



Improving prediction of helix–helix packing in membrane proteins using predicted contact numbers as restraints

Bian Li ^{1,2} Jeffrey Mendenhall,^{1,2} Elizabeth Dong Nguyen,² Brian E. Weiner,^{1,2} Axel W. Fischer ^{1,2} and Jens Meiler^{1,2*}

¹ Department of Chemistry, Vanderbilt University, Nashville, Tennessee 37232

² Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37232

ABSTRACT

One of the challenging problems in tertiary structure prediction of helical membrane proteins (HMPs) is the determination of rotation of α -helices around the helix normal. Incorrect prediction of helix rotations substantially disrupts native residue–residue contacts while inducing only a relatively small effect on the overall fold. We previously developed a method for predicting residue contact numbers (CNs), which measure the local packing density of residues within the protein tertiary structure. In this study, we tested the idea of incorporating predicted CNs as restraints to guide the sampling of helix rotation. For a benchmark set of 15 HMPs with simple to rather complicated folds, the average contact recovery (CR) of best-sampled models was improved for all targets, the likelihood of sampling models with CR greater than 20% was increased for 13 targets, and the average RMSD100 of best-sampled models was improved for 12 targets. This study demonstrated that explicit incorporation of CNs as restraints improves the prediction of helix–helix packing.

Proteins 2017; 85:1212–1221.

© 2017 Wiley Periodicals, Inc.

Key words: contact number; residue packing density; helix–helix packing; helical membrane protein; de novo protein structure prediction.

INTRODUCTION

Helical membrane proteins (HMPs) are essential components of a living cell. They play crucial roles in orchestrating the interactions of the cell with its environment; for example, by mediating cellular signaling, regulating ion gradients, and facilitating the transfer of molecules across the cell membrane. It was estimated that 20–30% of genes in most genomes encode HMPs.¹ HMPs are also very important therapeutic targets, about 50% of therapeutics on the market target HMPs.² The availability of a three-dimensional (3D) structure of a HMP not only improves our understanding of how the protein works at the atomic level³ but also facilitates the development of new therapeutics.^{4–6} Despite great progress in experimental techniques for determining HMP structures, only ~2% structures in the protein databank are HMPs,⁷ highlighting the fact that HMP structure characterization is still a challenge. Further, experimental data for HMPs are often of limited resolution, requiring computational methods to elucidate atomic-level details.

Similarly, not all biologically relevant conformations of HMPs—which tend to be very flexible—can be studied experimentally. Likewise, accurate computational methods for HMP structure prediction are a complement to existing experimental techniques to enable HMP structure determination from limited experimental data.^{8,9}

A commonly used computational approach for predicting protein tertiary structure is comparative modeling. However, a sequence identity of at least 25% between target and template proteins is recommended to give reliable models.¹⁰ Because the fold of most HMPs are unknown and it was estimated that comparative

Grant sponsor: NIH; Grant numbers: R01 GM080403, R01 GM099842, R01 DK097376, R01 HL122010, R01 GM073151; Grant sponsor: NSF; Grant number: CHE 1305874; Grant sponsor: American Heart Association; Grant number: 16PRE27260211.

*Correspondence to: Jens Meiler, Ph.D. Vanderbilt University, Departments of Chemistry and Pharmacology, Center for Structural Biology, 465 21st Ave South, BIOSCI/MRBIII, Room 5144B, Nashville, TN 37232-8725, USA. E-mail: jens.meiler@vanderbilt.edu

Received 4 August 2016; Revised 20 January 2017; Accepted 17 February 2017
Published online 6 March 2017 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.25281

modeling covers at most 10% of HMPs,¹¹ a few de novo methods have been developed, such as Rosetta–Membrane¹² and BCL::MP-Fold.⁷ Rosetta–Membrane assembles models helix-by-helix starting from a helix near the middle of the protein.¹² For HMPs with ~150 residues or less, Rosetta–Membrane achieved RMSD100 (root-mean-square distance normalized to a sequence of 100 residues) values of <4 Å to experimental structures. However, the prediction accuracy with respect to helix rotation around the main axis was either not evaluated or very poor.⁷ BCL::MP-Fold uses secondary structure element (SSE) pools and inserts helices across the membrane to build complete models. It achieved RMSD100 values to the experimental structure in the range of 3 to 8 Å for most benchmark HMPs.⁷ For models assembled by BCL::MP-Fold, even though TMHs are predicted to span the membrane with the correct topology, ~40% were reported to contain helices with incorrect rotation.⁷ For example, contact-forming, buried residues are sometimes rotated toward the membrane. For HMP models to be useful in applications such as structure-based drug design, accurate modeling of helix rotation is essential.

One approach to improving the accuracy of de novo tertiary structure prediction is to incorporate restraints.¹³ These restraints may be experimental, such as NMR chemical shifts⁸ and electron-paramagnetic resonance (EPR) accessibilities,⁹ or computational, such as predicted residue–residue contacts.^{11,13–16} For example, Fischer *et al.* recently showed that using either experimental or simulated EPR accessibility increases the likelihood of sampling native-like HMP folds and improves the accuracy of predicting helix rotations.⁹ Residue–residue contacts derived from experiments or accurate computational predictions also provide substantial guiding information for sampling. For instance, Evfold_membrane developed by Hopf *et al.* enables de novo prediction of tertiary structures of 25 HMPs by incorporating amino acid covariation extracted from evolutionary sequence record.¹¹

Residue contact number (CN) is a real-valued quantity that measures the degree of local packing of a residue within the protein tertiary structure. The CN of a given residue was originally computed by applying a clear distance cutoff and considering indiscriminately residues within the cutoff.^{17,18} Later improvements incorporated various distance-dependent weighting schemes to account for the distance-dependent nature of residue–residue interactions.^{19–21} CNs have been used to derive protein dynamic properties such as B-factor profile.²⁰ Studies have also shown that CN is the main structural determinant of site-specific substitution rates of proteins.²² Although it has been suggested that CNs could help in tertiary structure prediction, to our knowledge, no studies on tertiary structure prediction have explicitly incorporated CNs.

