CROATICA CHEMICA ACTA CCACAA **68** (1) 215–232 (1995)

ISSN 0011-1643
CCA–2224

Original Scientific Paper

A Scale of β-Preferences for Structure-Activity Predictions in Membrane Proteins

Davor Juretić and Roko Pešić

Department of Physics, N. Tesle 12, University of Split, HR-21000 Split, Croatia

Received August 16, 1994; revised November 17, 1994; accepted November 20, 1994

Physical chemists are often interested in predicting molecular properties from the known molecular structure. In this paper, a different approach is taken to predicting secondary structure features, membrane folding motifs and functionally important amino acid residues in membrane proteins, starting from physical, chemical or statistical properties of twenty natural amino acid types. Empirical knowledge of protein structure and property scales is combined through the preference function method so as to predict the secondary structure of membrane proteins in the three state model (α-helix, β-sheet, coil or turn). Of the 140 property scales examined, our own scale of β-sheet preferences, extracted from porins and defensins, is the best in predicting the sequence location of transmembrane helices. It is also shown that functionally important amino acids located in such helices can be predicted with satisfactory accuracy for the case of bacteriorhodopsin and the potassium channel subunit, where identity of amino acid residues involved in the ion transport activity is fairly well established.

INTRODUCTION

Membrane proton pumps, such as bacteriorhodopsin and cytochrome c oxidase, and membrane channels, such as the potassium channel, are integral membrane proteins with a crucial role in cellular free-energy transduction and cellular communication. According to the chemiosmotic hypothesis, dissipative proton currents are created by membrane proton pumps and subsequently used for adenosine triphosphate production.[1] Kinetic models for proton pump activity require a knowledge of their structure. Identity of proton donors and acceptors, which are parts of the proton-pumping mechanism must be known to propose the role of membrane-embedded helices in promoting vectorial proton transfer across membrane.

An attempt to integrate experimental data concerning the properties of amino acids, empirical knowledge of protein structure and function and theoretical structure-function prediction is presented here for membrane proteins in general and for bacteriorhodopsin and potassium channel in particular. Such a property-structure-function information flow is not circular, because the final property of interest, the

mechanism of ion transport through membrane protein, is quite difficult to predict from the properties of individual amino acids or amino acid residues in protein.

Bacteriorhodopsin from *Halobacterium salinarium* (old species name: *Halobacterium halobium*) froms 2-dimensional crystals that can be examined with electron microscopy, but no procedure for growing 3-dimensional bacteriorhodopsin crystals has been discovered and X-ray crystallography could not determine the secondary structure. Nevertheless, Henderson's results with electron microscopy[2] are accurate enough to locate transmembrane α-helices in the primary structure. These 7 helices are labelled from A to F, starting from the protein N-terminal:

A (10–32), B(38–62), C(80–101), D(108–127), E(136–157), F(167–193) and G(203–227). Numbers in the parentheses are sequence positions for the helix N-terminal and C-terminal cap, respectively.

Bacteriorhodopsin is the best known example of an »inside-out« protein, while soluble globular proteins have an predominantly hydrophilic surface and hydrophobic interior, the bacteriorhodopsin exterior is highly hydrophobic. Its surface hydrophobic groups are in contact with membrane phospholipids. Polar and charged amino acids in the bacteriorhodopsin interior often have an important structural or functional (proton donors or acceptors) role.

Amino acids directly involved in the proton pumping activity are known today as an result of intensive research effort during the past ten years in producing, expressing and examining bacteriorhodopsin mutants.[3] Site-specific mutagenesis studies leave no doubt that such amino acids are arginine 82, aspartate 85 and aspartate 96 in helix C, aspartate 115 in helix D, and aspartate 212 and lysine 216 in helix G.[3–5]

A method capable of predicting the location of transmembrane helices and of predicting functionally important amino acid residues within such helices would reduce the labour of producing hundreds of mutants containing single amino acid replacements. The method would not be of practical use for bacteriorhodopsin, but it may be useful for other integral membrane proteins involved in ion transport and bacteriorhodopsin could be used to test its predictive power.

In this work, we shall use the preference functions method[6] both to predict the primary structure location of membrane helices and to locate residue-candidates with unusual sequence environments inside such helices. Ionizable groups within the membrane buried active site certainly undergo very strong electrostatic interactions in their nonpolar three-dimensional environment. We wanted to see if it was possible to recognize from sequence patterns those charged residues that have a special role in the active site formed by transmembrane helices. One would expect to see a decrease in the helical propensity for such residues. We hoped that the low helical preference might be an indication of the important functional role of charged residues predicted to be membrane buried in the transmembrane helix. Sequence dependent conformational preferences derived from different amino acid property scales must be used to examine this problem. One hundred and forty scales of amino acid properties were examined with respect to their predictive power in predicting the membrane protein secondary structure, transmembrane helices in membrane proteins, bacteriorhodopsin transmembrane helices and amino acid residues involved in the proton pumping activity of bacteriorhodopsin. As it could be expected, a single property of 20 natural amino acids was not enough to produce the best predictor for each of the four goals listed above. All six amino acids involved in the pro-

ton pumping activity of bacteriorhodopsin were located with our own preference scale. Of the nine amino acids important for potassium transport *via* the potassium channel six were identified using our preference scale. In both cases the number of overpredictions is such that prediction accuracy for functionally important amino acids is not much better than 50%.

METHODS

Secondary Structure Prediction for Membrane Proteins

The secondary structure prediction method that makes use of property scales to extract preference functions has been described elsewhere.[6,7] The algorithm called PREF (previously PR and NORM) extracts preference functions from the database of proteins used for training, while the algorithm called SP is used to test prediction accuracy on a set (or one) of tested protein structures. The same property scale was used as input for PREF and SP. Class limits input for PREF[6] were calculated by using algorithm DIS with the same property scale applied to the list of 10 proteins: 155c, 1aapa, 1abp, 1acx, 1ak3a 1alc, 1bbpa, 1bp2, 1ca2 and 1cc5. These are Brookhaven Protein Data Bank codes[8] with the additional fifth symbol »a« (in some cases) indicating that only the first polypeptide chain has been taken from proteins having more than one polypeptide chain.

The structure was assigned to each residue according to the highest preference. The decision was taken after a simple smoothing procedure for preferences,[9] which eliminated many unrealistic predictions of single helical or sheet residue conformations surrounded with turn, coil or undefined conformations. The remaining of such cases were eliminated by the explicit requirement in the SP algorithm. The first and last two residues of each polypeptide were assigned undefined (not alfa or beta) conformation.

