Protein Engineering vol.11 no.4 pp.279–283, 1998

A novel, non-statistical method for predicting breaks in transmembrane helices

Gregory V.Nikiforovich

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

We have developed a novel, non-statistical procedure for predicting possible breaks in transmembrane helices based on energy calculations. The procedure consists of stepwise elongation of the 'core' helical fragment determined by consensus results of several available prediction procedures. At each step, we calculate the conformational energies corresponding to the regular 'frozen' helical conformer of the 'core' fragment elongated by two flanking residues, $E(\alpha)$, as well as those to several options for the fragment to enter or exit the helix by changing conformations of the flanking residues, E_i. The minimal values out of $E_i - E(\alpha)$, Δ_k, can be viewed as a profile of relative energies, where each minimum of Δ_k is a signal to start or to stop transmembrane helix. We suggest that boundaries of the transmembrane helix would be determined by the signals closest to the 'core' sequence in the Δ_k profiles. Our procedure was applied to prediction of the N- and C-termini for 45 transmembrane helices from the photosynthetic reaction center from *Rhodopseudomonas viridis*, bacteriorhodopsin and the cytochrome *c* oxidase from *Paracoccus denitrificans*. The results clearly showed that it is significantly more probable that a prediction accuracy within an error of ±2 residues will be obtained by our procedure than by three different statistical approaches.

Key words: energy calculations/α-helix/secondary structure prediction/transmembrane proteins

Introduction

Constructing 3D models for transmembrane G-protein-coupled receptors (GPCRs) starting from their sequences has emerged as one of the most challenging tasks facing computational biophysics and, simultaneously, as one of the most urgent problems for drug design. Recent reviews on 3D modeling of GPCRs (Donnelly and Findlay, 1994; von Heijne, 1994; Reithmeier, 1995) outline basically the same pathway for building a 3D model from the protein sequence; this model consists of three main steps: (i) location of the transmembrane helical fragments in the GPCR sequence (assuming they are helices); (ii) packing these fragments together in a bundle, and (iii) restoring interhelical loops and N- and C-termini. Predicting α-helical segments within the amino acid sequences is an important component in this general problem. Many methods for the prediction of α-helices (and other regular 3D structures) have been developed during the last three decades (for review see, e.g., Rost and Sander, 1995). Only a few use general physical and thermodynamical considerations explicitly; the vast majority of the methods developed employ various statistical approaches, sometimes based on rather sophisticated techniques such as neural networks or hidden Markov models. A general feature of these approaches consists of finding rules for the appearance of α-helical backbone conformations in the selected 'training set' of proteins with known 3D structures. Once established, the rules are validated by predicting helical segments for proteins with known 3D structures outside the training set. Obviously, the larger the training set, the more reliable are subsequent predictions; in some cases they are sufficiently accurate that 70–80% residues are correctly predicted as regards their α-helical conformational states.

All of those approaches have been most highly developed for soluble globular proteins. Their application to transmembrane proteins, including GPCRs, is virtually impossible for several reasons. First, the helical propensities for amino acid residues in transmembrane proteins immersed in a lipid bilayer are likely to be completely different from those in soluble proteins. [See, for instance, the discussion in Deber and Li (1995) on the possible role of the typical helix-breaking residue, proline, in transmembrane helical segments.] Second, as was noted above, statistical approaches require substantial experimentally defined training sets to establish the rules for predicting helices. Unfortunately, few reliable experimental data on transmembrane helical segments are available. In fact, the only experimental data on the 3D structure of transmembrane proteins with high helical content consist of three structures obtained by X-ray crystallography, namely, the photosynthetic reaction center (PRC) from *Rhodopseudomonas viridis* (Deisenhofer and Michel, 1989; Deisenhofer *et al.*, 1995), cytochrome *c* oxidase (OCC) from *Paracoccus denitrificans* (Iwata *et al.*, 1995) and cytochrome bc_1 complex from bovine heart mitochondria (Xia *et al.*, 1997); and two structures obtained by cryoelectron microscopy, namely bacteriorhodopsin (Henderson and Schertler, 1990) and rhodopsin (Schertler *et al.*, 1993). Of these proteins, only rhodopsin is a GPCR. In fact, only 45 transmembrane helices (28 for OCC, 10 for PRC and seven for bacteriorhodopsin) have been determined experimentally with sufficient resolution, and available in the Protein Data Bank (the atomic coordinates of cytochrome bc_1 complex from bovine heart mitochrondria are still unavailable). That is clearly not enough for training any statistical procedure, so various authors include in their training sets results of predictions of helical segments for different transmembrane proteins, mainly GPCRs, based on homology alignments and alignments of hydrophobic patterns (Rost *et al.*, 1995), thus reducing the reliability.

