Eur. J. Biochem. *190*, 207–219 (1990)

A critical evaluation of the hydropathy profile of membrane proteins

Mauro DEGLI ESPOSTI, Massimo CRIMI and Giovanni VENTUROLI

Department of Biology, University of Bologna, Bologna, Italy

(Received September 18/December 1, 1989) – EJB 89 1132

New membrane-preference scales are introduced for categories of membrane proteins with different functions. A statistical analysis is carried out with several scales to verify the relative accuracy in the prediction of the transmembrane segments of polytopic membrane proteins. The correlation between some of the scales most used and those calculated here provides criteria for selecting the most appropriate methods for a given type of protein. The parameters used in the evaluation of the hydropathy profiles have been carefully ascertained in order to develop a reliable methodology for hydropathy analysis. Finally, an integrated hydropathy analysis using different methods has been applied to several sequences of related proteins.

The above analysis indicates that (a) microsomal cytochrome P_{450} contains only one hydrophobic region at the N-terminus that is consistently predicted to transverse the membrane; (b) only four of the six or seven putative transmembrane helices of cytochrome oxidase subunit III are predicted and correspond to helices I, III, V and VI of the previous nomenclature; (c) the product of the mitochondrial ATPase-6 gene (or the chloroplast ATPase-IV gene) of F_0-F_1-ATPase shows that helix IV is not consistently predicted to traverse the membrane, suggesting a four-helix model for this family of proteins.

The clarification of the molecular structure of membrane (integral) proteins is a problem of great relevance in biological sciences due to the importance of these proteins in so many cellular processes [1]. Hydropathy methods are the most convenient tools for deducing the transmembrane folding of integral proteins when their sequence is known [1–9]. These methods measure the distribution of hydrophobic and hydrophilic regions (i.e the hydropathy [4]) through sequence using a reference scale of hydrophobicity of the amino acids [2–9]. From the hydropathy profile of the polypeptide, segments can be predicted to span the membrane using rules that are specific to each method [1–9]. Generally, it is assumed that these transmembrane segments form α-helical rods as in the known structure of bacteriorhodopsin [3, 7, 8] and in bacterial reaction centers [7, 10–12].

The procedure that is most widely used to predict the folding of newly sequenced proteins is that of Kyte and Doolittle [4]. The hydrophobicity scale of Kyte and Doolittle has also been used in subsequent works with the aim of improving the accuracy of the folding predictions [13–15]. However, other scales may provide a better estimation of the hydrophobic/hydrophilic balance of the individual amino acids than that of Kyte and Doolittle [6–9].

Recently, hydropathy methods have been derived from the statistical preference of the residues to form the integral regions in several membrane proteins [16, 17]. This statistical/empirical approach is similar to that previously employed for deducing the conformational preference [18] and the hydrophobicity [19] of the residues from known protein structures. Some statistical scales of hydrophobicity correlate much better than the physicochemical scales with the highly resolved crystal structures of globular proteins [20, 21]. No comparative and comprehensive study is available for evaluating the correlation of all the published hydrophobicity, hydropathy and statistical scales with the structures of membrane proteins.

This work represents a step towards a systematic evaluation of hydropathy procedures and, in particular, provides criteria for selecting the most appropriate scales for a given type of membrane protein. Three membrane-preference scales show the best correlation with the known transmembrane distribution of the residues in bacterial reaction centers, and have consequently been used to develop a carefully tested hydropathy scheme. Different predictive rules have been optimized for each scale after examining a variety of proteins with known membrane topology, and their results are integrated in order to reach a consensus picture of the likely number of transmembrane helices in polytopic proteins.

EXPERIMENTAL PROCEDURES

Elaboration of statistical scales of membrane preference

Many integral proteins possessing a well-characterized membrane topology have been used to determine a set of membrane-preference scales (MP) for the amino acid residues, according to the formula [16–18]:

$$MP_j = \frac{f_{j,MEM}}{f_{j,TOT}} \quad (1)$$

where MP_j is the membrane-preference value for residue j, $f_{j,MEM}$ is the frequency of occurrence of residue j in the transmembrane regions of the proteins and $f_{j,TOT}$ is the frequency of occurrence of residue j along the entire sequence of the protein, i.e. the molar fraction of residue j [16, 18].

Correspondence to Dr. M. Degli Esposti, Institute of Botany, Department of Biology, University of Bologna, Via Irnerio 42, I-40126 Bologna, Italy.

Abbreviations. AMP07, average membrane preference of seven scales; MP, membrane preference; MPH, membrane propensity for haemoproteins; NKD, Kyte-Doolittle scale which is normalized to the MPH scale.

Enzymes. Ubiquinol:cytochrome-*c* reductase (EC 1.10.2.2); cytochrome *c* oxidase (EC 1.9.3.1); 3-hydroxy-3-methylglutaryl-CoA reductase (EC 1.1.1.32).

Table 1. *Statistical scales of membrane preference that are used in this work*
The sequences have been taken from [12, 16, 17, 23–31, 54, 58, 63] and references therein. COI, COII, COIII, COIV, COX and COXI are all subunits of cytochrome oxidase

Definition of the scale	Acronym	Proteins	Membrane residues considered	Reference
Membrane-buried-helix parameter	RAOAR	7 acetylcholine receptors subunits; 4 ATPase 6; 7 cytochrome *b*; 5 COI; 6 COII; 3 COIII; 7 proteolipid ATP synthase; 8 cytochrome P_{450}	5632	16
Membrane propensity for haemoproteins	MPH	15 cytochrome *b* (haem-binding domain); 3 cytochrome b_6 subunit; 6 subunits of bacterial reaction centers	2885	17
	MPH89[a]	several cytochrome *b*	>4000	this work
Predicted membrane composition of redox proteins	PMCRE	cytochrome *f* spinach, tobacco; cytochrome b_{559} A, B spinach; *SDHC, SDHD, E. coli; SDHC B. subtilis; FMRC, FMRD E. coli; cyoB E. coli; NARI E. coli;* cytochrome b_5 calf; COI *P. denitrificans*; COI sea urchin; COIII *P. denitrificans*, sea urchin, bovine; COII human, sea urchin, maize; cytochrome *b D. yakuba*; subunit IV b_6f spinach, tobacco, *M. polymorpha*; cytochrome c_1 beef, yeast; subunit V, VI, VII b_1 beef, yeast; COIV, COX, COXI bovine, yeast	2810	this work
Membrane preference of D_1, D_2 thylakoid proteins	D12MP	D_1 spinach, *Synechococcus*, tobacco, *E. gracilis*; D_2 spinach, *C. reinhardtii*	683	this work
Membrane preference of opsins and related proteins	OPSMP	rhodopsin beef, human, *Drosophila*; β-adrenergic receptor human, hamster, pig	918	this work
Membrane composition of halorhodopsin and bacteriorhodopsin	RHODO	bacteriorhodopsin and halorhodopsin, *H. halobium*; octopus opsin	467	this work
Membrane composition of the L and M subnunits of bacterial reaction centers	MICHE	L and M subunits of reaction center, *R. monas viridis, R. capsulatus, R. sphaeroides*	825	12 and this work[b]
Average membrane preference	AMP07	arithmetical mean of the seven scales above	>14000	this work

[a] The membrane preference of histidine is 1.0 and this is the only variation from the original scale [17].
[b] Calculated from the frequencies reported by Michel et al. [12] by using Eqn (1); therefore, these values are different from the distribution factors measured in [12], with an overall correlation of 95% if cysteine is excluded.

New membrane-preference scales have been derived from groups of proteins which belong to two main categories: membrane proteins with redox function (category A); membrane proteins with transport or non-redox function (category B). A 'standard of truth' scale has been calculated for category A using the known structure of bacterial reaction centers [10–12], and for category B using the likely membrane composition of the rhodopsins in *Halobacterium halobium* [22–24]. A statistical scale has been elaborated from the membrane composition predicted by the consensus of the methods in [4, 13, 16, 17] for proteins which are structurally homologous to either the subunits of the reaction centers, i.e. the D_1 and D_2 thylakoid proteins [12, 25–27], or to the bacterial rhodopsins, i.e. the retinal opsins and the β-adrenergic receptors [28–31].

Another membrane-preference scale has been introduced for evaluating the influence of sequence similarity in the proteins of the data base. A strong sequence similarity is present in the proteins utilized for calculating the statistical scales above, and in [16, 17].

Theoretically, this may lead to limited accounts of the distribution of the amino acid residues in non-homologous proteins. Several membrane proteins with redox function that exhibit widely different topologies and little overall homology have been evaluated using a consensus of the methods in [13, 16, 17] yielding the statistical scale of predicted membrane composition of non-homologous redox proteins (PMCRE, Table 1).

Finally, an average scale of membrane preference (AMP07) has been computed as the arithmetical mean of the five scales above, and of the two statistical scales published previously [16, 17]. Table 1 lists all these scales, their acronyms and the proteins from which they are calculated. The acronyms of other scales have been taken from [20] or made up from names of the authors of the pertinent work (see Table 2).

