Eur Biophys J (1999) 28: 489–498

ARTICLE

Hyeon S. Son · Mark S.P. Sansom

Simulation of the packing of idealized transmembrane α-helix bundles

Received: 16 November 1998 / Revised version: 26 March 1999 / Accepted: 8 April 1999

Abstract The aim of this study is to investigate if the packing motifs of native transmembrane helices can be produced by simulations with simple potentials and to develop a method for the rapid generation of initial candidate models for integral membrane proteins composed of bundles of transmembrane helices. Constituent residues are mapped along the helix axis in order to maintain the amino acid sequence-dependent properties of the helix. Helix packing is optimized according to a semi-empirical potential mainly composed of four components: a bilayer potential, a crossing angle potential, a helix dipole potential and a helix-helix distance potential. A Monte Carlo simulated annealing protocol is employed to optimize the helix bundle system. Necessary parameters are derived from theoretical studies and statistical analysis of experimentally determined protein structures. Preliminary testing of the method has been conducted with idealized seven Ala_{20} helix bundles. The structures generated show a high degree of compactness. It was observed that both bacteriorhodopsin-like and δ-endotoxin-like structures are generated in seven-helix bundle simulations, within which the composition varies dependent upon the cooling rate. The simulation method has also been employed to explore the packing of $N = 4$ and $N = 12$ transmembrane helix bundles. The results suggest that seven and 12 transmembrane helix bundles resembling those observed experimentally (e.g., bacteriorhodopsin, rhodopsin and cytochrome *c* oxidase subunit I) may be generated by simulations using simple potentials.

H.S. Son (✉)[1] · M.S.P. Sansom
Laboratory of Molecular Biophysics, Rex Richards Building, University of Oxford, Oxford OX1 3QU, UK

Present address:
[1] CSM, 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

Key words Monte Carlo simulated annealing · Simulation · Membrane helix packing

Introduction

Integral membrane proteins (IMPs) are important to cells in functions such as transport, energy transduction and signalling. Three-dimensional molecular structures of proteins at the atomic level are needed to understand such processes. Prediction of membrane structure is necessary, especially because there are only a small number of membrane protein structures determined in atomic resolution. The existence of the lipid bilayer provides a powerful constraint on the methods employed to predict the structure; the two-dimensional nature of the lipid bilayer provides a constraint on the arrangement of the membrane spanning segments of IMPs (Donnelly et al. 1993). The α-helical or β-sheet conformation is favoured in a lipid bilayer environment because of the hydrophobic interior of the lipid bilayer, because unsatisfied hydrogen bonding within the bilayer is energetically costly (+5 kcal/mol per H-bond, Engelman et al. 1986). It is feasible to investigate methods for predicting IMP structure using the constraints by the lipid environment and by experiments such as the spin labelling experiment (Hubbell and Altenbach 1994), even though the experimental database of such structures is small. Only α-helical secondary structures are considered in this study. Methods developed to predict the transmembrane α helical region from its sequence (Kyte and Doolittle 1982; Eisenberg 1984; Engelman et al. 1986) would be useful if the α-helices are independently stable. The two-stage folding model proposed by Popot and Engelman (1990; Popot 1993) enables individual helices to be treated as independent folding units.

Simulated annealing via molecular dynamics (SA/MD) has been used successfully to refine structures from experiments and to model protein structures (Nilges and

Brünger 1991). SA/MD and molecular dynamics (MD) methods were used to study the behaviour of bacteriorhodopsin (BR) transmembrane helical segments in the lipid bilayer environment (Son 1997). The results from these studies suggest that these methods could be used for helix packing simulation. However, unlike the systems used in the studies of monomers and dimers, real systems contains several helical segments, in which determination of the initial structures is much more complicated. Furthermore, intuition to guess the initial structure is not plausible in this complicated case as it would introduce human prejudice towards the structure. Therefore a systematic method has to be developed to construct reasonable initial structure(s).

The aim of this study is (1) to see if the packing motifs of native transmembrane helices can be produced by simulations with simple potentials, and, if so, (2) to enable rapid generation of initial candidate model structures for bundles of transmembrane helices. Such models may then be further refined by, for example, SA/MD simulations. In this paper, the method of generating the initial structure(s) and some testing results will be discussed. Ala_{20} transmembrane segments will be used to verify the method and to develop a simulation protocol. The helix packing motifs generated by simulations of systems of bundles containing 4-, 7- and 12-Ala_{20} helices and simulation results of idealized BR and cytochrome *c* oxidase segments are also discussed.

Methods

The simulated annealing via Monte Carlo (SA/MC) protocol is used to study the packing of the helices to form a bundle. In general terms, a Monte Carlo search is a random selection procedure. The Metropolis method (Metropolis prescription; Metropolis et al. 1953) is employed to adopt the concept of SA/MC. The Metropolis prescription allows both downhill and uphill moves to overcome the local energy barrier according to the Boltzmann probability. A FORTRAN program TMH has been developed to implement the method described. Helix packing is optimized according to a simple semi-empirical potential composed of four components: a helix-helix distance potential, a crossing angle potential, a helix dipole potential and a bilayer potential. The distance potential, the helix dipole potential and the crossing angle potential represent the pairwise helix/helix interaction energy. The bilayer potential represents the sequence-dependent helix/bilayer interaction energy. Each helix is treated as a rigid body represented by a direction vector, the coordinates of its centre of mass and the coordinates of the Cα atoms. The two-stage folding model (Popot and Engelman 1990; Popot 1993) enables such a representation of helices. Also it has been reported that the BR helical segments, when SA/MD- and MD-simulated, stayed as intact helices during the simulation, i.e., they behaved as somewhat rigid objects (Sansom et al. 1995; Lakshmanan and Saraswathi 1996). Constituent residues are mapped onto the corresponding Cα atoms in order to maintain the amino acid sequence-dependent properties of the helix. This treatment of helices is convenient as it allows the reduced representation of helices, which significantly reduces the calculation time.