The CNs of interfacial TMHs (peripheral TMHs of a helical bundle) follow a signature periodic trend. Importantly, the CN signature of a TMH is tightly coupled to its rotation: even a small perturbation of the helix rotation will disrupt the CN signature. Hence, the CN signature of a TMH should give a strong constraint over its rotation. However, experimental CNs are not available until the tertiary structure of the protein is determined. Very recently, we developed a dropout neural network-based method, BCL::TMH-Expo, specifically for predicting CNs for HMPs.²³ CNs predicted by BCL::TMH-Expo correlate well with CNs computed from experimental structures and mirror exposure patterns of TMHs.²³ In this study, CNs predicted by BCL::TMH-Expo were incorporated into the empirical scoring function of BCL::MP-Fold in the form of restraints to improve prediction of helix–helix packing. We tested this method on a set of 15 benchmark HMPs that span a wide range of fold complexity.

MATERIALS AND METHODS

Benchmark set

A set of 15 multi-spanning HMP subunits were carefully selected to assess whether using CN restraints can improve the prediction of helix–helix packing. This set consists of HMP subunits that are both structurally and functionally diverse (Table I). Pairwise sequence identity is 30% or less. Sequence length ranges from 156 to 467 residues. The number of TMHs ranges from 4 to 10. As a measure of the size of transmembrane domains, the number of TMH residues was also computed for each target. None of these HMPs was used in the training set of BCL::TMH-Expo or had a sequence identity of >30% to any of the HMPs in the training set of BCL::TMH-Expo. This benchmark set contains diverse folds ranging from simplistic four-helix bundles and 7-TM receptors, up to proteins with 10 TMHs or helices in reentrant regions. Six of these HMPs are homo-oligomers. Because of the complexity of folding oligomers, we limited the scope of the present investigation to consider only a single subunit of each oligomer.

Computation of experimental and predicted contact numbers

The details of the algorithm for computing CNs from experimental structures can be found in two previous studies.^{21,23} Briefly, the experimental CN of residue *i* was computed as a weighted sum of contacts contributed by residues over the entire protein:

$$CN_i = \sum_{j \in [j-i] > 3}^n w_{ij}$$

Table 1
Summary of the Benchmark Set

PDB ID	Structure method	Resolution	Length	TMH	TMH residue	PCC	MAE	Oligomeric state
10ED	EM	4.0	227	4	104	0.35	2.23	Homopentamer
10KC	X-ray	2.2	292	6	214	0.39	2.37	Monomer
1PV6	X-ray	3.5	189	6	163	0.62	1.66	Monomer
1PY6	X-ray	1.8	249	7	177	0.72	1.29	Monomer
1U19	X-ray	2.2	348	7	173	0.58	1.63	Monomer
2BL2	X-ray	2.1	156	4	119	0.65	2.42	Homo 10-mer
2K73	NMR	NA	164	4	99	0.45	1.78	Monomer
2O9G	X-ray	1.9	234	6	166	0.69	1.74	Homotetramer
2Y01	X-ray	2.6	315	7	185	0.76	1.45	Monomer
3M71	X-ray	1.2	314	10	242	0.85	1.33	Homotrimer
3QAP	X-ray	1.9	239	7	168	0.69	1.35	Monomer
3UG9	X-ray	2.3	333	7	194	0.45	1.75	Homodimer
3UON	X-ray	3.0	467	7	183	0.66	1.60	Monomer
4A2N	X-ray	3.4	194	5	123	0.58	1.67	Monomer
4O6Y	X-ray	1.7	230	6	156	0.58	1.55	Homodimer
Mean			265	6.6	164	0.60	1.72	

PCC: Pearson correlation coefficient; MAE: mean absolute error; EM: electron microscopy; X-ray: X-ray diffraction; NMR: nuclear magnetic resonance.

where w_{ij} is the contribution made by residue j and is assigned in a distance-dependent manner such that short-range contacting residues have higher contribution than long-range contacting ones. Residues whose C_{β} atom is within 4.0 Å to the C_{β} atom of the residue of interest are assigned a contribution of 1.0; those with a distance longer than 11.4 Å are assigned a contribution of 0. Any residue 4–11.4 Å is assigned a contribution between 0.0 and 1.0 according to a smooth transition function.²¹ Only residues separated by more than three residues along the sequence were considered in the calculation to reduce the bias due to sequence proximity and local secondary structure. Experimental CNs were calculated based on structures retrieved from the OPM (Orientations of Proteins in Membranes) database.²⁴

Although a relatively low sequence identity (30%) was maintained while compiling a list of benchmark protein chains to reduce the homology between the modeling benchmark set and the training set for BCL::TMH-Expo, such level of sequence identity alone may not be sufficient to exclude homology among protein chains. In fact, substantial remote homology could still exist at this level placing HMPs in the same structural superfamily.²⁵ Such remote homology between proteins in the training set and proteins in the modeling benchmark set can lead to an optimistic estimate of the performance for new folds. As a way of preventing such optimism, the original training set for BCL::TMH-Expo was partitioned such that each SCOP superfamily²⁶ forms its own subset that contains all its members and no members from other SCOP superfamilies. Predicted CN of each residue of a modeling benchmark protein was then obtained through a specific variant of the neural network-based CN predictor BCL::TMH-Expo. This variant was trained using all remaining proteins after excluding the subset of proteins that share the same SCOP superfamily as the modeling

benchmark protein from the original training set of BCL::TMH-Expo. For example, for predicting the contact numbers for 3UON, all proteins that are in the same SCOP superfamily as 3UON were removed from the original training proteins of BCL::TMH-Expo and a neural network was trained using the remaining proteins. The contact numbers for 3UON were then predicted using this retrained neural network. This strategy was applied to each protein in the modeling benchmark set. Note that BCL::TMH-Expo is a dropout neural network-based algorithm that predicts CNs for HMPs. It uses the position-specific scoring matrix (PSSM)²⁷ derived from multiple sequence alignment (MSA) by PSI-BLAST²⁸ as predictive features and outputs residue-specific CN. The MSA for each protein chain in the benchmark set was obtained by searching the UniRef50²⁹ non-redundant sequence database with PSI-BLAST for five iterations.²⁸ The E value inclusion threshold was set to 10^{-2} . Floating point-valued PSSM was generated from PSI-BLAST checkpoint files using the source code (chkparse.c) adapted from PSIPRED.³⁰ Predicted CN was obtained by feeding the floating point-valued PSSM to BCL::TMH-Expo.

