Supplemental Material for:

**PredMP: a web server for *de novo* prediction and visualization of membrane proteins**

Sheng Wang\(^1\),\(^\dagger\),\(^*\), Shiyang Fei\(^3\),\(^\dagger\), Zongan Wang\(^4\),\(^\dagger\), Yu Li\(^1\), Jinbo Xu\(^5\), Feng Zhao\(^2\),\(^*\) and Xin Gao\(^1\),\(^*\)

\(^1\)Computational Bioscience Research Center (CBRC), Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Saudi Arabia.  
\(^2\)Prospect Institute of Fatty Acids and Health, Qingdao University, China.  
\(^3\)COMPASS, New York, USA.  
\(^4\)Department of Chemistry, University of Chicago, USA.  
\(^5\)Toyota Technological Institute at Chicago, USA.

\(^\dagger\)Contribute equally.

\(^*\)To whom correspondence should be addressed: sheng.wang@kaust.edu.sa, fengzhao21c@163.com or xin.gao@kaust.edu.sa

---

**Contents**

S1  Basic workflow of PredMP ................................................................. 2

S2  Dataset of non-redundant membrane proteins ........................................ 3

S3  Transmembrane region prediction by DeepCNF ..................................... 5

S4  Blind test of membrane protein cases in CAMEO .................................. 7

S5  Estimation of the 3D modeling accuracy ............................................. 18

S6  Input/output explanation of the PredMP server ..................................... 19

References ................................................................. 22
S1  Basic workflow of PredMP

Supplementary Figure S1. Illustration of the workflow of PredMP with three modules. Given an input membrane protein sequence, PredMP first uses HHblits [1] to generate the multiple sequence alignment (MSA). The MSA is used to (a) predict transmembrane regions by DeepCNF model, (b) predict secondary structure by RaptorX-Property server [2] through evolutionary analysis (i.e., the 1D annotation module), and (c) predict the contact map through Deep Transfer Learning (DTL) with co-evolutionary features [3]. The predicted secondary structure and contacts are fed into Crystallography & NMR System (CNS) suite [4] to de novo fold the 3D models by RaptorX-Contact server [5] (i.e., the 3D modeling module), which are then embedded into the bilayer membrane with the guide of predicted transmembrane regions and a depth- and residue-dependent membrane burial potential [6] in the visualization module.
**S2  Dataset of non-redundant membrane proteins**

*Supplemental Table S1.* A list of 510 non-redundant membrane proteins with solved structures in Protein Data Bank (PDB) from PDBTM database [7]. The entries highlighted with the bold (bold + underline) font indicate the model with TM-score larger than 0.5 (0.6). The entries shown in blue (italic) indicate the barrel membrane proteins (single-pass helical transmembrane proteins), whereas the others are multi-pass helical transmembrane proteins. Users may refer to the link [http://predmp.com/#/detail/1xxxA](http://predmp.com/#/detail/1xxxA) to check the details of the PredMP predictions, where 1xxx is the membrane protein id (PDB ID: 1xxx plus Chain ID: A).