Protein Data Bases

Ideally, membrane proteins of known structure should be used to train preference functions when membrane protein secondary structure is predicted. In practice, there are not enough membrane proteins of known structure to perform the training procedure. One can then use as the training database a) soluble proteins of well knwon structure,[7] b) membrane proteins of incompletely known structure[10] or c) a mixture of soluble and membrane proteins as in the present work.

The membrane proteins tested were bacteriorhodopsin (bR), and photosynthetic reactions centre (prc) L, M and H subunits from *Rhodopseudomonas viridis*(v) and *Rhodobacter sphaeroides*(s), a total of 7 polypeptides. We denote this testing list used by other authors as well[11] as P1. Experiments with different decision constants determined that the β-sheet structure must be suppressed when membrane proteins with transmembrane helices are predicted. Therefore, we used DCB=–0.6, DCH=0.35 and DCC=0.0 for the β-sheet, α-helix and coil decision constants, respectively, except when subunits H with only one transmembrane helix were tested. In the case of subunit H, no decision constant were used (DCB=DCH=DCC=0).

Only β-class soluble proteins were used in the mixture of soluble and membrane proteins for the training procedure. To achieve greater generality, most results were obtained by using two different training data bases of proteins. Training database

T1 : 37 proteins consisting of a total of 18 β-class soluble proteins known with resolution equal or better than 2.5 Å, and a total of 19 membrane proteins whose secondary structure or only transmembrane helices have been determined with different degree of accuracy. Soluble proteins are: 1bbpa, 1fdla, 1rei, 2rspa, 1fkf, 1pfc, 2stv, 3ebx, 2pcy, 1cd4, 1acx, 1paz, 2i1b, 1rbp, 2gcr, 2alp, 2fb4, 2fbj. Membrane polypeptides are 5 polypeptides from photosynthetic reaction centre (subunits L, M and H from *viridis* and subunit L and M from *sphaeroides*), light harvesting complex from pea leaves chloroplasts,[12] melittin, annexin,[13] light harvesting complex from tomato (LH11_LYCES), prostaglandin H2 synthase-1,[14] synaptophysin (SYPH_RAT), protein export protein (SECY$ECOLI), two maltose transaport proteins (MALF_ECOLI and MALG_ECOLI), lactose permease (LACY$ECOLI), glucose transporter (GTR1$HUMAN), potassium channel (CIK1$DROME), cytochrome b561 (C561$BOVIN) and Na/K ATP-ase (ATN1$SHEEP). Capital letters in the parentheses are the SWISPROT protein identification codes.

Training database T2: 58 proteins consisting of 37 β-class soluble proteins known with resolution equal or better than 2.5 Å, and a total of 21 membrane or membrane-active proteins with the known secondary structure or with location for transmembrane helices proposed by using different experimental methods. Database T2 *does not* contain photoshynthetic reaction centre polypeptides, nor bacteriorhodopsin. There are no melittin, light-harvesting complexes and Na/K ATP-ase in T2 either, so that only 10 membrane polypeptides are shared among T1 and T2. The new ones present only in T2 are the arabinose transporter (ARAE$ECOLI), muscarinic acetylcholine receptor's first 240 amino acids (ACM5$HUMAN), matrix M2 protein (VMT2$INABA), rhodopsin,[15] cytochrome b,[16] cytochrome oxidase subunit I,[17] glycophorin (GLP$PIG), β-2 adrenergic receptor (B2AR$MESAU), L-type calcium channel (CICA$RABIT), sodium channel first 800 amino acids (CINA$ELEEL), and colicin A pore-forming domain.[18] Of 37 soluble proteins, 15 are shared between T1 and T2. Proteins 1fkf, 1pfc and 2stv are present only in T1, while 2azaa, 2rhe, 1mcpa, 2soda, 2tbva, 2sga, 3sgba, 1tona, 2pkaa, 1trma, 2ptn, 2gcha, 3est, 1hnea, 3rp2, 1sgt, 2ltn, 2cna, 5pep, 4cmsa, 4ape and 2apr are present only in T2. Therefore, less than 50% of proteins in T2 (25) are shared with T1. The T2 data base was used to calculate all the reported results.

Performance Parameters

The success of predicting the secondary structure of membrane proteins, using each scale, was ranged according to the $Q3$ index, which is the percentage of residues predicted in correct conformation when the 3-state conformational model is used: residues are assumed to be in the α-helix or β-sheet or undefined (turn or coil) conformation. The performance parameters A_a and B_a, used in Table II, are discussed in Juretić *et al.*[19] In terms of positive correct predictions p, overpredictions o and underpredictions u, our performance parameters can be expressed as:

$$Q_3 = (p_\alpha + p_\beta + p_c)/N, \quad A_\alpha = p_\alpha/(p_\alpha + u_\alpha + o_\alpha), \quad \text{and} \quad B_\alpha = (p_\alpha - u_\alpha - o_\alpha)/N \qquad (1)$$

where α, β and c denote α-helix, β-sheet and undefined (turn or coil) conformations, respectively. The total number of residues considered is N. Table II and III results contain results for several additional performance parameters. These are the frac-

TABLE I

Correlation (r) between 11 best property scales for predicting transmembrane helices of bacteriorhodopsin and photosynthetic reaction centre (see Methods). The acronyms used are listed in the last part of the Methods section. The number in the parentheses identifies each scale as one of the 140 scales tested. All correlations are positive and are expressed as percentages.

Acronym											
KYTDO(1)	100										
SWEET(24)	75	100									
NNEIG(59)	85	77	100								
WERSC(58)	71	80	87	100							
CIDAB(105)	79	87	82	87	100						
BULDG(57)	60	88	63	75	81	100					
PRIFT(15)	81	86	89	88	86	77	100				
OPT21(109)	83	73	87	81	78	58	86	100			
FASBET(32)	65	81	82	74	74	77	75	66	100		
CHOUBB(80)	58	78	67	68	67	71	69	71	79	100	
PORBET(86)	53	68	69	64	61	67	63	63	70	72	100
	1	24	59	58	105	57	15	109	32	80	86

tion $Q_i = p_i/(p_i+u_i)$ of residues correctly predicted to be in class i relative to those observed to be in class i, Matthew's coefficients C_i^{20} and

$$C_g = (p - u)/N \tag{2}$$

where

$$p = p_\alpha + p_\beta + p_c, \quad u = u_\alpha + u_\beta + u_c .$$

The C_g and C_i parameters range is from –1 to +1.

TABLE II

The best property scales (out of 140) for 20 natural amino acids. Results obtained with the training database T2 (see Methods) and testing database P1 consisting of bacteriorhodopsin and the photosynthetic reaction centre: a total of seven membrane polypeptides (see text).

