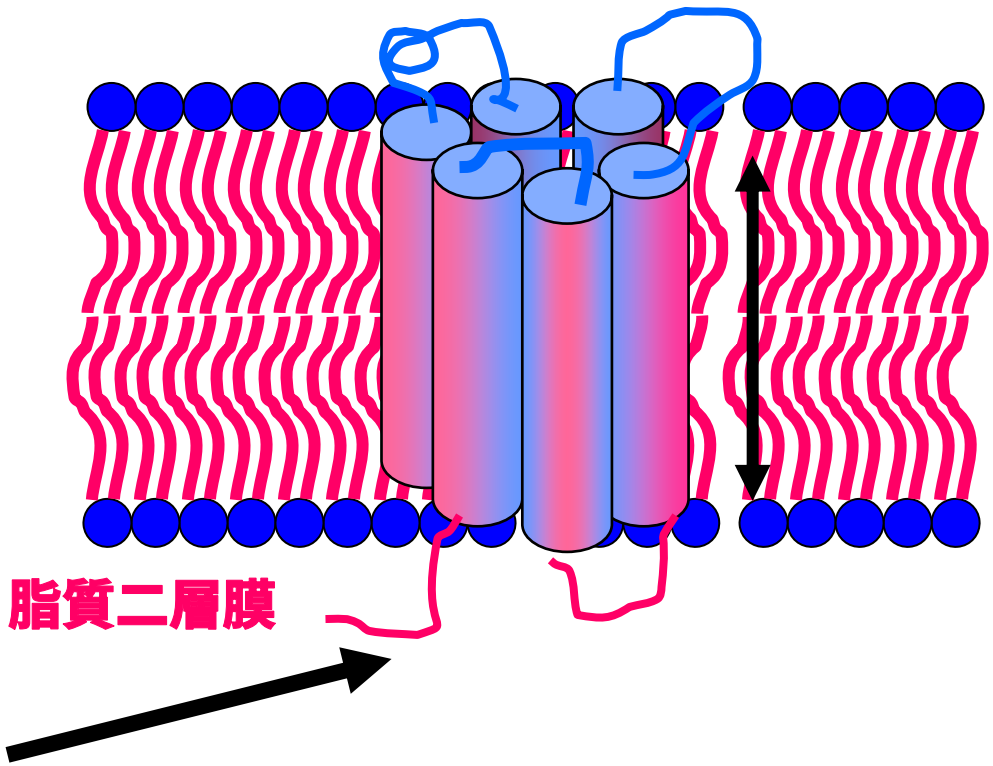
膜タンパク質判別、
膜貫通ヘリックス予測
トポロジー予測

Positive Inside Rule

細胞内側のループ部分には正電荷を持つ残基が多い



簡単な四則計算を用いた配列解析例

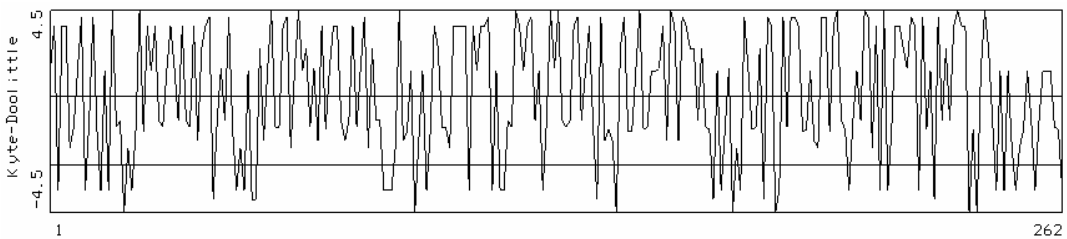
アミノ酸配列

```
1) mlellptavegvsqaqitgrpewiwlalgtalmglgtlyflvkgmgvsdp (50
51) dakkfyaittlvpaiaftmylsmllgygltmvpfggeqnpiywaryadwl (100
101) fttpllllldlallvdadqgtilalvgadgimigtglvgaltkvysyrfvw (150
```

アミノ酸疎水性指標

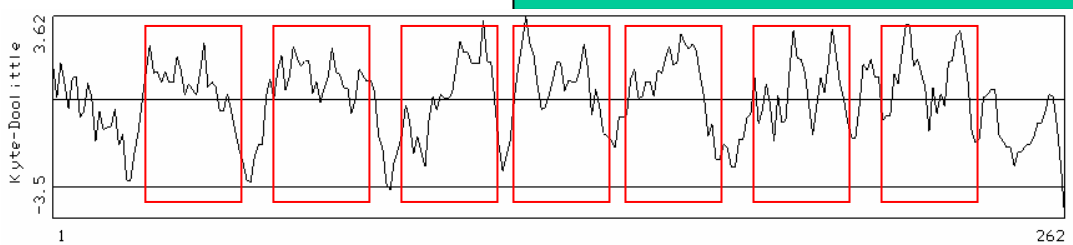
イソロイシン：I	4.5
バリン：V	4.2
ロイシン：L	3.8
フェニルアラニン：F	2.8
⋮	⋮