Simulated annealing/Monte Carlo simulation

SA/MC is employed to minimize the energy of the system. The Metropolis prescription is used to determine whether or not to accept each step of the selection:

$$q_i = \begin{cases} q_{i+1} & \text{if } \xi \le P \\ q_i & \text{if } \xi > P \end{cases} \quad (1)$$

where q_i is the structure selected by the ith step search, $P = \exp(-\Delta E(q)/kT)$, $\Delta E = E(q_{i+1}) - \mathrm{E}(q_i)$, $E(q_i)$ is the total energy of the system of structure q_i, k is the Boltzmann constant, T is the absolute temperature of the system and ξ is a random number uniformly distributed in the interval (0, 1). As shown in Eq. (1), the structure will be always accepted if the structure from the previous step has higher potential energy, because P in this case is always greater than 1.0 which is the maximum value of the random number. If the previous structure has lower energy, then it will be determined according to the Boltzmann probability P; comparing the probability with a random number, it decides whether or not to accept the new structure. For each temperature, N steps of conformational search are involved. The search space and N are given prior to the simulation. All the helical segments are confined within a simulation box whose dimensions are defined prior to the simulation. If a new position is selected outside the simulation box, the position is discarded and the previous valid position is held. The simulation is done with translational and rotational selections; the limitation of the simulation box is applied only to the centre of mass of individual helical segments, i.e., the N- or C-terminus of a helix is allowed to be outside the simulation box. Translational and a rotational stepsizes are specified prior to the simulations. The direction of the move is not biased nor restricted inside the simulation box. As long as the condition is fulfilled according to the Metropolis selection procedure, a move is always accepted.

Potentials

Helix packing is optimized by the SA/MC method employing a semi-empirical potential composed of four terms:

$$V_\mathrm{T} = w_\mathrm{d} V_\mathrm{d} + w_\mathrm{b} V_\mathrm{b} + w_\Omega V_\Omega + w_\mu V_\mu \quad (2)$$

where V_T is the total potential of the system, V_d is a potential function dependent on the inter-helix 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 and V_μ is the dipole potential; w_d, w_b, w_Ω and w_μ are the weights applied to each energy term. The parameters for V_b are from the GES hydrophobicity scale (Engelman et al. 1986); for V_μ, Eqs. (11) and (12) are used (see below). The V_d and V_Ω parameters are from the analysis of the structures in the Brookhaven protein databank (PDB). Given that the precise form and the absolute magnitude of the terms in the effective potential are not known, it is important to investigate whether the final structure is sensitive or not to small variations in the various weights, i.e., w_d, w_b, w_Ω and w_μ. We have accordingly performed some simple tests and found that the exact value of the weights are not important provided that they are chosen such that the typical magnitudes of the four terms in the potential are approximately the same.

Helix-helix distance potential

Transmembrane proteins consisting of more than two helices pack together to form a bundle structure. Secondary structures and their hydrophobic surfaces formed by sequences are known to be the primary determinants of the folding, i.e., packing, patterns (Chothia et al. 1981). A helix-helix distance potential is employed mainly to mimic the helix-helix packing interaction. A (12, 6) Lennard-Jones potential is used in order to provide a repulsive potential as well as an attractive potential:

$$E_d(r) = -4E_m\left\{\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right\} \tag{3}$$

$$V_d = \sum_{l=1}^{N-1}\sum_{m=l+1}^{N} E_d(r) \tag{4}$$

where E_d is the pair-wise helix-helix distance potential, E_m is the minimum energy value when the distance is $2^{1/6}\sigma$ and r is the inter-helical distance (the closest distance between two helices). The total helix-helix distance potential V_d is defined as the summation over all the possible helix pairs. N is the total number of helices, and l and m are the helix indices.

Bilayer potential

The lipid bilayer has two hydrophilic surfaces and a hydrophobic region between the surfaces. Edholm and Jähnig (1988; Jähnig and Edholm 1992) introduced a U-shaped potential into the system of bilayers to mimic the hydrophobic effect experienced by the amino acids of the transmembrane segments (Fig. 1). The same potential form is used except that the solvent exposed surface effect is not considered in this study. The potential is defined as:

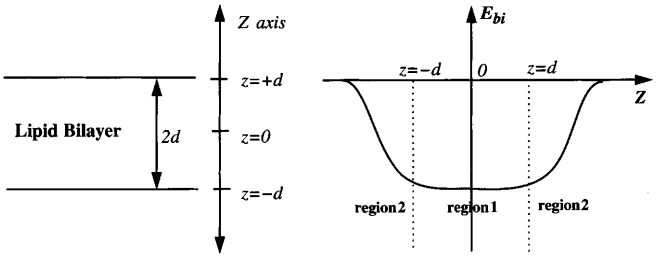

Fig. 1 The bilayer potential across the lipid membrane, where *2d* is the thickness of the membrane. The region 2 shows a transition between two different environments (the hydrophilic and the hydrophobic environments)

$$E_{bi} = h_i f(z_i) \tag{5}$$

where E_{bi} is the hydrophobic potential energy experienced by the ith amino acid and h_i is the residue hydrophobicity of the amino acid concerned derived from a suitable scale. $f(z_i)$ is defined as follows:

$$f(z_i) = \begin{cases} \frac{1}{2}e^{-(|z_i|-d)/\lambda} & \text{for } |z_i| \geq d \\ \frac{1}{2}(2 - e^{(|z_i|-d)/\lambda}) & \text{for } |z_i| < d \end{cases} \tag{6}$$

where z_i is the z coordinate (in this simulation system, the z-axis is parallel to the bilayer normal) of the residue concerned, $2d$ is the thickness of the bilayer and h_i is the hydrophobicity of the residue concerned. λ is a coefficient governing the shape of the curve in the region corresponding to the transition between hydrophilic and hydrophobic environments; $\lambda = 2.0$ Å is used to provide a smooth transition in all simulations described in this paper. Each segment has associated amino acids and, according to their z coordinates, the potential is calculated. The bilayer potential energy (E_{bl}) of a helix l is the summation over all the constituent residues in the helix:

$$E_{bl} = \sum_{i=1}^{n} E_{bi} \tag{7}$$

where the summation is over the n residues of a helix. The total bilayer potential V_b is the summation over all the helices:

$$V_b = \sum_{l=1}^{N} E_{bl} \tag{8}$$

where N is the number of helices in the system.