Correlation and normalization of the various scales

The correlation (r) among all the scales listed in Table 1, the hydrophobicity scales reported in [20] and the hydropathy scales in [2–9] has been computed according to the following equation [20]:

$$r = \frac{\sum_{i=1}^{20} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\left[\sum_{i=1}^{20} (x_i - \bar{x})^2 \cdot \sum_{i=1}^{20} (y_i - \bar{y})^2 \right]^{1/2}}, \quad (2)$$

Table 2. *Correlation (r) among several scales of hydrophobicity*

A selection of scales that correlate best with the average AMP07. The acronyms used are the following (see also [20]): LEVIT, physicochemical scale of Levitt [33]; WOLFE, experimental scale of partition of Wolfenden et al. [34]; FAUPL, experimental scale of partition into octanol measured by Fauchère and Pliška [35]. The solvation energy calculations of Eisenberg and McLachlan [67] and the physicochemical scale introduced by Roseman [37] are highly correlated with this experimental scale (results not shown); GUYME, mean of several statistical scales derived from the known structure of globular proteins [38]; ROSEF, statistical scale of Rose et al. [39]; PRIFT, statistical scale which optimizes the amphipathic character of α helices in soluble proteins [20]; VHEJB, the first hydropathy scale for membrane proteins introduced by Von Heijne and Blomberg [2]; ENGST, hydropathy scale based on physicochemical considerations introduced by Engelman [3, 7]. The scale IF05 recently reported in [68] is highly correlated with this scale; KYTDO, the hydropathy scale of Kyte and Doolittle [4]; EISEN, hydropathy scale proposed by Eisenberg et al. [5, 6]; LUNDE, modification of the VHEJB scale proposed by Lundeen et al. [9]; RAOAR, statistical scale of membrane-buried helix parameters [16]; MPH89, statistical scale of membrane propensity for haemoproteins [17] modified here (Tables 1 and 4). The scale derived from the predicted transmembrane structure of mitochondrial cytochrome *b* and incorporated in the MPH scale [17] shows a r of 78% with the MICHE scale; OPSMP, PMCRE, RHODO, MICHE and AMP07, membrane preference scales calculated in this work (Table 1). The MICHE values are nearly equivalent to the distribution factors calculated by Michel et al. [12]

Acronym	correlation (r)																	
	%																	
LEVIT	100																	
WOLFE	66	100																
FAUPL	92	69	100															
GUYME	78	65	90	100														
ROSEF	79	67	90	98	100													
PRIFT	63	57	76	90	88	100												
VHEJB	83	88	79	70	69	55	100											
ENGST	88	89	85	76	79	62	91	100										
KYTDO	68	89	81	83	84	81	80	85	100									
EISEN	86	91	87	79	81	64	92	94	88	100								
LUNDE	96	78	90	79	81	67	89	96	79	89	100							
RAOAR	77	88	83	84	87	77	85	93	94	90	88	100						
MPH89	64	74	70	84	87	81	65	79	84	71	77	91	100					
PMCRE	70	75	76	83	86	76	67	84	80	75	81	90	92	100				
OPSMP	77	80	83	83	84	79	82	84	89	86	83	90	81	78	100			
RHODO	70	58	70	72	65	59	75	68	64	68	71	65	51	61	74	100		
MICHE	44	59	57	70	75	67	48	66	71	51	61	78	87	85	60	42	100	
AMP07	75	82	82	89	91	82	78	89	90	83	86	96	96	95	88	69	87	100
	LEV	WOL	FAU	GUY	ROS	PRI	VHE	ENG	KYT	EIS	LUN	RAO	MPH	PMC	OPS	RHO	MIC	AMP

where $\bar{x}$ and $\bar{y}$ are the averages of the values of the residues in the scales $\{x_i\}_{i=1}^{20}$ and $\{y_i\}_{i=1}^{20}$ that are compared. The correlation coefficient is expressed as a percentage and is independent of the absolute values of hydrophobicity or membrane preference of the residues when linear images of the scales are compared [20].

It appears that the scale of Kyte and Doolittle [4] and that of membrane propensity for haemoproteins (MPH) [17] have a sufficiently similar symmetry in the distribution of the values of the residues to allow a meaningful linear normalization to each other. The following equation has been used for such a normalization:

$$x_{j,\mathrm{NKD}} = x_{j,\mathrm{KYTDO}} \cdot 0.198 + 1.1, \qquad (3)$$

where the factor 0.198 derives from the ratio of the maximal ranges spanned by the MPH [17] and the original Kyte and Doolittle scale (identified by the acronym KYTDO) [20]. In this way, the original Kyte and Doolittle scale is translated into a membrane-preference scale centered at a midpoint value of 1.1, differing from previous normalizations [6, 17].

Hydropathy algorithms

A computer program has been written in BASIC providing a large variety of options in the elaboration of the hydropathy plots. These options include the following: (a) choice of the scale of membrane preference or of the normalized Kyte and Doolittle scale (Table 4); (b) amplitude of the moving segment along the sequence, usually called scanning window [5–8], that averages the hydrophobicity of the residues; (c) value of the baseline (or cut-off) [1] which is selected for discriminating between hydrophilic and hydrophobic regions in the proteins [4, 16, 17]; (d) selection of the positive peaks above the chosen baseline that can be predicted to be transmembrane segments according to the rules specific to each scale.

Positive peaks in the hydropathy profiles are colored in black when they satisfy the rules for being predicted to transverse the membrane [17]. The plots are presented without smoothing of the data and using an arbitrary value of 0.2 for both the N-terminal and the C-terminal residue [17].

The computer program (MAGINT) containing the four methods of hydropathy in Table 3 is available from the authors. A large data base of membrane proteins which is particularly rich in polytopic proteins with redox function is also available from the authors.

RESULTS AND DISCUSSION

The rationale for evaluating the methodological aspects of a hydropathy analysis

Three major sources of uncertainty arise in the evaluation of the hydropathy profile of membrane proteins: the choice

Table 3. *Membrane-preference scale predictive rules*
The table shows the simplest formulation of the empirical rules employed here fore predicting transmembrane segments from positive peaks above the baseline in the hydropathy plots. The length refers to continuous segments having the mean value ≥ baseline. Except for NKD, whose plots are averaged by a window of 19 residues [4, 36], the rules are valid for a scanning window of seven residues and are simplified with respect to those originally proposed [16, 17]. AMP07 (A): rules are optimised for membrane proteins with redox function calibrated with the set of reference sequences quoted in the text and with the known structures of two bacterial reaction centers [11, 12]. AMP07 (B): rule for non-redox integral proteins calibrated with the set of reference sequences quoted in the text. For definition of the acronyms, see Table 1. NKD, normalized Kyte and Doolittle scale

Membrane-preference scale	Baseline value	Length	Heigth	Area
NKD	1.1	≥ 19		
RAOAR	1.05	≥ 12	≥ 1.13	
MPH89	1.1	≥ 12	≥ 1.3	≥ 2
AMP07 (A)	1.1	≥ 12		≥ 1.5
AMP07 (B)	1.0	≥ 13		

Table 4. *Membrane preference scales for the integrated hydropathy analysis of membrane proteins*
The original scale of Kyte and Doolittle [4] has been normalized (NKD) to the MPH scale as described in Experimental Procedures. In such a normalization, the baseline (midpoint) of 1.1 corresponds to 0.0 kJ/mol in the original dimensions, which is slightly higher than the original value suggested in [4]

Amino acids	Membrane-preference scales			
	NKD	RAOAR	MPH89	AMP07
Ala	1.46	1.36	1.56	1.26
Cys	1.60	1.27	1.80	1.60
Asp	0.41	0.11	0.23	0.27
Glu	0.41	0.25	0.19	0.23
Phe	1.65	1.57	1.42	1.46
Gly	1.02	1.09	1.03	1.08
His	0.47	0.68	1.00[a]	1.00
Ile	1.99	1.44	1.27	1.44
Lys	0.33	0.09	0.15	0.33
Leu	1.85	1.47	1.38	1.36
Met	1.48	1.42	1.93	1.52
Asn	0.41	0.33	0.51	0.59
Pro	0.78	0.54	0.27	0.54
Gln	0.41	0.33	0.39	0.39
Arg	0.21	0.15	0.59	0.38
Ser	0.94	0.97	0.96	0.98
Thr	0.96	1.08	1.11	1.01
Val	1.93	1.37	1.58	1.33
Trp	0.92	1.00	0.91	1.06
Tyr	0.84	0.83	1.10	0.89

[a] Only difference with the original MPH scale [17].

of the scale, the setting of the parameters for computing the profile and the use of rules for predicting the transbilayer segments [7, 13, 16]. The published procedures have been generally developed for optimizing the resolution of the helices and the precision in the identification of their ends, using bacteriorhodopsin and several monotopic proteins as the reference systems [3, 4, 6, 7, 12, 13, 16]. However, no thorough analysis has been dedicated to the evaluation of the influence of the selected scale on the accuracy of the deductions, perhaps because it is commonly considered that many scales have a similar effectiveness for membrane polypeptides [1, 3, 4, 12, 13]. In this study, the above uncertainties have been extensively assessed in order to elaborate a critical methodology of hydropathy analysis.

In particular, a detailed comparison of the scales has been undertaken by following the principles previously applied to globular proteins [20,21]. The linear correlation of the various scales with standards of known structure provides an useful indication of their validity [20]. A 'standard of truth' scale of membrane preference has been calculated by applying Eqn (1) to the distribution of the residues derived from the known atomic structures of the bacterial reaction centers [10–12] (MICHE scale, Table 1). Hence, the correlation with such a standard can measure the effectiveness of the various scales in describing the transmembrane composition of integral proteins, especially of those with redox function like the reaction centers.

Membrane proteins that do not have a redox function, however, may have a different distribution of the residues [12, 17]. Thus, the scale of membrane preference that is deduced as above from the current models of the rhodopsins [23, 24], i.e. proteins that transport ions across the membrane [23], can be used as an additional 'standard of truth' in the comparative correlation of the existing scales.

Owing to the limited number of sequences available for computing these standards, it is important to compare their parameters with those deduced from other proteins. This analysis is undertaken with the sequences listed in Table 1 and enables an evaluation of the statistical fluctuations of the preference parameters with respect to the type and number of the polypeptides examined. Since a statistical stabilization of the preference values is likely to occur by increasing the number of the sequences analyzed [12, 16, 20], a weighted scale, called AMP07, has been computed by averaging the data obtained previously [16, 17] and in this study (Table 1). The parameters of the AMP07 scale are indeed statistically stabilized since they do not significantly change after extending the analysis to over 25000 amino acid residues (results not shown).