数値列化

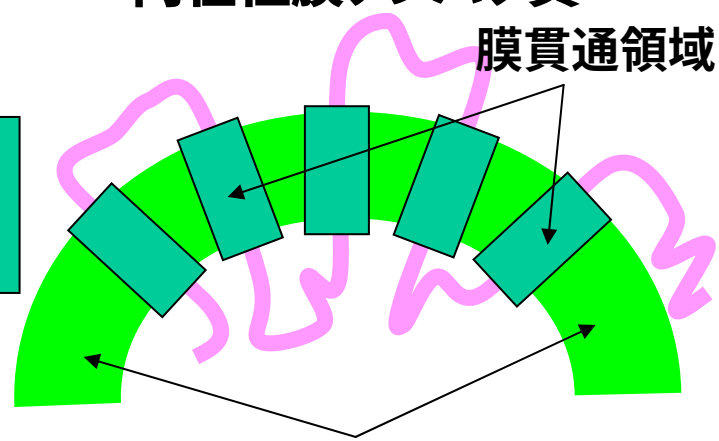


平均化（移動平均法）

$$\overline{H}(i) = \sum_{j=i-m}^{i+m} H(j) \Big/ 2m+1$$



内在性膜タンパク質



生体膜：親油性の環境

膜—水溶性タンパク質の判別予測

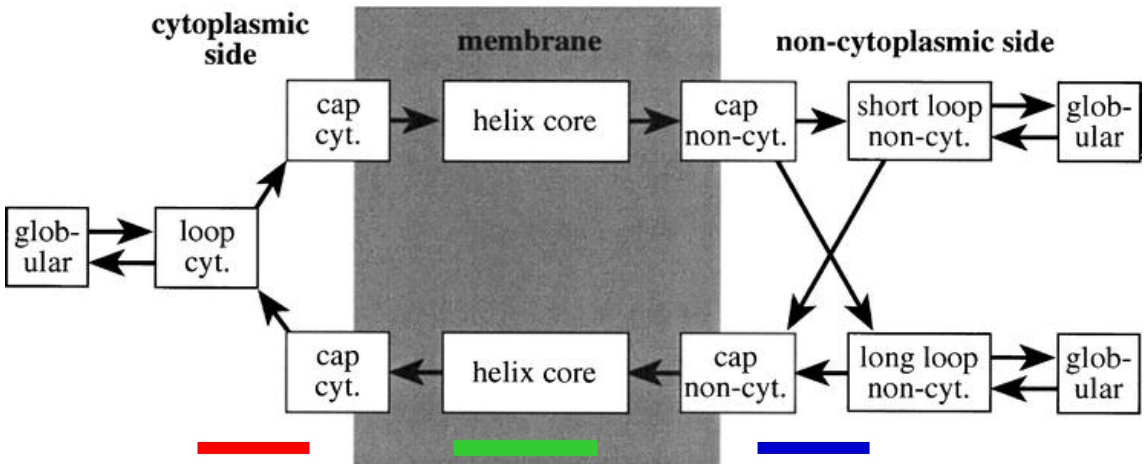
膜貫通ヘリックスを持っているものは、膜タンパク質

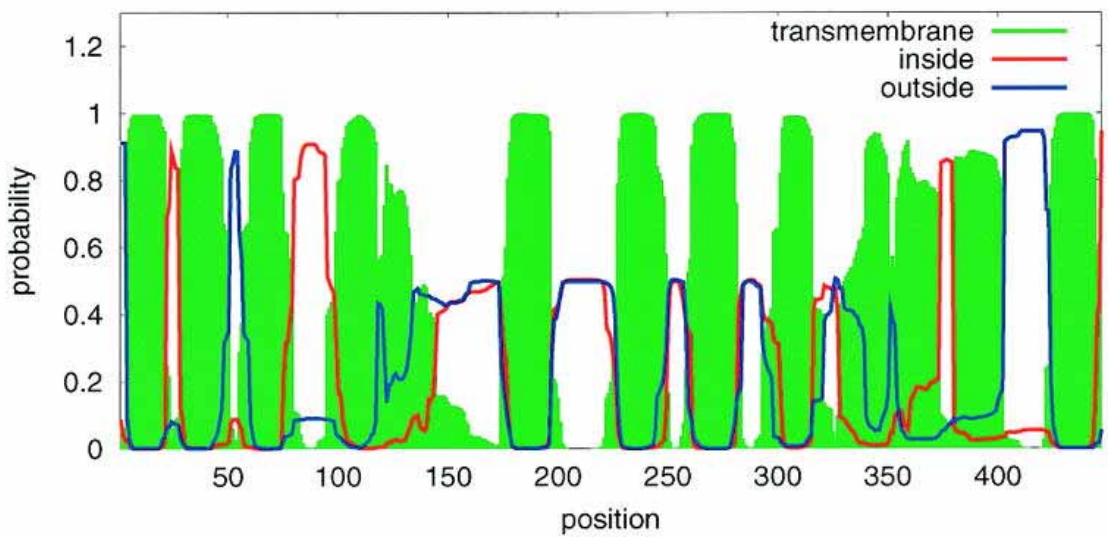
Prediction method (reference)	Feature (URL)
KKD (Klein *et al*., 1985)	hydrophobicity-based; discriminant function
SOSUI (Hirokawa *et al*., 1998)	hydrophobicity- and amphiphilicity-based; length of sequence (http://sosui.proteome.bio.tuat.ac.jp/cgi-bin/sosui.cgi?/sosui_submit.html)
TSEG (Kihara *et al*., 1998)	Mahalanobis distance with the average hydrophobicity and periodicity of hydrophobicity (http://www.genome.ad.jp/SIT/tsegdir/tseg_exe.html)
PRED-TMR2 (Pasquier and Hamodrakas, 1999)	artificial NN (http://biophysics.biol.uoa.gr/PRED-TMR2/input.html)
TMHMM 2.0 (Krogh *et al*., 2001)	HMM (http://www.cbs.dtu.dk/services/TMHMM-2.0/)
PRED-CLASS (Pasquier *et al*., 2001)	cascading artificial NN (http://o2.biol.uoa.gr/PRED-CLASS/input.html)
DAS-TMfilter (Cserzö *et al*., 2002; 2004)	comparison between transmembrane segments in a library of documented proteins (http://www.enzim.hu/DAS/DAS.html)

