Protein Engineering vol.12 no.4 pp.305–311, 1999

Isolated transmembrane helices arranged across a membrane: computational studies

Vladimir M.Tseitin and Gregory V.Nikiforovich[1]

Center for Molecular Design, Institute for Biomedical Computing, Washington University, Box 8036, 700 S. Euclid Avenue, St Louis, MO 63110, USA

[1]To whom correspondence should be addressed

A computational procedure for predicting the arrangement of an isolated helical fragment across a membrane was developed. The procedure places the transmembrane helical segment into a model triple-phase system 'water–octanol–water'; pulls the segment through the membrane, varying its 'global' position as a rigid body; optimizes the intrahelical and solvation energies in each global position by 'local' coordinates (dihedral angles of side chains); and selects the lowest energy global position for the segment. The procedure was applied to 45 transmembrane helices from the photosynthetic reaction center from *Rhodopseudomonas viridis*, cytochrome *c* oxidase from *Paracoccus denitrificans* and bacteriorhodopsin. In two thirds of the helical fragments considered, the procedure has predicted the vertical shifts of the fragments across the membrane with an accuracy of –0.15 ± 3.12 residues compared with the experimental data. The accuracy for the remaining 15 fragments was 2.17 ± 3.07 residues, which is about half of a helix turn. The procedure predicts the actual membrane boundaries of transmembrane helical fragments with greater accuracy than existing statistical methods. At the same time, the procedure overestimates the tilt values for the helical fragments.

Keywords: α-helix/energy calculations/helix–membrane interaction/transmembrane proteins

Introduction

Most biological results obtained using cloned and expressed transmembrane (TM) receptors interacting with G-proteins (GPCRs, G-protein coupled receptors) cannot be fully understood without reliable 3D models of TM receptors. Since no direct experimental data of high quality on 3D structures of GPCRs are available (the only exception is the low-resolution structure of rhodopsin; Schertler *et al*., 1993), construction of 3D models for TM GPCRs starting from their sequences is an important challenge for computational biophysics. Reviews on 3D modeling of GPCRs (Donnelly and Findlay, 1994; von Heijne, 1994; Reithmeier, 1995) outline the same steps to build a 3D model from the sequence, namely, (i) locate TM helical fragments in the GPCR sequence; (ii) pack these fragments together and (iii) restore interhelical loops and N- and C-termini.

This paper discusses a practical problem which arises between the first and second steps, namely how to find which residues in the TM helix lie within the membrane and which are in water on either side of the membrane. Suppose we already know the locations of TM helical fragments in a GPCR sequence (see, e.g. Nikiforovich, 1998). The next step would be to bring several helices in contact in order to obtain maximum complementarity at their interacting surfaces. As a measure of complementarity, one can use conformational energy (or a similar scoring function) of intra- and interhelical interactions. It seems reasonable (see, e.g., Tuffery *et al*., 1994) to assume that the backbone conformations of helices are more affected by their internal atom–atom interactions than by interactions between helices caused by packing. In other words, reasonable backbone conformations of TM helices can be obtained from energy calculations on isolated helices; this assumption correlates well with experimental observations on the packing of the individual helices in membranes (von Heijne, 1994). (However, some recent experimental data suggest that formation of TM helices occurs during insertion in the membrane along with their packing in a bundle; Riley *et al*., 1997.)

In practice, this means that energy calculations could be performed first for a helical bundle, including both intra- and interhelical atom–atom interactions, with the assumption of the 'hard cores' (backbones) and the 'soft shells' (side chains) for each helix. The variables for energy calculation would be the dihedral angles in the side chains for all helices and additional parameters describing the orientation of each helix relative to the others as a rigid body. The results of any procedure for energy minimization might be influenced by the choice of the starting point: either 'locally', i.e. in the space of the dihedral angles for side chains, or 'globally', i.e. in the space of coordinates describing the relative orientation of helices as rigid bodies. The proper choice of the global starting point seems to be most important. For instance, the 3D atomic structure of the bacteriorhodopsin helical bundle has been reproduced by a similar procedure with an accuracy of r.m.s. = 2.06 Å (Cα atoms), when starting tilts and shifts of helices (and also their backbone conformations) were taken from experimental data (Tuffery *et al*., 1994). Our initial results showed that such a procedure was able to reproduce the X-ray complex of the photosynthetic reaction center helices L1/L2 with the r.m.s. value of 1.61 Å (Cα atoms) when energy minimization was started from the experimental dihedral angle values and the experimental global parameters (Nikiforovich *et al*., 1998).

Therefore, to ensure proper reconstruction of a helical bundle, one needs the proper global starting point, i.e. one needs to know, at least, the approximate arrangement of helices across the membrane. Obviously, knowledge of which residues in a helix are situated on the membrane boundary can be of significant help in this respect. However, only a few experimental techniques can be used to approach such a problem in integral protein–membrane (or protein–lipid) systems. The relevant direct structural methods include FTIR (and, perhaps, NMR) measurements using pre-deuterated tyrosines (Sonar *et al*., 1994); ESR measurements using spin labels that can be placed on the cysteine residues which substitute residues in

the GPCR sequences (e.g. Faranbakhsh *et al*., 1995) (the same labels can be used to measure their accessibility to paramagnetic probes); fluorescence measurements using fluorescent labels on ligands and receptors (e.g. Turcatti *et al*., 1995) and measurements by high-resolution solid-state NMR (e.g. the recent study by Yamaguchi *et al*., 1998). There is also an opportunity to employ the fusion of large segments of TM proteins with some enzymes (such as alkaline phosphatase) which are active only on one side of membrane (Manoil and Beckwith, 1986). All these studies are technically complicated, expensive, mostly not detailed enough and, so far, not many of them have been performed. Theoretical methods for the prediction of transmembrane helical segments that also address the problem of which residues are situated on membrane boundaries were reviewed in a previous publication (Nikiforovich, 1998).