We have developed a novel procedure for locating the possible ends of transmembrane helical fragments, based on simple energetic assumptions. We have used the following basic considerations: (i) The direct experimental data describing the details of the complex molecular structure on the membrane surface which involves phospholipid molecules, water and transmembrane protein fragments are still not available. Under the circumstances, any attempts to reproduce such molecular details in modeling will inevitably result in a variety of

© Oxford University Press

assumptions, so the final predictions obtained by modeling would be still rather uncertain. Therefore, we will consider a rough model consisting of a separated helical fragment inserted into the membrane environment. (ii) The influence of the membrane environment is represented in our approach according to the following. To some extent, transmembrane helices are immobilized within the membranes. At the same time, the significance of interhelical interactions, especially those maximizing the number of possible hydrogen bonds, should increase in helical bundles (see, e.g., Pogozheva *et al.*, 1996). Therefore, a suitable computational model for an isolated transmembrane helix may consider a 'hard interior' (fixed values of dihedral angles for peptide backbone) and a 'soft exterior' (variable angles of dihedral angles for side chains). (iii) Transmembrane helical segments should be long enough to span the hydrophobic core of the membrane, the lower estimate being ~17 or 18 residues. (iv) We assume, according to Efimov (1986, 1991), that only a limited number of backbone conformers are available for the two N-terminal or C-terminal residues without steric hindrance where a peptide chain 'enters' an α-helix at the N-end or 'exits' from it at the C-end. These data were obtained employing very simple sterical considerations (mainly CPK models; see Efimov, 1986), and are equally valid for helices in soluble and transmembrane proteins.

Methods

In the initial step of the procedure, we locate an approximate position of the 'core' transmembrane helix in amino acid sequence by taking the intersection of predictions from several different statistical procedures. Specifically, we have used three different statistical procedures readily available via the Internet (the respective URL addresses are as follows: http://ulrec3.unil.ch/software/TMPRED_form.html; http://www.tuat.ac.jp/~mitaku/sosui/; http://www.biokemi.su.se/~server/DAS/). The first one (the TMPRED program) is basically the neural network application developed by Rost *et al.* (1995), the second one (the SOSUI program) was developed by Yanagihara *et al.* (1989) based mainly on studies on the spectral density of hydrophobicity plots, and the third one is based on the so-called dense alignment method (DAS) suggested by the von Heijne group which is a generalization of sequence alignment techniques proposed earlier (Cserzö *et al.*, 1994). The 'core' sequences can be seen in any row of Table I under the 'Statistical predictions' heading. For the TMPRED predictions listed in Table I, we selected the data with the highest scores and with transmembrane helix length limitations of 17–34 residues. For the DAS predictions we used those with the loose DAS cutoff of 1.7. The SOSUI program offers a single prediction. It is noteworthy that the 'core' sequences may be determined by other means, too, e.g., by the standard Kyte–Doolitle hydropathy plots (see the ProtScale procedure in the Swiss-Prot bank).

In the next step, two extensions of the transmembrane helix towards the N- and C-termini are defined by selecting sequence fragments not less than 30 residues in length; the first of these fragments would start at the beginning of the first extension and would end at the C-end of the 'core' helix; the second would start at the N-end of the 'core' helix and would end at the end of the second extension. Thus, the 'core' sequence belongs to both fragments. Energy calculations are then performed for both fragments, 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 are optimized in the energy calculations by the algorithm proposed earlier (Nikiforovich *et al.*, 1991). The minimum energy values obtained for the backbone dihedral angles are stored for further use.