Comparison of different statistical scales of membrane preference

The values of the various membrane-preference scales in Table 1 are represented in a graphical form in Fig. 1 for two reasons. Firstly, it is clear from this representation that the amino acids tend to segregate into three separate groups. Secondly, the correpondence of the membrane preference of the residues in each scale with the average value AMP07 can be visualized directly.

The group of amino acids that is most clearly defined includes the hydrophilic Asp, Glu, Lys, Asn, Gln, Pro and Arg (Fig. 1). It is interesting to note that proline is included among these hydrophilic residues (Fig. 1), contrary to common belief that considers such a residue to be hydrophobic [3, 6, 9, 32–38]. Indeed, proline is excluded from buried α helices of soluble proteins with known structure because of its helix-breaking character [18, 20, 39], and this may as well apply to transmembrane helices.

Serine and threonine, which are often believed to be hydrophilic [2, 5, 32–38], segregate with a group of amino acids which can be defined as 'neutral' since their average

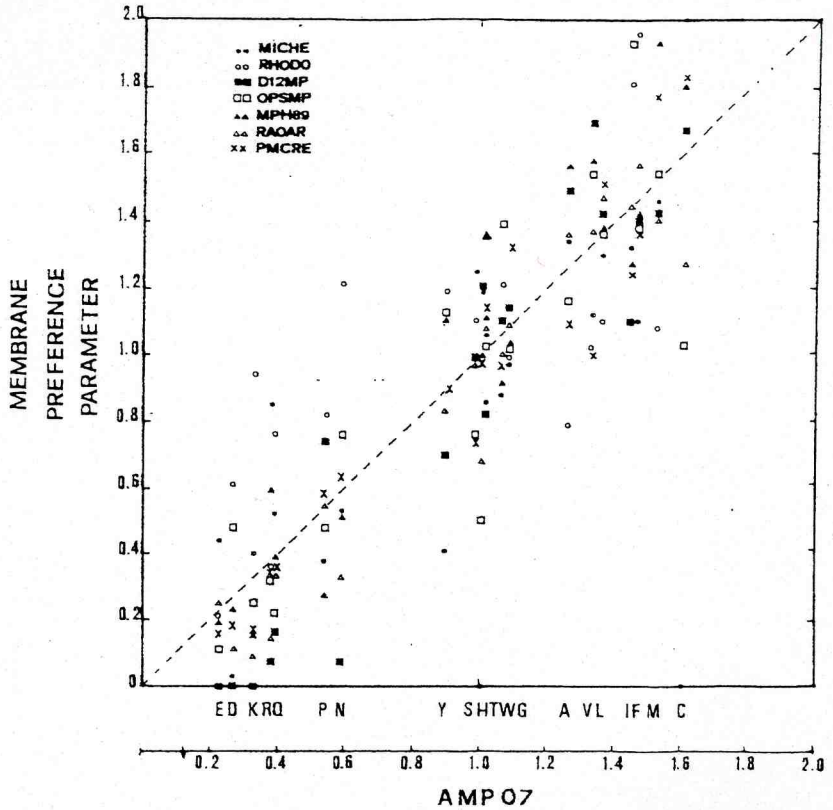

Fig. 1. *Comparison of the membrane-preference parameters of the statistical scales in Table 1.* The average membrane-preference (AMP07) scale is used as the reference in computing the graph and the dashed diagonal represents the graphical correlation with the parameters of such a scale. The value of 1.44 for histidine has been used for the MPH scale [17]. The single letter code identifies the amino acids

membrane propensity is ≈ 1 (Fig. 1). This group also includes tyrosine, histidine (which is commonly believed to be hydrophilic) [2–9, 32–38], tryptophan (which is generally considered to be rather hydrophobic [2, 3, 5, 7, 9, 19, 20, 32–39], and glycine. The neutrality of this latter amino acid is also evident in soluble proteins [20, 39].

The amino acids in the third group are hydrophobic since they have a MP >1 (Fig. 1), and include Ala, Cys, Phe, Ile, Leu, Met and Val. Cysteine usually has the largest hydrophobicity value, as noted previously [12, 17, 39].

Interestingly, the above segregation of the residues, except for the allocation of tryptophan, corrresponds to that obtained from the analysis of the fractional loss of surface area in globular proteins with known structure [39]. Indeed, a large number of the residues in membrane proteins having several transmembrane helices face other protein regions and not the lipids [11] and this renders the interior of proteins such as the photosynthetic reaction centers rather similar to the hydrophobic core of globular proteins [11, 19, 39].

It is worth noting that the common methods which are used for determining the polarity of proteins [4, 13, 32] consider an allocation of the hydrophilic residues that is significantly different from that in Fig. 1 or [39]. For instance, the average polarity of the integral proteins in Table 1 is 33.5% using the method of Capaldi and Varderkoii [32], and 23.9% by considering the hydrophilic residues in Fig. 1.

Correlation between the scales of hydrophobicity

Table 2 summarizes the correlation between some of the most used scales of hydrophobicity (both physicochemical and statistical) [20], of hydropathy [2–9] and of membrane preference (Table 1) [16,17]. In agreement with previous findings in globular proteins [20, 21], statistical scales generally correlate better than physicochemical scales with the 'standard of truth' deduced from the known structure of the bacterial reaction centers (MICHE scale, Table 2). The fact that the AMP07 and the MPH scale show the highest correlation with this standard (Table 2) is somewhat expected, since these scales incorporate the frequencies of amino acid residue distribution in the bacterial reaction centers (Table 1). However, the PMCRE scale (Table 2), which is derived from the analysis of proteins that are not homologous to the reaction centers (Table 1), also shows a similar high correlation with the MICHE scale (Table 2). This suggests that homology of the proteins in the data base cannot completely account for the high correlation with the 'standard of truth' of the bacterial reaction centers.

A completely different ranking is seen in the correlation with the 'standard of truth' of the rhodopsins (RHODO scale, Table 2). Very few scales, e.g. VHEJB (Table 2) [2], show correlations higher than 70% with the RHODO scale (Table 2), which is very poorly correlated with the MICHE scale and only modestly correlated with both the scale derived from the opsins (OPSMP, Table 2) and the average AMP07 scale (Table 2). This confirms the above hypothesis of the effect of sequence homology and also suggests that the rhodopsins of halobacteria may have a peculiar transmembrane composition. The latter suggestion conforms to the observation that, contrary to the situation of the redox proteins, some solubility scales correlate with the rhodopsin standard (Table 2) at the same level as many statistical scales.

From the overall pattern of correlation between scales (Table 2 and Appendix), it is inferred that the relative hydro-

phobicity of the residues does not vary much in soluble and membrane proteins and, within integral proteins, depends strongly on their function. This conclusion is supported by the different transmembrane frequencies of residues such as Cys, Asp, Lys, Asn, Pro, Arg, Ser and Tyr in the two categories of membrane proteins examined in this work, even if the overall content of hydrophobic residues is, on the average, very similar (see. Fig. 1 and [12]). For instance, Asp and Lys are essential for proton pumping and retinal binding in the rhodopsins [23], whereas Cys and His are required for metal binding in membrane proteins with redox function [12, 17] (see Michel et al. [12] for a detailed discussion of the structural differences between bacteriorhodopsin and the reaction center).

The correlations in Table 2 offer useful criteria for selecting the most appropriate scale(s) for a given protein. For instance, the AMP07 and the MPH scales are probably the best for describing the profile of integral proteins with redox function, whereas the scale of Von Heijne and Blomberg (VHEJB, Table 2), that of Rao and Argos (RAOAR, Table 2 [16]) and also the AMP07 scale are most suitable for describing the profile of membrane proteins with transport function. The AMP07 and the Rao and Argos scales [16], moreover, show the highest cumulative correlation with all the other scales ($\approx 83\%$ Table 2 and results not shown).

Optimization of the plotting parameters of the hydropathy profiles

The hydropathy profile of a sequence depends not only on the choice of the scale, but also on the averaging procedure(s) for the data presentation. The plotting of the profile is basically determined by a subjective choice of the amplitude of the scanning window. It is usually believed that an optimal scanning window should be as long as the minimal length of a membrane-spanning α helix, i.e. ≥ 19 residues [4, 6, 7, 12, 13, 15, 36]. Windows of length greater than 10 residues, however, produce a dramatic loss of local information, since they often lead to an apparent fusion of transmembrane segments that are closely spaced, such as some of the helices in bacteriorhodopsin [4, 7, 12, 16]. Moreover, large windows increase the errors of predicting hydrophobic regions of globular proteins as potential transmembrane helices [4, 16].

Clearly, it is preferable to have a relatively 'noisy', but well-resolved plot, than a nicely smoothed, but misleading one. Indeed, the best correspondence of the hydropathy profiles with the known structures of either globular or membrane proteins is seen when windows of 5–9 residues are used [4, 16, 17, 20]. This is also true for the AMP07 scale introduced here, which shows its best resolution of the known transmembrane helices of the proteins listed in Table 1 with a window of seven residues (results not shown). Therefore, a fixed scanning window of seven residues has been routinely used in this study without the further smoothing of the data that is suggested in [8, 16], since this smoothing also causes loss of information and unfavorably affects the reliability of the predictions (see below).

Calibration of rules for predicting transmembrane helices

The use of consistent rules for predicting the transbilayer nature of the positive peaks in the hydropathy profiles is a problem of crucial importance. First of all, these rules must be calibrated at a fixed value of the scanning window, because they are critically dependent on the shape of the profile. This circumstance is often dismissed in the literature. For instance, scanning windows of 11 residues have been used with the Kyte and Doolittle algorithm [14, 29], even if the original rules were calibrated for a window of 19 residues [4].

