BIOINFORMATICS

Vol. 17 no. 7 2001
Pages 646–653

Evaluation of methods for the prediction of membrane spanning regions

Steffen Möller [1], Michael D. R. Croning [1, 2] and Rolf Apweiler [1]

[1]EMBL-Outstation European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK and [2]School of Biological Sciences, The University of Manchester, Oxford Road, Manchester M13 9PT, UK

Received on December 15, 2000; revised on March 13, 2001; accepted on March 16, 2001

ABSTRACT

Motivation: A variety of tools are available to predict the topology of transmembrane proteins. To date no independent evaluation of the performance of these tools has been published. A better understanding of the strengths and weaknesses of the different tools would guide both the biologist and the bioinformatician to make better predictions of membrane protein topology.

Results: Here we present an evaluation of the performance of the currently best known and most widely used methods for the prediction of transmembrane regions in proteins. Our results show that TMHMM is currently the best performing transmembrane prediction program.

Contact: moeller@ebi.ac.uk; croning@ebi.ac.uk; apweiler@ebi.ac.uk

INTRODUCTION

Genome sequencing projects provide the scientific community with an ever-increasing rate of predicted protein sequences. To analyze these biochemically uncharacterized sequences, computer based methods have been established to provide researchers with an initial characterization. Many of these methods make use of sequence similarity to already described proteins. Other methods are used to predict certain properties like membrane spanning regions.

In this paper we have analyzed the performance of the different programs for the prediction of transmembrane regions in proteins. Such predictions are possible because of distinctive patterns of hydrophobic (intramembraneous) and polar (loops) regions within the sequence. The percentage of transmembrane proteins does not differ too much in various organisms (Wallin and Heijne, 1998; Stevens and Arkin, 2000) and about a fourth of all proteins in SWISS-PROT and TrEMBL (Bairoch and Apweiler, 2000) are predicted to be transmembraneous. Membrane proteins play important roles in the cell as key components of cell–cell signalling mechanisms, initiating signalling cascades. They also mediate the transmembrane transport of many ions and solutes, as well as being involved in the organism's recognition of self. The pharmaceutical industry has found them of particular interest, since membrane-bound receptors and channels have been repeatedly proven to be fruitful therapeutic targets. Additionally, membrane proteins often mediate acquired resistance to drugs.

Thorough structural analysis of membrane proteins is difficult to achieve since it is very hard to determine the structure due to the intrinsic difficulties involved in growing crystals of membrane proteins. It takes considerably less effort to biochemically determine just the membrane topology (Geest and Lolkema, 2000), which includes the determination of the localization of membrane spanning regions (MSRs) and the polarity of their integration into the membrane (sidedness).

Still, the topology of the vast majority of membrane proteins remains biochemically undetermined. Our group provides a collection of proteins with known biochemical characterizations of membrane topology (Möller *et al.*, 2000). However, this collection contains only ∼200 well-characterized sequences. Consequently, the characterization of the remaining membrane proteins requires an accurate method for the automated prediction of MSRs.

Reliable computational methods for topology predictions are very valuable as they provide the basis for further experimental analysis. A variety of tools have been implemented, with the first being about 20 years old. For an evaluation of predictions it is important not only to look at individual MSRs but at the whole protein. To make a prediction for proteins with seven MSRs 95% reliable, individual segments would need to be 99, 96% reliable, and additionally, the method must never over-predict. Current tools are far away from achieving this. The present study provides an evaluation of their actual performance.

EVALUATION

The following methods for prediction of MSRs have been evaluated: TMHMM 1.0, 2.0, and a retrained version

of 2.0 (Sonnhammer *et al.*, 1998; Krogh *et al.*, 2001), MEMSAT 1.5 (Jones *et al.*, 1994), Eisenberg (Eisenberg *et al.*, 1982), Kyte/Doolittle (Kyte and Doolittle, 1982), TMAP (Persson and Argos, 1997), DAS (Cserzo *et al.*, 1997), HMMTOP (Tusnády and Simon, 1998), SOSUI (Hirokawa *et al.*, 1998), PHD (Rost *et al.*, 1996), TMpred (Hofmann and Stoffel, 1993), KKD (Klein *et al.*, 1985), ALOM2 (Nakai and Kanehisa, 1992), and Toppred 2 (Claros and Heijne, 1994). Alom2, Eisenberg and Kyte/Doolittle, TMpred and DAS are methods investigating local properties of amino acid sequences to decide which sub-sequences are most likely to span the membrane, usually in a sliding window approach. PHD uses a neuronal network and should still be regarded as a local approach. Global approaches are HMMTOP and TMHMM, which both implement circular Hidden Markov Models. These approaches determine the statistically most probable topology for the whole protein according to the underlying model. TMAP, Toppred2 and also MEMSAT represent combined forms, in which results on a local level are evaluated by global heuristics such as the positive-inside rule or other differences in the distribution of amino acids. No additional input data was given, i.e. we did not provide sequence alignments to TMAP, to ensure the reproducibility of this evaluation.

