TMPfold: A Web Tool for Predicting Stability of Transmembrane α-Helix Association

Andrei L. Lomize[1], Kevin A. Schnitzer[2] and Irina D. Pogozheva[1]

1 - Department of Medicinal Chemistry, *College of Pharmacy, University of Michigan, 428 Church St., Ann Arbor, MI 48109-1065, USA*

2 - Department of Electrical Engineering and Computer Science, *College of Engineering, University of Michigan, 1221 Beal Ave, Ann Arbor, MI 48109-2102, USA*

Correspondence to Andrei L. Lomize: *almz@umich.edu*

https://doi.org/10.1016/j.jmb.2019.10.024

Edited by Michael Sternberg

Abstract

Estimating energies of transmembrane (TM) α-helix association is essential for understanding folding of membrane proteins and formation of their functional assemblies. A new physics-based method was developed and implemented in the TMPfold web server for the calculation of the free energy of TM helix association (ΔG_{asc}) in TM α-bundles of known structure. The method was verified using the experimental ΔG_{asc} values for 36 TM complexes, including dimers of 10 glycophorin A mutants. The calculated free energy changes ($\Delta\Delta G_{asc}$) caused by mutations in TM helices correlated with experimental changes in the stability of 42 mutants of bacteriorhodopsin and 25 mutants of rhomboid protease. TMPfold was applied for evaluation of ΔG_{asc} in 554 PDB structures of 85 seven-helical TM proteins and identification of stable two-helical folding intermediates. The proposed tentative paths of cotranslational helix assembly of several polytopic proteins were consistent with experimental studies of their folding. TMPfold is accessible at (https://opm.phar.umich.edu/tmpfold_server).

Introduction

Our understanding of folding and stability of transmembrane (TM) α-helical proteins remains very limited. The folding of multipass (*i.e.*, polytopic) TM proteins has been conceptualized as involving two stages: (1) insertion of TM helices into the lipid bilayer and (2) their lateral association [1,2]. Experimental evidence indicates the existence of thermodynamically controlled folding pathways *in vivo* [3–5]. The cotranslational membrane insertion, translocation, and topogenesis of the nascent polypeptide chain are assisted by TM protein-conducting channels, such as Sec translocon complexes or YidC/Oxa1/Alb3 and their partners [6]. In addition to individual TM α-helices, stable two-helical units (TM α-hairpins or parallel helices) may behave as stable folding units that self-assemble into a native protein structure. At later stages of biogenesis, many membrane proteins assemble into large functional complexes that may be further stabilized by bound ligands, lipids, and cofactors [5,7].

Experimental studies of TM peptides and single-pass (*i.e.* bitopic) TM proteins suggest that the coalescence of TM helices is driven mostly by the Van der Waals (vdW), hydrogen bonding, and ionic interactions between polar side chains, whereas hydrophobic forces are insignificant within the lipid bilayers [8–10]. Though possible models of cotranslational assembly of a few polytopic TM proteins have been proposed [4,7,11,12], the energetics of helix assembly at the second folding stage has not been fully understood and described.

Because of experimental challenges, the folding of membrane proteins is much less explored than that of small water-soluble proteins [11]. Dependence of the folding reaction on the heterogeneous lipid environment adds complexity to the process. Alternatively, computational methods can be beneficial for analysis of cotranslational insertion, folding, and mutational

Please cite this article as: A. L. Lomize, K. A. Schnitzer and I. D. Pogozheva, TMPfold: A Web Tool for Predicting Stability of Transmembrane α-Helix Association, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.10.024

effects in TM proteins. However, the existing computational methods perform poorly in predicting the impact of mutations on TM protein stability [13]. Furthermore, there are no adequate web tools for assessing stability of TM helical intermediates forming at the second stage of the folding process. Therefore, new methods are needed to reproduce available experimental data on conformational energetics and stability of native membrane proteins and their mutants and to improve our understanding of membrane protein folding *in vivo*.

To investigate the energetics of helix-helix interactions in membranes, we have developed a new computational method TMPfold (**TM P**rotein **fold**ing) based on our previously proposed approach [10] and using all-atom empirical energy functions that have been parameterized based on mutagenesis data for water-soluble proteins [14]. The method accounts for the formation of interhelical vdW interactions and H-bonds, solvation, electrostatics, and side-chain conformational entropy changes during helix association in membranes. We tested the method using available experimental data and made it publicly available by developing the TMPfold web server. The server calculates Gibbs free energies of association (ΔG_{asc}) of the individual TM α-helices or multihelical subunits within the lipid bilayer, using three-dimensional (3D) structures of TM proteins as input. Test results indicate that the method can be used for predicting the stability of TM α-helical complexes, including the impact of mutations, and for suggesting plausible folding nuclei.

Results and Discussion

Comparison with experimental stabilities of TM protein complexes

For each specified TM protein structure, the TMPfold web server calculated the following energies: (a) pairwise energies of association of all TM helices in every subunit, (b) total energy of association of all TM helices in each subunit, (c) pairwise energies of association of all TM subunits in a multiprotein complex, and (d) total energy of association of all subunits in the complex. These parameters were calculated and compared for the wild type or mutant proteins. The calculated energies were similar for different high-resolution crystal structures of the same protein or complex (standard deviations for pairwise ΔG_{asc} were less than 2 kcal/mol), but varied for loosely packed lower-resolution structures and NMR models, and alternative conformations.

To assess the accuracy and reliability of the TMPfold method, we tested it against experimental helix association energies of TM α-bundles that were determined in bicelles and micelles for complexes of 22 bitopic TM proteins, two TM peptides and two complexes of polytopic TM proteins (Table S1). The experimental ΔG_{asc} values were reproduced for 16 TM dimers and two tetramers whose 3D structures have been determined in NMR or X-ray studies (Fig. 1A). The slope of the linear regression line was close to 1 ($b = 1.14$) and the R^2 correlation coefficient was 0.81.

We also compared calculated and measured stabilities of parallel homodimers of 19 single-pass TM proteins generated by our TMDOCK web server [15] (Fig. 1B). In this case, the correlation coefficient was similar (R^2 of 0.81), but the slope of the correlation curve was 1.59. Thus, the absolute values of ΔG_{asc} for TMDOCK-generated models were systematically higher than the experimental values, probably because of overpacking of TM helices during local energy minimization with reduced vdW radii.

The root-mean-square error (RMSE) obtained for ΔG_{asc} of the experimental and the TMDOCK-generated structures were 2.34 and 2.42 kcal/mol, respectively. After scaling the calculated values to the experimental values to get the linear regression line with a slope equal to 1, $RMSE_{scal}$ were 2.19 and 1.20 kcal/mol, respectively. The Pearson's correlation coefficients (ρ) between the calculated and the measured helix association energies were 0.91 and 0.89 for experimental structures and TMDOCK models, respectively (Table S3).

Furthermore, using experimental data for dimers of 10 mutants of glycophorin A (Table S2), we tested the ability of TMPfold to reproduce stability of mutated protein complexes. The scatter plot showed a good correlation between calculated and experimental $\Delta\Delta G_{asc}$ values for these dimers with R^2 of 0.67, ρ of 0.85, RMSE of 0.37 kcal/mol, and $RMSE_{scal}$ of 0.32 (Fig. 1C, Table S3).

Changes in stability of polytopic TM proteins upon mutations

Some mutations can affect the stability of polytopic TM proteins. To investigate the impact of mutations on protein stability, we calculated $\Delta\Delta G_{asc}$ values as the differences in the total energy of association of all TM helix calculated for mutants and wild type proteins, using available data for 30 mutants of rhomboid protease GlpG and 52 mutants of bacteriorhodopsin (BRD) (Table S2). We found that the $\Delta\Delta G_{asc}$ values calculated by TMPfold for 25 GlpG mutants and 42 BRD mutants showed a reasonably good correlation with measured $\Delta\Delta G_{exp}$ values (Fig. 1D and E). The corresponding R^2 were 0.43 and 0.41, ρ were 0.66 and 0.70, and $RMSE_{scal}$ were 0.40 and 0.81 kcal/mol, respectively (Table S3).

Importantly, the calculated ΔG_{asc} values represent the difference between the native structure and a hypothetical state with all completely dissociated TM

Please cite this article as: A. L. Lomize, K. A. Schnitzer and I. D. Pogozheva, TMPfold: A Web Tool for Predicting Stability of Transmembrane α-Helix Association, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.10.024

