PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Evaluation of transmembrane helix predictions in 2014

Jonas Reeb,[1] Edda Kloppmann,[1,2]* Michael Bernhofer,[1] and Burkhard Rost[1,2,3,4]

[1] Department of Informatics & Center for Bioinformatics & Computational Biology—i12, Technische Universität München (TUM), Garching/Munich 85748, Germany
[2] New York Consortium on Membrane Protein Structure (NYCOMPS), New York Structural Biology Center, New York, New York 10027
[3] Institute of Advanced Study (TUM-IAS), Garching/Munich 85748, Germany
[4] Institute for Food and Plant Sciences WZW—Weihenstephan, Freising, Germany

ABSTRACT

Experimental structure determination continues to be challenging for membrane proteins. Computational prediction methods are therefore needed and widely used to supplement experimental data. Here, we re-examined the state of the art in transmembrane helix prediction based on a nonredundant dataset with 190 high-resolution structures. Analyzing 12 widely-used and well-known methods using a stringent performance measure, we largely confirmed the expected high level of performance. On the other hand, all methods performed worse for proteins that could not have been used for development. A few results stood out: First, all methods predicted proteins in eukaryotes better than those in bacteria. Second, methods worked less well for proteins with many transmembrane helices. Third, most methods correctly discriminated between soluble and transmembrane proteins. However, several older methods often mistook signal peptides for transmembrane helices. Some newer methods have overcome this shortcoming. In our hands, PolyPhobius and MEMSAT-SVM outperformed other methods.

Proteins 2015; 83:473–484.

Key words: membrane protein; transmembrane helices; transmembrane helix; α-helical membrane protein; transmembrane helix prediction; evaluation.

INTRODUCTION

Transmembrane proteins

Protein complexes that are embedded into the membrane carry out essential processes such as transport, signaling, or adhesion. Transmembrane proteins (TMPs) are involved in many diseases, including cancer, diabetes, or cystic fibrosis. For example, G protein-coupled receptors (GPCRs) are essential for cell signaling[1] and constitute one of the largest families of TMPs in eukaryotes, with almost 800 genes in human.[2] To highlight their relevance: two recent Nobel Prizes were awarded for studies on GPCRs and ∼30% of today's drugs target GPCRs.[3,4]

TMPs are assumed to pass the membrane either exclusively with transmembrane alpha-helices (TMHs) or with transmembrane beta-strands by forming beta-barrels. Beta-barrels have so far been found in gram-negative and acid-fast bacteria, as well as in chloroplasts and mitochondria.[5] Here, we focus on alpha-helical TMPs, the most abundant class of TMPs, assumed to constitute 20–30% of the proteome of any organism.[6–9]

Additional Supporting Information may be found in the online version of this article.

Abbreviations: DBN, dynamic Bayesian network; GPCR, G protein-coupled receptor; HMM, hidden Markov model; OPM, Orientations of Proteins in Membranes; PDB, Protein Data Bank; PDBTM, Protein Data Bank of Transmembrane Proteins; NN, (artificial) neural network; SVM, support vector machine; TMH, transmembrane alpha-helix; TMP, (alpha-helical) transmembrane protein.

Grant sponsor: Alexander von Humboldt foundation through the German Federal Ministry for Education and Research (BMBF); Grant sponsor: New York Consortium on Membrane Protein Structure (NYCOMPS) from the Protein Structure Initiative (PSI) of the National Institutes of Health (NIH); Grant number: U54 GM095315.

*Correspondence to: Edda Kloppmann, Department of Informatics & Center for Bioinformatics & Computational Biology—i12, Technische Universität München (TUM), Boltzmannstr. 3, Garching/Munich, Germany 85748. E-mail: edda.kloppmann@mytum.de

Received 5 September 2014; Revised 2 December 2014; Accepted 13 December 2014
Published online 26 December 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24749

Table I
Transmembrane Helix Prediction Methods

Name	Year	Method	Evolutionary information	Signal peptides	Topology
TopPred2	1994	—	No	No	Yes
PHDhtm	1995	NN	Yes	No	Yes
HMMTOP2	2001	HMM	No	No	Yes
TMHMM2	2001	HMM	No	No	Yes
SOSUI	2002	—	No	No	No
Phobius	2004	HMM	No	Yes	Yes
PolyPhobius	2005	HMM	Yes	Yes	Yes
MEMSAT3	2007	NN	Yes	Yes	Yes
Philius	2008	DBN	No	Yes	Yes
SCAMPI	2008	HMM	No	No	Yes
SPOCTOPUS	2008	NN+HMM	Yes	Yes	Yes
MEMSAT-SVM	2009	SVM	Yes	Yes	Yes

Listed are all TMH prediction methods evaluated in chronological order. We chose 12 methods based on popularity and availability for an overview of developments in TMH prediction during the last 20 years. Given is the name of the method, year of publication, and the machine learning approach used for the prediction (NN: neural network, HMM: Hidden Markov model, SVM: Support vector machine, DBN: dynamic Bayesian network). We indicate whether evolutionary information from multiple sequence alignments is used, as well as, whether signal peptides and topology (inside/outside location of non-TMH regions with respect to membrane) are predicted.

Experimental determination of high-resolution structures for TMPs continues to be challenging.[10,11] Therefore, these proteins are substantially underrepresented in the Protein Data Bank (PDB[12]): fewer than 2% of the PDB structures are TMPs.[11,13] Even when the experimental structures are known, the location of the membrane segments remains to be estimated by algorithms such as the ones used by OPM (Orientation of Proteins in the Membrane[14]) and PDBTM (Protein Data Bank of Transmembrane Proteins[13]) or by curated annotations such as in MPtopo (Membrane Protein Topology Database[15]).

The gap between what we want to know about TMPs (given their biomedical importance) and what we do know in terms of experimental structures makes prediction methods particularly important. Although the three-dimensional (3D) structure can be predicted *de novo*[16,17] or through comparative modeling,[18] for the majority of TMPs we have yet to be content with the prediction of secondary structure and topology.[19] Reliable methods for the prediction of secondary structure in nonmembrane proteins[20–23] also help in the annotation and understanding of TMPs. However, these predictors do not distinguish between alpha-helices within or outside the membrane. For TMPs, knowing the precise location of transmembrane regions is of high importance to understand protein function and to design experiments. Several independent evaluations revealed TMH predictions to be rather successful.[24–30]

A recent analysis focused on evaluating correctly predicted topology.[28] We left this perspective out, both to reduce publication redundancy and because obtaining reliable, nonambiguous topology data for our dataset is not trivial. Here, we complemented previous assessments by tapping into today's wealth of high-resolution data. We assembled distinct subsets to highlight different aspects of performance: (i) new *vs.* old compares performance for TMPs unknown and known during development, (ii) eukaryotic *vs.* bacterial allows to spot origin-specific aspects, and (iii) sets of proteins with different numbers of TMHs. We considered TMH annotations from two different databases (OPM and PDBTM) to prevent algorithmic bias. In addition, we applied a more stringent criterion for considering a TMH to be correctly predicted. Finally, we also used a set of soluble proteins and proteins (TMPs as well as soluble) with signal peptides.