<table>
<thead>
<tr>
<th>1a0sP</th>
<th>lw4A</th>
<th>2evuA</th>
<th>2ln1A</th>
<th>2wpvB</th>
<th>3cn5A</th>
<th>3kvA</th>
<th>3udcA</th>
<th>4chvA</th>
<th>4il3A</th>
<th>4or2A</th>
<th>4wgvA</th>
<th>5c8jI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1arlB</td>
<td>lq16C</td>
<td>2flcX</td>
<td>21omA</td>
<td>2wsc1</td>
<td>3cx5C</td>
<td>3ll1A</td>
<td>3ug9A</td>
<td>4csekA</td>
<td>4in5H</td>
<td>4p6vB</td>
<td>4wmzA</td>
<td>5cfbA</td>
</tr>
<tr>
<td>IbccE</td>
<td>1g90A</td>
<td>2f93B</td>
<td>21oqA</td>
<td>2wsc3</td>
<td>3d31C</td>
<td>31nmB</td>
<td>3ukmA</td>
<td>4czbA</td>
<td>4in5L</td>
<td>4p6vC</td>
<td>4x5mA</td>
<td>5ctgA</td>
</tr>
<tr>
<td>IbctA</td>
<td>1g90B</td>
<td>2f95B</td>
<td>21ora</td>
<td>2wscF</td>
<td>3dd1A</td>
<td>3lW54</td>
<td>3um7A</td>
<td>4d5bA</td>
<td>4j05A</td>
<td>4p6vD</td>
<td>4xk83</td>
<td>5d0yA</td>
</tr>
<tr>
<td>1bhaA</td>
<td>lqcrD</td>
<td>2fynB</td>
<td>21osa</td>
<td>2wscG</td>
<td>3dhwA</td>
<td>3lW5H</td>
<td>3uq7A</td>
<td>4d6tD</td>
<td>4j72A</td>
<td>4p6vE</td>
<td>4xnakA</td>
<td>5dirA</td>
</tr>
<tr>
<td>1c17M</td>
<td>1qd6C</td>
<td>2ge4A</td>
<td>21ota</td>
<td>2wscH</td>
<td>3dinE</td>
<td>3m71A</td>
<td>3ux4A</td>
<td>4d6tG</td>
<td>4j7ci</td>
<td>4p6vF</td>
<td>4xnvA</td>
<td>5d0qA</td>
</tr>
<tr>
<td>1e7pC</td>
<td>1qleC</td>
<td>2gfpA</td>
<td>2fynB</td>
<td>2losA</td>
<td>2wscF</td>
<td>3dllA</td>
<td>3lw54</td>
<td>3um7A</td>
<td>4d6uD</td>
<td>4k1cA</td>
<td>4pgrA</td>
<td>4xjaA</td>
</tr>
<tr>
<td>1ehkB</td>
<td>1rq1D</td>
<td>2fynB</td>
<td>21ota</td>
<td>2wscH</td>
<td>3dinE</td>
<td>3m71A</td>
<td>3ux4A</td>
<td>4d6tG</td>
<td>4j7ci</td>
<td>4p6vF</td>
<td>4xnvA</td>
<td>5d0qA</td>
</tr>
<tr>
<td>1fftB</td>
<td>1rh5C</td>
<td>2gr8A</td>
<td>2m20A</td>
<td>2wswA</td>
<td>3dwoX</td>
<td>3mk7C</td>
<td>3vmqA</td>
<td>4djiA</td>
<td>4kjrA</td>
<td>4phza</td>
<td>4xyd8</td>
<td>5e0KA</td>
</tr>
<tr>
<td>1fftC</td>
<td>1rwtA</td>
<td>2h8aA</td>
<td>2m67A</td>
<td>2wwbB</td>
<td>3dwwA</td>
<td>3mk7A</td>
<td>3vmqA</td>
<td>4djiA</td>
<td>4kjrA</td>
<td>4phza</td>
<td>4xyd8</td>
<td>5e0KA</td>
</tr>
<tr>
<td>1fw2A</td>
<td>1s51B</td>
<td>2h8pC</td>
<td>2m6bA</td>
<td>2wwbC</td>
<td>3dzmA</td>
<td>3mp7A</td>
