



# Computational Approaches for Revealing the Structure of Membrane Transporters: Case Study on Bilirubin Translocase

Katja Venko\*, A. Roy Choudhury, Marjana Novič\*

Department of Cheminformatics, National Institute of Chemistry, Ljubljana, Slovenia

## ARTICLE INFO

### Article history:

Received 2 November 2016

Received in revised form 19 January 2017

Accepted 20 January 2017

Available online 31 January 2017

### Keywords:

Membrane proteins

Bilirubin translocase

3D protein structure

Transmembrane region predictors

Helix–helix interactions

## ABSTRACT

The structural and functional details of transmembrane proteins are vastly underexplored, mostly due to experimental difficulties regarding their solubility and stability. Currently, the majority of transmembrane protein structures are still unknown and this presents a huge experimental and computational challenge. Nowadays, thanks to X-ray crystallography or NMR spectroscopy over 3000 structures of membrane proteins have been solved, among them only a few hundred unique ones. Due to the vast biological and pharmaceutical interest in the elucidation of the structure and the functional mechanisms of transmembrane proteins, several computational methods have been developed to overcome the experimental gap. If combined with experimental data the computational information enables rapid, low cost and successful predictions of the molecular structure of unsolved proteins. The reliability of the predictions depends on the availability and accuracy of experimental data associated with structural information. In this review, the following methods are proposed for *in silico* structure elucidation: sequence-dependent predictions of transmembrane regions, predictions of transmembrane helix–helix interactions, helix arrangements in membrane models, and testing their stability with molecular dynamics simulations. We also demonstrate the usage of the computational methods listed above by proposing a model for the molecular structure of the transmembrane protein bilirubin translocase. Bilirubin translocase is a bilirubin membrane transporter, which shares similar tissue distribution and functional properties with some of the members of the Organic Anion Transporter family and is the only member classified in the Bilirubin Transporter Family. Regarding its unique properties, bilirubin translocase is a potentially interesting drug target.

© 2017 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Transmembrane proteins play a crucial role in maintenance of normal cell functioning; acting as transporters and receptors, playing important roles in signaling pathways, the immune system, and energy production in form of ATP [1]. Therefore, they are of great interest as targets in drug designing, especially for having a double role in pharmacology: (i) indirect by influencing the absorption, distribution, metabolism and excretion (ADME) of drugs, (ii) direct as targets in the inhibition/overexpression of its primary function (the transport of molecules, switching on/off receptors). The membrane proteins are estimated to represent around a fourth of all genes and more than half of market drugs have a mode of action that targets membrane proteins [2]. Knowing the structures and functional mechanisms of more transmembrane proteins could significantly influence drug development and hopefully lead to better healthcare solutions.

Although, the first 3D atomic structure of a membrane protein was experimentally solved in 1984 [3], the majority of membrane protein structures are currently unknown and thus present a major experimental and computational challenge. In the last two years, statistics have shown an impressive increase in solved membrane protein structures by X-ray crystallography or NMR spectroscopy, but the number is still low in comparison with the soluble proteins (~380 unique structures, <1% of all known protein structures in PDB) [4]. The transmembrane proteins are polytopic integral membrane proteins with one or more chains that span the entire biological membrane. The major bottleneck in resolving the tertiary/quaternary structures of transmembrane proteins is the production of suitable stable crystals for X-ray diffraction studies. In this regard, experts in protein crystallography are faced with several technical difficulties in the synthesis, solubilization and maintenance of a functional and stable form of the protein during the crystallization process [5,6]. Since a considerable amount of effort is invested in breaking these barriers, new methodologies are published on a regular basis. For example, immanent crystallization screens have recently resulted in the development of new scaffold strategies for enhancing the stability of membrane proteins in detergent solutions [7,8]. Besides application of innovative methodological approaches the

Abbreviations: BTL, bilirubin translocase; TM, transmembrane.

\* Corresponding authors at: Department of Cheminformatics, National Institute of Chemistry, Hajdrihova ul. 19, SI-1000 Ljubljana, Slovenia.

E-mail addresses: [katja.venko@ki.si](mailto:katja.venko@ki.si) (K. Venko), [marjana.novic@ki.si](mailto:marjana.novic@ki.si) (M. Novič).

rapid emergence of genomics, proteomics and high-throughput technologies (automation, miniaturization, integration, third-generation synchrotrons, electron microscopy) support the enhanced protein structure determination rate [7,9]. With newly emerging methods, especially in atomic and cryo-electron microscopy, the information gaps in 3D structural datasets (e.g. CryoEM map) [10] will hopefully be filled in the near future. At present, the coverage of membrane protein fold space is relatively sparse and, therefore, the use of computational strategies is even more demanding [11].

Provisionally, given the current lack of experimental data, numerous highly sophisticated computational approaches are available for the elucidation of protein structures and functions [11–15]. The computational data-handling combined with existing experimental data inputs enable rapid, low-cost and successful approaches for the prediction of unsolved protein structures. The need for the acquisition of computational models can be directly demonstrated using a group of transporters defined as the solute carrier superfamily. In this superfamily, more than 15,100 human genes are currently annotated as transport proteins, but the X-ray structure only is resolved for 14 of them [4]. In the solute carriers (SLC) superfamily, the family of organic anion transporters (OAT) is the most dominant [16]. OAT members are of interest as potential drug targets due to their uptake of cancer drugs and also as tumor biomarkers [17–19]. In cases when we have only data on tissue expression and ligand-based experiments without synthesized protein and known gene annotation of protein, computational approaches are the only solutions to overcome the experimental gap. In this context, our work will shed light on the *in silico* research studies devoted to elucidating the three-dimensional (3D) structure of the bilirubin membrane transporter bilitranslocase (BTL, Uniprot O88750), which shares similar tissue distribution and functional properties with some of the members of the OAT family and is the only member classified in the Bilirubin Transporter Family (BRT) in the Transporter Classification Database (TCDB 2.A.65) [20]. As its name determines, it is strongly involved in the bilirubin uptake from the blood into the hepatocytes [21,22]. However, BTL is not responsible only for the transport of bilirubin into the hepatic cells; the biochemical studies performed so far reveal a much broader tissue expression (parietal cells of the gastric mucosa, proximal renal tubule cells and the vascular endothelium cells [23–27]), where BTL acts as an efficient influx device for a wide variety of vitally important endogenous substrates and drugs [28–30]. BTL is known to transport a wide variety of poly-aromatic compounds, but the most unique feature of BTL is the ability to transport nucleotides, which, to the best of our knowledge, no other membrane transporter is able to do [30]. These peculiarities make BTL a hot target for future investigations, especially in the context of endogenous drug delivery systems to intracellular targets. Consequently, there is an urgent need for the elucidation of its 3D molecular structure and the clarification of its substrate specificity, as well as its functional and mechanistic significance. Several experimental studies have already focused on BTL, but its 3D structure and functional mechanism still remain unclear due to problems with its synthesis. BTL is a 38.22 kDa transmembrane protein with 340 amino acids translated from a 1026 bp long mRNA sequence (GenBank: Y12178.1), which was isolated already in 1978 from liver cytoplasm of *Rattus norvegicus* [31]. The BTL gene remains unidentified until now. However, it is of interest that the mRNA of BTL shares 94% homology (RNO30106 coding/coding overlapping transcript in antiCODE database [32]) with antisense mRNA of ceruloplasmin, a plasma glycoprotein that has localized its functional gene on human chromosome 3q21–q24 and pseudogene on chromosome 8q21.13–q23 [27]. In spite of several efforts, the 3D structure of BTL is not yet resolved and no sequence homology with any other known membrane protein has been found [20]. Interestingly, BTL has a conserved bilirubin-binding motif, which plays a central role in ligand binding and is similar to globular alpha-phycoyanins [20], which are ancient biliproteins present in cyanobacteria.

The aim of this review is to demonstrate the usage of various computational methods for the structural determination of transmembrane proteins; examples are given for a case study of bilitranslocase. It needs to be emphasized that the selected case study is of special importance, as it demonstrates how to tackle the challenges of membrane protein structure analysis while lacking a template-based homolog. The structural studies include: sequence-dependent predictions of transmembrane regions, predictions of transmembrane helix–helix interactions, their arrangement in membrane models and testing their stability with molecular dynamics simulations.

## 2. Approaches for Predicting Protein Structures From Amino Acid Sequences

Protein structure prediction tools are nowadays the standard toolkit in life science research and a variety of *in silico* applications are available. The computational programs for protein structure elucidation include various statistical and supervised learning algorithmic techniques [13, 33] for solving the sequence–structure alignment and the identification of structural units: the very regular local sub-structures ( $\alpha$ -helices and  $\beta$ -sheets) in the secondary structure [34] and the protein domains (distinct and compact intra-chain units that could fold independently into other chain parts) in the tertiary structure [35]. All the computational algorithms were developed on the assumption that different proteins fold into similar 3D units because they have similar interaction patterns among their residues and between the residues and the environment [36].