The procedure itself can be divided into two parts. First, possible 'breaks' of transmembrane helices from the N-end are determined. Energy calculations are performed for several conformers of the 'core' fragment extended by two residues towards the N-end. Dihedral angles for the backbone of the core fragment are always frozen in the helical values obtained by energy minimization from the previous step. At the same time, starting values for the backbone dihedral angles of the first two residues are selected from a set consisting of the regular helical conformation and sterically allowed 'entries' to a helical segment (Efimov, 1986, 1991). Reference studies suggest the following possible backbone conformations of such two-residue entries: βα, βγ, δα, δγ, and, additionally, $\alpha_L\alpha$ and εα when the first residue is Gly [Efimov notations are used; see Efimov (1986, 1991)]. Translating this notation into the nearest known local minima in Ramachandran maps, we use the following starting points: $\{\phi_1 = -140°, \psi_1 = 140°; \phi_2 = -60°, \psi_2 = -60°\}$; $\{\phi_1 = -140°, \psi_1 = 140°; \phi_2 = -75°, \psi_2 = 80°\}$; $\{\phi_1 = -140°, \psi_1 = 80°; \phi_2 = -60°, \psi_2 = -60°\}$; $\{\phi_1 = -140°, \psi_1 = 80°; \phi_2 = -75°, \psi_2 = 80°\}$; and, additionally, $\{\phi_1 = 140°, \psi_1 = 140°; \phi_2 = -60°, \psi_2 = -60°\}$ and $\{\phi_1 = 60°, \psi_1 = 60°; \phi_2 = -60°, \psi_2 = -60°\}$ in the case when the first residue is Gly. These angles are allowed to vary in the process of energy minimization, as well as the dihedral angles in all side-chains. Thus, in this step we obtain conformational energies corresponding to the regular helical conformer, $E(\alpha)$, as well as to four (or six) options for a given peptide fragment to enter the helix, E_i. Then we report the smallest value of $E_i - E(\alpha)$. This value, Δ_k, characterizes the relative probability of the peptide fragment entering a helix at the kth residue: the lower the value of Δ_k, the greater the tendency for the helix to be disrupted at this residue. Increasing the size of the core helical fragment by one residue towards the N-end and repeating the entire procedure, we obtain the Δ_{k-1} value. (Note that now the frozen helical fragment is one residue longer.) Continuing the same process by one residue at a time along the amino acid sequence towards the N-terminus, we can obtain several low Δ_k values, which would represent several 'signals' to disrupt the helix at the N-terminus. These values can be viewed as a profile of relative energies, where each minimum of Δ_k should be regarded as a possible starting point for the transmembrane helix.

The procedure deals with exits from a helix in basically the same way, i.e. by elongating the core helical fragment towards the C-end. The studies by Efimov (1986, 1991) suggest the following backbone conformers for residues immediately after the helix: $\gamma\alpha_L\beta$, $\gamma\beta\alpha_L\beta$ (rarely), γβ, γδ, and, additionally, γε when the last residue is Gly. Some of these 'exits' are composed from more than two residues; however, examining possible hydrogen bonds in the backbone [see Figure 7 in Efimov (1991)] shows that, as a result of the presence of the *N*-methylamide at the C-terminus, all important hydrogen bonds are accounted for in our calculations, even for the $\gamma\alpha_L\beta$ 'exit'. Accordingly, we consider the following starting points for the last two residues in our energy calculations: $\{\phi_{n-1} = 60°, \psi_{n-1} = -60°; \phi_n = -60°, \psi_n = -60°\}$ (the regular helical conformer); $\{\phi_{n-1} = -75°, \psi_{n-1} = 80°; \phi_n = -140°, \psi_n = 140°\}$; $\{\phi_{n-1} = -75°, \psi_{n-1} = 80°; \phi_n = -140°,$

Table I. Predicted boundaries for transmembrane helical fragments in PRC, bacteriorhodopsin and OCC