TMHMM 膜貫通ヘリックス予測

膜貫通ヘリックス部分の隠れマルコフモデルを予測に応用。

(a)

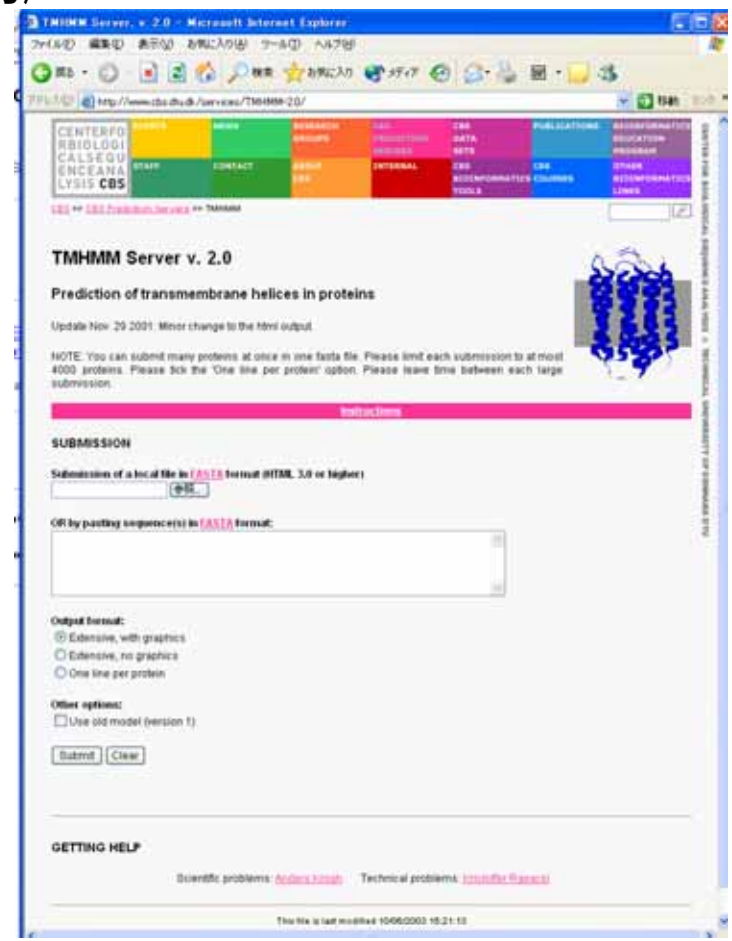




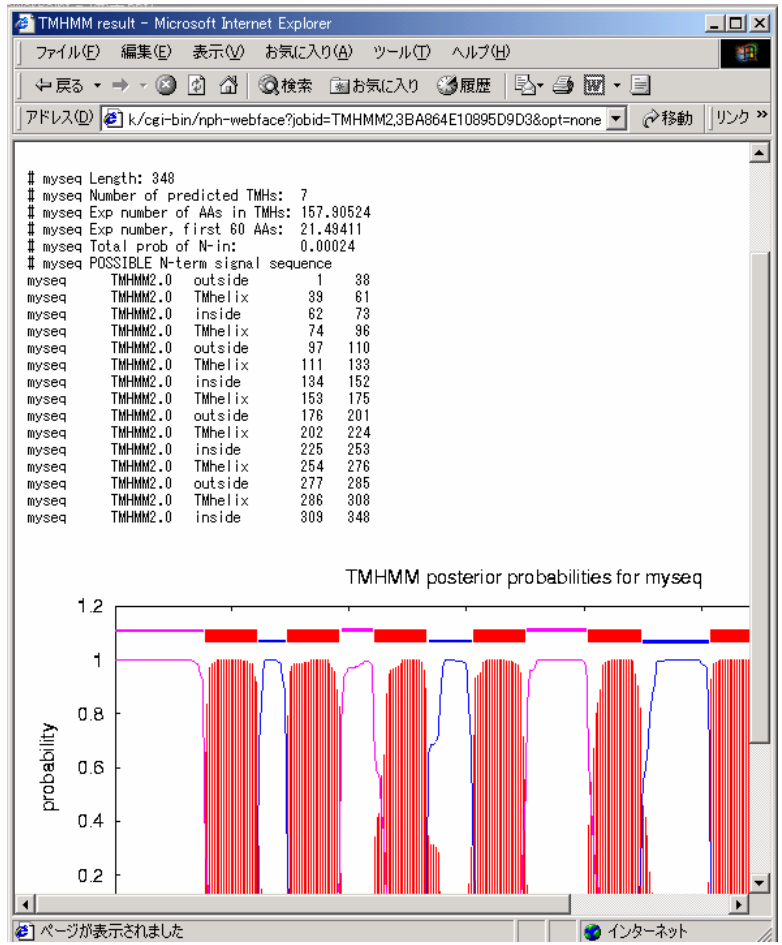
膜貫通部位予測ツール：TMHMM

http://www.cbs.dtu.dk/services/TMHMM-2.0/

入力)