Over 30 advanced methods predict TMHs; some have been used for over two decades.[31] We focused on the following 12 (Table I): HMMTOP2,[32] MEMSAT3,[33] MEMSAT-SVM,[34] PHDhtm,[35] Philius,[36] Phobius,[37] PolyPhobius,[38] SCAMPI,[39] SOSUI,[40] SPOCTOPUS,[41] TMHMM2,[42] and TopPred2.[43]

MATERIALS AND METHODS

TMPs from the PDB

A major challenge for the development and evaluation of transmembrane helix (TMH) prediction methods is the extraction of comprehensive and accurate datasets. In the past, such datasets relied on TMH annotations from biochemical experiments. In 2000, Möller *et al.* assembled a widely used dataset[44] that contained carefully extracted information from biochemical experiments, which was later referred to as low-resolution information—in contrast to high-resolution information from, for example, X-ray crystallography.[24] Surprisingly, such low-resolution biochemical data turned out to not be significantly more accurate than prediction methods.[24,27,34] Therefore, more recent evaluations focus on high-resolution structures.[24,34,41]

We used 462 TMPs (note: here exclusively denoting proteins with TMHs) common to three databases

Table II
Discrimination Between TMPs and Soluble Proteins As Well As TMHs and Signal Peptides

	no SPs			SPs			
	Sol			Sol			TMP
	Euk (5106) FPR	Gram− (356) FPR	Gram+ (911) FPR	Euk (1297) FPR	Gram− (400) FPR	Gram+ (204) FPR	Euk (332) Sens
TopPred2[a]	62	53	52	97	99.8	98	100
PHDhtm[a]	11	11	15	31	17	39	97
HMMTOP2[a]	21	17	15	56	77	89	96
TMHMM2[a]	1	1	1	20	25	63	97
SOSUI[a]	3	1	1	62	32	44	97
Phobius	3	1	1	4	2	13	99
PolyPhobius	6	3	2	5	3	12	97
MEMSAT3	8	4	3	60	47	67	100
Philius	2	1	1	3	1	12	93
SCAMPI[a]	30	27	21	92	95	97	100
SPOCTOPUS	11	24	25	13	3	13	97
MEMSAT-SVM	6	5	6	25	6	24	99

[a]Methods predict only TMHs, no option to predict SPs.

Shown are discrimination rates between soluble proteins (Sol) and TMPs and between signal peptides (SP) and TMHs. The datasets are assembled from SignalP4 training sets[49]: (i) three sets of soluble proteins without SPs: Sol eukaryotic, Gram-positive and Gram-negative bacteria, (ii) soluble proteins with SP (Sol(SP)) and (iii) a set of eukaryotic TMPs with SPs. Set sizes are given in the table in brackets. For soluble proteins, we consider as false positives all proteins with a predicted TMH and denote the false positive rate (FPR). For the TMP(SP) set the sensitivity is given. Rates and the sensitivity are given in percent (%) of all proteins of the respective set. Percentages are rounded to integers, except for values >99.5. For example PolyPhobius misclassifies 62 of 1297 proteins in the eukaryotic Sol(SP) set, that is, FPR = 5%.

(retrieved on September 2013): PDB,[12] OPM,[14] and PDBTM.[13] These 462 proteins have 1101 PDB chains. We mapped the TMH annotation based on the PDB ATOM record to the UniProt sequence using SIFTS.[45,46] The 1101 PDB chains do not contain chimeras or structure models and have a unique residue-level mapping between UniProt and ATOM record. For every sequence, all residues in an annotated TMH can be completely mapped. The resolution of the respective structures is at least 9 Å (average 2.9 Å).

These 1101 sequences were redundancy reduced at HVAL>0 using UniqueProt.[47] This implied that no pair of proteins had >20% pairwise sequence identity over alignment lengths of >250 residues. We weighted sequences to prefer smaller fractions of unmapped residues and longer sequences. We excluded the PDB chain 3n23:G (UniProt AC Q58K79) describing the Na+/K+-ATPase gamma subunit (annotated as TMP in PDBTM), because its pseudo-TMH is buried in a larger protein complex rather than contacting the lipid bilayer. This resulted in an independent test set of 190 TMPs (Supporting Information Table S1). PDBTM and OPM differ for several proteins, for example, in their annotations of the first (N-term) and last residues (C-term) in TMHs. We used both annotations by considering predictions correct if they matched any of the two. Although these databases might contain mistakes, we nevertheless prefer to use the systematic bias from consistent automated annotations as opposed to fully manual annotations which introduce a different set of challenges. As an exception, we manually edited annotations in six OPM entries and in one PDBTM entry, related to kinked and re-entrant helices as

Table III
Evaluation Scores

Name	Formula	Description
$Q_{\text{tmh}}^{\%\text{obs}}$	number of correctly predicted TMHs/number of TMHs observed	TMH recall
$Q_{\text{tmh}}^{\%\text{pred}}$	number of correctly predicted TMHs/number of TMHs predicted	TMH precision
Qok (%)	$\frac{100}{N_{\text{prot}}} \times \sum_{i=0}^{N_{\text{prot}}} \delta_i; \delta_i = \begin{cases} 1, & if\ Q_{\text{tmh}}^{\%\text{obs}} = 1 = Q_{\text{tmh}}^{\%\text{pred}} \\ 0, & else \end{cases}$	Percentage of proteins for which all TMH positions are correctly predicted
FPR (%)	100 false positives/(false positives + true negatives)	Percentage of false positives among negatives
Sens (%)	100 true positives/(true positives + false negatives)	Percentage of detected positives

Per-segment recall ($Q_{\text{tmh}}^{\%\text{obs}}$) and precision ($Q_{\text{tmh}}^{\%\text{pred}}$) are pooled for all TMHs of all proteins in the dataset and then averaged once. Qok, that combines recall and precision, is calculated for every protein.[24] False positive rate (FPR) and Sens(itivity) are employed for the scoring of discrimination between soluble proteins and TMPs.

detailed in Supporting Information Table S2. In addition to the TMH annotations from OPM and PDBTM, we used alpha-helix (H) annotations from DSSP (CMBI version, April 2010).[48] First, we extended the TMHs from OPM and PDBTM, to the full length of the DSSP-assigned alpha-helix, thereby ignoring membrane boundaries. Second, we exclusively used alpha-helix annotations from DSSP, ignoring OPM and PDBTM.

TMP subsets

We separated the 190 TMPs into three subsets. Subset 1, new *vs.* old: We BLASTed (*e*-value < 0.1) our TMP dataset against all datasets used for the development of the methods we analyzed. The sets for HMMTOP2 and PHDhtm had to be recreated according to the identifiers listed in the original publications. For HMMTOP2 UniProt ID tcr1_ecoli and for PHDhtm UniProt ID a1aa_human could not be mapped unambiguously and were excluded. In this way, we obtained a subset with 146 old TMPs (used for development) and 44 new TMPs (not used for development). Subset 2, eukaryotic *vs.* bacterial: We separately analyzed performance for 75 eukaryotic and 104 bacterial TMPs. The remaining 11 TMPs were viral (2) and archaeal (9), too few to support further separation. Subset 3, grouping by number of TMHs: We separated the TMPs by the number of TMHs annotated by OPM and PDBTM into three sets with 1 TMH, 2–5, and >5 TMHs. For this, five entries (2lp1:A, 3t9n:E, 1jb0:F, 4hzu:S, and 4i9w:B) had to be excluded, since they would fall into different bins due to differing numbers of annotated TMHs (Supporting Information Table S1).

