united nations
educational, scientific
and cultural
organization

international atomic
energy agency

the
abdus salam
international centre for theoretical physics

SMR/1499 - 11

INTERNATIONAL WORKSHOP ON PROTEOMICS: PROTEIN STRUCTURE, FUNCTION AND INTERACTIONS

(5 - 16 May 2003)

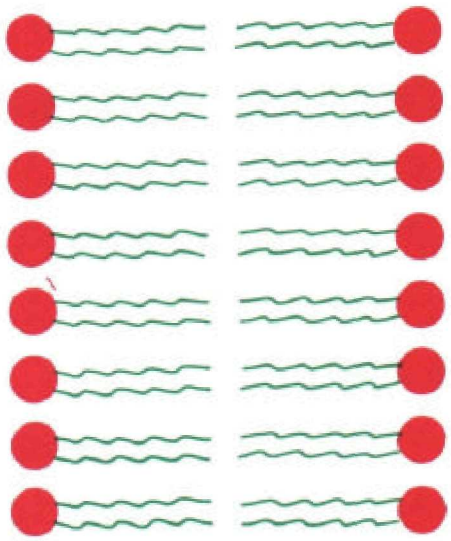
"Topology of transmembrane proteins"

presented by:

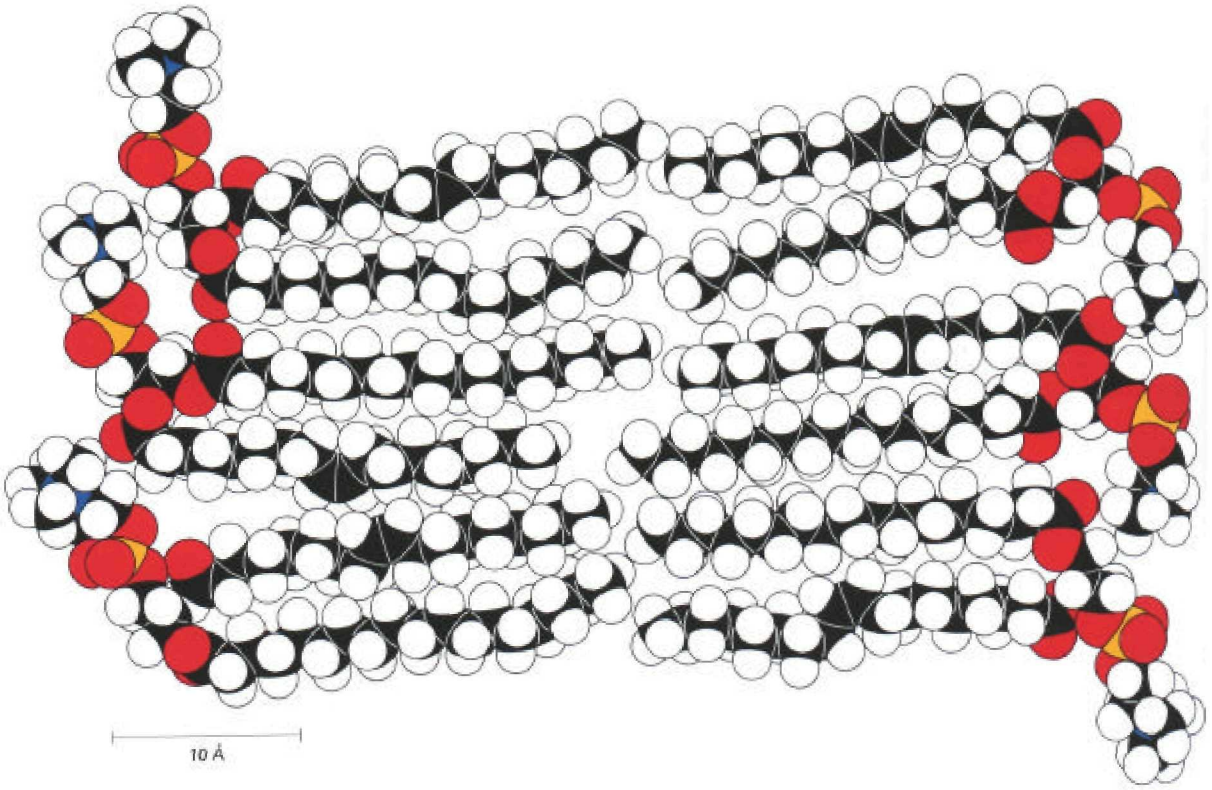
I. Simon
Institute of Enzymology, Budapest
Hungary

TOPOLOGY OF TRANSMEMBRANE PROTEINS

ISTVÁN SIMON



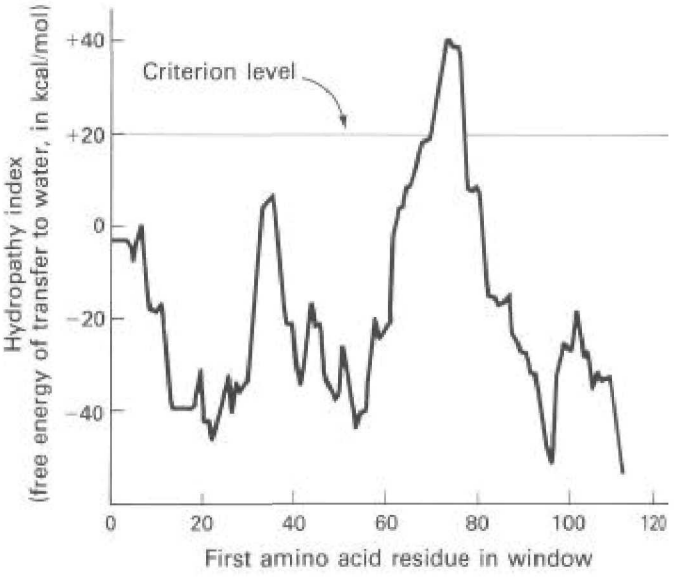
Phospholipid bilayer membrane



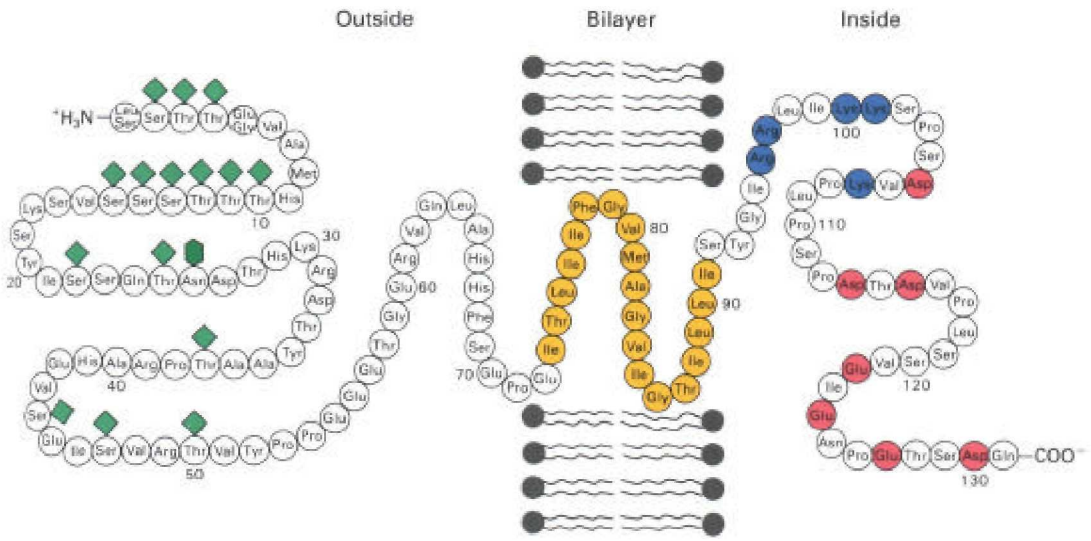
Space-filling model of a fluid
phospholipid bilayer membrane

Amino acid residue	*Transfer free energy (kcal/mol)*
Phe	3.7
Met	3.4
Ile	3.1
Leu	2.8
Val	2.6
Cys	2.0
Trp	1.9
Ala	1.6
Thr	1.2
Gly	1.0
Ser	0.6
Pro	−0.2
Tyr	−0.7
His	−3.0
Gln	−4.1
Asn	−4.8
Glu	−8.2
Lys	−8.8
Asp	−9.2
Arg	−12.3

