B), 17.2° (C) and 22.0° (F). The kink in helix B shows a minor disagreement with the EM-determined structure. The SA/MD refinement did not alter the overall shape of the structure, i.e., the shape of the refined structure is dependent on the initial TMH-generated structure. Therefore, TMH generation of initial structure(s) is pressingly important.

Discussion

Figures 3–6 show the structures generated from TMH simulations of different restraints imposed by loop potential. The restraint potential was employed to arrange the helices in the sequential order (simulation BR-6); further, to arrange them in the positions where the EM-determined BR structure show (simulations BR-7 and BR-11). The restraints were not based on existing EM experimental data, but based on the data which feasible experiments could produce; solid state NMR could give information about helix-helix distances as could other chemical methods such as fluorescence energy spectroscopy and cysteine scanning mutagenesis (Kono et al. 1996) experiments, and for some cases, short loops

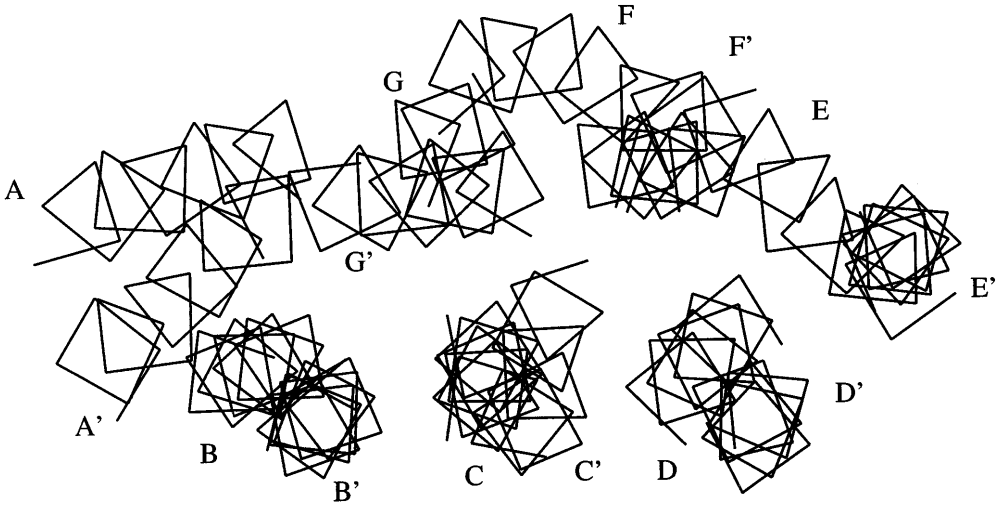

Fig. 8 Superimposed Cα-trace plot. The lowest energy structure from the BR-11 simulation (TMH-SA/MD) and the Brookhaven BR structure (2brd) are superimposed together

between helices could restrain the helix-helix distances (however, in the BR case, the loops are not sufficiently short to derive efficient positional restraints). From all the types of simulations, the shapes of the lowest energy structures are similar to the EM-determined structure. However, the positions of the helices in simulations BR-0 and BR-6 are different from those observed in the EM-determined BR structure. The simulations BR-7 and BR-11 exhibit the same shape and the same position of the helices. The BR-7 structure shows that the C-F restraints held the pair tightly together. Further positional restraints imposed in the BR-11 simulation produced a preferred structure in terms of the shape and positions of the helices. However, there also exists one other possible configuration that satisfies the condition: a clockwise ordered helix bundle. Among the simulated structures in the BR-11 simulation, it has been observed that some structures have clockwise oriented helices; among the clustered seven BR-like structures, two structures are anti-clockwise ordered, two are clockwise ordered and the rest are randomly ordered bundles. These clockwise ordered helix bundle structures could also have energies as low as the anti-clockwise oriented helix bundle (Fig. 6). The shapes of the structures are similar to the anti-clockwise helix bundle structure (EM-determined BR structure). Having implemented all the possible geometrical restraints, these clockwise ordered structures are energetically indistinguishable from the anti-clockwise helix bundle. This problem could be solved by experiments such as fluorescence energy spectroscopy and cysteine scanning mutagenesis (Kono et al. 1996) (the orientation of sidechains can be determined and this information can establish the order of helices). From the cluster analyses, it has been noticed that the introduction of restraints produced more randomly oriented structure (i.e., structures which are similar to none of the δ-endotoxin-like, BR-like and rhodopsin-like structures; the randomly oriented structures have higher potential values than other structures). This is because the restraints reduce the frequency of energy jumps between local energy minima, which results in structures trapped in one of the local energy minimal located in a high energy level.

