Biochimica et Biophysica Acta 815 (1985) 468–476
Elsevier

BBA 72604

The detection and classification of membrane-spanning proteins

Petr Klein *, Minoru Kanehisa and Charles DeLisi **

Laboratory of Mathematical Biology, National Cancer Institute, National Institues of Health, Bethesda, MD 20205 (U.S.A.)

(Received December 27th, 1984)

Key words: Membrane protein; Membrane-spanning segment; Discriminant analysis; Protein structure prediction

Discriminant analysis can be used to precisely classify membrane proteins as integral or peripheral and to estimate the odds that the classification is correct. Specifically, using 102 membrane proteins from the National Biomedical Research Foundation (NBRF) database we find that discrimination between integral and peripheral membrane proteins can be achieved with 99% reliability. Hydrophobic segments of integral membrane proteins can also be distinguished from interior segments of globular soluble proteins with better than 95% reliability. We also propose a procedure for determining boundaries of membrane-spanning segments and apply it to several integral membrane proteins. For the limited data available (such as on transplantation antigens), the residues at the boundaries of a membrane-spanning segment are predictable to within the error inherent in the concept of boundary. As a specific indication of resolution, seven membrane-spanning segments of bacteriorhodopsin are resolved with no information other than sequence, and the predicted boundary residues agree with the experimental data on proteolytic cleavage sites. Several definitive but yet to be tested predictions are also made, and the relation to other predictive methods is briefly discussed. A computer program in FORTRAN for prediction of membrane-spanning segments is available from the authors.

Introduction

Interest in computational methods for predicting the structure and disposition of membrane-associated proteins of known sequence arises in part from difficulties in applying standard biophysical tools for structural analysis, and, in part, from the recent rapid growth in sequence information. The general principles that determine the environment and structure most favorable to a given sequence appear to be understood in broad outline, if not in detail, and can often be used to advantage in planning experiments [1]. Nevertheless, serious difficulties remain. Rigorous comparison of the most popular methods for predicting even relatively simple secondary structural categories (α-helix, β-sheet or neither) against experimental data shows that little more than 50% of the residues in a sequence are predicted correctly [2,3]. The simpler problem of distinguishing integral (i.e., membrane spanning) segments, from segments in peripheral membrane proteins, or integral segments from segments buried in soluble proteins is also not fully understood (see, for example, Ref. 4).

In this paper we focus on primary structure analysis and introduce technical improvements in statistical methods for classifying membrane proteins. Although the development presented here is similar to those of previous authors in its use of various functions of sequence hydrophobicity as the primary variable for classifying the location of

* Present address: Division of Biological Sciences, National Research Council, Ottawa, Ontario, Canada K1A OR6.
** To whom reprint requests should be addressed.

a sequence, it differs in the method of analysis used to arrive at allocation rules and in the size of the data base upon which the rules are developed. These differences lead to higher levels of precision which, we will show, in some instances, translate into important predictive differences. They also permit the assignment of estimates on the reliability of the predictions, and this in turn leads to a new procedure for distinguishing membrane-spanning segments from segments buried in a soluble protein. FInally, in addition to comparing predictions of the method with available data, we make several unambiguous, but yet to be tested predictions, including positional classification of oncogene sequences, the disposition of cytochrome oxidase subunits and segments with respect to the membrane, and the membrane-spanning portion of the T-cell receptor beta chain.