出力)



PRED-TMR2 (Pasquier and Hamodrakas, 1999)

Table I. Propensity values and corresponding input used in the neural network for the 20 amino acid residue types that belong to transmembrane segments, calculated from the entire SWISS-PROT database

Residue		P_i	NN input
Phenylalanine	F	2.235	1.000
Isoleucine	I	2.083	0.929
Leucine	L	1.845	0.817
Tryptophan	W	1.790	0.791
Valine	V	1.756	0.775
Methionine	M	1.502	0.655
Alanine	A	1.383	0.599
Cysteine	C	1.202	0.514
Glycine	G	1.158	0.494
Tyrosine	Y	1.075	0.455
Threonine	T	0.879	0.362
Serine	S	0.806	0.328
Proline	P	0.597	0.230
Histidine	H	0.395	0.135
Asparagine	N	0.389	0.132
Glutamine	Q	0.273	0.078
Aspartic acid	D	0.153	0.021
Glutamic acid	E	0.131	0.011
Arginine	R	0.124	0.007
Lysine	K	0.108	0.000

A propensity for each residue to be in a transmembrane region was calculated using the formula

$$P_i = \frac{F_i^{TM}}{F_i}, \qquad (1)$$

where P_i is the propensity value (transmembrane potential) of residue type i and F_i^{TM} and F_i are the frequencies of the ith type of residue in transmembrane segments and in the entire SWISS-PROT database respectively. Values above 1 indicate a preference for a residue to be in the lipid-associated structure of a transmembrane protein, whereas propensities below 1 characterize unfavorable transmembrane residues.

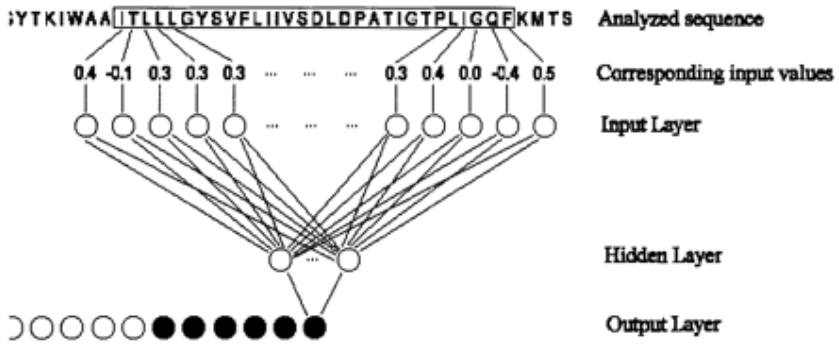
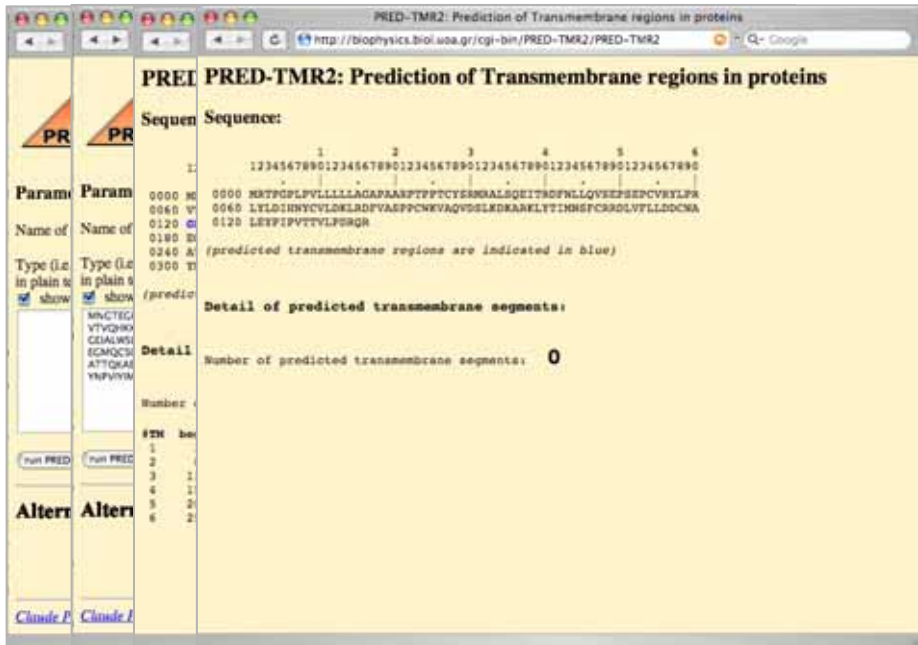


Fig. 1. Schematic architecture of the neural network. Amino acids of the input sequence are converted to unique input values corresponding to the propensity for each amino acid to be located inside a transmembrane region (see Table I). Output of the network consists of values between 0 and 1. Values above 0.9 (shown in black on the figure) indicate a detection of a potential transmembrane segment.