Incorporating CNs as restraints in folding simulations

The de novo membrane protein structure prediction algorithm BCL::MP-Fold⁷ developed by adapting the original algorithm BCL::Fold³¹ for membrane proteins was used to assemble 3D models. BCL::MP-Fold assembles 3D models by drawing TMHs from a pool of predicted TMHs. TMH pools were created from predictions made by the combined membrane association and secondary structure predictor BCL::MASP.³² A Monte Carlo minimizer with Metropolis criterion³³ was used to sample models with low energy. To use CNs to guide sampling of helix–helix packing, a CN-based penalty score

was added to the knowledge-based scoring function of BCL::MP-Fold, such that:

$$Score = \sum_i w_i \times S_i + w_p \times Penalty$$

where S_i represents each of the individual knowledge-based potentials previously derived and w_i is the associated weight. These potentials have been detailed in prior studies.^{7,34} The restraint scoring term was defined using the following formula:

$$Penalty = \sqrt{\frac{1}{n} \sum_{i=1}^n \delta_i^2}$$

where n is the number of residues in the assembled structural model, δ is the difference between the CN used as restraint and the CN calculated from the assembled structural model. w_p is the corresponding weight of the penalty. An optimal balance between the knowledge-based potentials and the penalty score is critical for correcting helix rotation while sampling native-like folds. If the weight for the restraint penalty is too low, its capacity of correcting helix rotation is reduced, if the weight is too high, it dominates other scoring terms. An empirical approach, in which a range of w_p values were systematically tested in preliminary sampling, was used to determine a near-optimal weight. Finally, five thousand models were assembled for each target in the benchmark set. The procedure for generating 3D models is summarized in Figure 1.

Metrics for measuring model quality

Root-mean-square distance (RMSD) gives a useful impression of the similarity between two structures if there is only a slight difference between their conformations. Unfortunately, a small perturbation in just one part of the protein (for instance, off position of a short loop) can lead to a large RMSD and it would seem that one structure substantially differs from the other. To address this issue, several quality measures have been introduced among which RMSD100³⁵ is commonly used. RMSD100 is a normalized, sequence length-independent version of RMSD calculated using:

$$RMSD100 = \frac{RMSD}{1 + \ln \sqrt{\frac{n}{100}}}$$

where n is the number of residues superimposed. Using RMSD100 as an indicator of structural variability reduces the influence of the intuition that larger proteins are more likely to differ from one another.³⁵ In this study, RMSD100 was computed over the C_{α} atoms of all TMH residues.

A metric called contact recovery (CR), defined as the percentage of native contacts recovered in the assembled 3D model, was used to measure the accuracy of helix

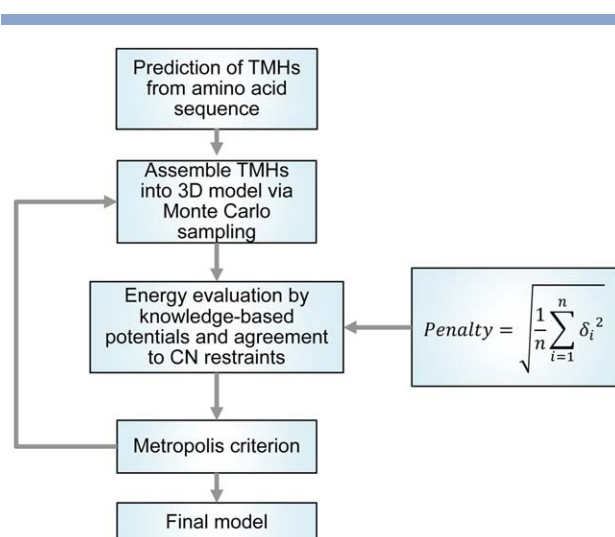


Figure 1

Protocol for assembling 3D models. BCL::MP-Fold predicts the tertiary structure of a HMP by assembling predicted TMHs in the 3D space. In the first step, the TMHs are predicted using the neural network-based membrane association and secondary structure prediction algorithm BCL::MASP. Predicted TMHs are assembled into a 3D model, and perturbed using a Monte Carlo sampling algorithm. The energy of the model after each perturbation is evaluated by knowledge-based potentials and agreement to CN restraints. The perturbation is subjected to the Metropolis criterion and is either accepted or rejected depending on the difference between the energies before and after the perturbation. This process is repeated for a specific number of iterations or until the maximum number of 2000 iterations without energy improvement is reached. [Color figure can be viewed at wileyonlinelibrary.com]

rotations in our previous study.⁷ However, the previous definition does not account for false positive contacts (FPC), which may be prevalent in 3D models assembled in a globular shape when the real shape of the protein is extended or rod-like and it has helices or strands that are somewhat “detached” from its main domain. In such cases, these “detached” secondary structure fragments could potentially be packed against the main domain of the protein by the folding algorithm, and thus, making a substantial fraction of FPCs. Thus, we redefined CR as the F1-score. Being the harmonic mean of precision and recall, the F1-score accounts for FPCs by weighting precision and recall equally:

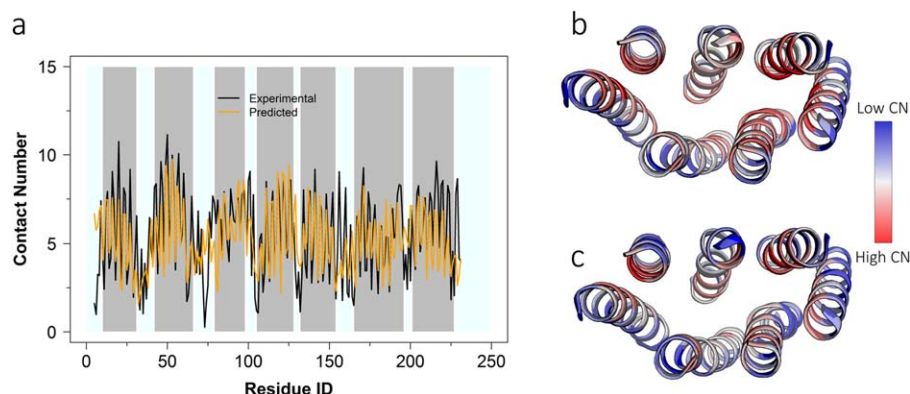
$$Contact\ Recovery = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

where

$$Precision = \frac{TPC}{TPC + FPC}$$

and

$$Recall = \frac{TPC}{TPC + FNC}$$

**Figure 2**

Agreement between experimental and predicted CNs of 1PY6. (a) Experimental and predicted CNs plotted against residues sequence positions, (b) Experimental CNs mapped onto structure; (c) Predicted CNs mapped onto structure. Color scheme in (b) and (c): as CN increases, color changes gradually from blue to red. Only TMHs are shown. [Color figure can be viewed at wileyonlinelibrary.com]