Human proteome

The complete human proteome with 20,258 sequences was retrieved from UniProtKB/Swiss-Prot (release 2014_04). Predictions were performed with PolyPhobius using the same settings as for the all other predictions. A single entry (UniProt AC Q8WZ42, Titin) had to be excluded because it was too long for that method. All statistics mentioned in Results and Discussion for the comparison against our dataset, are calculated for 6016 proteins that are predicted to be TMPs.

Soluble non-TM proteins

Sets of soluble proteins without signal peptides were assembled from the homology reduced SignalP4 data.[49] We excluded 12 proteins that could not be handled by all methods. Finally, we used three sets: a eukaryotic soluble set with 5106 proteins, a set from Gram-negative bacteria with 356 proteins, and another from Gram-positive bacteria with 911 proteins. The SignalP4 dataset has no entries from Archaea or viruses.

For an additional comparison on the human proteome, the complete proteome of 20,258 sequences was used. Sequences were subjected to prediction with SignalP4.1, which resulted in the following estimates: 14,149 sequences soluble without signal peptide, 3220 soluble with signal peptide, and 2889 TMPs. Error rates from Table II were then applied to these sets. For example, Phobius has error rates of 3%, 4%, and 1% (99% sensitivity) for the previous cases on eukaryotic proteins. Applying these to the estimates on the proteome results in $14{,}149 \times 0.03 = 424$ errors on soluble proteins without signal peptides, 129 on soluble with signal peptides and 29 on TMPs. Overall, this suggests that approximately 2.9% of the human proteome are predicted wrongly.

Datasets with signal peptides

Signal peptides are typically 15–30 residues long and mediate specific targeting of proteins to different cellular compartments.[50,51] The central section of a signal peptide is commonly highly hydrophobic, thereby resembling TMHs. Some signal peptides are, indeed, not cleaved off after reaching the target membrane but rather are embedded as a TMH.[52] Many TMH prediction methods predict signal peptides along with TMHs (Table I). We evaluated confusions between signal peptides and TMHs using the following sets of soluble proteins and TMPs with signal peptides.[49]

Four signal peptide datasets were built from the SignalP4 data. Soluble proteins and TMPs containing a signal peptide were collected and then split into eukaryotic, Gram-negative, and Gram-positive entries. We excluded 10 proteins that could not be handled by all methods. The final sets of soluble proteins with signal peptides included 297 eukaryotic, 400 Gram-negative, and 204 Gram-positive proteins. The TMP dataset with signal peptides contained 332 eukaryotic TMPs. There were too few bacterial TMPs with signal peptides (22 Gram-negative, 3 Gram-positive) to compile performance separately.

Evaluation measures

We assessed performance through the standard measures: We largely focused on the per-protein score Qok[24] (Table III). For this, a TMH counted as correctly predicted, when both TMH endpoints differed less than five residues between observation and prediction (Fig. 1A). In addition, any observed TMH matched maximally one predicted TMH and *vice versa.* This constraint implicitly handled the correct prediction of the number and placement of all TMHs simultaneously. We further calculated performance based only on OPM or PDBTM annotations and found performance to be higher and less deviating for OPM (mean Qok 43 ± 3) than for PDBTM (mean Qok 36 ± 7; data not shown). The above constraints, in particular the required overlap, yielded significantly more

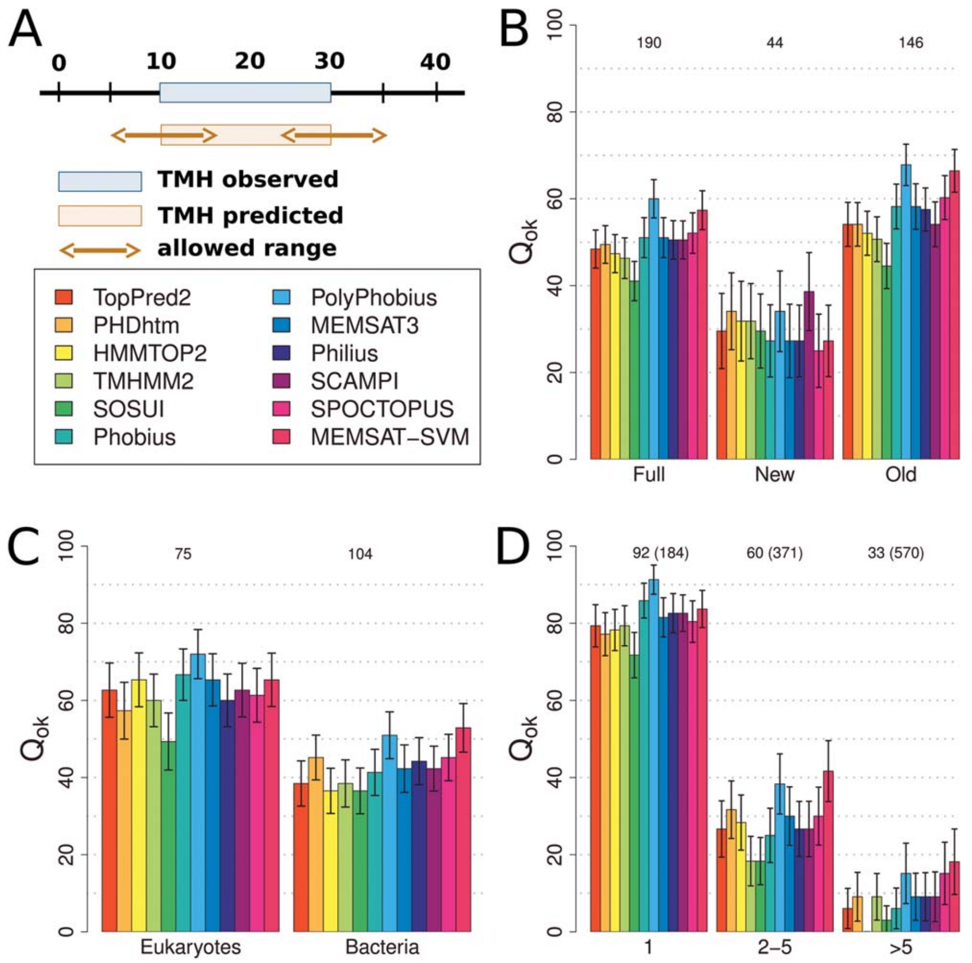

Figure 1

Transmembrane helix prediction performance. Qok scores for all 12 prediction methods on various sets of TMPs. Qok denotes the percentage of proteins for which all TMHs were correctly predicted (**A**: TMH endpoints within five or less residues of either OPM or PDBTM annotation for the whole protein, Methods). Above the bars are the numbers of proteins in each dataset. Error bars are the sample standard deviation generated by bootstrapping with 1000 draws of half the set size each (*cf.* Methods). **B**: Qok is plotted for 190 redundancy-reduced TMPs followed by 44 new (not used for development) and 146 old (used for development, either the protein itself or homologous proteins) TMPs. All methods clearly performed worse for more recently determined protein structures. The old–new difference for TopPred2 suggested that a significant fraction of the differences might not be explained by over-training. **C**: All methods reached higher Qoks for eukaryotes than for bacteria. Note that we excluded the nine archaeal and two sequences of viral origin. **D**: Performance declines from bitopic TMPs to those with 2–5 TMHs or more. For (**D**), the number in brackets behind the set size denotes the number of TMHs in the respective subset.

conservative estimates than those provided in previous evaluations.