The previously mentioned collection of well-characterized membrane proteins was used as the reference annotation to evaluate the predictions of the various methods. This test set contains 188 proteins with 883 MSRs that have been determined from either their elucidated structures or by fusion experiments. As described in Möller *et al.* (2000) the interpretation of experiments does not allow one to set unambiguous borders for transmembrane regions. Therefore some deviation of the prediction from the reference annotation must be tolerated. In accord with a previous study (Sonnhammer *et al.*, 1998), for an MSR to be evaluated as correct, we decided it must share at least nine residues with the reference annotation's MSR. This threshold is a little less than half that the ∼20 residues expected for an MSR.

Each program was rated by three values. Firstly, it was rated by the percentage of predicted transmembrane regions that could be assigned to a reference MSR (true positive predictions). Secondly, it was rated by the percentage of reference MSRs that were not predicted (false negatives) and thirdly, by the percentage of predicted MSRs that are not existent as MSRs in the reference protein test set (false positives). Also, but not applicable to all methods, we investigated the reliability of a prediction of the sidedness of the protein's membrane integration.

The reference annotation describes both mitochondrial inner membrane proteins and plasma membrane proteins, both of which are generally believed to be helical. For this reason a minimum length of 15 residues would be expected in order to span the membrane. The default parameter sets were used for the evaluation of all methods. For the evaluation of TMpred, parameter values of a minimum length of 15 and a maximum length of 25 residues for MSRs were utilized. It is likely that to achieve optimal results these values should be varied, depending on the organism (varying thickness of the membrane) or organelle (hypothetical influence of the length of MSRs on protein sorting). This optimization of parameters was not performed in the present study, in order to keep the evaluation straightforward, and subsequently easily reproducible.

We implemented a prediction upon the basis of the Kyte/Doolittle hydropathy scale with window lengths from 5 (KD5) to 9 (KD9) residues to the peak in hydropathy. A threshold of 1.6 was required for the predicition of an MSR. These were assumed to be of length 20 amino acid residues.

Table 1. Performance on transmembrane regions of all biochemically characterized proteins

Method	TP	FN	FP	(FN + FP)
TMHMM 2.0 (Sonnhammer *et al.*, 1998)	812	65	38	103
TMHMM 1.0 (Sonnhammer *et al.*, 1998)	818	63	45	108
TMHMM-Retrain*	811	70	38	108
MEMSAT 1.5 (Jones *et al.*, 1994)	772	110	78	188
Eisenberg (Eisenberg *et al.*, 1982)	809	72	163	235
KKD (Klein *et al.*, 1985)	719	164	72	236
KD5	773	139	125	259
TMAP (Persson and Argos, 1997)	675	191	82	273
DAS (Cserzo *et al.*, 1997)	829	38	243	281
HMMTOP (Tusnády and Simon, 1998)	639	243	65	308
SOSUI (Hirokawa *et al.*, 1998)	686	192	137	329
KD9	494	391	25	416
TMpred (Hofmann and Stoffel, 1993)	525	357	80	437
ALOM 2 (Nakai and Kanehisa, 1992)	429	545	17	471
PHD (Rost *et al.*, 1996)	564	319	207	526
Toppred 2 (Claros and Heijne, 1994)	468	417	123	540
Total number of MSRs	883			

TP stands for the number of correctly predicted MSRs, FN for MSRs that where not predicted and FP for predictions that where not confirmed by the reference annotation. The methods are sorted by the sum of false negative and false positive predictions. False negatives and true positives should sum up to the same number (883) for all the methods. This is not the case when a predicted MSR spans two reference regions. Also two predicted MSRs overlapping a single reference MSR would not be noticed in this table.
*This version of TMHMM was developed for the current evaluation only and is not available to the public.