Secondly, the prediction rules must be optimized with a large number of either membrane proteins with known topology (i.e. polypeptides that have an established number of transmembrane segments) or globular proteins of known structure [16]. Unfortunately, there are severe limits in such a verification, since the only membrane proteins with resolved crystal structures are the bacterial reaction centers [10–12], and most hydropathy methods clearly identify their transmembrane helices [7, 9, 12, 16, 17, 36, 37]. Furthermore, the majority of the other integral proteins which have an established topology possess one or two transbilayer segments, which are so hydrophobic that they are equally recognized by several hydropathy schemes [1, 4, 8, 13, 17, 36, 37]. Consequently, all these potential reference systems have a limited value for comparing the accuracy of the various methods and defining stringent prediction rules.

In this work, the profiles of Rao and Argos [16], the MPH [17] and the AMP07 scales have been examined by applying them to several proteins with well-characterized membrane topology for which, in contrast with the above systems, erroneous predictions are often obtained with the available hydropathy methods. Cytochrome P_{450} cam [40] and cytochrome-*c* peroxidase [41] are globular proteins with known structures that serve as valuable negative controls since they contain a large number of hydrophobic helices. The various mutants of the viral G proteins retain the ability to transverse the membrane until 12 uncharged residues are present in their membrane-anchoring domain [42], and are therefore ideal references for fine-tuning the rules.

Predictive rules have been calibrated in the above systems and in the following polytopic proteins: the photosynthetic D_1 and D_2 subunits from 12 species (listed in [25, 43]), the topology of which is inferred from the functional similarity and structural homology with the L and M subunits of the bacterial reaction centers [12, 25]; all the rhodopsins quoted in Table 1, which are considered to possess the same topology of bacteriorhodopsin [23]; 30 cytochromes *b* from mitochondria, chloroplasts and bacteria, whose transmembrane folding has recently been assessed by a combination of genetic and functional studies (see [17] and references therein); the *malF* permease [44] and the *secY* ribosomal protein from *Escherichia coli* [45], the topology of which has been elucidated by gene-fusion techniques [44].

The major criterion that has been followed in the empirical calibration of the rules is the definition of a minimal length of a hydrophobic stretch that enables its identification as a transmembrane helix. The baseline values of the profiles have been varied systematically between 1.0 and 1.2, and the length of the peaks above such baselines evaluated in view of the known topology of the reference proteins. The results of this wide screening indicate that a minimal length of 12 residues above a baseline at 1.05 for the non-smoothed Rao and Argos profile and at 1.1. for both the MPH and the AMP07 profile is the simplest and most powerful rule for minimizing the errors in predicting the transmembrane helices, as verified also for the reaction-center subunits (results not shown, see also [17]). In fact, the mutants of the viral G protein that possess a hydrophobic segment shorter than 12 residues [42] are correctly excluded from being transmembrane by such a rule.

Only a few additional rules have been devised for reducing the most typical errors that are obtained for each method with only the criterion of minimal length. These rules are condensed

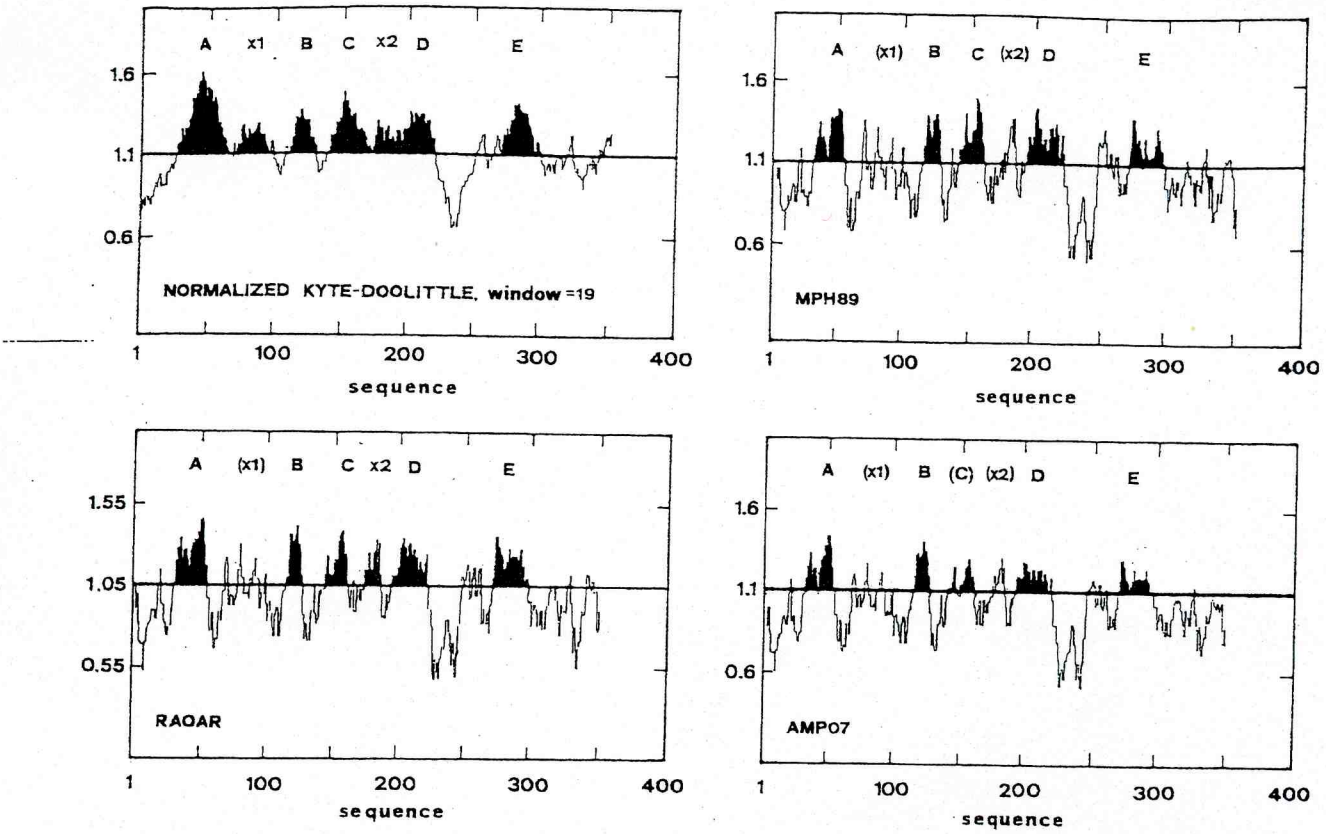

Fig. 2. *Hydropathy plots of the D_1 protein of* Prochloron *by the four methods used in this study.* The sequence is highly similar to the herbicide-binding protein of higher plants [25, 43]. Six or seven transmembrane helices are usually predicted in these proteins by the methods in [4, 8, 16] and only five by the MPH method (results not shown). The nomenclature of the transmembrane helices and of the false helices (X) is taken from [16]. Helix C is not predicted if the strict rule of minimal length is applied in the AMP07 plot. However, the positive peak of this protein region is flanked by clear minima and contains a single gap of two residues that have a mean just below the baseline. The scanning window is of seven residues except in the normalized Kyte and Doolittle plot

from those proposed previously [16, 17], and have been inserted, in their simplest formulation described in Table 3, in the computer program to obtain a consistent and objective selection of potential transmembrane segments. In this way, it is possible to prevent that a versatile evaluation of the results can lead to predictions influenced by subjective biases.

Although the rules in Table 3 minimize the false-positive and false-negative predictions of transmembrane helices, their stringent application is still prone to errors. Statistically, the errors are of a different nature in each method and for the two different categories of proteins examined. For instance, the AMP07 profile tends to underestimate the transmembrane segments in proteins with redox function, whereas the non-smoothed Rao and Argos [16] profile tends to overestimate the same helices. This tendency is most clearly illustrated by the photosynthetic D_1 subunits (Fig. 2), but is also seen in the *b* cytochromes (results not shown). In contrast, the non-smoothed Rao and Argos [16] profile and the AMP07 profile underestimate to comparable extents the known transmembrane segments of rhodopsins and other non-redox proteins by using the same baselines which minimize the false positive predictions in redox proteins (results not shown).

The above considerations indicate that it is impossible to achieve the maximal accuracy for both categories of proteins if a fixed cut-off value is maintained. In this study, the strategy for reducing the opposite type of errors consists of employing both the non-smoothed Rao and Argos [16] profile with its original baseline at 1.05 and the AMP07 profile with a different baseline for redox and non-redox proteins. The AMP07 profile resolves and accurately predicts all but one of the known transmembrane segments of the reference proteins with non-redox function by using a baseline at 1.0 and the simple criterion of a minimal length of 13 residues (results not shown). Therefore, such a variation of the predictive scheme is routinely employed when proteins without redox function are analyzed (Table 3).

It is consistently observed that with the present methods, an improved accuracy in identifying the ends of the transmembrane helices can be achieved only at the expense of an increase of the errors in predicting the correct number of the helices in polytopic proteins. Therefore, complementary procedures should be combined with hydropathy plots to avoid this problem. On the other hand, the rules described here do not recognize hydrophobic regions in extended conformation of either globular or integral proteins (e.g. bacterial porins) as possible transmembrane segments.

An integrated hydropathy analysis for deducing the membrane topology of proteins

Given the difficulty in obtaining crystal structures in membrane proteins, the evaluation of their possible molecular architectures relies on hydropathy deductions [1,16]. These deductions are also indispensable in the design of experimental approaches for probing the membrane topology of integral proteins [45].

