J. Mol. Biol. (1992) **225**, 487–494

Membrane Protein Structure Prediction

Hydrophobicity Analysis and the Positive-inside Rule

Gunnar von Heijne

Department of Molecular Biology
Karolinska Institute Center for Structural Biochemistry
NOVUM, S-141 57 Huddinge, Sweden

(Received 14 October 1991; accepted 20 January 1992)

A new strategy for predicting the topology of bacterial inner membrane proteins is proposed on the basis of hydrophobicity analysis, automatic generation of a set of possible topologies and ranking of these according to the positive-inside rule. A straightforward implementation with no attempts at optimization predicts the correct topology for 23 out of 24 inner membrane proteins with experimentally determined topologies, and correctly identifies 135 transmembrane segments with only one overprediction.

Keywords: membrane protein; protein structure; prediction

1. Introduction

The prediction of protein structure from sequence is a central problem in molecular biology. For globular proteins, even secondary structure prediction is fraught with difficulties and not very reliable (Kabsch & Sander, 1983); however, recent progress in the field of membrane protein assembly and folding has fuelled hopes that a solution to the folding problem may be within reach for this important class of proteins (von Heijne & Manoil, 1990).

The basic structural building-block in plasma membrane proteins of both prokaryotic and eukaryotic cells is the apolar, often slightly amphipathic, transmembrane α-helix (Jennings, 1989; Pattus, 1990), and the arguably most important event during the biogenesis of an integral membrane protein is the insertion of its transmembrane segment(s) into the lipid bilayer. Once this has been accomplished, the basic topology of the molecule is defined. What then remains is for the transmembrane helices to assemble into a membrane-embedded helix bundle and for the polar segments exposed outside the bilayer to fold into their proper tertiary structure. It should be noted that the situation may be very different for outer membrane proteins in Gram negative bacteria, where the canonical structural principle is that of a large, antiparallel β-barrel rather than a helical bundle (Jeanteur *et al.*, 1991; Weiss *et al.*, 1991); this latter type of protein will not be considered further here.

Thus, at least for multi-spanning (polytopic) membrane proteins with most of their mass buried within the bilayer in the form of transmembrane helices, the insertion event is decisive. For structure prediction, this means that a good part of the problem is solved if the transmembrane organization of the chain can be effectively calculated from the amino acid sequence. At the present state of the art, this is best attempted using some method of hydrophobicity analysis, where the amino acid sequence is scanned to locate segments rich in apolar residues. Typically, such an analysis will identify some segments of such high average hydrophobicity that they can be considered "certain" transmembrane segments, but will also produce one or more candidate segments of intermediate average hydrophobicity that cannot be confidently predicted as transmembrane.

Clearly, the ultimate hydrophobicity analysis method would be able to discriminate perfectly between transmembrane and non-membrane segments in any protein chain; this might be accomplished by better algorithms and better hydrophobicity scales. Here, however, I suggest a complementary approach: to "bootstrap" the output from a standard hydrophobicity analysis by a subsequent step of charge-bias analysis that ranks all possible structures on the basis of their conformity with the positive-inside rule; i.e. on the observation that positively charged amino acids are manifold more abundant in cytoplasmic, as compared to periplasmic, segments of integral membrane proteins (von Heijne, 1986; von Heijne & Gavel, 1988; Gavel *et al.*, 1991) and that positively charged residues can be used to manipulate the

0022-2836/92/100487-08 $03.00/0

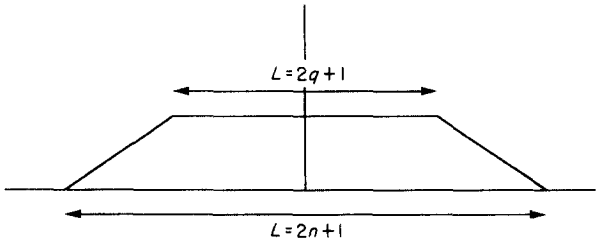

Figure 1. Sliding window used in the hydrophobicity analysis. For a given window-position, the hydrophobicity value h_i for each residue in the window is multiplied by the corresponding window-weight, w_i, and the sum over the window is taken. w_i is given by $w_i = \{i/S$ for $1 \leq i \leq n-q+1$; $(n-q+1)/S$ for $(n-q+1) < i < (n+q+1)$; $(2n+2-i)/S$ for $(n+q+1) \leq i \leq 2n+1\}$, where $S = (1+n)^2 - q^2$ is a normalization factor to make.

$$\sum_{i=1}^{2n+1} w_i = 1$$

In the calculations reported here, $n = 10$ and $q = 5$.