<td>3vr8C</td>
<td>4dveA</td>
<td>4kpA</td>
<td>4px7A</td>
<td>4y28G</td>
<td>5eulE</td>
</tr>
<tr>
<td>1fx8A</td>
<td>1s51E</td>
<td>2hdfA</td>
<td>2m7gA</td>
<td>2x4mA</td>
<td>3effk</td>
<td>3mp7B</td>
<td>3vr8D</td>
<td>4dxwa</td>
<td>4kt0F</td>
<td>4q2ea</td>
<td>4y28K</td>
<td>5e0MA</td>
</tr>
<tr>
<td>1gzmA</td>
<td>1s51X</td>
<td>2ibzG</td>
<td>2m8rA</td>
<td>2xq2A</td>
<td>3eh3A</td>
<td>3njtA</td>
<td>3wvwA</td>
<td>4elitA</td>
<td>4ktoK</td>
<td>4qncA</td>
<td>4y28L</td>
<td>5f1cA</td>
</tr>
<tr>
<td>1h2sB</td>
<td>1srqK</td>
<td>2ibzI</td>
<td>2mavA</td>
<td>2xutA</td>
<td>3ejzA</td>
<td>3nymA</td>
<td>3wdoA</td>
<td>4ea3A</td>
<td>4ky0A</td>
<td>4qnda</td>
<td>4y7ja</td>
<td>5fn2B</td>
</tr>
<tr>
<td>1h6s1</td>
<td>1t16A</td>
<td>2ibA</td>
<td>2mfrA</td>
<td>2y5yA</td>
<td>3emnx</td>
<td>3o0rB</td>
<td>3wmaA</td>
<td>4ezCA</td>
<td>4qtnA</td>
<td>4ymkA</td>
<td>5gaeh</td>
<td></td>
</tr>
<tr>
<td>1iz1A</td>
<td>1tl1A</td>
<td>2j58A</td>
<td>2myA</td>
<td>2y69D</td>
<td>3emoA</td>
<td>3op7A</td>
<td>3wmm1</td>
<td>4f35A</td>
<td>4l6v6</td>
<td>4quvA</td>
<td>4y5sc</td>
<td>5gaeq</td>
</tr>
<tr>
<td>1iz1C</td>
<td>1tqqa</td>
<td>2j7aC</td>
<td>2m8A</td>
<td>2y69G</td>
<td>3fhhA</td>
<td>3ohnA</td>
<td>3wmmM</td>
<td>4f41A</td>
<td>4l6v8</td>
<td>4rlia</td>
<td>4ytpC</td>
<td>5garO</td>
</tr>
<tr>
<td>1jb0K</td>
<td>1uuna</td>
<td>2jafA</td>
<td>2mmuA</td>
<td>2y69I</td>
<td>3fida</td>
<td>3orgA</td>
<td>3w07A</td>
<td>4fgeA</td>
<td>4ltoA</td>
<td>4rdqa</td>
<td>4ytpD</td>
<td>5hkiA</td>
</tr>
<tr>
<td>1k24A</td>
<td>1yuynX</td>
<td>2jlnA</td>
<td>2mmnA</td>
<td>2y69J</td>
<td>3g67A</td>
<td>3oufa</td>
<td>3wvfA</td>
<td>4fuva</td>
<td>4m58A</td>
<td>4rfsA</td>
<td>4z34A</td>
<td>5ilMV</td>
</tr>
<tr>
<td>1kf6C</td>
<td>1vclA</td>
<td>2jo1A</td>
<td>2mnpA</td>
<td>2y69K</td>
<td>3gi8C</td>
<td>3p5nA</td>
<td>3wxvA</td>
<td>4gluA</td>
<td>4m64A</td>
<td>4ri2A</td>
<td>4z3nA</td>
<td>5i20A</td>
</tr>
<tr>
<td>1kf6D</td>
<td>1v5fB</td>
<td>2jp3A</td>
<td>2mxbA</td>
<td>2y69L</td>
<td>3h6dA</td>
<td>3pjsK</td>
<td>3x29A</td>
<td>4g7vS</td>
<td>4mbsA</td>
<td>4rjwA</td>
<td>4z7fa</td>
<td>5i32A</td>
</tr>
</tbody>
</table>
S3 Transmembrane region prediction by DeepCNF