Errors were estimated through bootstrapping.[53] As the standard version with putting the data back "into the pot" tends to underestimate the scientific error, we used a version in which we randomly drew 1000 times half the dataset size (for example, 1000 sets with 95 TMPs each for assessing the error for the full dataset). The sample standard deviation was calculated as square root of the sample variance over the 1000 draws.

For the sets of soluble proteins, a prediction was considered as incorrect (FP: false positive), if any TMH was predicted. The calculation of false positive rate (FPR), and sensitivity for the TMP datasets, is defined in Table III. For the sets with signal peptides, we additionally checked if an incorrectly predicted TMH overlapped with an annotated signal peptide (FPR_SP). For this score, we tested 6 overlaps between 1 to 20 residues. In the reported results, we considered an overlap of predicted TMH and observed signal peptide by more than 7 residues as incorrect.

Transmembrane helix prediction methods

The 12 prediction methods were chosen according to various criteria. We sampled some methods of historical importance for comparison and focused on the more recent methods that were readily available and had been claimed to perform at the top. The contenders were as follows.

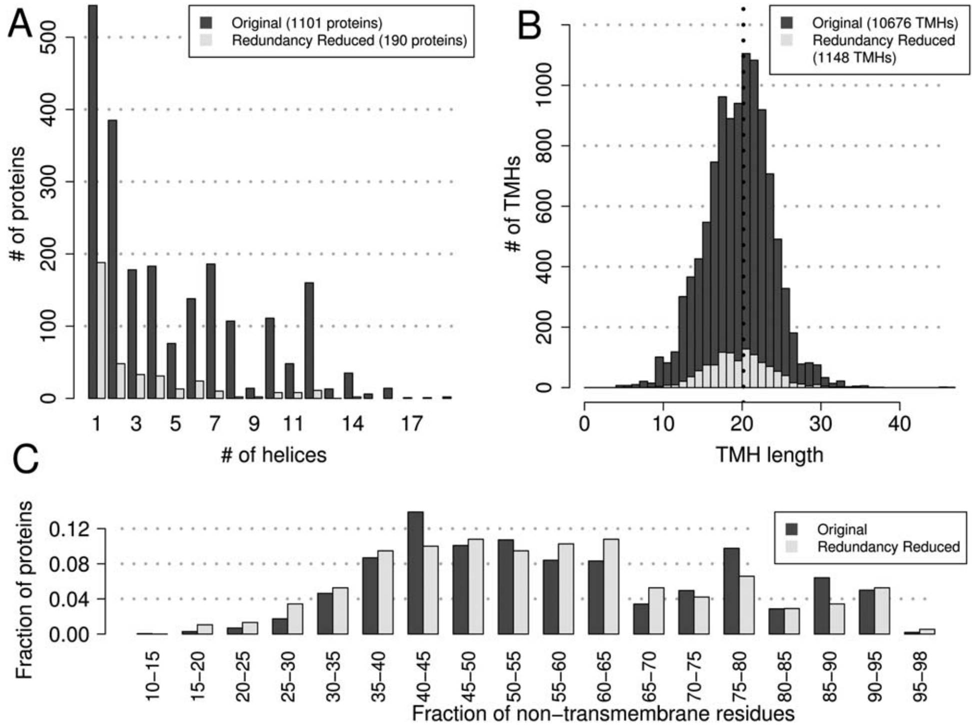

Figure 2

Transmembrane helix statistics. The plots show statistics for the original set of 1101 TMPs (plotted as dark grey bars in **A–C**) as well as the redundancy reduced set of 190 TMPs (plotted as light grey bars in **A–C**). We consider annotations from OPM and PDBTM for the evaluation. Therefore, the statistics shown are based on both annotations as well. **A**: Distribution of the number of TMHs per protein. **B**: Distribution of TMH lengths, where the 1101 proteins in the original set contain 10,676 TMHs and the 190 proteins in the redundancy contain 1148 TMHs. The average TMH length of 20 residues is indicated by the dotted line. **C**: Fraction of non-TM residues per protein. Statistics for the eukaryotic, bacterial, new, and old subsets can be found in Supporting Information Figures S1–S4.

No machine learning

TopPred2[43] uses hydrophobicity scales through a relatively simple sliding window approach and chooses the final topology based on the positive-inside rule.[19] *SOSUI*[40] uses the Kyte–Doolittle hydrophobicity scale.[54] Additionally, SOSUI employs an amphiphilicity index to account for weak and strong polar residues at the helix ends.[55]

Neural networks (NN)

PHDhtm[35,56] was one of the first TMH prediction methods based on machine learning. It also was the first to improve performance through the use of evolutionary information from multiple sequence alignments. *MEMSAT3* expanded this by also predicting signal peptides.[33] MEMSAT3 applies an external filter to distinguish between globular and membrane proteins. The most recent NN-based method in our evaluation was *SPOCTOPUS.*[41] SPOCTOPUS which further develops OCTOPUS[57] consists of several NNs for different residue types, one of which models re-entrant helices. It contains an additional NN with an adjunct hidden Markov model to improve discrimination between signal peptides and TMHs.

Hidden Markov models (HMM)

HMMTOP2[32] was the second TMH prediction method based on HMMs (TMHMM was the first); it relies on the divergence in amino acid distribution between structural components of the protein. *TMHMM2*[42] is based on a cyclic HMM with modules for a helix core, capping regions and loops on either side of a transmembrane helix as well as a single state for globular sequence parts, that is, loops longer than 20 residues. A major limitation that was already pointed out by the authors in the original publication is the high FPR due to the confusion of signal peptides and TMHs. This shortcoming was tackled by *Phobius*[37] which essentially combines the HMMs of TMHMM and the signal peptide model of SignalP-HMM.[58] *PolyPhobius*[38] additionally incorporates homology information through a new HMM decoder. *SCAMPI*[39] was developed with the premise that the biological machinery does not know about amino acid distributions and that successful predictions should be possible from physical principles. In contrast to the other machine learning-based methods that optimize many parameters, SCAMPI, although technically an HMM, only optimizes two parameters and bases the prediction on the estimated free energy of membrane insertion, given a segment of 21 residues and based on a previously experimentally developed propensity scale.

Bayesian networks

Dynamic Bayesian networks can be considered generalizations of HMMs. We evaluated one method based on this concept, namely, *Philius*[36] which uses a state transition diagram equal to that employed by Phobius.

Support vector machines (SVMs)

MEMSAT-SVM, the most recent method evaluated here, consists of five separate SVMs and incorporates homology information as well as discrimination between TMHs and signal peptides. It also treats re-entrant helices separately.

For all methods, we used the default parameters and ran the methods on our machines where stand-alone versions were available. For all others (SOSUI, Philius), the respective webservers were employed. As proposed by the authors, multiple sequence alignments for PolyPhobius were created using Kalign2.[59] All methods requiring a BLAST database were run on UniProtKB/Swiss-Prot release 2013_09. For tests regarding the effect of the database size, UniProtKB/Swiss-Prot release 2014_04 was used, as well as a database assembled from merging PDB, UniProtKB/Swiss-Prot and TrEMBL of the same release, redundancy reduced with CD-HIT[60] at 98% pairwise sequence identity.