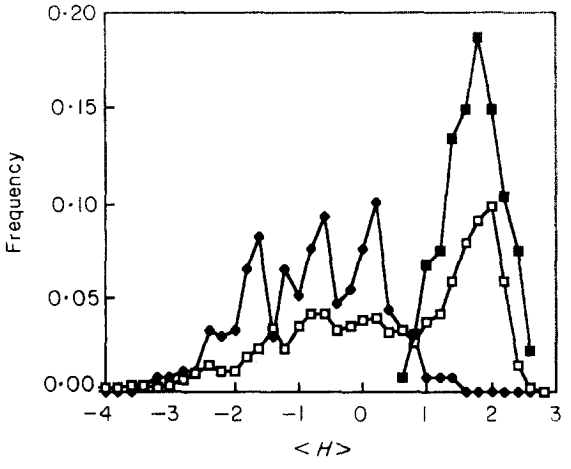

Figure 2. Peak-height distributions derived from hydrophobicity analysis of 92 bacterial inner membrane proteins (□), 25 periplasmic *Escherichia coli* proteins (◆), and the 135 transmembrane segments identified in the 24 bacterial inner membrane proteins discussed in the text (■).

topology of such proteins (Boyd & Beckwith, 1990; Dalbey, 1990; von Heijne & Manoil, 1990). As shown below, this strategy leads to a considerable improvement in prediction accuracy, at least for bacterial inner membrane proteins.

2. Methods

(a) *Collection of sequence data*

Twenty-four bacterial inner membrane proteins of known sequence and with experimentally well-characterized topologies were collected from the literature (Table 1). In most cases, the topology has been determined by fusion protein analysis (Manoil & Beckwith, 1986). Although this method seems to be a very reliable way of obtaining topological information, it should be kept in mind that the 3-dimensional structure has been determined for only 3 of the proteins listed in Table 1, namely bacteriorhodopsin, and the recation center L and M subunits.

(b) *Hydrophobicity analysis*

Hydrophobicity analysis was carried out using the GES-scale (Engelman *et al.*, 1986) and a trapezoid sliding window composed of a central, 11-residue rectangular section and 2 flanking wedge-like sections, each 5 residues long (Fig.1). This window-shape was chosen to combine the favorable noise-reduction of a triangular window (Claverie & Daulmiere, 1991) with a physically more realistic rectangular window representing the central apolar part of a lipid bilayer. In this way, one obtains a physically reasonable, soft transition from the apolar interior to the polar surface of the membrane.

From the hydrophobicity profile, candidate transmembrane segments were extracted automatically by a procedure that 1st identifies the highest peak, records its average hydrophobicity $\langle H \rangle$, and removes the part of the profile that corresponds to the 21 residues in this segment. This is repeated until either no stretch longer than 20 residues remains, or until no peak higher than the lower cutoff ($\langle H \rangle = 0{\cdot}5$) remains. All peaks with $\langle H \rangle \geq 1{\cdot}0$ are considered "certain" transmembrane segments, and all with $0{\cdot}5 \leq \langle H \rangle < 1{\cdot}0$ are considered putative candidates. These cutoff values were chosen on the basis of an initial analysis of the bimodal peak-height distribution for the whole sequence sample obtained with no cutoffs (Fig. 2: see von Heijne, 1986).

Finally, all possible topologies that include the certain transmembrane segments and either include or exclude each of the candidate segments were automatically generated, and the difference in the number of positively charged amino acid residues (Arg and Lys) between the 2 sides of each structure was calculated. If the penultimate N-terminal amino acid was neither Arg, Lys, Leu, Phe nor Ile (i.e. residues that block f-Met removal in prokaryotes (Flinta *et al.*, 1986)), an additional positive charge representing the free N-terminal amino group was added to the 1st polar segment. Polar segments longer than 70 residues were not included in the charge-bias calculation, since their content of charged residues does not seem to be dependent on their cytoplasmic or periplasmic location (von Heijne & Gavel, 1988). The possible topologies were then ranked in decreasing order of charge bias, and their orientation was predicted as the one with the more highly charged side facing the cytoplasm. A program called TOP-PRED (TOPology PREDiction program, written in THINK Pascal for Macintosh computers) implementing these steps is available upon request.