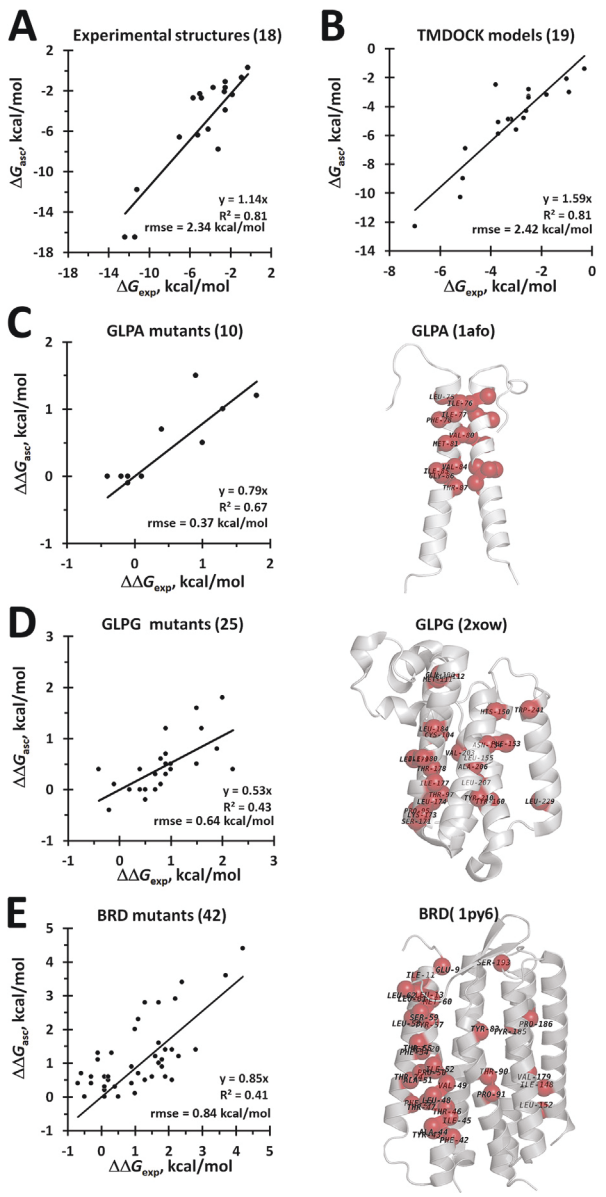

Fig. 1. Correlations between TMPfold-calculated helix association energy and experimentally measured stability of membrane proteins and their mutants. Experimental data were obtained for 18 TM protein dimers and tetramers with known 3D structures (A), 19 TM homodimers of bitopic proteins generated by TMDOCK web server (B), TM homodimers of 10 mutants of glycophorin A (C), 25 mutants of rhomboid protease GlpG (D), and 42 mutants of bacteriorhodopsin (E). Right panels for C–F demonstrate cartoon representations of corresponding proteins with position of Cα-atoms of mutated residues shown by red spheres. Numbers of proteins studied are indicated in parenthesis.

helices. Such a state may not exist in unfolding experiments used for the measurement of $\Delta\Delta G$ values. In the reference state of a protein denatured in detergent, some helices may be aggregated, whereas others may unfold or be located at the surface of micelles to provide the exposure of polar residues to water [5]. Therefore, one can expect a disagreement between the calculated ($\Delta\Delta G_{asc}$) and the experimental ($\Delta\Delta G_{folding}$) values if mutated residues are exposed to water or located in disordered protein segments in the detergent-denatured state. Indeed, we observed several outliers (i.e., data points that deviated by more than two RMSE from the values predicted by the regression line) which represented mutations of polar, charged, or bulky residues. Nevertheless, the obtained correlations indicate that TMPfold can be useful for a crude evaluation of the possible impact of mutations on the stability of polytopic TM proteins.

Comparison with other web servers

Accurate theoretical estimation of protein-protein binding affinity is a challenging problem [16]. To compare the performance of TMPfold with well-known predictors of protein-protein interaction energies, such as PDBePISA [16], and PRODIGY [17], we calculated protein binding energies of 18 TM protein complexes with known structures (Fig. S1) using these three servers. We found that TMPfold outperformed both servers in reproducing ΔG_{exp}. Indeed our method showed ρ of 0.91 between the predicted and measured binding affinities, R^2 of 0.81, and $RMSE_{scal}$ of 2.19 kcal/mol. For the same protein set (Table S4a), PDBePISA and PRODIGY demonstrated ρ of 0.82 and 0.81, R^2 of 0.46 and 0.65, $RMSE_{scal}$ of 7.25 kcal/mol and 2.37 kcal/mol, respectively (Table S4).

The performance of existing methods and web tools for predicting the effect of mutations on the stability of polytopic TM proteins has been recently assessed [13]. It was not surprising that these methods, which were parameterized on data obtained for water-soluble proteins, performed rather poorly in reproducing changes in the stabilities ($\Delta\Delta G$) of membrane protein mutants. To compare the performance of TMPfold with these computational methods, we produced scatter plots for calculated *vs* experimental data for 42 BRD mutants and assessed the R^2, ρ, and RMSE values (Fig. S2, Table S5). Even though our method was not intended to assess the impact of mutations on membrane protein stability, it outperformed all 11 computational methods for assessment of $\Delta\Delta G$ of protein mutants with respect to ρ, R^2, and RMSE values. On the other hand, the applicability of TMPfold is limited by mutations only in TM α-helical segments.

Application of TMPfold to TM proteins of known structure

The application of the server to different quaternary TM protein complexes from the Protein Data

Please cite this article as: A. L. Lomize, K. A. Schnitzer and I. D. Pogozheva, TMPfold: A Web Tool for Predicting Stability of Transmembrane α-Helix Association, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.10.024

Bank [18] helped us to identify complexes with strongly bound subunits. For example, the stability of aquaporin tetramer (PDB ID: 1h6i) was −25.3 kcal/mol, and pairwise association energies of subunits H-M and L-M of bacterial photoreaction centers (PDB ID: 1dxr) were −10.5 and −15.0 kcal/mol, respectively. The majority of complexes in the absence of bound lipids and cofactors had weak to moderate helix association energies of −2 to −8 kcal/mol. Conformational rearrangements of helices may result in significant changes of complex stability. For example, the tetrameric viral M2 channels appeared to be highly stable only in the "closed" conformation (ΔG_{asc} of −11.8 kcal/mol; PDB IF: 6bkk), but the ΔG_{asc} value for the TM part of the channel becomes close to zero in the "open" conformation (−0.5 kcal/mol; PDB ID: 3c9j).

TMPfold was also used for analysis of helix-helix association energies within the same polypeptide chain. As a test case, we compared the calculated energies of helix-helix interactions in 85 seven-TM helical proteins represented by 554 PDB entries: 56 G protein-coupled receptors (GPCRs), 2 adiponectins, 16 microbial and algae rhodopsins, and 11 other proteins. For each protein, we evaluated the free energy of pairwise associations of TM helices and identified stable α-helical units formed by sequential or structurally adjacent helices that may serve as folding nuclei (Fig. 2).

We found that a number of TM helical pairs associate strongly with free energies, ΔG_{asc}, in the range of −5 to −20 kcal/mol. The most stable are TM α-hairpins with left-handed helix arrangement (e.g., in bacteriothodopsin), as well as antiparallel helixes packed through GxxxG-like motifs (e.g., helical pairs "5-6," "6-7" in Ni^{2+}/Co^{2+} transporter). Furthermore, the strong associations of helices are often conserved in homologous proteins. For instance, three TM α-hairpins, "1-2," "2-3," and "6-7," are stable in GPCRs, with an average ΔG_{asc} for 58 proteins (284 structures) of −9.4, −8.7, and −9.9 kcal/mol, respectively (Table S6). Four TM α-hairpins, "1-2," "2-3," "4-5," and "6-7," are highly stable in microbial and algae rhodopsins with an average ΔG_{asc} for 16 proteins (236 structures) of −9.4, −10.0, −12.0, and −10.3 kcal/mol, respectively (Table S7).

The formation of stable TM α-hairpins has been supported by many experimental studies. For example, the highly stable N-terminal "1-2" hairpin of proteorhodopsin (ΔG_{asc} of −11.8 kcal/mol) was seen in the cryoelectron microscopy structure of ribosome-SecY-proteorhodopsin complex (PDB ID: 5ABB) as a folding intermediate located near the SecY lateral gate [19]. TM α-hairpins were observed by NMR in micelles and proposed to serve as folding intermediates of the yeast Ste2p receptor [20]. TM α-hairpins from subunit C of F_1F_0-ATPase appeared to be stable even in a chloroform-methanol-water mixture [21]. Furthermore, single-molecule force microscopy demonstrated that mechanical unfolding of bacteriorhodopsin predominantly occurs in pairs of adjacent helices ("6-7," "4-5," "2-3," 1) [22–24].

Tentative folding pathways and folding intermediates

The analysis of pairwise helix association energies for a polytopic TM protein suggests the tentative path of helix assembly to form the final 3D structure (Fig. 2, Table S8). Helix assembly may start from the formation of strongly interacting pairs of sequential TM α-helices (e.g., α-hairpins). The formation of such folding nuclei may be followed by their mutual association or growth, i.e., one-by-one attachment of

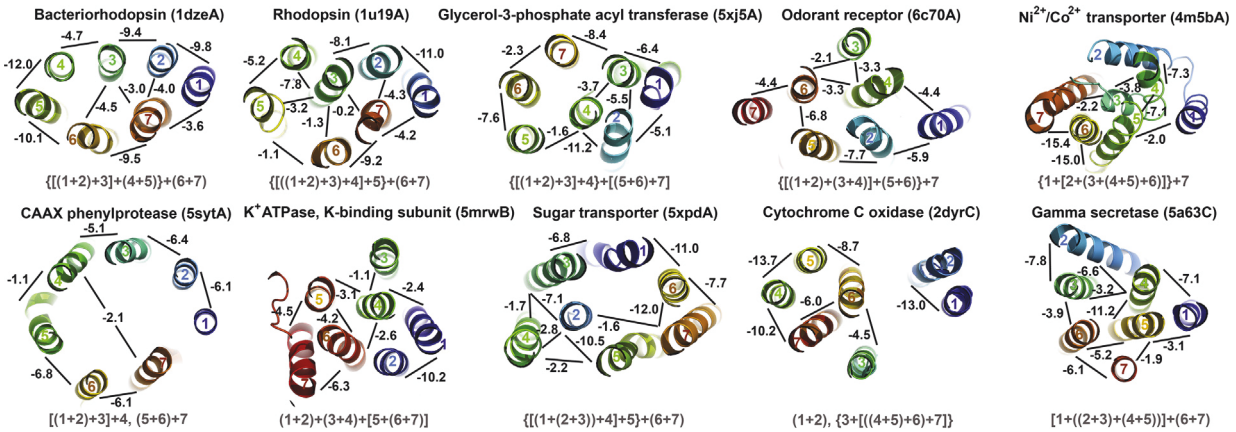

Fig. 2. Diverse folds of seven-TM helical membrane proteins. TM helix arrangements (from the extracellular membrane side) are shown by cartoons (colored rainbow) for representatives of 10 families of 7 TM-helical membrane proteins. Numbers near bars indicate pairwise helix association energies. Plausible "helix assembly paths" are shown below each picture.

sequentially adjacent single TM α-helices or α-hairpins. To illustrate this idea, we selected several TM proteins whose folding was experimentally studied (Table S8). Importantly, the order of association of TM appeared to depend not only on the strength of helix-helix interactions but also on helix stability and hydrophobicity (Table S9).