Crossing angle potential

The helix-helix crossing angle (Ω) is the angle between a pair of helices (Chothia et al. 1981). As shown in Fig. 2, the crossing angle potential for each pair of helices forms a quadratic function in region 2, and it has zero values in regions 1 and 3. This potential term is valid only for the helices within a cutoff distance:

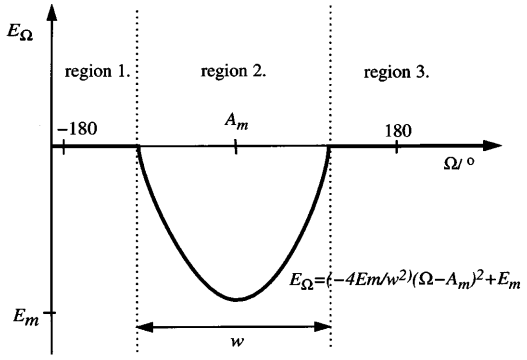

Fig. 2 The crossing angle potential. E_Ω is quadratic in region 2 and zero potential in other regions. By definition, the crossing angle varies within a range between –180° and +180°

$$E_\Omega(\Omega) = (-4E_\mathrm{m}/w^2)(\Omega - A_\mathrm{m})^2 + E_\mathrm{m} \quad (9)$$

$$V_\Omega = \sum_{l=1}^{N-1} \sum_{m=l+1}^{N} E_\Omega(\Omega) \quad (10)$$

where the potential is at its minimum (E_m) when the crossing angle (Ω) is at A_m, which is the angle for the minimum potential; w is the width of the potential; E_m works for negative values only; E_Ω takes effect only for those helix pairs within a cutoff range; V_Ω is the total crossing angle potential; N is the total number of helices. Summation is over all possible helix pairs, where l and m are the helix indices.

The quadratic form of the potential comes from the statistical treatment of crossing angle data sampled from the α and the $\alpha + \beta$ class protein structures deposited in the Brookhaven PDB (Son 1997). The statistical analysis showed that the observed crossing angles formed a normal distribution pattern. Conversion of the distribution to a potential form was done using Boltzmann's equation.

Helix dipole potential

This potential is introduced to mimic the helix dipole effect. The α-helix dipole is due to the aligned dipoles of the single peptide units. It has been shown that the effect of the helix dipole is equivalent to the effect of half a positive unit charge at the N-terminus of the helix and half a negative charge at the C-terminus (Hol et al. 1978, 1981). The dielectric potential is expressed by following equation:

$$E_\mu(\mu_l, \mu_m, \alpha, \beta, \gamma, r) = \frac{\mu_l \mu_m}{4\pi\epsilon_0\epsilon_r} \frac{(3\cos\beta\cos\gamma - \cos\alpha)}{|r|^3} \quad (11)$$

$$V_\mu = \sum_{l=1}^{N-1} \sum_{m=l+1}^{N} E_\mu(\mu_l, \mu_m, \alpha, \beta, \gamma, r) \quad (12)$$

where E_μ is the dipole energy, μ_1 and μ_2 are helix dipoles, ϵ_0 is the permittivity of a vacuum, ϵ_r is the dielectric constant of the medium, α, β and γ are angles and $r = |\underline{r}|$, the centre-to-centre distance between helix l and helix m (illustrated in Fig. 3). V_μ is the total helix dipole potential; summation is over all the possible pair-wise combinations of helical segments; l and m are the helix indices; N is the total number of helices (Hol et al. 1978, 1981).

Simulation protocol

One of the main objectives of this study is to see whether the transmembrane segments pack together according to the parameters using simplified potentials. The validity of the reduced representation method employed in the TMH simulation should be investigated in order to generate the initial model structures for further SA/MD refinement. SA/MD-simulated 7-Ala_{20} helix bundles were reported to be similar to those experimentally determined (Sansom et al. 1995). Therefore it is also interesting to see if the TMH simulation could generate structures similar to the SA/MD-produced ones.

Initial models for the TMH simulation were generated using the "helix building" module of TMH. For each helix, only Cα atoms were used to build an idealized helix structure (3.6 residues per turn and 1.5 Å rise per residue along the helix axis). The bundle system was first heated to a very high temperature (90 000 K) in order to randomize the initial structure; then the system was cooled abruptly to 7000 K and then slowly cooled to 300 K. This process was repeated to generate ensembles of 50 structures. From 7000 K to 300 K, 50 K was taken to be a step temperature unit, and for each step temperature, 600 times of sampling was done (300 times of translational sampling and 300 times of rotational). The parameters were set to generate the minimized structures having an inter-helical distance (the parameter was from

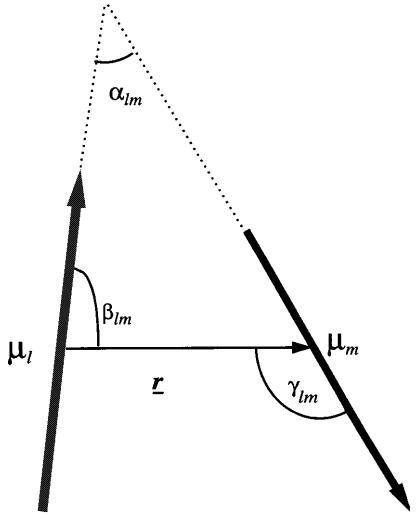

Fig. 3 The helix-helix dipole: r is the vector joining the centers of helices l and m

the analysis of PDB protein structures) of 8.8 Å, and crossing angles of 14° for parallel and −173° for antiparallel pair of helices [i.e., $A_m = 14°$ for parallel and −173° for antiparallel pair of helices (Sansom et al. 1995)]. The width of the potential represents the standard deviation of the sample set from which the parameters were deduced ($w = 16$ for parallel and $w = 24$ for anti-parallel pairs of helices). The GES scale (Engelman et al. 1986) was used to assign a hydrophobic value (h_i) to each amino acid; λ was assigned to be 2.0 Å (Jänig and Edholm 1992); ϵ_r was set to be 2.0 to mimic the lipid bilayer medium. The weights were adjusted so that the minimum energy values of constituent potentials could be −10.0 kcal/mol. Therefore every energy term was set to contribute equally towards the total energy when it was at its optimum. The simulation protocol has been described only generally. Different sets of parameters were used in some simulations. In these cases, the protocol and the parameter values are described in detail in relevant sections.