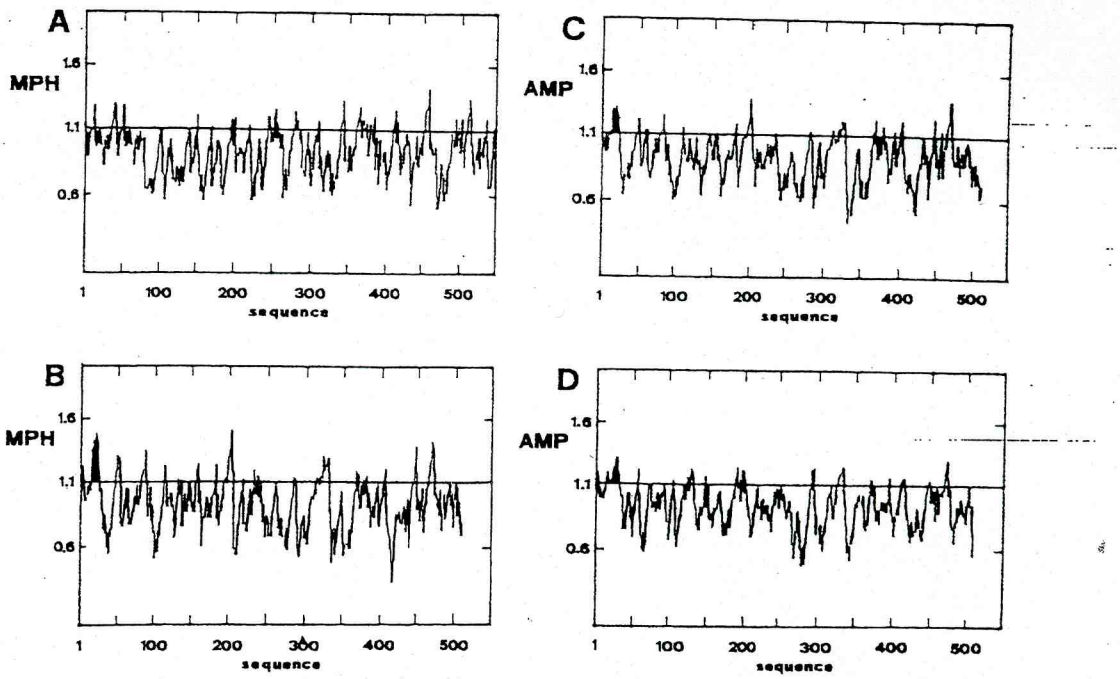

Fig. 3. *Hydropathy plots of four different cytochrome* P_{450}. All the plots are obtained with a window of seven residues. The sequences are taken from the list of Nelson and Strobel [47] and belong to different families of cytochrome P_{450} that show little sequence similarity [46, 47]. (A) Side-chain-cleavage cytochrome P_{450} from adrenal mitochondria, that lacks the hydrophobic region at the N-terminus which is present in all the microsomal forms [47, 51, 52]. (B) microsomal cytochrome P_{450} d from rat liver. (C) microsomal cytochrome P_{450} form 4 from rabbit liver (this protein has been topologically characterized in [52]); (D) cytochrome P_{450} lanosterol ω-hydroxylase from rat

The integration of the methods in Tables 3 and 4 is likely to implement the accuracy of the hydropathy predictions, as illustrated in Fig. 2 for the photosynthetic protein D_1 of *Prochloron* [43]. This protein is an example of how easy it is to overestimate the number of transmembrane segments in proteins with redox function, particularly with the Kyte and Doolittle scale (see also [13, 16, 17]). Due to its wide use in the literature, the Kyte and Doolittle method has been utilized here as a reference with the window of 19 residues [4, 13, 36], but with a slightly higher baseline for reducing its common error of overestimating the number of transmembrane segments [17] (Table 3, see [4]).

The consensus profile of related proteins is less sensitive to local sequence variations that may critically affect the hydropathy deduction in a single polypeptide [16, 17, 21]. Thus, predictive errors are expected to be reduced by comparing the hydropathy plots of homologous proteins, as in the case of mitochondrial cytochrome *b* and chloroplast cytochrome b_6 [16, 17]. Accordingly, the present analysis has been mainly applied to proteins where many homologous sequences are known. Different indications from those obtained previously with limited hydropathy evaluations are obtained for the proteins summarized in Table 5, as detailed below.

The membrane topology of cytochrome P_{450} *is very simple*

The superfamily of cytochrome P_{450} includes soluble (in bacteria) [40] and membrane-bound (in eukaryotes) [46, 47] haemoproteins that are involved in a variety of degradative and metabolic processes (for a review see [46]). Despite the very large number of sequences available, the membrane topology of eukaryotic cytochrome P_{450} has not yet been assessed [46–52]. The present hydropathy analysis indicates that the form of cytochrome P_{450} which is present in adrenal mitochondria [46] is unlikely to be an integral protein since no transmembrane segment is predicted (Fig. 3A). This finding is in agreement with recent views [47] and is apparently supported by the successful crystallization of this protein in a solution without lipid-like reagents [48].

On the other hand, the strong hydrophobicity displayed by all types of microsomal cytochrome P_{450} leads to large overestimations of the number of transmembrane segments with all the methods employed so far [4, 7, 16, 18, 33, 47]. The minimal number of predicted transmembrane segments is generally five [16, 47], but models with eight [49] or even ten [50] transmembrane segments have been proposed (Table 5). In contrast, recent topological studies demonstrate that most of cytochrome P_{450} is exposed at the cytoplasmic side of the membrane, and probably a single start stop [1] transmembrane segment is present at the N-terminus [51–53]. The integrated hydropathy analysis performed in microsomal cytochromes P_{450} belonging to several families shows that only one hydrophobic region at the N-terminus is consistently predicted to transverse the membrane (Fig. 3). Such a region is not present in the bacterial [40] and in the mitochondrial form [47, 48]. Hence, the most likely membrane topology of microsomal cytochrome P_{450} consists of one transmembrane helix which anchors the protein to the bilayer and of a large catalytic domain that protrudes into the cytoplasm [51–53]. This topology is similar to that of both cytochrome P_{450} reductase and cytochrome b_5 [47].

Surprisingly, the Rao and Argos method [16] as used here (Table 3) generally confirms the predictions obtained with the MPH or AMP07 method, whereas in the original work five transmembrane helices were predicted in microsomal cytochrome P_{450} [16]. The discrepancy arises solely from the use of an iterative smoothing procedure [8, 16] in calculating the hydropathy plots. This smoothing artificially enlarges positive

Table 5. *Comparison of the predicted transmembrane segments in several integral proteins by different methods*
Except 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) reductase, at least six sequences of each family have been analyzed. See legends of Tables 1 and 2 and also [20] for the meaning of the acronyms. CHOFA identifies the Chou and Fasman method [18]. The segments predicted to transverse the membrane either in α or β conformation are according to the references cited, whereas the likely number of transmembrane helices is derived from clear topological data. The question mark signifies that such topological data are not available

Protein function	Protein (super)-family	Hydropathy method	Predicted trans-membrane segments			Likely number of helices	
			α	β	reference	number	references
Redox	Cytochrome P_{450} microsomal	ENGST	10		50	1	51–53
		CHOFA	8		49		
		RAOAR	5		16		
		KYTDO/ENGST	≥ 4		47		
		MPH/AMP07[a]	1		this work		
	CO III mitochondria	KYTDO	7		54	?	
		RAOAR	6		16		
		RAOAR/MPH/AMP07[a]	4		this work		
	HMG-CoA reductase	KYTDO/CHOFA	7	2	55	?	
		RAOAR/MPH/AMP07[a]	6		this work		
Non-redox	mitochondrial carriers	KYTDO	6	≥ 1	14	?	
		RAOAR/AMP07[b]	4		this work		
	F_0 F_1-ATPase subunit 6 or IV or subunit a (*E. coli*)	VHJEB	6–7		60	?	
		EISEN	8		69		
		KYTDO	5		61		
		RAOAR	4		16		
		RAOAR/AMP07[b]	4		this work		

[a] Integrated hydropathy analysis performed here using the AMP07 and MPH methods that correlate best with the standards of category A (Table 2). The predicted number of transmembrane helices derives from the consensus of the predictions obtained in different sequences.
[b] Integrated hydropathy analysis performed here using the RAOAR and AMP methods.

peaks above the baseline by condensing together some closely spaced hydrophobic segments that are present in the haem-binding domain of the protein (Fig. 4 and [40, 47]), thereby causing an overestimation of the number of transmembrane helices (results not shown).

Other membrane proteins with redox function

Despite its uncertain role in the enzyme, subunit III of cytochrome oxidase can be broadly classified among membrane proteins with redox function since it is coded by the mitochondrial genome as the other two subunits that carry the metal centers of the oxidase [54]. The availability of many sequences for this subunit of cytochrome oxidase allows a detailed hydropathy analysis for assessing the likely number of its transmembrane helices. Only four of the seven [54] or six [16] putative transmembrane helices are consistently predicted to span the membrane (Table 5) and correspond to helices I, III, V and VI of the previous nomenclature [54].

The sequence of 3-hydroxy–3-methylglutaryl-CoA reductase has been analyzed and the consensus of the predictions obtained by the MPH, AMP07 and Rao and Argos [16] methods indicates that the protein probably contains six transmembrane helices (Table 5). Thus, one region in the α helix and two in β-sheet conformation, which have been previously proposed to span the membrane in this protein [55], are unlikely to transverse the lipid bilayer.

A valuable test of the integrated hydropathy analysis will be provided by the imminent resolution of the crystal structure of the photosystem I of *Synechococcus* [56, 57]. Both subunits A (Fig. 4) and B (not shown) of the system in chloroplasts [58] are predicted to possess ten transmembrane helices. Since these photosynthetic proteins are highly conserved in all organisms [25, 57, 58], such predictions are tentatively extended to the subunits of the *Synechococcus* photosystem.