Polarity scale for identifying trans-membrane helices



Hydropathy plot for glycophorin



Amino acid sequence and transmembrane disposition of glycophorin. The carboxyl-terminal part of the molecule, located on the cytosolic side of the membrane is rich in charged residues (red: negative, blue: positive)

J Chem Inf Comput Sci 2001 Mar-Apr;41(2):364-8

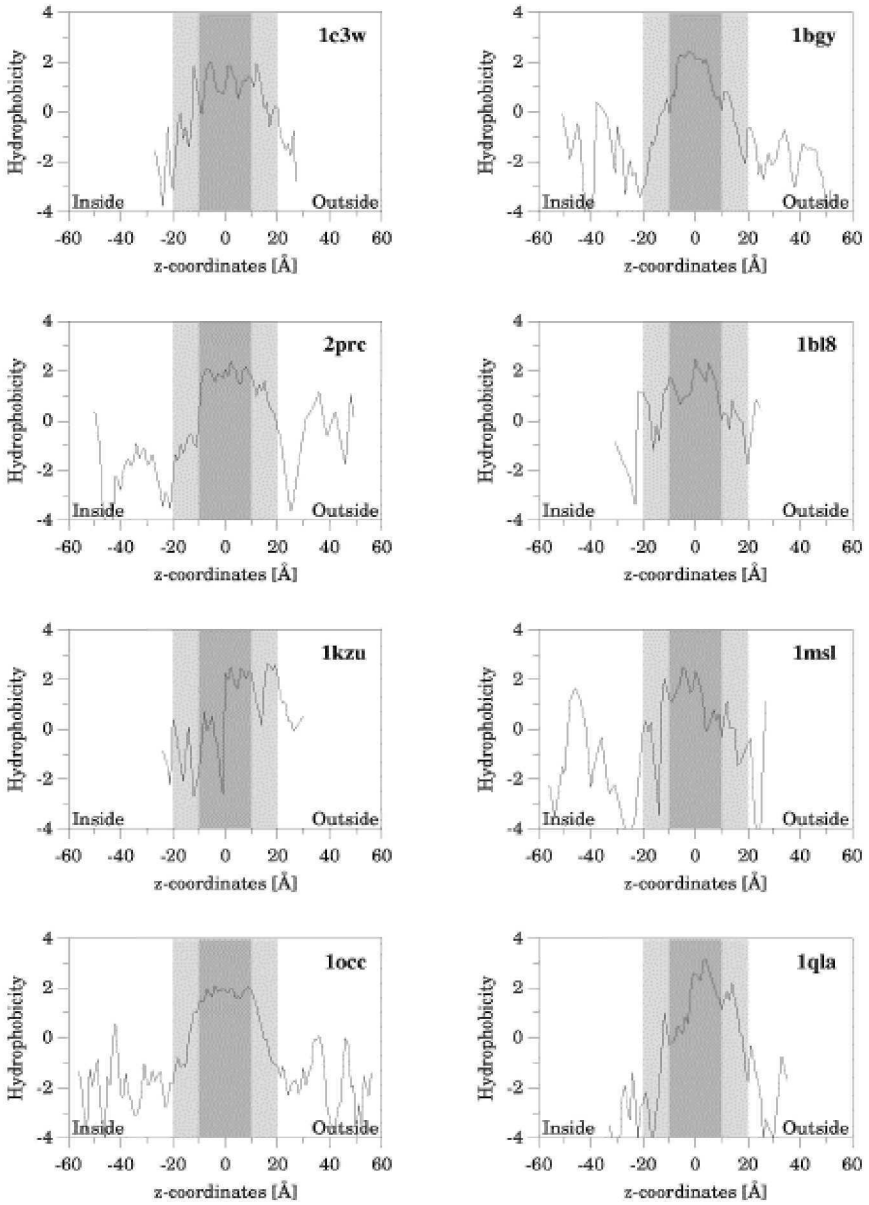
Topology of membrane proteins.

Tusnady GE, Simon I.

Institute of Enzymology, BRC, Hungarian Academy of Sciences, Budapest, Hungary.

Integral membrane proteins play important roles in living cells. Due to difficulties of experimental techniques, theoretical approaches, i.e., topology prediction methods, are important for structure determination of this class of proteins. Here we show a detailed comparison of transmembrane topology prediction methods. According to this comparison, we conclude that the topology of integral membrane proteins is determined by the maximum divergence of the amino acid composition of sequence segments. These segments are located in different areas of the cell, which can be characterized by different physicochemical properties. The results of these prediction methods compared to the X-ray diffraction data of several transmembrane proteins will also be discussed.

PMID: 11277724



Average hydrophobicity of various transmembrane proteins measured parallel to the average direction of the transmembrane helices in 1 Å slices. Zero is set to the average middle point of the transmembrane helices. The dark gray area marks the region defined by distances of ±10 Å, while the lighter gray area shows the region defined by distances of ±20 Å from the zero point.

J Mol Biol 1994 Oct 28;243(3):388-96

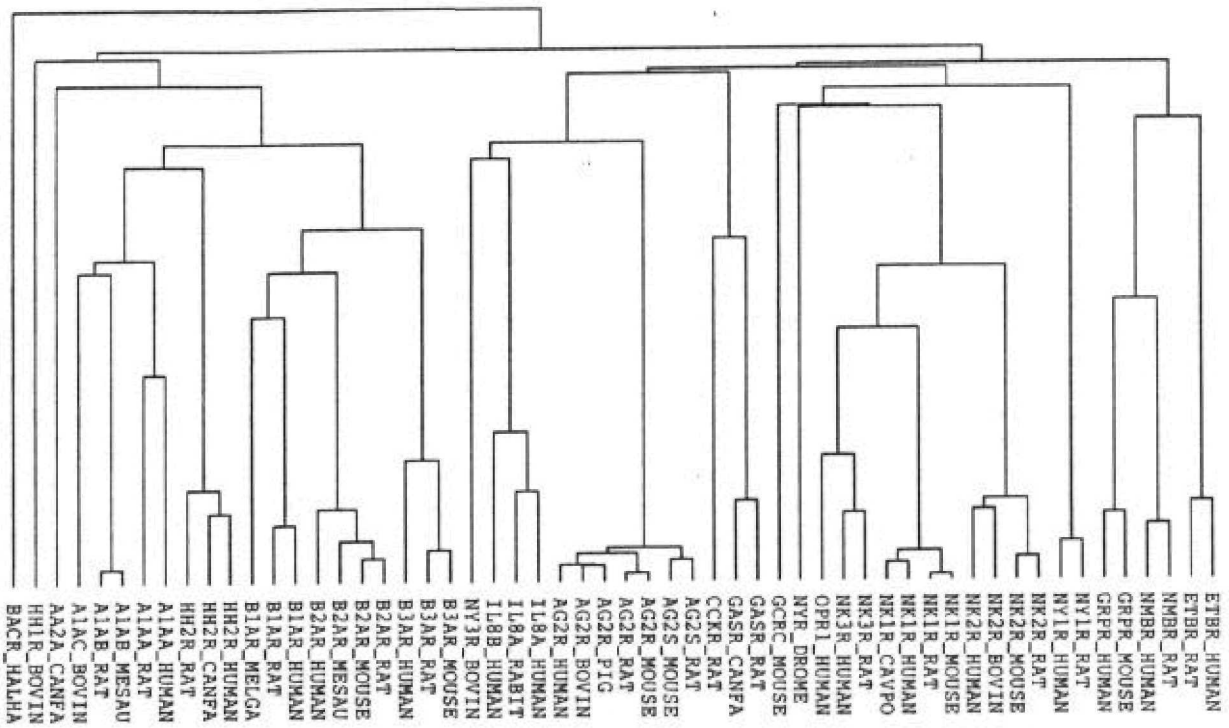
New alignment strategy for transmembrane proteins.

Cserzo M, Bernassau JM, Simon I, Maigret B.

Laboratoire de Chimie Theorique URA CNRS No. 510 Universite de Nancy-1-BP239, France.