So far, the method of predicting BR structure has been investigated. The RMSD value between the BR-11 structure and the EM structure is 0.43 Å. The method of calculating the RMSD value is similar to the one used for cluster analysis; both structures are projected onto a two-dimensional plane (the bilayer plane), and the pairwise RMSD value is derived by calculating the minimum distances between centres of corresponding helices. Figure 8 shows the superimposed structures. The next

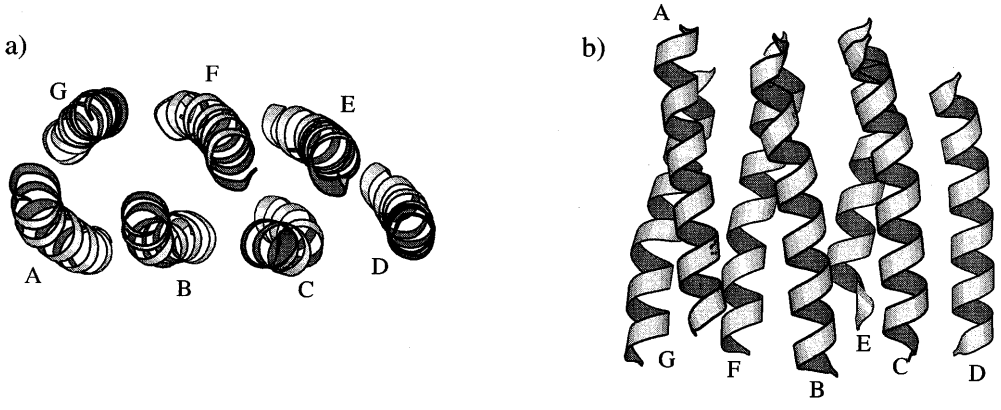

Fig. 9a, b Average RMSD BR structure at the end of the SA/MD simulation. The structure is the RMSD averaged ensemble structure

stage is to refine the representative structure generated by the TMH method. The SA/MD procedure is chosen to be the refinement method.

The lowest energy structures from the TMH simulations were analysed in terms of bundle geometry, and the result of the analysis is shown in Table 2 and discussed in the Results section. The shapes of the structures are similar to that of the EM-determined BR structure. Having employed the positional restraints, the lowest energy structures (simulations BR-7 and BR-11) showed preferred structures in terms of shapes and helix positions in the bundles. The lowest energy structure from the BR-11 simulation was then subjected to a SA/MD simulation. This procedure is to refine the representative structure (from the BR-11 simulation). The SA/MD simulation method generally searches for wide conformational space; therefore this refinement procedure enables the helix bundle to adjust itself to form a more stable conformation. Particularly, other atoms previously restrained to the Cα atoms were released to move in order to introduce the internal flexibility of a helix bundle. The SA/MD simulation results show that the prolonged kidney shape of the structure has been retained throughout the simulation. Therefore the shape of the structure was more or less pre-determined before the SA/MD simulation, i.e., by the TMH simulation, and the structure was then refined by the subsequent SA/MD simulation.

The helix-helix distance and the crossing angle analyses show that the predicted bundle of helices are reasonably similar to the EM-determined structure. The kinks in the proline-containing helices were observed in the simulated structures; this kink was clearly induced during the SA/MD stage. The helix-helix orientation was supporting the similarity between the simulated and the EM-determined structure.