SCALE	Q_3	Q_α	C_α	A_α	B_α	A_{m1}	A_{m2}	A_{m3}	Q_2
PORBET	71	80	.49	.63	.22	.73	.79	.81	92
EDEL21	70	74	.51	.61	.19	.64	.74	.77	90
WERSC	70	75	.47	.61	.21	.51	.63	.70	86
NNEIG	69	76	.49	.62	.21	.66	.74	.78	90
PRIFT	69	75	.46	.60	.19	.61	.72	.76	90
BULDG	69	77	.43	.60	.19	.41	.58	.66	84
FASBET	69	71	.47	.59	.19	.36	.56	.65	84
CHOUBB	68	74	.46	.59	.18	.62	.74	.77	90
SWEET	68	74	.43	.60	.17	.48	.63	.70	86
CIDAB	67	76	.40	.58	.16	.62	.72	.76	90
KYTDO	65	69	.37	.54	.10	.54	.70	.74	89

TABLE III

Prediction of transmembrane helices and of complete secondary structure of membrane proteins with our preference scale PORBET (see Methods for the definition of the performance quality parameters used in the first column and for the explanation of protein acronyms used in the first row). N_{TM} is the total number or residues observed in the transmembrane helix conformation

PROT:	*bR*	prcl(v)	prcm(v)	prch(v)	prcl(s)	prcm(s)	prch(s)	TOTAL
$Q_3(\%)$	87	69	77	63	69	66	64	71
C_g	0.74	0.38	0.53	0.26	0.39	0.32	0.28	0.41
$C_\alpha(\%)$	0.71	0.35	0.56	0.43	0.41	0.42	0.55	0.49
Q_α	97	85	84	53	83	80	73	79
A_α	0.83	0.65	0.71	0.39	0.68	0.66	0.46	0.63
B_α	0.51	0.25	0.30	–0.07	0.30	0.25	–0.02	0.22
u_m	5	8	9	2	6	9	5	44
o_m	25	10	32	1	29	15	0	122
p_m	159	117	118	22	124	121	21	682
N_{TM}	164	125	127	24	130	130	26	726
A_{m1}	0.79	0.79	0.61	0.79	0.69	0.75	0.62	0.73
A_{m2}	0.82	0.86	0.68	0.88	0.73	0.82	0.81	0.79
A_{m3}	0.84	0.87	0.74	0.88	0.78	0.83	0.81	0.82
$Q_2(\%)$	88	93	87	99	88	92	98	92

We also use the performance parameters for predicting transmembrane helices:

$$Q_2=(p_m + p_c)/N, \quad A_{m1} = (p_m - u_m - o_m)/N_m, \quad A_{m2} = (N_m - u_m - o_m)/N_m \quad \text{and} \quad A_{m3} = p_m/(p_m + u_m + o_m) \tag{3}$$

where Q_2 is the percentage (when multiplied by 100) of corectly predicted conformations in the 2-state model: residues in transmembrane α-helices and all other residues. Numbers p_m, u_m, o_m and p_c are, respectively, the numbers of residues observed and predicted in transmembrane helices, observed but not predicted, not observed but predicted, and neither observed nor predicted in transmembrane helices. The total number of residues observed in transmembrane helices is N_m.

Recognition of Observed and Predicted Transmembrane Helices

Transmembrane α-helices are defined as observed helices in integral membrane proteins of α-class which contain at least 19 sequential residues in helical conformation. Predicted transmembrane α-helices are defined as having a continuous stretch of at least 18 residues predicted to be in the α-helix conformation. In the present version of SP algorithm, there was no attempt to rescue as predicted transmembrane helices hydrophobic segments of about right size were interrupted with only one or two polar residues not predicted in helical conformation. Predicted continuous helical segments longer than 36 amino acid residues are either broken into two transmembrane helices or shortened at the N- or C-terminal. Sequence positions of maximal turn propensity are used to find new N- and C-terminals of predicted

transmembrane helices. Neither transmembrane helix-surface amphipathic helix-transmembrane helix nor transmembrane helix-β-strand-transmembrane helix motifs are recognized.

Property Scales

The complete list of property scales, each consisting of 20 parameters for 20 natural acid types, is not presented here, because among 140 such scales considered, including almost all hydrophobicity scales proposed in the literature, only 19 scales have been selected as top runners for secondary structure prediction of membrane proteins. In the subsequent selection cycle, correlation coefficients among all previously selected property scales were calculated. From two or more very similar property scales (with a correlation coefficient higher than 0.90) only one was selected. This procedure left 11 property scales which are listed below. Table I lists the correlation (r) between all those scales. The results obtained with such selection of scales are presented and discussed in the next sections.

Amino acid property scales used in this work for different prediction goals (the number next to each acronym denotes which one of 140 property scales has been selected):

BULD (57): change in surface tension of water.[21]

FASBET (32): Chou-Fasman's β-sheet preferences.[21]

PORBET (86): β-sheet preferences extracted by us from 7 porins and 2 defensins. The nonstandard code that we used for these proteins is followed by reference numbers for primary and secondary structures: OMPF,[23] PORI[24,25] OMPA,[26] OMP32,[27] PHOE,[23,28] VDAC,[29] VDACH,[30] DEFE,[31] HNP3.[32]

SWEET (24): Optimal matching hydrophobicity scale.[33]

WERSC (58): protein-specific hydrophobicity scale based on the ratio of in/out residues.[34]

KIDBET (136): β-structure preference.[35]

PRIFT (15): Optimal scale for amphipathic helices.[36]

NNEIG (59): protein specific hydrophobicity scale based on hydrophobicities of nearest neighbours.[36]

CIDAB (105): Hydrophobicity scale for proteins of α+β class.[37]

CHOUBB (80) β-sheet preferences from β-class proteins.[38]

EDEL21 (109): Optimal predictor of Edelman.[11]

RESULTS

Secondary structure prediction of membrane proteins: the performance of different property scales

Our goal was to find the best properties for predicting the membrane protein secondary structure, transmembrane helices in such proteins and functionally important amino acids in such helices. From our database of 140 property scales we first found 11 best predictors for the membrane protein secondary structure (Table II). All 11 best scales can be grouped into two categories: β-sheet preference scales and hydrophobicity scales for amino acid residues in a protein (Methods).