Protein	Subunit and helix	Statistical predictions				Our predictions	Experimental data
		TMPRED	SOSUI	DAS	core		
Photosynthetic reaction center	L1	29–53	32–56	25–50	32–50	32–51	33–53
	L2	83–105	83–107	89–99	89–99	83–104	84–111
	L3	119–140	117–141	117–139	119–139	117–139	116–139
	L4	174–194	177–201	178–196	178–194	172–198	171–198
	L5	232–256	221–245	224–247	232–245	232–250	226–249
	M1	52–71	51–75	52–73	53–71	52–74	52–76
	M2	110–132	110–133	114–125	114–125	110–128	111–137
	M3	146–166	140–164	147–163	147–163	146–165	143–166
	M4	199–219	204–228	205–223	205–219	204–222	198–223
	M5	268–289	266–290	268–288	268–288	264–288	260–284
Bacteriorhodopsin	1	10–29	11–35	13–29	13–29	10–32	10–32
	2	42–66	45–69	46–67	46–66	39–68	38–62
	3	80–101	80–101	89–101	89–101	83–101	80–100
	4	108–133	107–131	109–126	109–126	108–127	108–127
	5	135–154	136–159	139–156	139–154	138–157	137–157
	6	177–193	176–200	178–190	178–190	178–198	167–191
	7	206–224	206–224	203–224	206–224	205–224	203–225
Cytochrome *c* oxidase	A1	15–37	19–43	18–36	19–36	18–36	12–40
	A2	56–78	62–86	57–78	62–78	59–82	51–86
	A3	97–123	97–120	81–116	97–116	95–119	95–117
	A4	150–169	146–170	145–167	150–167	147–167	141–170
	A5	186–206	185–209	184–208	186–206	182–206	183–212
	A6	243–261	240–264	243–255	243–255	240–262	228–261
	A7	270–290	270–289	274–289	274–289	274–294	270–286
	A8	303–324	307–329	309–316	309–316	302–321	299–327
	A9	337–359	338–362	339–359	339–359	339–362	336–357
	A10	379–397	379–403	374–397	379–397	378–398	371–400
	A11	412–433	414–436	414–422	414–422	408–433	407–433
	A12	452–473	455–478	458–474	458–473	453–478	447–478
	B1	27–46	25–48	27–47	27–46	16–47	15–45
	B2	63–85	64–88	65–83	65–83	63–84	60–87
	C1	17–35	10–33	21–31	21–31	16–35	16–34
	C2	nd	40–64	39–49	40–49	34–54	41–66
	C3	83–100	81–104	82–101	83–100	79–104	73–105
	C4	129–146	124–147	132–144	132–144	130–148	129–152
	C5	159–176	159–178	162–175	162–175	160–181	156–183
	C6	205–223	196–220	196–220	205–220	197–225	191–223
	C7	237–259	240–257	244–257	244–257	nd	233–255
	D1	80–98	75–98	81–98	81–98	78–103	77–103
	G1	18–36	nd	17–34	18–34	13–37	13–37
	I1	19–40	21–38	17–35	21–35	19–39	12–52
	J1	34–52	29–52	33–51	34–51	29–52	26–54
	K1	16–38	13–37	17–34	17–34	12–34	9–35
	L1	22–40	nd	22–38	22–38	17–39	18–44
	M1	15–35	17–39	18–34	18–34	13–34	12–35

nd, this helical fragment was not predicted by this particular method.

$\psi_n = 80°\}$; $\{\phi_{n-1} = -75°, \psi_{n-1} = 80°; \phi_n = 60°, \psi_n = 60°\}$, and, additionally, $\{\phi_{n-1} = -75°, \psi_{n-1} = 90°; \phi_n = -140°, \psi_n = -140°\}$ in the case when the last residue is Gly. Again, the backbone dihedral angles for the two last residues and for all side-chains are allowed to rotate, whereas the backbone angles within the core fragment are frozen. Thus, we obtain Δ_k values by extending the core fragment towards the C-terminus exactly as described above, but now these values characterize the relative probability of the peptide fragment exiting a helix at the kth residue.

Energy calculations for each peptide fragment were performed as described earlier (Nikiforovich, 1994) employing the ECEPP/2 force field (Dunfield *et al.*, 1978; Nemethy *et al.*, 1983). The end groups of all fragments were the acetyl groups at the N-termini and the *N*-methylamides at the C-termini. Obviously, the final results would depend on the chosen force field. However, the results would be even more dependent on other assumptions, e.g. on the rather strong limitations for backbone conformational mobility (the fixed dihedral angles in the core sequence), so the choice of the force field does not seem so important.