As an alternative, this paper suggests a computational procedure to model the 'isolated helix–membrane' integral system. The procedure places a TM helical segment into a model triple-phase system, 'water–octanol–water', where the octanol phase mimics the membrane; pulls the segment through the membrane phase, varying its global positions; optimizes the entire intrahelical and helix–solvent energies in each global position by local coordinates (dihedral angles of side chains); and selects the best global position for the segment corresponding to the lowest energy value. A similar procedure has been used successfully to study orientations of peptides and peptide-like compounds at a lipid–water phase boundary (Galaktionov and Marshall, 1993).

Obviously, the uniform triple-phase system 'water–octanol–water' is a very rough approximation of the real situation on membrane boundaries, which includes complex interactions between the transmembrane peptide, lipids and water molecules. However, the very complexity of this system is, in our view, the source of inevitable errors for molecular modeling using too detailed energy calculations. First, it is still not clear which set of atomic parameters is more adequate for describing peptide–lipid and peptide–water interactions for TM peptides and proteins. For instance, a detailed study of several sets of parameters was not conclusive as to which set is preferable in this situation (Nolde *et al*., 1997). Second, detailed energy calculations require significant computational resources and depend on their starting points [e.g. recent multi-nanosecond MD simulations for a single $(Ala)_{32}$ TM helix in a fully solvated dimyristoylphosphatidylcholine bilayer; Shen *et al*., 1997]. Third, current experimental techniques still do not present sufficient resolution to determine, for instance, how many residues at the C-terminal helix of bacteriorhodopsin protrude from the membrane (Yamaguchi *et al*., 1998); consequently, the results of detailed energy calculations are hard to validate with equally detailed experimental data on the 3D structure of a peptide–membrane–water system.

On the other hand, fairly simple modeling involving, e.g., the empirically derived pore-restraining parameters allowed a reasonable description of TM ion channels consistent with the data on ionic permeability (Grice *et al*., 1997). Even coarser modeling of helical bundles based on complementarity of amphiphilic surfaces of helices (Baldwin, 1993) showed consistency with available experimental data in the case of rhodopsin (Baldwin *et al*., 1997). Therefore, we view our approach as a practical tool, the results of which will not overinterpret the available experimental data, but could be helpful in determining starting points for more detailed modeling procedures (Shen *et al*., 1997).

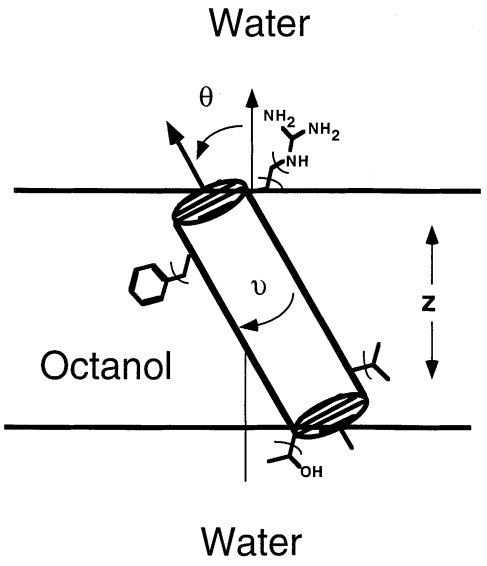

Fig. 1. Schematic diagram showing placing a helical fragment in water–octanol–water system. Global variables (shift *z*, tilt angle θ and rotation angle υ) and local variables (curves around simplified side chains) are shown.

Methods

Figure 1 illustrates the main features of the developed approach. We consider the 'membrane' as a space with a given normal distance *D* ('membrane thickness') between two infinite parallel planes filled inside with a homogeneous octanol environment. Similarly, the homogeneous water environment occupies space from both sides outside the planes. Octanol and water environments differ in their contributions to solvation energy for each particular atom transferred from one environment to another.

Three global coordinates which describe the spatial position of a helical fragment as a rigid body are defined according to diagram in Figure 1. The vertical *z* axis is calculated by averaging the spatial positions of all Cα atoms of the fragment. The axis is oriented normal to the two planes representing the membrane surfaces. Computational procedures pull the helical fragment through the membrane along the *z* axis in the direction of the N-terminus with step increments of 1.5 Å. At each step, the helical fragment is tilted in the direction normal to the axis (tilt angle θ) up to ±40° in 20° increments and is rotated around the axis (rotation angle υ) from 0 to 150° in 30° increments. This corresponds to the total circle of rotation, since both positive and negative tilts are considered. In total, at each vertical step, 25 positions of the helical fragment are examined (24 tilted positions plus one position with zero tilt). Obviously, the above three coordinates fully describe the spatial position of the helical fragment, as a whole, in the membrane; in each of these positions, the computational procedure also performs energy optimization of side chain arrangements by the algorithm developed previously (Nikiforovich *et al*., 1991).