Membrane proteins with transport function

The integrated hydropathy analysis has been carried out only in membrane proteins with transport function that show an estimated membrane composition that is at least 70% correlated with either the AMP07 or the RAOAR scale. By combining the predictions obtained from several methods it is found that the bacterial and chloroplast permeases [44, 58, 59], the family of the mitochondrial carriers [14] and of the ATPase-6 subunit of F_0-F_1-ATPase [60, 61] show this minimal correlation. The predicted transmembrane composition of these proteins is also better correlated with that of the bacterial reaction centers than with that of rhodopsins (results not shown). These observations indicate that the methods introduced here (Table 3), which are based on scales that show high correlation with the reaction-center standard (Table 2), may also provide a reliable estimation of the folding of these non-redox integral proteins.

Mitochondrial solute carriers

Various mitochondrial carriers such as the adenine-nucleotide translocator and the uncoupling protein of brown adipose tissue can be grouped into a single superfamily [14]. The integrated hydropathy analysis performed on the available sequences of these carriers (Fig. 5 and results not shown) suggests that two of the six transmembrane helices that have been previously proposed, namely helices B and D [14], are unlikely to span the membrane. The carriers may also possess

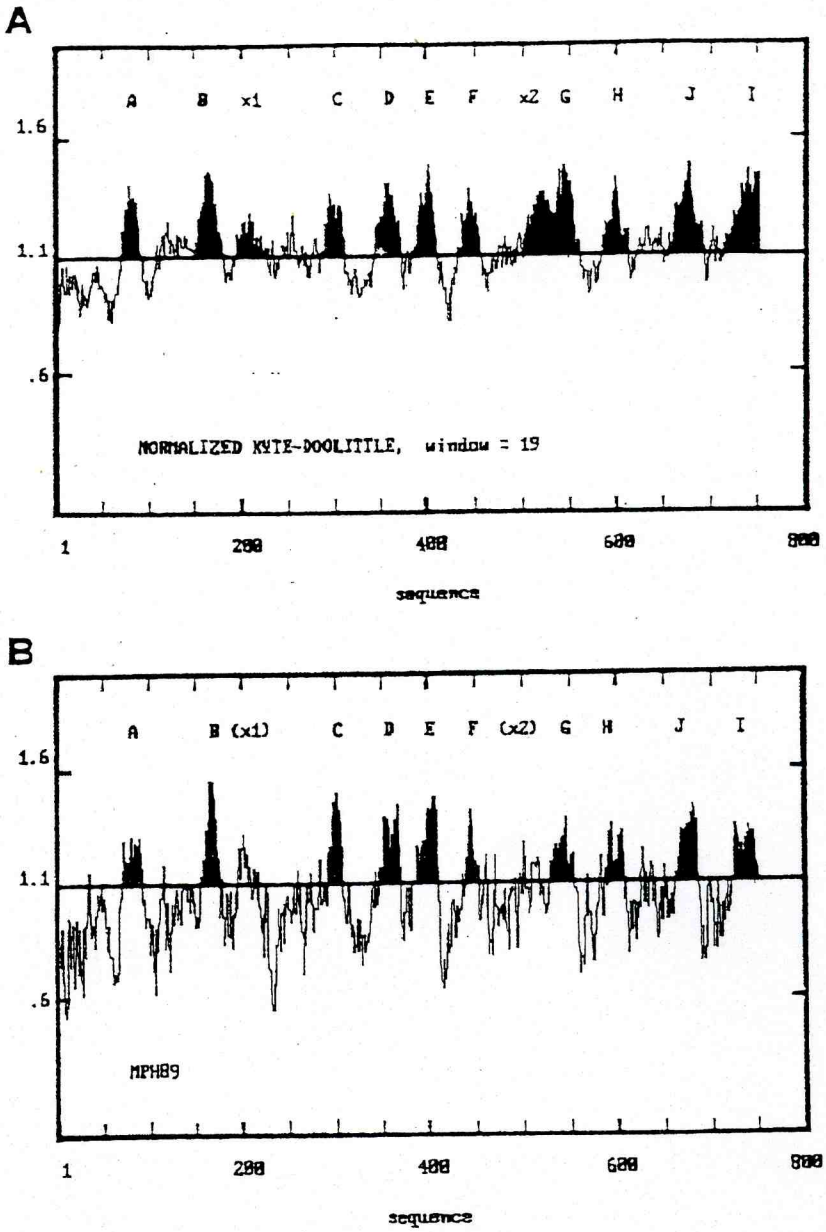

Fig. 4. *Hydropathy plots of the subunit A of photosystem I from the chloroplast of liverwort.* The sequence of the *psaA* gene from the chloroplasts of the liverwort *Marchantia polymorpha* [58] is analyzed with the normalized Kyte and Doolittle scale (A) and the MPH89 scale (B). Ten transmembrane segments are predicted by the MPH method and this is confirmed by the analysis of the homologous protein in tobacco and also with the AMP07 method (results not shown). The normalized Kyte and Doolittle method (Tables 3 and 4), however, predicts two additional transmembrane segments (identified as X1 and X2). It is interesting to note that this protein may contain twice the transmembrane segments of the D_1 subunit (see Fig. 2)

transmembrane segments in β conformation [14], but this remains to be experimentally proven.

The F_0-ATPase subunits

Whereas the membrane topology of the proteolipid (subunit c in *E. coli* [60]) of the F_0-F_1-ATPase is fairly established [60, 61], that of the larger subunit of F_0 that is homologous to subunit a in *E. coli* (i.e. the product of the mitochondrial ATPase-6 gene [60, 61] or of the chloroplast ATPase-IV gene [58]) is still unclear. Current models postulate that all the members of the F_0 subunit a possess five transmembrane segments [61, 62], but Rao and Argos [16] concluded that only four transmembrane segments are consistently predicted. The integrated analysis performed with the AMP07 plus the modified Rao and Argos procedures (Table 3) has been extended to ten ATPase-6 proteins (besides those in [16] also those listed in [63]), to subunit a of *E. coli* [60] and to the ATPase-IV proteins listed in [58].

Helices I, II, III and V are consistently predicted to span the membrane in all cases, even if helix II is sometimes not clearly resolved from helix III (Fig. 6). However, helix IV is less hydrophobic than the others in several cases and is clearly excluded from transversing the membrane in the chloroplast proteins (Fig. 6). Hence, it is inferred that helix IV does not traverse the membrane in all the members of this F_0-F_1-ATPase subunit, and possibly runs parallel to the membrane plane where it is involved in protein contacts with other sub-

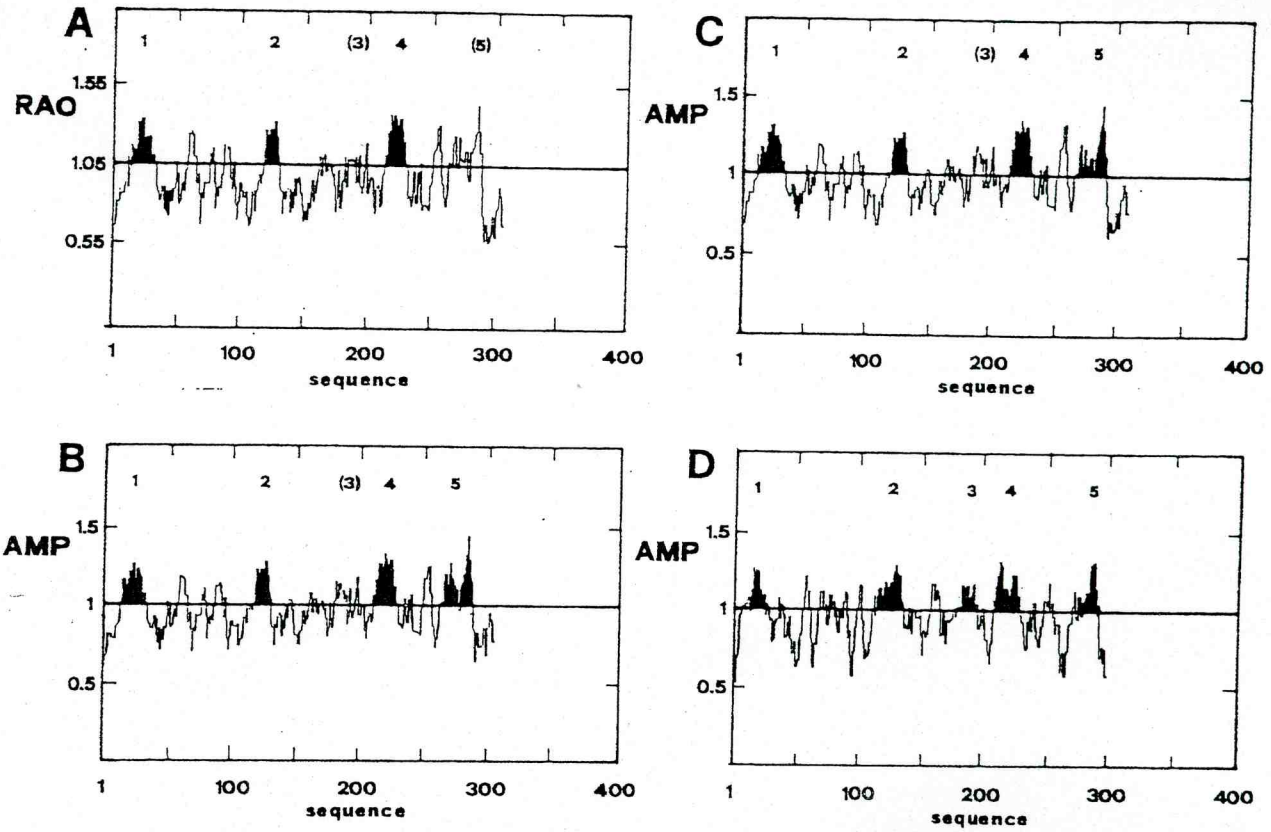

Fig. 5. *Hydropathy plots of mitochondrial carriers.* The sequence of the uncoupling protein from hamster is analyzed with the Rao and Argos scale (abbreviated as RAO) in (A) and with the AMP07 scale (abbrevated as AMP) in (B). The AMP07 hydropathy plot of the mouse uncoupling protein is shown in (C) and that of the beef ATP/ADP translocator in (D). The sequences are taken from [14]. The numbers refer to the transmembrane segments that are predicted by the normalized Kyte and Doolittle method (not shown) and they correspond to helices A and C—F that were previously predicted [14]. The results are basically the same in the beef phosphate carrier and in the rat uncoupling protein (results not shown). Note that helix 5 is not always predicted to be transmembrane