Methods

Discriminant analysis allows the classification of any membrane protein as integral (intrinsic, I) or peripheral (extrinsic, E) according to the characteristics of its sequence. The characteristics are sometimes simple such as maximum local hydrophobicity, or they might be more subtle and obtained by a Fourier analysis of sequence properties as indicated previously [3–5 and Refs. cited therein). Leaving aside for the moment their precise nature, we let x denote the vector of their values, and let $P(I/x)$ and $P(E/x)$ be, respectively, the conditional probabilities that a protein is integral or peripheral, given that it has attribute vector x. Further, let $P(I)$ and $P(E)$ denote the prior probabilities of intrinsic and extrinsic membrane proteins, these being estimated by the relative abundance of each group. Then a sequence is assigned to E if

$$P(E/x) > P(I/x). \tag{1}$$

By invoking Bayes theorem

$$P(I/x) = P(x/I)P(I)/[P(x/I)P(I) + P(x/E)P(E)] \text{ and}$$

$$P(E/x) = P(x/E)P(E)/[P(x/I)P(I) + P(x/E)P(E)], \tag{2}$$

condition (1) is equivalent to

$$P(E)P(x/E) > P(I)P(x/I), \tag{3}$$

where $P(x/I)$ and $P(x/E)$ are determined from estimates of probability distributions of attributes in the two groups of proteins. If probability density functions are known, $P(x/I)$ and $P(x/E)$ are replaced by the values of the density functions $f(x/I)$ and $f(x/E)$, respectively [6].

Linear discriminant rules arise when the density functions are multivariate normal with different means but equal variance matrices. When the assumption of equality of variance matrices is relaxed, the discriminant function is quadratic in the attributes [7]. Errors are estimated by allocating each membrane protein according to the discriminant rule and counting the number of misallocations.

It is evident from the above development that in order to distinguish two categories on the basis of some set of shared properties, we need to be able to quantitate those properties, and to find their joint distribution for each of the categories. An example relevant to integral/peripheral discrimination will help fix ideas. Consider a protein of length L, and to each of the $L - l + 1$ blocks of contiguous residues of length l, assign an average hydrophobicity to that block. Find the block of maximum hydrophobicity

$$\max H(l) = \max_{k=1,\ldots,L-l+1} (1/l) \sum_{i=k}^{k+l-1} H_i, \tag{4}$$

repeat the procedure for each membrane protein in the training set, and thus determine the way in which max $H(l)$ is distributed within the integral and peripheral membrane proteins. A typical discriminant rule would be of the form: allocate a test protein to the peripheral category if Eqn. 3 is satisfied with $x = \max H(l)$, otherwise allocate to the integral category.

The success of the above procedure relies on finding suitable parameters and a reliable method for quantitating them, and on an adequate training set of proteins upon which to develop the rules. We used 66 integral membrane proteins (i.e., proteins having at least one buried helical segment) [8–17] and 36 peripheral [10,14,18,19] membrane proteins (Tables I and II) from the NBRF protein

TABLE I

INTEGRAL PROTEINS USED TO DERIVE EQNS. 5 AND 6

Numbers in square brackets indicate how many different sources of the same protein were taken.

cytochrome *c*1
cytochrome *b* [54]
cytochrome *b*5 [5]
cytochrome oxidase polypeptide I [4]
cytochrome oxidase polypeptide II [5]
cytochrome oxidase polypeptide III [4]
cytochrome oxidase polypeptide IV
cytochorme oxidase polypeptide VIIIa
isomaltase
neuraminidase
ATPase lipid-binding protein [5]
ATPase lipid-binding protein (c chain)
ATPase protein 6 [3]
ATPase protein 3 or 6
ATPase protein 6 (a chain)
ATPase b chain
Ig γ-2b chain C region membrane-bound segment
Ig γ-1 chain C region membrane-bound segment
Ig μ chain C region membrane-bound segment
histocompatibility antigen H-2Ld heavy chain
histocompatibility antigen H-2Kb heavy chain
histocompatibility antigen HLA-DR α chain precursor
histocompatibility antigen-related DC1 α chain
rhodopsin
hepatic lectin
acetylcholine receptor protein, 40k α chain
acetylcholine receptor protein, 50k β chain
acetylcholine receptor protein, 60k γ chain
acetylcholine receptor protein, 65k δ chain
glycophorin A (human)
glycophorin (pig)
cytochrome *b* mRNA maturase
histidine permease gene M protein
histidine permease membrane gene Q protein
lactose permease
bacteriorhodopsin precursor
light-harvesting protein
probable colicin E1 immunity protein
hemagglutinin precursor
coat and membrane polyprotein
spike glycoprotein (G)
coat protein b precursor