A backbone conformation for the helical fragment was found by independent energy minimization performed for a segment comprising the amino acid sequence of the given fragment extended by four residues (approximately one helical turn) in both directions. We introduced this extension, since the membrane boundaries are difficult to define in the experi-

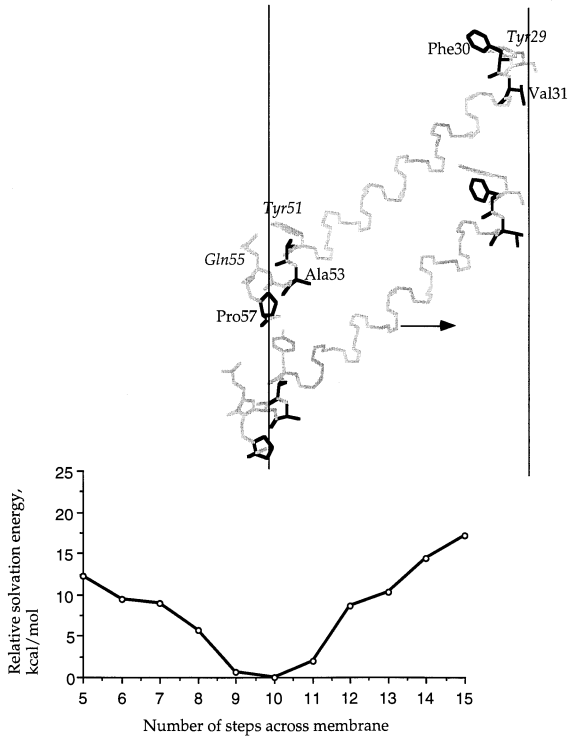

Fig. 2. Graphic images of calculated structures of the PRC L1 helix along with the profile of relative solvation energy for various steps across the membrane. The images correspond to steps 10 and 11 (from top to bottom). For the middle part of the helix, only Cα, C′ and N backbone atoms are shown as thick light gray lines. For the terminal parts, hydrophobic residues are shown in black (labels in roman font) and hydrophilic residues are shown in light gray (labels in italics). Vertical lines represent membrane boundaries. The arrow shows the positive direction across the membrane. The same notations are used also in Figures 3–6.

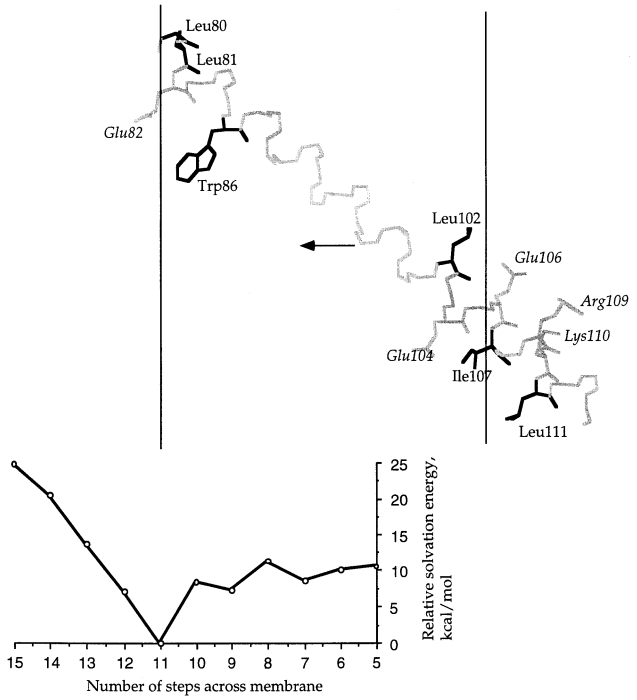

Fig. 3. Graphic image of calculated structure of the PRC L2 helix along with the profile of relative solvation energy for various steps across the membrane. The image corresponds to step 11.

mental 3D structures used for validation (see below). Elongation of the helical segments could account for these uncertainties. Energy minimization was performed employing the standard ECEPP/2 force field upon all dihedral angles, the starting point being the regular helical conformation, i.e. $\{\phi_i = -60°, \psi_i = -60°\}$ with inherent limitations of $-30° \leqslant \phi_i, \psi_i \geqslant -90°$. Side chain spatial locations were optimized by the same algorithm before energy minimization (Nikiforovich *et al*., 1991). The values obtained for the backbone dihedral angles were then 'frozen' in all further calculations.

For each of 25 global spatial positions corresponding to every vertical step, the optimal solvation energy of the helical fragment, E_{sol}, was calculated. Energy optimization was performed by finding the lowest energy arrangement of side chains by the same algorithm (Nikiforovich *et al*., 1991). Energy values themselves are the sum of two parts. The first was regular intramolecular atom–atom interactions inside the helical fragment calculated by the ECEPP/2 force field with the effective dielectric constant equal to 80 (to suppress unwanted electrostatic interactions between ionogenic groups within the membrane). The second was a sum of solvation energies for each atom calculated by the known procedure of the 'excluded volume' (Hopfinger, 1973). This procedure is based on the solvation shell model which assumes that, for the *i*th atom, transition from a completely solvated state to a partially solvated state is associated with a change in free energy, F_i, which is as follows:

$$F_i = \Delta_f \left[n - \left(\frac{V_i^0}{V_s + V_f} \right) \right] \quad \text{when } V_i^0 < n(V_s + V_f)$$

$$F_i = 0 \quad \text{when } V_i^0 \geqslant n(V_s + V_f)$$

where Δ_f is the change in free energy associated with the removal of one solvent molecule from the solvation shell, n is the maximum number of solvent molecules which can occupy the solvation shell; V_i^0 is the total volume excluded from the solvation shell by all other non-bonded atoms of solute, V_s is the volume of the solvent molecule and V_f is the free volume of packing associated with one solvent molecule in the solvation shell.

Obviously, Δ_f, n and V_f depend on the type of atom and on the type of solvent. Our calculations use values derived from water–octanol partition coefficients (Hopfinger and Battershell, 1976). These coefficients are consistent with the ECEPP force field parameters. We consider an atom to be entirely transferred from octanol to water if the Cartesian coordinates of its center are above (or below) the membrane boundaries, no matter how close to the boundaries they may be; there are no atoms partially solvated by water and partially by octanol.