Sequence-based tools for secondary and tertiary protein structure prediction are an important and challenging field of research in structural bioinformatics. Over the last two decades, the methods for the prediction of secondary structural elements in proteins have obviously progressed. Nowadays, secondary structure prediction methods are seldom used alone, but are mostly used to provide constraints for the tertiary structure prediction methods [11]. Consequently, the accurate secondary structure identification, the ever-increasing size of the training sets and the optimized alignment methods have also improved the sensitivity and usability of tertiary structure prediction methods [37]. The most widely used techniques for tertiary structure prediction are homology modeling, fold recognition (protein threading), and *de novo* (*ab initio*) protein structure prediction. The techniques vary in their basic concepts; (i) the first two are based on fitting a protein sequence to a structural template, requiring the identification of one or more homologs that the structure is based on (e.g. PROSPECT [38], RAPTOR [39], pGenTHREADER [40], pro-sp3-TASSER [41], MODELER [42], I-TASSER [43], iMembrane [44]), (ii) the third involves the use of simulated annealing algorithms, which are based on the general principles that govern protein structure and energetics and therefore does not require a template (e.g. ROSETTA [45], FRAGFOLD [46], PROINFO [47], FILM [48]). Each software tool has its specificities and advantages, and use different energy functions and computational methods, but protein 3D structures' building methods generally achieve relatively similar levels of accuracy. Especially among template-based methods, accurate predictions were observed when sequence identity was above 30%. Therefore, among the tertiary structure prediction methods, it is interesting to note the really outstanding popularity of homology modeling with MODELER [42]. However, for TM protein prediction Kelm et al. [49] presented the TM-specific homology-based tool MEDELLER and used performance analyses to demonstrate that MEDELLER outperforms the most popular homology modeling tool MODELER. The low usability of *ab initio* techniques is reasonable because they are much more computationally intensive thus limited to smaller proteins and are less accurate than template-based methods. Since template-based methods are currently the most widely used computational approaches for protein structure prediction [11], they are unfortunately still mostly applicable for soluble proteins, for which several crystal structure templates are available, while only few or no homologs of known structures exist for

majority of TM protein families. In this regard, *ab initio* methods seem more suitable for membrane protein structure predictions, but due to the size restriction associated with these methods, they are now mostly used for building the non-homologous loop regions only [33]. Some exceptions exist, such as the methods FILM [48] and RosettaMembrane [50], which were specially modified for TM proteins and are capable of predicting TM structures in sizes up to 300 residues. Therefore, in membrane protein structural bioinformatics, the topology prediction methods dominate, as the accuracy of models built using template-based methods is highly restricted by the availability of good structural templates and near-perfect alignments [11]. The structure elucidation of polytopic integral membrane proteins starts with especially sophisticated tools for the topology prediction of TM regions. Several TM-specific predictors are available for this step and will be further compared in this review with a practical representation of use on the BTL transporter. To date, the tertiary structure prediction of membrane proteins remains a significant challenge, but the recent development of methods specifically optimized to align membrane proteins (AlignMe [51], MP-T [52], PRALINE™ [53], TM-Coffee [54]) enable promising progress. Significant advances have been made in the last decade on methods for predicting helix–helix interactions, but the methods for the optimal packing of  $\beta$ -sheets are still in the development phase due to the severe shortage of known structures and higher diversity in topology [12].

However, among membrane protein families, some successful cases of the use of template-based and *ab initio* methods have also been presented [49,50,55,56]. G protein-coupled receptors (GPCRs) [56] and SLC transporters [37,55] are widely studied and already structurally resolved membrane proteins, therefore members of these groups can now be modeled with an accuracy sufficient for virtual screening. While, there is concern about the reliabilities of the structural models of membrane proteins that were attempted using methods that are sophisticated for soluble proteins [49]. The major concerns and drawbacks of such models are: (i) the models were built from template structures with low sequence identity to the target sequence, (ii) the alignment methods were based on statistical potentials derived from soluble proteins and did not take into account the fact that membrane proteins have different amino acid substitution preferences from their soluble counterparts, (iii) unique properties like specific interhelical interactions were not considered [11]. Recent studies demonstrate that if such preferences were included, the alignments were improved and consequently the membrane protein models were also improved [52, 55].

### 2.1. Sequence Dependent Predictions of Transmembrane Regions

The first step in transmembrane protein structure prediction is the identification of the  $\alpha$ - or  $\beta$ -TM regions. TM regions are, by concept, all the peptide chains within the nonpolar region of the lipid bilayer. The TM regions have in common the organization of chains into a plethora of predominantly hydrophobic residues that are energetically suitable for the hydrophobic membrane environment and have aromatic/charged residues at the terminal positions, which are more suitable for the membrane-water interface. The predictions of the TM regions are based on inputs of amino acid sequences and can in general be approached from two distinct perspectives: (i) pattern-based (hydrophobic or sequence pattern classification), (ii) homology-based (making comparison to existing data of homologs). In the first concept, the structural features are predicted based on algorithms using hydrophobicity scales or sequence similarity [31], therefore applicability is theoretically equal for homologs and non-homologs. The second concept is based on algorithms, which besides sequence patterns also include evolutionary information, so the probability of prediction is dependent on or biased with the homology rate [13,33,57]. Interestingly, all integral membrane proteins with currently known high-resolution structures are strictly homomers and not mix assemblies of both TM structural

units. Therefore, membrane proteins are classified in two highly distinct structural classes: bundles of  $\alpha$ -helices or  $\beta$ -barrels [58]. In this regard, the predictors for each class were diverged and developed separately to achieve better precision in the predictions of the TM structural units.

Both types of predictors apply various methods based on different algorithmic techniques, which can in general be categorized into three classes: physicochemical methods, statistical methods and machine learning methods [13,33,59]. Several physicochemical methods based on hydrophobicity indexes [60–62] and several hydropathy analyses that identify long stretches of hydrophobic residues are available for the topology prediction of TM helices (KD [63], PRED-TMR [64], SOSUI [65], TM Finder [66], TopPred2 [67]) or barrels (BBF [68], BOMP [69]). These methods are successful at identifying the hydrophobic core of TM regions, but cannot precisely determine the ends of TM residues, therefore they are commonly corrected with the cutoff values for membrane-spanning residues (values calibrated against known structures of membrane proteins) [70]. The propensity-based methods rely on statistical analysis of the occurrence of certain residues in the secondary structures of known protein structures (e.g. for helices (MEMSAT3 [71], SPLIT [72], TMPred [73]) and for barrels (Freeman–Wimley approach [74])). The applicability of these methods is mostly tempered due to the limited number of known atomic resolution structures of transmembrane proteins. The machine learning methods are rapidly evolving and are based on learning algorithms like Support Vector Machines, Hidden Markov Models and Neural Networks. Regarding various performance analyses, these methods are recognized as the most advanced and accurate [57,59,75]. Some of them are able to predict at least three structural states: cytoplasmic region, TM region and extracellular region. The basic principle of these methods is to train based on a set of input/output pairs and to detect correlations that facilitate pattern recognition and then evaluate the prediction probabilities (e.g. for helices HMMTOP [76], ENSEMBLE [77], Philius [78], TMHMM [79], MEMSAT-SVM [59], TOPCONS [80], Pred $\alpha$ TM [81] and for barrels B2TMPRED [82], ConBBPRED [75], PRED-TMBB [83], TMBpro [84], Pred $\beta$ TM [57], TMBHMM [85] etc.). Machine learning approaches are best for solving problems in the absence of general theories when there is a large amount of data with noisy patterns and are thus ideal for use in the explication of protein complexity.

Due to the shortage of the high-resolution TM protein structures, current computational methodologies probably do not yet cover all the properties of the TM domains in the proteome. Therefore, currently available algorithms are based on datasets that do not completely sample the TM protein data space of living organisms. The lack of information is especially substantial in the case of  $\beta$ -TM proteins. Nevertheless, the current predictors in most cases seems robust and relatively accurate, but the performances are significantly overestimated [57,59]. Consequently, the tendency in structural bioinformatics is to develop algorithms independent of explicitly resolved evolutionary information, which are potentially capable of giving accurate TM region predictions for the novel non-homologous membrane protein sequences. Even more, besides identifying the TM regions, the interest in using predictors for genome annotation is also increasing [57,59,86]. In spite of the abundance of predictors, the researcher's intent is always to use the most up to date methods based on all the currently available sequence and structural data. Thus prediction methods are gradually evolving by including the new protein structure information that becomes available [87]. However, with new experimental strategies, the availability of high-resolution structures is rapidly increasing and, consequently, coverage of the complete transmembrane protein space is approaching.

#### 2.1.1. Accuracy Measurements and Performance Analyses of Secondary Structure Predictors

Comparing the accuracy of different structure prediction methods is not a univocal task due to a number of methodological challenges, such as the use of different protein data sets or homologous proteins, various secondary structure definitions, and various accuracy measures. For TM

proteins in particular, these problems are of special relevance according to a relatively limited high-resolution structure database, which often leads to the overestimation of prediction accuracy [33]. Thus the best strategy to improve prediction accuracy is to make a consensus prediction based on compiling the prediction results from different methods [75,81].

In general, measurements of the accuracy of predictors are based on a comparison of the predicted and experimentally observed TM segments. Most of the measures are defined for a two-state classification scheme (TM segment or not TM segment), which are further differentiated based on per-residue and per-segment accuracy. Accuracy measurements are evaluated using confusion matrix data, which include the numbers of correct/incorrect predictions (true/false positive, true/false negative) and are defined as accuracy, sensitivity, specificity, and the Matthews correlation coefficient. Per-residue accuracy involves evaluating the accuracy of prediction for particular residues (the percentage of residues predicted correctly in TM regions). Per-segment accuracy is more commonly used and based on a comparison of predicted versus experimentally observed TM segments in a data set and are defined as per-segment sensitivity, positive predictive value and segment overlap [88,89]. The sensitivity is the percentage calculated from the number of correctly predicted TM regions divided by the number of experimentally observed TM regions in the data set. The precision is specified as a percentage of all the TM regions that are correctly predicted. The segment overlap (SOV) is a measure of the similarity of the predicted and experimentally observed TM segments [88].