In this paper an algorithm which locates helical transmembrane segments is described. It is shown that given the location of transmembrane helices of a protein, corresponding helices in another membrane related protein can be pinpointed. The method seems to be extremely insensitive to sequence identity but highly sensitive to the property of a sequence to assume transmembrane helical structure. As an example, using the present method, a sequence alignment between bacteriorhodopsin and human rhodopsin is carried out and it provides a good starting point for homology modeling of this G-protein coupled receptor. It is difficult to obtain this particular alignment using the traditional methods because of poor sequence homology. There are indications that hint at the broader range of applicability of the presented method.

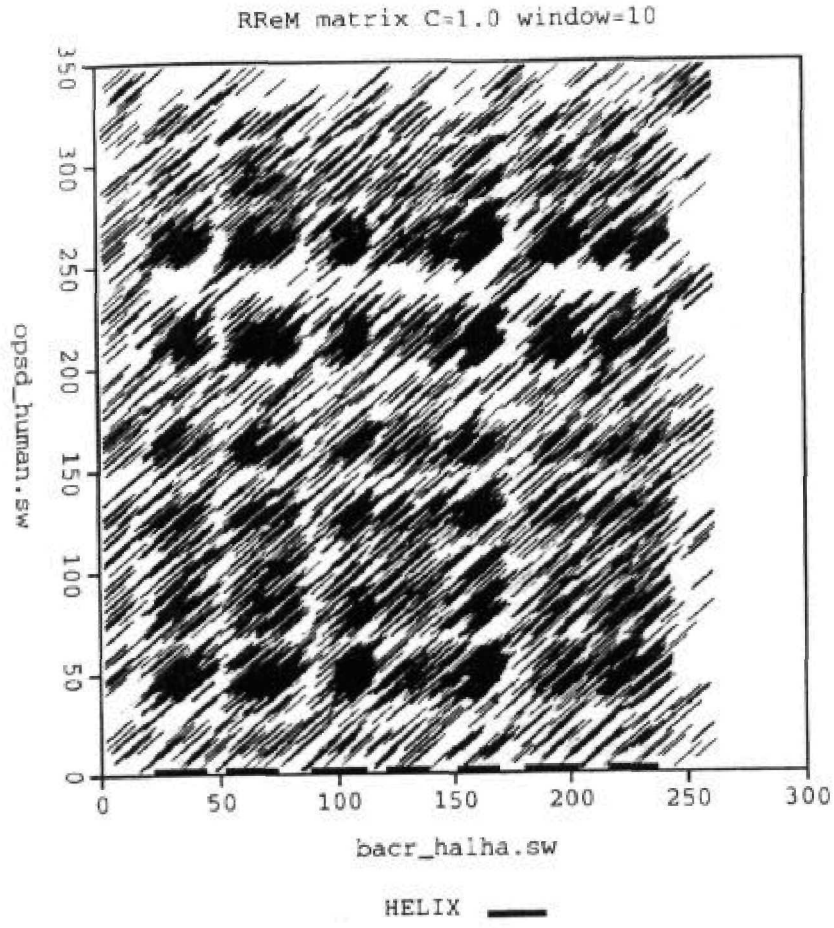
PMID: 7966267



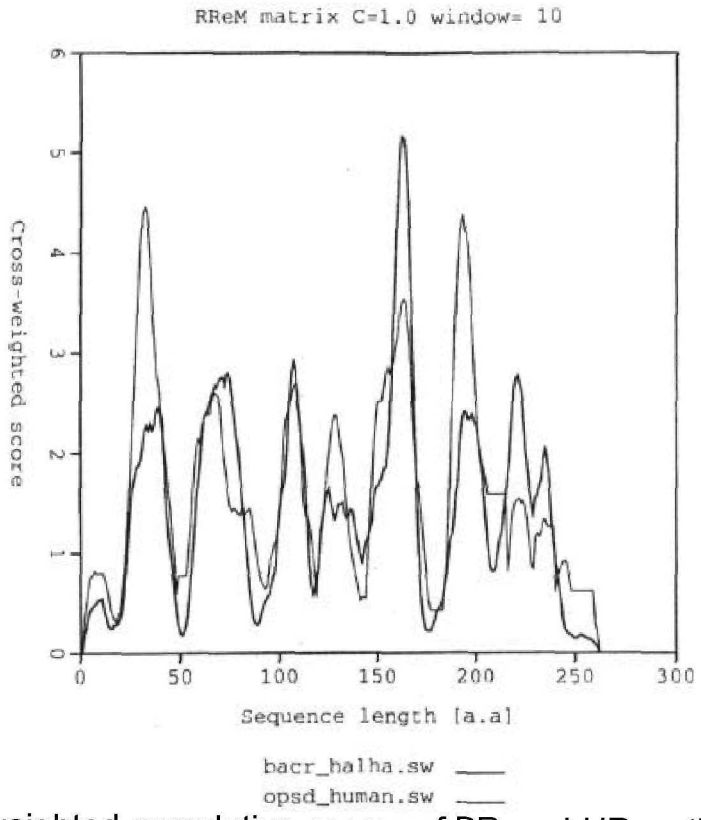
The homology clustering of 73 transmembrane proteins. The BR (SWISS-PROT code: bacr_halha) appears as a stand alone branch on the tree.
Figure is made by the 'pileup' utility of the GCG package.

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W
C	0.04																		
D	−0.08	−0.24																	
E	0.08	−0.51	0.37																
F	−0.25	0.30	−0.40	−0.50															
G	0.11	0.25	0.11	−0.23	−0.09														
H	−0.11	0.18	0.22	−0.19	−0.06	0.18													
I	−0.10	0.26	−0.54	−0.51	0.60	−0.22	−0.14												
K	−0.07	−0.43	0.18	0.54	−0.41	−0.23	−0.09	−0.40											
L	0.20	0.23	−0.55	−0.45	0.44	−0.14	−0.12	0.60	−0.34										
M	0.15	0.10	−0.25	−0.14	0.22	−0.01	−0.19	0.36	−0.16	0.38									
N	−0.14	−0.14	0.56	0.13	−0.33	0.22	0.32	−0.37	0.22	−0.39	−0.11								
P	−0.07	0.02	0.09	−0.11	−0.12	0.25	0.05	−0.19	−0.17	−0.20	−0.17	0.05							
Q	−0.01	−0.29	0.19	0.51	−0.44	−0.13	0.05	−0.43	0.52	−0.30	−0.18	0.20	−0.03						
R	−0.07	−0.20	−0.01	0.34	−0.31	−0.11	0.13	−0.29	0.47	−0.20	−0.17	0.12	−0.17	0.57					
S	0.14	0.10	0.22	−0.17	−0.01	0.29	0.14	−0.27	−0.24	−0.19	−0.17	0.14	0.27	−0.06	−0.16				
T	0.11	0.14	0.23	−0.27	0.02	−0.02	0.08	0.00	−0.24	−0.04	−0.02	0.13	0.20	−0.15	−0.32	0.45			
V	0.06	0.33	−0.48	−0.44	0.47	−0.12	−0.18	0.70	−0.36	0.49	0.23	−0.46	−0.17	−0.38	−0.29	−0.18	0.07		
W	−0.25	0.18	−0.41	−0.22	0.41	−0.08	−0.11	0.36	−0.22	0.22	0.16	−0.39	−0.13	−0.17	−0.03	−0.17	−0.12	0.29	
Y	−0.36	0.21	−0.34	−0.32	0.54	−0.24	0.08	0.43	−0.20	0.17	0.07	−0.27	−0.23	−0.28	−0.03	−0.21	−0.14	0.31	0.43

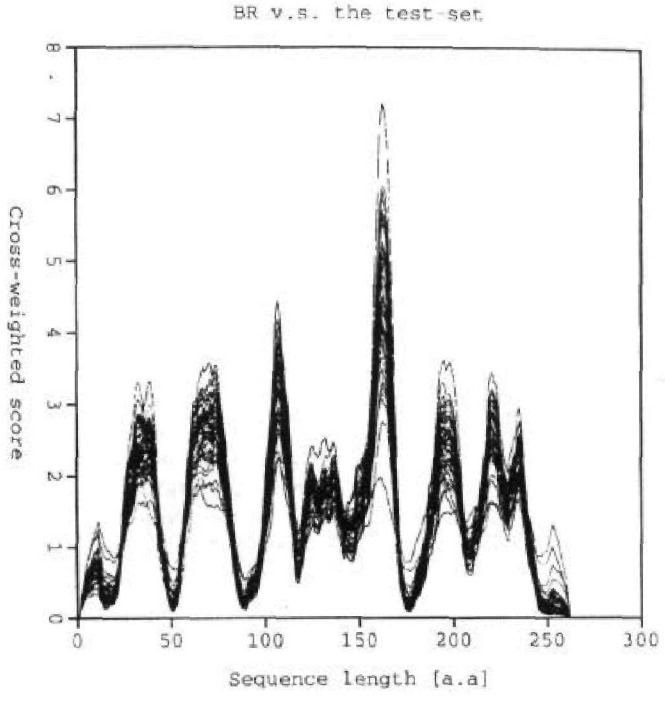
Residue Replace-ability Matrix



The alignment surface of the BR versus HR.
A 10 residue window size has been applied



The crossweighted cumulative scores of BR and HR as the modified Needleman & Wunsch algorithm aligns them



The crossweighted cumulative score profiles of BR obtained with the alignment against the 56 other proteins of the test-set. The positions of the peaks are conserved. The height of the peaks deviate considerably, but typically higher than 2.