RESULTS

Performance on transmembrane regions of all biochemically characterized proteins

Table 1 shows the performance of the evaluated methods on individual MSRs. The methods are ranked by the number of errors detected (FN + FP). The method TMHMM

Table 2. Performance on proteins with characterized MSRs

Method	All MSRs found	Additionally correct sidedness
TMHMM-Retrain*	129 (69%)	102 (79% of 129)
TMHMM 2.0	128 (68%)	89 (70%)
TMHMM 1.0	126 (67%)	91 (72%)
MEMSAT 1.5	100 (53%)	77 (77%)
KKD	85 (45%)	n/a
HMMTOP	83 (44%)	68 (82%)
TMAP	80 (43%)	21 (26%)
Eisenberg	72 (38%)	n/a
DAS	70 (37%)	n/a
TMpred	70 (37%)	12 (17%)
SOSUI	68 (36%)	n/a
KD5	61 (32%)	n/a
KD9	49 (26%)	n/a
PHD	49 (26%)	34 (69%)
Toppred 2	48 (26%)	23 (48%)
ALOM 2	14 (7%)	n/a
Total number of proteins	188 (100%)	

This table presents an analysis of the program's performance in predicting all MSRs within a transmembrane protein. It displays in the second column the number of predictions that had all MSRs correctly assigned. This was defined as being the case when a sequence had no false positives, no false negatives and also the correct number of MSRs predicted. The third column shows how often the sidedness of the integration was predicted correctly.
*This version of TMHMM was developed for the current evaluation only and is not available to the public.

in all its three versions is by far the best in this comparison. MEMSAT is the second best method, although it produces twice as many errors as TMHMM. The only additional interesting result here is the low number of false positives assigned by ALOM. Its FP/TP ratio is even slightly lower than the one of TMHMM.

Performance on all MSRs within a protein

Table 2 shows the performance of the evaluated method on all MSRs within a protein and basically confirms the results of Table 1. The TMHMM versions predicted in approximately two thirds of the reference proteins all MSRs correctly. In about 70–80% of these correctly predicted proteins the sidedness was correctly predicted, too. The retrained TMHMM performs better in the determination of the sidedness. MEMSAT was able to predict all MSRs correctly in 53% of the cases. While HMMTOP is the best method to predict the sidedness of a transmembrane protein, Toppred, TMAP and TMpred decide the sidedness less reliably than by random choice.

Performance on transmembrane regions of proteins unknown to the method

Table 3 presents a variation of the analysis shown in Table 1, by being based on only those MSRs that were not presented to the respective program for its training or analysis.

Table 3. Performance on known MSRs not used in the training sets of the method

Method	TP + FN	TP	FN	FP	FN + FP	% correct
TMHMM-Retrain*	322	294	28	20	48	85.1
TMHMM 2.0	469	415	54	27	81	82.7
TMHMM 1.0	471	413	58	36	94	80
MEMSAT 1.5	722	620	102	69	171	76.3
Eisenberg	881	809	72	163	235	73.3
KKD	883	719	164	72	236	73.3
KD5	907	773	134	125	259	71.4
TMAP	696	538	158	68	226	67.5
DAS	626	598	28	210	238	62
SOSUI	829	638	191	137	328	60.4
KD9	885	494	391	25	416	53
TMpred	882	525	357	80	437	50.5
HMMTOP	453	251	202	33	235	48.1
ALOM 2	883	429	454	17	471	46.7
PHD	883	564	319	207	526	40.4
Toppred 2	885	468	417	123	540	39

TP stands for the number of correctly predicted MSRs, FN for MSRs that where not predicted and FP for predictions that where not confirmed by the reference annotation. The methods are sorted by the percentage of correct predictions. False negatives and true positives should sum up to the same number (883) for all the methods. This is not the case when a predicted MSR spans two reference regions. Also two predicted MSRs overlapping a single reference MSR would not be noticed in this table. Sums larger than 883 are explained by two predicted MSRs overlapping with a single reference MSR of the collection. In this case both predicted MSRs are counted as true positive. Please be aware that the number of MSR differs for different methods since the training/evaluation set of the methods differ. The set is smallest for the newer versions of TMHMM and HMMTOP.
*This version of TMHMM was developed for the current evaluation only and is not available to the public.

Again, the TMHMM versions performed best (80–85% correct predictions), slightly ahead of MEMSAT, which confirmed 76% of the MSRs correctly. The Eisenberg and Kyte/Doolittle methods are very close runner-ups with 73.3% each. The low number of false negatives of the Eisenberg method (8.1%) and especially of DAS (4.5%) should be mentioned. The false negative rates of the best performing TMHMM version and of MEMSAT are 8.6 and 14%, respectively.

Performance on all MSRs within the proteins that were not used for training

Table 4a presents, like Table 2, a view upon the prediction performance for whole proteins rather than on individual MSRs. The intersection of proteins that were not used for training or analysis by any of the programs contains only 87 proteins. A larger data set optimizes the reliability of this analysis for all individual methods. Hence Table 4a presents the analysis of Table 2 on the basis of different protein sets, the respective maximal sets of proteins unknown to the method.