The possible folding pathway of bovine rhodopsin (PDB ID: 1u19) can be described by the following steps: (1) the formation of the stable "1-2" nucleus at N-terminus; (2) its growth by sequential addition of TM helices 3, 4, and 5; (3) the formation of the "6-7" hairpin; and (4) the coalescence of folding intermediates from steps 2 and 3. The folding intermediates formed by helices "1-5" and "6-7" of rhodopsin identified by TMPfold are supported by the experimentally observed formation of the functional rhodopsin from gene fragments corresponding to TM helical domains "1-5" and "6-7" [25]. Similar folding domains have been detected in other G protein-coupled receptors [26].

TMPfold calculations suggested a similar assembly pathway for bacteriorhodopsin, BRD (PDB ID: 1dze), except the formation of the strong additional "4-5" hairpin. Thus, the N-terminal "1-2" hairpin grows through the addition of the TM3 and the "4-5" hairpin, and then the "1-5" domain may associate with the "6-7" hairpin. Based on the supplemental analysis of stability and hydrophobicity of individual helices (Table S9), we found that the individual TMs 6 and 7 are unstable, so the "6-7" hairpin likely folds and inserts into the membrane during the last steps of the folding pathway, which could be driven by interaction with the "1-5" intermediate. This is consistent with the experimental model of BRD folding [11].

TMPfold predicted identical pathways for six-TM helical aquaporin-1 (AQ1) and aquaporin 4 (AQ4) with sequential immersion of all six TM helices into membrane. This is consistent with the experimental model of the cotranslational folding and assembly of AQ4 [12]. However, the experimental model of the cotranslational folding of AQ1 suggests a more complex process, where at the first stage, only four helices (except the TMs 2 and 4) insert into the membrane, whereas the TM 3 has an inverted orientation. At the second stage, AQ1 undergoes a cotranslational maturation step, where TMs 2 and 4 immerse into the membrane, TM3 flips 180°, and the final six-helical structure is formed [11,12]. An estimation of stability and hydrophobicity of individual TM helices in both aquaporins may explain this difference: TMs 2, 4, and 5 of AQ1 appeared to have much lower stabilities and hydrophobicities than the corresponding TM helices of AQ4.

TMPfold calculations suggested the simple assembly pathways of the six-TM helical rhomboid protease GlpG and of the four-TM helical disulfide oxidoreductase B (DsbB). The proposed helix assembly of both proteins starts from the formation of the stable N-terminal "1-2" hairpin followed by one by one attachment of the subsequently immersed TM helices (Table S8). In accordance with our prediction, the experimental model of the cotranslational folding also suggests the initial formation of the "1-2" hairpin and subsequent sequential assembly of the GlpG helices [4,11]. However, the experimental model suggests that folding of DsbB may occur in two stages: the cotranslational membrane insertion of the TMs 1 and 4; and the posttranslational insertion of the unstable TMs 2 and 3 [4,11]. Our observation that only the TMs 1 and 4 of DsbB are highly stable and hydrophobic (Table S9) may support this model.

Conclusions

We have developed the first public web tool for the prediction of the free energy of TM α-helix association in 3D structures of membrane proteins and their complexes. TMPfold was able to reproduce experimental stability (ΔG) of TM complexes and free energy changes ($\Delta\Delta G$) caused by mutations in TM segments of polytopic proteins, whereas outperforming other computational methods. The server also automatically produced probable assembly pathways of polytopic TM proteins that appeared to be consistent with experimental models of cotranslational folding of several well-studied TM proteins. Importantly, the underlying theoretical approach is physics-based, general, and does not require any training datasets.

We expect that the agreement between the calculated and measured binding energies may be improved in the future by using a more advanced representation of the anisotropic lipid environment and by improving the treatment of residues located at the membrane interface. Furthermore, we plan to advance a procedure for generating protein folding pathways by including the stability and hydrophobicity of individual TM helices, topology-related factors, lengths of connecting loops, and the presence of peripheral amphiphilic helices or folded water-soluble domains. The development of the TMPfold method is a step towards the creation of the energy-based approach to ab initio modeling of TM α-bundles. We anticipate that the availability of this method online will spark the interest of investigators in the fields related to structural biology, membrane proteins, and personal medicine.

Methods

Calculation of free energy of association of TM α-helices or subunits

The TMPfold web server calculates free energy of association of arbitrary TM α-bundles (systems of one or several TM α-helices) using the 3D structure of their

Please cite this article as: A. L. Lomize, K. A. Schnitzer and I. D. Pogozheva, TMPfold: A Web Tool for Predicting Stability of Transmembrane α-Helix Association, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.10.024

complex as input. The approach is based on the previously developed methodology [10] and energy functions [14]. To reproduce experimental free energies of protein binding the following energy contributions were included: (a) environment-dependent energies of vdW interactions and H-bonds; (b) solvation energy changes during transfer of different protein groups from membrane to the protein interior; (c) changes in the conformational entropy of side chains during helix association; and (d) energies of electrostatic interactions of α-helical "macrodipoles" in the nonpolar environment. To account for the conformational flexibility of side chains, we implemented a fast model for averaging contributions of different side-chain conformers for each residue individually, whereas keeping the rest of the structure unchanged. The solvation energy was calculated using a simple model where TM α-helices are immersed into the cyclohexane-like solvent [10], rather than our more advanced anisotropic solvent model of the lipid bilayer [27]. The calculations included only TM segments, without taking into account loops, surface helices, and some cofactors that can be present in experimental structures. Equations for energy calculations are in the Supplemental data section.

Generation of helix assembly maps for polytopic TM proteins

For a specified TM protein structure, TMPfold calculates free energies of association (ΔG_{asc}) for all pairwise combinations of TM helices. Based on this energy matrix and assuming the sequential insertion and association of stable TM helices in membranes, TMPfold produces a tentative assembly order (i.e., pathway) of TM helices during cotranslational protein folding. A simple automatic procedure to determine an assembly "equation" (Fig. 2) uses the following rules: (a) TM α-helices insert into the membrane sequentially; (b) the addition of each helix represents a separate step of the process; (c) the newly synthesized TM α-helix associates with the previously inserted individual TM α-helices or larger nuclei, if the gain in ΔG_{asc} value exceeds a cutoff; and (d) if several different association events satisfy the cutoff, then the sequentially adjacent structures will associate first. The value of the cutoff was taken as −7 kcal/mol; however, it is gradually decreased up to −1 kcal/mol if at least one helix in the subunit remains unattached to the rest of the protein.

Protein 3D structures and modeling of mutations

For verification of the method, 3D structures of TM proteins and their complexes were taken from the Protein Data Bank [18]. They included oligomeric complexes of 16 single-pass TM proteins or TM peptides, complex of sensory rhodopsin with transducer, dimer of Cl^--channel, and two multipass TM proteins, bacteriorhodopsin (BRD) and rhomboid protease GlpG (Tables S1 and S2). Structures of mutants were generated using QUANTA modeling software (Accelrys, Inc) by replacing the mutated residues in the original structure (PDB IDs: 1py6 for BRD, 2xow for GlpG, and 1afo for glycoprotein A dimer) without energy minimization. The experimental structures were used only for three BRD mutants: T46S, A51P, and I148V (PDB IDs: 1s53, 1tn0, and 3har, respectively).

Web server

The TMPfold method was implemented as a public web tool accessible at https://opm.phar.umich.edu/tmpfold_server that can be used for calculating helix association energies (ΔG_{asc}) and of the tentative cotranslational folding pathway for an arbitrary TM α-bundle with known structure. The server uses as input the 3D structure (experimental or modeled) of a polytopic TM α-helical protein or a TM multiprotein complex complemented by the list of TM helical fragments. These fragments can be obtaining by running the PPM web server [28] for a given 3D structure. Inaccurately defined TM segments may affect calculations. Thus, water-accessible polar and charged residues at the ends of α-helices are usually excluded from TM segments.

The output provides (1) the matrix of calculated pairwise helix association energies; (2) the predicted cotranslational assembly pathway for polytopic TM proteins; and (3) the visualization of structures and stabilities of plausible folding intermediates by GLMol. To evaluate the impact of a mutation on the association free energy of TM helices or subunits ($\Delta\Delta G_{asc}$), the user can run the server twice, for native and mutated proteins, respectively, and subtract the total ΔG_{asc} values of the native protein from that of the mutant. A positive $\Delta\Delta G_{asc}$ value would indicate the destabilization of a protein or a protein complex because of a mutation. The calculation for a single molecule requires 1−4 min (single CPU), depending on the molecular size. The results of the calculations are displayed on a web page and can be received *via* email.

The web server was developed using Python3 with the Flask framework and an executable script that runs the corresponding program. The virtual server that hosts our programs and web applications is equipped with Gunicorn, a Python WSGI HTTP Server for UNIX.

Acknowledgment

This work was supported by the Division of Biological Infrastructure of the National Science Foundation [Award #1855425 to A.L., I.P.].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2019.10.024.