Protein Eng 1997 Jun;10(6):673-6

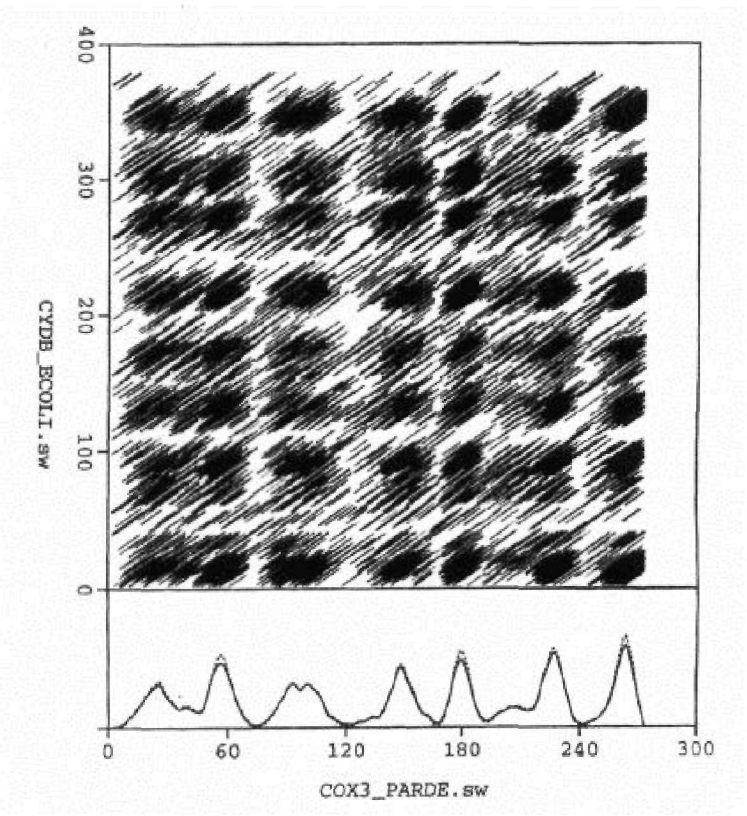
Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method.

Cserzo M, Wallin E, Simon I, von Heijne G, Elofsson A.

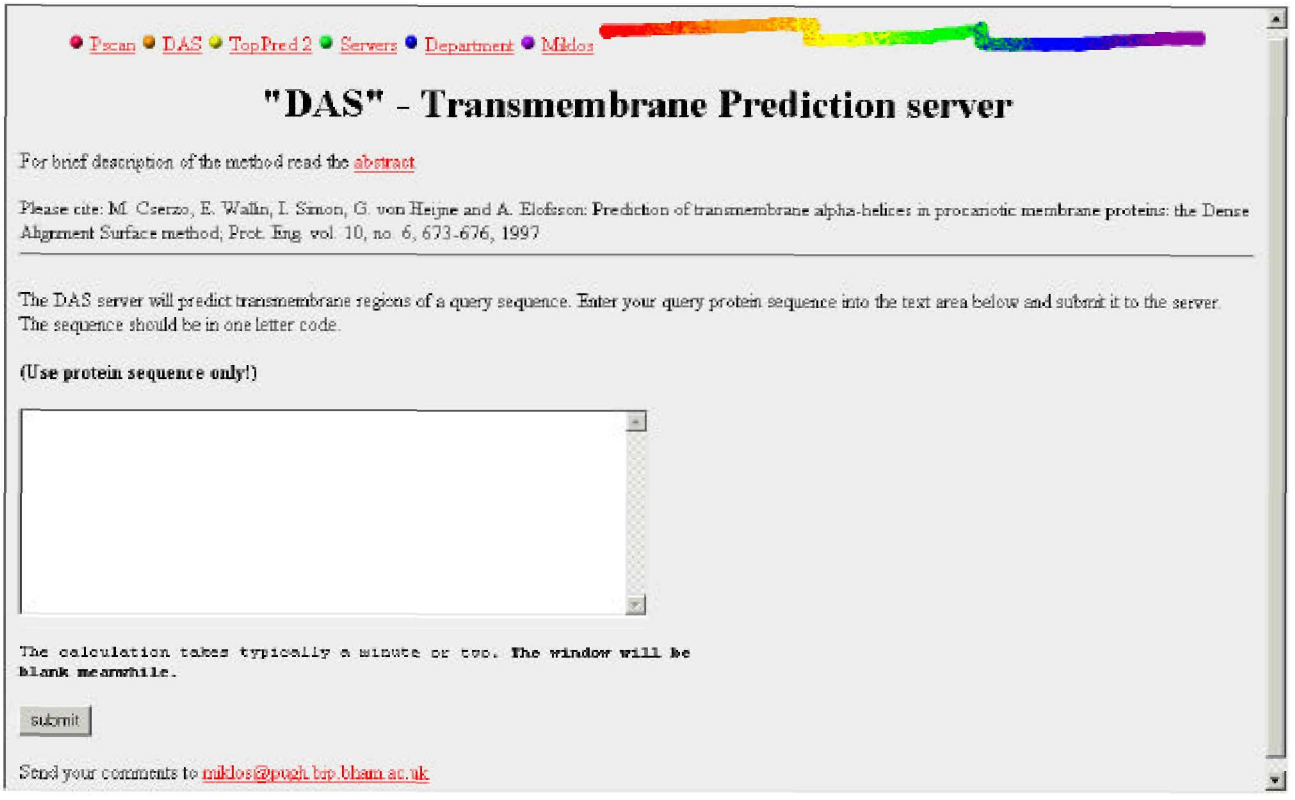
Institute of Enzymology, Biological Research Center Hungarian Academy of Sciences, Budapest.

A new, simple method for predicting transmembrane segments in integral membrane proteins has been developed. It is based on low-stringency dot-plots of the query sequence against a collection of non-homologous membrane proteins using a previously derived scoring matrix [Cserzo et al., 1994, J. Mol. Biol., 243, 388-396]. This so-called dense alignment surface (DAS) method is shown to perform on par with earlier methods that require extra information in the form of multiple sequence alignments or the distribution of positively charged residues outside the transmembrane segments, and thus improves prediction abilities when only single-sequence information is available or for classes of membrane proteins that do not follow the 'positive inside' rule.