Table 4a. Performance on proteins with characterized MSRs not known to the method

Method	No. of proteins	All MSRs found	Additionally correct sidedness
TMHMM 2.0	108	64 (59%)	40 (63%)
TMHMM 1.0	108	57 (53%)	21 (53%)
TMHMM-Retrain	69	35 (51%)	22 (62%)
MEMSAT 1.5	159	80 (50%)	58 (73%)
KKD	188	85 (45%)	n/a
TMAP	156	69 (44%)	18 (26%)
Eisenberg	188	72 (38%)	n/a
TMpred	188	70 (37%)	12 (17%)
KD5	188	61 (32%)	n/a
SOSUI	147	53 (36%)	n/a
HMMTOP	106	37 (35%)	29 (78%)
DAS	148	50 (33%)	n/a
PHD	151	49 (33%)	34 (70%)
Toppred 2	188	48 (26%)	23 (48%)
KD9	188	48 (26%)	n/a
ALOM 2	188	14 (7%)	n/a
Total number of proteins	188		

This table presents an analysis of the program's performance on the whole transmembrane protein. Methods are sorted by the percentage of correctly predicted proteins. The second column shows the number of proteins that could be used for the evaluation since they were not presented to the respective program for its training or analysis. The third column shows the number of proteins whose MSRs were all correctly predicted. This was defined as being the case when a sequence had no false positives, no false negatives and also the correct number of MSRs predicted. The fourth column shows how often the sidedness of the integration was predicted correctly.

The drawback of this approach is that the methods are not constrained to the identical weaknesses and difficulties present in the evaluation set. Table 4b shows therefore the same analysis on the set of 87 proteins that were not involved in the training of any of these methods.

Both Table 4a and b confirm the dominance of TMHMM. The three versions of this method predict all MSRs within proteins that were not used for training in 51–60% of the cases correctly. MEMSAT correctly predicted 47% of all MSRs within proteins that are not used for training of the program.

Influence of signal peptides and transit peptides

Transmembrane prediction programs have the tendency to interpret the hydrophobic parts of signal sequences and transit peptides as MSRs. The transmembrane test set contains 34 proteins with a cleavable signal and eight proteins with transit peptides. Table 5 shows that only ALOM 2 correctly predicted not a single signal sequence as transmembrane. ALOM 2 is followed by PHD with one error and Toppred 2 with three errors. The seven errors of TMHMM 2.0 account for 16% of the total TMHMM false positives from Table 1. Only the Kyte/Doolittle hydropathy analysis methods (KD5–KD9) predicted the 8 mitochondrial transit peptides as transmembranous.

Table 4b. Comparison of performance on an identical set of proteins unknown to methods

Method	All MSRs found	Additionally correct sidedness
TMHMM-Retrain	52 (60%)	43 (83% of 52)
TMHMM 2.0	48 (55%)	36 (75% of 48)
TMHMM 1.0	45 (52%)	33 (73% of 45)
MEMSAT 1.5	41 (47%)	33 (80% of 41)
KKD	39 (45%)	n/a
TMAP	35 (40%)	12 (34% of 35)
KD8	33 (37%)	n/a
Tmpred	29 (33%)	9 (31% of 29)
Eisenberg	27 (31%)	n/a
SOSUI	27 (31%)	n/a
KD5	26 (30%)	n/a
KD9	25 (29%)	n/a
DAS	24 (28%)	n/a
HMMTOP	23 (26%)	19 (83% of 23)
KD6	21 (24%)	n/a
PHD	18 (21%)	17 (94% of 18)
Toppred 2	16 (18%)	6 (38% of 16)
ALOM 2	9 (10%)	n/a

This table presents an analysis of the program's performance on the whole transmembrane protein. The set of 87 proteins not involved in the training of any of the prediction methods was used as the basis for this analysis. Methods are sorted by the percentage of correctly predicted proteins. The second column shows the number of proteins whose MSRs were all correctly predicted. This was defined as being the case when a sequence had no false positives, no false negatives and also the correct number of MSRs predicted. The third column shows how often the sidedness of the integration was predicted correctly.

Table 5. Discriminative performance on signal and transit peptides

Method	No. of signal sequences predicted as MSRs
ALOM 2	0
PHD	1
Toppred 2	3
TMHMM 1.0	7
TMHMM 2.0	7
TMHMM-Retrain	9
MEMSAT 1.5	12
SOSUI	14
TMAP	20
HMMTOP	25
Eisenberg	26
KKD	26
TMpred	31
DAS	33
KD5	34
KD9	34
Maximum	34 of 34

The second column displays the number of proteins in which a signal sequence was predicted to be an MSR.