Finally, the minimum value out of 25 values for E_{sol} was obtained for each particular vertical step. As a result, the procedure calculates a profile of the energetic cost for transferring the helical fragment from one side of the membrane to another. The lowest point in this profile may be used for indicating boundaries of the intramembrane helical segment.

Results and discussion

Example of typical energy profiles: PRC L subunit

The general calculation procedure was applied to predict the intramembrane regions of TM helices in the L-subunit of the photosynthetic reaction center (PRC) from *Rhodopseudomonas viridis* (Deisenhofer and Michel, 1989; Deisenhofer *et al*.,

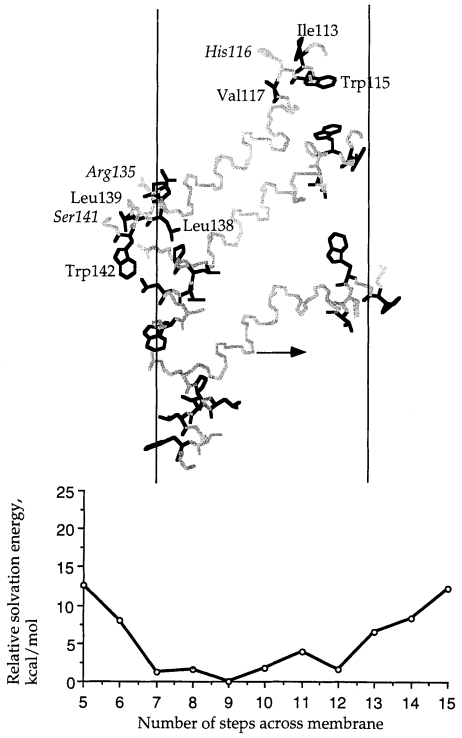

Fig. 4. Graphic images of calculated structures of the PRC L3 helix along with the profile of relative solvation energy for various steps across the membrane. The images correspond to steps 7, 9 and 12 (from top to bottom).

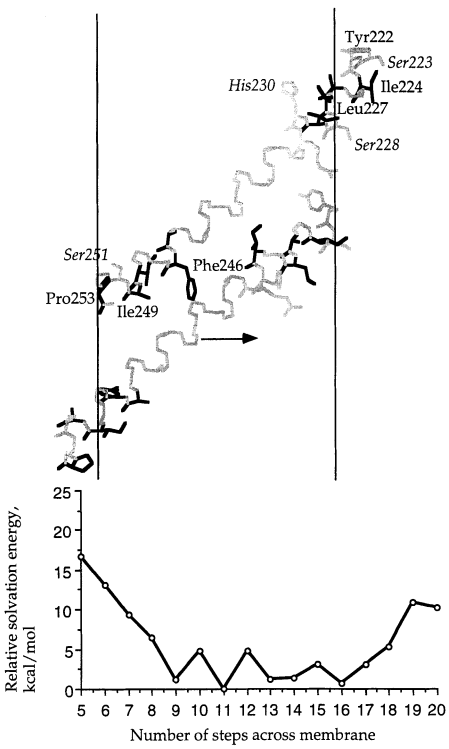

Fig. 6. Graphic images of calculated structures of the PRC L5 helix along with the profile of relative solvation energy for various steps across the membrane. The images correspond to steps 16 and 9 (from top to bottom).

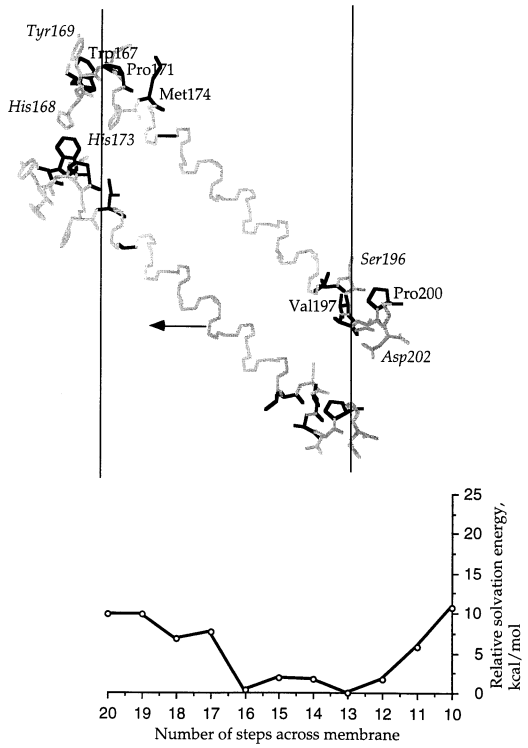

Fig. 5. Graphic images of calculated structures of the PRC L4 helix along with the profile of relative solvation energy for various steps across the membrane. The images correspond to steps 13 and 16 (from top to bottom).

1995). The membrane thickness D is taken to be 30 Å. The optimal tilt obtained was $\pm 40°$, i.e. the maximum allowed value, for all five helices. Profiles of relative E_{sol} values across the membrane for different helices are displayed in Figures 2–6. The profiles range from more or less clear one-minimum curves for the PRC L1 and PRC L2 helices (minima corresponding to steps 10 and 11, respectively) to almost flat curves such as the profile for PRC L5 at steps from 9 to 16. We discuss these results in detail, since their interpretation is typical of our results obtained for other proteins (see below).