To train a machine learning model for predicting the transmembrane region at each residue given a protein primary sequence, we performed the following procedures. We first collected 510 non-redundant transmembrane proteins (shown in Table S1) at the chain level from PDBTM [7]. To label each residue from a given transmembrane protein sequence, we used the following 9 labels extracted from PDBTM: 1 (Side1), 2 (Side2), B (Beta-strand), H (alpha-helix), C (coil), I (membrane-inside), L (membrane-loop), F (interfacial helix), and U (unknown localizations). We then trained a deep learning model, DeepCNF [8, 9], on this annotated sequence dataset.

As shown in Figure S2, DeepCNF has two modules: (i) the Conditional Random Fields (CRF) [10], and (ii) the Deep Convolutional Neural Network (DCNN) [11]. DeepCNF can model not only complex relationship between the sequence and transmembrane regions by a deep hierarchical architecture, but also interdependency between adjacent transmembrane region labels [9]. To deal with the imbalanced distribution of some transmembrane region labels, such as interfacial helix and membrane-inside, we trained DeepCNF by maximizing AUC [6]. According to [9], the DCNN architecture is set as follows: it consists of five layers where each layer has 100 neurons and the window size at each layer is set to 11.

We used the following 68 input features: 20 one-hot encoding from the primary sequence, 20 position specific scoring matrix (PSSM) from PSI-BLAST [12] with E-value threshold 0.001 and three iterations to search UniRef90 [13], 20 PSSM from HHblits [1] with E-value threshold 0.001 and three iterations to search UniProt20 [13], and 8 predicted eight-state secondary structure element by DeepCNF_SS [9]. Note that although we used DeepCNF_SS to generate the predicted secondary structure features for transmembrane region prediction, the training data for DeepCNF_SS only come from non-MPs.

This method achieved 62% cross-validation predictive accuracy on classifying a residue into the nine categories of the transmembrane region. If we merged label B, H, and C as ‘transmembrane region’ label, and all other labels as ‘non-transmembrane region’ label, then this method could achieve 89% predictive accuracy, as well as AUC and AUPRC 0.94 and 0.89, respectively. Finally, using forward-backward algorithm in CRF [10], we assigned to each residue position a reliable ‘transmembrane’ or ‘non-transmembrane’ label based on the computed probability.

It should be noted that other transmembrane region (or, membrane protein topology) predictors could also be used here, such as TOPCONS [14], MEMSAT-SVM [15], PHOBIUS [16], or OCTOPUS [17], just name a few. We will add these third-party tools for predicting and visualizing transmembrane regions in the next release version of PredMP.
Last but not least, this transmembrane region prediction module will be added to RaptorX-Property [2] in the near future. Currently, users may refer to the source code at GitHub https://github.com/realbigws/RaptorX_Property_Fast.

Supplemental Figure S2. Illustration of DeepCNF. Here \( i \) is the position index and \( X_i \) the associated input features, \( H^k \) represents the \( k \)-th hidden layer, and \( Y \) is the output label. All the layers from the 1\(^{st}\) to the \( K \)-th form a deep convolutional neural network (DCNN) with parameter \( W^k(k=1,2,\ldots,K) \), which is shown in blue. The \( K \)-th layer and the label layer form a Conditional Random Fields (CRF), which is shown in red. The parameter \( U \) specifies the relationship between the \( K \)-th layer and the label layer, and \( T \) the binary relationship between adjacent labels. This figure is taken from Wang S. et al. [2].
S4  Blind test of membrane protein cases in CAMEO

Blind and live test in CAMEO

CAMEO [18] can be interpreted as a fully automated CASP [19], but has a smaller number (about 40) of participating servers since many CASP-participating servers are not fully automated. By “blind” it means that the experimentally solved structure of a test protein has not been released in PDB when it is used as a test target. By “live” it means that every weekend CAMEO releases about 20 sequences for prediction test. The test proteins used by CAMEO have no publicly available native structures before it finishes collecting models from servers. The CAMEO server ID of RaptorX-Contact (the main module in PredMP server to generate the 3D models) is Server60, and it has been fully functioning since September 2016.

Since experimentally solving the structures of membrane proteins (MPs) is challenging, starting from September 2016 and up to January 2018, we have observed 10 non-homologous MPs among all CAMEO hard targets, as shown in Table S2.