In the case of  $\alpha$ -TM region predictors, several performance analyses have been implemented [57,89–91,87]. All have in common that the simple predictors based on hydrophobicity scales perform fairly well for the prediction of  $\alpha$ -TM helices, but are less accurate than the algorithms based on amino acid distribution preference, sequence alignments and evolutionary information. Overall, hydrophobicity-based methods overestimate the TM helices, mostly because of the inability to distinguish TM helices from signal peptides and hydrophobic regions outside of membranes. To avoid these problems, some authors propose firstly to use the prediction methods that are effective in identifying signal peptides as a pre-filter [92], prior to analyzing with TM topology predictors. Furthermore, some  $\alpha$ -TM region predictors already contain specialized sub-models for the recognition of signal peptides [78,59,87,93]. The studies have shown that advanced machine learning methods with complex algorithms using multiple sequence alignments outperform hydrophobicity-based predictors, but have a tendency to

underestimate the TM helices [59]. Based on a benchmark analysis of ten machine learning methods by Nugent and Jones [59], the MEMSAT-SVM method was demonstrated as the most accurate one. In another benchmark analysis, Tsirigos et al. [87] reported that the updated version of TOPCONS, a consensus method that combines several prediction methods, offers state-of-the-art performance. On the other hand, an outstanding methodological exception is the Pred $\alpha$ TM algorithm based on an effective mathematical representation of a peptide sequence. In this algorithm, the amino acid composition of TM regions and sequence patterns are characterized by mathematical descriptors derived from the amino acid adjacency matrix [81,91,94]. The recent performance analysis of 17 state-of-art  $\alpha$ -TM region predictors carried out on 38 benchmark sequences including both single-spanning and multi-spanning membrane proteins from both prokaryotes and eukaryotes is presented in Table 1. It is clear that some of the advanced methods perform better than others. Those achieving the highest sensitivity and precision are Pred $\alpha$ TM [81,91], TOPCONS [80], HMMTOP [76], Philius [78], OCTOPUS [95], SCAMPI [96] and SVMtop [97]. The Pred $\alpha$ TM predictor [91] is independent of any explicitly expressed evolution information, thus has an added advantage over the aforementioned algorithms when predicting TM helices for proteins that show very low or no sequence homology and it could also be useful in the context of entire proteomes analysis. Nugent and Jones [59] also demonstrated on a number of complete genomes that the MEMSAT-SVM method can effectively discriminate between TM and globular proteins and thus could be suited to the whole genome annotation of  $\alpha$ -helical TM proteins. Furthermore, Tsirigos et al. [86,87] examined the ability of prediction methods to distinguish between membrane proteins and soluble ones, and obtained the best performances with the methods utilizing multiple sequence alignments like TOPCONS [87] and Phobius [93]. The low agreement between predictions obtained using different methods shows that the global properties of the membrane proteome are still a hot matter of concern [86]. Therefore, for a relevant differentiation between globular and membrane proteins, it is suggested to apply a consensus approach based on the analysis of results from various currently available predictors.

The old performance analyses for the prediction of  $\beta$ -TM regions [75, 98] showed that methods based on the Hidden Markov Model like ProfTMB [80], HMM-B2TMR [81] and PRED-TMBB [68] were the best in terms of SOV, per-residue and per-segment accuracy. Furthermore, Bagos et al. [75] demonstrated that the consensus prediction method ConBBPred performed significantly better than each individual available

**Table 1**

The comparative performance of state-of-art freely available predictors for:  $\alpha$ -transmembrane regions - based on a benchmark dataset of 38  $\alpha$ -TM proteins [81,91],  $\beta$ -transmembrane regions - based on a benchmark dataset of 35  $\beta$ -TM proteins [57].

$\alpha$ -transmembrane region predictors					$\beta$ -transmembrane region predictors				
Program	SE	PR	Method	Reference	Program	SE	PR	Method	Reference
DAS-TMfilter	79	76	HA & DAS	Cserzo et al., 2004 [107]	B2TMPred	83	42	NN	Jacoboni et al., 2001 [82]
HMMTOP	90	89	HMM	Tusnady & Simon, 2001 [76]	ConBBPred	56	86	consensus	Bagos et al., 2005 [75]
MemBrain	83	87	NN	Shen & Chou, 2008 [108]	Pred $\beta$ TM	84	73	SVM & NN	Roy Choudhury et al., 2015 [57]
MEMSAT3	91	82	NN	Jones, 2007 [71]	TBBpred	74	41	SVM & NN	Natt et al., 2004 [98]
OCTOPUS	90	89	NN	Viklund & Elofsson, 2008 [95]	TMBETA-NET	72	46	NN	Gromiha et al., 2005 [102]
Philius	92	89	DBN	Reynolds et al., 2008 [78]	TMBpro	75	70	NN	Randall et al., 2008 [84]
Phobius	91	86	HMM	Kall et al., 2004 [93]					
Pred $\alpha$ TM	92	90	SVM & NN	Roy Choudhury et al., 2013 [91,103]					
PRED-TMR	92	82	HA	Pasquier et al., 1999 [64]					
SCAMPI	90	89	HMM	Bernsel et al., 2008 [96]					
SOSUI	88	80	HA	Hirokawa et al., 1998 [65]					
SVMtm	90	80	SVM	Yuan et al., 2004 [109]					
SVMtop	90	89	SVM	Lo et al., 2008 [110]					
TMHMM	91	87	HMM	Krogh et al., 2001 [79]					
TMpred	87	87	SA	Hofmann & Stoffel, 1993 [73]					
TOPCONS	90	90	consensus	Bernsel et al., 2009 [80]					
TopPred II	86	88	HA	Claros & von Heijne, 1994 [67]					

HA - hydrophathy analysis, SA - statistical analysis, DAS - Dense Alignment Surface, DBN - Dynamic Bayesian network.

SE (sensitivity) - % of all observed TM regions predicted correctly.

PR (precision) - % of all TM regions that are correctly predicted.

predictor. On the other hand, Freeman and Wimeley [74] showed that a good statistical approach can surpass the accuracy of machine learning methods. However, due to the amphipathic nature of the  $\beta$ -TM regions, the hydrophobicity was recognized as an inefficient discriminating factor and prediction accuracy was improved with the incorporation of non-linear statistics and evolutionary profiles. For most algorithms, the accuracy was overestimated regarding a limited benchmark set and the probable inclusion of homolog proteins in the original models dataset. The newest benchmarking analysis for  $\beta$ -TM predictors presented in Table 1 shows that the machine-learning-based Pred $\beta$ TM predictor currently outperforms all the state-of-art  $\beta$ -TM region prediction methods [57]. Although, the consensus method ConBBPred, which makes predictions by combining outputs from several algorithms (Pred-TMBB [83], ProfTMB [99], HMM-B2TMR [100], etc.) achieves a considerably high degree of precision, it radically fails in sensitivity, being the lowest among all the compared methods. In general, the  $\beta$ -TM proteins have little sequence identity, even if only the TM regions are considered. Therefore, the Pred $\beta$ TM algorithm, which is independent of any explicitly given evolutionary information [43], has an important advantage over most of the methods based on sequence profile data like B2TMPred [82], TBBpred [101], TMBpro [74] and TMBETA-NET [102]. Moreover, besides concerning the identification of TM proteins, it can also be useful for gene annotation in gram-negative bacterial genomes [57,74]. We can emphasize that the prediction of  $\beta$ -TM regions still remains a difficult problem. The majority of predictors have limited prediction accuracy due to the limited availability of non-homologous  $\beta$ -TM protein structures [4] and the probability of the existence of  $\beta$ -TM protein families that have not yet been described.

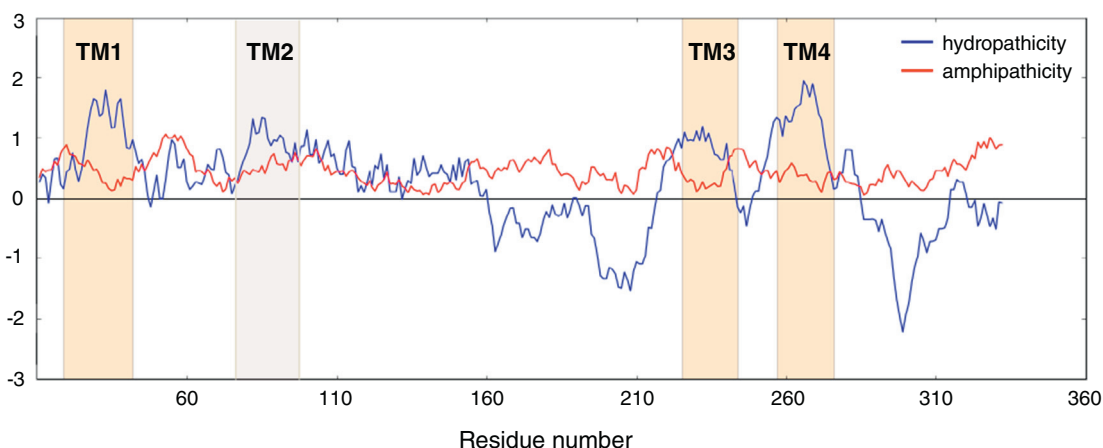
### 2.1.2. Case Study: Step 1 – BTL TM Regions Prediction

BTL is a 340 amino acids long TM protein of an uncharacterized gene and unknown 3D structure. Knowing only the mRNA amino acid sequence, Roy Choudhury et al. [81,91] made a prediction of the BTL TM regions by testing 17 freely available  $\alpha$ -TM region predictors. The analysis of the predictions resulted in the identification of four TM helices: TM1 (24–45), TM2 (73–95), TM3 (221–238) and TM4 (258–277). Only three algorithms (Pred $\alpha$ TM, TMpred and TopPred II) predict all four TM regions. The algorithms PRED-TMR, MemBrain and Philius predict the transmembrane regions TM1, TM2 and TM4. On the other hand, the predictors HMMTOP, SCAMPI and TOPCONS predict the transmembrane regions TM1, TM3 and TM4. The rest of the algorithms only predicted the regions TM1 and TM4. Interestingly, the SOUSI predictor

classifies BTL as a globular protein. The hydrophathy sequence profile shown in Fig. 1 can give an explanation why all the predictors reliably identified the TM1 and TM4 regions, while the identification of TM2 and TM3 regions was more ambiguous. Furthermore, based on these computational analyses of BTL secondary structure units, experiments were designed and the ambiguous TM2 and TM3 helices were successfully confirmed using multidimensional NMR spectroscopy [104–106].