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, a CPU time of approximately 10 min was required for the 4-Ala_{20} simulation, 30 min for the 7-Ala_{20} simulation, 45 min for the 12-Ala_{20} simulation, 35 min for the BR simulation and 50 min for the simulation of cytochrome *c* oxidase.

Sampling method

In this Monte Carlo simulation there are two different types of sampling procedures: one is translational and the other is rotational. Combining these two is essential, though CPU intensive. If a helix is allowed to move freely in a box of dimensions of 10 Å × 10 Å × 10 Å, and the stepsize is fixed to be 1.0 Å, it has 1000 different locations and each location has 360° × 360° rotational spaces if the rotational stepsize is fixed to be 1.0°. In this situation it is difficult to optimize both properties at the same time. To overcome this problem a "step selection method" (as opposed to a "whole selection method") was used. This is done by performing the two sampling steps one after the other repeatedly, and doing the Metropolis selection after each sampling step. This procedure is repeated over a given number of sampling frequencies to ensure that the final structure is properly optimized. In real terms, helices are packed according to the packing potentials (e.g., the helix-helix distance potential, the bilayer and the dipole potentials) and then the tilts of the helices are optimized by the crossing angle, the dipole and the bilayer potentials. To evaluate the methods, Fig. 4 is provided. In Fig. 4, the energy (total potential energy only) trajectories from 7-Ala_{20} simulations using two different selection procedures are shown. The "step selection" energy trajectory shows earlier convergence of total energy than the "whole selection" trajectory, and also shows many more uphill energy jumps in the high temperature region. Further, the "step selection" results in finding lower energy structures than the "whole selection". In other words, in the whole selection procedure, much of the sampling movement is wasted because of the mismatch of two types of sampling procedures. It is also observed that the "step selection" method shows a higher rate of acceptance than the "whole selection" method in the high temperature region. Therefore, the "step selection" method is more effective than the "whole selection" method in selecting lower energy structures in the TMH simulation. From this result it was decided to employ the "step selection" protocol for all TMH simulations.

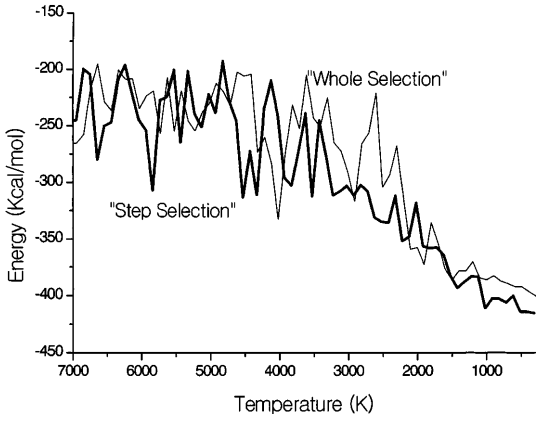

Fig. 4 Energy trajectories for two different "selection procedures": the *thick solid line* represents the "step selection" procedure, the *thin line* the "whole selection"

Results

Simulations on Ala_{20} helix bundles

Idealized Ala_{20} helix bundles consisting of 4, 7 and 12 helices were simulated. The 4-Ala_{20} helix bundle is a simple testing system to investigate various aspects of helix packing. 7-Ala_{20} and 12-Ala_{20} are convenient simulation systems to verify the methods because BR and cytochrome *c* oxidase subunit I have seven and 12 transmembrane helical segments, respectively. From the simulations with bundles containing 4, 7 and 12 Ala_{20} helices, the lowest energy structures from each ensemble are shown in Fig. 5. To analyse the bundles generated, three properties were investigated:

1. The orientation and the location of the bundle with respect to the lipid bilayer.
2. The compactness of the bundle.
3. The relative orientation of constituent helices.

In particular, it has been investigated whether favoured low energy structures emerge as a result of the helix packing considerations which have been described. The lowest energy structures from ensembles were subjected

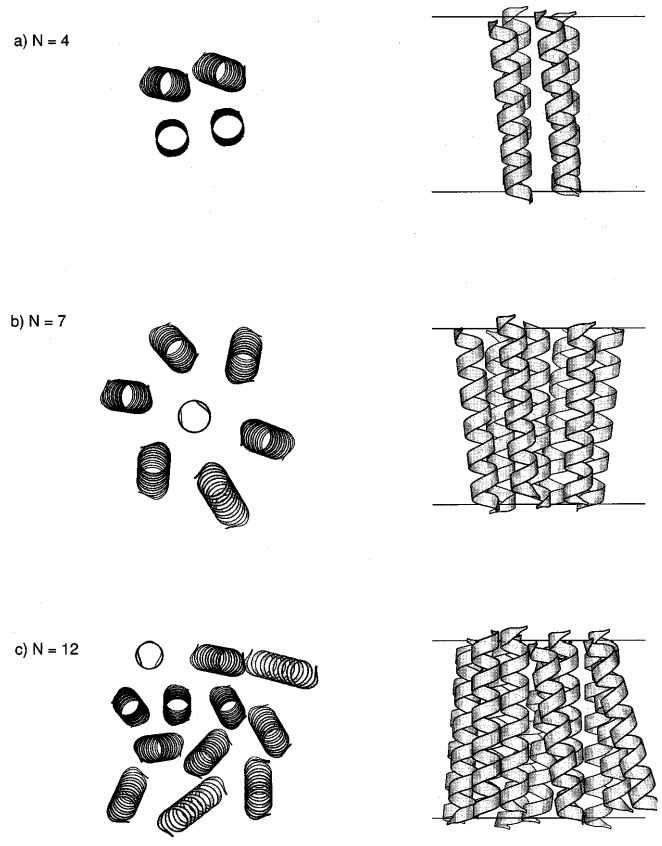

Fig. 5 Simulated poly-Ala_{20} bundles containing **a** 4, **b** 7 and **c** 12 helices. The lowest energy structures from each ensemble are shown