PRED-CLASS (Pasquier *et al.*, 2001)

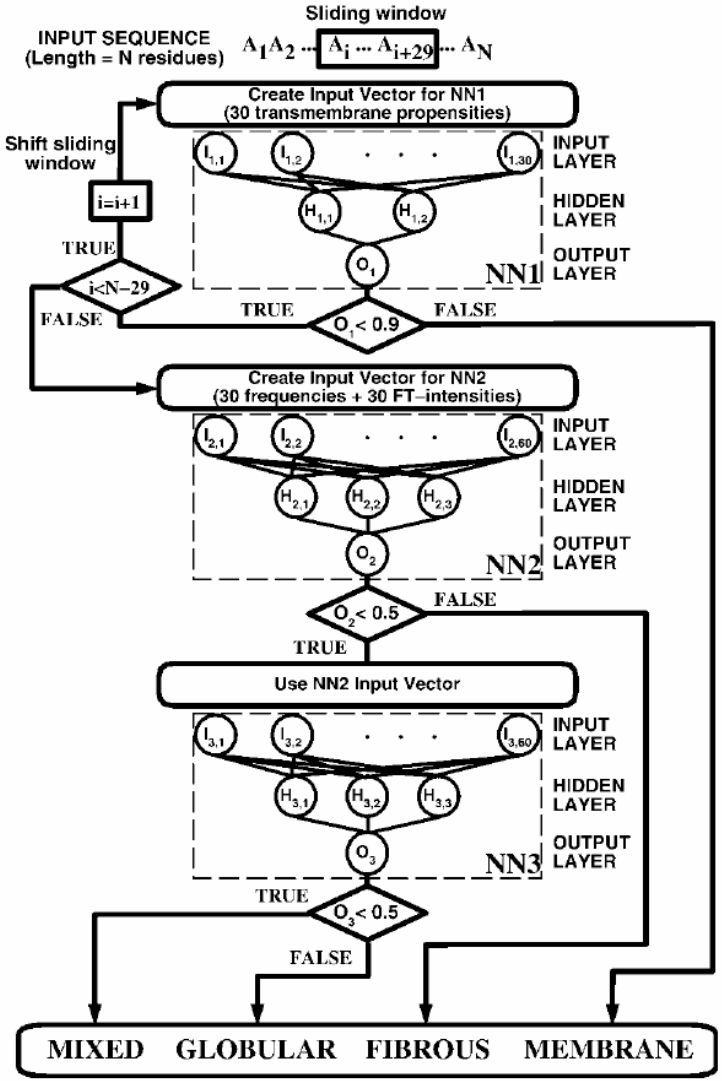
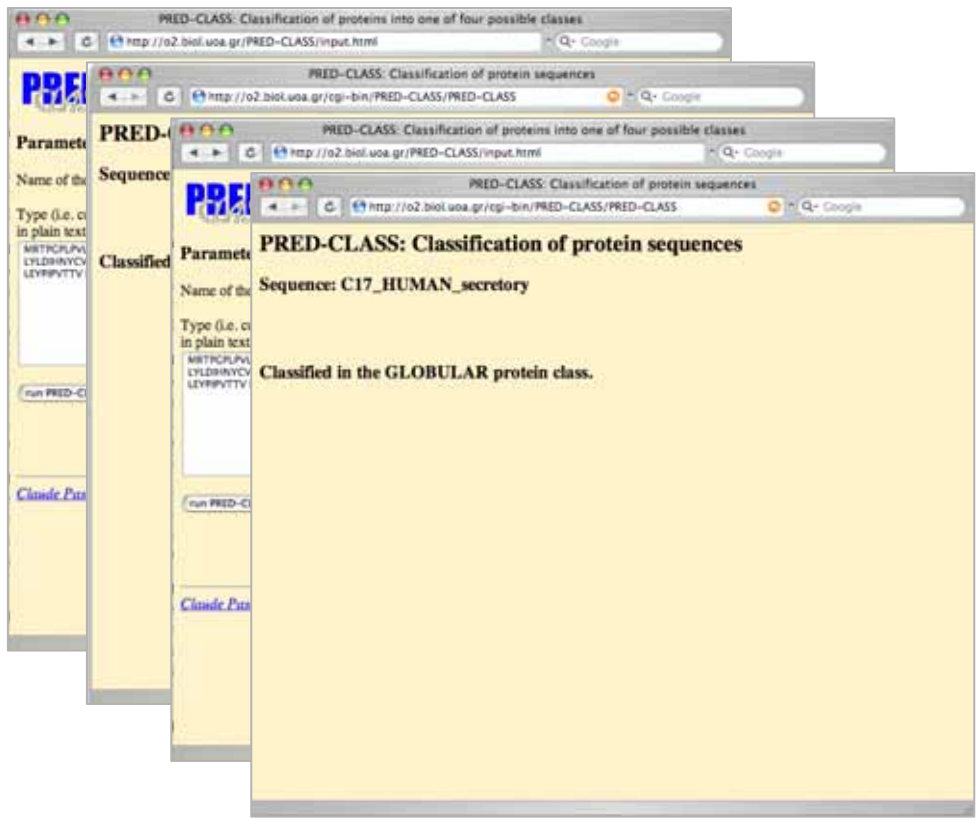


Fig. 1. PRED-CLASS architecture: individual component NNs and their layered structure. The input type for each subsystem, network connectivities, information flow, and decision scheme for the output layer of each NN are indicated.