### 3. Determining the Structure of Assembly of the TM Regions

For the elucidation of the membrane protein structure and functional mechanism, knowing the exact assembly and arrangement of its structural domains is crucial. The interactions between TM regions play a critical role in the function, assembly and oligomerization of membrane proteins. In these interactions, physical forces (e.g. polar and aromatic interactions, salt bridges) among two or more helices or strands are included [12,14]. The determination of the optimal packing of the TM regions includes effective sampling strategy and the usage of complex optimization algorithms following the principles from the laws of quantum mechanics. Theoretically, the optimal arrangements have the minimal free energy geometry [13]. In general, the simplified two-stage protein folding model of membrane proteins enhances the solving of this complex folding problem by including the results from TM regions predictors, which provide reliable information on the assembly of secondary structure units. Further, with mathematical operations, the minimal inter-unit interaction energy is calculated based on scoring functions that represent the molecular determinants of helix–helix/strand–strand interactions [12,14,111,112]. The representation of such determinants is the most challenging part and over the years, significant progress has been made in the field of tertiary structure prediction for helical TM proteins [113–115], while for  $\beta$ -TM proteins the identification of molecular determinants is moderate [33,85,116]. While the folding of globular proteins can be predicted using methods like homology modeling and threading, such methods should be applied very carefully and critically for the prediction of membrane protein folding, due to the shortage of atomic resolution TM protein structures. The accuracy of prediction is questionable if specific membrane preferences are not included in the protocol [115,37,49]. Once the optimal packing is determined, the stability of the predicted arrangement in the lipid environment could be further studied by using molecular dynamics (MD) simulations [117]. The MD simulations can be computationally expensive and thus the size of the simulated system should be reasonably limited or a multiscale approach used [117,118]. Regarding the



**Fig. 1.** The prediction of the bilitranslocase transmembrane regions. Hydrophathy analyses of the amino acid sequence using the Kyte-Doolittle scale (WHAT 2.0, [63]): the blue line represent hydrophathy and the red line represent amphipathicity. The columns represent the position of the predicted transmembrane regions (helices): orange – regions predicted using HMMTOP, Pred $\alpha$ TM, SCAMPI, TMpred, TopPredII, TOPCONS; gray – extra region predicted using Pred $\alpha$ TM, TMpred, TopPredII, Philius, MemBrain, PRED-TMR [81,91]. The transmembrane regions TM1, TM3 and TM4 are consistent with hydrophilic peaks, while the detection of TM2 is indefinite due to the ambiguous hydrophathy profile. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

case study protein, the details of practical examples of analyses of the arrangement of the four helices of biltranslocase is presented.

### 3.1. Predicting the TM Helix–Helix Interactions

So far, the known membrane proteins (ion channels, transporters and receptors) in living organisms are mostly built of helices [119], therefore substantial research is done on studying helix–helix interactions [12,14,114,120] and less is known about strand–strand interactions. The correct folding of polytopic membrane proteins firstly involves individual helix–helix interactions and then multiple helix–dimer interactions need to be controlled and aligned to result in the final higher-ordered oligomeric structure. In contrast to soluble proteins, our knowledge of the factors that control the oligomerization of membrane helices is limited. Taking into account that the membrane environment has unique chemical and physical properties, the rules applicable for interactions between soluble segments are not necessarily valid within the membrane [121]. Of interest is the recent extensive research by Zhang et al. [115] on helix–helix interactome, which provides the first comparison of TM and water-soluble helical pairs. Although both types of proteins have a similar pair geometry, the interactions are different. The involvement of larger charged residues and more interhelical hydrogen bonds are observed in water-soluble structures [115]. Among the best known TM helix dimerization factors are the conserved motifs GxxxG, QxxS and WxxW [14,114,120], interhelical residue pairs/triplets [122], polar clamp, and serine/leucine/glycine zippers [12–14]. Although it was believed that TM helix–helix interactions are mostly dominated by the rule of amino acid sequence motifs, experimental evidence suggests that interactions between TM helices are much more complex, resulting in an aromatic pattern (aromatic-XX-aromatic) [123] or interactions between TM sequences that do not contain any recognizable motifs [14]. Cymer et al. [114] highlighted the sequence context in their study, as well as lipid bilayer properties that highly modulate the structure and stability of a helix dimer. Recent studies have also revealed the particular physical properties of lipids or membranes that play roles in the protein folding process [124]. Most of the studies evaluate the helix–helix interactions by molecular dynamics (MD) simulations [12], but also several *in silico* methods exist for the prediction of pairwise helix–helix interactions from the primary sequence. In general, they are generated using various machine-learning-based algorithms involving either residue contacts (per-residue) or complete TM regions (per-segment) (e.g. MemBrain [125], MEMPack [126], PROFcon [127], SVMcon [128], TMHcon [112], TMhit [113], TMhhcp [130], IMP [131]). However, the tools available for the prediction of helix–helix interactions in globular proteins [127,128] perform relatively poorly in TM proteins. The reason is probably the differences between TM and globular interaction motifs [14,114,115,121]. For example, the helix–helix interaction prediction performance for five predictors (the TM protein contact predictors MEMPack, TMHcon and TMhit, and the globular protein contact predictors PROFcon and SVMcon) was assessed based on a data set of 74 sequences, which contained at least two TM helices [106]. The TM proteins' specific predictors TMhit and MEMPack achieved the highest accuracy (>60%). These predictors are not solely able to predict residue contacts and helix–helix interactions, but can also present a modest estimation of the lipid exposure and helical packing arrangement of TM proteins [129]. For example, studies on archaerhodopsin [126] and cytochrome C oxidase [129] revealed a high level of consensus among the predicted interactions and the observed helical packing arrangements derived from the crystal structure. Since such predictors are only capable of constructing packing arrangements of proteins with up to 13 TM helices [129], more advanced tools for predicting the assemblies of proteins were developed, which are further described in Section 3.2.

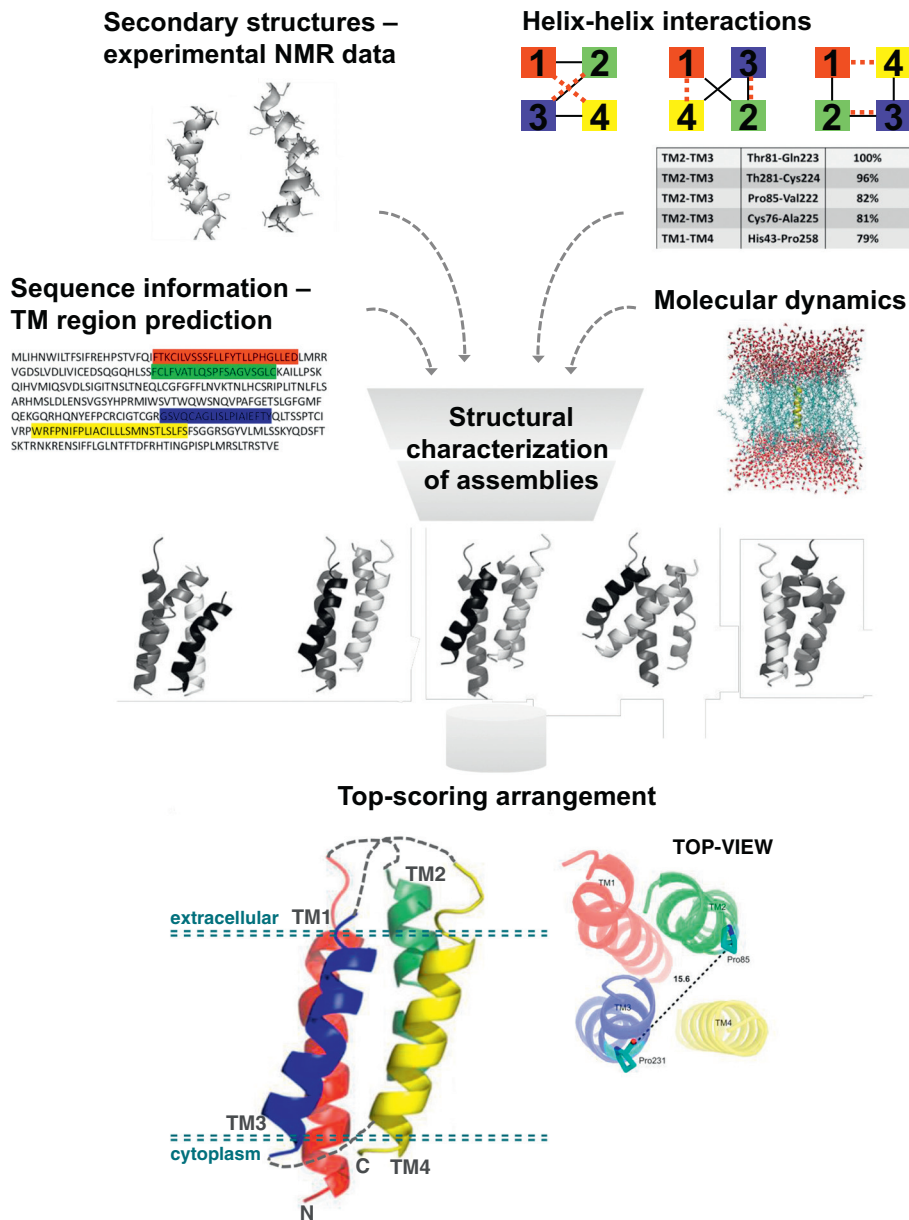
#### 3.1.1. Case Study: Step 2 – the Prediction of Transmembrane Helix–Helix Interactions in BTL

The Integrative Modeling Platform (IMP) [131] was used to predict the per-segment helix–helix interactions in BTL. The sequences of TM regions with defined topologies predicted by the TM region predictors served as the input. The rigid body representation for each TM region was built based on Discrete Optimized Protein Energy (DOPE) [132] using specified statistical potentials for TM proteins from SaliLab and without considering the experimental NMR or MD data that are available. All possible TM helix pairs were considered to predict the probable interactions, because TM helices can interact in multiple ways [122]. Each member of a helix–helix pair was allowed to access all the permissible combinations of rotational, translational and tilting movements. The permitted range of tilting angles, diameter, excluded volume and packing restraints for each TM helix pair were based on the Orientations of Proteins in Membranes (OPM) database [133], which lists the spatial arrangements of alpha TM proteins with respect to the hydrocarbon core of the lipid bilayer. The stability of the possible transmembrane helix–helix interactions were optimized using the scoring function based on interaction data of TM helices from the OPM database. It was assumed that all the TM helices could interact with each other; therefore each helix–helix pair interaction was optimized and scored independently. The highest scoring transmembrane pairs with scores above a predefined threshold were considered to be interacting. The four BTL predicted TM regions give rise to six possible helix–helix interaction pairs. Among them, the pairs TM2–TM3 and TM1–TM4 have the lowest energy scores and therefore showed the most optimized configurations and were reported to be interacting (Fig. 2) [106]. The TMhit algorithm [113], which predicts per-residue helix–helix interactions, was also applied. The TM2–TM3 and TM1–TM4 helix–helix pairs were found to interact, and these are the same pairs as those previously predicted by IMP. The existence of the TM2–TM3 pair in the SDS micellar environment was also independently supported experimentally by NMR spectroscopy and FRET efficiency [106].