to geometrical analyses [a FORTRAN program ACTIVE (H. S. Son and M. S. P. Sansom, unpublished) was used] and the results are shown in Table 1. The crossing angles and the helix-helix distances were optimized successfully for "ridges into grooves" packing of helices. The tilt angle (ζ) is defined as the angle between the bilayer normal and the helix axis. The tilt angles (ζ) show clearly that the helix-bilayer interaction energy holds the helix bundle approximately perpendicular to the plane of the lipid bilayer. The presence of the lipid bilayer and the successful implementation of the bilayer potential implies that the helix packing is essentially a two-dimensional optimization problem. If helices are to be optimized in two-dimensional space in terms of the distances between the helices, the best optimized structures should have the maximum number of contacts allowed for a particular configuration, e.g., a trimer (3-Ala_{20}) bundle has a maximum of three helix-helix contacts in a triangular conformation, while a seven-helix bundle has maximum of 12 helix-helix contacts. The compactness of a helix bundle can be defined in terms of the number of contacts of neighbouring helices in the bundle. To determine whether two helices are in contact or not, the closest distance was calculated and this value was compared to a threshold value. The results of this theoretical consideration have been found to be true in the simulation results presented in Fig. 5. All the structures observed are compact according to this criteria.

Figure 6 shows the energy trajectory for the lowest energy structure of the 7-Ala_{20} ensemble. At high temperature, i.e., $T > 3000$ K, the energy of the selected structure fluctuates as the temperature changes, indicating that up and down transitions are occurring. However, at low temperature, i.e., $T < 2000$ K, only the structures having lower energy are selected, which shows "simulated annealing" works effectively.

Cluster analysis is useful to investigate the potential surface of the system. It is also interesting to see whether the generated structures represent different local energy minima. Clustering by energy values was considered

Table 1 Geometrical analyses of TMH simulated transmembrane helix bundles of Ala_{20}, BR and cytochrome *c* oxidase subunit I. The geometrical properties for the lowest energy structures from each ensemble are summarized. Ω is the crossing angle between two helices, d_c is the closest distance between two helices and ζ is the tilt angle (the angle between a helix and the bilayer normal) of the structure

	Ω (°) for parallel pair	Ω (°) for anti-parallel pair	d_c (Å)	ζ (°)
4-Ala_{20}	7.3 (0.3)	−175.9 (3.3)	9.1 (1.1)	5.8 (2.2)
7-Ala_{20}	14.3 (3.9)	−168.4 (5.1)	8.8 (0.3)	8.4 (3.3)
12-Ala_{20}	13.4 (2.4)	−171.5 (2.8)	8.7 (1.1)	11.1 (4.9)
BR	10.5 (0.6)	−171.1 (3.0)	8.7 (0.2)	8.7 (3.6)
Cytochrome *c* oxidase SUI	13.4 (8.6)	−168.6 (4.9)	7.9 (1.7)	11.1 (3.8)

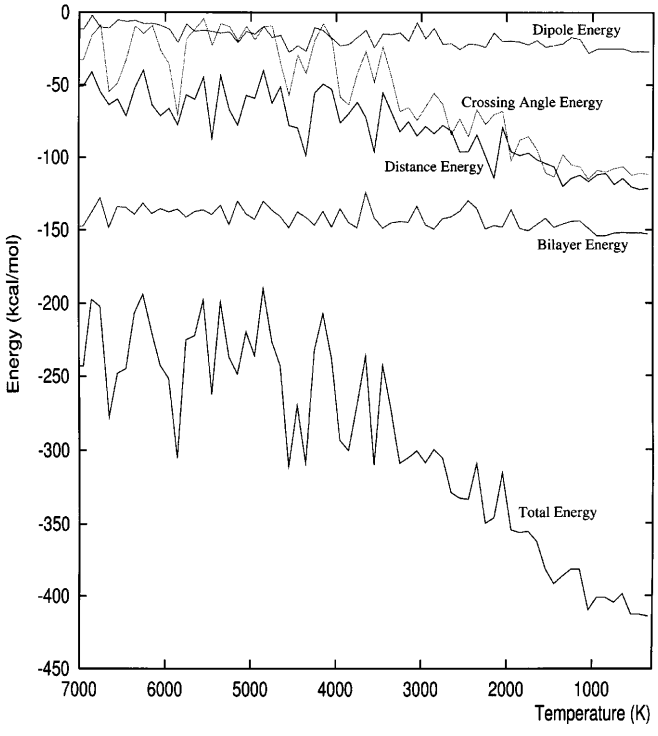

Fig. 6 Energy trajectory for the lowest energy structure of the 7-Ala_{20} ensemble

initially. However, it was found to be inappropriate since clear clusters of distinguishable structures were not found with this method. This means that the energy gap separating different structures is not wide enough to allow visible clustering. Consequently, a structure clustering was considered instead. 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 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 7-Ala_{20} simulation generated clusters of structures: δ-endotoxin-like, BR-like, along with other structures. The occurrence of structural clustering was confirmed by visual inspection of the structures. This indicates that those two naturally observed structures exist as low energy states.

The effect of the cooling rate on the frequency of occurrence of different energy level structures has also been investigated. The terms "slow cooling" and "fast cooling" are used as relative definitions between cooling protocols. Altering the cooling rates can be achieved by changing the sampling frequencies or the step temperature, or both. The same protocol previously described was used except that this time the rate of cooling was set to be faster than before (the sampling frequency was one third of the one used in the previous 7-Ala_{20} simulation). Analysis shows that the newly generated ensemble of structures contains more higher energy structures. The number of δ-endotoxin-like structures found in this fast cooling simulation is far less than in the slow cooling simulation. Because the δ-endotoxin-like structure has the lowest energy conformation in the 7-Ala_{20} system, it is not surprising to observe more of the δ-endotoxin-like structures when cooled slowly because a wider conformational search provided by the slow cooling rate results in finding more low energy structures. Generally, a sufficient number of samplings is necessary to search for wide conformation space because the Monte Carlo method relies on random sampling. Also, it was suggested that slow cooling would be important in crystallographic refinement by simulated annealing (Brünger et al. 1990). Figure 7 shows the difference between two cooling rates in terms of energy values during the simulations. Average values of total energy of the system were calculated across the ensemble of 50 structures and plotted against cooling temperature for each simulation. The graph shows the intermediate energy values with standard deviations. It is clearly shown that the final mean value of the total energy for the slowly cooled ensemble is lower than the fast cooled one. More importantly, the slow cooling energy trajectory shows more active up and down moves into the high temperature region ($T > 4000$ K) than the fast cooling one. In other words, simulated annealing is done more effectively with the slow cooling protocol than the other, i.e., the slow cooling searches for a wider range of conformational space, which consequently finds more lower energy structures.