The overall results obtained (maximal $Q3$ around 71% and maximal A_{m2} of 0.79) are similar in prediction quality to the neural network results for soluble globular proteins,[39] and to the results for transmembrane helices obtained with best predictors for the location of such helices in membrane proteins.[11,40] The frequently used Kyte-Doolittle hydrophobicity scale[41] gives a much better prediction of transmembrane helices when used through the mechanism of preference functions (A_{m2} = 0.76 for photosynthetic reaction centre from *R. viridis* transmembrane segments) than directly[40] (A_{m2} = 0.45) or as the sieved Kyte-Doolittle method[40] (A_{m2} = 0.56) applied at the same testing set of three membrane polypeptides. The EDEL21 scale[11] is even better according to all performance parameters, but that scale was derived by optimizing the predictor on the same data set of seven membrane polypeptides that was used for testing. Almost equally good scales are the self-consistent average surrounding hydrophobicity scale NNEIG,[36] β-preference scale CHOUBB extracted from β-class proteins,[38] and the optimum scale PRIFT for amphipathic structures.[36] Another scale based on the Manavalan and Ponnuswamy idea of average surrounding hydrophobicities[42] is the CIDAB scale extracted from the class of α+β proteins.[37] Its performance is also associated with an A_{m2} index higher than 0.70 (Table II).

The rank order of property scales depends on the choice of a limited number of membrane proteins of known secondary structure to test the algorithm, on the choice of feature predicted in such proteins (complete secondary structure or transmembrane helices only) and on the choice of performance parameters. This is the reason why we used such a large set of performance parameters, including several parameters that are specific for expressing prediction accuracy for transmembrane helices.

Results are also sensitive to the choice of the training database of proteins. One would expect better results with the T1 training database which contains proteins to be tested (see Methods). Instead, prediction accuracy drops to 69% while A_{m2} drops to 0.74 when the PORBET scale is used with the T1 training database. It may be that the smaller number of proteins in the data base (37 instead of 58) offsets the advantage of having test proteins in the training data base.

The choice of class limits for separation of the sequence environment into classes, the obligatory step during the derivation of preference functions,[6] can also be the cause of a small change in the reported prediction accuracy. When our training list T2 of 58 proteins is used to extract class limits, instead of the shorter list of 10 proteins, the prediction accuracy with our best scale PORBET decreases to A_{m2} = 0.78. The prediction accuracy is actually better for the photosynthetic reaction centre subunits, but transmembrane helix D from bacteriorhodopsin is not predicted because aspartate 115 is predicted in the turn conformation.

Protein by protein comparison of prediction results with our best scale PORBET is shown in Table III. The proteins in the test list contain 29 transmembrane helices. All are predicted but one helix is predicted where none is found in the subunit M sequence from *viridis*. The low prediction accuracy of the helix conformation in the prch(v) (Q_α = 53%) is due to the poor prediction of the helical segment (not in transmembrane conformation) near the C-terminal of that polypeptide. A more detailed comparison of predictions for each transmembrane helix (underpredictions u_m and overpredictions o_m) illustrates that our PORBET scale generally overpredicts residues in the transmembrane helix conformation even when approximate position of such helices and separation between helices is correctly located. If prediction accu-

racy does not suffer too much, predicting slightly longer transmembrane helices is not a disadvantage when it is desired to locate all functionally important residues in such helices.

Predicting transmembrane helices: the bacteriorhodopsin example

A similar analysis using a large number of property scales was performed for bacteriorhodopsin alone. It turned out that only four scales are able to predict all seven helices of the bacteriorhodopsin: PORBET, NNEIG, EDEL21 and KYTDO. Other scales fail in the prediction of the D helix, which extends from residue 108 to 127 and contains a high number of glycines. Our best scale PORBET also misses helix D if our algorithm is trained on the T1 data set instead on the T2 data set of proteins.

The problem of recognizing the transmembrane helix D by visual inspection of the helix preference profile often does not exist even when our algorithm fails to recognize it. The case in point is such a profile obtained by the CHOUBB scale. All relevant residues from 108 to 127 have the helical preference (after the decision constant is added) higher or equal to 1.44 and are predicted in the helical conformation. In fact, there is only one minimum in the helix preference profile for that segment. Its value is 1.44 and it is located precisely at the functionally important residue D115. The problem arises when our algorithm tries to separate helix D from helix E, which are glued together as the prediction result of almost all property scales examined. For instance, in the case of CHOUBB scale, the extra-membrane segment 128–135 consisting of residues TKVYSYRF, which separates helix D from helix E, has a very high helix preference in the range from 1.92 to 2.38 and a very low turn preference from .25 to .52. It is almost certainly a short amphipathic helix attached to the extracellular membrane side in the native protein, as already seen by Henderson *et al.*[2] as a dense feature in electron diffraction patterns linking helix D with helix E. Our algorithm cannot recognize it as such because the present version of the algorithm cuts all helices predicted to be longer than 36 amino acids into two pieces by introducing several turn residue conformations in the sequence position of maximal turn propensity.

Predicting functionally important amino acids residues

The definition of performance parameters for measuring prediction accuracy for predicting functionally important amino acid residues requires specification of the range of preference values and a choice of residue types to be predicted. As an example, we can test the idea that charged residues of arginine, lysine and aspartate with low helical propensity, but found in the observed transmembrane helix, are functionally important. For instance, let us try as requirement $pxh < 0.6$ for lysine and $pxh < 1.1$ for arginine and aspartate, where pxh is the non-smoothed helix preference. The choice of upper bound for the pxh values for positive prediction is connected with the choice of decision constants. With DCH = 0 instead of 0.35, the upper limit for pxh values should be lowered to 0.25 and 0.75 for lysine and arginine or aspartate, respectively. Positive correct predictions are for those lysine, arginine and aspartate residues in the predicted transmembrane helices which satisfy those requirements and are experimentally observed to be important in membrane protein activity. Overpredictions and underpredictions are defined in the same manner.

TABLE IV

Predicting transmembrane segments in bacteriorhodopsin.

A) Ranking of property scales with respect to the prediction quality of bacteriorhodopsin transmembrane segments.