Table 6. Membrane proteins whose MSRs were not correctly predicted by any program

Trust level	Number of problematic proteins	Number of test set proteins	Test set entries: SWISS-PROT ID [SWISS-PROT AC]*
A	1 (3%)	34	PGH1_SHEEP[P05979]
B	5 (22%)	23	ARSB_ECOLI[P37310], DTPT_LACLA[P36574], HLYB_ECOLI[P08716], PTNC_ECOLI[P08187], PTND_ECOLI[P08188]
C	17 (15%)	108	ADT2_YEAST[P18239], ALKB_PSEOL[P12691], B3AT_HUMAN[P02730], CYB_RHOSH[Q02761], CYDA_ECOLI[P11026], CYOE_ECOLI[P18404], FLO1_HUMAN[P41440], PMA1_NEUCR[P07038], RBSC_ECOLI[P04984], S61A_YEAST[P32915], SCAA_RAT[P37089], STE6_YEAST[P12866], [LEP00030], [LEP00130], [LEP00330], [LEP03300], [LEP03303]
C*	3 (13%)	23	GAA4_BOVIN[P20237], GRA1_HUMAN[P23415], GRA3_RAT[P24524]
Sum	26 (14%)	188	

Column one shows the category of trust as set in the collection of transmembrane proteins for individual entries. Trust level A stands for an available crystal structure, B for strong biochemical evidence and C for less reliable biochemical evidence. C* denotes entries with MSR annotation labelled in SWISS-PROT as highly reliable. The second column lists the entries of the test set with their entry name and the accession number in brackets.
*The constructed LEP0xxx proteins are not in SWISS-PROT/TrEMBL.

Table 7. Membrane proteins whose sidedness was not correctly predicted by any program but had their MSRs predicted correctly by at least one method

Trust level	Number of problematic proteins	Number of test set proteins	Test set entries: SWISS-PROT ID [SWISS-PROT AC]*
A	3 (9%)	34	ATPL_ECOLI[P00844], CB22_PEA[P07371], COX3_PARDE[P06030]
C	12 (11%)	108	CITN_KLEPN[P31602], CLC1_HUMAN[P35523], CYOA_ECOLI[P18400], CYOC_ECOLI[P18402], GAB1_HUMAN[P18505], IM23_YEAST[P32897], MDFA_ECOLI[Q46966], ROM1_BOVIN[P52205], [LEP00000], [LEP00003], [LEP00300], [LEP00303]
C*	16 (70%)	23	GAA1_CHICK[P19150], GAA2_HUMAN[P47869], GAA3_HUMAN[P34903], GAA5_HUMAN[P31644], GAA6_MOUSE[P16305], GAB2_HUMAN[P47870], GAB3_HUMAN[P28472], GAB4_CHICK[P24045], GAC1_RAT[P23574], GAC3_MOUSE[P27681], GAC4_CHICK[P34904], GAD_MOUSE[P22933], GAR1_HUMAN[P24046], GAR2_HUMAN[P28476], GRB_RAT[P20781], SSRG_RAT[Q08013]
Sum	31 (16%)	188	

Column one shows the category of trust as set in the collection of transmembrane proteins for individual entries. Trust level A stands for an available crystal structure, B for strong biochemical evidence and C for less reliable biochemical evidence. C* denotes entries with MSR annotation labelled in SWISS-PROT as highly reliable. The second column lists the entries of the test set with their entry name and the accession number in brackets.
*The constructed LEP0xxx proteins are not in SWISS-PROT/TrEMBL.

Summary of evaluation based on reference TM annotation

162 proteins (85%) of the reference test set's 188 proteins have their MSRs correctly predicted by at least one program. When the sidedness is included in this analysis, this reduces the number of correct predictions to 131 (70%). Table 2 shows TMHMM to be the best performing. Its versions were able to predict at least 89 (48% of all proteins) entries completely correctly, including their sidedness. In its retrained variant, it was even predicting 54% of the entries completely correct, though this is only due to the better performance of the retrained version on the determination of the sidedness. Table 6 shows the entries from the collection for which the MSRs could not be correctly assigned by any method.

All proteins in the test set, except the LEP0xxxx proteins, are SWISS-PROT entries. The LEP0xxxx proteins are artificial proteins, resulting from fusions of the *E.coli* leader peptidase with itself. Polar residues were introduced in the loops, which led to topologically 'frustrated' membrane regions (Gafvelin and Heijne, 1994). These are membrane spanning regions that maintained all their hydrophobicity but are not integrated after the modification since the integrated polar residues in the connecting loops are incompatible with the final topology of the insertion process. None of the current methods seems sensitive enough for these subtle changes.