TABLE II

PERIPHERAL PROTEINS USED TO DERIVE EQNS. 5 AND 6

Numbers ins square brackets indicate how many different sources of the same protein were taken.

cytochrome *c* [5]
ribonuclease
ribonuclease T1
ribonuclease U2
α-lactalbumin [3]
ATPase β chain [2]
ATPase δ chain
ATPase ε chain [3]
ATPase γ chain (version 1)
ATPase γ chain (version 2)
desmin [2]
myosin heavy chain, skeletal muscle
actin 3 (fruit fly)
actin [3] (yeasts and soy bean)
myosin A1 and A2 catalytic light chains, skeletal muscle
myosin L1 (A1) and L4 (A2) catalytic light chains, skeletal muscle
myelin basic protein [2]
myelin basic protein S
sulfate-binding protein
phosphocarrier protein HPr
LIV-binding protein
L-arabinose-binding protein
D-galactose-binding protein

sequence database [13] to derive the rules, and additional 167 proteins to test them. In order to avoid introducing biases that would result from too many similar sequences from different species, we did not include all known membrane proteins in the training set. The test set, however, did contain some proteins homologous to those in the training set (and is therefore not fully orthogonal to the training set) as well as additional non-homologous proteins.

Results

Allocation rules

Various methods have been proposed for assigning hydrophobicity values, each having its own strengths and limitations (see for example the review by Edsall and McKenzie [20]). We used and compared four of these: the scales of Nozaki and Tanford [21], Hopp and Woods [22], Kyte and Doolittle [23] and Eisenberg et al. [24] (Table III).

The ability to develop a rule that will cleanly partition all membrane-associated proteins into integral and peripheral, depends only slightly on the choice of hydrophobicity scale and on whether the

TABLE III

ERRORS OF DISCRIMINATION WITH DIFFERENT ATTRIBUTES

Attribute	Hydrophobicity scale	Error (%) Discriminant analysis	
		Quadratic	Linear
(1) max $H(17)$	Kyte-Doolittle [23]	0	1
max $H(19)$		1	3
average hydrophobicity		22	25
(2) max $H(17)$	Nozaki-Tanford [21]	3	3
max $H(19)$		2	5
average hydrophobicity		12	12
(3) max $H(17)$	Hopp-Woods [22]	2	5
max $H(19)$		5	5
average hydrophobicity		16	15
(4) max $H(17)$	Eisenberg et al. [24]	0	2
max $H(19)$		3	2
average hydrophobicity		20	26

discriminant function is linear or non-linear, but it is heavily dependent on the particular function of hydrophobicity used (Table III). The training set of 102 proteins can be partitioned into integral and peripheral categories with no misallocations using the Kyte-Doolittle scale and a quadratic function of max $H(17)$. Thus, a protein is allocated to the peripheral category (E) if

$$1.05\,x^2 - 12.30\,x + 17.49 > 0 \qquad (5)$$

where x is max $H(17)$ according to the scale of Kyte and Doolittle; otherwise it is allocated to the integral category (I). The odds, $P(E/x):P(I/x)$, are given by e^b where b is the left-hand side of the inequality 5. When the discrimination error is plotted against block length (Fig. 1), we find that there is also no error when the block of length 9 is used. We chose a length of 17 because it is closer to the minimum length of an α-helical membrane-spanning segment, and provides us with better estimate of the residues at the boundaries, as discussed below.

The corresponding linear inequality which allocates 101 out of 102 membrane proteins correctly is

$$-9.02\,x + 14.27 > 0 \qquad (6)$$

The odds are again given by the exponential function of the left-hand side of this inequality.