Received 26 August 2019;
Received in revised form 21 October 2019;
Accepted 23 October 2019
Available online xxxx

Keywords:
Membrane proteins;
Folding;
Mutation;
Free energy;
Pathway

References

[1] J.L. Popot, D.M. Engelman, Membrane-protein folding and oligomerization - the two-stage model, Biochemistry 29 (1990) 4031–4037.
[2] J.L. Popot, D.M. Engelman, Helical membrane protein folding, stability, and evolution, Annu. Rev. Biochem. 69 (2000) 881–922.
[3] F. Cymer, G. von Heijne, S.H. White, Mechanisms of integral membrane protein insertion and folding, J. Mol. Biol. 427 (2015) 999–1022.
[4] N.J. Harris, E. Reading, K. Ataka, L. Grzegorzewski, K. Charalambous, X. Liu, X. Liu, R. Schlesinger, J. Heberle, P.J. Booth, Structure formation during translocon-unassisted co-translational membrane protein folding, Sci. Rep. 7 (2017) 8021.
[5] J.T. Marinko, H. Huang, W.D. Penn, J.A. Capra, J.P. Schlebach, C.R. Sanders, Folding and misfolding of human membrane proteins in health and disease: from single molecules to cellular proteostasis, Chem. Rev. 119 (2019) 5537–5606.
[6] T. Tsukazaki, Structural basis of the Sec translocon and YidC revealed through X-ray crystallography, Protein J. 38 (2019) 249–261.
[7] J. Neumann, N. Klein, D.E. Otzen, D. Schneider, Folding energetics and oligomerization of polytopic α-helical transmembrane proteins, Arch. Biochem. Biophys. 564 (2014) 2812–2896.
[8] K.R. MacKenzie, D.M. Engelman, Structure-based prediction of the stability of transmembrane helix-helix interactions: the sequence dependence of glycophorin A dimerization, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 3583–3590.
[9] W.F. DeGrado, H. Gratkowski, J.D. Lear, How do helix-helix interactions help determine the folds of membrane proteins? Perspectives from the study of homo-oligomeric helical bundles, Protein Sci. 12 (2003) 647–665.
[10] A.L. Lomize, I.D. Pogozheva, H.I. Mosberg, Quantification of helix-helix binding affinities in micelles and lipid bilayers, Protein Sci. 13 (2004) 2600–2612.
[11] G.A. Pellowe, P.J. Booth, Structural insight into co-translational membrane protein folding, Biochim. Biophys. Acta (2019), https://doi.org/10.1016/j.bbamem.2019.07.007 pii: S0005-2736(19)30157-30159.
[12] M.T. Virkki, N. Agrawal, E. Edsbacker, S. Cristobal, A. Elofsson, A. Kauko, Folding of Aquaporin 1: multiple evidence that helix 3 can shift out of the membrane core, Protein Sci. 23 (2014) 981–992.
[13] B.M. Kroncke, A.M. Duran, J.L. Mendenhall, J. Meiler, J.D. Blume, C.R. Sanders, Documentation of an imperative to improve methods for predicting membrane protein stability, Biochemistry 55 (2016) 5002–5009.
[14] A.L. Lomize, M.Y. Reibarkh, I.D. Pogozheva, Interatomic potentials and solvation parameters from protein engineering data for buried residues, Protein Sci. 11 (2002) 1984–2000.
[15] A.L. Lomize, I.D. Pogozheva, TMDOCK: an energy-based method for modeling alpha-helical dimers in membranes, J. Mol. Biol. 429 (2017) 390–398.
[16] E. Krissinel, Crystal contacts as nature's docking solutions, J. Comput. Chem. 31 (2010) 133–143.
[17] L.C. Xue, J.P. Rodrigues, P.L. Kastritis, A.M. Bonvin, A. Vangone, PRODIGY: a web server for predicting the binding affinity of protein-protein complexes, Bioinformatics 32 (2016) 3676–3678.
[18] S.K. Burley, H.M. Berman, G.J. Kleywegt, J.L. Markley, H. Nakamura, S. Velankar, Protein Data Bank (PDB): the single global macromolecular structure archive, Methods Mol. Biol. 1607 (2017) 627–641.
[19] L. Bischoff, S. Wickles, O. Berninghausen, E.O. van der Sluis, R. Beckmann, Visualization of a polytopic membrane protein during SecY-mediated membrane insertion, Nat. Commun. 5 (2014) 4103.
[20] M. Poms, P. Ansorge, L. Martinez-Gil, S. Jurt, D. Gottstein, K.E. Fracchiolla, L.S. Cohen, P. Güntert, I. Mingarro, F. Naider, O. Zerbe, NMR investigation of structures of G-protein coupled receptor folding intermediates, J. Biol. Chem. 291 (2016) 27170–27186.
[21] D.V. Tulumello, R.M. Johnson, I. Isupov, C.M. Deber, Design, expression, and purification of de novo transmembrane "hairpin" peptides, Biopolymers 98 (2012) 546–556.
[22] D.J. Muller, M. Kessler, F. Oesterhelt, C. Moller, D. Oesterhelt, H. Gaub, Stability of bacteriorhodopsin alpha-helices and loops analyzed by single-molecule force spectroscopy, Biophys. J. 83 (2002) 3578–3588.
[23] H. Yu, M.G.W. Siewny, D.T. Edwards, A.W. Sanders, T.T. Perkins, Hidden dynamics in the unfolding of individual bacteriorhodopsin proteins, Science 355 (2017) 945–950.
[24] R. Petrosyan, C.A. Bippes, S. Walheim, D. Harder, D. Fotiadis, T. Schimmel, D. Alsteens, D.J. Müller, Single-molecule force spectroscopy of membrane proteins from membranes freely spanning across nanoscopic pores, Nano Lett. 15 (2015) 3624–3633.
[25] K.D. Ridge, S.S. Lee, L.L. Yao, In vivo assembly of rhodopsin from expressed polypeptide fragments, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 3204–3208.
[26] T. Gudermann, T. Schoneberg, G. Schultz, Functional and structural complexity of signal transduction via G-protein-coupled receptors, Annu. Rev. Neurosci. 20 (1997) 399–427.
[27] A.L. Lomize, I.D. Pogozheva, H.I. Mosberg, Anisotropic solvent model of the lipid bilayer. 2. Energetics of insertion of small molecules, peptides, and proteins in membranes, J. Chem. Inf. Model. 51 (2011) 930–946.
[28] M.A. Lomize, I.D. Pogozheva, H. Joo, H.I. Mosberg, A.L. Lomize, OPM database and PPM web server: resources for positioning of proteins in membranes, Nucleic Acids Res. 40 (2012) D370–D376.

Please cite this article as: A. L. Lomize, K. A. Schnitzer and I. D. Pogozheva, TMPfold: A Web Tool for Predicting Stability of Transmembrane α-Helix Association, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.10.024

Supplemental Data

TMPfold: a web tool for predicting stability of transmembrane α-helix association

Andrei L. Lomize[†]*, *Kevin A. Schnitzer*[‡], *Irina D. Pogozheva*[†]

[†] Department of Medicinal Chemistry, College of Pharmacy, University of Michigan, 428 Church St., Ann Arbor, MI 48109-1065

[‡] Department of Electrical Engineering and Computer Science, College of Engineering, University of Michigan, 1221 Beal Ave, Ann Arbor, MI 48109-2102

Table of contents

METHOD

The association free energy of the TM α-bundles, ΔG_{asc}, represents a difference between energies calculated for the whole complex and energies of individual α-helices (or helical groups), in accordance with eq. (1):

$$\Delta G_{asc} = \Delta G^{complex} - \sum_{p=1}^{N} \Delta G_p^{Bundle}$$

The energies of the individual α-helices or subunits are calculated as described in eq. (2):

$$\Delta G_p^{Bundle} = \sum_{k=1}^{N} \Delta G_k + E_{helix-electr}$$

where $E_{helix-electr}$ s energy of electrostatic interactions of TM α-helices, ΔG_k is the contribution of residue *k* to stability of the structure, and *N* is number of residues.

The contribution of each residue *k* includes the solvation/transfer energy of its atoms from membrane to the protein interior, interaction energies of its atoms (van der Waals and hydrogen bonds), and side chain conformational entropy of the residue. These interactions are averaged over all side-chain conformers of the residue *k* in accordance with eq. (3) :

$$\Delta G_k = \sum_{m=1}^{M_k} P_m \left(E_m^{transf} + \frac{1}{2} E_m^{inter} - RTlnP_m\right)$$

where P_m, E_m^{transf}, and E_m^{inter} are the occupancy, the transfer energy, and the interaction energy of the conformer *m*, respectively, and M_k is the number of side-chain conformers of residue *k*. The ½ multiplier appears in equation (3) because the interaction energy of residues *k* and *l* will be summed twice.

Transfer energy of residue *k* in conformer *m* from the external environment (water or lipid) to the protein interior is defined by eq. (4):

$$E_m^{transf} = \sum_{i=1}^{I_k} \sigma_i \, ASA_i$$

where σ_i are the corresponding atomic solvation parameters to define the transfer energy of different types of atoms from the lipid to the protein environment (as described previously [1], and ASA_i is the accessible surface area of atom *i*.

Interaction energy of conformer *m* of residue *k* with all surrounding residues is defined by eq. (5):

$$E_m^{inter} = \sum_{i=1, i \neq k}^{N_{contact}} E_{kl}(m)$$

where the interaction energy of residues *k* and *l* is defined by eq. (6):

$$E_{kl}(m) = \sum_{i=1}^{I_k} \sum_{j=1}^{J_l} e_{ij}\,(m)$$

Here, e_{ij} is the interaction energy of atoms *i* and *j* that belong to residues *k* and *l*, respectively, and I_k and J_l are numbers of atoms in residues *k* and *l*. During the calculations, the atomic coordinates of all surrounding residues *l* are taken from the provided 3D structure without any changes.