PMID: 9278280



DAS plot of two arbitrarily chosen proteins (COX3_PARDE versus CYDB_ECOLI). The cross weighted cumulative score profile (dotted line) and the global DAS profile calculated as the average of the cumulative score profile obtained for comparisons with the other 43 proteins in the test set are also shown for COX3. COX3 has 7 and CYDB has 8 transmembrane segments



The DAS server home page available at http://www.sbc.su.se/~miklos/DAS/

J Mol Biol 1998 Oct 23;283(2):489-506

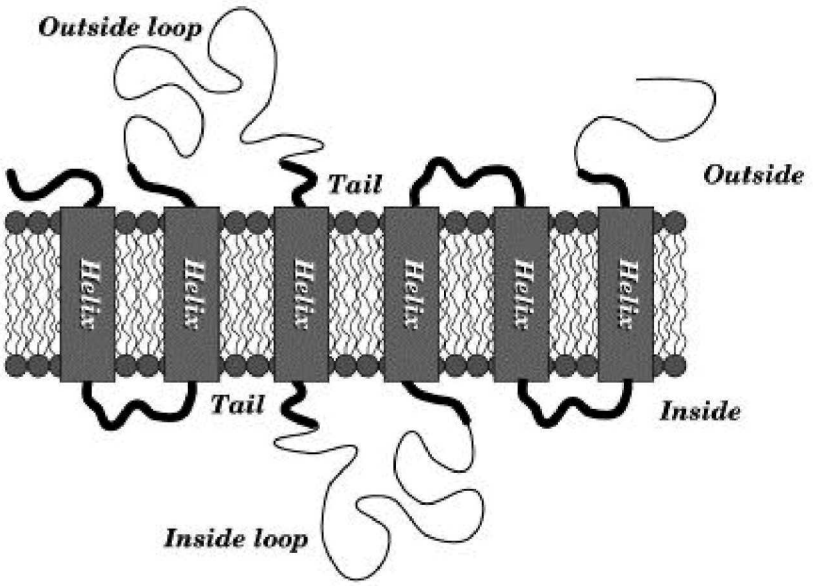
Principles governing amino acid composition of integral membrane proteins: application to topology prediction.

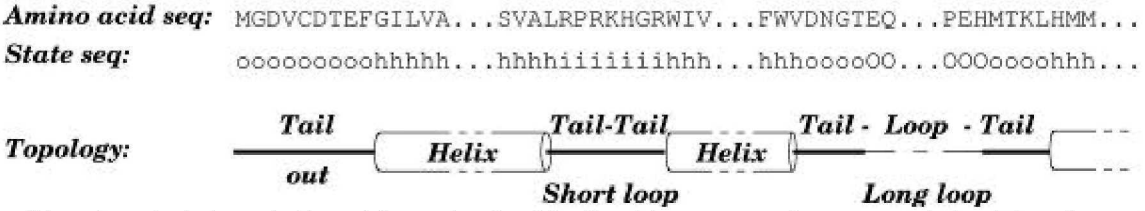
Tusnady GE, Simon I.

Institute of Enzymology. Biological Research Center, Hungarian Academy of Sciences, H-1518 Budapest, Hungary.

A new method is suggested here for topology prediction of helical transmembrane proteins. The method is based on the hypothesis that the localizations of the transmembrane segments and the topology are determined by the difference in the amino acid distributions in various structural parts of these proteins rather than by specific amino acid compositions of these parts. A hidden Markov model with special architecture was developed to search transmembrane topology corresponding to the maximum likelihood among all the possible topologies of a given protein. The prediction accuracy was tested on 158 proteins and was found to be higher than that found using prediction methods already available. The method successfully predicted all the transmembrane segments in 143 proteins out of the 158, and for 135 of these proteins both the membrane spanning regions and the topologies were predicted correctly. The observed level of accuracy is a strong argument in favor of our hypothesis.

PMID: 9769220





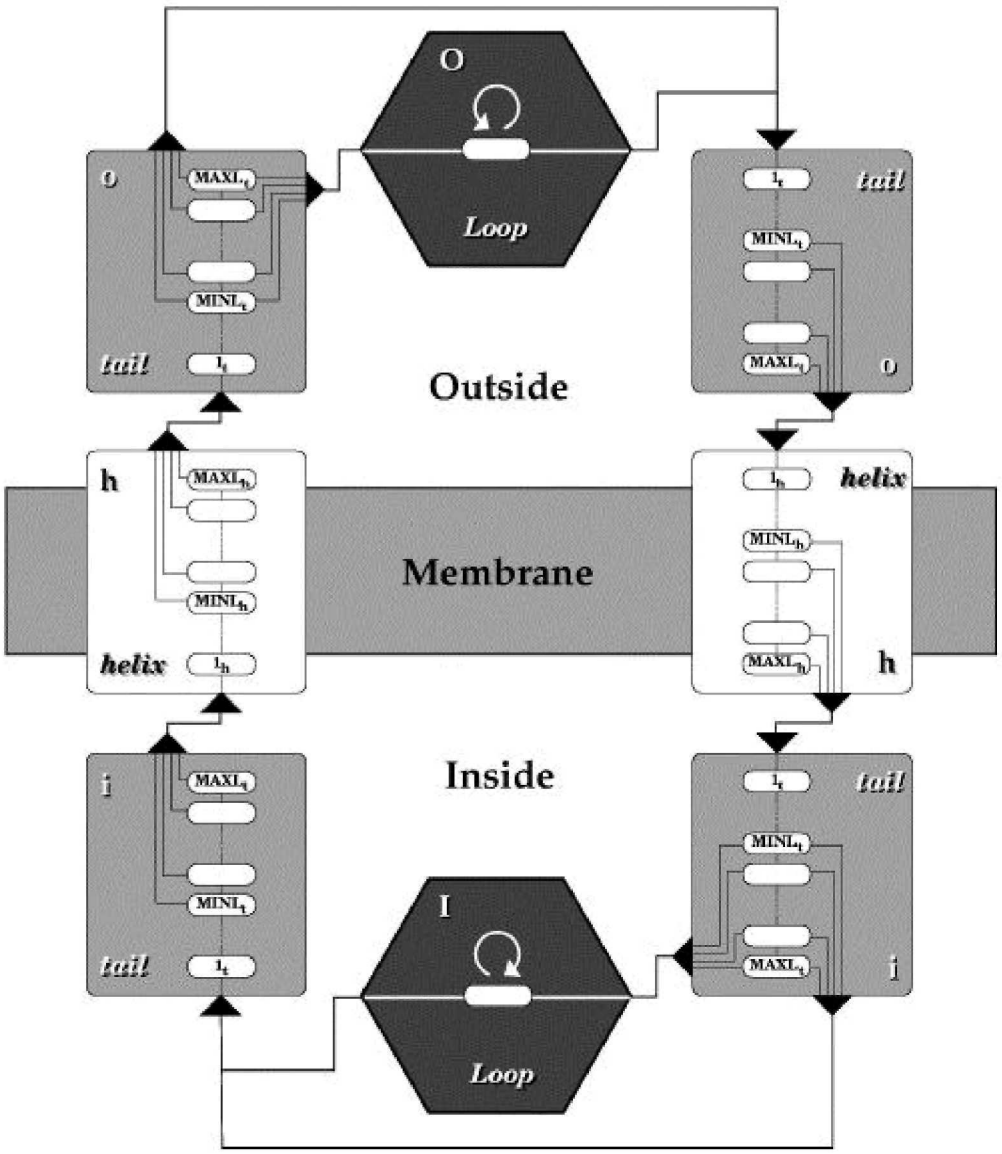
Structural states defined for a typical helical transmembrane protein. The five states are: inside loop (I), inside tail (i), membrane helix (h), outside tail (o) and outside loop (O). Tails (thick lines) are thought to interact with the inside or outside parts of the membrane, while loops (thin lines) do not. Two tails between helices can form a short loop, but longer loops are formed by tail-loop-tail sequences

$$F = \sum_{i=1}^{5} \sum_{j=1}^{20} q_{ij} \log\left(\frac{q_{ij}}{p_j}\right)$$

Sum of the difference of amino acid compositions

where p_j is the frequency of residue j in the whole protein and q_{ij} is the frequency of residue j in the structural part i.

Architecture of HMM used for topology prediction



States with the same transition matrices are colored in the same way: white, helix states; light gray, tail states; dark gray, loop states. Rectangular areas FL type states; hexagonal ones, NFL type states. The observation-symbol probabilities used by states are marked in each state. The structure of substates in the case of the FL type is drawn within states. Lines and arrows show the possible transition between states or substates.