Of the 167 membrane proteins in the test set (not shown), the linear rule correctly allocates 83 of the 84 proteins believed to be integral, and 81 out of the 83 believed to be peripheral.

The odds parameter turns out to play a useful role in discriminating between integral membrane segments and other hydrophobic segments, presumably those buried in the interior of soluble proteins. This becomes apparent when Eqn. 6 is

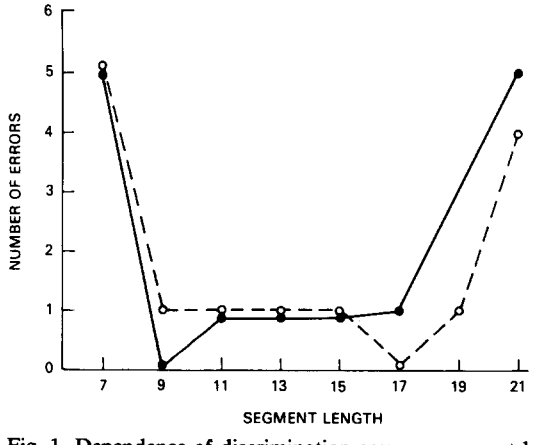

Fig. 1. Dependence of discrimination error on segment length using linear (●) and quadratic (○) discriminant analysis.

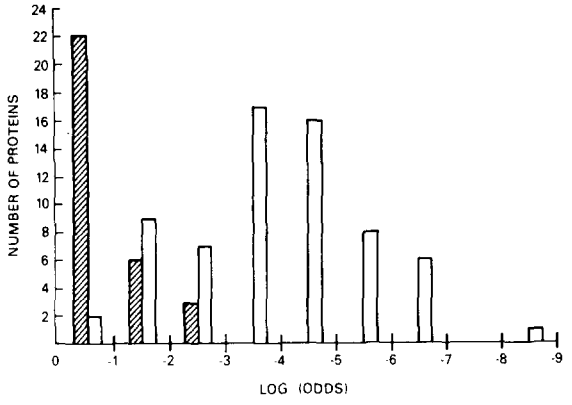

Fig. 2. The number of proteins classified as integral as a function of odds of correct allocation. Integral proteins (unhatched bars) are from Table I. The remaining proteins (hatched bars) were obtained by a systematic search of some 2000 proteins in the NBRF database, excluding proteins believed to be membrane-associated.

applied to all proteins in the NBRF database. The few soluble proteins that classify as integral membrane do so with low odds (Fig. 2). In particular, the two distributions (number of integral proteins as a function of odds and number of non-membrane-associated proteins classified as integral as a function of odds) overlap very little. If, for instance, we require odds of at least 1:10 for a protein to be classified as integral, only two membrane proteins out of 66 from Table I do not satisfy this requirement, and only 9 soluble proteins (seven lactate dehydrogenases and two hemoglobins) from the entire NBRF database do satisfy it.

For the purpose of completeness, we note that the discriminant rules themselves have a statistical uncertainty in their numerical coefficients. Specifically, for the linear rule which can be formulated as $x < 1.58$, 95% confidence in the numerical factor occurs at 1.3 and 1.9.

There are two previous methods which bear a strong relation to different aspects of the work presented here, one by Kyte and Doolittle [23] and one by Barrantes [19]; we comment on these briefly to provide perspective. Barrantes developed a discriminant function based on amino acid composition rather than sequence properties, and hence it is, not unexpectedly, considerably less accurate than those presented here. In particular the allocatioan error for the proteins in Tables I and II is larger than 10%. Kyte and Doolittle, on the other hand, did not use discriminant analysis, but arrived empirically at a rule by comparing the most hydrophobic regions of 9 membrane spanning segments and 9 soluble globular proteins. Their suggested rule is $x < 1.6$ where x is max $H(19)$ rather than our max $H(17)$. The resemblance to Eqn. 6 is striking, though perhaps fortuitous, since the two protein categories used to obtain it (integral and buried), are different from the categories used to obtain Eqn. 6.