3. Results

(a) *The positive-inside rule holds for all exposed loops in bacterial inner membrane proteins*

The positive-inside rule was originally suggested by the observation that positively charged amino acid residues (Arg and Lys) are much more abundant in cytoplasmic as compared to periplasmic regions of bacterial inner membrane proteins (von Heijne, 1986), and has since been confirmed by a number of experimental as well as further statistical studies (von Heijne, 1989; Boyd & Beckwith, 1990; Dalbey, 1990; Nilsson & von Heijne, 1990; von Heijne & Manoil, 1990). However, to be able to make use of this rule for prediction purposes, it was

Table 1

Membrane segments identified by hydrophobicity and charge-bias analysis

Protein	Membrane segments	Segment category†	Charge bias‡	Reference
Bacteriorhodopsin	22–42	c+	−2, 0, +3	Henderson *et al.* (1990)
	57–77	c+	−5*	
	96–116	t+		
	119–139	c+		
	151–171	c+		
	189–209	c+		
	217–237	t+		
ColA immunity protein	17–37	c+	+4*	Geli *et al.* (1989)
	72–92	c+		
	104–124	c+		
	143–163	c+		
ColE1 immunity protein	6–26	c+	−5, +7*	Song & Cramer (1991)
	37–57	t+		
	90–110	c+		
CyoA	9–29	c+	+7*	Chepuri & Gennis (1990)
	46–66	c+		
	88–108	c+		
CyoB	16–36	c+	−16*, +6,	Chepuri & Gennis (1990)
	57–77	c+	−6, −4	
	105–125	c+		
	143–163	c+		
	194–214	c+		
	235–255	c+		
	265–285	t−		
	286–306	c+		
	313–333	c+		
	342–362	c+		
	380–400	c+		
	421–441	c+		
	456–476	c+		
	494–514	c+		
	539–559	t−		
	590–610	c+		
	611–631	c+		
CyoC	35–55	c+	+6*	Chepuri & Gennis (1990)
	71–91	c+		
	97–117	c+		
	145–165	c+		
	184–204	c+		
CyoD	19–39	c+	+3*	Chepuri & Gennis (1990)
	40–60	c+		
	78–98	c+		
CyoE	11–31	c+	+19*	Chepuri & Gennis (1990)
	36–56	c+		
	85–105	c+		
	107–127	c+		
	133–153	c+		
	160–180	c+		
	208–228	c+		
	231–251	c+		
	264–284	c+		
CytB-558	12–32	c+	+4*	Fridén (1989)
	60–80	c+		
	93–113	c+		
	137–157	c+		
	178–198	c+		
LacY	7–27	c+	+9, +17*	Calamia & Manoil (1990)
	45–65	c+		
	78–98	c+		
	102–122	c+		
	145–165	c+		
	167–187	c+		
	219–239	c+		
	263–283	c+		
	291–311	c+		
	315–335	t+		
	349–369	c+		
	382–402	c+		
Phage M13 procoat protein	5–25	c+	+6*	Kuhn *et al.* (1986)
	44–64	c+		

Table 1
Contd.

Protein	Membrane segments	Segment category†	Charge bias‡	Reference
MalF	17–37	c+	+11*, +8	Froshauer *et al.* (1988)
	39–59	c+		
	72–92	c+		
	276–296	c+		
	319–339	c+		
	346–366	t−		
	372–392	c+		
	417–437	c+		
	484–504	c+		
Phage pf3 coat protein	17–37	c+	−1*	Rohrer & Kuhn (1990)
Lep	2–22	c+	−8, −2	Moore & Miura (1987)
	60–80	c+		
	83–103	c−		
	254–274	t−		
Lsp	4–24	c+	−1, +3*	Munoa *et al.* (1991)
	69–89	c+		
	94–114	t+		
	139–159	c+		
Light-harvesting complex LH1	20–40	c+	+4*	Drews (1985)
Light-harvesting complex LH2	25–45	c+	+1*	Drews (1985)
Reaction center complex	29–49	c+	+8*, +2	Michel *et al.* (1986)
L-subunit	83–103	c+		
	114–134	c+		
	135–155	t−		
	174–194	c+		
	232–252	c+		
Reaction center complex	54–74	c+	+9*, +3	Michel *et al.* (1986)
M-subunit	109–139	c+		
	144–164	c+		
	170–190	t−		
	199–219	c+		
	166–286	c+		
SecE	16–36	c+	+6*	Schatz *et al.* (1989)
	45–65	c+		
	91–111	c+		
SecY	23–43	c+	+26*, +8	Akiyama & Ito (1987)
	75–95	c+		
	122–142	c+		
	155–175	c+		
	184–204	c+		
	217–237	c+		
	271–291	c−		
	292–312	c+		
	317–337	c+		
	371–391	c+		
	397–417	c+		
TetA	7–27	c+	+13*	Eckert & Beck (1989)
	42–62	c+		
	81–101	c+		
	102–122	c+		
	133–153	c+		
	159–179	c+		
	201–221	c+		
	240–260	c+		
	276–296	c+		
	298–318	c+		
	337–357	c+		
	369–389	c+		
UhpT	25–45	c+	+27*	Lloyd & Kadner (1990)
	59–79	c+		
	97–117	c+		
	118–138	c+		
	167–187	c+		
	190–210	c+		