Bioinformatics 2001 Sep;17(9):849-50

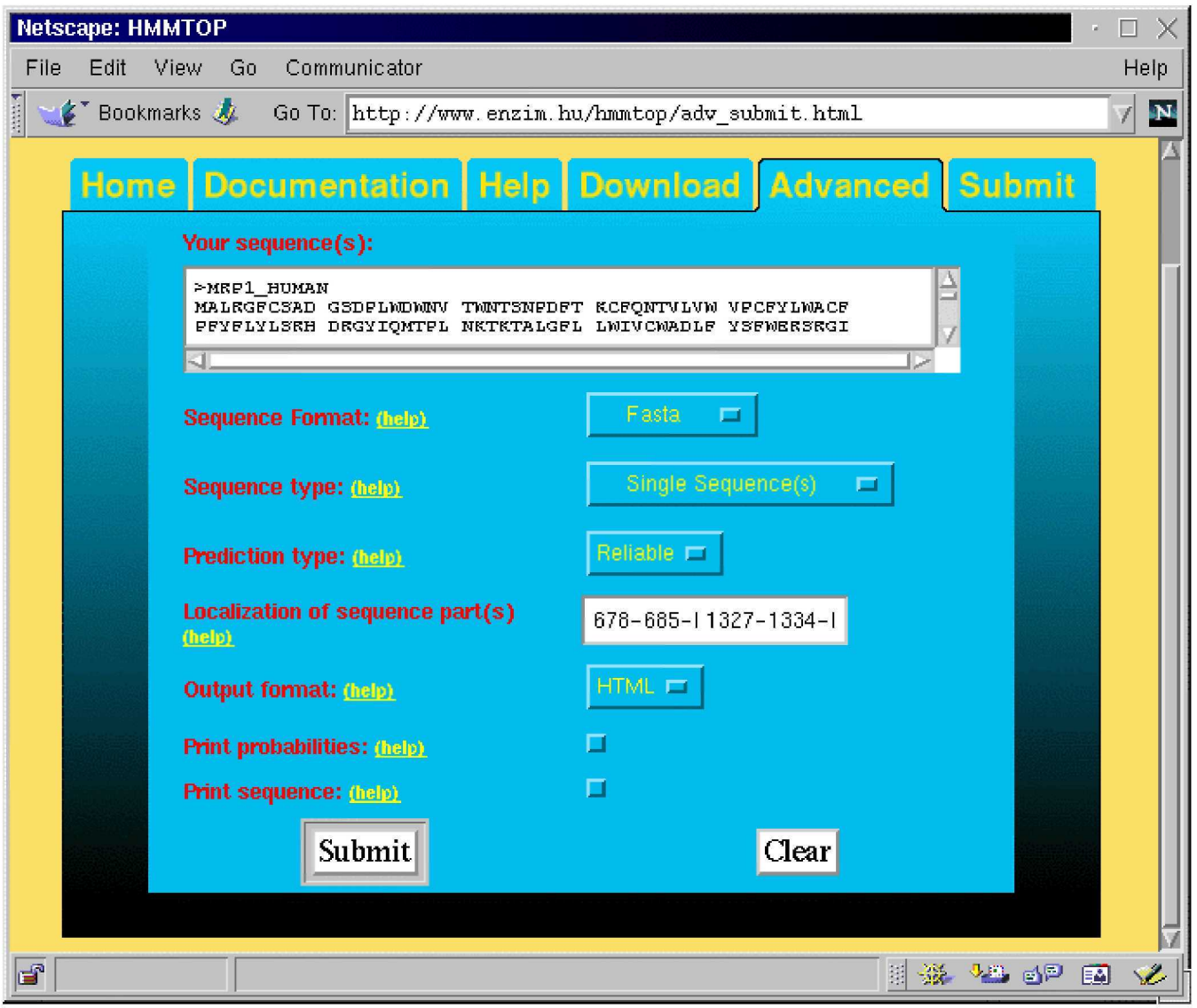
The HMMTOP transmembrane topology prediction server.

Tusnady GE, Simon I.

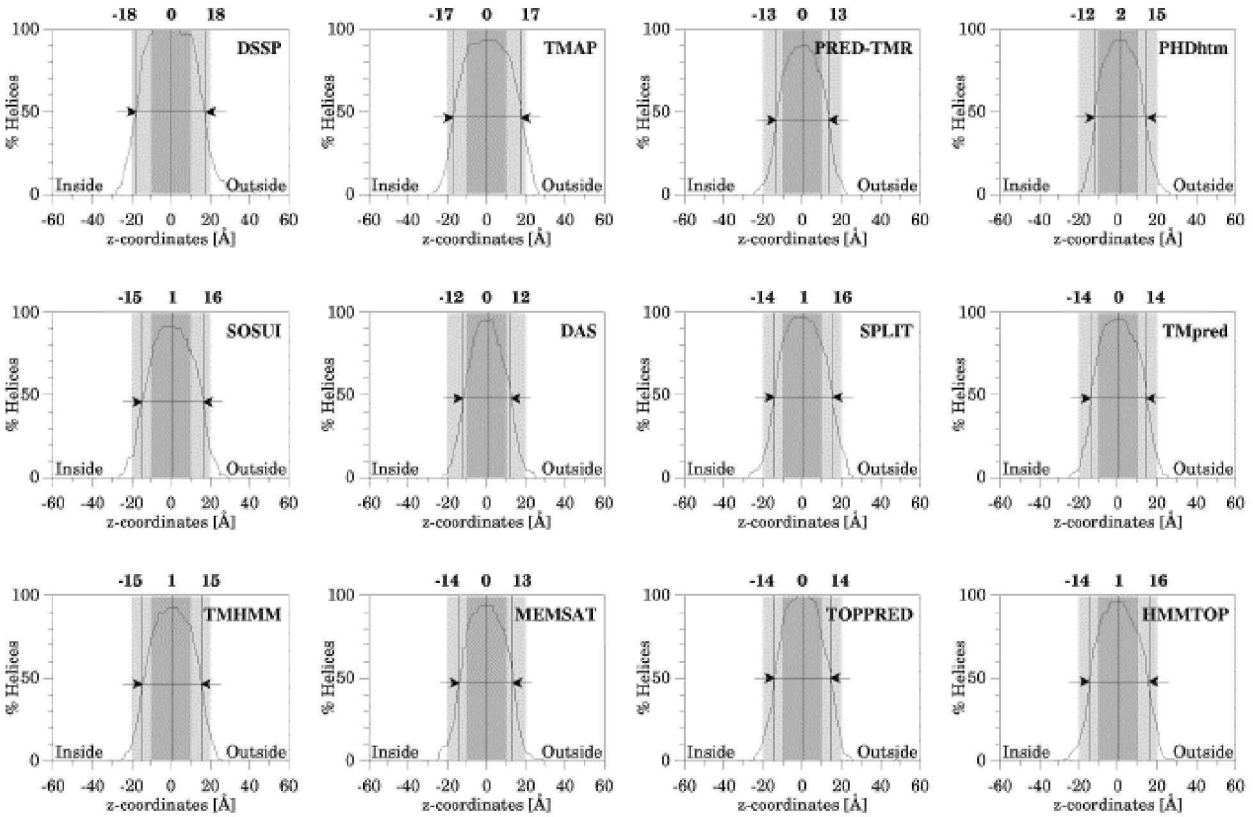
Institute of Enzymology, BRC, Hungarian Academy of Sciences, H-1518 Budapest, PO Box 7, Hungary. tusi@enzim.hu

The HMMTOP transmembrane topology prediction server predicts both the localization of helical transmembrane segments and the topology of transmembrane proteins. Recently, several improvements have been introduced to the original method. Now, the user is allowed to submit additional information about segment localization to enhance the prediction power. This option improves the prediction accuracy as well as helps the interpretation of experimental results, i.e. in epitope insertion experiments. Availability: HMMTOP 2.0 is freely available to non-commercial users at http://www.enzim.hu/hmmtop. Source code is also available upon request to academic users.

PMID: 11590105



The HMMTOP-2 server home page available at „http://www.enzim.hu/hmmtop"



Distribution of transmembrane helices measured parallel to the average direction of the trans-membrane helices in 1Å slices. Transmembrane helices are defined by the DSSP program and predicted by various methods. The z-coordinates at the value of 50% and at the mean of the curves are shown above the graphs

Proc Natl Acad Sci U S A 2001 Apr 10;98(8):4431-6

Prion protein: evolution caught en route.

Tompa P, Tusnady GE, Cserzo M, Simon I.