TABLE II. System Performance on a Test Set of 387 Protein Sequences

	Predicted					
	TM	FIBR	GLOB	MIX	Total observed	SEL (%)
Observed						
TM	139	0	8	0	147	94.5
FIBR	1	72	0	0	73	98.6
GLOB	0	1	54	0	55	98.2
MIX	3	3	0	106	112	94.6
Total predicted	143	76	62	106	387	
SENS (%)	97.2	94.7	87.1	100.0		



予測プログラムの評価

BioInformatics
17, 646－653
2001

Method	TP	FP	FN
TMHMM2.0	812	65	38
TMHMM1.0	818	63	45
TMHMM-Retain	811	70	38
MEMSAT1.5	772	110	78
Eisenberg	809	72	163
KKD	719	164	72
KD5	773	139	125
TMAP	675	191	82
DAS	829	38	243
HMMTOP	639	243	65
SOSUI	686	192	137
KD9	494	391	25
TMpred	525	357	80
ALOM 2	429	545	17
PHD	564	319	207
Toppred 2	468	417	123

Total 883

膜貫通ヘリックス, トポロジー予測法 (1)

Prediction method (reference)	Feature (URL)
KKD (Klein *et al*., 1985)	Kyte and Doolittle hydropathy scale; discriminant function
ALOM 2 (Nakai and Kanehisa, 1992)	filtering hydrophobic scale; discriminant function (http://psort.nibb.ac.jp/form.html)
TMpred (Hofmann and Stoffel, 1993)	statistical preferences of transmembrane segment (http://www.ch.embnet.org/software/TMPRED_form.html)
TopPred II (Claros and von Heijne, 1994)	GES hydropathy scale; positive-inside rule (http://bioweb.pasteur.fr/seqanal/interfaces/toppred.html)
HTP (Fariselli *et al*., 1996)	artificial NN
PHDhtm (Rost *et al*., 1996)	artificial NN; homology search (http://maple.bioc.columbia.edu/predictprotein/submit_def.html#top)
DAS (Cserzö *et al*., 1997)	dense alignment surface; RReM scoring matrix (http://www.sbc.su.se/~miklos/DAS/)
TMAP (Persson and Argos, 1997)	multiple alignment-based (http://www.mbb.ki.se/tmap/index.html)
SOSUI (Hirokawa *et al*., 1998)	Kyte and Doolittle hydropathy- and amphiphilicity-based (http://sosui.proteome.bio.tuat.ac.jp/sosui_submit.html)
TSEG (Kihara *et al*., 1998)	Mahalanobis distance with the average hydrophobicity and the periodicity of hydrophobicity (http://www.genome.ad.jp/SIT/tsegdir/tseg_exe.html)

膜貫通ヘリックス, トポロジー予測法 (2)

Prediction method (reference)	Feature
MEMSAT 2 (Jones, 1998)	dynamic-programming-based (http://bioinf.cs.ucl.ac.uk/psiform.html)
PRED-TMR (Pasquier *et al*., 1999)	propensity of optimized hydropathy (http://o2.db.uoa.gr/PRED-TMR/input.html)
TMHMM 2.0 (Krogh *et al*., 2001)	HMM (http://www.cbs.dtu.dk/services/TMHMM-2.0/)
TM Finder (Deber *et al*., 2001)	combination of hydrophobicity and nonpolar phase helical propensity scales (http://www.bioinformatics-canada.org/TM/login.html)
MPEx (Jayasinghe *et al*., 2001)	Wimley-White hydropathy scale (http://blanco.biomol.uci.edu/mpex)
HMMTOP 2.0 (Tusnády and Simon, 2001)	HMM (http://www.enzim.hu/hmmtop/html/submit.html)
DAS-TMfilter (Cserzö *et al*., 2002; 2004)	comparison between transmembrane segments in a library of documented proteins (http://www.enzim.hu/DAS/DAS.html)
THUMBUP (Zhou and Zhou, 2003)	mean burial propensity and HMM (http://www.smbs.buffalo.edu/phys_bio/service.htm)
ENSEMBLE (Martelli *et al*., 2003)	combination of cascading artificial NN and HMM (http://www.biocomp.unibo.it)

膜貫通トポロジー予測法性能評価
(Ikeda *et al.*, 2002; 2003)

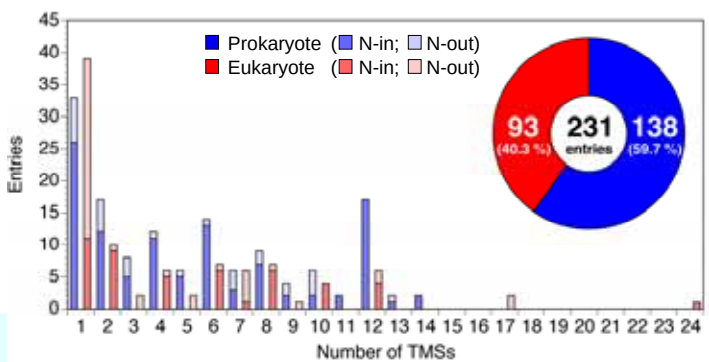
1,074 articles

↓ experimentally-determined data only

TMPDB (302 entries)

↓ α-helical TM proteins < 30% sequence similarities

TMPDB_alpha_ non-redundant (231 entries)