Table 1
Contd.

Protein	Membrane segments	Segment category†	Charge bias‡	Reference
	299–319	c+		
	326–346	c+		
	351–371	c+		
	405–425	c+		
	427–447	c+		
UncE	11–31	c+	−2*	Deckers-Hebestreit & Altendorf (1986)
	59–79	c+		

†c, certain; t, tentative; +, a membrane segment present in the known structure; −, a membrane segment not present in the known structure.

‡The total difference in the number of Arg+Lys residues between the 2 sides of the structure for all possible topologies generated as described in Methods are given, with the value corresponding to the known topology marked with an asterisk. Positive values correspond to topologies where the N terminus is predicted to be cytoplasmic, negative values to topologies predicted to have the N terminus facing the periplasm.

important to establish that it holds for all segments of a transmembrane protein, and not just for the N-terminal or C-terminal ends, say.

To this end, 24 bacterial inner membrane proteins of known sequence and with experimentally mapped topologies were analyzed (see Table 1). For each protein, the number of Arg and Lys residues in each cytoplasmic and periplasmic segment was recorded, and the counts were pooled according to the position of the segment (counting from the N terminus) and its cytoplasmic or periplasmic location. As shown in Figure 3, the bias in the distribution of Arg+Lys is equally strong throughout the sequences; i.e. all cytoplasmic loops have a similarly high content (~15%) whereas all periplasmic loops have a low content (~5%). It is thus clear that all parts of a multi-spanning inner membrane protein conform to the positive-inside rule.

(b) *The positive-inside rule can be used to improve the prediction of transmembrane segments*

Since the positive-inside rule apparently applies to any part of a bacterial inner membrane protein, it should provide a strong criterion for testing whether a putative transmembrane topology is likely to be correct or not. One is thus led to consider the following procedure. First, make a list of all possible transmembrane segments in the protein based on some standard method of hydrophobicity analysis and a liberal cutoff criterion; second, decide which of these candidates are certain and which must be regarded as tentative based on a more stringent cutoff criterion; third, construct all possible transmembrane structures always including the certain candidates but either including or excluding each of the tentative segments; and fourth, rank these structures according to their degree of bias in the distribution of positively charged residues. The rationale behind the last step, is that any structure with an incorrect number of transmembrane segments would have one or more polar domains placed on the wrong side of the membrane, and hence would be likely to have a smaller charge bias than the correct structure.

As an example, consider the LacY protein, which is known to have a cytoplasmic N terminus followed by 12 transmembrane segments (Calamia & Manoil, 1990). The hydrophobicity plot shown in Figure 4(a), predicts 11 certain segments and one putative candidate. Thus, two topologies can be generated, one excluding and one including the putative candidate (Fig. 4(b)). The former has a bias in the distribution of positively charged residues of nine, the latter of 17, and it is clear from the Figure that the C-terminal part of the second, correct model has a much better correspondence with the positive-inside rule.