The conformer occupancies, P_m in equation (3) is defined as follows:

$$P_m = \exp\left(-\frac{\Delta E_m}{RT}\right) / \sum_{m=1}^{M_i} \exp(-\frac{\Delta E_m}{RT}) \quad (7)$$

where ΔE_m is the relative energy of a conformer *m*:

$$\Delta E_m = E_m - E_0 \quad (8)$$

E_0 is the lowest energy, and

$$E_m = E_m^{transf} + E_m^{inter} + E_m^{repuls} \quad (9)$$

The potentials for vdW interactions and H-bonds are calculated as 6-12 and 10-12 functions, respectively, with softened repulsions [2]:

$e_{ij}^{vdW}(m,n) / e^0_{ij} = (r^0_{ij}/r_{ij})^{12} - 2(r^0_{ij}/r_{ij})^6$, for $r_{ij} > r^0_{ij}$ (10)

$e_{ij}^{vdW}(m,n) / e^0_{ij} = (1-r_{ij}/r^0_{ij})^2 - 1$, for $r_{ij} < r^0_{ij}$

for vdW interactions, and

$e_{ij}^{Hb}(m,n) / e^0_{ij} = 5(r^0_{ij}/r_{ij})^{12} - 6(r^0_{ij}/r_{ij})^{10}$, for $r_{ij} > r^0_{ij}$ (11)

$e_{ij}^{Hb}(m,n) / e^0_{ij} = (1-r_{ij}/r^0_{ij})^2 - 1$, for $r_{ij} < r^0_{ij}$

for H-bonds, where r_{ij} is the distance between atoms *i* and *j*, e^0_{ij} is energy at the minimum of the potential, and the equilibrium distances, r^0_{ij}, were taken from ECEPP/2 [3]. As an additional criterion of an A-B…C-D H-bond formation, both A-B…C or B...C-D angles had to be >90°. In this work, we allowed formation of only one possible H-bond for every donor and acceptor group.

The electrostatic interactions between large systems of oriented peptide dipoles in α-helices are calculated with partial atomic charges from CHARMM:

$$E_{helix-electr} = 332 \sum_i \sum_j q_i q_j / \varepsilon r_{ij} \quad (12)$$

where atoms *i* and *j* belong to backbones of different α-helices (including polar hydrogens). The repulsion energy, E_m^{repuls}, is added to exclude all sterically forbidden side-chain conformers. It is calculated similar to E_m^{inter} (equations (11) and (12)), except that the interaction energy of atoms *i* and *j* are defined differently:

$e_{ij}^{repuls} = \rho(r_{ij} - R_0^{vdW})^2$ for $r_{ij} < R_0^{vdW}$ (13)

$e_{ij}^{repuls} = 0$ for $r_{ij} > R_0^{vdW}$

where R_0^{vdW} are radii of Chothia reduced by 0.1 Å, and the weight factor, ρ, is equal to 3 kcal/molÅ^2. The repulsions are softened, because the experimental coordinates of atoms are determined with a precision of ~0.3 A, which may produce significant errors in the calculated energies.

Table S1. Comparison of experimental (ΔG_{exp}) and TMPfold-calculated helix association energies for experimental structures of TM dimers and tetramers ($\Delta G_{calc\text{-}PDB}$) and TMDOCK-generated models of parallel TM homodimers ($\Delta G_{calc\text{-}TMDOCK}$).

Protein Name	Oligomeric state	PDB ID	$\Delta G_{calc\text{-}PDB}$ kcal/mol	$\Delta G_{calc\text{-}TMDOCK}$ kcal/mol	ΔG_{exp} kcal/mol	Ref.[a]
GLPA	2	1afo	-6.60	-12.30	-7.00	[4]
TNR16 C257A	2	2mjo	-0.70	-3.00	-0.90	[5]
VGFR2	2	2m59	-1.70	-3.30	-2.50	[6]
ERBB1	2	2m0b	-1.10	-2.80	-2.50	[7]
ERBB1 M650L/M668I	2	2m20	-3.90	-3.40	-2.50	[7]
ERBB4	2	2l2t	-2.10	-4.30	-2.60	[8]
ERBB3	2	2l9u	0.30	-1.40	-0.30	[9]
EPHA1	2	2k1l	-1.70	-5.90	-3.70	[10]
EPHA2	2	2k9y	-2.30	-6.90	-5.00	[11]
TLR3 (AB)	2	2mka	-2.40	-3.20	-1.80	[12]
FGFR3	2	2lzl	-6.40	-10.30	-5.20	[13]
VGFR G770E	2	2meu	-7.80	-4.90	-3.20	[14]
$(AALALAA)_3$	2	-	-	-2.10	-1.00	[15]
AALALAA-AGLALGA-AALALAA	2	-	-	-9.00	-5.10	[15]
MCP(1-50)	2	-	-	-4.80	-2.70	[16]
MCP-TM	2	-	-	-4.90	-3.30	[16]
MCP-TM M28L/V31L	2	-	-	-2.50	-3.80	[16]
EPO mouse	2	-	-	-5.10	-3.70	[17]
EPO human	2	-	-	-5.60	-3.00	[17]
E5 protein-TM	2	-	-	-11.80	-5.00	[18]
M2 channel	4	6bkk	-11.80	-	-11.20	[19]
M2 channel S31N	4	6mjh	-16.50	-	-12.40	[19]
Integrin α2β3	2	2knc	-2.70	-	-4.84	[20]
Integrin α2β3 A711P	2	2n9y	-2.70	-	-5.66	[20]
Sensory Rhod. II-Transducer	4	5jje	-5.80	-	-4.20	[21]
Cl-channel	2	1ots	-16.50	-	-11.40	[22]

[a] references to publications with experimentally obtained helix association energies

Table S2. Comparison of experimental ($\Delta\Delta G_{exp}$) and TMPfold-calculated changes in helix association energies ($\Delta\Delta G_{calc}$) caused by mutations in glycophorin A (GLPA) dimer, rhomboid protease (GlpG), and bacteriorhodopsin (BRD).

Mutant	PDB ID[a]	ΔG_{exp} kcal/mol	ΔG_{calc} kcal/mol
GLPA dimer (main set: 10)			
L75A	1afo	1.3	1.0
I76A	1afo	1.8	1.2
I77A	1afo	0.1	0.0
F78A	1afo	-0.1	0.0
V80A	1afo	0.4	0.7
M81A	1afo	-0.2	0.0
V84A	1afo	1.0	0.5
I85A	1afo	-0.4	0.0
G86A	1afo	-0.1	-0.1
T87A	1afo	0.9	1.5
GplG (main set: 25)			
P95A	2xow	0.5	-0.2
T97A	2xow	2.0	1.8
C104A	2xow	0.9	0.5
M111A	2xow	0.7	0.0
Q112A	2xow	0.8	0.1
H150A	2xow	0.8	0.1
F153A	2xow	0.7	0.3
N154A	2xow	1.5	1.6
L155A	2xow	1.0	0.4
Y160A	2xow	0.5	0.0
S171A	2xow	-0.1	0.1
L174A	2xow	1.9	0.8
I177A	2xow	2.2	0.4
T178A	2xow	0.9	0.3
L179A	2xow	0.4	0.0
I180A	2xow	0.4	0.4
L184A	2xow	-0.4	0.4
Q190A	2xow	0.2	0.0
V203A	2xow	1.5	0.5
A206G	2xow	1.0	0.5
L207A	2xow	0.9	0.7
Y210A	2xow	1.6	1.2
L229A	2xow	-0.2	-0.4
W241A	2xow	0.9	1.2

K173A	2xow	0.8	0.6
GlpG (outliers: 5)			
M100A	2xow	1.7	0.0
S201A	2xow	0.3	1.8
Y205A	2xow	0.6	2.9
W236A	2xow	0.5	3.4
D268A	2xow	1.1	1.8
BRD (main set: 42)			
E9A	1py6	0.1	0.2
L13A	1py6	1.8	1.1
M20A	1py6	2.8	1.4
T24S	1py6	0.2	0.5
T24V	1py6	-0.3	0.6
F27A	1py6	2.1	0.5
F42A	1py6	2.0	1.2
Y43A	1py6	2.1	1.4
Y43F	1py6	1.7	1.6
Y43P	1py6	-0.1	1.3
A44P	1py6	0.5	0.0
I45A	1py6	1.9	0.6
T46A	1py6	2.2	2.9
T46P	1py6	1.1	2.3
T46S	1s53	0.1	0.6
T47A	1py6	1.1	0.7
T47P	1py6	0.9	0.4
L48A	1py6	0.1	0.3
V49A	1py6	0.3	1.3
P50A	1py6	-0.6	0.7
A51P	1tn0	2.4	3.4
I52A	1py6	1.5	0.6
F54A	1py6	0.4	0.3
T55A	1py6	0.1	0.0
Y57A	1py6	3.7	3.6
L58A	1py6	-0.3	0.4
S59A	1py6	0.1	0.0
M60A	1py6	1.0	0.1
L61A	1py6	-0.7	0.4
L62A	1py6	-0.5	0.0
Y83A	1py6	1.7	2.8
Y83F	1py6	1.0	2.0
T90A	1py6	1.3	2.8

P91A	1py6	1.3	0.5
L111A	1py6	1.7	1.0
I148A	1py6	2.3	1.2
I148V	3har	0.2	0.6
L152A	1py6	1.9	1.4
L174A	1py6	1.8	0.9
Y185A	1py6	4.2	4.4
P186A	1py6	0.9	0.9
S193A	1py6	-0.1	1.1
BRD (outliers: 10)			
T24A	1py6	-0.6	1.2
M56A	1py6	-1.6	0.4
L94A	1py6	3.1	0.9
D96A	1py6	1.5	3.1
L97A	1py6	2.9	0.2
D115A	1py6	-0.5	3.4
Y185F	1py6	0.4	3.2
W189F	1py6	-1.0	2.2
E204A	1py6	1.9	4.0
D212A	1py6	1.2	4.5

[a] PDB IDs of structures used for mutated proteins. Ala-mutants were modeled from original structures of GlpG (PDB ID: 2xow), BRD (PDB ID: 1py6), and GLPA dimer (PDB ID: 1afo). Experimental data are taken from compilation in reference [23].