Boundary predictions

The above procedure allocates known membrane proteins to integral and peripheral categories and provides estimates of the odds for correct allocation, but it does not explicitly consider predicting the length of the membrane-spanning segment, nor the identification of all possible membrane spanning segments. With respect to the latter problem, if a sequence is allocated to the integral category, the segment of maximum hydrophobicity is stored, and then, in effect, removed from the sequence by assigning negative values of hydrophobicity to all residues in that segment. The remaining sequence is again tested and, if it is integral, the procedure is repeated, with the process continuing until a sequence remains that is finally classified as peripheral. In this way, all membrane spanning segments (or at least their essential parts of length 17) are identified. The ends of these segments (i.e., the first and seventeenth residues) can be viewed as inner boundaries i_1, i_2 of the membrane-spanning segments, where counting begins at the amino terminus. In other words, the membrane-spanning segments are considered to be at least 17 residues long and to encompass the segment of maximum hydrophobicity. To obtain a consistent upper bound on the length of the membrane-spanning segment (outer boundaries), we take the first segment of 17 to the left of the inner bounds (its boundaries are $i_1 - 1$, $i_2 - 1$) and calculate the average hydrophobicity. If this value classifies as integral, we take the next segment to the left and so on, until a segment is reached that classifies as peripheral. The left boundary of the last segment classified as integral is referred to as the outer boundary o_1. By a

similar process we obtain o_2, the outer boundary to the right of i_2.

We assume that the actual left-hand boundary falls between o_1 and i_1, and that the actual right-hand boundary falls between i_2 and o_2. Our experience indicates that the inner and outer estimates usually do not differ by more than 6 or 7 residues and are often closer. Consequently, the prediction cannot be too far wrong if the boundary residues b_1 and b_2 are taken as half way between the inner and outer estimates, with any fractional ambiguities resolved by taking the residues that maximize the length of the membrane-spanning segment.

For proteins that span a membrane several times, the two intervals defined by outer boundaries of neighboring membrane-spanning segments could, in principle, overlap. If the intervals bounded by b_1 and b_2 also overlapped, we would shorten both membrane-spanning segments as little as possible to eliminate the overlap.

One further observation is worth indicating in connection with this assignment. The ratio of hydrophilic to hydrophobic residues in the neighborhood of o_j ($j = 1, 2$) shows a striking peak at positions $o_1 - 1$ and $o_2 + 1$ (Fig. 3). These strong peaks indicate that the most hydrophilic residues tend to occur just outside the lipid-polar interface, perhaps providing an energy barrier that stabilizes the segment against movement transverse to the membrane. These peaks suggest that o_1 and o_2 occur beyond rather than within the membrane boundaries, and, consequently, may well be regarded as true outer bounds.

TABLE IV

PREDICTION OF MEMBRANE LOCATION OF ONCOGENE PRODUCTS

Oncogene	peripheral (E) or integral (I)	$r = \frac{P(E/x)}{P(I/x)}$
abl	E	9:5
fes (strain Gardner-Arnstein)	E	8:1
fes (strain Snyder-Theilen)	E	8:1
fps	E	2:1
v-src (strain Prague C)	E	22:5
v-src (strain Schmidt-Ruppin)	E	22:5
c-src	E	12:1
yes	E	10:1
v-fgr	E	17:1
erb-B	I	1:10000
v-fms	I	1:100000
v-raf	E	58:1
c-mos (human)	E	71:1
c-mos (mouse)	E	87:1
v-mos	E	87:1

Applications

As an example of an application of the distinction between integral and peripheral proteins, we have applied Eqn. 6 to all oncogene products that are believed to function as protein kinase (Table IV) [25]. The products of fms and erb-B, which are known to possess a membrane-spanning segment [26], are clearly classified as integral; all others are predicted to be peripheral. For abl and fbs, the odds are not, however, sufficiently high to make the allocation unambiguously. These observations,