The base performance was estimated by three simple prediction methods: RANDOM first randomly predicted the number of TMHs (according to the TMH distribution in the full dataset). The randomly predicted TMHs were then inserted consecutively, nonoverlapping, with lengths according to a normal distribution approximating helix lengths in the dataset ($\mu = 20.5$, $\sigma = 4.3$). Another random prediction method, *SEMIRANDOM*, started with the correct number of TMHs for every protein, and then placed TMHs at random along the sequence. The third background method, *HYDRO*, used hydrophobicity values from the Eisenberg scale.[61] For a fixed window size of 21, the sum of hydrophobicity values was calculated at each position in the sequence. A helix was predicted for all windows with a sum > 4, beginning from the highest values until all TMHs were placed. Where two helices overlapped, the initial prediction was retained while all subsequent overlapping TMHs were removed.

RESULTS AND DISCUSSION

Assembling a curated dataset of high-resolution TMP structures

To assess the performance of transmembrane helix (TMH) prediction, we assembled a new dataset of TMPs. Our final curated and redundancy reduced dataset contained 190 unique TMPs from the PDB (Supporting Information Table S1, available along with our evaluation protocol through https://rostlab.org/resources/tmh_eval). We annotated the observed TMHs using OPM and PDBTM.[13,14] For 15 cases, the database annotations disagreed on the number of TMHs they annotated (Supporting Information Table S1). For these cases, prediction methods seem to favor the annotations provided by OPM (Supporting Information Table S3).

Our TMP dataset contained 190 proteins with 1–14 TMHs (Fig. 2A). This distribution was similar before and after redundancy reduction. The largest differences were found for TMPs with one TMH (49% after redundancy-reduction, 25% before). The distribution of TMH lengths (mean ~20 residues) also did not change on redundancy reduction (Fig. 2B). The standard deviation in TMH length is slightly larger for structure-derived annotations than for predictions (Supporting Information Table S4). Six TMHs in our dataset were $\leq$10 residues long; five of the six came from OPM. The longest TMH with 38 residues is annotated by OPM, and describes a very angular TMH (PDB ID 3tij:A, TMH 7). About 37% of all residues in TMPs were in TMHs, similar to the percentage before the redundancy-reduction (Fig. 2C). The few cases with <10% TM residues were TMPs with only one or two TMHs. Statistics for the eukaryotic, bacterial, new and old subsets can be found in Supporting Information Figures S1–S4.

Performance estimated through stringent overlap criterion

Measured by two-state per-residue scores, that is, fraction of residues correctly predicted as TMH or non-TMH, many prediction methods appeared to be very good (~80% precision and recall), and also very similar (Supporting Information Table S5 and Figs. S5 and S6). This has been noted before and it is debatable whether experimental annotation is accurate enough to consider these measures meaningful.[27] Therefore, we measured performance through the Qok score, that is, the percentage of proteins for which all TMHs were correctly predicted (Table III).

Typically, TMHs have been considered as correctly predicted when as few as 3–5 of the ~20-residues-long TMH were matched.[24,34,38] This tends to overestimate performance. Since most TMH predictions are indeed rather reliable, we introduced a more stringent criterion: if the start and end points of a predicted TMH differ by at most n residues from the start and end points of an observed TMH, this helix is considered as correctly predicted (schematic in Fig. 1A). Although the method ranking changed only slightly depending on this parameter (we compared values $n = 3$–7), the actual values differed substantially: from average Qok = 28% for $n = 3$ to average Qok = 63% for $n = 7$. For the remainder, we focused on results for $n = 5$. All 12 methods tested reached Qok>40% (average Qok = 50%, Fig. 1B). Random predictions reached Qok = 5%; simple hydrophobicity-based predictions reached Qok = 38% (*cf.* RANDOM and HYDRO in Methods). When the OPM and PDBTM annotations differed, we used whichever yielded the more optimistic performance estimate for each prediction method and each TMP. Assessing OPM annotations like a prediction method against the annotations from PDBTM and *vice versa*, resulted in an overall Qok near the lower-end performing methods (Qok = 44%; data not shown).

Using alpha-helix annotations from DSSP instead of TMH annotations results in a significantly lower performance (Supporting Information Fig. S7: average Qok = 16% compared to 50% for TMH annotations). We also used DSSP helix annotations to extend the TMHs given by OPM and PDBTM to the full length of the respective alpha-helix. This extended 62% of all TMHs on the N-terminus, and 59% on the C-terminus, both by five residues on average. Extended helices result in a drop in performance (Supporting Information Fig. S7: average Qok = 29%). These additional analyses underline the unique features of TMHs and the need for specialized prediction methods. Another result in the same direction was that a state-of-the-art method for the prediction of general secondary structure in proteins performed worse by all three measures (Supporting Information Fig. S7), that is, for the TMH annotation (ΔQok = 43 percentage points (*pp*) with respect to the best TMH prediction method), for DSSP elongated helices (ΔQok = 17 *pp*) and for DSSP alpha-helix annotation (ΔQok = 4 *pp*).

Error estimates remain challenging

Our TMP dataset contained only TMPs of known 3D structure. Although the number of TMPs with known 3D structure has grown substantially over the last decades, the resulting dataset is still small.[11] However, experimental TMH annotations from other biochemical assays tend to contain errors on a similar scale as prediction methods.[24] For the final 190 TMPs in our dataset, we estimated errors by bootstrapping.[53] Almost all 12 methods fell within one standard deviation interval of each other given those error estimates. However, this did not imply the differences to be statistically insignificant.[62]

TMH predictions largely successful

TopPred2 was developed more than 20 years ago and still reached Qok = 48%, that is, almost the average performance level. More recent methods tended to perform slightly better (Qok > 50%). However, the differences were often insignificant. Two methods peaked at Qok > 55%: PolyPhobius performed best on the complete dataset with Qok = 60%, followed by MEMSAT-SVM (Qok = 57%). For more than 70% of the proteins all methods correctly predicted the number of TMHs (Supporting Information Table S6 and Fig. S8). PolyPhobius (84%) and SCAMPI (83%) performed best. Methods tended to not detect TMHs more often than to over-predict. Most mistakes were minor: for >93% of the proteins the predicted number of TMHs was either correct or only off by one TMH (Supporting Information Table S6 and Fig. S8).

We applied PolyPhobius to the human proteome to estimate bias in our experimental dataset (Fig. 2) with respect to all proteins in a complete proteome. We observed only one major difference in the number of TMHs: PolyPhobius predicted 16% of the TMPs in the human proteome to have seven TMHs as opposed to only 3% in our dataset (data not shown). The average observed TMH in our dataset was 20 residues long; this was similar to the average of 22 residues for the human proteome predicted by PolyPhobius (*cf.* Supporting Information Table S4). The amount of non-TMH residues was higher for the human proteome: 25% of the proteins had 95–100% non-TMH residues and very few cases were predicted with <35% non-TMH residues.

New proteins predicted less accurately

Prediction performance significantly decreased for new proteins, that is, those that have no homologs in any of the methods' training sets (Fig. 1B). Since all prediction methods optimize some parameters, differences in performance between new (unused for training) and old (used for training) might point at over-training. However, TopPred2 did not really train, and even for TopPred2 we observed a substantial difference in Qok of 25 *pp* (Fig. 1B, Qok(old) = 54% *vs.* Qok(new) = 29%).