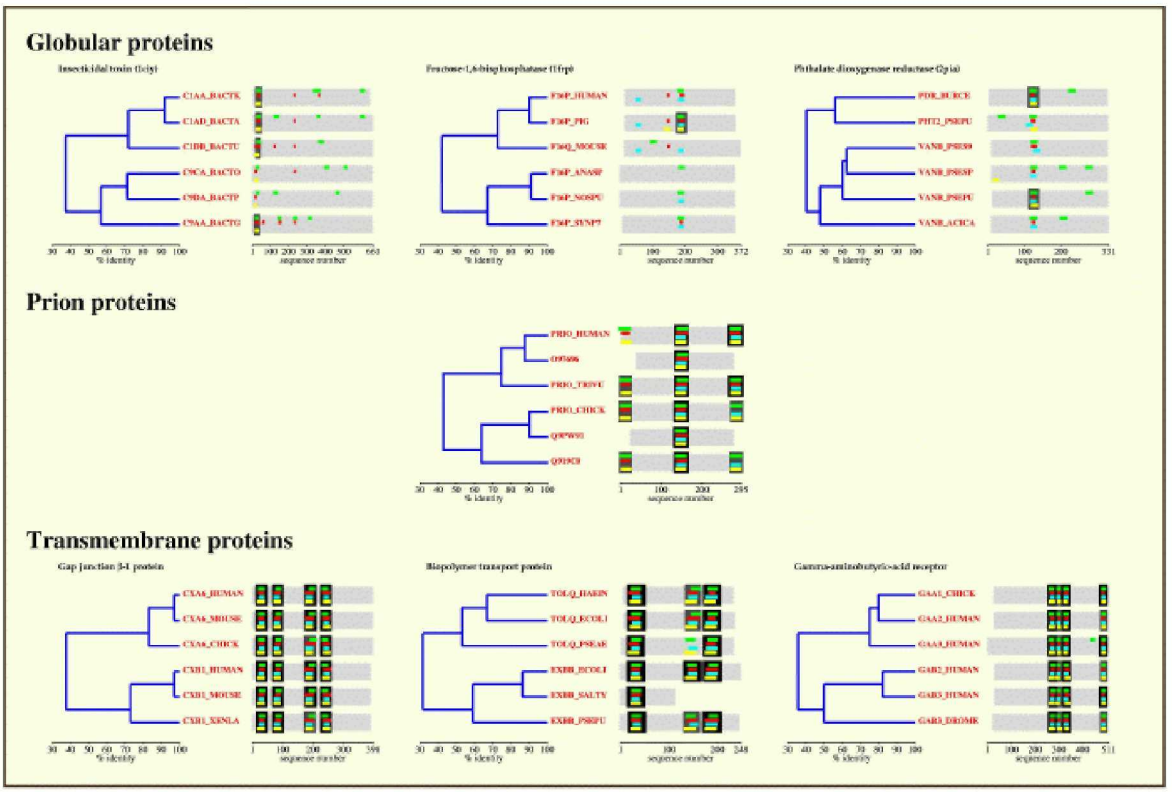
Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, P.O. Box 7, H-1518 Budapest, Hungary.

The prion protein displays a unique structural ambiguity in that it can adopt multiple stable conformations under physiological conditions. In our view, this puzzling feature resulted from a sudden environmental change in evolution when the prion, previously an integral membrane protein, got expelled into the extracellular space. Analysis of known vertebrate prions unveils a primordial transmembrane protein encrypted in their sequence, underlying this relocalization hypothesis. Apparently, the time elapsed since this event was insufficient to create a "minimally frustrated" sequence in the new milieu, probably due to the functional constraints set by the importance of the very flexibility that was created in the relocalization. This scenario may explain why, in a structural sense, the prion protein is still en route toward becoming a foldable globular protein.

PMID: 11287647

Prediction method	TOPPRED	DAS	PHDhtm	HMMTOP	TOPPRED	DAS	PHDhtm	HMMTOP	TOPPRED	DAS	PHDhtm	HMMTOP	TOPPRED	DAS	PHDhtm	HMMTOP	TOPPRED	DAS	PHDhtm	HMMTOP
Number of TMHs	0				1				2				3				more than 3			
Globular	523				0				0				0				0			
proteins	269	394	475	453	158	107	35	56	62	17	13	12	21	5	0	2	13	0	0	0
Transmembrane	0				43				7				6				102			
proteins	0	0	1	1	25	30	42	39	12	9	6	10	13	17	10	7	108	102	99	101
Prion proteins	0	0	0	0	2	2	2	2	0	2	4	0	4	2	0	4	0	0	0	0

Predictions by four different algorithms were made via the Internet by using the corresponding servers for 523 globular proteins, 158 membrane proteins with known transmembrane topology, and six prion proteins. For each structural class, the number of proteins with a given number of transmembrane helices (TMHs) observed (red) or predicted (blue) is shown.



The localization of TM helices in globular, transmembrane, and prion proteins. For comparison, 6–6 homologs of three TM and three globular proteins that have a similar relative similarity dendogram to that of the selected six prion proteins were selected (proteins are given by SWISS-PROT accession no. or ID). For each protein, TM helices are predicted by four prediction methods and are shown by color coding as follows: TOPPRED (green), DAS (red), PHDhtm (blue), and HMMTOP (yellow). A TM region is boxed if predicted by three (gray) or four (black) methods. Please note that there are only 11 globular proteins of 523 for which at least three methods predict a TM region; only three could be found with a relative similarity dendogram as shown

Biophys J 2002 Apr;82(4):1711-8

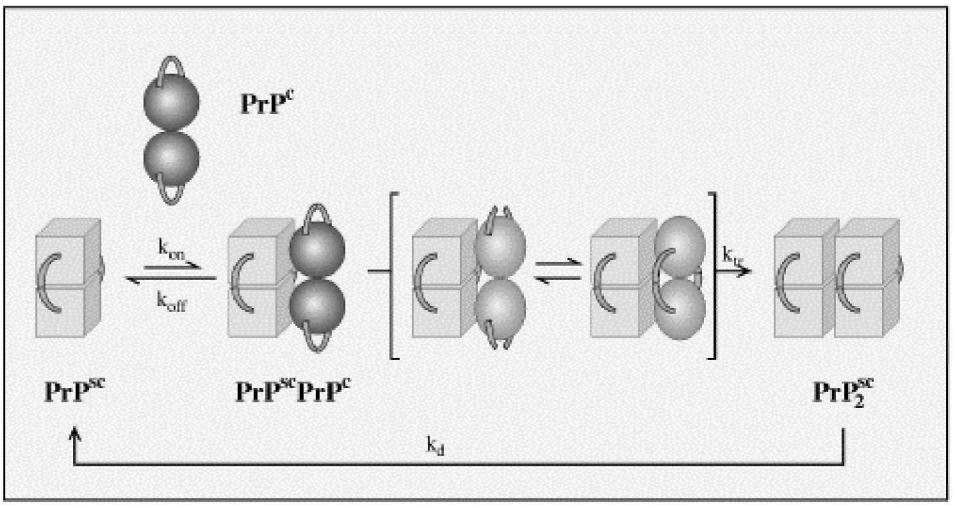
The role of dimerization in prion replication.

Tompa P, Tusnady GE, Friedrich P, Simon I.

Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, Budapest.

The central theme in prion diseases is the conformational transition of a cellular protein from a physiologic to a pathologic (so-called scrapie) state. Currently, two alternative models exist for the mechanism of this autocatalytic process; in the template assistance model the prion is assumed to be a monomer of the scrapie conformer, whereas in the nucleated polymerization model it is thought to be an amyloid rod. A recent variation on the latter assumes disulfide reshuffling as the mechanism of polymerization. The existence of stable dimers, let alone their mechanistic role, is not taken into account in either of these models. In this paper we review evidence supporting that the dimerization of either the normal or the scrapie state, or both, has a decisive role in prion replication. The contribution of redox changes, i.e., the temporary opening and possible rearrangement of the intramolecular disulfide bridge is also considered. We present a model including these features largely ignored so far and show that it adheres satisfactorily to the observed phenomenology of prion replication.