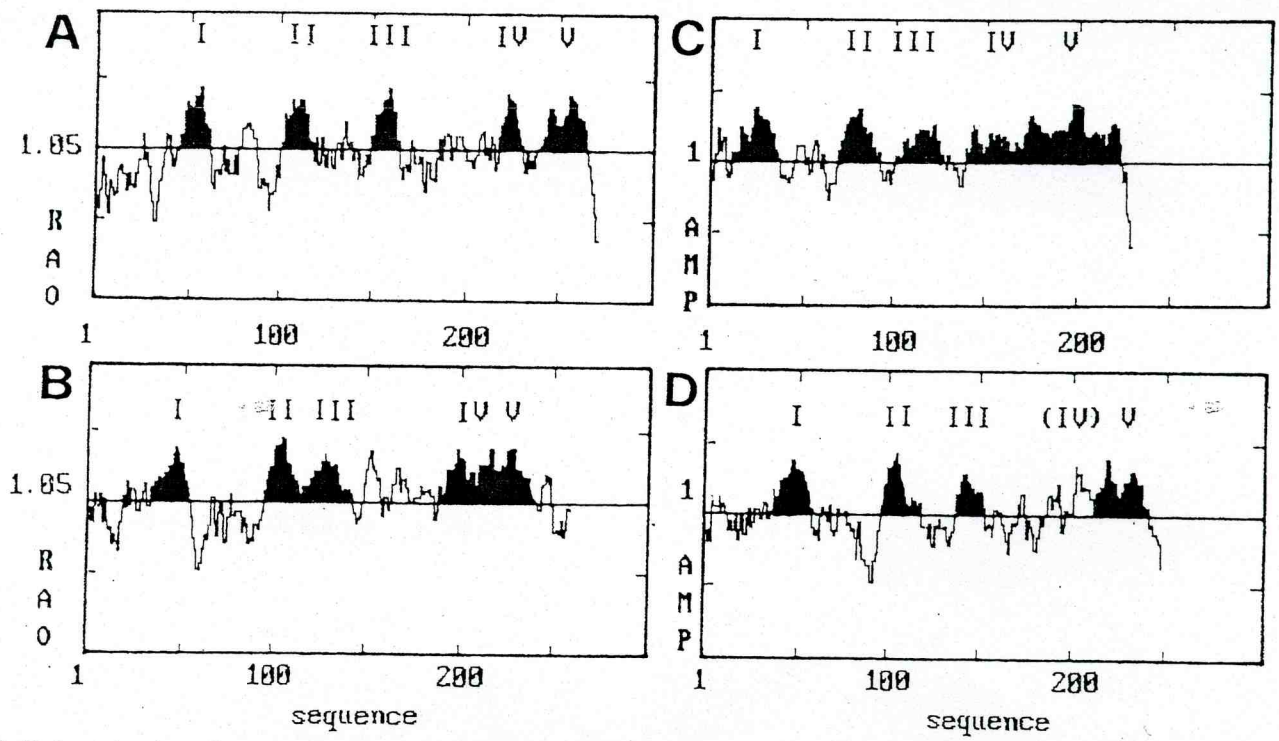

Fig. 6. *Hydropathy plots of various F_0-ATPase subunits.* (A) Subunit a of *E. coli* [60] with the Rao and Argos method (Table 3 and 4). (B) ATPase-6 subunit from yeast mitochondria (listed in [16]) with the same method as in (A); (C) ATPase-6 subunit from sea urchin mitochondria [63]. (D) ATPase IV subunit (homologous to the above proteins) [58, 61] of liverwort chloroplasts [58]. In both (C) and (D) the AMP07 method (abbreviated to AMP) is applied (Tables 3 and 4). The MPH plot at a baseline of 1.1 identifies only four transmembrane helices in (A), (B) and (D) (results not shown)

units. Indeed, the region corresponding to helix IV contains the most conserved stretch of all the sequences [60, 61], a situation that would be uncommon if the helix were buried in the lipid membrane [11].

Three points support this putative topology: (a) a transmembrane four-helical bundle is a rather stable structural motif in the low-dielectric lipid phase and is found in the core of the bacterial reaction centers [11] and in the acetylcholine receptor [16]; (b) the situation of hydropathy predictions for the F_0-subunit family is very much the same as that found previously in the family of cytochrome *b*, in that the chloroplast members of the family (i.e. cytochrome b_6) show a hydrophilic region corresponding to a stretch that is moderately hydrophobic in the mitochondrial members [16, 17] and rather hydrophobic in the bacterial members [17, 64] (see Fig. 6 and [16,17,64]). In the cytochrome-*b* family it is now established that this putative transmembrane helix IV does not actually transverse the membrane in all cases [17], the major evidence being the genetic location of the point mutations responsible for the resistance to inhibitors that bind to regions exposed at opposite sides of the membrane in the protein [64, 65]; (c) two groups of mutations leading to resistance to the antibiotic oligomycin [61, 66] can be located at the same side of the membrane if helix IV is excluded from transversing the bilayer. These mutations, in fact, map at the beginning of both helix IV and V and would lie at the opposite side of the membrane if helix IV is transmembrane [61]. This situation is again very similar to that existing in the cytochrome-*b* family, where mutations leading to resistance to antibiotics such as antimycin would lie at the opposite side of the membrane if helix IV is transmembrane [64, 65].

CONCLUSIONS

Some important conclusions may be drawn from this investigation. Firstly, it is clear that some hydrophobicity scales, particularly those based on statistical analysis (Table 1 and [16, 17]), correlate much better than others with the transmembrane distribution of the residues in some types of integral proteins (Table 2 and Fig. 2). This contrasts with common views that hydropathy scales might be nearly equivalent in evaluating the hydrophobicity of membrane proteins [1, 13, 36, 37].

Secondly, an integration of the most appropriate methods is required to achieve a reliable deduction of the membrane topology in a given polytopic protein, since all methods are prone to errors. Whenever a single procedure is to be adopted, that of Rao and Argos [16], particularly with the modifications introduced here (Table 3), should be preferred because of its highest cumulative correlation and general accuracy.

A third conclusion addresses the fundamental problem of the circular arguments in the statistical/empirical approach [37–39]. The frequencies for elaborating the statistical scales used here (Table 1) contain a significant contribution from helices that were erroneously considered to be transmembrane, such as those previously predicted in cytochrome P_{450} [16] or helix IV in mitochondrial cytochrome *b* [17]. In spite of this, these helices are excluded from the membrane by an optimal computation of the data and by using consistent rules in the interpretation of the hydropathy profiles (Fig. 3) [17]. Therefore, the predictions are not biased towards the *a priori* evaluation of the topology of the proteins, and the circularity in the statistical approach is actually less important than might have been thought.

In perspective, it is hoped that this critical evaluation of the hydropathy methodology will provide a step forward in the implementation of structural deductions for membrane proteins and will guide future experimental approaches for probing the membrane topology of integral polypeptides.

We thank A.Ghelli and R.Luchetti for their help in compiling the data base. Discussions with Dr. P. Argos (Heidelberg), Prof. G. Lenaz, Prof. R. Casadio and Prof. B. Melandri (Bologna) are gratefully acknowledged. This work has been partially sponsored by the *Ministero della Pubblica Istruzione*, Rome, Italy.