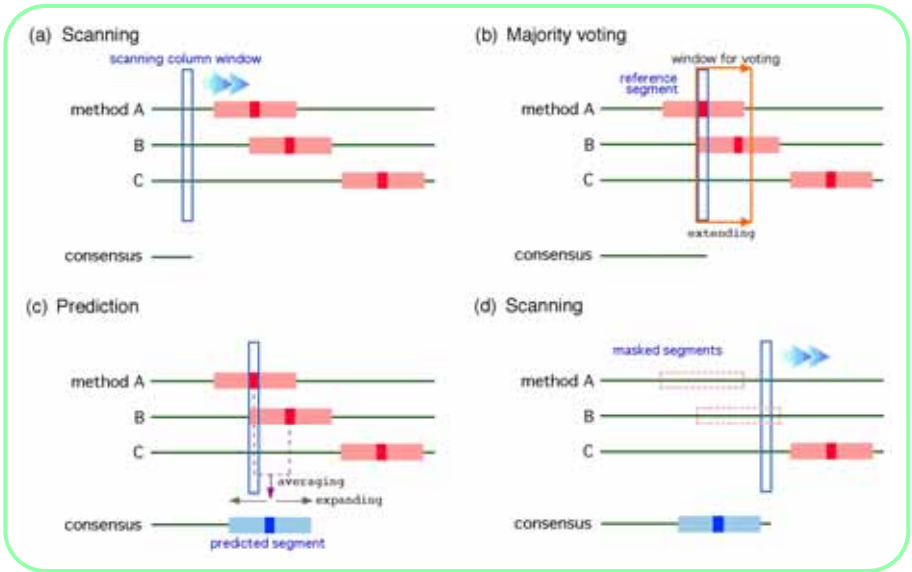


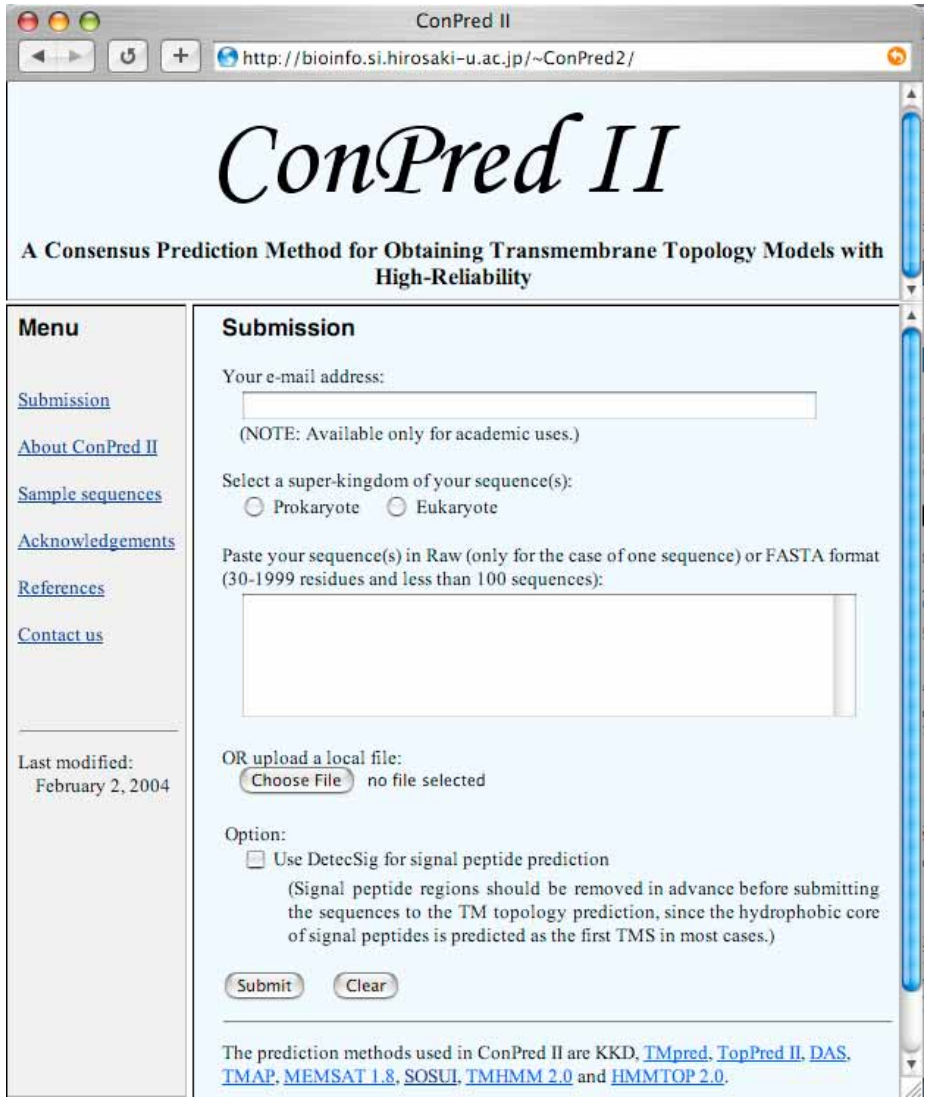
	Prediction accuracy (%)			
	#TMS	#TMS&position	N-tail location	TM topology
Prokaryote				
KKD	60.1	55.1	-	-
TMpred	56.5	50.7	61.6	36.2
TopPred II	56.5	47.1	73.9	38.4
DAS	41.3	34.8	-	-
TMAP	52.9	45.7	57.2	29.0
MEMSAT 1.8	69.6	65.2	84.1	56.5
SOSUI	65.2	59.4	-	-
PRED-TMR2	52.9	50.0	76.8	44.2
TMHMM 2.0	65.2	60.9	73.9	53.6
HMMTOP 2.0	69.6	63.8	79.7	56.5
Eukaryote				
KKD	54.8	49.5	-	-
TMpred	59.1	53.8	64.5	35.5
TopPred II	51.6	48.4	65.6	36.6
DAS	31.2	29.0	-	-
TMAP	59.1	52.7	47.3	26.9
MEMSAT 1.8	57.0	54.8	63.4	39.8
SOSUI	57.0	53.8	-	-
PRED-TMR2	55.9	50.5	58.1	33.3
TMHMM 2.0	59.1	58.1	75.3	46.2
HMMTOP 2.0	68.8	64.5	72.0	51.6