The lower performance for new proteins might therefore indicate that newer structures reveal new and complex structural features such as transmembrane segments that are bent or kinked. This view is supported by the performance of our hydrophobicity-based predictor HYDRO (*cf.* Methods) which behaved similar to the other methods (ΔQok(old–new) = 27 *pp*), while the random predictor RANDOM remained unaffected (ΔQok = 3 *pp*). If true, differences significantly larger than 25 *pp* could be explained by over-training which seemed more detrimental for new than for old methods: The average ΔQok(old–new) for the five methods published before 2002 (Table I) was 20 *pp*, while that for the seven newer methods was 30 *pp*. The exception was SCAMPI (ΔQok = 15 *pp*), a new method that optimized only two free parameters.

The subset of new proteins was too small to support clear conclusions. Nevertheless, we observed that the ranking of methods shifted: only SCAMPI reached higher Qok values than the other methods for the new structures. Given the error rates, this difference was within the range of one standard deviation (Fig. 1B), that is, we observed a trend, not a statistically significant difference.

Prediction performance higher for eukaryotic sequences

All methods performed better for eukaryotic than for bacterial proteins (Fig. 1C). This surprising new finding contradicted previous reports.[24,25] We found the smallest difference in Qok for MEMSAT-SVM and PHDhtm (ΔQok = 12 *pp*). As these two represent new and old methods, the improved prediction for eukaryotes seemed unaffected by over-training or novel TMP structures. Overall, PolyPhobius and MEMSAT-SVM stood out amongst the best for eukaryotes and bacteria.

Worse predictions for proteins with many TMHs

Performance tended to decrease for proteins with more TMHs (Fig. 1D). Numerically, all methods reached the highest values for proteins with a single TMH. First, all proteins in our experimental set had TMHs. Second, since most methods get the number of TMHs right most of the time (Supporting Information Fig. S8), the odds are much higher for proteins with a single TMH to be fully correct. At the same time, proteins with more TMHs can contain different signals and therefore pose an additional challenge for prediction. Qok dropped strongly from one TMH to two TMHs (Supporting Information Fig. S9). The same logic applied for the comparison between proteins with 2–5 and those with >5 TMHs. We found it impossible to separate the statistical effect from a possible inherent characteristic of polytopic TMPs. The reason was that different random models gave different answers to the question whether or not proteins with one TMH were predicted better (Supporting Information Figs. S10 and S11).

A similar correlation between the number of TMHs and Qok has been observed before.[24] Possibly, performance decreases because bitopic TMPs are, on average, more hydrophobic than polytopic TMPs.[63] This hypothesis is supported by the methods that arguably rely most on hydrophobicity: TopPred2 and our simple hydrophobicity-based prediction. Their difference in Qok (ΔQok = 47 *pp* and 51 *pp*, respectively) is among the largest between bitopic (Qok(TMHs = 1)) and polytopic (Qok(TMHs = 2–5) + Qok(TMHs > 5)) TMPs. We observed the lowest difference for the newest method (MEMSAT-SVM, ΔQok = 24 *pp*).

Re-entrant helices did not significantly impact our overall results

Recently determined structures shed light on the prevalence of an additional type of membrane helices, namely re-entrant helices that do not cross the membrane but instead enter and exit on the same side. Here, we treated re-entrant regions as TMHs throughout, since only two of the evaluated methods (SPOCTOPUS and MEMSAT-SVM) distinguish re-entrant from transmembrane helices. However, we investigated a subset of 175 proteins without re-entrant helices. For this subset, performance appeared slightly higher (average Qok = 53% compared to 50% for the full set). The ranking of the methods was not significantly affected (data not shown).

Discrimination between soluble proteins and TMPs clearly sets methods apart

Up to this point, our analysis focused on proteins known to contain TMHs, that is, we have provided estimates that hold for about one-fourth of all proteins in most organisms. Do the same methods also correctly recognize the other three-fourths as non-TMP, that is, soluble proteins? Many proteins entering the secretory pathway have signal peptides that resemble TMHs (and are occasionally also embedded as TMHs).[50,51] More recent methods account for this phenomenon (Table I). However, older methods not accounting for signal peptides often confuse these with TMHs (Table I). Therefore, we differentially assessed nonmembrane proteins with and without signal peptides. For proteins without signal peptides, TMHMM2 and Philius appeared best at distinguishing soluble proteins and TMPs (FPR<2%, Table II). For proteins with signal peptides, the FPR for TMHMM2 shot up (>20%) because it does not account for signal peptides. In contrast, Philius still remained below 3% error except for proteins with signal peptides from Gram-positive bacteria for which the error increased to 12% (which was still the best performance, followed by PolyPhobius, SPOCTOPUS, and Phobius; Table II). Phobius, PolyPhobius, and Philius performed best overall for proteins with and without signal peptides. Similar conclusions have been suggested previously.[28]

On the flip side of the coin, Philius had the lowest sensitivity in detecting TMPs, that is, its power in recognizing proteins without TMHs came at the price of missing many TMPs. To put Table II in context of an entire organism, we applied the FPRs and sensitivity for eukaryotic proteins to the complete human proteome (Methods). Based on this estimate, Phobius and Philius gave the best compromise between missing (Table II: sensitivity) and over-predicting TMPs (Table II: FPRs for soluble eukaryotic proteins with and without signal peptides). Both predict 2.9% of the proteome incorrectly as either TMPs or soluble proteins, followed by TMHMM2 (4%) and PolyPhobius (6%). Given the good PolyPhobius prediction for the TMH locations (Fig. 1), and the very fast runtime of Phobius it might be an option to run Phobius first to find TMPs and PolyPhobius afterward to predict where the TMHs are.

Several of the older methods that had access to significantly less data during development than more recent methods were among the top performers in terms of distinction between TMPs and soluble proteins without signal peptides. Namely, SOSUI (from 1998) and TMHMM2 (2001) with average FPRs<2%. On the other hand, very recent methods, using the most up to date datasets, did not necessarily distinguish well between TMPs and soluble proteins, as exemplified by MEMSAT-SVM.

Among the methods that do not account for signal peptides, TopPred2 and SCAMPI stood out as very poor filters: they incorrectly predicted at least one TMH for almost every soluble protein with a signal peptide. In both cases, most misclassifications originated from mistaking a signal peptide for a TMH (Supporting Information Table S7: FPR(SP)). In fact, SCAMPI was developed on a datasets of soluble proteins and TMPs that did not contain signal peptides.[39] As suggested before,[49] MEMSAT3 consistently performed worst among the methods accounting for signal peptides.

For most methods, signal peptides in proteins from Gram-positive bacteria constituted the toughest challenge. This difference might be explained by the fact that signal peptides in Gram-positives are longer and more hydrophobic than those in Gram-negatives.[58] However, many of the mistakes in Gram-positives could not be attributed to confusing signal peptides with TMHs (directly measured by FPR(SP)). In fact, most mistakes (FPR) were made outside the regions of the signal peptides (Supporting Information Table S7: FPR(SP) much smaller than FPR). The differences between eukaryotic and Gram-negative proteins were similar albeit less substantial, yielding the overall tendency for the FPR in proteins with signal peptides: Gram-positives≫eukaryotes>Gram-negatives.

Some methods are more CPU intensive than others

The 12 TMH prediction methods could be separated into three groups according to their runtime (Supporting Information Table S8): Fast methods (TopPred2, SCAMPI, TMHMM2, HMMTOP2, and Phobius) predicted >100 proteins per minute on our machines (single core, AMD Opteron 2431, Supporting Information Table S8). Slow methods (PHDhtm, PolyPhobius, MEMSAT3, and SPOCTOPUS) predicted 4–11 proteins per minute, and the slowest method (MEMSAT-SVM) took several minutes for one protein.