Table S3. Correlations between the TMPfold-calculated helix association energies and experimentally measured stabilities of TM complexes and mutants of glycophorin A (GLPA) dimer, rhomboid protease (GlpG), and bacteriorhodopsin (BRD).

Plot	Equation	n	R^2	RMSE, kcal/mol	$RMSE_{scal}$[a], kcal/mol	ρ [b]
ΔG_{calc_PDB}	$\Delta G_{calc\_PDB} = 1.14\ \Delta G_{exp}$	18	0.81	2.34	2.19	0.91
ΔG_{calc_TMDOCK}	$\Delta G_{calc\_TMDOCK} = 1.59\ \Delta G_{exp}$	19	0.81	2.42	1.20	0.89
$\Delta\Delta G_{calc\_GLPA}$	$\Delta\Delta G_{calc\_GLPA} = 0.79\ \Delta\Delta G_{exp\_GLPA}$	10	0.67	0.37	0.32	0.85
$\Delta\Delta G_{calc\_GplG}$	$\Delta\Delta G_{calc\_GlpG} = 0.53\ \Delta\Delta G_{exp\_GlpG}$	25	0.43	0.64	0.40	0.66
$\Delta\Delta G_{calc\_BRD}$	$\Delta\Delta G_{calc\_BRD} = 0.85\ \Delta\Delta G_{exp\_BRD}$	42	0.41	0.85	0.81	0.70
$\Delta\Delta G_{calc\_ALL\_MUT}$	$\Delta\Delta G_{calc\_ALL\_MUT} = 0.77\ \Delta\Delta G_{exp\_ALL\_MUT}$	77	0.47	0.74	0.65	0.70

[a] The RMSE values after scaling the calculated values to the experimental values to get the linear regression with slop at 1.

[b] The Pearson's correlation coefficient.

Table S4. Comparison of the performance of TMPfold with PDBePISA [24] and PRODIGY [25] in prediction of protein-protein binding energy for the set of 18 TM protein complexes (see Table S3a below).

Method	R^2	RMSE, kcal/mol	$RMSE_{scal}$[a], kcal/mol	ρ [b]
TMPfold	0.81	2.34	2.19	0.91
PDBePISA	0.46	15.95	7.25	0.82
PRODIGY	0.65	2.39	2.37	0.81

[a] The RMSE values after scaling the calculated values to the experimental values to get the linear regression with slop at 1.
[b] The Pearson's correlation coefficient.

Table S4a. Set of 18 TM protein complexes used for comparing the performance of TMPfold with PDBePISA and PRODIGY.

Protein Name	Oligomeric state	PDB ID	$\Delta G_{calc\text{-}TMPfold}$ kcal/mol	$\Delta G_{calc\text{-}PDBePISA}$ kcal/mol	$\Delta G_{calc\text{-}PRODIGY}$ kcal/mol	ΔG_{exp} kcal/mol	Ref.[a]
GLPA	2	1afo	-6.60	-12.60	-3.90	-7.00	[4]
TNR16 C257A	2	2mjo	-0.70	-9.50	-4.50	-0.90	[5]
VGFR2	2	2m59	-1.70	-15.40	-1.40	-2.50	[6]
ERBB1	2	2m0b	-1.10	-17.70	-3.60	-2.50	[7]
ERBB1 M650L/M668I	2	2m20	-3.90	-12.40	-3.30	-2.50	[7]
ERBB4	2	2l2t	-2.10	-13.70	-2.20	-2.60	[8]
ERBB3	2	2l9u	0.30	-13.40	-4.10	-0.30	[9]
EPHA1	2	2k1l	-1.70	-10.20	-0.30	-3.70	[10]
EPHA2	2	2k9y	-2.30	-12.00	-1.40	-5.00	[11]
TLR3 (AB)	2	2mka	-2.40	-9.40	-2.20	-1.80	[12]
FGFR3	2	2lzl	-6.40	-20.10	-2.90	-5.20	[13]
VGFR G770E	2	2meu	-7.80	-13.90	-4.90	-3.20	[14]
M2 channel	4	6bkk	-11.80	-36.40	-10.20	-11.20	[19]
M2 channel S31N	4	6mjh	-16.50	-32.40	-15.20	-12.40	[19]
Integrin α2β3	2	2knc	-2.70	-15.90	-2.50	-4.84	[20]
Integrin α2β3 A711P	2	2n9y	-2.70	-22.50	-2.40	-5.66	[20]
Sensory Rhod. II-Transducer	4	5jje	-5.80	-30.50	-4.30	-4.20	[21]
Cl-channel	2	1ots	-16.50	-44.00	-13.60	-11.40	[22]

[a] references to publications with experimentally obtained helix association energies

Table S5. Comparison of the performance of TMPfold with 11 computational methods for predicting changes in membrane protein stability ($\Delta\Delta G$) due to mutations.

Method	Equation [a]	R^2 [a]	RMSE, kcal/mol	Equation [b]	R^2 [b]	ρ [c]
TMPfold	$\Delta\Delta G_{asc} = 0.66\Delta\Delta G_{exp} + 0.43$	0.49	0.84	$\Delta\Delta G_{asc} = 0.85\Delta\Delta G_{exp}$	0.41	0.91
PROVEAN	$\Delta\Delta G_{calc} = 1.16\Delta\Delta G_{exp} + 1.36$	0.34	2.38	$\Delta\Delta G_{calc} = 1.76\Delta\Delta G_{exp}$	0.15	0.59
PPSC (M8)	$\Delta\Delta G_{calc} = 0.73\Delta\Delta G_{exp} + 1.31$	0.34	1.57	$\Delta\Delta G_{calc} = 1.31\Delta\Delta G_{exp}$	-0.10	0.58
Rosetta[d]	$\Delta\Delta G_{calc} = 0.93\Delta\Delta G_{exp} + 0.98$	0.27	1.88	$\Delta\Delta G_{calc} = 1.25\Delta\Delta G_{exp}$	0.13	0.52
I Mutant 3.0	$\Delta\Delta G_{calc} = 0.32\Delta\Delta G_{exp} + 1.00$	0.27	1.01	$\Delta\Delta G_{calc} = 0.76\Delta\Delta G_{exp}$	-0.80	0.52
ELASPIC	$\Delta\Delta G_{calc} = 0.36\Delta\Delta G_{exp} + 0.34$	0.23	1.10	$\Delta\Delta G_{calc} = 0.50\Delta\Delta G_{exp}$	0.15	0.48
EASE MM	$\Delta\Delta G_{calc} = 0.34\Delta\Delta G_{exp} + 0.97$	0.22	1.06	$\Delta\Delta G_{calc} = 0.77\Delta\Delta G_{exp}$	-0.53	0.47
mCSM	$\Delta\Delta G_{calc} = 0.27\Delta\Delta G_{exp} + 1.60$	0.16	1.38	$\Delta\Delta G_{calc} = 0.96\Delta\Delta G_{exp}$	-1.90	0.30
FoldX	$\Delta\Delta G_{calc} = 0.55\Delta\Delta G_{exp} + 1.82$	0.11	2.27	$\Delta\Delta G_{calc} = 1.36\Delta\Delta G_{exp}$	-0.38	0.33
DUET	$\Delta\Delta G_{calc} = 0.25\Delta\Delta G_{exp} + 1.48$	0.07	1.52	$\Delta\Delta G_{calc} = 0.91\Delta\Delta G_{exp}$	-0.84	0.26
SDM	$\Delta\Delta G_{calc} = 0.40\Delta\Delta G_{exp} - 0.07$	0.06	2.05	$\Delta\Delta G_{calc} = 0.36\Delta\Delta G_{exp}$	0.06	0.24
PPSC (M47)	$\Delta\Delta G_{calc} = 0.14\Delta\Delta G_{exp} + 1.47$	0.05	1.31	$\Delta\Delta G_{calc} = 0.79\Delta\Delta G_{exp}$	-2.11	0.22

Experimental ($\Delta\Delta G_{exo}$) and TMPfold-calculated values ($\Delta\Delta G_{calc}$) of change of protein stability for 42 bacteriorhodopsin (BRD) mutants are collected in Table S2. The $\Delta\Delta G_{calc}$ values obtained using 11 computational methods are collected in the Table S1_ddg_5_26_2018 of the publication by Kroncke et al. [23].

[a] Equations and correlation coefficients (R^2) for linear regression lines that do not cross the intercept (x=0, y=0)

[b] Equations and correlation coefficients (R^2) or linear regression lines that cross the intercept (x=0, y=0)

[c] The Pearson's correlation coefficient (ρ) between calculated and measured binding affinity on the set of 42 BRD mutants, except for the Rosetta method (37 mutants).

[d] The $\Delta\Delta G_{calc}$ values were calculated by Rosetta for 37 BRD mutants, excluding Y43P, A44P, T46P, T47P, A51P.

Table S6. Free energy (ΔG_{asc}) of pairwise TM α-helix association calculated by TMPfold for G protein-coupled receptors (GPCRs) and adiponectin.