SCALE	$Q_3(\%)$	A_{m2}	# helices preodicted
PORBET	87.1	0.817	7
NNEIG	85.1	0.780	7
EDEL21	84.3	0.774	7
KYTDO	83.1	0.756	7
CHOUBB	85.9	0.726	6
FASBET	84.3	0.707	6
SWEET	83.9	0.683	6
WERSC	83.5	0.677	6
PRIFT	83.1	0.652	6
BULDG	82.3	0.646	6
CIDAB	82.3	0.640	6

B) Prediction quality with property scale PORBET:

HELIX #	PREDICTED	OBSERVED	u_m	o_m	p_m	N
1	9–31	10–32	1	1	22	23
2	41–70	38–62	3	8	22	25
3	78–102	80–101	0	3	22	22
4	107–131	108–127	0	5	20	20
5	135–159	136–157	0	3	22	22
6	168–193	167–193	1	0	26	27
7	202–231	203–227	0	5	25	25

C) Scale PORBET (One letter amino acid code is used)

A	C	L	M	E	Q	H	K	V	I
1.094	1.230	1.228	1.097	0.905	1.046	0.753	0.999	1.177	1.092
F	Y	W	T	G	S	D	N	P	R
1.155	1.372	1.363	0.967	1.013	0.956	0.561	0.668	0.419	0.933

Table V is an analysis of the power of different amino acid property scales to perform such prediction in the case of bacteriorhodopsin. A total of sixteen charged amino acids are considered as candidates for functionally important amino acids (first column). There are other charged amino acid in bacteriorhodopsin but these are never predicted on transmembrane location by any of the 11 property scales considered. Three different performance parameters are listed in the last three rows of Table V. Only aspartate 212 is always correctly predicted. Aspartate 85 is predicted by all property scales with the exception of KYTDO. All 6 amino acid residues involved in the proton pumping activity are predicted only by our PORBET scale. According to A and B parameters, the best scale is FASBET, while the second best are SWEET, EDEL21 and NNEIG. The FASBET scale predicts aspartate 115 in the turn conformation and misses entirely the D helix.

TABLE V

Prediction of amino acid residues involved in proton pumping. Code numbers for property scales in the first row are listed in Table I and explained in the Methods section. Positive correct prediction (p) is scored only in the case when residues R82, D85, D96, D115, D212 and K216 are inside the predicted membrane spanning segment with their preference for the α-helix conformation less than 1.1 (D or R) or less than 0.6 (K). Underprediction (u) is scored when these residues do not satisfy the above mentioned requirements. Overprediction (o) is scored when any other D, R or K residues (among N = 16) satisfy these requirements. Reported performance parameters are $B=(p-u-o)/N$, $A=p/(p+u+o)$ and prediction accuracy $Q = (p/6)$.

PROPERTY SCALE #	86	59	109	24	80	32	58	57	15	105	1
RESIDUE											
K41	o				o						
R82	p	p	u	p	p	p	p	u	p	u	u
D85	p	p	p	p	p	p	p	p	p	p	u
D96	p	p	p	p	p	p	u	p	u	u	u
D115	p	p	p	u	u	u	u	u	u	u	u
D102	o	o	o		o		o	o	o	o	o
D104					o			o			
K129											
R134	o				o				o		
K159	o										
K172	o									o	
R175	o	o	o	o	o		o	o	o	o	o
D212	p	p	p	p	p	p	p	p	p	p	p
K216	p	u	p	p	p	p	u	p	u	p	u
R225	o	o		o	o	o					
R227	o	o	o	o	o	o	o	o	o	o	o
B	–0.13	0	.06	.06	–.19	.13	–.19	–.13	–.25	–.25	–.44
A	0.43	0.50	0.56	0.56	0.38	0.63	0.33	0.40	0.30	0.30	0.11
Q	1.00	0.83	0.83	0.83	0.83	0.83	0.50	0.67	0.50	0.50	0.17

Can the same approach be used for voltage-gated ion channels which play a central role in the propagation and transduction of cellular signals? We utilized the 220–500 segment of the potassium channel membrane subunit from *Drosophila Shaker B*, which contains six membrane-spanning segments very well determined and extensively explored using the site-directed mutagenesis.[43] A substantially reduced channel activity or changed voltage-dependent activation is achieved with point-mutations at arginine 297, asparatate 316, arginine 362, 365, 368 and 371, lysine 374, arginine 377 and lysine 380.[44] Since these resideus are all located in the observed transmembrane segments and are of the same type already used for our structure-function studies with bacteriorhodopsin, it would be interesting to repeat exactly the same analysis. All six transmembrane segments were correctly predicted with our PORBET scale. The reported performance parameters were Q_3 = 80% and A_{m2} = 0.72. The P region between S5 and S6, which appears to line a narrow portion

of the central pore, was predicted as the segment of 12 residues of very high helical propensity. Sequence analysis with eleven hydrophobicity and preference scales (listed in Table I) selected 15 arginine, lysine and aspartate residues frequently predicted in transmembrane helices. Positive correct prediction for ion transport function was obtained by using the PORBET scale for R297, D316, R362, R365, R368 and R371. Underprediction for arginine 377 and lysine 380 occurred because the S4 transmembrane helix predicted by our algorithm ended too early, while underprediction for lysine 374 was only marginal (pxh = 0.6). Arginine residues 227 and 394 were overpredicted to be functionally important and to be located in transmembrane segments. This result corresponds to the performance parameters: A = 0.55, B = 0.07, Q = 67%. (see Table V legend). Other 10 scales (from Table I) were not able to predict arginine residues 362, 365 and 368 because of the surprisingly high α-helix propensity of that charged transmembrane segment, and their performance was not satisfactory in spite of the high quality of prediction for different folding motifs. For example, Q = 85% and A_{m2} = 0.81 for the EDEL21 scale but A = 0.17, B = –0.47 and Q = (2/9)100 = 22%.

DISCUSSION

In this paper, we asked the question about the optimal properties of natural amino acids that can be used to predict transmembrane helices and to find crucial residues involved in the ion transport activity. Proton donors and acceptors are known to be some charged amino acid residues buried in the transmembrane segments.[3] We first wanted to be sure that we have a reliable automated procedure for predicting such segments in the α-class integral membrane proteins and for predicting the overall secondary structure of such proteins. The straightforward sequence analysis, using the preference functions method,[6] selected 11 best property scales (Table I).

As expected, several different (with less than r = 0.90 correlation factor between them) hydrophobicity scales are excellent predictors for the sequence location of transmembrane helices. The preference function method uses hydrophobicity scales in a novel manner and enhances their predictive power, as seen with the example of the Kyte-Doolittle scale. A surprising result is that several β-preference scales are so good for predicting transmembrane helices and the overall secondary structure of membrane proteins lacking any β-sheet structure. This is the case with the PORBET scale (Table IV) extracted by us from porins and defensins, which is the best property scale (out of the 140 scales considered) for predicting the overall secondary structure, transmembrane helices and even functionally important charged amino acids in such helices. The β-sheet preference scales, derived from the database of soluble proteins (scales CHOUBB and FASBET) are also very good. One possible answer is that some folding motifs among β-sheet super secondary structures[45] in soluble proteins and all outside surfaces of β-barrel motifs in porins are very hydrophobic. That β-sheet preference scale is useful in predicting membrane topology of membrane proteins was observed before.[7] This should not surprise research workers such as Deber,[46] who discovered, almost a decade ago, that glycine and β-branched residues (valine, isoleucine, threonine), known to disfavour α-helices in soluble proteins, often account for nearly 50% of amino acids in membrane-spanning α-helical segments.