Supplemental Table S2. A list of 10 non-homologous membrane proteins among all CAMEO hard targets from Sep 2016 to Jan 2018.

- 5h35E (CAMEO ID: 2017-01-07_00000030_3)
- 5jkiA (CAMEO ID: 2017-02-18_00000075_1)
- 510wA (CAMEO ID: 2017-03-18_00000059_2)
- 5khnA (CAMEO ID: 2017-06-10_00000043_1)
- 5kymB (CAMEO ID: 2017-07-22_00000026_1)
- 5mm0A (CAMEO ID: 2017-08-05_00000083_1)
- 5gufA (CAMEO ID: 2017-10-07_00000005_1)
- 5ogkH (CAMEO ID: 2017-11-18_00000021_1)
- 6bmsB (CAMEO ID: 2018-01-06_00000139_1)
- 5vkvA (CAMEO ID: 2018-01-27_00000035_1)

We show in the following sections that RaptorX-Contact successfully modeled all ten MPs belonging to the hard category of CAMEO.
Supplemental Figure S3. Case study of CAMEO target 5h3SE. This protein is an intracellular cation channel ortholog from *Sulfolobus solfataricus*. (A) The long- and medium-range contact prediction accuracy of our methods, MetaPSICOV, and Gremlin. (B-C) The overlap between the native contact map and contact maps predicted by our method, Gremlin, and MetaPSICOV. Top L predicted all-range contacts are displayed. A gray, red, and green dot represents a native contact, a correct prediction, and a wrong prediction, respectively. (D) The superimposition between our predicted model (in red) and the native structure (in blue). (E) The list of top models submitted by CAMEO servers and their quality scores.
Supplemental Figure S4. Case study of CAMEO target SjkiA. This protein is a transmembrane PAP2 type phosphatidylglycerolphosphate phosphatase from *Bacillus subtilis*. (A) The long- and medium-range contact prediction accuracy of our methods, MetaPSICOV, and Gremlin. (B-C) The overlap between the native contact map and contact maps predicted by our method, Gremlin, and MetaPSICOV. Top L predicted all-range contacts are displayed. A gray, red, and green dot represents a native contact, a correct prediction, and a wrong prediction, respectively. (D) The superimposition between our predicted model (in red) and the native structure (in blue). (E) The list of top models submitted by CAMEO servers and their quality scores.
Supplemental Figure S5. Case study of CAMEO target 510wA. This protein is a post-translational translocation Sec71/Sec72 complex from *Escherichia coli*. (A) The long- and medium-range contact prediction accuracy of our methods, MetaPSICOV, and Gremlin. (B-C) The overlap between the native contact map and contact maps predicted by our method, Gremlin, and MetaPSICOV. Top L predicted all-range contacts are displayed. A gray, red, and green dot represents a native contact, a correct prediction, and a wrong prediction, respectively. (D) The superimposition between our predicted model (in red) and the native structure (in blue). (E) The list of top models submitted by CAMEO servers and their quality scores.
Supplemental Figure S6. Case study of CAMEO target 5khnA. This protein is the Burkholderia multivorans hopanoid transporter HpnN. (A) The long- and medium-range contact prediction accuracy of our methods, MetaPSICOV, and Gremlin. (B-C) The overlap between the native contact map and contact maps predicted by our method, Gremlin, and MetaPSICOV. Top L predicted all-range contacts are displayed. A gray, red, and green dot represents a native contact, a correct prediction, and a wrong prediction, respectively. (D) The superimposition between our predicted model (in red) and the native structure (in blue). (E) The list of top models submitted by CAMEO servers and their quality scores.
Supplemental Figure S7. Case study of CAMEO target SkymB. This protein is the 1-acyl-sn-glycerophosphate (LPA) acyltransferase, PlsC, from *Thermotoga maritima*. (A) The long- and medium-range contact prediction accuracy of our methods, MetaPSICOV, and Gremlin. (B-C) The overlap between the native contact map and contact maps predicted by our method, Gremlin, and MetaPSICOV. Top L predicted all-range contacts are displayed. A gray, red, and green dot represents a native contact, a correct prediction, and a wrong prediction, respectively. (D) The superimposition between our predicted model (in red) and the native structure (in blue). (E) The list of top models submitted by CAMEO servers and their quality scores.