TPC (true positive contacts) denotes the number of contacts observed in the experimental structures that are correctly predicted in the assembled model and FNC (false negative contacts) is the number of contacts in the experimental structure that are missed in the assembled model. Two residues are considered in contact if they are separated along the sequence by at least 12 residues and the distance between their C_{β} atoms is within 8 Å. CR reaches its best value at 100% and worst at 0%.

Computation of enrichment

The enrichment was used to measure how capable a scoring function is to select the most accurate models from a pool of models. To calculate enrichment, models of a given set S are sorted by their CR values. The top 10% of the models with the highest CR values are put into the set T (true) and the rest of the models are put into the set F (false). The models in S are then sorted by their evaluated score. The top 10% of models with the lowest score are put into the set P (positive) and the rest are put into the set N (negative). The intersection of sets T and P are models that are correctly identified by the scoring function and referred to as TP (true positives). The intersection of sets F and P are models that are incorrectly identified by the scoring function and are referred to as FP (false positives). The enrichment value is then computed using the following formula:

$$\text{Enrichment} = \frac{TP}{TP + FP} / \frac{P}{P + N}$$

Intuitively, $\frac{P}{P+N}$ represents that probability of obtaining a native-like model when choosing a model from S at random, whereas $\frac{TP}{TP+FP}$ represents the probability of obtaining a native-like model when choosing from a set of models below an energy cutoff. By our experimental

design, $\frac{P}{P+N}$ has a constant value of 0.1, and therefore, the maximum enrichment value that can be achieved is 10.

RESULTS AND DISCUSSION

Predicting CNs for HMPs in the benchmark set

Table I shows the Pearson correlation coefficient (PCC) between experimental and predicted CNs as well as the mean absolute error (MAE) of predicted CNs for each target in the modeling benchmark set. The average PCC and the average MAE over the modeling benchmark set were 0.60 and 1.72, respectively. Notably, the CNs for three proteins, namely 1PY6, 2Y01, and 3M71, were predicted with a $PCC > 0.70$. Whereas, for 1OED, 1OKC, 3UG9, and 4A2N, the PCCs were below 0.50. Factors affecting the accuracy of CN prediction include oligomeric state, whether the protein chain is bitopic, and other factors that had been discussed previously in detail.²³ To illustrate the agreement between experimental and predicted CNs and visualize the predictions, the experimental and predicted CNs of 1PY6 were plotted and mapped onto its experimental structure. As shown in Figure 2(a), the predicted CNs of 1PY6 are in close agreement with experimental CNs, particularly in trans-membrane regions (vertical gray bars). As expected, predicted CNs generally distinguish between the exposed and buried faces of helices [Figs. 2(b, c)]. We, therefore, reasoned that the native rotation of helices can be confined by forcing them to satisfy predicted CNs, thus improving the prediction of helix–helix packing.

Incorporation of CNs significantly improved CR

The following three CR-based parameters were compared among the three simulation groups (E: with

Table II

Summary of Contact Recovery

Target	β_{CR} (%)			μ_{CR} (%)			Relative improvement in μ_{CR} (%) $\frac{\mu_{CR(P)} - \mu_{CR(N)}}{\mu_{CR(N)}} \times 100$	π_{20} (%)		
	<i>E</i>	<i>P</i>	<i>N</i>	<i>E</i>	<i>P</i>	<i>N</i>		<i>E</i>	<i>P</i>	<i>N</i>
10ED	73.10	38.02	28.93	70.59	32.88	23.47	40.09	51.77	5.17	0.33
1OKC	18.34	10.78	9.96	14.81	9.14	8.17	11.87	0.00	0.00	0.00
1PV6	31.55	30.02	21.90	26.93	23.09	17.06	35.35	1.37	0.35	0.03
1PY6	54.65	41.86	22.31	44.45	35.56	20.08	77.09	13.01	10.31	0.11
1U19	30.28	25.31	20.46	26.98	23.59	16.57	42.37	2.43	1.87	0.04
2BL2	68.40	59.29	54.50	66.78	55.22	49.63	11.26	76.26	50.98	29.80
2K73	59.49	49.33	30.70	57.04	44.13	27.82	58.63	72.04	33.58	1.45
2O9G	14.65	14.15	11.67	11.47	11.92	10.76	10.78	0.00	0.00	0.00
2Y01	36.15	21.97	19.42	30.60	20.70	17.30	19.65	1.94	0.19	0.00
3M71	23.46	23.58	17.14	21.77	20.35	14.42	41.12	0.54	0.20	0.00
3QAP	48.24	43.64	26.16	39.67	39.48	22.24	77.52	15.63	10.86	0.32
3UG9	38.38	35.90	24.16	35.98	30.08	20.66	45.60	14.71	6.77	0.14
3UON	32.37	21.81	19.16	25.93	20.02	16.01	25.05	1.11	0.09	0.00
4A2N	49.64	42.45	27.24	41.56	38.93	24.65	57.93	11.49	11.26	0.48
4O6Y	57.08	31.92	35.29	46.25	29.25	24.94	17.28	8.09	2.84	0.47
Mean	42.39	32.67	24.60	37.39	28.96	20.92	38.11	18.03	8.96	2.21

E: contact numbers computed using experimental structure; P: contact numbers predicted by neural network; N: no contact numbers; μ_{CR} improved by 5% or more (**bold**) and <5% (*italic*) when folded with predicted CNs.

experimental CNs, P: with predicted CNs, N: without CNs):

- β_{CR} : the highest CR achieved,
- μ_{CR} : the average of the 10 highest CR values,
- π_{20} : the percentage of models with a CR >20%,

β_{CR} and μ_{CR} measure how accurate the best-assembled models can be, whereas π_{20} measures how often an accurate model can be sampled.