#### 3.2. The Prediction of TM Assemblies

To elucidate protein functions the models of high order assemblies are essential. Using current computational models, the determination of the positions and orientations of all components can be unambiguous, yet continuous improvements using the implementation of new scoring functions are in progress [10,12,125]. However, such models are hopefully valuable for directing experimental studies on TM proteins where structural data is currently unavailable. Various prediction methods that generate the pseudo-atomic models of the assembly are available (e.g. MultiFit [10], RosettaMembrane [134], FRAGFOLD [46], FILM3 [135], IMP [131]). Nowadays they are gradually increasing in number due to the rise in availability of datasets from atomic- and low-resolution (e.g. cryo-electron microscopy (cryo-EM) density maps, small-angle-X-ray scattering (SAXS) profiles), genomic and proteomic techniques. In general, the use of such tools is less user-friendly due to the complex computational protocols and is not so appropriate for beginners. For an accurate and careful performance, the knowledge of an expert in structural bioinformatics is indispensably needed.

The methods for predicting high order arrangements of TM regions are based on the constraints of the interacting TM helix–helix pairs and moderate helical packing arrangements identified in advance using the methods described in the Section 3.1. This information is especially valuable for *ab initio* modeling methods [46,134,135], as it reduces the conformational search space and computational requirements [129]. In general, the probable arrangements of the predicted TM helices are determined with the identification of the most populated and energetically favorable arrangement type. The arrangement with the most favorable energy can be identified with an effective sampling strategy defined in an optimization algorithm. Probably the most popular



**Fig. 2.** The structural characterization of bilitranslocase, a membrane protein with no structural homolog. A hybrid approach that integrates data from diverse computational and experimental methods [91,104–106] was used for determining the structure of the transmembrane assemblies. At the bottom, the best scoring arrangement of the four transmembrane helices of bilitranslocase (Top2 [106]) is presented; TM1 (24–45), TM2 (73–95), TM3 (221–238) and TM4 (258–277).

optimization algorithm is based on Monte Carlo simulations, which can perform an optimized sampling of the protein search space [131]. The algorithms include scoring functions, which are composed of restraints and evaluate how well a model agrees with the information from which the restraint was derived. In most cases the restraints encode both what is generally known about protein structures and what is known about a particular structure of interest. Thus, a candidate model with the highest score is consistent with all the input information. Normally, by increasing the amount and quality of information included in restraints, the precision and accuracy of the resulting model increases [15].

In recent years, researchers realized that the detailed structural characterization of complex proteins or macromolecular assemblies is practically impossible to achieve using any single experimental or computational method. Therefore, scientists were and are still searching for a way to overcome this barrier. Until now, hybrid approaches that integrate data from diverse experiments (e.g. X-ray crystallography, cryo-EM, immuno-electron microscopy, NMR and FRET spectroscopy, SAXS,

immunoprecipitation, genetic and proteomic interactions) are highly efficient and prospective [15]. For example, integrated approaches like MultiFit [10] and IMP [131] can contribute to the structural characterization of biomolecules ranging in size and complexity from small peptides to large macromolecular assemblies, by integrating data from diverse experiments. MultiFit [10] is based on a protocol for the simultaneous fitting of high-resolution structures of components into a density map of the whole assembly. On the other hand, IMP [131] is even more advanced and has scoring functions suitable for integrating various types of experimentally obtained data like SAXS profiles, NMR spectroscopy, cryoEM and immunoEM images and density maps, proteomics data, fingerprints or affinity purification. Numerous complex protein structures have been successfully solved using this platform. The 26S proteasome structure [136] was determined with the integration of data from an electron microscopy map, proteomics data, and comparative protein structure models. The structure of the bacterial type II pilus was identified from NMR data and X-ray crystallographic structures of

constituent proteins [137]. The value of IMP especially, is demonstrated by solving the human RNA Polymerase II [138] and yeast nuclear pore complex [139] with integrating information from multiple experimental and computational sources.

### 3.2.1. Case Study: Step 3 – the Prediction of BTL TM Regions Arrangement

In the case of predicting the arrangement of the TM regions of BTL, the input parameters were the TM region sequences, topology, loop connectivity, conformations from MD simulations [104,105], the predicted interactions in the pairs TM2-TM3 and TM1-TM4 [106], and the filter parameters [121]. The discrete conformation space was defined with restrictions, such as the transport channel diameter and the tilt and depth of the TM helices [122,133]. The chosen value for the transport channel diameter is dependent on the number of TM regions present in BTL [133]. So defined initial configurations, were used as constraints in the Monte Carlo simulations coupled with the DOMINO algorithm in the MultFit toolkit [10]. The DOMINO algorithm applied an approach to efficiently discover global optimal solutions within the discrete sampling space [10]. The resulting subset solutions were then combined to obtain the global solution. The distribution analysis illustrates that even though the highest scoring structure favors the TM1-TM4-TM3-TM2 arrangement, the most frequently observed was the TM1-TM2-TM4-TM3 arrangement type (shown in Fig. 2), which has the predicted interacting TM helix pairs positioned diagonally opposite each other [106].

### 3.3. Stability Predictions of TM Assemblies with Molecular Dynamics (MD) Simulations

MD simulations can be used for stability predictions concerning the TM regions in a lipid bilayer and even more, the internal motions and conformational changes of the TM protein domains can be assessed [118]. MD simulations have ever-increasing potential in the elucidation of all biological dynamic systems by generating dynamic models that illustrate contact interactions and internal motions that are playing a functional role. In contrast to crystal structures, membrane proteins in their native membrane environments are not static, but rather have a variety of conformations [140]. Therefore, MD simulations allow the *in silico* reconstitution of structures and their movements in a bilayer environment. Bondar et al. [141] reported that the membrane protein conformational dynamics is associated with changes in the interhelical hydrogen bonding. As hydrogen bond interactions can be extracted from MD simulations, running MD simulations is crucial for understanding the H-bond dynamics and the stability of protein domains. Various computational dynamics methods can be used to investigate and study the dynamics of the biomolecular system [142]. Since membrane systems are large in size, the atomistic simulations are generally time-consuming, therefore the introduction of coarse-grained models in MD simulations speeds up this process [117,143]. In general, methodologies are highly dependent upon setting the role of the solvent in protein dynamics and the availability of a suitable potential-energy function to describe the energy landscape of the system with respect to the relative coordinates of the constituent atoms, which represent the structure of the studied system [117]. In this regard, the Lipidbook, a public database for force-field parameters used in simulating biological membrane systems [144], and the Course Grained Database (CGDB) [143] can be useful. Significant progress has been made in the performance of membrane protein MD simulations in recent years. For example, important functional implications of TM protein interactions were discovered in the KscA potassium channel [145], the Mscl mehanosensitive channel [146], the sugar transporter LacY [117], aquaporin [140], etc. [12]. In this way, specific residues that play important structural and functional roles or influence selectivity have been characterized [13,118,147,148]. Moreover, since a key element in the refinement of 3D protein structure prediction is to incorporate reliable structural information into the protocol, it is essential that various

prediction tools that enable the inclusion of MD data in the computation of 3D structural models already exist [149,150,131]. Moreover, to make MD simulations in bilayers easily available to interested stakeholders, the MemProtMD software was recently developed. This is a completely automated high-throughput computational pipeline for the insertion of membrane protein structures into an atomic resolution model of the membrane environment, and the results of simulations are publicly accessible in the MemProtMD database [151].

### 3.3.1. Case Study: Step 4 – MD Simulations of BTL TM Regions

The preparation of the system of BTL TM helices for MD simulation was done using CHARMM, one of the popular molecular modeling programs [152]. The lipid system was based on the standard dipalmitoylphosphatidylcholine (DPPC); additional layers of water molecules were added above and below the lipid bilayer to enable full system solvation. The Monte Carlo simulations are commonly used for generating a set of representative configurations under given thermodynamic conditions such as temperature and volume [153]. Thus, they were also performed in this study. Numerous 20 ns trajectories were obtained using different force constraints; all the trajectories were accurately analyzed to get the overall view. The stability of the TM regions was confirmed, if the RMSD values and backbone torsion angles were stable. In BTL, during the 20 ns long MD simulations, an overall stable RMSD was observed for each helix. The analysis of the torsion angles confirmed that the  $\alpha$ -helical secondary structure of the TM2 and TM3 regions remained preserved in the membrane environment [104,105]. In addition, the formation of Proline-induced kinks was observed in TM2-TM3 [106] and the authors suppose that this is crucial for the conformational change during BTL substrate transport. Kinks are well known to have a functional role in membrane proteins [147] and the detailed research on TM helices showed that majority of kinks are associated with Pro residue [148].

## 4. Summary and Outlook

The interdisciplinary approach is essential in complex bio-molecular studies, where experimentalists collaborate with computational chemists, who analyze large quantities of existing data and may quickly provide structural and functional information that is fundamental to guiding the further directions of investigation. Since the 3D structure of transmembrane proteins in general are mostly unknown due to technical difficulties in experimental sample preparations, computational approaches are the only way to elucidate the potential structural features of these proteins. This is a very challenging task due to the gap in the known atomic structures, which is especially notable in TM proteins. Although a variety of computational methods for the *in silico* determination of 3D protein structures are available, not all of them are sufficiently applicable for membrane proteins. The methods that are really sophisticated for membrane protein structure predictions are currently limited, but this field of research is of high interest and existing methods are constantly being updated or new ones developed with integrated modifications for TM proteins. Currently, the use of various template-based methods like comparative modeling and threading was able and surprisingly efficient for membrane proteins that share sequence homology with structurally resolved proteins. But what can be done in most of the cases when no structural membrane template exist that can be reliably aligned to the selected target? The use of *ab initio* methods is one of the possible choices, but is nowadays still limited to small protein systems. In contrast, an interesting hybrid approach, primarily used for the structural determination of macromolecular assemblies, is available for use in the structural characterization of membrane proteins. This approach was demonstrated by a case of structural elucidation of bilitranslocase, a transmembrane protein with a known amino acid sequence but with no available information about its 3D structure or structural homolog. The integration of diverse existing data obtained from computational and experimental methods was the



only way to make further progress in the elucidation of its structure. Such data integration with its iterative series of satisfaction to the input spatial restraints resulted in the 3D computational model of TM region arrangement of BTL, which represents an optimal arrangement of four membrane  $\alpha$ -helices. The methodology based on integrative approaches can maximize the accuracy and efficiency of structural characterization using constraints from diverse experimental and computational data. Thus, this methodology is recommended for the challenging elucidation of TM protein structures, when single computational and experimental methods are insufficient or not available. Finally, we would like to draw attention to an alternative direction of investigation of protein function when its' 3D structure is not known. When the 3D structure of the protein or at least of close homologs is not available, there are still some methods for the investigation of functions of such proteins based on structure-free approaches such as pharmacophore and structure–activity modeling. The *in silico* ligand-based models designed based on quantitative structure activity relationships (QSAR) are one of very few methods that can be used in the absence of the resolved 3D protein structure. Once the QSAR model is constructed, it can offer valuable information on the structural features of substrates that are of significance for their transport ability (e.g. drug absorption/excretion). For example, despite of lack of structural data, such effective QSAR models were also possible to develop for this case study protein BTL [29, 30,154]. However, this is completely new story for another review.