Results and discussion

Generally speaking, the stronger the signal (i.e. the smaller Δ_k), the higher should be the probability of entering or exiting a helix. Our aim is to model transmembrane helices, which is why we keep the backbone dihedral angles of the core sequences frozen. In fact, our procedure checks the probabilities of the core transmembrane helices of various lengths being broken at both sides. Typically, it gives several 'weak points' where a helix can be broken (i.e. several minima along the Δ_k profile), corresponding to several lengths of possible transmembrane helices. Obviously, the actual lengths of the transmembrane helices depend on the thickness of the

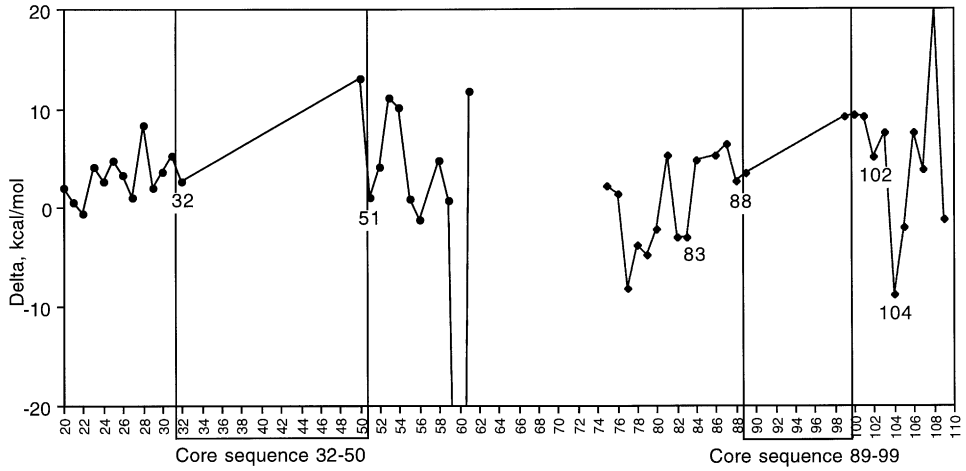

Fig. 1. Calculated 'signals' for boundaries of PRC L1 and L2 helices.

membrane they span. These unknown values would vary from one transmembrane protein to another, so it is impossible to include in our procedure the proper length of the 'core' transmembrane helix *a priori*.

At the same time, it seems reasonable to assume that immobilization of the helical structure of the transmembrane fragment depends not only on interactions with the lipid environment but also on the entire system of intramolecular interactions within a fragment with that particular amino acid sequence. Therefore, we have suggested that boundaries of the transmembranes helix modeled by the rigid 'core' sequence would be determined by the signals closest to the core sequence in the Δ_k profiles. In other words, we suggest that the transmembrane helix tends to be broken at the very first opportunity during its elongation. Sometimes, however, the resulting helix may be too short (less than 17 or 18 residues); in that case we choose N-terminal and C-terminal signals among those next to the closest ones corresponding, at the same time, to the lowest Δ_k values.

As an example, Figure 1 depicts the Δ_k profiles obtained for the first and second transmembrane helices of the L-subunit of PRC (PRC L1 and PRC L2). The core sequences are shown in boxes; they have been predicted at positions 32–50 and 89–99 as the intersection of results by three statistical methods mentioned above. For PRC L1, the Δ_k minima closest to the core correspond to residue 32 at the N-terminus, and to residue 51 at the C-terminus, respectively (that is, residue 32 is the first in the helix, and residue 51 is the last in the helix). Accordingly, the length of the predicted helical fragment would be 20 residues, which is quite satisfactory. However, in the case of PRC L2, the minima closest to the core correspond to residues 88 and 102, i.e. to a fragment with a length of 15 residues. Therefore, out of three possible options, namely 88–104, 83–102 and 83–104, we select the last one as corresponding to the lowest values of Δ_k.

Table I summarizes results of predictions for 45 transmembrane helices from PRC, bacteriorhodopsin and OCC by our procedure as well as by the three statistical methods, and compares them with experimental data. (Note that the transmembrane helices of PRC and bacteriorhodopsin have already been included in the training sets for all three statistical methods.) Some helical fragments were not predicted by some methods: the TMPRED predictions did not score the OCC C2 helix high enough; the OCC G- and L-subunits were not recognized as membrane proteins by the SOSUI procedure; and our procedure failed to show a clear Δ_k minimum at the C-terminus of the OCC C7 helix. Accordingly, all further comparisons between different prediction methods have been made employing the results obtained for the remaining 41 helices.