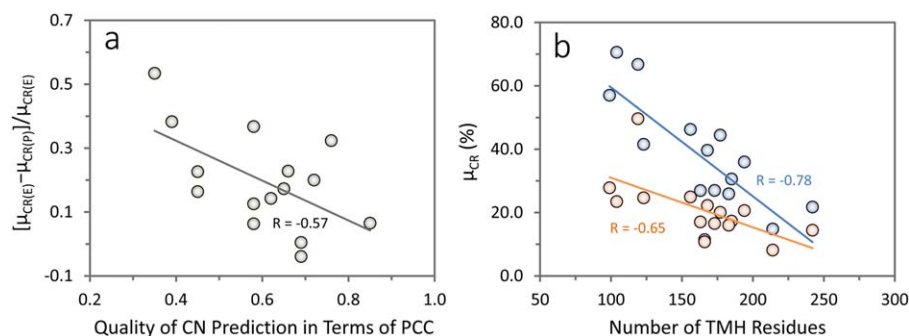
As summarized in Table II, model quality was generally improved using CNs as restraints. Specifically, μ_{CR} was improved for all targets when models were assembled using predicted CNs as restraints, and π_{20} was improved for all but two targets (1OKC and 2O9G). β_{CR} was improved for all targets except 4O6Y and by an average amount of 8.07% and μ_{CR} was improved by an average amount of 8.04% compared to folding without CN restraints. A substantial increase in μ_{CR} (>5%) was seen for 10 of the 15 targets, with 4 of the targets (1PY6, 2K73, 3QAP, and 4A2N) showing >10% of improvement. By using CN restraints, not only the best models were more accurate, but the probability that accurate models were sampled was also increased. For example, comparison of π_{20} among groups shows that π_{20} was increased by 6.75% on average when folded with predicted CNs compared to folded without CNs. It is worth noting that for three targets (2Y01, 3M71, and 3UON), models with CR >20% were not sampled ($\pi_{20} = 0$) without CN restraints but sampled with noticeable frequency with predicted CNs as restraints. Experimental CNs further improved CR, for example, β_{CR} was improved by an average amount of 17.78% when using experimental CNs as restraints. In summary, both experimental and predicted CNs enable strongly significant improvements in CR of folded protein models ($P < 0.01$, paired t test).

Accurate prediction of CNs is not sufficient for improving prediction of TMH rotations

Though a consistent improvement in CR was observed (Table II) when folded with predicted CNs as restraints, the improvement was not as substantial as with experimental CNs. In fact, the higher the PCC of CN prediction is, the closer the μ_{CR} obtained with predicted CNs ($\mu_{CR(P)}$) is to that obtained with experimental CNs ($\mu_{CR(E)}$). This relationship is illustrated by a scatter plot [Fig. 3(a)] of the PCCs of CN prediction and the values of $\frac{\mu_{CR(E)} - \mu_{CR(P)}}{\mu_{CR(E)}}$, which measures the relative difference between $\mu_{CR(E)}$ and $\mu_{CR(P)}$. And the correlation shows that there is still the need to improve the accuracy of CN prediction if one is to make the best of using contact number as restraints.

Intuitively, one might also expect that more accurate prediction of CNs leads to larger relative improvements in CR relative to folding without CNs. However, the correlation between PCCs and the values of $\frac{\mu_{CR(P)} - \mu_{CR(N)}}{\mu_{CR(N)}}$, which measures the relative improvement in $\mu_{CR(P)}$ compared to $\mu_{CR(N)}$, is only very weak (0.28). For instance, $\mu_{CR(P)}$ is improved by 58.63% relative to $\mu_{CR(N)}$ for 2K73 although the accuracy of CN prediction for it is low (PCC: 0.45). Whereas for 2BL2 and 2O9G for which CN predictions are comparably accurate (PCCs are 0.65 and 0.69, respectively), $\mu_{CR(P)}$ is improved by only 11.26% and 10.78% relative to $\mu_{CR(N)}$, respectively. This suggests that other factors besides accurate CN prediction affect improvement in CR.

One intuitive factor is the size of proteins. In fact, as the size of transmembrane domain (measured by the number of TMH residues) increases it becomes more difficult to predict the correct rotation of helices. To illustrate this, the values of $\mu_{CR(E)}$ and $\mu_{CR(N)}$ are plotted

**Figure 3**

(a) Negative correlation ($R = -0.57$) between PCCs of CN prediction and relative differences between $\mu_{CR(E)}$ and $\mu_{CR(P)}$. $\mu_{CR(P)}$ values obtained with better CN predictions is closer to $\mu_{CR(E)}$ than those obtained with poorer CN predictions. (b) μ_{CR} is negatively correlated with number of TMH residues. Orange dots indicate $\mu_{CR(N)}$ values and blue dots indicate $\mu_{CR(E)}$ values. [Color figure can be viewed at wileyonlinelibrary.com]

against number of TMH residues. As shown in Figure 3(b), μ_{CR} is negatively correlated with number of TMH residues ($R = -0.78$ for $\mu_{CR(E)}$ and -0.65 for $\mu_{CR(N)}$). In addition to this negative correlation, improvement in μ_{CR} also becomes less substantial as transmembrane domain becomes larger. This is reflected on the fact that the gap between the two fitted lines shrinks as TMH residues increases. It is also worth noting that $\mu_{CR(N)}$ is below 20% for 7 out of 11 targets with >150 TMH residues, whereas $\mu_{CR(E)}$ is above 20% for all but two targets (1OKC and 2OG9).