Table 8. Performance on G-protein coupled receptors

Program	Number of proteins with specific number of predicted membrane spanning regions (percentage of all GPCRs) Number without correction of overlap with signal sequence									
	0 MSRs predicted	1	2	3	4	5	6	**7 correct**	8	>8
ALOM 2	0	6	20	57	176	291	248	**29**	6	0
	(0)	(1)	(2)	(7)	(21)	(35)	(30)	**(3)**	(1)	(0)
	0	6	16	57	170	271	269	**35**	9	0
DAS	2	0	5	42	212	369	173	**24**	3	1
	(0)	(0)	(1)	(5)	(25)	(44)	(21)	**(3)**	(0)	(0)
	2	0	5	42	194	357	156	**62**	9	4
HMMTOP	0	0	0	0	1	1	27	**712**	88	4
	(0)	(0)	(0)	(0)	(0)	(0)	(3)	**(85)**	(11)	(0)
	0	0	0	0	1	1	25	**644**	154	8
MEMSAT 1.5	0	23	21	22	14	40	100	**551**	56	6
	(0)	(3)	(3)	(3)	(2)	(5)	(12)	**(66)**	(7)	(1)
	0	23	21	0	16	33	106	**531**	73	10
TMHMM 1.0	0	0	1	0	2	12	98	**707**	13	0
	(0)	(0)	(0)	(0)	(0)	(1)	(12)	**(85)**	(2)	(0)
	0	0	1	0	2	12	96	**696**	26	0
TMHMM 2.0	0	0	0	1	3	12	98	**711**	8	0
	(0)	(0)	(0)	(0)	(0)	(1)	(12)	**(85)**	(1)	(0)
	0	0	0	1	3	12	96	**698**	23	0

The table is set up according to the correct number of predicted MSRs. All GPCR proteins should have seven MSRs, this 'correct' group is showin in bold. The first number shows the number of MSRs that do not overlap with a signal sequence as annotated in the SWISS-PROT database. The second number gives the percentage of the first value of all 833 GPCRs. The third number shows the original number of predicted MSRs before correcting for signal peptides.

The *E.coli* leader peptidase in its native form is among the proteins of the test set and gets correctly predicted. The LEP-LEP fusions though irritate the prediction methods, especially for the determination of their sidedness.

Other proteins involved in the integration of membrane proteins into the membrane, e.g. SecY and SecE, seem to be reliably predicted. Exceptions are the yeast Sec61A (Table 6) and the mitochondrial IM23 (Table 7).

It is not too surprising that proteins within larger membrane complexes are harder to predict, since their properties are less constrained by the membrane than by their interaction with other proteins within their complex. Also it is not clear if they are integrated into the membrane by the same mechanism. The problems with COX and CYO proteins can possibly be explained in this way.

The remaining problematic proteins of Tables 6 and 7 have in common that they have at least four transmembrane regions. Most of them are ion transporters, which have polar residues within their MSRs. This may have contributed to the difficulty of an automated prediction of MSRs and the sidedness.

Evaluation on seven-transmembrane proteins

In the following analysis we have used a subset of the available methods to predict the MSRs of a set of 833 G-protein coupled receptors (GPCRs). They are determined by the database reference of SWISS-PROT to the GPCRDB (Horn *et al.*, 1998). Table 8 shows the prediction results of ALOM 2, DAS, HMMTOP, MEMSAT, TMHMM 1.0 and TMHMM 2.0. MSRs predicted N-terminal of a potential signal peptide cleavage point (as annotated in SWISS-PROT) were ignored.

One should note that the numbers of MSRs possessed by these proteins have not been biochemically determined. However, GPCRs are generally accepted to have seven transmembrane regions with an extracellular N-terminus.

The Hidden Markov Model based methods TMHMM and HMMTOP performed well in this evaluation, reaching 85% correct MSR assignments. MEMSATs performance was less satisfying with 66%. Alom 2 and DAS failed completely with only 3% correct MSR assignments. An explanation may be that the membrane topology of GPCRs is rather hard to predict, possibly reflecting a high

proportion of polar residues within their transmembrane helices (Ji *et al.*, 1998).

Negative set of soluble proteins

We mentioned before that MSR prediction methods often predict hydrophobic parts of the N-terminal signal as transmembraneous. This error can easily be corrected by an additional run of a tool for signal prediction. What should not happen, though, is that hydrophobic regions within soluble proteins or globular loops of transmembrane proteins are predicted as transmembrane. To evaluate the ability of transmembrane prediction programs in their discrimination of transmembrane proteins from soluble proteins, all the programs were run on a set of 634 known cytoplasmic or periplasmic soluble proteins derived from the SWISS-PROT release 38. Accordingly, not a single MSR should have been assigned to any one of these proteins.