A helix dipole moment is calculated assuming a half charge at the N-terminus (positive charge) of the helix and at the C-terminus (negative charge) (Hol et al. 1981). The dielectric constant was set as $\epsilon = 2$ for lipid bilayer media. A helix-helix dipole moment potential was introduced to generate anti-parallel-ordered (i.e., neighbouring helices are in anti-parallel orientation) helix bundles as low energy structures. 4-Ala_{20} simula-

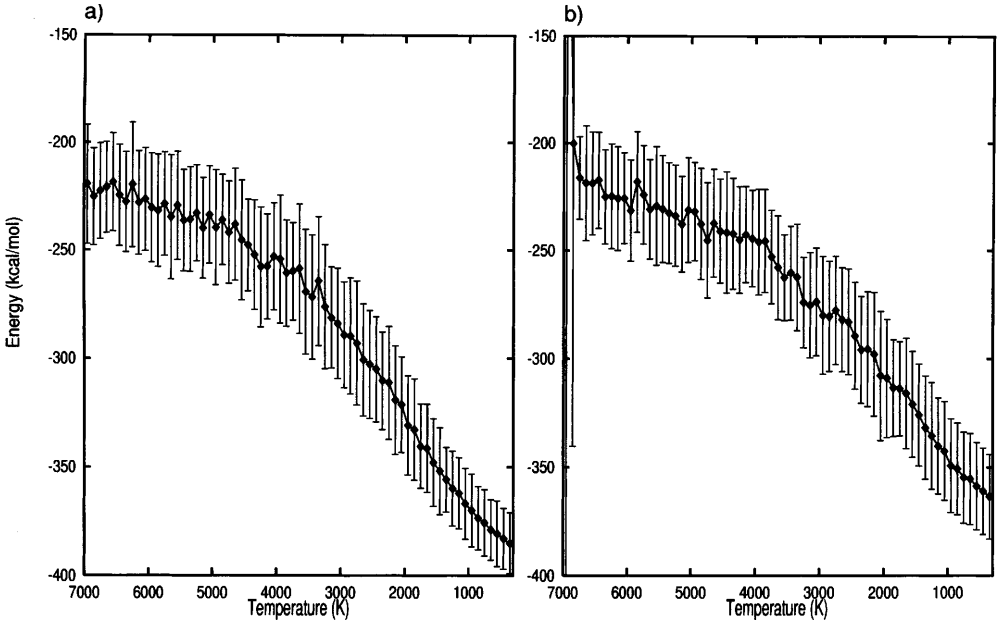

Fig. 7 Energy trajectories of **a** slow and **b** fast cooling 7-Ala_{20} TMH simulations. Plotted energy values are ensemble averages of 50 structures. *Error bars* represent the standard deviation

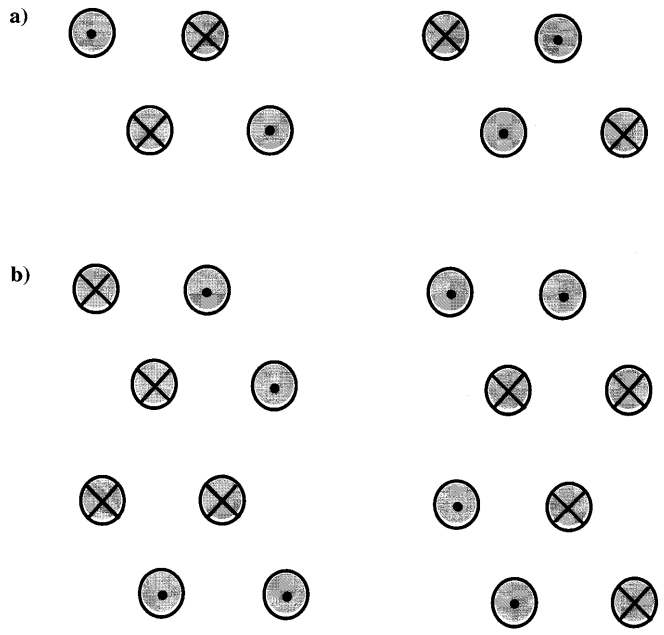

Fig. 8a, b Configurations from the 4-Ala_{20} simulation. The *arrowhead* represents the direction of the helices out of the membrane surface. **a** Anti-parallel-ordered, **b** less-ordered configurations in terms of helix-helix orientation

tion is best suited to investigate the effect of the helix dipole potential. Figure 8 shows six possible configurations from the 4-Ala_{20} TMH simulation. The group A (two structures in Fig. 8a) configuration has four anti-parallel contacts and one parallel contact. The group B (four structures in Fig. 8b) configuration has three anti-parallel contacts and two parallel contacts. In both groups, the total number of contacts is five. If the dipole energy term effectively works, then the group A configuration should dominate the ensemble of structures because of its dominant anti-parallel contacts. Two 4-Ala_{20} simulations were conducted, with and without the dipole potential term. For each simulation, the number of occurrences of each configuration was counted. In the absence of the dipole potential, 20% of structures were of the group A type, while in the presence of the dipole potential, 58% were of the group A type. In other words, simulation with the helix-helix dipole potential term effectively generated anti-parallel-ordered helical segments as desired. This suggests that helix packing is mediated by the dipole moment of the helices, as well as other factors. It is also interesting to note that the anti-parallel-order helix bundle was suggested to be the most favourable four-helix bundle conformation, because of the dielectric effect of the electrostatic interactions, from statistical analysis of four-helix bundles in the PDB (Presnell and Cohen 1989).