Another factor is that some proteins might just represent easy cases whereas others difficult cases for the BCL::MP-Fold algorithm no matter whether CN restraints are incorporated or not. For easy cases, on the one hand, BCL::MP-Fold samples models with high CR even without CN restraints and for them it is difficult to improve substantially upon such a high CR with the current level of accuracy of CN prediction. For example, the membrane rotor of the V-type ATPase 2BL2 whose subunit adopts a four-helical bundle fold³⁶ can be considered an easy case for BCL::MP-Fold. As mentioned previously, its $\mu_{CR(N)}$ is as high as 49.63% even without CN restraints and the relative improvement in CR in terms of μ_{CR} is a comparably low value of 11.26%. For difficult cases on, the other hand, BCL::MP-Fold is not able to sample models with comparably high CR even experimental CN restraints partially due these proteins' intrinsic topological complexity. In this modeling benchmark set, 1OKC and 2OG9 represent such cases as all TMHs of 1OKC are kinked and 2OG9 has two helices located in reentrant regions.^{37,38}

RMSD100 is improved using CN restraints

While the primary motivation to introduce CNs as restraints was to improve prediction of helix rotation, an improvement in RMSD100 was also expected after helix rotations are improved. To verify this, we compared the following parameters among the three simulation groups

(E: with experimental CN, P: with predicted CN, N: without CN):

β_{RMSD100} : the lowest RMSD100 achieved,

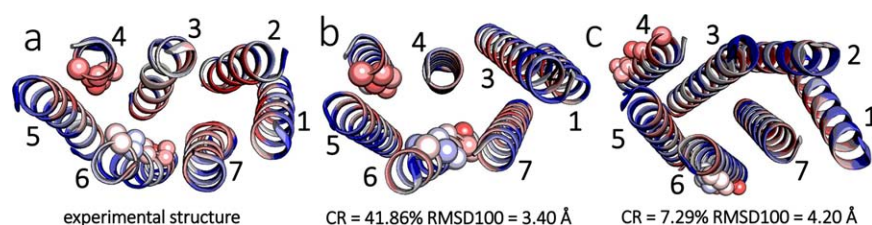
μ_{RMSD100} : the average of the lowest 10 RMSD100 values,

π_5 : the percentage of models with an RMSD100 value lower than 5 Å.

When using experimental CNs as restraints, β_{RMSD100} was decreased for all targets and μ_{RMSD100} was decreased for all but 2OG9 Table III. As discussed in the previous section, the subunit of the tetrameric aquaporin 2OG9 is a special case in that it has two reentrant helices sitting on top of each other.³⁷ When using predicted CNs as restraints, a decrease in β_{RMSD100} is seen for 13 targets and a decrease in μ_{RMSD100} is seen for 12 cases. In terms of μ_{RMSD100} , a decrease of 0.5 Å or more is achieved for 4 targets and the most substantial improvement is a 0.79 Å decrease for 4A2N. Use of predicted CNs yields smaller, albeit still statistically significant ($P < 0.05$, paired t test), improvements to μ_{RMSD100} . It is also interesting to note that models with RMSD100 within 5 Å to experimental structures were assembled with noticeable frequencies for three targets (1PV6, 1U19, and 3UON) when using predicted CNs as restraints, whereas no such models were assembled without CNs as restraints.

Helix rotation accuracy is improved by predicted CN restraints

To visualize the refinement of helix rotation in models with good accuracy at the fold level, experimental CNs were mapped onto the experimental structure and 3D models with the lowest RMSD100 values. Helices with incorrect rotation would have buried residues exposed and exposed residues buried, thus by coloring buried and exposed residues differentially, incorrectly rotated helices in models can be readily identified. 1PY6 was selected, in part because its fold was generally

**Figure 4**

Experimental CNs mapped onto experimental structures and folded models. (a) experimental structure; (b) model with lowest RMSD100 folded with predicted CNs as restraints; (c) model with lowest RMSD100 folded without CN restraints. Color scheme: gradient from blue—fully exposed, red—fully buried. Only TMHs are shown for clarity. Spheres represent C_{α} atoms of buried residues of helices 4 and 6 in the experimental structure.

predicted correctly even without the CN restraints. The CR values of the 1PY6 models with the lowest RMSD100 values are 41.86% and 7.29%, respectively when folded with predicted CNs and without CNs. Without CN restraints, the buried face of TMH4 and that of TMH6 were modeled to exposed. This can be readily seen by comparing the rotation of their buried face with that in the experimental structure [Fig. 4(a,c)]. The incorrect rotation of these two helices disrupts many native contacts between the buried residues of TMH4 and TMH6 (exemplified by red spheres), and likewise, leading to a significantly lower CR. With CN restraints, the rotations of TMH4 and TMH6 were consistent with the experimental structure [Fig. 4(b)].

Increased ability of the scoring function at selecting accurate models

When folded without CN restraints, the average enrichment value over the benchmark set was 1.12. Using

predicted CNs as restraints, enrichment was increased for 14 targets and the average enrichment was improved to 1.64 (Table IV). Paired t test showed that enrichment is improved with statistical significance when folding with predicted CNs ($P < 0.01$). Indeed, enrichment exceeded 1.50 when folding with predicted CNs for eight targets, versus only three targets when folding without CN restraints. Enrichment was improved even further by using experimental CNs as restraints. For example, the average enrichment was increased to 1.92 and 13 targets had enrichment ≥ 1.50 . Because of the intrinsic inaccuracy of the scoring function in the approximation to the potential energy surface, it should be admitted that these relatively low enrichment values are indicative of a difficulty in selecting the most accurate models of the BCL::MP-Fold algorithm.³⁹ Nevertheless, the statistically significant improvement in enrichment indicates that CN restraints provide the scoring function with critical information about residue burial, often corresponding to misrotated helices.