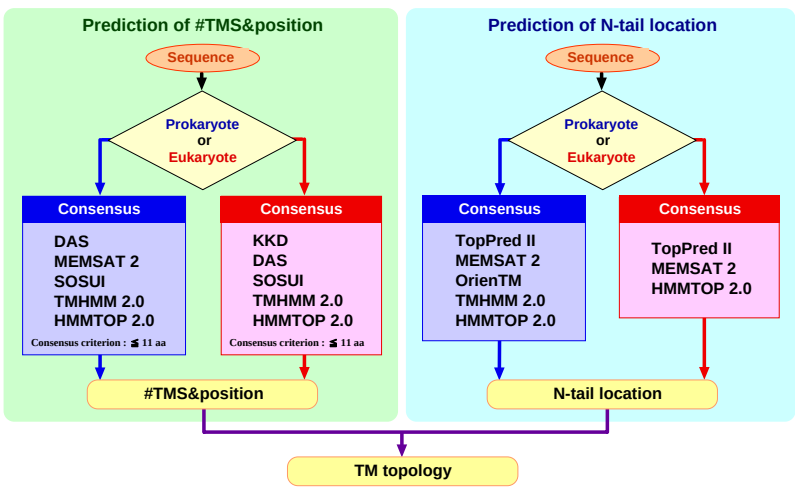
参照文献: Ikeda, M., Arai, M., Lao, D. M. and Shimizu, T. (2002) Transmembrane topology prediction methods: a re-assessment and improvement by a consensus method using a dataset of experimentally-characterized transmembrane topologies. *In Silico Biol.*, **2** (1), 19-33.

予測法の組み合わせによる精度向上

— ConPred II (http://bioinfo.si.hirosaki-u.ac.jp/~ConPred2/) —







参照文献: Ikeda *et al*. (2002) *In Silico Biol*., **2** (1), 19-33; Ikeda *et al*. (2003) *Nucleic Acids Res*., **31** (1), 406-409.

膜貫通トポロジー予測法性能評価 (Möller *et al*., 2001)

予測法名 (参照文献)	ヘリックス本数･位置	ヘリックス本数･位置･膜貫通方向
TMHMM-Retrain	**69%**	**54%**
TMHMM 2.0 (Krogh *et al*., 2001)	**68%**	**47%**
TMHMM 1.0 (Sonnhammer *et al*., 1998)	**67%**	**48%**
HMMTOP (Tusnády and Simon, 1998)	**55%**	**45%**
MEMSAT 1.5 (Jones *et al*., 1994)	**53%**	**41%**
KKD (Klein *et al*., 1995)	**45%**	**n/a**
TMAP (Persson and Argos, 1997)	**43%**	**11%**
Eisenberg (Eisenberg *et al*., 1982)	**38%**	**n/a**
DAS (Cserzö *et al*., 1997)	**37%**	**n/a**
TMpred (Hofmann and Stoffel, 1993)	**37%**	**6%**
SOSUI (Hirokawa *et al*., 1998)	**36%**	**n/a**
KD5 (Kyte and Doolittle, 1982)	**32%**	**n/a**
KD9	**26%**	**n/a**
PHDhtm (Rost *et al*., 1996)	**26%**	**18%**
TopPred II (Claros and von Heijne, 1994)	**26%**	**12%**
ALOM 2 (Nakai and Kanehisa, 1992)	**7%**	**n/a**

参照文献: Möller, S., Croning, M. D. and Apweiler, R. (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*, **17** (7), 646-653.

膜貫通トポロジー予測法性能評価
(Jayasinghe *et al.*, 2001)

Table 1. *General characteristics of the MPtopo database*

	MPtopo subset		
	3D_helix	1D_helix	3D_other
No. of proteins[a]	41	38	11
No. of total residues	8960	15018	4171
Average sequence length[b]	218	395	379
No. of residues in TM segments	4186	5426	1671
No. of total TM segments	150	242	142
Average TM segment length[b]	28 ± 5	22 ± 4	12 ± 3
TM segment length range[b]	17 – 43	9 – 46	4 – 20

[a] Includes protein subunits.
[b] Given as the number of residues.

Table 2. *Prediction accuracy of various algorithms using MPtopo*

		No. of transmembrane helices[a]		
MPtopo subset	Algorithm	$N_{predicted}$	$N_{correct}$	Q (%)[b]
3D_helix (N_{known} = 150)				
	PHDhtm	152	146	97
	HMM	154	145	95
	TopPred II	162	148	95
	TMAP[f]	139	136	96
1D_helix (N_{known} = 242)				
	PHDhtm	250	228	93
	HMM	264	240	95
	TopPred II	259	224	89
	TMAP	241	221	92

[a] N_{known}, $N_{predicted}$, $N_{correct}$ are, respectively, number of experimentally known helices, total number of predicted, and number predicted correctly. $N_{correct}$ is defined as predicted helices that exhibited at least a 50% overlap with known transmembrane helices.
[b] Prediction accuracy Q was determined as described in Tusnády and Simon (1998).

$$Q = 100 \sqrt{\frac{N_{correct}}{N_{known}} \frac{N_{correct}}{N_{pred}}}.$$

参照文献: Jayasinghe, S., Hristova, K. and White, S. (2001)
MPtopo: A database of membrane protein topology. *Protein Sci.*, **10** (2), 455-458.