Simulation of transmembrane BR helices

The main purpose of this simulation is to evaluate the method by comparing the resultant structure to the known experimentally determined structure. BR-like structures were observed in the 7-Ala_{20} bundle simulation. The sequence of BR transmembrane helices was used to build seven idealized transmembrane helices (only Cα atoms were used). This helix bundle was subjected to a TMH simulation to see if the sequence difference affects the simulation results compared to the 7-Ala_{20} bundle simulation results. It should be noted that the sequence difference affected only the bilayer potential (the hydrophobicity related potential) term. Fig. 9 shows three distinctive structures generated from the simulation. The δ-endotoxin-like structure (Fig. 9a) is the lowest energy clustered structure, possessing lower energy than the BR-like structure (Fig. 9b) and the rhodopsin-like structure (Fig. 9c). The δ-endotoxin-like structure was observed as the lowest energy structure, as in the 7-Ala_{20} bundle simulation.

Simulation of transmembrane cytochrome *c* oxidase SU I domain helices

The cytochrome *c* oxidase structure has been determined at 2.8 Å by Iwata et al. (1996) and Tsukihara et al. (1996). Subunit I of cytochrome *c* oxidase has 12 transmembrane helices which make a compact bundle. The cytochrome *c* oxidase sequence was used to simulate a 12-helix bundle structure. The protocol was the same as the one used for 7-Ala_{20} simulations. Since the main idea of the TMH simulation is to generate compact

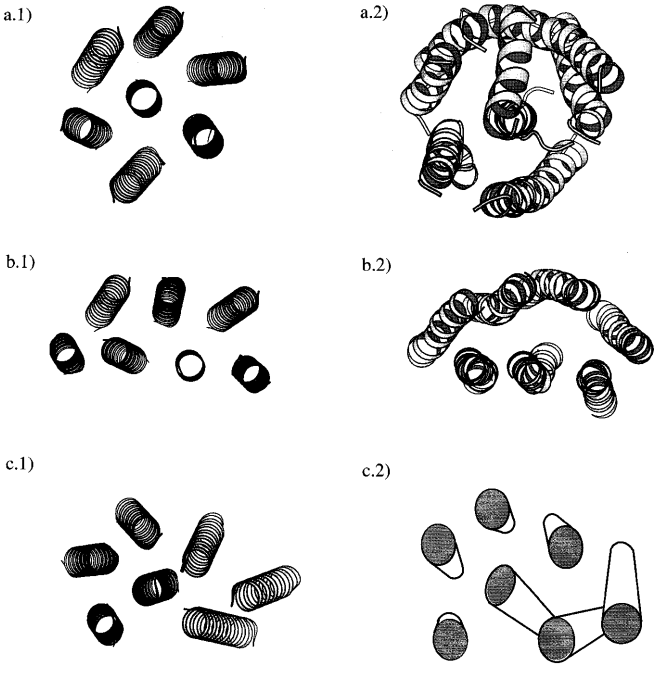

Fig. 9a.1–c.2 Three of the structures generated from the TMH simulations (*left*): **a.1** δ-endotoxin-like, **b.1** BR-like and **c.1** rhodopsin-like structures, and the corresponding structures from experiments (*right*): **a.2** δ-endotoxin (Li et al. 1991), **b.2** BR (Henerson and Unwin 1975; Grigorieff et al. 1996), **c.2** rhodopsin (Schertler et al. 1993; Vinzenz and Schertler 1995) structures

bundles of helices, this simulation should be able to generate cytochrome *c* oxidase-like compact structures as a low energy structure. Figure 10 shows one of the low energy structures from the simulation. The bundle structure is compact and it resembles the theoretically considered lowest energy configuration. Geometrical analyses show that the helices were well optimized (Table 1). Even though this simulation was not expected to generate the native structure of cytochrome *c* oxidase, it produced promising results.

Discussion

This work is intended to generate initial models for further simulation. Priority is given to the rapid generation of structures so that reasonable statistical analyses can be done to select representative structure(s). Consequently, a reduced representation approach was adopted to minimize the calculation time in generating structures. This simplification resulted in neglecting some important features essential in describing native structures, e.g., side chain coordinates, dihedral flexibility, loop regions. Therefore the generated structures would not represent detailed native structures. There were pre-assumptions in this simulation study:

1. All the transmembrane segments are α helical.
2. Helix packing is mediated by the existence of the lipid bilayer.
3. All the parameters are available either from experimental data or from theoretical considerations.
4. Detailed helical properties such as helix kink or side chain coordinates may be left to the next refinement stage using, for example, SA/MD and MD simulations.

Based upon these assumptions, a FORTRAN program TMH has been developed to simulate the packing of transmembrane helix bundles. The algorithm is based on simulated annealing via the Monte Carlo search method by Metropolis prescription. Four simple potential terms were implemented to simulate the helix packing. Simulation protocol and parameters were set prior to the simulations. The protocol employed was established after many simulation trials. To optimize the multi-variant function effectively, a "step selection method" was used. This method showed clear advantage over "whole selection method" as described previously. Poly-Ala_{20} bundles were simulated mainly for testing purposes.