A simulation method to model seven transmembrane helix bundles has been reported (Herzyk and Hubbard 1995). The method was to pack the transmembrane helices according to the restraints that were derived from experiments. A penalty function was constructed to evaluate the inter-helical restraints. BR has been used to test the method; a 7 Å EM density map of BR (Henderson and Unwin 1975) was used to derive the positional restraints. The simulation produced a model structure with a RMSD of 1.87 Å. This simulation study showed that experimental data were an invaluable source of information in predicting transmembrane protein structures. The method used, however, included neither the lipid environment nor the helix-helix orientation terms. This is a reasonable assumption because the restraints from the density map were sufficient to generate the model without including the lipid environment. However, in the absence of the EM density map, the method would have difficulty in predicting the packing characteristics of the helices such as helix-helix orientation and helix orientation with respect to the lipid membrane. The TMH method employed no geometrical restraints on the initial conformation, i.e., no EM density map is needed to predict the structure (if a density map is provided, then the CPU time will be greatly reduced as more positional restraints could be given to the simulation system). The TMH simulation-produced structure is analogous to the density map by EM experiments. However, the TMH-generated structure has no means of determining the correct positions of constituent helices [for EM-determined BR, the positions of the helices were determined by careful inspection of aromatic sidechains in the sequence and the helix density map (Henderson et al. 1990)]. Therefore a set of geometrical restraints are needed to position the helices in a bundle. The TMH simulation system includes a lipid bilayer environment as a continuum function of hydrophobic interaction with amino acids. This certainly reduces the CPU time but more sophisticated representation of the bilayer could be beneficial. As the reduced representation method has been employed for helical segments in the TMH simulation method, a simple physical representation of the lipids would be a possibility. One such representation has been reported (Baumgärtner 1996). The effect of lipid membrane was studied by Monte Carlo simulation of helix insertion. The membrane was represented by a monolayer of hard cylinders. The cylinders were allowed lateral movement only and the hydrophobic region was incorporated in the cylinders to mimic the hydrophobic region of the lipid bilayer membrane.

Conclusion

A series of BR simulations (TMH and SA/MD simulations) from the sequence analysis to the final generation of a tertiary structure has been completed. The analyses showed that the structure generated at the end of the simulations plausibly resembled the EM-determined structure. The potential terms included in the potential function have been successfully optimized. As the testing case, the methods employed so far have been successful in predicting the BR helix bundle; the same method could be applied to unknown systems with the same parameter set.

Acknowledgements This work was supported by a grant from the Wellcome Trust. Our thanks to the Oxford Centre for Molecular Science for the use of computational facilities.

References

Arseniev AS, Maslennikov IV, Bytrov VF, Kozhich AT, Ivanov VT, Ovchinnikov YA (1988) Two-dimensional ^{1}H-NMR study of bacterioopsin-(34-65)-polypeptide conformation. FEBS Lett 231: 81–88

Barlow DJ, Thornton JM (1988) Helix geometry in proteins. J Mol Biol 201: 601–619

Barsukov L, Nolde DE, Lomiz AL, Arseniev AS (1992) Three-dimensional structure of proteolytic fragment 163–231 of bac-