Figure 2 shows calculated structures for the PRC L1 helix at steps 10 (the lower image) and 11 (the upper image). It is clear that further movement of PRC L1 across the membrane starting from step 10 would place hydrophobic residues Val31 and Phe30 outside the membrane and hydrophilic residue Gln55 inside the membrane, which will increase the solvation energy. On the other hand, movement backwards from step 10 would place hydrophobic residues Pro57 and Ala53 outside the membrane, which also will increase the solvation energy. That explains the existence of the distinct minimum of solvation energy at step 10.

A similar situation exists for the PRC L2 helix (Figure 3). Here further movement of PRC L1 across the membrane starting from step 11 will expose hydrophobic residues Leu80 and Trp86 to the water phase and the hydrophilic residue Glu106 would be placed inside the membrane. As a result, solvation energy will increase. Movement backwards from step 11 will expose hydrophobic residues Ile107 and Leu102 to water, but the corresponding increase in solvation energy would be partially compensated by placing hydrophilic residues Glu106 and Glu104 in the water phase. That may explain the almost flat profile of E_{sol} at steps from 5 to 10 (Figure 3).

Figure 4 depicts three images of calculated structures of the PRC L3 helix at steps 7, 9 and 12 (from the upper to the lower). The corresponding values of solvation energy are almost the same for all three steps. This is due to a specific

Table I. Relative energies as functions of tilt angle θ and rotation angle υ for PRC L helices in the lowest points of energy profiles of Figures 2–6

Helix	Tilt angle θ (°)	Rotation angle υ (°)					
		0	30	60	90	120	150
L1	–40	0.6	9.1	2.7	3.1	0.0	4.0
	–20	13.4	16.8	16.8	16.8	16.3	16.5
	0	21.4	21.4	21.4	21.4	21.4	21.4
	20	14.9	14.1	14.2	14.3	14.3	14.3
	40	5.5	10.4	9.9	9.5	8.0	10.3
L2	–40	56.1	36.1	24.7	0.0	7.2	53.2
	–20	13.5	21.7	15.0	10.7	14.5	11.7
	0	30.7	30.7	30.7	30.7	30.7	30.7
	20	22.8	19.1	19.1	25.1	34.6	19.8
	40	59.0	69.3	71.6	43.3	56.8	54.4
L3	–40	8.6	7.6	0.0	1.1	12.3	12.1
	–20	17.8	17.8	17.9	18.1	18.1	20.5
	0	27.4	27.4	27.4	27.4	27.4	27.4
	20	20.5	21.2	15.1	19.0	19.0	17.6
	40	10.4	4.3	2.7	8.3	7.6	10.4
L4	–40	1.9	0.0	3.1	5.8	3.0	4.4
	–20	28.5	31.6	32.7	33.9	32.6	31.6
	0	33.4	33.4	33.4	33.4	33.4	33.4
	20	30.8	29.8	24.8	27.5	24.5	27.3
	40	6.5	8.5	9.0	10.3	14.1	10.3
L5	–40	8.6	6.1	5.2	6.0	6.7	0.0
	–20	22.1	19.3	18.5	19.0	19.0	16.9
	0	20.7	20.7	20.7	20.7	20.7	20.7
	20	19.1	19.6	19.9	20.1	18.3	17.2
	40	5.4	11.0	12.4	14.1	12.2	10.5

alternating distribution of amino acid residues along the sequence at the N- and C-termini of the helix. Since hydrophobic residues (Ile113, Trp115, Val117, Leu137, Leu138, Trp142) alternate with hydrophilic residues (His116, Arg135, Ser141), movement across the membrane in any direction will cause both negative and positive contributions to solvation energy, which compensate each other at steps ranging from 7 to 12.

The alternating order of hydrophobic and hydrophilic residues along the sequence is also seen in the PRC L4 helix (Figure 5). At the N-terminus, hydrophobic residues Trp167, Pro171 and Met174 alternate with hydrophilic residues His168, Tyr169 and His 173 and at the C-terminus, hydrophobic residues Val197 and Pro200 are surrounded by hydrophilic residues Ser196 and Asp202 (see the upper image in Figure 5, which corresponds to step 13). Therefore, the values of solvation energy at steps 13 and 16 are nearly the same. Similarly, for the PRC L5 helix, solvation energy values at steps 9, 11 and 16 are almost the same (Figure 6), since hydrophobic residues Ile224, Leu227, Phe246, Ile249 and Pro253 alternate in the sequence with hydrophilic residues Ser223, His230, Ser228 and Ser251. This can be clearly seen at the images of PRC L5 in Figure 6.

The energy profiles in Figures 2–6 correspond to the lowest energy values for every step across membrane taken out of 25 energy values obtained for different tilt and rotation angles. Table I contains those energy values for all five PRC L helices. One can see that in all cases the lowest energy corresponds to the maximum possible tilt angle, $\theta = \pm 40°$. The other important feature to note is that whereas at $\theta = 0°$ all energy values corresponding to various rotation angles υ are the same which is natural in a uniform system, the energy profiles $E(\upsilon)$ became more and more non-uniform with increasing deviations of θ from zero. Therefore, it is legitimate to conclude that

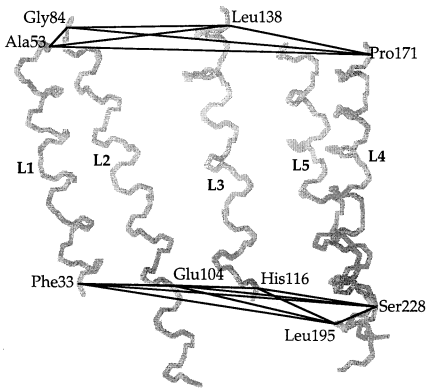

Fig. 7. X-ray structures of TM helices for the L-subunit of PRC (thick light gray lines, only Cα, C′ and N backbone atoms shown) with two parallel planes (thin black lines) designating 'experimental membrane surfaces'.

the energy profiles in Figures 2–6 (and, accordingly, spatial positions of residues at membrane boundaries) depend not only on the number of steps across the membrane, but also on the values of the rotation angle υ.