GPCRs	Num.of PDBs	Num.of proteins	ΔG_{asc} for pairs of TM helices										Total ΔG_{asc}
			1&2	1&7	2&3	2&4	2&7	3&4	3&6	4&5	5&6	6&7	
Family A	246	45	-9.7	-4.3	-8.5	-0.6	-5.6	-9.1	-1.1	-2.8	-6.3	-9.8	-60.9
Family B	14	7	-5.7	-7.5	-12.1	-2.6	-9.6	-3.0	-2.5	-2.8	-4.0	-7.8	-57.4
Family C	5	2	-5.1	1.5	-12.0	-6.9	-9.3	-5.5	-1.9	-6.9	-6.0	-7.0	-59.8
Family F	14	2	-7.4	-8.9	-7.2	-5.0	-8.8	-4.1	-2.2	0.3	-5.5	-13.9	-62.0
Adiponectin	5	2	-13.8	-8.8	-9.6	-4.8	-7.4	0.2	-2.2	-3.5	-6.0	-11.7	-68.1
ALL	284	58	-9.4	-4.7	-8.7	-1.1	-6.0	-8.3	-1.2	-2.7	-6.1	-9.9	-60.9

Average values of ΔG_{asc} (kcal/mol) are shown for each protein family.
Stable TM α-hairpins that are conserved for all GPCR-type proteins are shadowed gray.

Table S7. Free energy (ΔG_{asc}) of pairwise TM α-helix association calculated by TMPfold for microbial and algae rhodopsins.

Protein name	Num.of PDBs	Num.of proteins	ΔG_{asc} for pairs of TM helices										Total ΔG_{asc}
			1&2	1&7	2&3	2&7	3&4	3&6	3&7	4&5	5&6	6&7	
Archaerhodopsin1	1	1	-14.4	0.0	-9.4	-3.8	-6.2	-2.5	-5.3	-11.6	-5.1	-9.0	-67.1
Archaerhodopsin2	4	1	-10.9	0.1	-8.1	-3.4	-6.0	-2.7	-2.9	-12.2	-4.5	-5.9	-56.1
Bacteriorhodopsin	139	1	-9.3	-1.9	-9.6	-3.6	-4.3	-3.9	-2.6	-13.1	-9.9	-10.6	-68.6
Halorhodopsin	20	1	-4.8	-8.3	-10.9	-6.7	-6.7	-2.6	-6.3	-13.5	-5.5	-7.6	-72.8
Cruxrhodopsin 3	2	1	-8.5	-0.6	-17.7	-3.7	-6.6	-2.8	-1.8	-12.2	-9.2	-16.7	-80.2
Deltarhodopsin	1	1	-12.3	-2.6	-9.1	-3.5	-4.6	-3.8	-1.6	-14.7	-9.7	-21.2	-83.1
Proteorhodopsin	7	1	-13.6	-1.9	-12.6	-5.5	-8.3	-1.6	-9.9	-6.2	-8.6	-12.4	-81.9
Rhodopsin I	2	1	-15.3	-0.4	-8.2	-5.5	-6.3	-2.9	-4.9	-6.7	-6.6	-7.9	-64.2
Xanthorhodopsin	1	1	-6.7	0.9	-12.3	-5.5	-7.9	-0.7	-6.6	-10.7	-8.4	-8.2	-66.9
Blue-light proteorhodopsin	7	1	-12.6	-2.5	-11.9	-6.1	-7.6	-1.2	-9.2	-6.6	-7.1	-12.4	-77.6
Sensory rhodopsin II	23	1	-11.9	-0.1	-6.4	-3.1	-8.1	-1.3	-3.6	-7.0	-4.6	-7.6	-54.6
Channel rhodopsin 1,2	4	1	-8.5	-3.9	-7.4	-7.0	-4.8	-1.9	-5.1	-19.0	-4.8	-9.8	-70.6
Chloride pump rhodopsin	7	1	-9.9	0.8	-16.2	-4.7	-9.5	-1.1	-7.1	-10.7	-5.3	-11.3	-76.2
Proton pump rhodopsin 2	1	1	-9.4	-0.7	-4.6	-5.8	-5.4	-5.1	-3.0	-4.9	-7.2	-8.2	-53.8
Light-gated anion channel	3	1	-8.6	-6.5	-5.8	-7.8	-4.6	-1.6	-5.9	-14.9	-5.9	-9.6	-70.6
Sodium pump rhodopsin, NaR	14	1	-8.2	0.5	-14.9	-2.0	-7.4	-1.0	-6.1	-12.7	-4.5	-14.0	-75.0
ALL	236	16	-9.4	-2.1	-10.0	-4.0	-5.5	-3.0	-3.9	-12.0	-8.1	-10.3	-68.8

Average values of ΔG_{asc} (kcal/mol) are shown for each protein.

Stable TM α-hairpins that are conserved for all GPCR-type proteins are shadowed gray.

Table S8. Prediction by TMPfold of stable TM α-hairpins, total association energy of TM α-bundles (kcal/mol) and tentative assembly pathway of selected polytopic TM proteins.

Protein name	PDB ID	ΔG_{asc} for pairs of TM helices								Helix assembly	Total ΔG_{asc}
		1&2	**2&3**	**3&4**	**4&5**	**5&6**	**6&7**				
RH_{bov}	**1u19**	-11.0	-8.1	-7.8	-5.2	-1.1	-9.2			(((((1+2)+3)+4)+5)+(6+7))	-62.5
BRD	**1dze**	-9.8	-9.4	-4.7	-12.0	-10.1	-9.5			((((1+2)+3)+(4+5))+(6+7))	-69.4
		1&2	**2&3**	**3&4**	**4&5**	**5&6**	**1&3**	**2&5**	**4&6**		
$AQ1_{bov}$	**4j4n**	-7.1	-0.6	-0.4	-9.3	-0.6	-4.1	-11.3	-5.1	(((1+2)+3)+((4+5)+6))	-37.9
$AQ4_{rat}$	**2d57**	-5.1	-0.6	-0.4	-5.2	-0.8	-5.3	-9.5	-7.8	(((1+2)+3)+((4+5)+6))	-34.4
		1&2	**2&3**	**3&4**	**4&5**	**5&6**	**3&6**	**4&6**			
GlpG	**2xow**	-12.5	-4.9	-2.7	-2.0	-1.8	-8.8	-8.1		(((((1+2)+3)+4)+5)+6)	-47.1
		1&2	**2&3**	**3&4**	**2&3**						
DsbB	**2hi7**	-7.2	-4.1	-0.7	-4.1					((1+2)+3),4	-11.1

The TM α-hairpins with low pairwise helix association energy, ΔG_{asc} <- 5 kcal/mol, we defined as stable folding nuclei (shadowed gray). Tentative helix assembly paths were automatically produced based on sequential formation of stable nuclei and subsequent grow of the TM α-bundle by adding individual helices or helical pairs. Selected polytopic TM proteins include: bovine rhodopsin (RH_{bov}), bacteriorhodopsin (BRD), bovine aquaporin-1 ($AQ1_{bov}$), rat aquaporin-4 ($AQ4_{rat}$), rhomboid protease (GlpG), and disulfide oxidoreductase B (DsbB).

Table S9. Stability (ΔG_{hel}) and hydrophobicity (ΔG_{transf}) of individual TM α-helices of polytopic TM proteins calculated by FMAP.

Name	PDB ID	TM1		TM2		TM3		TM4		TM5		TM6		TM7	
		ΔG_{hel}	ΔG_{transf}	ΔG_{hel}	ΔG_{transf}	ΔG_{hel}	ΔG_{transf}	ΔG_{hel}	ΔG_{transf}	ΔG_{hel}	ΔG_{transf}	ΔG_{hel}	ΔG_{transf}	ΔG_{hel}	ΔG_{transf}
RH_{bov}	**1u19**	-26.0	-27.0	-15.9	-26.9	-8.7	-18.5	-13.5	-20.1	-25.7	-28.0	-23.5	-28.5	-	-
BRD	**1dze**	-17.7	-18.5	-19.2	-24.7	-	-	-5.9	-18.6	-23.9	-27.9	-	-	-	-
$AQ1_{bov}$	**4j4n**	-23.3	-30.5	-5.1	-11.9	-12.5	-19.5	-6.6	-14.5	-4.6	-11.7	-20.5	-21.8		
$AQ4_{rat}$	**2d57**	-16.7	-25.7	-10.0	-15.5	-9.2	-15.4	-13.3	-21.4	-11.4	-13.9	-12.8	-16.0		
GlpG	**2xow**	-17.4	-20.5	-21.2	-19.2	-9.6	-16.6	-	-	-18.8	-24.1	-9.8	-12.3		
DsbB	**2hi7**	-21.6	-20.6	-7.6	-15.5	-	-	-22.5	-25.7						

[a] the extended helix that includes both TM3 and TM4.

' – ' indicates the unstable TM helices. Unstable helices and TM with decreased stability and hydrophobicity are shadowed gray.

ΔG_{hel} and ΔG_{transf} of individual TM α-helices of polytopic TM proteins were calculated by the FMAP web server (https://membranome.org/fmap "Transmembrane protein" option).

Selected polytopic TM proteins include: bovine rhodopisin (RH_{bov}), bacteriorhodopin (BRD), bovine aquaporin-1 ($AQ1_{bov}$), rat aquaporin-4 ($AQ4_{rat}$), rhomboid protease (GlpG), and disulfide oxidoreductase B (DsbB).