Effect of larger sequence database is ambiguous

Using evolutionary information as input can improve predictions.[6,35,38] For the five evaluated methods that use such information (*cf.* Table I), we compared performance for generating alignments from the small Swiss-Prot database (~545,000 proteins, release 2014_04) and from a larger database based on the complete UniProtKB merged with PDB, redundancy reduced at 98% sequence identity (~21.7 million proteins, release 2014_04, Methods). The effect varied with MEMSAT-SVM benefiting most from the 40 times larger database (ΔQok>5 *pp*). PHDhtm and MEMSAT3 remained largely unaffected (ΔQok = 0.5 *pp*), while PolyPhobius performed slightly worse (ΔQok = −2.6 *pp*). For the subset of new proteins, SPOCTOPUS and MEMSAT3 reached Qok = 34%, while the other three methods reached Qok = 31%. Overall, the reasons for these observations remain unclear. One issue could be that on such large databases the number of returned hits exceeds the number of hits expected during development of the method and therefore the majority of them are discarded resulting in a set of low variance. Since the search parameters, the methods use are in nearly all cases not easily customizable and since the larger database also significantly increased the runtime, we recommend using a small database. The prediction results presented in this work were based on alignments built from Swiss-Prot.

CONCLUSION

We have evaluated how well 12 transmembrane helix (TMH) prediction methods correctly identify all experimentally observed TMHs and how well they distinguish between proteins with and without TMHs (Table I). The analysis used a newly created dataset of 190 nonredundant high-resolution alpha-helical TMPs. Forty-four of these 190 proteins were not sequence similar to any protein used for development of the 12 methods.

For our analysis, we used a performance measure which reflects the fraction of proteins in a dataset for which all TMHs are correctly predicted according to annotations of either OPM or PDBTM (Qok). Further, we introduced more stringent criteria than previous scores (difference between predicted and observed TMH endpoints maximally five residues, Qok, Table III) to account for the fact that subsequent experimental and computational methods based on TMH predictions often need precise membrane boundaries.[64,65] A global score such as Qok with our new stringent criteria was crucial, because traditional per-residue scores suggested unrealistically high performance (Supporting Information Figs. S5 and S6).

When combining several aspects of performance and runtime, PolyPhobius appeared to be the best method, followed by MEMSAT-SVM (Fig. 1). Our observation that performance dropped for many methods when evaluated on proteins for which structures were published after the method development, once again, highlighted the continued problem of over-training. Conversely, some of this drop could be explained by the fact that newly determined structures contain important novel information that will be needed to improve prediction methods of the future.

Earlier analyses suggested lower performance for eukaryotic TMPs than for bacterial TMPs.[25,33] Our result, in contrast, suggested the opposite: performance was lower for bacterial than for eukaryotic TMPs (Fig. 1C).

We confirmed that the best methods accurately distinguish between proteins with and without TMPs (Table II). Problems arose when testing soluble proteins with signal peptides. But even for those, at least for eukaryotes and Gram-positive bacteria, the best methods (Phobius and PolyPhobius) maintained error rates below 10% (Table II). Interestingly, the incorporation of evolutionary information did not help to improve much in this respect (PolyPhobius adds evolutionary information to Phobius and performs slightly worse, Table II). In contrast, all methods performing best in correctly predicting TMHs used evolutionary information. The fact that some methods need much less CPU-resources than others (differences of over two orders of magnitude) complicated the overview even more.

We did not explicitly evaluate re-entrant helices. These are predicted by only two of the 12 methods and we did show that our overall results did not change significantly for the subset of proteins without re-entrant helices. However, the growth in experimental information about re-entrant helices and their importance for topology prediction should encourage future method developers to account for this important type of membrane helices more explicitly. Another open problem was underlined by the comparisons between the TMH and the DSSP annotations, and between methods optimized to predict alpha-helices in general and TMHs: we do not have a tool that combines TMH prediction with a general secondary structure prediction in one model, although many proteins do.

Our analysis was limited by several constraints, most importantly by the limitation in the dataset size. Nevertheless, we answered our initial question by providing a good impression what method to use for which task. Finally, we concluded that future improvements are both feasible and needed.

ACKNOWLEDGMENTS

The authors thank Tim Karl for technical assistance and Marlena Drabik for administrative support. They particularly thank Marco Punta for many helpful discussions. They also thank the anonymous reviewers for helpful suggestions and all authors who made their methods openly available and provided versions to run on our own machines. Last but not least the authors thank all who practice open science and deposit their data into public databases and those who maintain these excellent databases.

REFERENCES

1. Bjarnadóttir TK, Gloriam DE, Hellstr SH, Kristiansson H, Fredriksson R, Schioth HB. Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse. Genomics 2006;88:263–273.
2. Pieper U, Schlessinger A, Kloppmann E, Chang GA, Chou JJ, Dumont ME, Fox BG, Fromme P, Hendrickson WA, Malkowski MG, Rees DC, Stokes DL, Stowell MHB, Wiener MC, Rost B, Stroud RM, Stevens RC, Sali A. Coordinating the impact of structural genomics on the human alpha-helical transmembrane proteome. Nat Struct Mol Biol 2013;20:135–138.
3. Jacoby E, Bouhelal R, Gerspacher M, Seuwen K. The 7 TM G-protein-coupled receptor target family. ChemMedChem 2006;1:761–782.
4. Overington JP, Al-Lazikani B, Hopkins AL. How many drug targets are there? Nat Rev Drug Discov 2006;5:993–996.
5. Kessel A, Ben-Tal N. Introduction to proteins. London, UK: CRC Press; 2011.
6. Rost B, Fariselli P, Casadio R. Topology prediction for helical transmembrane proteins at 86% accuracy. Protein Sci 1996;5:1704–1718.
7. Liu J, Rost B. Comparing function and structure between entire proteomes. Protein Sci 2001;10:1970–1979.
8. Fagerberg L, Jonasson K, von Heijne G. Prediction of the human membrane proteome. Proteomics 2010;10:1141–1149.
9. Stevens TJ, Arkin IT. Do more complex organisms have a greater proportion of membrane proteins in their genomes? Proteins 2000; 39:417–420.
10. White SH. The progress of membrane protein structure determination. Protein Sci 2004;13:1948–1949.
11. Kloppmann E, Punta M, Rost B. Structural genomics plucks high-hanging membrane proteins. Curr Opin Struct Biol 2012;22:326–332.
12. Berman HM, Westbrook J, Feng Z, Gillil G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res 2000;28:235–242.
13. Tusnády GE, Dosztányi Z, Simon I. PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank. Nucleic Acids Res 2005;33(Database issue):D275–D278.
14. Lomize Ma, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Res 2012;40(Database issue):D370–D376.
15. Jayasinghe S, Hristova K, White SH. MPtopo: a database of membrane protein topology. Protein Sci 2001;10:455–458.
16. Arnold K, Kiefer F, Kopp J, Battey JND, Podvinec M, Westbrook JD, Berman HM, Bordoli L, Schwede T. The protein model portal. J Struct Funct Genomics 2009;10:1–8.
17. Yarov-Yarovoy V, Schonbrun J, Baker D. Multipass membrane protein structure prediction using Rosetta. Proteins 2006;62:1010–1025.
18. Goldberg T, Hamp T, Rost B. LocTree2 predicts localization for all domains of life. Bioinformatics 2012;28:i458–i465.
19. von Heijne G, Gavel Y. Topogenic signals in integral membrane proteins. Eur J Biochem 1988;174:671–678.
20. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 1999;292:195–202.
21. Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, Hönigschmid P, Schafferhans A, Roos M, Bernhofer M, Richter L, Ashkenazy H, Punta M, Schlessinger A, Bromberg Y, Schneider R, Vriend G, Sander C, Ben-Tal N, Rost B. PredictProtein—an open resource for online prediction of protein structural and functional features. Nucleic Acids Res 2014;42(Web Server issue):W337–W343.
22. Rost B. PHD: predicting one-dimensional protein structure by profile based neural networks. Methods Enzymol 1996;266:525–539.