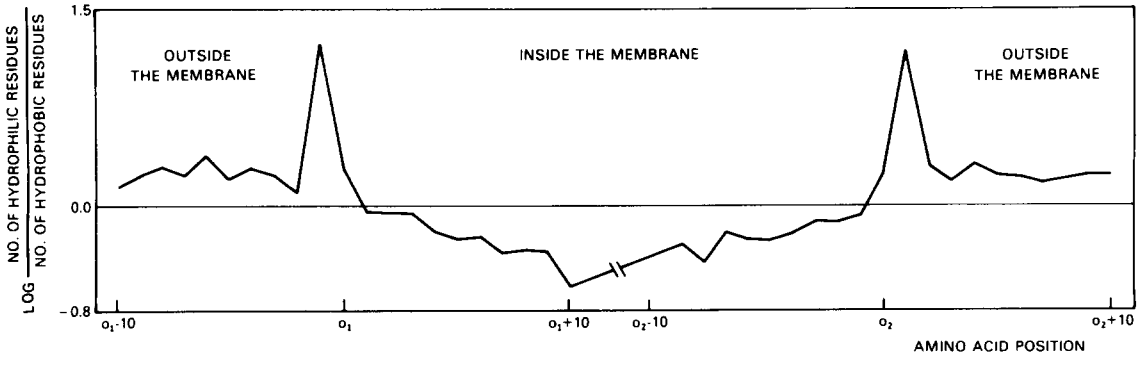

Fig. 3. The logarithm of the ratio of the number of hydrophilic to the number of hydrophobic residues appearing in integral membrane proteins at 10 positions to the left and right of the predicted outer boundaries o_1 and o_2. Positive values indicate prevailing hydrophilicity, negative values prevailing hydrophobicity. The peaks are at positions $o_1 - 1$ and $o_2 + 1$.

including the spread in odds, suggest pronounced heterogeneity in the functional kinases, some probably residing at or near one of the membrane surfaces, others likely being well within the membrane.

An important protein for which different methods make different predictions is the T-cell receptor recently sequenced by Hedrick et al. [27]. The proposed membrane-spanning segment of approx. 33 residues places a lysine well within the apolar portion of the membrane. An alternative prediction by the method reported here, has the 17 residue segment between the two charged lysines symmetrically placed within, and transverse to, the apolar portion of the membrane. In an alpha helix, this segment, not counting the lysines, would span about 25–26 Å, although the extension might be slightly longer, due to a repulsive interaction between the two positively charged residues which could contribute 13 kJ/mol, assuming an intervening dielectric constant of 4. In any case, the extended segment would likely be a few angstroms

TABLE V

BOUNDARIES b_1 AND b_2 FOR MEMBRANE-SPANNING SEGMENTS PREDICTED BY OUR METHOD COMPARED WITH PREVIOUS ASSIGNMENTS

Number in parentheses are distances from inner and outer boundaries.