## Acknowledgements

The authors confirm that this article content has no conflicts of interest.

The authors acknowledge the financial support from the Slovenian Research Agency (research core funding No. P1-0017).

## References

- [1] Elofsson A, von Heijne G. Membrane protein structure: prediction versus reality. *Annu Rev Biochem* 2007;76:125–40.
- [2] Overington JP, Al-Lazikani B, Hopkins AL. How many drug targets are there? *Nat Rev Drug Discov* 2006;5(12):993–6. <http://dx.doi.org/10.1038/nrd2199>.
- [3] Deisenhofer J, Epp O, Miki K, Huber R, Michel H. X-ray structure analysis of a membrane protein complex. Electron density map at 3 Å resolution and a model of the chromophores of the photosynthetic reaction center from *Rhodospseudomonas viridis*. *J Mol Biol* 1984;180:385–98.
- [4] Protein Data Bank Statistics. <http://www.rcsb.org/pdb/results/results.do?tabshow=Current&sqid=B42BA2BF>.
- [5] Carpenter EP, Beis K, Cameron AD, Iwata S. Overcoming the challenges of membrane protein crystallography. *Curr Opin Struct Biol* 2008;18(5):581–6.
- [6] Baker M. Making membrane proteins for structures: a trillion tiny tweaks. *Nat Methods* 2010;7:429–34. <http://dx.doi.org/10.1038/nmeth0610-429>.
- [7] Kang HJ, Lee C, Drew D. Breaking the barriers in membrane protein crystallography. *Int J Biochem Cell Biol* 2013;45:636–44.
- [8] Gorrec F. Protein crystallization screens developed at the MRC Laboratory of molecular biology. *Drug Discov Today* 2016;21(5):819–25.
- [9] Moraes I, Evans G, Sanchez-Weatherby J, Newstead S, Stewart PDS. Membrane protein structure determination – the next generation. *Biochim Biophys Acta* 2014;1838:78–87.
- [10] Lasker K, Topf M, Sali A, Wolfson HJ. Inferential optimization for simultaneous fitting of multiple components into a CryoEM map of their assembly. *J Mol Biol* 2009;388:180–94.
- [11] Nugent T, Jones DT. Membrane protein structural bioinformatics. *J Struct Biol* 2012;179:327–37.
- [12] Fink A, Sal-Man N, Gerber D, Shai Y. Transmembrane domains interactions within the membrane milieu: principles, advances and challenges. *Biochim Biophys Acta* 2012;1818:974–83.
- [13] Liang J, Naveed H, Jimenez-Morales D, Adamian L, Lin M. Computational studies of membrane proteins: models and predictions for biological understanding. *Biochim Biophys Acta* 2012;1818:927–41. <http://dx.doi.org/10.1016/j.bbame.2011.09.026>.
- [14] Li E, Wimley WC, Hristova K. Transmembrane helix dimerization: beyond the search for sequence motifs. *Biochim Biophys Acta* 2012;1818:183–93.
- [15] Alber F, Forster F, Korkin D, Topf M, Sali A. Integrating diverse data for structure determination of macromolecular assemblies. *Annu Rev Biochem* 2008;77:443–77.
- [16] Koeppell H, Endou H. The SLC22 drug transporter family. *Pflugers Arch Eur J Physiol* 2004;447:666–76.
- [17] Golijanin DJ, Marshall J, Cardin A, Singer EA, Wood RW, et al. Bilitranslocase (BTL) is immunolocalised in proximal and distal renal tubules and absent in renal cortical tumors accurately corresponding to intraoperative near infrared fluorescence (nirf) expression of renal cortical tumors using intravenous indocyanine green (ICG). *J Urol* 2008;179:137.
- [18] Montanic S, Terdoslavich M, Rajcevic U, De Leo L, Bonin S, et al. Development and characterization of a novel mAb against bilitranslocase – a new biomarker of renal carcinoma. *Radiol Oncol* 2013;47(2):128–37.
- [19] Buxhofer-Ausch V, Secky L, Wlcek K, Svoboda M, Kounnis V, Briasoulis E, et al. Tumor-specific expression of organic anion-transporting polyptides: transporters as novel targets for cancer therapy. *J Drug Deliv* 2013. <http://dx.doi.org/10.1155/2013/863539>.
- [20] Battiston L, Passamonti S, Macagno A, Sottocasa GL. The bilirubin-binding motif of bilitranslocase and its relation to conserved motifs in ancient biliproteins. *Biochem Biophys Res Commun* 1998;247:687–92.
- [21] Sottocasa GL, Passamonti S, Battiston L, Pascolo L, Tiribelli C. Molecular aspects of organic anion uptake in liver. *J Hepatol* 1996;24:36–41.
- [22] Passamonti S, Terdoslavich M, Medic N, Margon A, Cocolo A, et al. Uptake of bilirubin into HepG2 cells assayed by thermal lens spectroscopy. Function of bilitranslocase. *FEBS J* 2005;272:5522–35.
- [23] Elias MM, Lunazzi GC, Passamonti S, Gazzin B, Miccio M, et al. Bilitranslocase localization and function in basolateral plasma membrane of renal proximal tubule in rat. *Am J Physiol* 1990;259:559–64.
- [24] Terdoslavich M, de Graaf IA, Proost JH, Cocolo A, Passamonti S, et al. Bilitranslocase is involved in the uptake of bromosulphophthalein in rat and human liver. *Drug Metab Lett* 2012;6:165–73.
- [25] Brandoni A, Di Giusto G, Franca R, Passamonti S, Torres AM. Expression of kidney and liver bilitranslocase in response to acute biliary obstruction. *Nephron Physiol* 2010;114:35–40.
- [26] Nicolini V, Grill V, Micali F, Narducci P, Passamonti S. Immunolocalisation of bilitranslocase in mucosecretory and parietal cells of the rat gastric mucosa. *J Mol Histol* 2005;36:45–50.
- [27] Maestro A, Terdoslavich M, Vanzo A, Kuku A, Tramer F, et al. Expression of bilitranslocase in the vascular endothelium and its function as a flavonoid transporter. *Cardiovasc Res* 2010;85:175–83.
- [28] Passamonti S, Terdoslavich M, Franca R, Vanzo A, Tramer F, et al. Bioavailability of flavonoids: a review of their membrane transport and the function of bilitranslocase in animal and plant organisms. *Curr Drug Metab* 2009;10:369–94.
- [29] Karawajczyk A, Drgan V, Medic N, Oboh G, Passamonti S, et al. Properties of flavonoids influencing the binding to bilitranslocase investigated by neural network modelling. *Biochem Pharmacol* 2007;73:308–20.
- [30] Župerl Š, Fornasaro S, Novič M, Passamonti S. Experimental determination and prediction of bilitranslocase transport activity. *Anal Chim Acta* 2011;705:322–33.
- [31] Tiribelli C, Lunazzi GC, Luciani M, Panfilii E, Gazzin B, et al. Isolation of a sulfobromophthalein-binding protein from hepatocyte plasma membrane. *Biochim Biophys Acta* 1978;532:105–12.
- [32] Yin Y, Zhao Y, Wang J, Liu C, Chen S, et al. antiCODE: a natural sense-antisense transcripts database. *BMC Bioinform* 2007;8:319.
- [33] Xu X, Xu D, Liang J, editors. Computational methods for protein structure prediction and modeling. Springer; 2007. p. 65–108.
- [34] Pauling L, Corey RB, Branson HR. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *PNAS* 1951;37(4):205–11. <http://dx.doi.org/10.1073/pnas.37.4.205>.
- [35] Richardson IS. The anatomy and taxonomy of protein structure. *Adv Protein Chem* 1981;34:167–339.
- [36] Pierri CL, Parisi G, Porcelli V. Computational approaches for protein function prediction: a combined strategy from multiple sequence alignment to molecular docking-based virtual screening. *Biochim Biophys Acta* 2010;1804:1695–712.
- [37] Schlessinger A. Characterizing the structure, function, and evolution of human solute carrier (SLC) transporters using computational approaches. In: Krämer R, Ziegler C, editors. Membrane transport mechanism: 3D structure and beyond. Springer; 2014. p. 23–58.
- [38] Xu Y, Xu D. Protein threading using PROSPECT: design and evaluation. *Proteins* 2000;40:343–54.
- [39] Xu J, Li M, Kim D, Xu Y. RAPTOR: optimal protein threading by linear programming. *Bioinform Comput Biol* 2003;1:95–117.
- [40] Lobley A, Sadowski MI, Jones DT. pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. *Bioinformatics* 2009;25:1761–7.
- [41] Zhou H, Skolnick J. Protein structure prediction by pro-sp3-TASSER. *Biophys J* 2009;96(6):2119–27.
- [42] Webb B, Sali A. Comparative protein structure modeling using Modeller. Current protocols in bioinformatics. John Wiley & Sons; 2014.
- [43] Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 2010;5:725–38.
- [44] Kelm S, Shi J, Deane CM. iMembrane: homology-based membrane-insertion of proteins. *Bioinformatics* 2009;25(8):1086–8.
- [45] Simons KT, Bonneau R, Ruczinski I, Baker D. Ab initio protein structure prediction of CASP III targets using ROSETTA. *Proteins* 1999;3:171–6.
- [46] Jones DT. Predicting novel protein folds by using FRAGFOLD. *Proteins Suppl* 2001;5:127–32.
- [47] Hung LH, Samudrala R. PROTFIN: secondary and tertiary protein structure prediction. *Nucleic Acids Res* 2003;31:3296–9.
- [48] Pellegrini-Calace M, Carotti A, Jones DT. Folding in lipid membranes (FILM): a novel method for the prediction of small membrane protein 3D structures. *Proteins* 2003;50:537–45.
- [49] Kelm S, Shi J, Deane CM. MEDELLER: homology-based coordinate generation for membrane proteins. *Bioinformatics* 2010;26(22):2833–40.