23. Bau D, Martin AJ, Mooney C, Vullo A, Walsh I, Pollastri G. Distill: a suite of web servers for the prediction of one-, two- and three-dimensional structural features of proteins. BMC Bioinformatics 2006;7:402.
24. Chen CP, Kernytsky A, Rost B. Transmembrane helix predictions revisited. Protein Sci 2002;11:2774–2791.
25. Ikeda M, Arai M, Lao DM, Shimizu T. Transmembrane topology prediction methods: a re-assessment and improvement by a consensus method using a dataset of experimentally-characterized transmembrane topologies. In Silico Biol 2002;2:19–33.
26. Cuthbertson JM, Doyle DA, Sansom MSP. Transmembrane helix prediction: a comparative evaluation and analysis. Protein Eng Des Sel 2005;18:295–308.
27. Elofsson A, von Heijne G. Membrane protein structure: prediction versus reality. Annu Rev Biochem 2007;76:125–140.
28. Tsirigos KD, Hennerdal A, Käll L, Elofsson A. A guideline to proteome-wide alpha-helical membrane protein topology predictions. Proteomics 2012;12:2282–2294.
29. Rath EM, Tessier D, Campbell AA, Lee HC, Werner T, Salam NK, Lee LK, Church WB. A benchmark server using high resolution protein structure data, and benchmark results for membrane helix predictions. BMC Bioinformatics 2013;14;111.
30. Kernytsky A. Static benchmarking of membrane helix predictions. Nucleic Acids Res 2003;31:3642–3644.
31. Punta M, Forrest LR, Bigelow H, Kernytsky A, Liu J, Rost B. Membrane protein prediction methods. Methods (San Diego, Calif.) 2007;41:460–474.
32. Tusnády GE, Simon I. The HMMTOP transmembrane topology prediction server. Bioinformatics 2001;17:849–850.
33. Jones DT. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. Bioinformatics 2007; 23:538–544.
34. Nugent T, Jones DT. Transmembrane protein topology prediction using support vector machines. BMC bioinformatics 2009;10:159.
35. Rost B, Casadio R, Fariselli P, Sander C. Transmembrane helices predicted at 95% accuracy. Protein Sci 1995;4:521–533.
36. Reynolds SM, Käll L, Riffle ME, Bilmes JA, Noble WS. Transmembrane topology and signal peptide prediction using dynamic bayesian networks. PLoS Comput Biol 2008;4:e1000213.
37. Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol 2004; 338:1027–1036.
38. Käll L, Krogh A, Sonnhammer ELL. An HMM posterior decoder for sequence feature prediction that includes homology information. Bioinformatics 2005;21 (Suppl 1):i251–i257.
39. Bernsel A, Viklund H, Falk J, Lindahl E, Von Heijne G, Elofsson A. Prediction of membrane-protein topology from first principles. Proc Natl Acad Sci USA 2008;105:7177–7181.
40. Hirokawa T, Boon-Chieng S, Mitaku S. SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics 1998;14:378–379.
41. Viklund H, Bernsel A, Skwark M, Elofsson A. SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. Bioinformatics 2008;24:2928–2929.
42. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. J Mol Biol 2001;305:567–580.
43. Claros MG, Von Heijne G. TopPred II: an improved software for membrane protein structure predictions. Comput Appl Biosci 1994; 10:685–686.
44. Möller S, Kriventseva EV, Apweiler R. A collection of well characterised integral membrane proteins. Bioinformatics 2000;16:1159–1160.
45. The Uniprot Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res 2012; 40(Database issue):D71–D75.
46. Velankar S, McNeil P, Mittard-Runte V, Suarez A, Barrell D, Apweiler R, Henrick K. E-MSD: an integrated data resource for bioinformatics. Nucleic Acids Res 2005;33(Database issue):D262–D265.
47. Mika S, Rost B. UniqueProt: creating representative protein sequence sets. Nucleic Acids Res 2003;31:3789–3791.
48. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
49. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 2011;8:785–786.
50. Blobel G, Dobberstein B. Transfer of proteins across membranes. II. Reconstitution of functional rough microsomes from heterologous components. J Cell Biol 1975;67:852–862.
51. von Heijne G. The signal peptide. J Membr Biol 1990;3:195–201.
52. Xie K, Dalbey RE. Inserting proteins into the bacterial cytoplasmic membrane using the sec and YidC translocases. Nat Rev Microbiol 2008;6:477–487.
53. Efron B. Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. Biometrika 1981;68:589–599.
54. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982;157:105–132.
55. Mitaku S, Hirokawa T, Tsuji T. Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane–water interfaces. Bioinformatics 2002;18:608–616.
56. Rost B, Casadio R, Fariselli P. Refining neural network predictions for helical transmembrane proteins by dynamic programming. Proc Int Conf Intell Syst Mol Biol 1996;4:192–200.
57. Viklund H, Elofsson A. OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. Bioinformatics 2008;24:1662–1668.
58. Nielsen H, Krogh A. Prediction of signal peptides and signal anchors by a hidden Markov model. Proc Int Conf Intell Syst Mol Biol 1998;6:122–130.
59. Lassmann T, Frings O, Sonnhammer ELL. Kalign2: high-performance multiple alignment of protein and nucleotide sequences allowing external features. Nucleic Acids Res 2009;37:858–865.
60. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 2006;22:1658–1659.
61. Hopf TA, Colwell LJ, Sheridan R, Rost B, Sander C, Marks DS. Three-dimensional structures of membrane proteins from genomic sequencing. Cell 2012;149:1607–1621.
62. Payton ME, Greenstone MH, Schenker N. Overlapping confidence intervals or standard error intervals: what do they mean in terms of statistical significance? J Insect Sci 2003;3:34.
63. Hedin LE, Ojemalm K, Bernsel A, Hennerdal A, Illergård K, Enquist K, Kauko A, Cristobal S, von Heijne G, Lerch-Bader M, et al. Membrane insertion of marginally hydrophobic transmembrane helices depends on sequence context. J Mol Biol 2010;396:221–229.
64. Alber F, Förster F, Korkin D, Topf M, Sali A. Integrating diverse data for structure determination of macromolecular assemblies. Annu Rev Biochem 2008;77:443–477.
65. Fleishman SJ, Harrington S, Friesner RA, Honig B, Ben-Tal N. An automatic method for predicting transmembrane protein structures using cryo-EM and evolutionary data. Biophys J 2004;87:3448–3459.