	Prediction	Previous assignment	Source (Ref.)
Neuraminidase	(5−1,+2)–29(−6,+6)	7– 35	13
Histocompatibility Ag			
H-2Kb heavy chain	283(−4,+4)–306(−3,+3)	282–308	13
HLA-DR α-chain prec.-	216(−1,+2)–239(−5,+5)	217–239	13
related DC1 α chain	194(−2,+2)–218(−6,+5)	195–217	13
Hepatic lectin	23(−2,+3)– 48(−6,+6)	24– 48	13
Glycoprotein D	346(−4,+4)–362(−4,+3)	340–364	13
Bacteriorhodopsin	13(0, 0)– 29(0, 0)	8– 31	29
	45(0, 0)– 63(−2,+1)	41– 64	
	86(−1,+1)–103(0, 0)	78–101	
	107(0,+1)–125(−, 0)	107–130	
	137(−2,+3)–157(−1,+1)	133–157	
	177(0, 0)–193(0, 0)	167–191	
	205(0,+1)–223(−1,+1)	198–224	
Band 3 anion transport	213(−1,+2)–232(−1, 0)	211–234	13
protein fragment	248(−2,+3)–269(−2,+2)	244–267	
Bovine cytochrome oxidase			
subunit I	20(0,+1)– 37(0, 0)		
	57(−5,+5)– 79(−1,+1)		
	102(0, 0)–118(0, 0)		
	152(0,+1)–169(0, 0)		
	184(−4,+4)–208(−4,+4)		
	242(0,+1)–259(0, 0)		
	270(0,+1)–289(−2,+1)		
	338(−3,+4)–360(−2,+1)		
	373(0, 0)–389(0, 0)		
	454(−1,+2)–475(−3,+3)		
II	26(−1,+2)– 47(−3,+3)		
	63(−3,+4)– 86(−3,+2)		
III	81(0,+1)–100(−2,+2)		
	159(0,+1)–176(0, 0)		
	202((0,+1)–221(−2,+2)		
IV	80(−1,+1)– 98(−1,+1)		
VIIIa	23(0,+1)– 40(0, 0)		

shorter than the commonly used value of 30 Å for the apolar portion of the membrane. This would leave both lysines just inside the apolar portion, where they would, however, be up against negatively charged phosphates. Such a location contributes very favorably to the overall free energy of the system. While it is not our intention here to attempt a calculation of the energetics of polypeptide segments in different environments, we note that with three angstroms between a positive and negative charge (and hence a dielectric constant in the vicinity of 1) each pair can contribute over 418 kJ/mol. This difference is undoubtedly an upper bound since the low local dielectric environment in which the charges interact is bounded by a high dielectric polar environment which can set up an opposing field. Simple models of cavities in dielectrics indicate that such effects might attenuate the actual field by a factor of two or three [28]. Notwithstanding the absence of a rigorous calculation, the estimate indicates the potential importance of electrostatic effects in considering the assignment of a segment. In any case, the difference in predicted locations is striking and provides an unambiguous distinction between two methods of assigning segment locations.

Finally, we have summarized in Table V locations and boundaries of membrane-spanning segments assigned subjectively on the basis of amino acid sequence and other information [13] (no exact boundaries are known to date), and those predicted by our method. Of these, little is known about the disposition of cytochrome oxidase, but a relatively large amount of experimental information is available for bacteriorhodopsin. The former molecule, therefore, provides a prospective, and the latter a retrospective, test of the method. With respect to cytochrome oxidase, 5 out of 7 subunits –I, II, III, IV and VIIIa (also denoted as VII) – are believed to be integral [11]. Which segments are in the membrane, however, is not currently agreed upon; our predictions are listed in Table V.

Electron scattering and electro microscopy indicate that bacteriorhodopsin is composed of seven membrane spanning stalks, each believed to consist of one or more alpha helices, more or less transverse to the plane of the membrane. The resolution of the data is not, however, sufficient to determine which residues are in which stalks, and, consequently, a number of investigators have approached this question by using various types of computational analyses [29,30,23]. Most of these methods start from the experimental observation that seven stalks exist, whereas the metod reported here appears to be the only one which, in addition to assigning residues to each stalk, predicts that seven stalks exist. We also did not use the available pronase digestion data [14] in the predictions, but find that such data is completely in accord with the predicted assignment, which places amino acids 1–4, 65–73, 161–162 and 230–247 outside the apolar portion of the membrane. Finally, the Schiff-base linkage to lysine 216 is predicted to be well within the apolar portion of the membrane, also in agreement with experiment, and strongly suggesting that the retinal ring is centered in the membrane [31].