- [50] Barth P, Wallner B, Baker D. Prediction of membrane protein structures with complex topologies using limited constraints. *Proc Natl Acad Sci U S A* 2009;106(5):1409–14.
- [51] Stamm M, Staritzbichler R, Khafizov K, Forrest LR. AlignMe—a membrane protein sequence alignment web server. *Nucleic Acids Res* 2014;42:W246–51. <http://dx.doi.org/10.1093/nar/gku291>.
- [52] Hill JR, Deane CM. MP-T: improving membrane protein alignment for structure prediction. *Bioinformatics* 2013;29(1):54–61.
- [53] Pirovano W, Feenstra KA, Heringa J. PRALINETM: a strategy for improved multiple alignment of transmembrane proteins. *Bioinformatics* 2008;24(4):492–7.
- [54] Chang JM, Di Tommaso P, Taly JF, Notredame C. Accurate multiple sequence alignment of transmembrane proteins with PSI-coffee. *BMC Bioinform* 2012;13. <http://dx.doi.org/10.1186/1471-2105-13-54-51>.
- [55] Schlessinger A, Yee SW, Sali A, Giacomini KM. SLC classification: an update. *Clin Pharmacol Ther* 2013;94(1):19–23.
- [56] Michino M, Chen J, Stevens RC, Brooks CL. FoldGPCR: structure prediction protocol for the transmembrane domain of G protein-coupled receptors from class A. *Proteins* 2010;78(10):2189–201.
- [57] Roy Choudhury A, Novič M. Pred $\beta$ TM: a novel  $\beta$ -transmembrane region prediction algorithm. *PLoS One* 2015;10(12):e0145564.
- [58] White SH, Wimley WC. Membrane protein folding and stability: physical principles. *Annu Rev Biophys Biomol Struct* 1999;28:319–65.
- [59] Nugent T, Jones DT. Transmembrane protein topology prediction using support vector machines. *BMC Bioinform* 2009;10:159.
- [60] Eisenberg D, Weiss RM, Terwilliger TC. The helical hydrophobic moment—a measure of the amphiphilicity of a helix. *Nature* 1982;299:371374.
- [61] Engelman DM, Steitz TA, Goldman A. Identifying nonpolar transbilayer helices in amino-acid-sequences of membrane-proteins. *Annu Rev Biophys Chem* 1986;15:321–53.
- [62] Kessel A, Ben-Tal N. Free energy determinants of peptide association with lipid bilayers. In: Simon S, McIntosh T, editors. *Peptide-Lipid Interactions*, vol. 52. San Diego: Academic Press; 2002. p. 205–53.
- [63] Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 1982;157:105–32.
- [64] Pasquier C, Promponas VJ, Palaios GA, Hamodrakas SJ. A novel method for predicting transmembrane segments in proteins based on statistical analysis of the SwissProt database: the PRED-TMg algorithm. *Protein Eng* 1999;12:381–5.
- [65] Hirokawa T, Boon-Chieng S, Mitaku S. SOSUL: classification and secondary structure prediction of membrane proteins. *Bioinformatics* 1998;14:378–9.
- [66] Deber CM, Wang C, Liu LP, Prior AS, Agrawal S, et al. TM finder: a prediction program for transmembrane protein segments using a combination of hydrophobicity and nonpolar phase helicity scales. *Protein Sci* 2001;10:212–9.
- [67] Claros MG, von Heijne G. TopPred II: an improved software for membrane protein structure predictions. *Comput Appl Biosci* 1994;10(6):685–6.
- [68] Zhai YF, Saier MH. The beta-barrel finder (BBF) program, allowing identification of outer membrane beta-barrel proteins encoded within prokaryotic genomes. *Protein Sci* 2002;11:2196–207.
- [69] Berven FS, Flikka K, Jensen HB, Eidhammer I. BOMP: a program to predict integral beta-barrel outer membrane proteins encoded within genomes of gram-negative bacteria. *Nucleic Acids Res* 2004;32(13):W394–9.
- [70] Ulmschneider MB, Sansom MS, Di Nola A. Properties of integral membrane protein structures: derivation of an implicit membrane potential. *Proteins Struct Funct Bioinform* 2005;59:252–65.
- [71] Jones DT. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics* 2007;23:538–44.
- [72] Juretic D, Jeroncic A, Zucic D. Sequence analysis of membrane proteins with the web server SPLIT. *Croat Chem Acta* 1999;72(4):975–97.
- [73] Hofmann K, Stoffel W. TMbase—A database of membrane spanning protein segments. *Biol Chem Hoppe Seyler* 1993;374:166.
- [74] Freeman TC, Wimley WC. A highly accurate statistical approach for the prediction of transmembrane  $\beta$ -barrels. *Bioinformatics* 2010;26(16):1965–74.
- [75] Bagos PG, Liakopoulos TD, Hamodrakas SJ. Evaluation of methods for predicting the topology of beta-barrel outer membrane proteins and a consensus prediction method. *BMC Bioinform* 2005;6:7.
- [76] Tusnády GE, Simon I. The HMMTOP transmembrane topology prediction server. *Bioinformatics* 2001;17:849–50.
- [77] Martelli L, Fariselli P, Casadio R. An ENSEMBLE machine learning approach for the prediction of all-alpha membrane proteins. *Bioinformatics* 2003;19:i205–11.
- [78] 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.
- [79] Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 2001;305:567–80.
- [80] Bernsel A, Viklund H, Hennerdal A, Elofsson A. TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res* 2009;37:W465–8.
- [81] Roy Choudhury A, Novič M. Data-driven model for the prediction of protein transmembrane regions. *SAR QSAR Environ Res* 2009;20:741–54.
- [82] Jacoboni I, Martelli PL, Fariselli P, De Pinto V, Casadio R. Prediction of the transmembrane regions of beta barrel membrane proteins with a neural network based predictor. *Protein Sci* 2001;10:779–87.
- [83] Bagos G, Liakopoulos TD, Spyropoulos IC, Hamodrakas S. PRED-TMBB: a web server for predicting the topology of beta-barrel outer membrane proteins. *Nucleic Acids Res* 2004;32:W400–4.
- [84] Randall A, Cheng J, Sweredoski M, Baldi P. TMBpro: secondary structure, beta-contact and tertiary structure prediction of transmembrane beta-barrel proteins. *Bioinformatics* 2008;24(4):513–20.
- [85] Singh NK, Goodman A, Walter P, Helms V, Hayat S. TMBHMM: a frequency profile based HMM for predicting the topology of transmembrane beta-barrel proteins and the exposure status of transmembrane residues. *BBA* 2011;1814(5):664–70.
- [86] Tsirigos KD, Hennerdal A, Käll L, Elofsson A. A guideline to proteome-wide  $\alpha$ -helical membrane protein topology predictions. *Proteomics* 2012;12(14):2282–94.
- [87] Tsirigos KD, Peters C, Shu N, Käll L, Elofsson A. The TOPCONS web server for consensus prediction of membrane protein A guideline to proteome-wide  $\alpha$ -helical membrane protein topology predictions. *Nucleic Acids Res* 2015;43:W401–7.
- [88] Zemla A, Venclovas C, Fidelis K, Rost B. A modified definition of SOV, a segment-based measure for protein secondary structure prediction assessment. *Proteins Struct Funct Genet* 1999;34:220–3.
- [89] Chen CP, Kernysky A, Rost B. Transmembrane helix predictions revisited. *Protein Sci* 2002;11:2774–91.
- [90] Viklund H, Elofsson A. Best alpha-helical transmembrane protein topology predictions are achieved using hidden Markov models and evolutionary information. *Protein Sci* 2004;13:1908–17.
- [91] Roy Choudhury A. Development of transmembrane region prediction algorithms and structural elucidation of biltranslocase. Doctoral dissertation Ljubljana: Jožef Stefan International Postgraduate School; 2013[112 pp].
- [92] Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* 2007;2(4):953–71.
- [93] Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* 2004;338:1027–37.
- [94] Roy Choudhury A, Novič M. Amino acid distribution in transmembrane regions: a statistical analysis and comparison with globular proteins. *Int J Chem Model* 2012;4:205–19.
- [95] Viklund H, Elofsson A. A method that improved topology prediction for transmembrane proteins by using two-track ANN-based preference scores and an improved topological grammar. *Bioinformatics* 2008;24:1662–8.
- [96] Bernsel A, Viklund H, Falk J, Lindahl E, von Heijne G, et al. Prediction of membrane-protein topology from first principles. *PNAS* 2008;105(20):7177–81.
- [97] Lo A, Chiu HS, Sung TY, Lyu PC, Hsu WL. Enhanced membrane protein topology prediction using a hierarchical classification method and a new scoring function. *J Proteome Res* 2008;7:487–96.
- [98] Nilsson I, Persson B, von Heijne G. Consensus predictions of membrane protein topology. *FEBS Lett* 2000;486:267–9.
- [99] Bigelow HR, Petrey DS, Liu J, Przybylski D, Rost B. Predicting transmembrane beta-barrels in proteomes. *Nucleic Acids Res* 2004;32:2566–77.
- [100] Martelli PL, Fariselli P, Krogh A, Casadio R. A sequence-profile based HMM for predicting and discriminating beta barrel membrane proteins. *Bioinformatics* 2002;18:546–53.
- [101] Natt NK, Kaur H, Raghava GPS. Prediction of transmembrane regions of beta-barrel proteins using ANN and SVM based method. *Proteins Struct Funct Bioinform* 2004;56:11–8.
- [102] Gromiha MM, Ahmad S, Suwa M. TMBETA-NET: discrimination and prediction of membrane spanning  $\beta$ -strands in outer membrane proteins. *Nucleic Acids Res* 2005;33:W164–7.
- [103] Pred $\alpha$ TM and Pred $\beta$ TM: transmembrane region predictors. Available: <http://transpred.ki.si>.
- [104] Perdih A, Roy Choudhury A, Župerl Š, Sikorska E, Zhukov I, et al. Structural analysis of a peptide fragment of transmembrane transporter protein biltranslocase. *PLoS One* 2012;7(6):e389671–14].
- [105] Roy Choudhury A, Perdih A, Župerl Š, Sikorska E, Šolmajer T, et al. Structural elucidation of transmembrane transporter protein biltranslocase: conformational analysis of the second transmembrane region TM2 by molecular dynamics and NMR spectroscopy. *BBA-Biomembranes* 2013;1828(11):2609–19.
- [106] Roy Choudhury A, Sikorska E, van den Boom J, Bayer P, Popena Ł, et al. Structural model of the biltranslocase transmembrane domain supported by NMR and FRET data. *PLoS One* 2015;10(8):e0135455.
- [107] Cserzo M, Eisenhaber F, Eisenhaber B, Simon I. TM or not TM: transmembrane protein prediction with low false positive rate using DAS-TMfilter. *Bioinformatics* 2004;20:136–7.
- [108] Shen H, Chou JJ. MemBrain: improving the accuracy of predicting transmembrane helices. *PLoS One* 2008;3:e2399.
- [109] Yuan Z, Mattick JS, Teasdale RD. SVMtm: support vector machines to predict transmembrane segments. *J Comput Chem* 2004;25:632–6.
- [110] Lo A, Chiu HS, Sung TY, Lyu PC, Hsu WL. Enhanced membrane protein topology prediction using a hierarchical classification method and a new scoring function. *J Proteome Res* 2008;7(2):487–96.
- [111] Lo A, Chiu HS, Sung TY, Lyu PC, Hsu WL. Enhanced membrane protein topology prediction using a hierarchical classification method and a new scoring function. *J Proteome Res* 2008;7(2):487–96.
- [112] Fuchs A, Kirschner A, Frishman D. Prediction of helix–helix contacts and interacting helices in polytopic membrane proteins using neural networks. *Proteins* 2009;74:857–71.
- [113] Lo A, Chiu YY, Rødland EA, Lyu PC, Sung TY, et al. Predicting helix–helix interactions from residue contacts in membrane proteins. *Bioinformatics* 2009;25(8):996–1003.
- [114] Cymer F, Veerappan A, Schneider D. Transmembrane helix–helix interactions are modulated by the sequence context and by lipid bilayer properties. *Biochim Biophys Acta* 2012;1818:963–73.
- [115] Zhang SQ, Kulp DW, Schramm CA, Mravic M, Samish I, et al. The membrane- and soluble-protein helix–helix interactome: similar geometry via different interactions. *Structure* 2015;23:527–41.
- [116] Thomas JF, Wimley CWC. A highly accurate statistical approach for the prediction of transmembrane beta-barrels. *Bioinformatics* 2010;26(16):1965–74.