To test this strategy more thoroughly, hydrophobicity analysis was carried out on the 24 proteins listed in Table 1 as described in Methods, with a lower cutoff set to $\langle H \rangle = 0{\cdot}5$ and a higher cutoff set

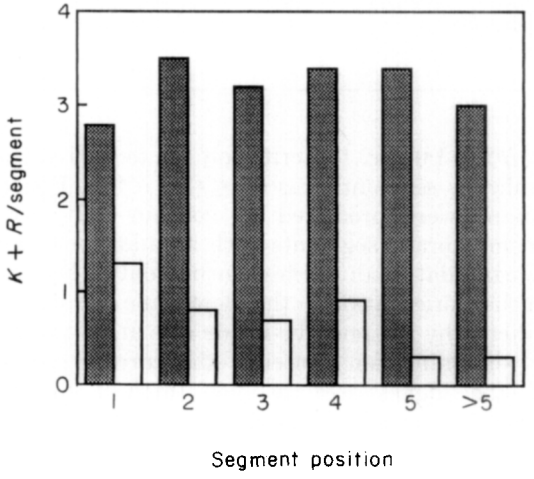

Figure 3. Mean number of Lys+Arg in cytoplasmic (stippled bars) and periplasmic (open bars) polar segments as a function of the position of the segment (1st cytoplasmic segment, 2nd cytoplasmic segment, etc., counting from the N terminus) in a sample of 24 bacterial inner membrane proteins with known topology.

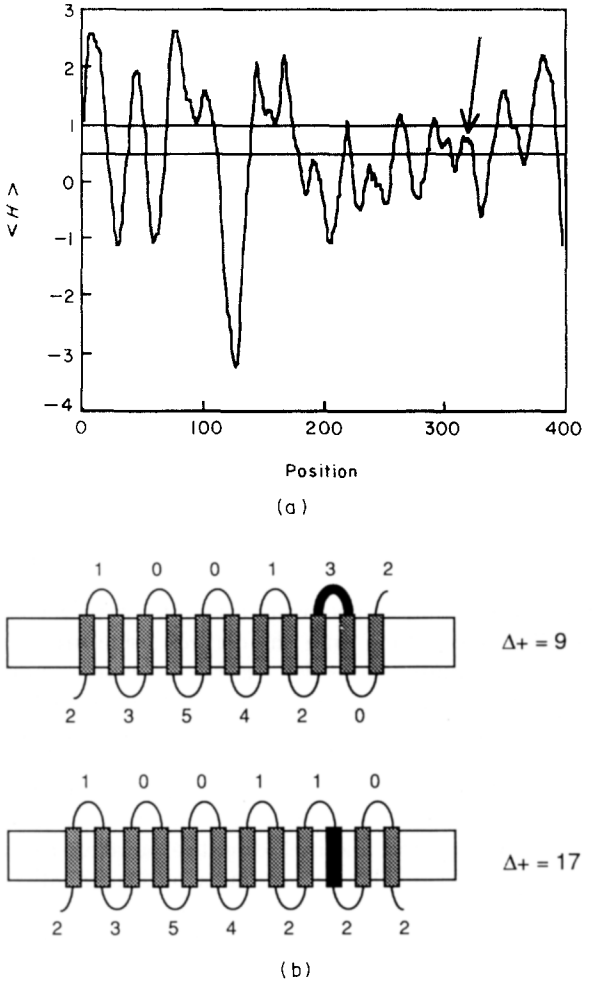

Figure 4. (a) Hydrophobicity plot for the SecY protein. The upper and lower cutoffs are marked. A tentative transmembrane segment with a mean hydrophobicity falling between the 2 cutoffs is marked by an arrow. (b) Two possible topologies for the SecY protein based on the hydrophobicity plot. The putative transmembrane segment is shown in black. The number of Arg+Lys residues is shown next to each polar segment. Note that the correct alternative (bottom, including the putative transmembrane segment) has a much higher charge-bias than the incorrect one.

to $\langle H \rangle = 1{\cdot}0$ (i.e. the criterion for tentative transmembrane segments was $0{\cdot}5 \leq \langle H \rangle < 1{\cdot}0$). Fourteen proteins were predicted to contain only certain transmembrane segments with no additional tentative candidates; all of these turned out to be correct and their orientation in the membrane was correctly predicted by the positive-inside rule in all cases. For the remaining ten proteins, the correct structure had the highest charge bias (highest difference between the number of periplasmic and cytoplasmic Arg+Lys) in eight cases (Table 1). An extra, incorrect certain transmembrane segment ($\langle H \rangle = 1{\cdot}06$) was predicted for the SecY protein; closer inspection revealed that this resulted from a "shoulder" on a higher, somewhat broad peak. Inclusion of this segment gave a topology with a charge bias of eight, and its exclusion gave a charge bias of 26. Thus, although the fully automated method goes wrong in this case (upper cutoff is set marginally too low), a subsequent check of the hydrophobicity plot would have suggested the correct topology. Finally, one protein (Lep) was erroneously and irretrievably predicted to contain one more certain transmembrane segment ($\langle H \rangle = 1{\cdot}15$) than the real protein. It should be noted that this unusually hydrophobic non-transmembrane segment has been suggested to be functionally important (Bilgin *et al.*, 1990), and that a single Asp→Leu mutation in it seems to convert it into a transmembrane segment (our unpublished data).