The performance of the majority of these tools in Table 9 seems disappointing. Except for TMHMM (8 = 1% false annotations) and SOSUI (19 = 3%) they all have the tendency to strongly over-predict. Even ALOM 2 (61 = 10%), which has the lowest FP rate on real MSRs, does not perform well in this evaluation against soluble proteins.

Table 9. Performance on set of soluble proteins

Method	No. of FP proteins	No. of FP MSR (−signals)	No. of entries/100s
TMHMM 1.0	8 (1.26%)	8 (−1)	37
TMHMM 2.0	8 (1.26%)	8 (−2)	37
SOSUI	19 (2.99%)	27 (−3)	10
KD9	49(7.73%)	53(−1)	3963
ALOM 2	61 (9.6%)	65 (−0)	2438
HMMTOP	70 (11.0%)	84 (−9)	72
Eisenberg	84 (13.0%)	290 (−2)	3993
PHD	120 (18.9%)	212 (−1)	18
KKD	136 (21.5%)	166 (−7)	5835
Tmap	203 (32.0%)	276 (−6)	352
TMpred	350 (55.2%)	434 (−3)	n/a
MEMSAT 1.5	431 (68.0%)	784 (−8)	84
Toppred 2	472 (76.0%)	1198 (−8)	40
KD5	489 (77.1%)	1034 (−7)	1650
DAS	524 (82.6%)	1257 (−9)	5

The first column presents the method's name, the second the number of proteins that are false positive and the third presents the number of false positive MSRs. The number of signal sequences predicted as transmembrane is stated as a negative number in parentheses behind the total number of false positive MSRs. The fourth column compares the CPU time.

DISCUSSION

We compared the performance of current methods for the prediction of MSRs and their sidedness. The tools were run on a set of well-characterized transmembrane proteins as a positive control, a set of GPCRs for which only the total number of MSRs within each protein could be compared, and a set of soluble proteins as a negative control. We found that the performance of some of these tools is good, while not perfect, in determining the location of transmembrane regions. Though it seems that the determination of the sidedness of transmembrane proteins is not well modelled by most of the tools.

Overall, TMHMM performs best, followed by MEMSAT. TMHMM is especially good at reliably distinguishing between soluble and transmembrane proteins. Also for proteins known to be transmembrane it performs best, followed by MEMSAT. ALOM 2 performed well in confirming transmembrane regions with a very low number of false positives.

It was surprising to see that the KKD analysis (Klein *et al.*, 1985), or the analysis of the hydrophobic moment (Eisenberg *et al.*, 1982), are relatively reliable predictors for the MSRs of membrane proteins. Their main weakness, which they share with other window-based methods, is their lack of specificity for membrane proteins.

No method was able to predict more than 52% of the proteins correctly. However, 86% of the proteins had all their MSRs correctly predicted by at least one method and for 70% a correct prediction that includes the sidedness could be achieved by at least one method.

From a technical standpoint, there is no difference between TMHMM 1.0 and TMHMM 2.0 except for the latter being retrained on the identical data set. The developers of TMHMM kindly provided us with an additional version, TMHMM-Retrain, that was retrained on a non-redundant subset of the reference annotation used for this evaluation. This version had some slight advantages over the versions 1.0 and 2.0, especially for the prediction of the sidedness, but otherwise this demonstrates that the choice of training data has only a limited impact on the performance. This study does not confirm any significant superiority of TMHMM 2.0 over its predecessor.

Our results from the prediction of MSRs within GPCRs reveal how varying the performance of the prediction methods can be. It also demonstrates that Hidden Markov Models have a superiority over sliding window approaches in such difficult cases.

Although TMHMM proves fairly robust against signal sequences, the topology prediction should not be performed without the consultation of signal sequence prediction methods like SignalP 2.0 (Nielsen and Krogh, 1998; Nielsen *et al.*, 1999). TMHMM is the first choice to decide if a protein is transmembraneous or not. It has the best overall performance but with a tendency to underpredict. When there is doubt in the correctness of the TMHMM prediction, additional evidence like determined

protein domains or post-translational modifications should be considered and additional tools should be consulted to arrive at a consensus. The strongly underpredicting tool ALOM 2 might serve to increase the degree of confidence in individual MSRs, while more sensitive tools can be used to increase the number of candidates for an MSR. Recently an integration of multiple prediction methods evaluated in this paper was published (Nilsson *et al.*, 2000). SPLIT (Juretic and Lucin, 1998; Juretic *et al.*, 1999) and TM Finder (Deber *et al*., 2001) integrate multiple scales for amino acids for the prediction of MSRs. This suggests an evaluation of strategies for the integration of multiple predictors should be carried out.

Finally, we suggest all the tools should be considered merely an aid to the biologist in making an educated guess as to the whereabouts of MSRs with a protein.