Figure 2 displays distributions of the distances between experimental and predicted endpoints (i.e. errors of predictions) for 82 ends of transmembrane helices. Positive or negative deviations mean that predictions shift the boundaries for helical fragments to make them longer, or shorter, respectively. All distributions in Figure 2 have multiple peaks. Our results show the main peak at the (–1; 0) region, and two smaller peaks at the values (–3) and (–6). The main peaks for the DAS and TMPRED results are both located at the value (–3), and the second peaks are at the (–1; 0) region. In other words, these three methods tend to underestimate the length of transmembrane helices. In contrast, both of the largest peaks for the SOSUI results correspond to positive deviation values, namely to (3) and (1). This distribution is almost flat in the region (–6; −1).

The average values and standard deviations calculated for prediction errors in Figure 2 are (–2.07 ± 3.86) for our results, (–2.57 ± 4.24) for TMPRED, (–1.26 ± 4.62) for SOSUI and (–4.31 ± 4.65) for DAS. According to these numbers, the average results of all four prediction methods are not statistically different. It is obvious, however, that these numbers are not very informative in the case of multipeak distributions. In our view, Figure 3 is more informative; it depicts histograms of probabilities to predict the actual boundaries of transmembrane helical fragments by four discussed methods with the accuracy of ±2, ±4 and ±6 residues, respectively. (These probabilities could be regarded as the values of integrals over the curves in Figure 2 at the corresponding intervals.) The histograms clearly show that our procedure is significantly more likely to give an accurate prediction within ±2 residues than is any of the three statistical approaches. In this respect, our predictions also outperform the equally weighted combined predictions of all three statistical approaches (the last histogram in Figure 3). This advantage is less pronounced at the error level of ±4 residues and completely disappears at the error level of ±6 residues (Figure 3). Keeping in mind that locating transmembrane fragments in the sequence is just the first rough step in the process of 3D modeling of transmembrane proteins, the

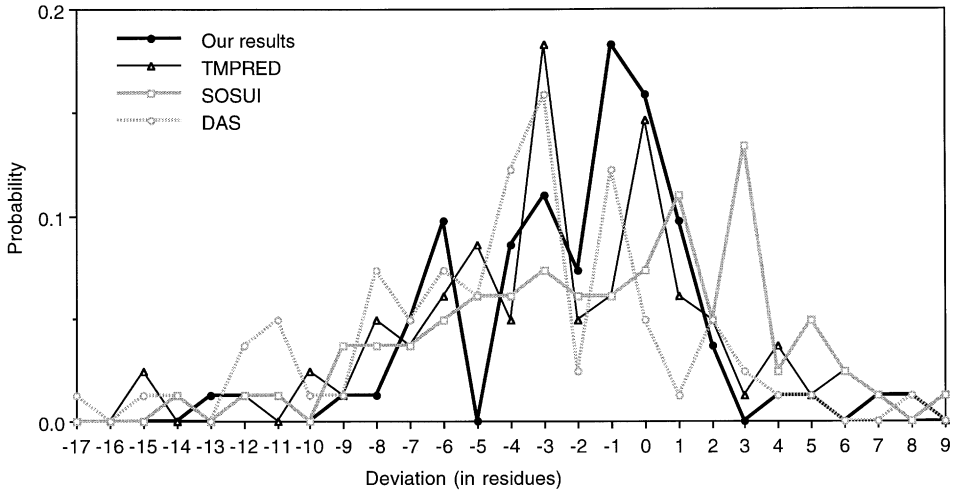

Fig. 2. Distributions of deviations between predicted and experimentally determined boundaries for transmembrane helices in PRC, bacteriorhodopsin and OCC.

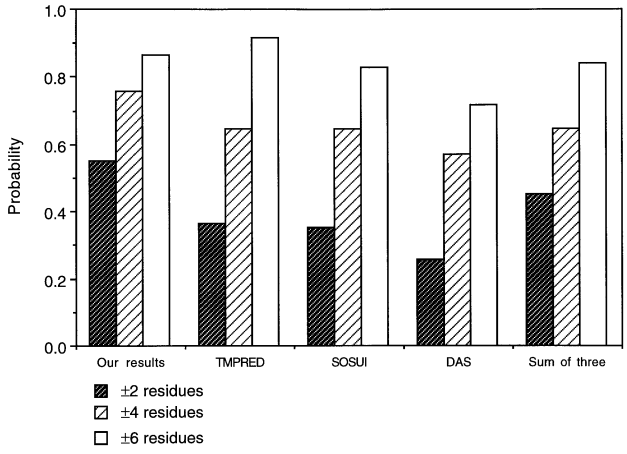

Fig. 3. Probability of predicting helix endpoints within ±2, ±4 and ±6 residues of experimental location, employing our approach and three statistical approaches separately and as a combination.

accuracy of ±0.5 helical turns (i.e. ±2 residues) may be considered fairly satisfactory.