Problem of determining experimental membrane boundaries

We validated our procedure using all available experimental data of high resolution on 3D structures of TM proteins with a high helical content. They include two structures obtained by X-ray crystallography, namely the photosynthetic reaction center (PRC) from *Rhodopseudomonas viridis* (Deisenhofer and Michel, 1989; Deisenhofer *et al*., 1995) and cytochrome *c* oxidase (OCC) from *Paracoccus denitrificans* (Iwata *et al*., 1995). [The atomic coordinates for the cytochrome bc_1 complex from bovine heart mitochondria (Xia *et al*., 1997) are still unavailable in the Protein Data Bank.] We also used one structure obtained by electron cryo-microscopy, namely bacteriorhodopsin (BR) (Henderson *et al*., 1990). Recently, several studies describing X-ray data on BR have been published (e.g. Pebay-Peyroula *et al*., 1997); the experimentally determined TM helices in these studies, are very close to those determined by electron cryo-microscopy (Henderson *et al*., 1990).

Since the above data do not contain any direct information on the spatial positions of membrane boundaries, we established 'experimental membrane surfaces' by visual examination of experimental atomic coordinates preserving two requirements: (i) both outer and inner surfaces should be as parallel to each other as possible and (ii) the experimentally determined boundary residues of TM helical fragments should be placed at the established surfaces as often as possible.

To clarify this point, Figure 7 depicts the 'experimental membrane surfaces' established for the X-ray structure of the L-subunit of PRC. The lines connecting the Cα atoms of the labeled residues in Figure 1 clearly form two parallel planes which can be regarded as 'membrane boundaries'. The lower plane in Figure 7 involves residues belonging to all helices in the PRC L subunit. Two of these residues, namely Phe33 and His116, are the residues which terminate the PRC L1 and PRC L3 helices according to X-ray data (Deisenhofer and Michel, 1989; Deisenhofer *et al*., 1995). The upper plane in Figure 7 involves residues belonging to only four of the helices, since the Ile249 residue in the PRC L5 helix is located below the plane. All four residues forming the upper plane (i.e. Ala53, Gly84, Leu138 and Pro171) are the residues which terminate helices. Thus, six out of 10 residues which mark the ends of

Table II. Intramembrane boundaries for TM helical fragments in PRC, BR and OCC

Protein, subunit, helix	Experimentally defined TM helices	'Experimental' intramembrane fragments[a]	Intramembrane fragments according to our predictions[a]
PRC L1	33–53	33–53	29–54
PRC L2	84–111	84–104	87–112
PRC L3	116–139	116–138	(112)–138
PRC L4	171–198	171–195	170–198
PRC L5	226–249	228–(249)	222–247
PRC M1	52–76	52–76	49–74
PRC M2	111–137	111–134	107–132
PRC M3	143–166	(143)–164	142–167
PRC M4	198–223	198–223	197–225
PRC M5	260–284	260–284	262–287
BR1	10–32	10–32	7–32
BR2	38–62	40–62	39–64
BR3	80–100	80–100	80 –101
BR4	108–127	(108)–121	106–(131)
BR5	137–157	137–157	133–158
BR6	167–191	170–191	167–192
BR7	203–225	203–225	202–226
OCC A1	12–40	12–40	15–36
OCC A2	51–86	51–84	58–87
OCC A3	95–117	95–117	(91)–115
OCC A4	141–170	141 –170	146–171
OCC A5	183–212	183–212	185–211
OCC A6	228–261	228–259	229–255
OCC A7	270–286	270–(286)	(266)–(290)
OCC A8	299–327	299–327	302–(331)
OCC A9	336–357	(336)–(357)	334–359
OCC A10	371–400	(371)–398	371–396
OCC A11	407–433	407–433	408–433
OCC A12	447–478	447–477	448–473
OCC B1	15–45	19–45	23–(49)
OCC B2	60–87	60–85	62–88
OCC C1	16–34	(16)–34	(12)–(38)
OCC C2	41–66	(41)–64	(37)–59
OCC C3	73–105	76–105	78–103
OCC C4	129–152	(129)–152	(125)–151
OCC C5	156–183	157–183	158–178
OCC C6	191–223	193–222	195–220
OCC C7	233–255	233–(255)	234–259
OCC D1	77–103	77–103	79–99
OCC G1	13–37	13–37	14–31
OCC I1	12–52	17–45	12–37
OCC J1	26–54	26–55	24–58
OCC K1	9–35	13–35	9–34
OCC L1	18–44	18–44	15–43
OCC M1	12–35	12–35	15–(39)

[a]The numbers in parentheses correspond to terminal residues found inside membrane.

helices can be placed at the two parallel 'membrane boundaries' in this case. Any other way of tracing the planes would violate their parallel arrangement.

Exactly the same arguments were used to establish 'experimental membrane boundaries' for all other protein subunits considered in this study. Comparison of the second and the third columns of Table II shows that, in total, the helix-terminated residues have been placed at the boundaries in 55 cases out of 90.

Parameters of procedure: membrane thickness and octanol–water partition coefficients

The next set of predictions was performed for the L- and M-subunits of PRC and BR assuming various values for the membrane thickness *D*. The goal of these calculations was to clarify which value of *D* can be regarded as optimal in our procedure and whether we can find the expected dependence of the helix tilt angle θ on the *D* value (the larger is the *D* value, the smaller absolute values of θ can be obtained).