PMID: 11916832



Replication of the scrapie state based on dimer formation and disulfide rearrangement. The figure is a schematic rendering of a PrP c - PrP sc conversion model based on both dimer formation and disulfide rearrangement. It is assumed that PrP sc is a dimer of predominantly beta-structure, stabilized by two intermolecular disulfide bridges. The critical step in replication is the recruitment of a PrP c dimer with an alfa-helical structure. Upon binding, the structure of PrP c dimer unfolds to a large extent; within this transient structure the disulfide bonds open up and reform in an intermolecular fashion. This initiates the structures to collapse into the more stable scrapie state with prevailing beta-sheet(s). The newly formed scrapie dimer either diffuses away or remains in place, serving as a seed for amyloid.

Protein Eng 2002 Sep;15(9):745-52

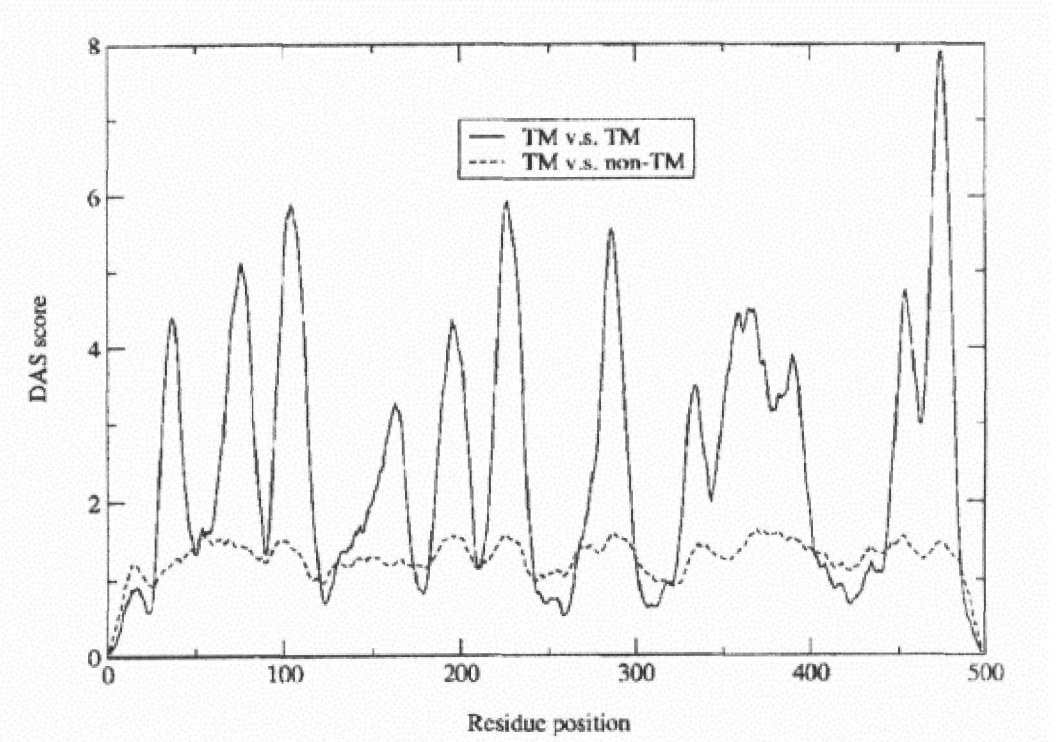
On filtering false positive transmembrane protein predictions.

Cserzo M, Eisenhaber F, Eisenhaber B, Simon I.

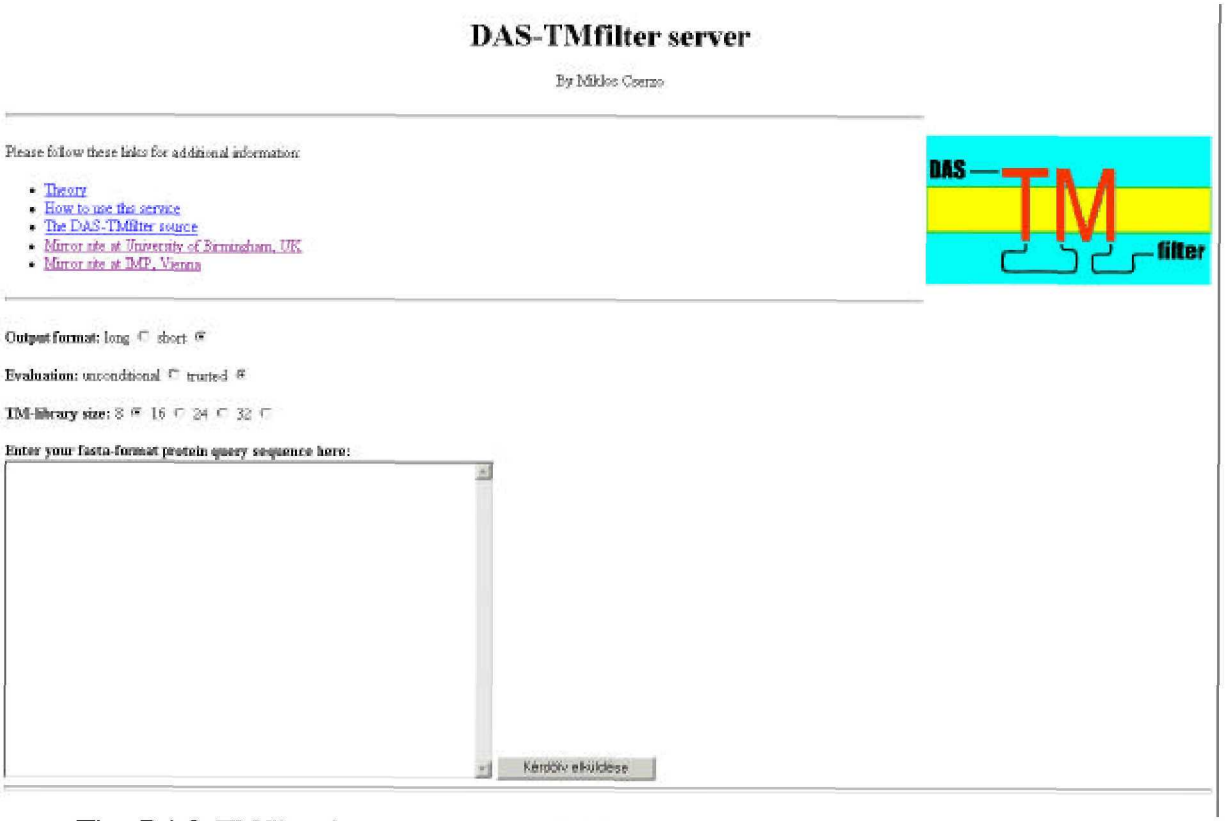
University of Birmingham, School of Biosciences, Edgbaston, Birmingham B15 2TT, UK.
miklos@bip.bham.ac.uk

While helical transmembrane (TM) region prediction tools achieve high (>90%) success rates for real integral membrane proteins, they produce a considerable number of false positive hits in sequences of known nontransmembrane queries. We propose a modification of the dense alignment surface (DAS) method that achieves a substantial decrease in the false positive error rate. Essentially, a sequence that includes possible transmembrane regions is compared in a second step with TM segments in a sequence library of documented transmembrane proteins. If the performance of the query sequence against the library of documented TM segment-containing sequences in this test is lower than an empirical threshold, it is classified as a non-transmembrane protein. The probability of false positive prediction for trusted TM region hits is expressed in terms of E-values. The modified DAS method, the DAS-TMfilter algorithm, has an unchanged high sensitivity for TM segments (approximately 95% detected in a learning set of 128 documented transmembrane proteins). At the same time, the selectivity measured over a non-redundant set of 526 soluble proteins with known 3D structure is approximately 99%, mainly because a large number of falsely predicted single membrane-pass proteins are eliminated by the DAS-TMfilter algorithm.

PMID: 12456873



DAS profiles of a TM protein as a function of residue number. The library DAS profile lambda (TLCA_RICPR, ***Q***) for the SWISS-PROT sequence with ac.num. P19568, a member of the TM library, has been averaged (i) over all library sequences as query ***Q*** (full line) and (ii) over all sequences of the non-TM set as query ***Q*** (dashed line). This example shows that non-TM sequences have a clear tendency to produce low library profiles whereas the reported TM regions can be recognized as peaks of the profile computed with true TM regions.



The DAS-TMfilter home page available at „http://www.enzim.hu/DAS/DAS.html"