Thus, using the positive-inside rule as a screen, the transmembrane topologies of 22 (or, with some good-will, 23) out of 24 proteins were correctly predicted, and all of the 135 transmembrane segments were identified with only one overprediction (Lep).

(c) *Possibilities for improvements in the hydrophobicity scales*

In the above analysis, the standard GES hydrophobicity scale (Engelman *et al.*, 1986) based ultimately on physico-chemical considerations was used and no attempt at optimization was made. With this scale and through application of the positive-inside rule, 135 transmembrane segments were correctly identified in the 24 proteins analyzed. From these data, we derived a new, statistically based scale by comparing the amino acid frequencies f_i^M in the central 11-residue stretch of the transmembrane segments with the amino acid frequencies f_i^A in the non-membraneous segments (taking the membrane-embedded segments to be 17 residues long): $h_i = \ln(f_i^M/f_i^A)$, Table 2. With an upper cutoff of 0·05 and a lower of −0·1 and the

Table 2

Hydrophobicity scale based on 135 transmembrane segments from 24 bacterial inner membrane proteins

One-letter amino acid code	Hydrophobicity value
A	0·267
C	1·806
D	−2·303
E	−2·442
F	0·427
G	0·160
H	−2·189
I	0·971
K	−2·996
L	0·623
M	0·136
N	−1·988
P	−0·451
Q	−1·814
R	−2·749
S	−0·119
T	−0·083
V	0·721
W	−0·875
Y	−0·386

same sliding window as above, this scale seems to perform slightly better than the original GES-scale, with 18 out of the 24 proteins (including SecY) now being predicted to contain only certain transmembrane segments (all correct). Again, ranking based on the positive-inside rule produced the correct topology for all but one (Lep) of the remaining proteins.

4. Discussion

Two important conclusions can be drawn from the data presented here: first, the positive-inside rule applies with equal strength to all parts of polytopic bacterial transmembrane proteins, and, second, the positive-inside rule can very efficiently be used to rank possible predicted transmembrane topologies and thus in a sense to bootstrap the standard hydrophobicity analysis.

The first conclusion has certain mechanistic implications, and suggests a mode of assembly of bacterial membrane proteins where locally formed "helical hairpins" composed of two neighboring hydrophobic helices and a connecting loop containing few positively charged residues insert into the membrane more or less independently of the remaining parts of the polypeptide chain (Engelman & Steitz, 1981; von Heijne, 1986). A similar model is also suggested by recent experimental data on the topological consequences of deleting one or more transmembrane segment (Yamane *et al.*, 1990; Bibi *et al.*, 1991; Ehrmann & Beckwith, 1991; McGovern *et al.*, 1991). This is in contrast to a "linear" insertion model, where the most N-terminal transmembrane segment(s) insert first, and the rest of the chain simply follows the topological dictate of this first insertion (Wessels & Spiess, 1988; Hartmann *et al.*, 1989). It may be, though, that the latter model more accurately describes membrane protein assembly into the endoplasmic reticulum membrane, which is thought to be obligatorily co-translational and hence likely to proceed in an N to C-terminal direction (High *et al.*, 1991). It should be noted in this regard that the bias in the distribution of positively charged residues, although clearly there, is nevertheless less marked in eukaryotic plasma membrane proteins (von Heijne & Gavel, 1988).

For prediction purposes, a combination of hydrophobicity analysis and charge-bias analysis clearly improves the results, almost to the point where very little experimental work would have to be done to confirm a predicted topology. In this study, the topology of 95% of the proteins was correctly predicted by a fully automatic method, even when the initial hydrophobicity analysis method had not been consciously optimized. Thus, the topology of bacterial inner membrane proteins seems to be rather straightforward to predict directly from their sequence, setting the stage for attempts to develop methods for modeling their full three-dimensional structure.