ACKNOWLEDGEMENTS

Many thanks go to Anders Krogh, Henrik Nielsen, Miklos Cserzo, Kenta Nakai and G.E.Tusnády for their comments and co-operation.

REFERENCES

Bairoch,A. and Apweiler,R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res*, **28**, 45–48.

Claros,M.G. and Heijne,G.v. (1994) TopPred II: an improved software for membrane protein structure predictions. *Comput. Appl. Biosci.*, **10**, 685–686.

Cserzo,M., Wallin,E. *et al.* (1997) Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng.*, **10**, 673–676.

Deber,C.M., Wang,C. *et al.* (2001) TM Finder: a prediction program for transmembrane protein segments using a combination of hydrophobicity and nonpolar phase helicity scales. *Protein Sci.*, **10**, 212–219.

Eisenberg,D., Weiss,R.M. *et al.* (1982) The helical hydrophobic moment: a measure of the amphiphilicity of a helix. *Nature*, **299**, 371–374.

Gafvelin,G. and Heijne,G.v. (1994) Topological 'frustration' in multispanning *E.coli* inner membrane proteins. *Cell*, **77**, 401–412.

Geest,M.v. and Lolkema,J.S. (2000) Membrane topology and insertion of membrane proteins: search for topogenic signals. *Microbiol. Mol. Biol. Rev.*, **64**, 13–33.

Hirokawa,T., Boon-Chieng,S. *et al.* (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*, **14**, 378–379.

Hofmann,K. and Stoffel,W. (1993) TMBASE—a database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler*, **374**, 166.

Horn,F., Weare,J. *et al.* (1998) GPCRDB: an information system for G protein-coupled receptors. *Nucleic Acids Res.*, **26**, 275–279.

Ji,T.H., Grossmann,M. *et al.* (1998) G protein-coupled receptors. I. Diversity of receptor–ligand interactions. *J. Biol. Chem.*, **273**, 17 299–17 302.

Jones,D.T., Taylor,W.R. *et al.* (1994) A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry*, **33**, 3038–3049.

Juretic,D. and Lucin,A. (1998) The preference functions method for predicting protein helical turns with membrane propensity. *J. Chem. Inf. Comput. Sci.*, **38**, 575–585.

Juretic,D., Jeroncic,A. *et al.* (1999) Sequence analysis of membrane proteins with the web server SPLIT. *Croatica Chemica Acta*, **77**, 975–997.

Klein,P., Kanehisa,M. *et al.* (1985) The detection and classification of membrane-spanning proteins. *Biochim. Biophys. Acta*, **815**, 468–476.

Krogh,A., Larsson,B. *et al.* (2001) Predicting transmembrane protein topology with a hidden Markov model: application to compete genomes. *J. Mol. Biol.*, in press.

Kyte,J. and Doolittle,R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, **157**, 105–132.

Möller,S., Kriventseva,E. *et al.* (2000) A collection of well characterised integral membrane proteins. *Bioinformatics*, **16**, 1159–1160.

Nakai,K. and Kanehisa,M. (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics*, **14**, 897–911.

Nielsen,H. and Krogh,A. (1998) Prediction of signal peptides and signal anchors by a hidden Markov model. *Int. Conf. Intell. Syst. Mol. Biol.* AAAI Press, Montreal, Canada, pp. 122–130.

Nielsen,H., Brunak,S. *et al.* (1999) Machine learning approaches to the prediction of signal peptides and other protein sorting signal. *Protein Eng.*, **12**, 3–9.

Nilsson,J., Persson,B. *et al.* (2000) Consensus prediction of membrane protein topology. *FEBS Lett.*, **486**, 267–269.

Persson,B. and Argos,P. (1997) Prediction of membrane protein topology utilizing multiple sequence alignments. *J. Protein Chem.*, **16**, 453–457.

Rost,B., Casadio,R. *et al.* (1996) Refining neural network predictions for helical transmembrane proteins by dynamic programming. *Int. Conf. Intell. Syst. Mol. Biol.* AAAI Press, St Louis, USA, pp. 192–200.

Sonnhammer,E.L.L., Heijne,G.v. *et al.* (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Int. Conf. Intell. Syst. Mol. Biol.* AAAI Press, Montreal, Canada, pp. 176–182.

Stevens,T.J. and Arkin,I.T. (2000) Do more complex organisms have a greater proportion of membrane proteins in their genomes? *Proteins*, **39**, 417–420.

Tusnády,G.E. and Simon,I. (1998) Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J. Mol. Biol.*, **283**, 489–506.

Wallin,E. and Heijne,G.v. (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.*, **7**, 1029–1038.