Table III

Summary of RMSD100

Target	β_{RMSD100} (Å)			μ_{RMSD100} (Å)			π_5 (%)		
	<i>E</i>	<i>P</i>	<i>N</i>	<i>E</i>	<i>P</i>	<i>N</i>	<i>E</i>	<i>P</i>	<i>N</i>
1OED	1.88	3.69	3.70	2.08	3.85	3.89	13.47	3.80	4.28
1OKC	10.93	11.73	11.75	11.85	12.25	12.05	0	0	0
1PV6	4.34	4.14	5.09	4.92	4.72	5.49	0.16	0.16	0
1PY6	3.13	3.40	4.20	3.99	4.38	4.70	0.63	0.35	0.22
1U19	3.83	4.44	5.10	5.17	5.42	5.80	0.07	0.04	0
2BL2	2.14	2.36	2.77	2.25	2.84	2.86	11.99	8.14	13.67
2K73	3.01	3.59	3.82	3.06	3.72	4.03	32.44	10.76	7.07
2O9G	10.42	12.21	11.41	12.60	12.72	12.41	0	0	0
2Y01	4.94	5.06	5.26	5.21	5.46	5.76	0.04	0	0
3M71	5.63	5.75	5.94	6.05	6.26	6.36	0	0	0
3QAP	3.33	3.89	4.26	4.25	4.50	4.65	0.56	0.39	0.3
3UG9	3.36	3.24	4.57	3.76	4.19	4.83	1.54	0.77	0.28
3UON	3.70	4.94	5.30	5.17	5.30	5.81	0.13	0.02	0
4A2N	3.51	3.56	4.30	3.94	3.79	4.58	1.28	1.53	0.55
4O6Y	2.71	4.21	3.59	3.36	4.90	4.04	1.04	0.07	0.45
Mean	4.46	5.08	5.40	5.18	5.62	5.82	4.22	1.74	1.79

E: contact numbers computed using experimental structure; *P*: contact numbers predicted by neural network; *N*: no contact numbers; μ_{RMSD100} improved by 0.5 Å or more (**bold**), 0.0–0.5 Å (*italic*), and no improvement (normal) when folded with predicted CNs.

Table IV
Enrichment Achieved With and Without CN Restraints

Target	Enrichment		
	<i>E</i>	<i>P</i>	<i>N</i>
1OED	2.79	1.76	0.88
1OKC	0.51	0.97	0.71
1PV6	1.60	1.27	0.98
1PY6	1.95	1.97	1.16
1U19	1.61	1.29	1.05
2BL2	2.22	2.19	0.43
2K73	2.40	2.43	1.36
2O9G	2.22	1.05	1.08
2Y01	1.50	1.48	1.10
3M71	2.17	1.69	1.45
3QAP	2.20	1.75	1.60
3UG9	2.57	1.67	1.25
3UON	1.43	1.41	0.86
4A2N	1.61	1.89	0.80
4O6Y	1.97	1.80	2.14
Mean	1.92	1.64	1.12

E: contact numbers computed using experimental structure; *P*: contact numbers predicted by neural network; *N*: no contact numbers.

LIMITATIONS AND FUTURE DIRECTIONS

Incorporating the burial status of residues has been shown to improve de novo structure prediction for soluble proteins.^{21,31,40} It is thought that the benefit of incorporating burial status in de novo structure prediction is even larger for HMPs^{41,42} because distinguishing buried from exposed residues in the apolar membrane environment is more challenging for nonspecific scoring functions. Our results indicate that explicit incorporation of CN restraints into the BCL::MP-Fold algorithm significantly improves the prediction of TMH rotations and increases the accuracy of helix–helix packing.

Our results also shows that using experimental CNs as restraints resulted in significantly more improvement in modeling performance than using predicted CNs. This indicates that the performance of the CN predictor BCL::TMH-Expo is an important factor in the BCL::MP-Fold algorithm for HMPs, especially for simple folds such as 1OED. Although using predicted CNs improved modeling outcomes for most targets, we found that accurate prediction of CNs does not guarantee a substantial improvement in CR or RMSD100 for every target. For example, only marginal improvement in CR was seen for 2O9G (Table II) even though its CNs were predicted with high PCC (Table I) and using predicted CNs did not improve RMSD100 for 1OKC or 2O9G (Table III).

1OKC and 2O9G represent intrinsically difficult targets for BCL::MP-Fold and probably for other methods too. The mitochondrial ADP/ATP carrier (1OKC) has its three odd-numbered TMHs kinked substantially by the presence of prolines,³⁸ whereas the aquaporin (2O9G) contains two reentrant regions.³⁷ Tertiary structure

prediction for them was either not benchmarked by methods such as Rosetta-Membrane¹² or Evfold_membrane,¹¹ or proved to be poor with BCL::MP-Fold. BCL::MP-Fold was not able to sample models remotely similar to their experimental structure. The best RMSD100 values for both are >10 Å (Table III). BCL::MP-Fold does not typically accurately represent bent helices. It starts with an idealized, perfectly straight, pool of TMHs. While there are bending moves during the Monte Carlo sampling that bend the TMHs, the current algorithm does not adequately capture the kinks and bends that are commonly seen in native TMHs. This limitation can be overcome with increased probabilities for the bending Monte Carlo moves or more sophisticated bending moves that perturb several ϕ/ψ angles simultaneously by fitting to observed TMH fragments.

CONCLUSIONS

Contact number is a key property of amino acid residues that indicate their local packing density. We have demonstrated that explicitly incorporating contact numbers as restraints into the membrane protein structure prediction algorithm, BCL::MP-Fold, significantly improved prediction of helix–helix packing. Specifically, contact number restraints helped sample more accurate helix rotation and fold, and improved the ability of the scoring function to select native-like models. The relative improvement from using CN restraints was often greatest for proteins with relatively simple folds, though improvements in contact recovery were observed across all proteins in the benchmark set when using predicted CNs. More accurate contact number predictors and structure sampling algorithms that can sample the correct fold of large proteins will be critical to future development of de novo tertiary structure prediction for HMPs.

SOFTWARE AVAILABILITY

BCL::MP-Fold has been integrated into the Biochemical Library (BCL) software suite that is being actively developed. It is available at <http://www.meilerlab.org/bclcommons> under academic and business site licenses. The BCL source code is published under the BCL license and is available at <http://www.meilerlab.org/bclcommons>. Contact numbers can be readily predicted for novel HMPs using BCL::TMH-Expo via its webserver: http://www.meilerlab.org/servers/tmh_expo.

REFERENCES

- 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.
- Overington JP, Al-Lazikani B, Hopkins AL. How many drug targets are there?. *Nat Rev Drug Discov* 2006;5:993–996.