REFERENCES

1. Von Heijne, G. (1988) *Biochim. Biophys. Acta 947*, 307–333.
2. Von Heijne, G. & Blomberg, C. (1979) *Eur. J. Biochem. 97*, 175–181.
3. Engelman , D. M. & Steitz, T. A. (1981) *Cell 23*, 411–422.
4. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.
5. Eisenberg, D., Weiss, R. M., Terwilliger, T. C. & Wilcox, W. (1982) *Faraday Symp. Chem. Soc. 17*, 109–120.
6. Eisenberg, D. (1984) *Annu. Rev. Biochem. 53*, 595–623.
7. Engelman , D. M., Steitz, T. A. & Goldman, A. (1986) *Annu. Rev. Biophys. Biophys. Chem. 15*, 321–353.
8. Argos, P., Rao, J. K. M. & Hargrave, P. A. (1982) *Eur. J. Biochem. 128*, 565–575.
9. Lundeen, M., Chance, B. & Powers, L. (1987) *Biophys J. 51*, 693–695.
10. Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. (1985) *Nature 318*, 618–624.
11. Yeates, T. O., Komya, H., Rees, D. C., Allen, J. P. & Feher, G. (1987) *Proc. Natl Acad. Sci. USA 84*, 6438–6442.
12. Michel, H., Weyer, K. A., Gruenberg, H., Dunger, I., Oesterhelt, D. & Lottspeich, F. (1986) *EMBO J. 5*, 1149–1158.
13. Klein, P., Kanehsa, M. & De Lisi, C. (1985) *Biochim. Biophys. Acta 815*, 468–476.
14. Aquila, H., Link, T. A. & Klingenberg, M. (1987) *FEBS Lett. 212*, 1–9.
15. Bangham, J. A. (1988) *Anal. Biochem. 174*, 142–145.
16. Rao, M. J. K. & Argos, P. (1986) *Biochim. Biophys. Acta 869*, 197–214.
17. Degli Esposti, M., Ghelli, A. M., Luchetti, R., Crimi, M. & Lenaz, G. (1989) *Ital. J. Biochem. 38*, 1–22.
18. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry 13*, 211–222.
19. Chothia, C. (1984) *Annu. Rev. Biochem. 53*, 537–572.
20. Cornette, J. L., Cease, K. B., Margalit, T. H., Spouge, J. L., Berzofsky, J. A. & De Lisi, C. (1987) *J. Mol. Biol. 195*, 659–685.
21. Argos, P. (1987) *J. Mol. Biol. 193*, 385–396.
22. Dunn, R. J., Hackett, N. R., Huang, K. S., Jones, S. S., Khorana, H. G., Liao, M. J., Lo, K. M., Satterthwait, A., Seehra, J. S & Yatsunami, K. (1983) *Cold Spring Harbor Symp. Quant. Biol. 48*, 853–862.
23. Khorana, H. G. (1988) *J. Biol. Chem. 263*, 7439–7442.
24. Schobert, B., Lanyi, G. & Oesterhelt, D. (1988) *EMBO J. 7*, 905–911.
25. Gray, J. C. (1987) in *Photosynthesis* (Amesz, J., ed.) p. 319–342, Elsevier, Amsterdam.
26. Karabin, G. D., Farley, M. & Hallick, R. B. (1984) *Nucleic Acid Res. 12*, 5801–5812.
27. Erickson, J. M., Rhaire, M., Malnoe, P., Girard-Bascou, J., Pierre, Y., Bennoun, P. & Rochaix, J. D. (1988) *EMBO J. 7*, 1745–1752.
28. Nathans, J. & Hogness, D. S. (1983) *Cell 34*, 807–814.
29. O'Tousa, J. E., Baehr, W., Martin, R. L., Hirsh, J., Pak, W. L. & Applebury, M. L. (1985) *Cell 40*, 839–850.
30. Chung, F. Z., Lentes, K. U., Gocayne, J., Fitzgerald, M., Robinson, D., Kerlavage, A. R., Fraser, C. M. & Venter, J. C. (1987) *FEBS Lett. 211*, 200–206.
31. Kubo, T., Fukuda, K., Mikami, A., Maeda, A., Takahashi, H., Mishina, M., Haga, T., Haga, K., Ichiyama, A., Kangawa, K.,

Kojima, M., Matsuo, H., Hirose, T. & Numa, S. (1986) *Nature 323*, 411–416.
32. Capaldi, R. A. & Vanderkooi, G. (1972) *Proc. Natl. Acad. Sci. 69*, 930–932.
33. Levitt, M. (1976) *J. Mol. Biol 104*, 59–107.
34. Wolfenden, R., Andersson, L., Cullis, P. M. & Southgate, C. C. B. (1981) *Biochemistry 20*, 849–855.
35. Fauchère, J. & Pliška, V. (1983) *Eur. J. Med. Chem. 18*, 369–375.
36. Von Heijne, G. (1986) *EMBO J. 5*, 3021–3027.
37. Roseman, M. A. (1988) *J. Mol. Chem. 200*, 513–522.
38. Guy, H. R. (1985) *Biophys. J. 47*, 61–70.
39. Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H. & Zehfus, M. H. (1985) *Science 229*, 834–838.
40. Poulos, T. L., Finzel, B. C. & Howard, A. J. (1987) *J. Mol. Biol. 195*, 687–700.
41. Finzel, B. C., Poulos, T. L. & Kraut, J. (1984) *J. Biol. Chem. 259*, 13027–13036.
42. Adams, G. A. & Rose, J. K. (1985) *Cell 41*, 1007–1015.
43. Morden, C. W. & Golden, S. S. (1989) *Nature 337*, 382–385.
44. Boyd, D., Manoil, C. & Beckwith, J. (1987) *Proc. Natl. Acad. Sci. 84*, 8525–8529.
45. Akiyama, Y. & Ito, K. (1987) *EMBO J. 6*, 3465–3470.
46. Nebert, D. W. & Gonzales, F. J. (1987) *Ann. Rev. Biochem. 56*, 945–993.
47. Nelson, D. R. & Strobel, H. W. (1988) *J. Biol. Chem. 263*, 6038–6050.
48. Iwamoto, Y., Tsubaki, M., Hiwatashi, A. & Ichikawa, Y. (1988) *FEBS Lett. 233*, 31–36.
49. Tarr, G. E., Black, S. D., Fujita, V. S. & Coon, M. J. (1983) *Proc. Natl Acad. Sci. USA 80*, 6552–6556.
50. Heinemann, F. S. & Ozols, J. (1983) *J. Biol. Chem. 258*, 4195–4201.
51. Sakaguchi, M., Mihara, K. & Sato, R. (1987) *EMBO J. 6*, 2425–2434.
52. De Lemos-Chiarandini, C., Frey, A. B., Sabatini, D. O. & Kreibich, G. (1987) *J. Cell. Biol. 104*, 209–219.
53. Vergères, G., Winterhalter K. H. & Richter, C. (1989) *Biochemistry 28*, 3650–3655.
54. Finel, M., Haltia, T., Holm, L., Jalli, T., Metso, T., Pvustinen, A., Raitio, M., Saraste, M. & Wikström, M. (1987) in *Cytochrome systems* (Papa, S., Chance, B. & Ernster, L., eds), pp. 247–252, Plenum Press, New York.
55. Liscum, L., Finer-Moore, J., Stroud, R. M., Luskey, K., Brown, M. S. & Goldstein, J. L. (1985) *J. Biol. Chem. 260*, 522–530.
56. Witt, I., Witt, H. T., Di Fiore, D., Roguer, M., Hinrichs, W, Saenger, W., Granzin, J., Betzel, C. & Dauter, Z. (1988) *Ber. Bunsenges. Phys. Chem. 92*, 1503–1506.
57. Ford, B. (1983) *Nature 337*, 510–511.
58. Umesono, K., Inokuchi, H., Shiki, Y., Takeuchi, M., Chang, Z., Fukuzawa, H., Kohchi, T., Shirai, H., Ohyama, K. & Ozeki, H. (1988) *J. Mol. Biol. 203*, 299–331.
59. Maiden, M. C. J., Davis, E. O., Baldwin, S. A., Moore, D. C. M. & Henderson, P. J. F. (1987) *Nature 325*, 641–643.
60. Hoppe, J. & Sebald, W. (1984) *Biochim. Biophys. Acta 768*, 1–27.
61. Senior, A. E. (1988) *Physiol. Rev. 68*, 177–231.
62. Lightowlers, R. N., Howitt, S. M., Hatch, L., Gibson, F. & Cox, G. (1988) *Biochim. Biophys. Acta 933*, 241–248.
63. Jacobs, H. T., Elliott, D. J., Math, V. B. & Farquharson, A. (1988) *J. Mol. Biol. 202*, 185–217.
64. Crofts, A., Robinson, H., Andrews, K., Van Doren, S. & Berry, E. (1987) in *Cytochrome Systems* (Papa, S., Chance, B. & Ernster, L., eds), pp. 617–624, Plenum Press, New York.
65. Di Rago, J. P. & Colson, A. M. (1988) *J. Biol. Chem. 263*, 12564–12570.
66. Macino, G. & Tzagoloff, A. (1980) *Cell 20*, 507–517.
67. Eisenberg, D. & McLachlan, A. D. (1986) *Nature 333*, 345–567.
68. Jacobs, R. E. & White, S. H. (1989) *Biochemistry 28*, 3421–3437.
69. Eisenberg, D., Shwarz, G., Komarony, M. & Wall, R. (1984) *J. Mol. Biol. 179*, 125–142.

APPENDIX

Table 6. *Other correlations among scales not included in Table 2*

Acronym	correlation (r)															
	%															
JONES	100															
ROSEM	65	100														
CHOFA	7	14	100													
CHOTH	42	87	19	100												
WERSC	66	72	23	72	100											
PONNU	54	72	18	83	88	100										
MIJER	65	83	24	84	94	93	100									
NIOII	57	72	14	80	89	95	89	100								
OMHSE	73	71	5	66	80	80	87	76	100							
PRIFT	54	65	10	77	88	91	91	89	86	100						
ARGOS	43	75	32	72	67	72	84	65	73	75	100					
MPH00	23	64	16	76	77	85	79	82	65	79	72	100				
D12MP	35	77	42	84	65	73	69	71	49	59	60	82	100			
RHODO	55	66	13	63	67	56	68	54	74	59	49	51	37	100		
MICHE	10	57	14	71	67	71	67	75	43	67	63	87	72	42	100	
AMP07	38	82	8	88	81	86	87	85	73	82	78	94	87	69	87	100
	JON	ROS	CHO	CHO	WER	PON	MIJ	NIO	OMH	PRI	ARG	MPH	D12	RHO	MIC	AMP

The acronyms used identify the following scales: JONES, see [20]; ROSEM, physicochemical scale recently introduced by Roseman [37]; CHOFA, α-helix preference parameters of Chou and Fasman [18]; CHOTH, first statistical scale of hydrophobicity derived from the known structure of globular proteins [19]; WERSC, PONNU, MIJER, NIOII, statistical scales derived from the evaluation of the crystal structures of globular proteins that show the highest correlation with the PRIFT scale (see [20] for the pertinent references); OMHSE, so-called optimal mean hydrophobicity introduced by Sweet and Eisenberg (1983) and identified in [20] by the acronym SWEET (See [20] for the reference); ARGOS, first statistical scale of membrane preference introduced by Argos et al. [8]; MPHOO, orginal scale of membrane propensity for haemoproteins, where the value of histidine is 1.44 [17]; this scale is correlated at 98% with the MPH89 scale utilized here (Table 4); D12MP, RHODO. MICHE and AMP07, statistical scales introduced in this work (Table 1)