An implicit requirement in predicting transmembrane helices is to define what should be recognized by algorithm as observed and as predicted transmembrane helix. Our definition of predicted transmembrane helices (Methods) tends to glue together two or more transmembrane helices observed close by in the sequence into giant »transmembrane« helix and to miss the transmembrane helix broken with only one or two residues predicted in turn conformation. However, four best property scales: PORBET and three hydrophobicity scales (Table IV), predict all transmembrane helices in bacteriorhodopisn with high accuracy so that refinement of our predictor for such test-protein was not necessary.

Scale SWEET is the only property scale that combines physical information (hydrophobicity) with the known-how of Darwinian evolution through eons. Its range of application should be wider than what was proposed by the authors: finding similarity between two sequences through calculating the correlation of numerical hydrophobicities.[33] In that scale, hydrophobicity of each amino acid is calculated as the average of the amino acids observed to substitute for it in point mutations. Weighing is also introduced according to the frequency of substitutions.[47] To achieve self-consistency, the hydrophobicity scale calculated in this manner is used as input for another cycle of computations. This iterative procedure is repeated until the final scale does not change any more. The authors claim that their optimal matching scale is independent of the choice of the initial hydrophobicity scale.

Scale NNEIG is another solution (in addition to scale SWEET) for how to construct a self-consistent hydrophobicity scale starting with an iterative procedure with any other hydrophobicity scale as input.[36] Instead of the Dayhoff *et al.*[47] matrix of numbers of accepted point mutations, the nearest neigbour matrix is used to perform iterations. All amino acids found within the 8 Å distance of a given residue in three-dimensional protein structure are considered to be neighbours to that residue. Starting with input hydrophobicities , the average surrounding hydrophobicity of each residue type can be found.[42] This procedure is repeated until the computed scale converge.

Scale WERSC is also the protein answer about the appropriate hydrophobicity scale. It is derived by residue classification as being inside or outside and computing the frequencies of inside and outside resideus for all residue types in 20 soluble globular proteins.[34]

Scale PRIFT was derived from the observed helices in soluble globular proteins as a scale that maximizes the amphipathic index defined by the authors.[36] Since any hydrophobicity scale can be used to detect periodicity in hydrophobicity through the Fourier transform power spectrum and amphipathic index, scale PRIFT can be also considered as a hydrophobicity scale. Since most transmembrane helices in the bacteriorhodopsin and photosynthetic reaction centre exhibit characteristic helical periodicity in hydrophobicity,[48] it is expected that a scale designed to recognize amphipathic helices (PRIFT)) would be a good predictor for such a feature.

Scale BULDG arose from experimental measurements of the changes in water surface tension caused by increased concentration of a chosen amino acid.[21] The authors suggested that the slopes of the surface tension dependence on amino acid concentration constitute a hydrophobicity scale.

The spread of accuracy values for different proteins in both parts of Table III is such (from 87% to 63%) that it becomes of obvious importance to use exactly the

same testing protein list of membrane proteins when comparing prediction results from different laboratories. As mentioned before, the rank order of scales depends very much on the choice of membrane proteins of known structure for testing different property scales. Since the number of such proteins is very restricted at present, this situation gives undue weight to one protein or even one helix which is difficult to predict in its observed conformation. One example is helix D from bacteriorhodopsin, which is not predicted as transmembrane helix whenever aspartate 115 is predicted in the turn conformation. Another example are two helical segments found in subunit H, transmembrane being only the one close to the N-terminal. Omitting subunits H can »improve« predictions measured by Q_3 and Q_α. On the other hand, if transmembrane helix prediction parameters are used, omitting subunits H will »worsen« the prediction. When transmembrane helices are only a small part of protein sequence, then accuracy parameters for predicting such helices need not be correlated with the quality factors for predicting regular secondary structure.

When performance parameters A_α and B_α were used to test the prediction accuracy of the PHD method[39] the results were $A_\alpha = 0.57$, and $B_\beta = 0.05$ for helices in soluble proteins.[19] As it can be seen from Table II, our results for membrane proteins are better when expressed as A and B parameters. The undefined configuration and beta configuration (which is very rare in the tested membrane proteins) are both predicted less well than by the PHD method in soluble proteins.[19]

Helix prediction accuracy Q_α is just the ratio of correctly predicted helical residues p_α to all observed helical residues N_α. This ratio can be increased at will by overprediction of helical residues, which is not the case with A and B parameters that »punish« both overprediction and underprediction. Matthew's correlation coefficient[20] has also its deficiencies as discussed recently,[19] but is included here because the majority of other authors in this field use it.

Our Q_3 results for subunits L and M and H (accoroding to Table III 69% for prcl(v), 77% for prcm(v) and 63% for prch(v) are superior to Fariselli *et al.* neueral network results[49] for the same subunits of the photosynthetic reaction centre.

Performance parameter A_{m2} is commonly used to express the prediction accuracy for predicting transmembrane helices.[11,40] It is called selectivity by Edelman[11] or percentage of accuracy by Ponnuswamy.[40] As pointed out by Edelman,[11] it can have negative values so that A_{m2} (and A_{m1}) are not percentages. Both parameters are similar to parameter B_α (Eq. 1), which we called frequency balance[19] because it subtracts all wrong prediction frequencies from all right prediction frequencies. Parameter A_{m3} is completely analogous to A_α (Eq. (1)) in the three state model.

Our prediction accuracy parameters for transmembrane helices do not measure the prediction quality for each individual helix in membrane proteins. Since the number of possible wrong predictions is for some proteins greater than the total number of all possible right predictions (N_m), neither A_{m2} nor A_{m1} have a lower bound –1 in spite of having the upper bound fixed at +1. In other words, the weakness of these parameters is the possibility of values such as –3 or even –8 for very poor predictions.

It would be a mistake to use only one such parameter, for instance Q_2, to express the prediction accuracy for predicting transmembrane helices. Some hydrophobicity scales cannot predict a single subunit's H transmembrane helix at all, but the Q_2

factor for subunit H is still around 90% because 90% of residues not in the transmembrane helix conformation are indeed predicted not to be in such conformation. All scales listed in Table II are good predictors of the single transmembrane helix of subunit H and poor predictors of the subunit's secondary structure.