<table>
<thead>
<tr>
<th></th>
<th>Long range accuracy</th>
<th></th>
<th></th>
<th></th>
<th>Medium range accuracy</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L</td>
<td>L/2</td>
<td>L/5</td>
<td>L/10</td>
<td>L</td>
<td>L/2</td>
<td>L/5</td>
</tr>
<tr>
<td>Our method</td>
<td>0.773</td>
<td>0.927</td>
<td>0.918</td>
<td>1.000</td>
<td>0.275</td>
<td>0.472</td>
<td>0.673</td>
</tr>
<tr>
<td>metaPSICOV</td>
<td>0.534</td>
<td>0.748</td>
<td>0.939</td>
<td>0.958</td>
<td>0.231</td>
<td>0.398</td>
<td>0.592</td>
</tr>
<tr>
<td>Gremlin</td>
<td>0.563</td>
<td>0.780</td>
<td>0.837</td>
<td>0.917</td>
<td>0.142</td>
<td>0.260</td>
<td>0.551</td>
</tr>
</tbody>
</table>
Supplemental Figure S8. Case study of CAMEO target 5mm0A. This protein is a Dolichyl phosphate mannose synthase. (A) The long- and medium-range contact prediction accuracy of our methods, MetaPSICOV, and Gremlin. (B-C) The overlap between the native contact map and contact maps predicted by our method, Gremlin, and MetaPSICOV. Top L predicted all-range contacts are displayed. A gray, red, and green dot represents a native contact, a correct prediction, and a wrong prediction, respectively. (D) The superimposition between our predicted model (in red) and the native structure (in blue). (E) The list of top models submitted by CAMEO servers and their quality scores.
Supplemental Figure S9. Case study of CAMEO target 5gufA. This protein is a CDP-archaeol synthase (CarS). (A) The long- and medium-range contact prediction accuracy of our methods, MetaPSICOV, and Gremlin. (B-C) The overlap between the native contact map and contact maps predicted by our method, Gremlin, and MetaPSICOV. Top L predicted all-range contacts are displayed. A gray, red, and green dot represents a native contact, a correct prediction, and a wrong prediction, respectively. (D) The superimposition between our predicted model (in red) and the native structure (in blue). (E) The list of top models submitted by CAMEO servers and their quality scores.
Supplemental Figure S10. Case study of CAMEO target SogkH. This protein is a nucleotide sugar
transporter. (A) The long- and medium-range contact prediction accuracy of our methods, MetaPSICOV,
and Gremlin. (B-C) The overlap between the native contact map and contact maps predicted by our
method, Gremlin, and MetaPSICOV. Top L predicted all-range contacts are displayed. A gray, red, and
green dot represents a native contact, a correct prediction, and a wrong prediction, respectively. (D) The
superimposition between our predicted model (in red) and the native structure (in blue). (E) The list of
top models submitted by CAMEO servers and their quality scores.
Supplemental Figure S11. Case study of CAMEO target 6bmsB. This protein is a DHHC (Asp-His-His-Cys) palmitoyltransferases. (A) The long- and medium-range contact prediction accuracy of our methods, MetaPSICOV, and Gremlin. (B-C) The overlap between the native contact map and contact maps predicted by our method, Gremlin, and MetaPSICOV. Top L predicted all-range contacts are displayed. A gray, red, and green dot represents a native contact, a correct prediction, and a wrong prediction, respectively. (D) The superimposition between our predicted model (in red) and the native structure (in blue). (E) The list of top models submitted by CAMEO servers and their quality scores.
**Supplemental Figure S12.** Case study of CAMEO target 5vkvA. This protein is the membrane electron transporter CcdA. (A) The long- and medium-range contact prediction accuracy of our methods, MetaPSICOV, and Gremlin. (B-C) The overlap between the native contact map and contact maps predicted by our method, Gremlin, and MetaPSICOV. Top L predicted all-range contacts are displayed. A gray, red, and green dot represents a native contact, a correct prediction, and a wrong prediction, respectively. (D) The superimposition between our predicted model (in red) and the native structure (in blue). (E) The list of top models submitted by CAMEO servers and their quality scores.
S5  Estimation of the 3D modeling accuracy