terioopsin determined from nuclear magnetic resonance data in solution. Eur J Biochem 206: 665–672
Barton GJ (1990) Protein multiple sequence alignment and flexible pattern matching. Methods Enzymol 183: 403–428
Baumgärtner A (1996) Insertion and hairpin formation of membrane proteins: a Monte Carlo study. Biophys J 71: 1248–1255
Blaurock AE, Stoeckenius W (1971) Structure of the purple membrane. Nature 233: 152–155
Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: a program for macromolecular energy minimization, and dynamics calculations. J Comput Chem 4: 187–217
Brünger AT (1992) X-PLOR version 3.0. System for X-ray crystallography and NMR. Yale University Press, New Haven
Brünger AT, Krukowski A, Erickson JW (1990) Slow-cooling protocols for crystallographic refinement by simulated annealing. Acta Crystallogr A 46: 585–593
Chothia C, Levitt M, Richardson D (1981) Helix to helix packing in proteins. J Mol Biol 145: 215–250
Donnelly D, Overington JP, Ruffle SV, Nugent JHA, Blundell TL (1993) Modelling α-helical transmembrane domains: the calculation and use of substitution tables for lipid-facing residues. Protein Sci 2: 55–70
Edholm O, Jähnig F (1988) The structure of a membrane-spanning polypeptide studied by molecular dynamics. Biophys Chem 30: 279–292
Engelman DM, Steitz TA, Goldman TA (1986) Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu Rev Biophys Chem 15: 321–353
Grigorieff N, Ceska TA, Downing KH, Baldwin JM, Henderson R (1996) Electron-crystallographic refinement of the structure of bacteriorhodopsin. J Mol Biol 259: 393–421
Henderson R, Unwin PNT (1975) Three-dimensional model of purple membrane obtained by electron microscopy. Nature 257: 28–32
Henderson R, Baldwin JM, Ceska TA, Zemlin F, Beckmann E, Downing KH (1990) Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. J Mol Biol 213: 899–929
Herzyk P, Hubbard RE (1995) Automated method for modelling seven-helix transmembrane receptors from experimental data. Biophys J 69: 2419–2442
Hubbell WL, Altenbach C (1994) Investigation of structure and dynamics in membrane proteins using side-directed spin labelling. Curr Opin Struct Biol 4: 566–573
Jähnig F, Edholm O (1992) Modelling of the structure of bacteriorhodopsin, a molecular dynamics study. J Mol Biol 226: 837–850
Kahn TW, Engelman DM (1992) Bacteriorhodopsin can be refolded from two independently stable transmembrane helices and the complementary five-helix fragment. Biochemistry 31: 6144–6151
Karplus M, Pesko GA (1990) Molecular dynamics simulations biology. Nature 347: 631–639
Kirkpatrick S, Gellatt CD Jr, Vecchi MP (1983) Optimization by simulated annealing. Science 220: 671–680
Kono M, Yu H, Oprian DD (1996) Interactions between helix 5 and helix 6 of rhodopsin using cysteine, scanning mutagenesis of spilt receptor mutants. Biophys J Abstr 70: 395: W-Pos94
Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH (1953) Equation of state calculations by fast computing machines. J Chem Phys 21: 1087–1092
Oesterhelt D, Stoecknius W (1973) Functions of a new photoreceptor membrane. Proc Natl Acad Sci USA 70: 2853–2857
Orekhov VYu, Konstantine VP, Arseniev AS (1994) Backbone dynamics of (1-71) bacterioopsin studied by two-dimensional ^{1}H-^{15}N NMR spectroscopy. Eur J Biochem 219: 887–896
Persson B, Argos P (1994) Prediction of transmembrane segments in proteins utilising multiple sequence alignments. J Mol Biol 237: 182–192
Pervushin KV, Arseniev AS (1992) Three-dimensional structure of (1-36) bacterioopsin in methanol-chloroform mixture and SDS micelles determined by 2D ^{1}H-NMR spectroscopy. FEBS Lett 308: 190–196
Popot J-L, Engelman DM (1990) Membrane protein folding and oligomerization: the two-stage model. Biochemistry 29: 4031–4037
Rees DC, Deantonio L, Eisenberg D (1989) Hydrophobic organization of membrane proteins. Science 245: 510–513
Sankararamakrishnan R, Vishveshwara S (1992) Geometry of proline-containing alpha-helices in proteins. Int J Pep Protein Rev 39: 356–363
Sansom MSP, Son HS, Sankararamakrishnan R, Kerr ID, Breed J (1995) Seven-helix bundles: molecular modelling via restrained molecular dynamics. Biophys J 68: 1295–1310
Sayle R (1996) RasMol: molecular visualisation program
Son HS, Sansom MSP (1999) Simulation of the packing of idealized transmembrane α-helix bundles. Eur Biophys J 28: 489–498
Stoeckenius W, Rawen R (1967) A morphological study of *Halobacterium halobium* and its lysis in media of low salt concentration. J Cell Biol 34: 365–393
Yeates TO, Komiya H, Rees DC, Allen JP, Feher G (1987) Structure of the reaction center from *Rhodobacter spaeroides* R-26: membrane-protein interactions. Proc Natl Acad Sci USA 84: 6438–6442