Recent results for predicting transmembrane helices in the photosynthetic reaction centre of *R. viridis*[11,40] with A_{m2} around 0.80 are similar to our results. Our result is $A_{m2} = 0.78$ for eleven transmembrane helices of the photosynthetic reaction centre with the PORBET property scale. It is better than the results obtained with the Kyte-Doolittle,[41,50] Sieved Kyte-Doolittle,[50,51] Klein-Kanehisa-DeLisi,[52] von Heijne,[53] Engelman-Steitz-Goldman,[54] Esposti-Crimi-Venturoli[55] and Ponnuswamy-Gromiha methods without end correction.[40]

The last goal of this paper, to predict functionally important amino acids for active or passive ion transport is connected to several of the first goals. Charged residues must be located as being inside or outside transmembrane helices, just as transmembrane helices must be recognized among other secondary structures. Assuming that some charged residues inside transmembrane helices are crucial for the ion transport mechanism, one would like to have an algorithm capable of identifying such residues.

In the case of bacteriorhodopsin, β-preference scales such as FASBET, CHOUBB, PORBET and KIDBET[35] (see Methods) are able to identify most proton acceptors and donors in the protein sequence (Table V). Bacteriorhodopsin has a total of 23 lysine, arginine and aspartate residues but only 14 are located in transmembrane helices and only 16 listed in Table V are predicted to be located in transmembrane helices according to any property scale. Prediction accuracy of 97% (Q_α in the Table III) with our PORBET scale (Table IV) ensures that all 14 interesting residues will be located properly in transmembrane helices, with the possible exception of one or, at most, two such residues. The next obstacle to predicting 6 out of those 14 residues which are functionally important (see Introduction and Figure 1) is more difficult to overcome as testifeid by Table V. Still, some residues, such as D85 and D212, are always predicted as functionally important. These amino acids are first and second proton acceptors after a very fast light-activated proton release step from the lysine Schiff's base connecting the retinal to opsin molecule (Figure 1). In fact, all of the 6 crucial residues involved in the proton pumping-activity of bacteriorhodopsin are correctly identified by our PORBET scale. Therefore, amino acids that form the obligatory part of the mechanism for proton pumping to extracellular space could have been predicted with a fairly high degree of accuracy before the experimental evidence obtained with site directed point mutations.[5]

Overpredictions are often associated with charged aspartate or arginine residues found at helix C-cap locations (D102, R225, R227). Underpredictions are sometimes associated with the internal helical residue having so low helical propensity (a necessary condition in our approach for being predicted as functionally important) that its conformation is predicted as turn or coil. This is often the case of aspartate 115. Another reason for frequent underpredictions of aspartate 115 is the underprediction of transmembrane helix D in which it is located. Both error conditions can be corrected in the future by using a more accurate procedure for predicting transmembrane helices.

The question remains whether our algorithm is so specific that it can work only for bacteriorhodopsin. One reason for optimism is the observation that cooperation

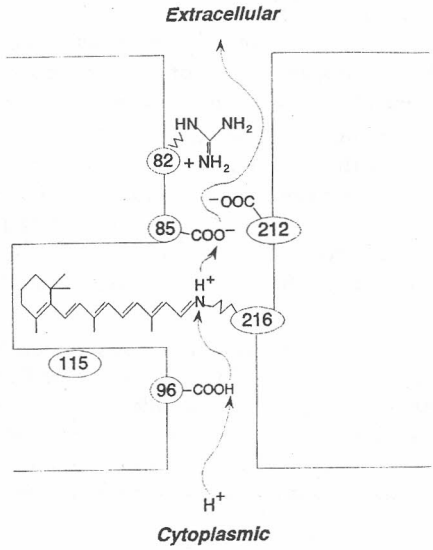

Figure 1. Proton pathway through bacteriorhodopsin. Only essential amino acid residues (bold and circled numbers) for proton pumping activity are shown. Photon absorption causes release of the proton from Schiff's base connection between the retinal and lysine 216. Proton release to extracellular space is facilitated by proton acceptors aspartate 85 and aspartate 212.

of transmembrane helices in promoting charge transfer through membrane seems to be a widespread mechanism.[56] For instance, some arginine, lysine and aspartate residues, located in the predicted transmembrane segments of the potassium channel,[43] are crucial for the voltage-gated ion transport function of that membrane protein.[44] Our PORBET scale correctly predicts 6 of the crucial 9 of such potassium channel residues and all 6 of such residues in the bacteriorhodopsin.

Acknowledgements. – Some references cited below were brought to our attention by B. Lučić and D. Zučić. Many discussions with B. Lučić are gratefully acknowledged. Computer drawing was done by V. Juretić. This work was supported by the Croatian Ministry of Science Grant 1-03-171.

REFERENCES

1. W. A. Cramer and D. B. Knaff, *Energy Transduction in biological membranes: A textbook of bioenergetics*. Springer-Verlag, New York, 1991.
2. R. Henderson, J. M. Baldwin, T. A. Ceska, F. Zemlin, E. Beckmann and K. H. Downing, *J. Mol. Biol.* **213** (1990) 899.
3. H. G. Khorana, *Proc. Natl. Acad. Sci. U.S.A.* **90** (1993) 1166.
4. T. Marinetti, S. Subramaniam, T. Mogi, T. Marti, and H. G. Khorana, *Proc. Natl. Acad. Sci. U.S.A.* **86** (1989) 529.
5. L. J. Stern and H. G. Khorana, *J. Biol. Chem.* **264** (1989) 14202.
6. D. Juretić, B. K. Lee, N. Trinajstić, and R. W. Williams, *Biopolymers* **33** (1993) 255.
7. D. Juretić, *Croat. Chem. Acta* **65** (1992) 921.

8. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. J. Meyer, J. R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, *J. Mol. Biol.* **112** (1977) 535.
9. D. Juretić, B. Lučić, and N. Trinajstić, *Croat. Chem. Acta* **66** (1993) 201.
10. D. Juretić, *Acta Pharmaceutica* **43** (1993) 223.
11. J. Edelman, *J. Mol. Biol.* **232** (1993) 165.
12. W. Kühlbrandt, D. N. Wang and Y. Fujiyoshi, *Nature* **367** (1994) 614.
13. R. Huber, R. Berendes, A. Burger, M. Schneider, A. Karshikov, H. Luecke, J. Romisch, and E. Paques, *J. Mol. Biol.* **223** (1992) 683.
14. D. Picot, P. J. Loll, and M. Garavito, *Nature* **367** (1994) 243.
15. P. A. Hargrave, *FASEB J.* **6** (1992) 2323.
16. A. -M. Colson, B. Edderkaoui, and. J. -Y. Coppee, *Biochim. Biophys. Acta* **1101** (1992) 157.
17. R.B. Gennis, *Biochim. Biophys. Acta* **1101** (1992) 184.
18. M.W. Parker, J. P. M. Postma, F. Pattus, A. D. Tucker, and D. Tsernoglou, *J. Mol. Biol.* **224** (1992) 639.
19. D. Juretić, B. Lučić and N. Trinajstić: "Secondary structure prediction quality for naturally occurring amino acids in soluble proteins". To be published in *Theochem*, 1994.
20. B. W. Matthews, *Biochim. Biophys. Acta* **405** (1975) 442.
21. H. B. Bull and K. Breese, *Arch. Biochem. Biophys.* **161** (1973) 665.
22. P. Y. Chou and G.D. Fasman, *Adv. Enzymol. Annu.* **47** (1978) 45.
23. S. W. Cowan, T. Schirmer, G. Rummel, M. Steiert, R. Ghosh, R. A. Pauptit, J. N. Jansonius, and. J. P. Rosenbusch, *Nature* **358** (1992) 727.
24. E. Schiltz, A. Kreusch, U. Nestel, and G. E. Schulz, *Eur. J. Biochem.* **199** (1991) 587.
25. M. S. Weiss, A. Kreusch, E. Schiltz, U. Nestel, W. Welte, J. Weckesser, and G. E. Schulz, *FEBS Letters* **280** (1991) 379.
26. H. Vogel and F. Jahnig, *J. Mol. Biol.* **190** (1986) 191.
27. S. Gerbl-Rieger, H. Engelhardt, J. Peters, M. Kehl, F. Lottspeich, and W. Baumeister, *J. Structural Biology* **108** (1992) 14.
28. N. Overbeeke, H. Bergmans, F. van Mansfeld, and B. Lugtenberg, *J. Mol. Biol.* **163** (1983) 513.
29. C. A. Mannella, M. Forte, and M. Colombini, *J. Bioenergetics and Biomembranes* **24** (1992) 7.
30. V. De Pinto and F. Palmieri, *J. Bioenergetics and Biomembranes* **24** (1992) 21.
31. P. Bulet, S. Cociancich, M. Reuland, F. Sauber, R. Biscfoff, G. Hegy, A. Van Dorsselaer, C. Hetru, and J. A. Hoffmann, *Eur. J. Biochem.* **209** (1992) 977.
32. C. P. Hill, J. Yee, M. E. Selsted, and D. Eisenberg, *Science* **251** (1991) 1481.
33. R. M. Sweet and D. Eisenberg, *J. Mol. Biol.* **171** (1983) 479.
34. D. H. Wertz and H. A. Scheraga, *Macromolecules* **11** (1978) 9.
35. A. Kidera, Y. Konishi, M. Oka, T. Ooi, and H. A. Scheraga, *J. Prot. Chem.* **4** (1985) 23.
36. J. L. Cornette, K. B. Cease, H. Margalit, J. L. Spouge, J. A. Berzofsky, and C. DeLisi, *J. Mol. Biol.* **195** (1987) 685.
37. H. Cid, M. Bunster, M. Canales and F. Gazitua, *Protein Engineering* **5** (1992) 373.
38. G. D. Fasman (Ed.) *Prediction of protein structure and the principles of protein conformation.* Plenum Press, New York, 1989, pp. 569.
39. B. Rost and C. Sander, *J. Mol. Biol.* **232** (1993) 584.
40. P. K. Ponnuswamy and M. M. Gromiha, *Int. J. Peptide Protein Res.* **42** (1993) 326.
41. J. Kyte and R. F. Doolittle, *J. Mol. Biol.* **157** (1982) 105.
42. P. Manavalan and P. K. Ponnuswamy, *Nature* **275** (1978) 673.
43. F. J. Sigworth, *Quarterly Reviews of Biophysics* **27** (1993) 1.
44. D. M. Papazian, L. C. Timpe, Y. N. Jan and L. Y. Jan, *Nature* **349** (1991) 305.
45. W. R. Taylor and J. M. Thorton, *J. Mol. Biol.* **173** (1984) 487.
46. C. M. Deber, C. J. Brandl, R. B. Deber, L. C. Hsu, and X. K. Young, *Arch. Biochem. Biophys.* **251** (1986) 68.
47. M. O. Dayhoff, R. M. Schwartz and B.C. Orcutt, in: *"Atlas of protein sequence and structure 1978"*, National Biomedical Research Foundation, Silver Spring, MD, Vol. 5, Supp. 3, pp. 345–352.
48. S. Sun and R. Parthasarathy, *Biophysical J.* **66** (1994) 2092.
49. P. Fariselli, M. Compiani, and R. Casadio, *Eur. Biophys. J.* **22** (1993) 41.
50. G. D. Fasman and W. A. Gilbert, *Trends Biochem. Sci.* **15** (1990) 89.

51. J. A. Bangham, *Anal. Biochem.* **174** (1988) 142.
52. P. Klein, M. Kanehisa, C. DeLisi, *Biochim. Biophys. Acta* **815** (1985) 468.
53. G. von Heijne, *EMBO J.* **5** (1986) 3021.
54. D. M. Engelman, T. A. Steitz and A. Goldman, *Annu. Rev. Biophys. Biophys. Chem.* **15** (1986) 321.
55. M. D. Esposti, M. Crimi, and G. Ventruoli, *Eur. J. Biochem.* **190** (1990) 207.
56. M. S. P. Sansom, *Prog. Biophys. & Molec. Biol.* **55** (1991) 139.

SAŽETAK

Skala β-sklonosti za pretkazivanje odnosa struktura-aktivnost u membranskih proteina

Davor Juretić i Roko Pešić

Predviđa se sekundarna struktura membranskih proteina, položaj transmembranskih uzvojnica i aminokiselinske ostatke u takvim uzvojnicama koji su važni za funkciju proteina, polazeći od fizikalnih, kemijskih ili statističkih svojstava dvadeset prirodnih aminokiselina u proteinima. U metodi sklonosnih funkcija znanje o strukturi proteina kombinira se s odgovarajućim skalama svojstava aminokiselina da bi se predvidjela sekundarna struktura membranskih proteina. Od ukupno 140 skala svojstava aminokiselina naša skala sklonosti za β-nabranu plohu, koju smo dobili koristeći porine i defensine, pokazala se najboljom u predviđanju položaja transmembranskih uzvojnica. Također se pokazuje da je moguće identificirati sa zadovoljavajućom točnošću one aminokiseline u transmembranskim uzvojnicama koje su važne za funkciju transportnih proteina poput bakteriorodopsina i kanala za kalij, u kojima je položaj i identitet takvih aminokiselina dobro poznat.