Acknowledgements

We are deeply grateful to Dr. John Jacquez who has been a continuing source of guidance and advice. We also thank Drs. Jacquez, Robert Blumenthal and Robert Homer Guy for helping us in the collection of integral and peripheral membrane proteins, and Drs. Guy and Jack Dainty for reading and commenting on the manuscript.

References

1 Kaiser, E.T. and Kezdy, F.J. (1984) Science 223, 249–255
2 Kabsch, W. and Sander, C. (1983) FEBS Lett. 155, 179–182
3 DeLisi, C., Klein, P. and Kanehisa, M. (1985) in Molecular Basis of Cancer (Rein, R., ed.), part A, pp. 431–442, Liss, New York
4 Eisenberg, D. (1984) Annu. Rev. Biochem. 53, 595–623
5 Klein, P., Kanehisa, M. and DeLisi, C. (1984) Biochim. Biophys. Acta 787, 221–226
6 Anderson, T.W. (1958) An Introduction to Multivariate Statistical Analysis (Wiley, New York)
7 Kendall, M. and Stuart, A. (1976) The Advanced Theory of Statistics, Vol. 3 (Hafner Press, New York)
8 DePierre, J.W. and Ernster, L. (1977) Annu. Rev. Biochem. 46, 201–262.
9 Klingenberg, M. (1981) Nature 290, 449–454
10 Singer, S.J. (1974) Annu. Rev. Biochem. 43, 805–833
11 Capaldi, R.A. (1982) Trends Biochem. Sci. 7, 292–295
12 Engelman, D.M. and Steitz, T.A. (1981) Cell 23, 411–422
13 Barker, W.C., Hunt, L.T., Orcutt, B.C. et al. (1984) Protein Sequence Database of the Protein Identication Resource,

Version 1, National Biomedical Research Foundation, Washington, D.C.
14 Ovchinnikov, Y.A., Abdulaev, N.G. and Modyanov, N.N. (1982) Annu. Rev. Biophys. Bioeng. 11, 445–463
15 Noda, M., Takahashi, H., Tanabe, T., Toyosato, M., Kikyotani, S., Furutani, Y., Hirose, T., Takashima, H., Inayama, S., Miyata, T. and Numa, S. (1983) Nature 302, 528–532
16 Cramer, W.A., Dankert, J.R. and Uratani, Y. (1983) Biochim. Biophys. Acta 737, 173–193
17 Von Heijne, G. (1981) Eur. J. Biochem. 120, 275–278
18 Geiger, B. (1983) Biochim. Biophys. Acta 737, 305–341
19 Barrantes, F.J. (1975) Biochem. Biophys. Res. Commun. 62, 407–414
20 Edsall, J.T. and McKenzie, H.A. (1983) Adv. Biophys. 16, 53–183
21 Nozaki, Y. and Tanford, C. (1971) J. Biol. Chem. 246, 2211–2217
22 Hopp, T.P. and Woods, K.R. (1981) Proc. Natl. Acad. Sci. USA 78, 3824–3828
23 Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132
24 Eisenberg, D., Weiss, R.M. and Terwilliger, T.C. (1984) Proc. Natl. Acad. Sci. USA 81, 140–144
25 Marx, J.L. (1984) Science 223, 673–676
26 Hunter, T. (1984) Sci. Amer. 251(2), 70–79
27 Hedrick, S.M., Nielsen, E.A., Kavaler, E., Cohen, D.I. and Davis, M.M. (1984) Nature 308, 153–158
28 Jackson, J.D. (1962) Classical Electrodynamics, p. 115, John Wiley and Sons, New York
29 Engelman, D.M., Henderson, R., McLachlan, A.D. and Wallace, B.A. (1980) Proc. Natl. Acad. Sci. USA 77, 2023–2027
30 Argos, P., Rao, J.K.M. and Hargrave, P.A. (1982) Eur. J. Biochem. 128, 565–575
31 King, G.I. Stoeckenius, W. Crespi, H. and Schoenborn, P.P. (1979) J. Mol. Biol. 130, 395–404