We performed a statistical study to show the relationship between 3D model quality and the number of effective sequence homologs (i.e., Meff) using 356 multi-pass helical MPs from the 510 dataset (as shown in Table S1).

We used Meff to measure the amount of homologous information in an MSA (multiple sequence alignment). It can be interpreted as the number of non-redundant (or effective) sequence homologs in an MSA when 70% sequence identity is used as cutoff [20].

We measured the quality of a 3D model by TM-score [21], which ranges from 0 to 1 indicating the worst and the best quality, respectively. A 3D model with TM-score≥0.6 is likely to have a correct fold while a 3D model with TM-score<0.5 usually does not. TM-score = 0.5 is also used by the community as a cutoff to judge if a model has a correct fold or not [22].

Figure S13 shows the TM-score of the 356 MPs with respect to the length-normalized Meff (or, Neff which is defined as Meff/L^{0.7}). When ln(Neff) is larger than 1.5 and 3.5, the predicted models on average have TM-score >0.5 and >0.6, respectively.

![Supplemental Figure S13](image)

Supplemental Figure S13. 3D modeling accuracy of transmembrane proteins (measured by TM-score) with respect to the number of effective sequence homologs in MSA (measured by Neff defined as Meff/L^{0.7}). The blue curve shows the mean and standard deviation at each ln(Neff) bin at 0.5 unit, whereas the red line is a fitted curve of the blue curve.
Input/output explanation of the PredMP server

Input of the PredMP server

Supplemental Figure S14. The only required input of PredMP is the membrane protein sequence. The "Job Submission" section also allows users to provide an email address to be used for notification when the job is done. An email is not required, but strongly recommended since it can be used to retrieve the results of your job.

Output of the PredMP server

The outputs of the PredMP server include:

1) Five full-length de novo constructed 3D models of the input membrane protein sequence. These models are ranked according to the energy function of Crystallography & NMR System (CNS) [4]. These models are then embedded into the bilayer membrane by the Positioning of Proteins in Membranes (PPM) method [23], as shown in Figure S15.

2) Estimated accuracy of the predicted 3D models in three categories: high confidence, medium confidence, and low confidence. The confidence score is calculate based on Neff (defined as Meff/L^{0.7}) which measures the amount of homologous information in the multiple sequence alignment (MSA), as explained in Figure S13.
3) 1D annotation of local structural properties, including the predicted secondary structure, the disordered region, and the transmembrane topology, as illustrated in Figure S16.

Supplemental Figure S15. The result page of the PredMP server for the 3D model prediction followed by the embedding into the bilayer membrane. PredMP will remotely call RaptorX-Contact server to provide five full-length de novo constructed 3D models of the input membrane protein sequence. PredMP also estimates the accuracy of the predicted 3D models in three categories: high confidence (in green), medium confidence (in yellow), and low confidence (in red), respectively.
Supplemental Figure S16. The result page of the PredMP server for the 1D annotation of local structural properties. PredMP will remotely call RaptorX-Property server to provide these local structural properties. Specifically, the upper section shows the summary predicted results, with the first row showing the result of order/disorder regions, and the remaining rows showing the prediction of transmembrane topology, and 3-state secondary structure, respectively. By clicking on a specific summary result, such as the predicted transmembrane topology, the detailed annotation on the input sequence is shown in the lower section.
References