This work was supported by grants from the Swedish Natural Sciences Research Council and the Swedish Board for Technical Development.

References

Akiyama, Y. & Ito, K. (1987). Topology analysis of the SecY protein, an integral membrane protein involved in protein export in *Escherichia coli*. *EMBO J.* **6**, 3465–3470.

Bibi, E., Verner, G., Chang, C. Y. & Kaback, H. R. (1991). Organization and stability of a polytopic membrane protein—deletion analysis of the lactose permease of *Escherichia coli*. *Proc. Nat. Acad. Sci., U.S.A.*, **88**, 7271–7275.

Bilgin, N., Lee, J. I., Zhu, H., Dalbey, R. & von Heijne, G. (1990). Mapping of catalytically important domains in *Escherichia coli* leader peptidase. *EMBO J.* **9**, 2717–2722.

Boyd, D. & Beckwith, J. (1990). The role of charged amino acids in the localization of secreted and membrane proteins. *Cell*, **62**, 1031–1033.

Calamia, J. & Manoil, C. (1990). Lac permease of *Escherichia coli*—topology and sequence elements promoting membrane insertion. *Proc. Nat. Acad. Sci., U.S.A.* **87**, 4937–4941.

Chepuri, V. & Gennis, R. B. (1990). The use of gene fusions to determine the topology of all of the subunits of the cytochrome o terminal oxidase complex of *Escherichia coli*. *J. Biol. Chem.* **265**, 12978–12986.

Claverie, J.-M. & Daulmiere, C. (1991). Smoothing profiles with sliding windows: better to wear a hat! *CABIOS* **7**, 113–115.

Dalbey, R. E. (1990). Positively charged residues are important determinants of membrane protein topology. *Trends Biochem. Sci.* **15**, 253–257.

Deckers-Hebestreit, G. & Altendorf, K. (1986). Accessibility of F_0 subunits from *Escherichia coli* ATP synthase. *Eur. J. Biochem.* **161**, 225–231.

Drews, G. (1985). Structure and functional organization of light-harvesting complexes and photochemical reaction centers in membranes of phototrophic bacteria. *Microbiol. Rev.* **49**, 59–70.

Eckert, B. & Beck, C. F. (1989). Topology of the transposon Tn*10*-encoded tetracycline resistance protein within the inner membrane of *Escherichia coli*. *J. Biol. Chem.* **264**, 11633–11670.

Ehrmann, M. & Beckwith, J. (1991). Proper insertion of a complex membrane protein in the absence of its amino-terminal export signal. *J. Biol. Chem.* **266**, 16530–16533.

Engelman, D. M. & Steitz, T. A. (1981). The spontaneous insertion of proteins into and across membranes: The helical hairpin hypothesis. *Cell*, **23**, 411–422.

Engelman, D. M., Steitz, T. A. & Goldman, A. (1986). Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biophys. Chem.* **15**, 321–353.

Flinta, C., Persson, B., Jörnvall, H. & von Heijne, G. (1986). Sequence determinants of cytosolic N-terminal protein processing. *Eur. J. Biochem.* **154**, 193–196.

Fridén, H. (1989). Studies on the cytochrome *B-558* subunit of *Bacillus subtilis* succinate:quinone oxidoreductase. PhD thesis, Lund University, Sweden.

Froshauer, S., Green, G. N., Boyd, D., McGovern, K. & Beckwith, J. (1988). Genetic analysis of the membrane insertion and topology of MalF, a cyto-

plasmic membrane protein of *Escherichia coli*. *J. Mol. Biol.* **200**, 501–511.

Gavel, Y., Steppuhn, J., Herrmann, R. & von Heijne, G. (1991). The positive-inside rule applies to thylakoid membrane proteins. *FEBS Letters*, **282**, 41–46.

Geli, V., Baty, D., Pattus, F. & Lazdunski, C. (1989). Topology and function of the integral membrane protein conferring immunity to colicin A. *Mol. Microbiol.* **3**, 679–687.

Hartmann, E., Rapoport, T. A. & Lodish, H. F. (1989). Predicting the orientation of eukaryotic membrane proteins. *Proc. Nat. Acad. Sci., U.S.A.* **86**, 5786–5790.

Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E. & Downing, K. H. (1990). A model for the structure of bacteriorhodopsin based on high resolution electron cryo-microscopy. *J. Mol. Biol.* **213**, 899–929.