- [117] Bond PJ, Holyoake J, Ivetac A, Khalid S, Sansom MSP. Coarse-grained molecular dynamics simulations of membrane proteins and peptides. *J Struct Biol* 2007;157:593–605.
- [118] Lindahl E, Sansom MS. Membrane proteins: molecular dynamics simulations. *Curr Opin Struct Biol* 2008;18(4):425–31. <http://dx.doi.org/10.1016/j.sbi.2008.02.003>.
- [119] Greene LH, Lewis TE, Addou S, Cuff AL, Dallman T, et al. The CATH domain structure database: new protocols and classification levels give a more comprehensive resource for exploring evolution. *Nucleic Acids Res* 2007;35:D291–7.
- [120] Teese MG, Langosch D. Role of GxxxG motifs in transmembrane domain interactions. *Biochemistry* 2015;54:5125–35.
- [121] Walters RFS, DeGrado WF. Helix-packing motifs in membrane proteins. *Proc Natl Acad Sci U S A* 2006;103(37):13658–63.
- [122] Adamian L, Liang J. Helix–Helix packing and interfacial pairwise interactions of residues in membrane proteins. *J Mol Biol* 2001;311:891–907.
- [123] Sal-Man N, Gerber D, Bloch I, Shai Y. Specificity in transmembrane helix–helix interactions mediated by aromatic residues. *J Biol Chem* 2007;282:19753–61.
- [124] Mitchell DC. Progress in understanding the role of lipids in membrane protein folding. *Biochim Biophys Acta* 2012;1818:951–6. <http://dx.doi.org/10.1016/j.bbamem.2011.12.029>.
- [125] Yang J, Jang R, Zhang Y, Shen HB. High-accuracy prediction of transmembrane inter-helix contacts and application to GPCR 3D structure modeling. 2013;29(20):2579–87.
- [126] Nugent T, Ward S, Jones D. The MEMPack alpha-helical transmembrane protein structure prediction server. *Bioinformatics* 2011;27(10):1438–9. <http://dx.doi.org/10.1093/bioinformatics/btr096>.
- [127] Punta M, Rost B. PROFcon: novel prediction of long-range contacts. *Bioinformatics* 2005;21:2960–8.
- [128] Cheng J, Baldi P. Improved residue contact prediction using support vector machines and a large feature set. *BMC Bioinform* 2007;8:113.
- [129] Nugent T, Jones DT. Predicting transmembrane helix packing arrangements using residue contacts and a force-directed algorithm. *PLoS Comput Biol* 2010;6:e1000714.
- [130] Wang XF, Chen Z, Wang C, Yan RX, Zhang Z, et al. Predicting residue-residue contacts and helix–helix interactions in transmembrane proteins using an integrative feature-based random forest approach. *PLoS One* 2011;6:e26767.
- [131] Russel D, Lasker K, Webb B, Schneidman D, Velázquez-Muriel J, et al. Putting the pieces together: integrative structure determination of macromolecular assemblies. *PLoS Biol* 2012;10:e1001244.
- [132] Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. *Protein Sci* 2006;15:2507–24.
- [133] Lomize MA, Lomize AL, Pogozheva D, Mosberg HI. OPM: orientation of proteins in membrane database. *Bioinformatics* 2006;22:623–5.
- [134] Barth P, Schonbrun J, Baker D. Toward high-resolution prediction and design of transmembrane helical protein structures. *PNAS* 2007;104:15682–7.
- [135] Nugent T, Jones DT. Accurate de novo structure prediction of large transmembrane protein domains using fragment assembly and correlated mutation analysis. *PNAS* 2012;E1540–7.
- [136] Lasker K, Förster F, Bohn S, Walzthoeni T, Villa E, et al. Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proc Natl Acad Sci U S A* 2012;109(5):1380–7.
- [137] Simon B, Madl T, Mackereth CD, Nilges M, Sattler M. An efficient protocol for NMR-spectroscopy-based structure determination of protein complexes in solution. *Angew Chem* 2010;49:1967–70.
- [138] Lasker K, Phillips JL, Russel D, Velázquez-Muriel J, Schneidman-Duhovny D, et al. Integrative structure modeling of macromolecular assemblies from proteomics data. *Mol Cell Proteomics* 2010;9:1689–702.
- [139] Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, et al. Determining the architectures of macromolecular assemblies. *Nature* 2007;450:683–94.
- [140] Janosi L, Ceccarelli M. The gating mechanism of the human aquaporin 5 revealed by molecular dynamics simulations. *PLoS One* 2013;8(4):e59897. <http://dx.doi.org/10.1371/journal.pone.0059897>.
- [141] Bondar AN, White SH. Hydrogen bond dynamics in membrane protein function. *Biochim Biophys Acta* 2012;1818:942–50. <http://dx.doi.org/10.1016/j.bbamem.2011.11.035>.
- [142] Haile JM. MOLECULAR DYNAMICS SIMULATION: elementary methods. New York: Wiley-Interscience Publication; 1997[481 pp.].
- [143] Chetwynd AP, Scott KA, Mokrab Y, Sansom MS. CGDB: a database of membrane protein/lipid interactions by coarse-grained molecular dynamics simulations. *Mol Membr Biol* 2008;25:662–9.
- [144] Domański J, Stansfeld PJ, Sansom MS, Beckstein O. Lipidbook: a public repository for force-field parameters used in membrane simulations. *J Membr Biol* 2010;236(3):255–8. <http://dx.doi.org/10.1007/s00232-010-9296-8>.
- [145] Stewart JJ. Application of the pm6 method to modeling proteins. *J Mol Model* 2009;15(7):765–805.
- [146] Jeon J, Volth GA. Gating of the mechanosensitive channel protein mscl: the interplay of membrane and protein. *Biophys J* 2008;94(9):3497–511.
- [147] Hall S, Roberts K, Vaidehi N. Position of helical kinks in membrane protein crystal structures and the accuracy of computational prediction. *J Mol Graph Model* 2009;27(8):944–50. <http://dx.doi.org/10.1016/j.jmglm.2009.02.004>.
- [148] Yohannan S, Faham S, Yang D, Grosfeld D, Chamberlain A, et al. A C alpha–H...O hydrogen bond in a membrane protein is not stabilizing. *J Am Chem Soc* 2004;126(8):2284–5. <http://dx.doi.org/10.1021/ja0317574>.
- [149] Chen J, Brooks CL. Can molecular dynamics simulations provide high-resolution refinement of protein structure? *Proteins* 2007;67(4):922–30.
- [150] Moulton J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP) – round x. *Proteins* 2014;82(2):1–6. <http://dx.doi.org/10.1002/prot.24452>.
- [151] Stansfeld PJ, Goose JE, Caffrey M, Carpenter EP, Parker JL, et al. MemProtMD: automated insertion of membrane protein structures into explicit lipid membranes. *Structure* 2015;23(7):1350–61. <http://dx.doi.org/10.1016/j.str.2015.05.006>.
- [152] Brooks BR, Brooks CL, MacKerell Jr AD, Nilsson L, Petrella RJ, et al. CHARMM: the biomolecular simulation program. *J Comput Chem* 2009;30:1545–615.
- [153] Paquet E, Viktor HL. Molecular dynamics, Monte Carlo simulations, and Langevin dynamics: a computational review. *Biomed Res Int* 2015;183918.
- [154] Martinčič R, Venko K, Župerl Š, Novič M. Chemometrics approach for the prediction of structure – activity relationship for membrane transporter bilitranslocase. *SAR QSAR Environ Res* 2014;25(11):853–72.