3. Li B, Li W, Du P, Yu KQ, Fu W. Molecular insights into the D1R agonist and D2R/D3R antagonist effects of the natural product (-)-stepholidine: molecular modeling and dynamics simulations. *J Phys Chem B* 2012;116:8121–8130.
4. Zhan C, Li B, Hu L, Wei X, Feng L, Fu W, Lu W. Micelle-based brain-targeted drug delivery enabled by a nicotine acetylcholine receptor ligand. *Angew Chem Int Ed English* 2011;50:5482–5485.
5. Li B, Xu L, Shen Q, Gu X, Fu W. Discovery of novel small-molecule Src kinase inhibitors via a kinase-focused druglikeness rule and structure-based virtual screening. *Mol Simul* 2014;40:341–348.
6. Xiong ZJ, Du P, Li B, Xu LL, Zhen XC, Fu W. Discovery of a Novel 5-HT_{2A} inhibitor by pharmacophore-based virtual screening. *Chem Res Chinese U* 2011;27:655–660.
7. Weiner BE, Woetzel N, Karakas M, Alexander N, Meiler J. BCL::MP-fold: folding membrane proteins through assembly of transmembrane helices. *Structure* 2013;21:1107–1117.
8. Weiner BE, Alexander N, Akin LR, Woetzel N, Karakas M, Meiler J. BCL::Fold—protein topology determination from limited NMR restraints. *Proteins* 2014;82:587–595.
9. Fischer AW, Alexander NS, Woetzel N, Karakas M, Weiner BE, Meiler J. BCL::MP-Fold: membrane protein structure prediction guided by EPR restraints. *Proteins* 2015;83:1947–1962.
10. Cavasotto CN, Phatak SS. Homology modeling in drug discovery: current trends and applications. *Drug Discov Today* 2009;14:676–683.
11. 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.
12. Yarov-Yarovoy V, Schonbrun J, Baker D. Multipass membrane protein structure prediction using Rosetta. *Proteins* 2006;62:1010–1025.
13. Barth P, Wallner B, Baker D. Prediction of membrane protein structures with complex topologies using limited constraints. *Proc Natl Acad Sci USA* 2009;106:1409–1414.
14. Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R, Sander C. Protein 3D structure computed from evolutionary sequence variation. *PLoS One* 2011;6:e28766.
15. Nugent T, Jones DT. Accurate de novo structure prediction of large transmembrane protein domains using fragment-assembly and correlated mutation analysis. *Proc Natl Acad Sci USA* 2012;109:E1540–E1547.
16. Kosciolk T, Jones DT. De novo structure prediction of globular proteins aided by sequence variation-derived contacts. *PLoS One* 2014;9:e92197.
17. Nishikawa K, Ooi T. Prediction of the surface-interior diagram of globular proteins by an empirical method. *Int J Pept Protein Res* 1980;16:19–32.
18. Nishikawa K, Ooi T. Radial locations of amino acid residues in a globular protein: correlation with the sequence. *J Biochem* 1986;100:1043–1047.
19. Kinjo AR, Horimoto K, Nishikawa K. Predicting absolute contact numbers of native protein structure from amino acid sequence. *Proteins* 2005;58:158–165.
20. Lin CP, Huang SW, Lai YL, Yen SC, Shih CH, Lu CH, Huang CC, Hwang JK. Deriving protein dynamical properties from weighted protein contact number. *Proteins* 2008;72:929–935.
21. Durham E, Dorr B, Woetzel N, Staritzbichler R, Meiler J. Solvent accessible surface area approximations for rapid and accurate protein structure prediction. *J Mol Model* 2009;15:1093–1108.
22. Echave J, Spielman SJ, Wilke CO. Causes of evolutionary rate variation among protein sites. *Nat Rev Genet* 2016;17:109–121.
23. Li B, Mendenhall J, Nguyen ED, Weiner BE, Fischer AW, Meiler J. Accurate prediction of contact numbers for multi-spanning helical membrane proteins. *J Chem Inf Model* 2016;56:423–434.
24. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI. OPM: orientations of proteins in membranes database. *Bioinformatics* 2006;22:623–625.
25. Jaakkola T, Diekhans M, Haussler D. Using the Fisher Kernel Method to Detect Remote Protein Homologies. In *Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology*, AAAI Press; 1999. pp 149–158.
26. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 1995;247:536–540.
27. Gribskov M, McLachlan AD, Eisenberg D. Profile analysis: detection of distantly related proteins. *Proc Natl Acad Sci USA* 1987;84:4355–4358.
28. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402.
29. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* 2007;23:1282–1288.
30. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics* 2000;16:404–405.
31. Karakas M, Woetzel N, Staritzbichler R, Alexander N, Weiner BE, Meiler J. BCL::Fold—de novo prediction of complex and large protein topologies by assembly of secondary structure elements. *PLoS One* 2012;7:e49240.
32. Jeffrey LM, Meiler J. Prediction of transmembrane proteins and regions using fourier spectral analysis and advancements in machine learning. Nashville, TN: SERMACS; 2014.
33. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *J Chem Phys* 1953;21:1087–1092.
34. Woetzel N, Karakas M, Staritzbichler R, Muller R, Weiner BE, Meiler J. BCL::Score—knowledge based energy potentials for ranking protein models represented by idealized secondary structure elements. *PLoS One* 2012;7:e49242.
35. Carugo O, Pongor S. A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein Sci Publ Protein Soc* 2001;10:1470–1473.
36. Murata T, Yamato I, Kakinuma Y, Leslie AG, Walker JE. Structure of the rotor of the V-Type Na⁺-ATPase from *Enterococcus hirae*. *Science* 2005;308:654–659.
37. Savage DF, Stroud RM. Structural basis of aquaporin inhibition by mercury. *J Mol Biol* 2007;368:607–617.
38. Pebay-Peyroula E, Dahout-Gonzalez C, Kahn R, Trezeguet V, Lauquin GJ, Brandolin G. Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside. *Nature* 2003;426:39–44.
39. Fischer AW, Heinze S, Putnam DK, Li B, Pino JC, Xia Y, Lopez CF, Meiler J. CASP11—an evaluation of a modular BCL::fold-based protein structure prediction pipeline. *PLoS One* 2016;11:e0152517.
40. Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J Mol Biol* 1997;268:209–225.
41. Adamian L, Liang J. Prediction of transmembrane helix orientation in polytopic membrane proteins. *BMC Struct Biol* 2006;6:13.
42. Park Y, Hayat S, Helms V. Prediction of the burial status of transmembrane residues of helical membrane proteins. *BMC Bioinform* 2007;8:302.