We examined three values for *D*, namely, 24, 30 and 32 Å. For each value, we calculated the average and standard deviation of prediction errors, i.e. of the differences (in numbers of residues) between predicted and experimentally observed residues at membrane boundaries (e.g. differences between the third and the fourth columns in Table II). The sign of the difference (positive or negative) was defined depending on whether a predicted boundary residue was located outside or inside the membrane, respectively, relative to an observed boundary residue. Calculations of the average errors did not include the cases where either the predicted or observed terminal residue of the helical fragment (bracketed numbers in Table II) was located inside the membrane.

The absolute value of the average error calculated for results obtained with $D = 24$ Å was higher (–2.81 residues) than the errors calculated for results obtained with $D = 30$ and 32 Å (1.66 and 1.92 residues, respectively). Also, in almost all cases, the tilt angle values corresponding to spatial positions of helices with the lowest energy values were the same despite the difference in *D* values, namely, most of helices possessed maximum tilt angles, i.e. $\theta = \pm 40°$; in some cases (e.g. BR3 helix) a tilt angle of $\pm 20°$ was found for all values of *D*. This finding does not prove that the optimum tilt angles do not depend, as a rule, on membrane thickness; however, it seems to be the case in our procedure with strict limitations on the maximum tilt angles. Therefore, we decided to perform further calculations for OCC helices with $D = 30$ Å.

The advantages of employing octanol–water partition coefficients found previously (Hopfinger and Battershell, 1976) have been discussed already. However, we performed some additional calculations for PRC L helices ($D = 24$ Å), using another way of calculating solvation energies. We used the known Connolly algorithm for solvating atomic surfaces with parameters borrowed from a more recent paper (Juffer *et al*., 1995). The results obtained were at the same level of the prediction error, requiring, at the same time, much longer computations; therefore, these parameters and also this computational procedure have been abandoned.

Prediction results: general overview

Table II combines results of all our predictions for PRC, BR and OCC performed with $D = 30$ Å. One can see that both sides of a helical fragment have been placed either outside the membrane or immediately at the membrane surfaces for only 30 out of 45 fragments described in Table II. For the other 15 fragments, at least one terminal residue is immersed in the membrane according either to experimental or to calculated estimates. The average value and standard deviation for the prediction errors in the case of the former 30 fragments was (-0.15 ± 3.12) residues. The same value for the latter 15 fragments calculated for both bracketed and unbracketed numbers in Table II was (2.17 ± 3.07) residues, i.e. roughly half of a helix turn.

The calculations predicted the largest possible tilt values ($\pm 40°$) for almost all helical fragments, except BR3, BR5, OCC A1, OCC C5 and OCC D1, where these values were $\pm 20°$. Those overtilts are the main source of errors in our predictions. In particular, they lead to immersion in the

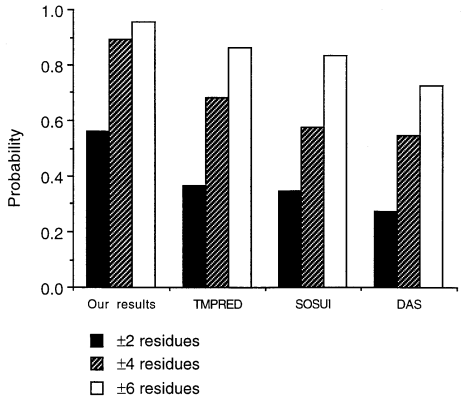

Fig. 8. Histograms of probabilities of predicting the actual boundaries of transmembrane helical fragments by our results and by the three statistical methods with accuracies of ±2, ±4 and ±6 residues, respectively

membrane of the extra four residues which have been added from each side of a helical fragment in our calculations (see Methods). That is why the prediction data tend to increase the length of the intramembrane segments for these helical fragments (note the positive value of the average error). At the same time, it may be reasonable to suggest that the extra four residues are more likely to retain their helical conformation in the membrane rather than in water; keeping the helical conformation intact in our calculations adds to the effect of over-predicting the length of the intramembrane segments.

Nevertheless, we can consider our results as fairly successful compared with similar results obtained by others. In fact, we have predicted boundaries of transmembrane segments, as did many other authors reviewed in our previous publication (Nikiforovich, 1998). For comparison purposes, we borrowed results of predictions for the same proteins, PRC, BR and OCC, performed by three different advanced statistical methods (TMPRED based on Rost *et al.*, 1995, SOSUI based on Yanagihara *et al*., 1989, and DAS based on Cserzo *et al*., 1994) described previously (Nikiforovich, 1998; Table I, third, fourth and fifth columns, respectively). Figure 8 depicts histograms of probabilities of predicting the actual boundaries of transmembrane helical fragments by our results and by the three discussed methods with an accuracy of ±2, ±4 and ±6 residues, respectively. To obtain the histograms in Figure 8, we calculated absolute values of differences (in residue numbers) between boundaries predicted by all four methods and those 'experimentally measured' (third column in Table II in the present study). Only unbracketed numbers were used for the calculations. The histograms clearly show that our procedure determines the actual membrane boundaries of transmembrane fragments significantly more accurately than any of the three statistical approaches at all three levels of accuracy, i.e. for ±2, ±4 or ±6 residues.