High, S., Flint, N. & Dobberstein, B. (1991). Requirements for the membrane insertion of signal-anchor type proteins. *J. Cell. Biol.* **113**, 25–34.

Jeanteur, D., Lakey, J. H. & Pattus, F. (1991). The bacterial porin superfamily—sequence alignment and structure prediction. *Mol. Microbiol.* **5**, 2153–2164.

Jennings, M. L. (1989). Topography of membrane proteins. *Annu. Rev. Biochem.* **58**, 999–1027.

Kabsch, W. & Sander, C. (1983). How good are predictions of protein secondary structure? *FEBS Letters*, **155**, 179–182.

Kuhn, A., Kreil, G. & Wickner, W. (1986). Both hydrophobic domains of M13 procoat are required to initiate membrane insertion. *EMBO. J.* **5**, 3681–3685.

Lloyd, A. D. & Kadner, R. J. (1990). Topology of the *Escherichia coli uhp*T sugar-phosphate transporter analyzed by using Tn*phoA* fusions. *J. Bacteriol.* **172**, 1688–1693.

Manoil, C. & Beckwith, J. (1986). A genetic approach to analyzing membrane protein topology. *Science*, **233**, 1403–1408.

McGovern, K., Ehrmann, M. & Beckwith, J. (1991). Decoding signals for membrane protein assembly using alkaline phosphatase fusions. *EMBO J.* **10**, 2773–2782.

Michel, H., Weyer, K. A., Gruenberg, H., Dunger, I., Oesterhelt, D. & Lottspeich, F. (1986). The 'light' and 'medium' subunits of the photosynthetic reaction centre from *Rhodopseudomonas viridis*: isolation of the genes, nucleotide and amino acid sequence. *EMBO J.* **5**, 1149–1158.

Moore, K. E. & Miura, S. (1987). A small hydrophobic domain anchors leader peptidase to the cytoplasmic membrane of *Escherichia coli*. *J. Biol. Chem.* **262**, 8806–8813.

Munoa, F. J., Miller, K. W., Beers, R., Graham, M. & Wu, H. C. (1991). Membrane topology of *Escherichia coli* prolipoprotein signal peptidase (signal peptidase II). *J. Biol. Chem.* **266**, 17667–17672.

Nilsson, I. M. & von Heijne, G. (1990). Fine-tuning the topology of a polytopic membrane protein. Role of positively and negatively charged residues. *Cell*, **62**, 1135–1141.

Pattus, F. (1990). Membrane protein structure. *Curr. Opinion Cell. Biol.* **2**, 681–685.

Rohrer, J. & Kuhn, A. (1990). The function of a leader peptide in translocating charged amino acyl residues across a membrane. *Science*, **250**, 1418–1421.

Schatz, P. J., Riggs, P. D., Jacq, A., Fath, M. J. & Beckwith, J. (1989). The *secE* gene encodes an integral membrane protein required for protein export in *Escherichia coli*. *Genes Develop.* **3**, 1035–1044.

Song, H. Y. & Cramer, W. A. (1991). Membrane Topography of ColE1 gene products—the immunity protein. *J. Bacteriol.* **173**, 2935–2943.

von Heijne, G. (1986). The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J.* **5**, 3021–3027.

von Heijne, G. (1989). Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature (London)*, **341**, 456–458.

von Heijne, G. & Gavel, Y. (1988). Topogenic signals in integral membrane proteins. *Eur. J. Biochem.* **174**, 671–678.

von Heijne, G. & Manoil, C. (1990). Membrane proteins —from sequence to structure. *Protein Eng.* **4**, 109–112.

Weiss, M. S., Kreusch, A., Schiltz, E., Nestel, U., Welte, W., Weckesser, J. & Schulz, G. E. (1991). The structure of porin from *Rhodobacter capsulata* at 1.8Å resolution. *FEBS Letters*, **280**, 379–382.

Wessels, H. P. & Spiess, M. (1988). Insertion of a multispanning membrane protein occurs sequentially and requires only one signal sequence. *Cell*, **55**, 61–70.

Yamane, K., Akiyama, Y., Ito, K. & Mizushima, S. (1990). A positively charged region is a determinant of the orientation of cytoplasmic membrane proteins in *Escherichia coli*. *J. Biol. Chem.* **265**, 21166–21171.

Edited by B. W. Matthews