Conclusions

We have developed a novel procedure for predicting possible breaks in transmembrane helices. The procedure is conceptually simple and does not require sophisticated statistical approaches. Most important, it does not require any training sets, since our results depend solely on the particular amino acid sequence. Despite the rather rough initial model employed, the obtained results are certainly good enough to serve as a starting point for further refinement when more structural experimental data on transmembrane proteins become available. This procedure also provides satisfactory starting points for arranging isolated transmembrane helices across a membrane (Tseitin and Nikiforovich, 1998), as well as for computational procedures dealing with the process of helix packing, which is the obligatory next step in the 3D modeling of transmembrane proteins. Once implemented, the procedure will be available for the use of scientific community (e.g., via the Internet).

Acknowledgements

The author is grateful to Dr Garland R.Marshall and Dr Stan Galaktionov for stimulating this work and for continuous discussions. Dr Tom Blackwell is gratefully acknowledged for reading the manuscript and for discussing results. This work was partly supported by the NIH grant 2 RO1 GM48184-04A1.

References

Cserzö,M., Bernassau,J.-M., Simon,I. and Maigret,B. (1994) *J. Mol. Biol.*, **243**, 388–396.
Deber,C.M. and Li,S.-C. (1995) *Biopolymers (Peptide Science)*, **37**, 295–318.
Deisenhofer,J. and Michel,H. (1989) *Science*, **245**, 1463–1473.
Deisenhofer,J., Epp,O., Sinning,I. and Michel,H. (1995) *J. Mol. Biol.*, **246**, 429–457.
Donnelly,D. and Findlay,J.B.C. (1994) *Curr. Opin. Struct. Biol.*, **4**, 582–589.
Dunfield,L.G., Burgess,A.W. and Scheraga,H.A. (1978) *J. Phys. Chem.*, **82**, 2609–2616.
Efimov,A.V. (1986) *Molekuliarnaia Biologia*, **20**, 250–260 [in Russian].
Efimov,A.V. (1991) *Protein Engng*, **4**, 245–250.
Henderson,R. and Schertler,G.F.X. (1990) *Phil. Trans. Roy. Soc. Lond.*, **B326**, 379–389.
Iwata,S., Ostermeier,C., Ludwig,B. and Michel,H. (1995) *Nature*, **376**, 660–669.
Nemethy,G., Pottle,M.S. and Scheraga,H.A. (1983) *J. Phys. Chem.*, **87**, 1883–1887.
Nikiforovich,G.V. (1994) *Int. J. Peptide Protein Res.*, **44**, 513–531.
Nikiforovich,G.V., Hruby,V.J., Prakash,O. and Gehrig,C.A. (1991) *Biopolymers*, **31**, 941–955.
Pogozheva,I.D., Lomize,A.L. and Mosberg,H.I. (1996) In Kaumaya,P.T.P. and Hodges,R.S. (eds), *Peptides: Chemistry, Structure and Biology. Proceedings of the Fourteenth American Peptide Symposium*. Mayflower Scientific, Kingswinford, West Mids, UK, pp. 350–351.
Reithmeier,R.A.F. (1995) *Curr. Opin. Struct. Biol.*, **5**, 491–500.
Rost,B. and Sander,C. (1995) *Proteins*, **23**, 295–300.
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.
Tseitin,V.M. and Nikiforovich,G.V. (1998) In Tam,J. and Kaumaya,P.T.P. (eds), *Peptides: Chemistry, Structure, Biology, Proceedings of the Fifteenth American Peptide Symposium*. Mayflower Scientific, Kingswinford, West Mids, UK, in press.
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.
Yanagihara,N., Suwa,M. and Mitaku,S. (1989) *Biophys. Chem.*, **34**, 69–77.

Received September 16, 1997; revised October 23, 1997; accepted November 12, 1997