Summarizing, we have developed a computational procedure which predicts the arrangement of an isolated helical fragment across a membrane. We applied this procedure to the available experimental data and examined adjustments of the unknown parameters such as membrane thickness. In two thirds of the helical fragments considered, the procedure predicted the vertical shifts of the fragments across membrane in excellent agreement with the experimental data. In the remaining fragments, the agreement with the experimental data was less good (error of about half of a helix turn), but was still good enough for the suggested purpose, namely for deducing global starting points for helix packing as far as the relative shifts of helices are concerned. The procedure predicts actual membrane boundaries of transmembrane helical fragments with greater accuracy than existing statistical methods. At the same time, our procedure overestimates the tilt values for the helical fragments. This is due to the fact that the current procedure considers only *isolated* helical fragments, whereas in reality these fragments can interact within the transmembrane bundles. Accordingly, further development of this procedure will involve the simultaneous handling of several helical fragments in the membrane model.

Acknowledgements

The authors are grateful to Dr Rohit Pappu for making available the subroutine for calculation of the vertical axes for the helical fragment and to Drs Garland R.Marshall and Stan Galaktionov for general support and fruitful discussions. The authors also acknowledge Dr Tom Blackwell for reading the manuscript and for discussing the results. This work was supported in part by NIH grant RO1 GM48184.

References

Baldwin,J.M. (1993) *EMBO J.*, **12**, 1693–1703.
Baldwin,J.M., Schertler,G.F.X. and Unger,V.M. (1997) *J. Mol. Biol.*, **272**, 144–164.
Cserzo,M., Bernassau,J.-M., Simon,I. and Maigret,B. (1994) *J. Mol. Biol.*, **243**, 388–396.
Deisenhofer,J., Epp,O., Sinning,I. and Michel,H. (1995) *J. Mol. Biol.*, **246**, 429–457.
Deisenhofer,J. and Michel,H. (1989) *Science*, **245**, 1463–1473.
Donnelly,D. and Findlay,J.B.C. (1994) *Curr. Opin. Struct. Biol.*, **4**, 582–589.
Faranbakhsh,Z.T., Ridge,K.D., Khorana,H.G. and Hubbel,W.L. (1995) *Biochemistry*, **34**, 8812–8819.
Galaktionov,S.G. and Marshall,G.R. (1993) *Biophys. J.*, **65**, 608–617.
Grice,A.L., Kerr,I.D. and Sansom,M.S.P. (1997) *FEBS Lett.*, **405**, 299–304.
Henderson,R., Baldwin,J.M., Ceska,T.A., Zemlin,F., Beckmann,E. and Downing,K.H. (1990) *J. Mol. Biol.*, **213**, 899–929.
Hopfinger,A.J. (1973) *Conformational Properties of Macromolecules*. Academic Press, New York.
Hopfinger,A.J. and Battershell,R.D. (1976) *J. Med. Chem.*, **19**, 569–573.
Iwata,S., Ostermeier,C., Ludwig,B. and Michel,H. (1995) *Nature*, **376**, 660–669.
Juffer,A.H., Eisenhaber,F., Hubbard,S.J., Walther,D. and Argos,P. (1995) *Protein Sci.*, **4**, 2499–2509.
Manoil,C. and Beckwith,J. (1986) *Science*, **233**, 1403–1408.
Nikiforovich,G.V. (1998) *Protein Engng*, **11**, 279–283.
Nikiforovich,G.V., Hruby,V.J., Prakash,O. and Gehrig,C.A. (1991) *Biopolymers*, **31**, 941–955.
Nikiforovich,G.V., Galaktionov,S., Tseitin,V.M., Lowis,D.R., Shenderovich,M.D. and Marshall,G.R. (1998) *Lett. Pept. Sci.*, **5**, 413–415.
Nolde,D.E., Arseniev,A.S., Vergoten,G. and Efremov,R.G. (1997) *J. Biomol. Struct. Dyn.*, **15**, 1–18.
Pebay-Peyroula,E., Rummel,G., Rosenbush,J.P. and Landau,E.M. (1997) *Science*, **277**, 1676–1681.
Reithmeier,R.A.F. (1995) *Curr. Opin. Struct. Biol.*, **5**, 491–500.
Riley,M.L., Wallace,B.A., Flitsch,S.L. and Booth,P.J. (1997) *Biochemistry*, **36**, 192–196.
Rost,B., Casadio,R., Fariselli,P. and Sander,C. (1995) *Protein Sci.*, **4**, 521–533.
Schertler,G.F., Villa,C. and Henderson,R. (1993) *Nature*, **362**, 770–772.
Shen,L., Bassolino,D. and Stouch,T. (1997) *Biophys. J.*, **73**, 3–20.
Sonar,S., Lee,C.-P., Coleman,M., Patel,N., Liu,X., Marti,T., Khorana,G.G., RajBhandary,U.L. and Rothschild,K.J. (1994) *Struct. Biol.*, **1**, 512–517.
Tuffery,P., Etchebest,C., Popot,J.L. and Lavery,R. (1994) *J. Mol. Biol.*, **236**, 1093–1104.
Turcatti,G., Vogel,H. and Chollet,A. (1995) *Biochemistry*, **34**, 3972–3980.
von Heijne,G. (1994) *Annu. Rev. Biophys. Biomol. Struct.*, **23**, 167–192.
Xia,D., Yu,C.-A., Kim,H., Xia,J.-Z., Kachurin,A.M., Zhang,L., Yu,L. and Deisenhofer,J. (1997) *Science*, **277**, 60–66.
Yamaguchi,S., Tuzi,S., Seki,T., Tanio,M., Needleman,R., Lanyi,J.K., Naito,A. and Saito,H. (1998) *J. Biochem.*, **123**, 78–86.
Yanagihara,N., Suwa,M. and Mitaku,S. (1989) *Biophys. Chem.*, **34**, 69–77.

Received August 8, 1998; revised 17 November, 1998; accepted December 28, 1998