Individual potential functions were verified, and the results clearly showed that the helix-helix distance potential, the bilayer potential and the helix crossing angle potential were successfully implemented. The dipole potential term was intended to pack helices in anti-parallel fashion. A comparison study was done using the results of 4-Ala_{20} simulations with and without the potential. The simulation with the dipole potential generated dominantly anti-parallel-ordered structures. The 7-Ala_{20} simulation produced interesting results. The constraint imposed by the biological membrane reduces the transmembrane helix packing problem to a two-dimensional (2D) one. A δ-endotoxin-like structure is considered to be the lowest energy structure when only 2D conformational space is assumed. The ensemble generated contains the δ-endotoxin-like structures, the BR-like structures and other conformational structures. According to the cooling rate, the numbers of each type of structure varied. Fast cooling simulation generated less δ-endotoxin-like structures than the slow cooling simulation. It is important to notice, in this simulation study, that native structures observed by experiments were also obtained by the simulation with extremely simple potentials. However, in this case the BR-like structure is not the lowest energy structure. Subsequent simulation with real BR sequence showed similar results (this simulation also generated the δ-endotoxin-like structure as the lowest energy conformation and the BR-like structures were also observed), which means that the amino acid sequence did not visibly influence the helix packing in the simulation system, which employed only four potentials. In other words, the existing four potentials were not sufficient to reflect all the amino acid properties and to generate a unique structure from the sequence. On the contrary, the 12-Ala_{20} simulation, and the simulation with the cytochrome *c* oxidase sequence, produced structures resembling native-like structures even with the existing four potentials only. They generated tightly packed structures resembling native cytochrome *c* oxidase. Undoubtedly, a cytochrome *c* oxidase-like structure, i.e., a compact 12-helix bundle, is the lowest energy conformation. These results imply that, for some structures, the lowest energy structure is the native conformation, while it is not true for some others. Therefore another type of potential which could generate the desired configuration according to the amino acid composition is needed. The new potential should be able to generate BR-like structures as the lowest energy conformation when the BR sequence is applied to the simulation system, while the cytochrome *c* oxidase-like structure should be observed as the lowest energy conformation with its own sequence. One possible sequence-specific potential would be a helix-lipid

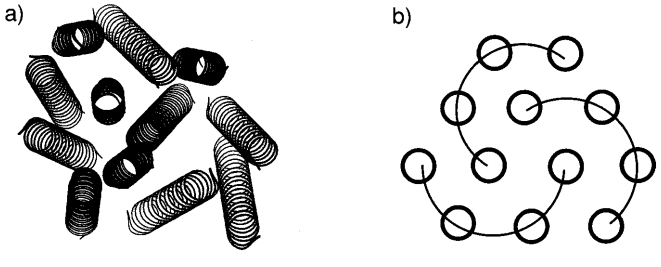

Fig. 10 **a** A low energy structure generated from the TMH simulation with cytochrome *c* oxidase sequence, **b** structural motif suggested by X-ray crystallography (Iwata et al. 1996; Tsukihara et al. 1996)

orientation potential (i.e., which face of the helix will be oriented towards the lipid bilayer to give minimum hydrophilic contact with the lipid).

Having generated an ensemble of structures, determination of the representative structure is the next important step prior to further refinement. In order to select the representative structure, two selection protocols may be considered: one is to obtain the lowest energy structure, the other is to obtain an average structure from the structures in the lowest energy cluster of structures. The representative structure could be subjected to a further refinement using, for example, SA/MD and/or MD.

Conclusion

The SA/MC method adopted in this study produced promising results. The problem of optimizing two different kinds of variable, i.e., translational and rotational, has been solved using a ''step selection method''. Compact transmembrane helix bundles were generated with simple potential functions. The parameters were derived from either statistical analyses or simulation studies. One question is whether global/local energy conformations reflect the native protein conformation. Both δ-endotoxin-like and BR-like structures were observed in the structures from the 7-Ala_{20} simulation, and a cytochrome *c* oxidase-like structure was obtained when 12 transmembrane helices were simulated. The results suggest that native protein structures exist as global/local minimum energy conformations which can be found by simulations with simple potentials. However, further development in selecting representative configurations is necessary. This could be done by introducing another potential function into the simulation system.

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

Brünger AT, Krukowski A, Erickson JW (1990) Slow-cooling protocols for crystallographic refinement by simulated annealing. Acta Crystallogr A46: 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 polypepetide studied by molecular dynamics. Biophys Chem 30: 279–292

Eisenberg D (1984) Three dimensional structure of membrane and surface proteins. Annu Rev Biochem 53: 595–623

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

Hol WGJ, Duijnen PT van, Berendsen HJC (1978) The α-helix dipole and the properties of proteins. Nature 273: 443–446

Hol WGJ, Halie LM, Sander C (1981) Dipoles of the α-helix and β-sheet: their role in protein folding. Nature 294: 532–536

Hubbell WL, Altenbach C (1994) Investigation of structure and dynamics in membrane proteins using site-directed spin labelling. Curr Opin Struct Biol 4: 566–573

Iwata S, Östermeier C, Ludwig B, Michel H (1996) Structure at 2.8 Å resolution of cytochrome *c* oxidase from *Paracoccus denitrificans*. Nature 376: 660–669

Jähnig F, Edholm O (1992) Modelling of the structure of bacteriorhodopsin: a molecular dynamics study. J Mol Biol 226: 837–850

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathics character of a protein. J Mol Biol 157: 105–132

Lakshmanan KL, Saraswathi V (1996) The stability of transmembrane helices: a molecular dynamics study on the isolated helices of bacteriorhodopsin. Biopolymers 38: 401–421

Li J, Carroll J, Ellar DJ (1991) Crystal structure of insecticidal δ-endotoxin from *Bacillus thuringiensis* at 2.5 Å resolution. Nature 353: 815–821

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH (1953) Equation of state calculations by fast computing machines. J Chem Phys 21: 1087–1092

Nilges M, Brünger AT (1991) Automated modelling of coiled coils: application to the GCN4 dimerization region. Protein Eng 4: 649–659

Popot J-L (1993) Integral membrane protein structure: transmembrane α-helices as autonomous folding domains. Curr Opin Struct Biol 3: 532–540

Popot J-L, Engelman DM (1990) Membrane protein folding and oligomerization: the two-stage model. Biochemistry 29: 4031–4037

Presnell SR, Cohen FE (1989) Topological distribution of four-α-helix bundles. Proc Natl Acad Sci USA 86: 6592–6596

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

Schertler GFX, Villa C, Henderson R (1993) Projection structure of rhodopsin. Nature 362: 770–772

Son HS (1997) Prediction of membrane protein structure. DPhil Thesis. Oxford University

Tsukihara T, Aoyama H, Yamashita E, Tomizaki T, Yamaguchi H, Shinzawa-Itoh K, Nakashima R, Yaono R, Yoshikawa S (1996) The whole structrue of the 13-subunit oxidized cytochrome *c* oxidase at 2.8 Å. Science 272: 1136–1144

Vinzenz M, Schertler GFX (1995) Low resolution structure of bovine rhodopsin determined by electron cryo-microscopy. Biophys J 68: 1776–1786