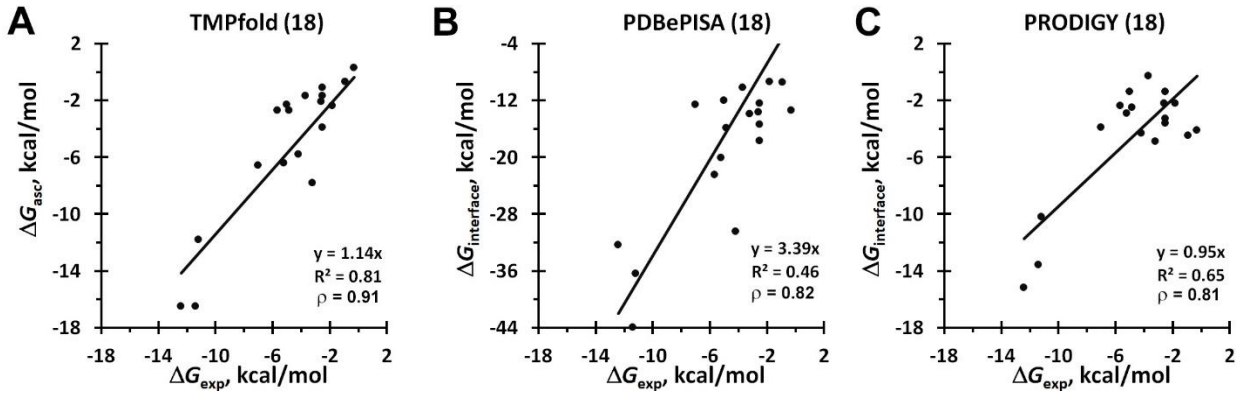

Figure S1. Comparison of the performance of TMPfold (A) with PDBePISA (B), and PRODIGY (C) in prediction of the protein-protein binding energy for 18 protein complexes. The PDBePISA web server [24] was operating in the "interfaces" mode, the PRODIGY web server [25] was running for protein-protein complexes (25° C). Experimental and calculated values of protein-protein binding energies are collected in Table S3.

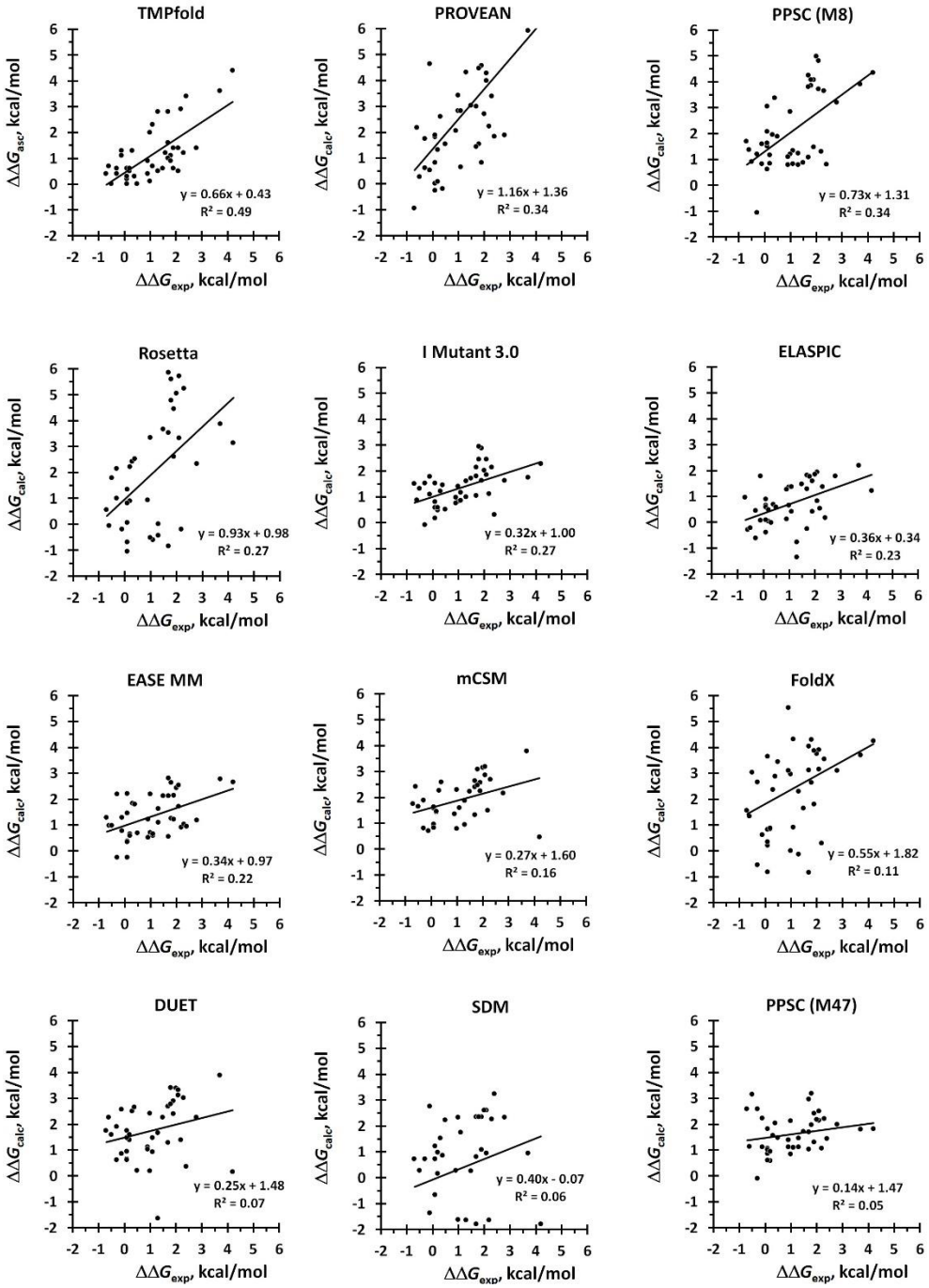

Figure S2. Comparison of the performance of the TMPfold method with 11 computational methods for predicting stability of membrane proteins and their mutants that have been assessed by Kroncke et al. [23]. Calculated values ($\Delta\Delta G_{calc}$) were plotted against experimental values obtained for 42 BRD mutants ($\Delta\Delta G_{exp}$ are in Table S2). The Rosetta method was applied for 37 mutants, excluding five X→Pro mutants. Linear regression line calculated for the TMPfold plot (upper left panel) showed a better correlation coefficient (R^2=0.49) than plots for other computational methods.

REFERENCES

[1] A.L. Lomize, I.D. Pogozheva, H.I. Mosberg, Quantification of helix-helix binding affinities in micelles and lipid bilayers, Protein Sci. 13 (2004) 2600-2612.

[2] A.L. Lomize, M.Y. Reibarkh, I.D. Pogozheva, Interatomic potentials and solvation parameters from protein engineering data for buried residues, Protein Sci. 11 (2002) 1984-2000.

[3] F.A. Momany, L.M. Carruthers, H.A. Scheraga, Intermolecular potentials from crystal data. IV. Application of empirical potentials to the packing configurations and lattice energies in crystals of amino acids, J Phys Chem. 78 (1974) 1621-1630.

[4] K.G. Fleming, Standardizing the free energy change of transmembrane helix-helix interactions, J Mol Biol. 323 (2002) 563-571.

[5] K.D. Nadezhdin, I. Garcia-Carpio, S.A. Goncharuk, K.S. Mineev, A.S. Arseniev, M. Vilar, Structural basis of p75 transmembrane domain dimerization, J Biol Chem. 291 (2016) 12346-12357.

[6] K.S. Mineev, D.M. Lesovoy, D.R. Usmanova, S.A. Goncharuk, M.A. Shulepko, E.N. Lyukmanova, M.P. Kirpichnikov, E.V. Bocharov, A.S. Arseniev, NMR-based approach to measure the free energy of transmembrane helix-helix interactions, Biochim Biophys Acta. 1838 (2014) 164-172.

[7] E.V. Bocharov, D.M. Lesovoy, K.V. Pavlov, Y.E. Pustovalova, O.V. Bocharova, A.S. Arseniev, Alternative packing of EGFR transmembrane domain suggests that protein-lipid interactions underlie signal conduction across membrane, Biochim Biophys Acta. 1858 (2016) 1254-1261.

[8] E.V. Bocharov, K.S. Mineev, M.V. Goncharuk, A.S. Arseniev, Structural and thermodynamic insight into the process of "weak" dimerization of the ErbB4 transmembrane domain by solution NMR, Biochim Biophys Acta. 1818 (2012) 2158-2170.

[9] K.S. Mineev, N.F. Khabibullina, E.N. Lyukmanova, D.A. Dolgikh, M.P. Kirpichnikov, A.S. Arseniev, Spatial structure and dimer--monomer equilibrium of the ErbB3 transmembrane domain in DPC micelles, Biochim Biophys Acta. 1808 (2011) 2081-2088.

[10] E.O. Artemenko, N.S. Egorova, A.S. Arseniev, A.V. Feofanov, Transmembrane domain of EphA1 receptor forms dimers in membrane-like environment, Biochim Biophys Acta. 1778 (2008) 2361-2367.

[11] D.R. Singh, E.B. Pasquale, K. Hristova, A small peptide promotes EphA2 kinase-dependent signaling by stabilizing EphA2 dimers, Biochim Biophys Acta. 1860 (2016) 1922-1928.

[12] K.S. Mineev, S.A. Goncharuk, A.S. Arseniev, Toll-like receptor 3 transmembrane domain is able to perform various homotypic interactions: an NMR structural study, FEBS Lett. 588 (2014) 3802-3807.

[13] S. Sarabipour, K. Hristova, Mechanism of FGF receptor dimerization and activation, Nat Commun. 7 (2016) 10262.

[14] S. Manni, K.S. Mineev, D. Usmanova, E.N. Lyukmanova, M.A. Shulepko, M.P. Kirpichnikov, J. Winter, M. Matkovic, X. Deupi, A.S. Arseniev, K. Ballmer-Hofer, Structural and functional characterization of alternative transmembrane domain conformations in VEGF receptor 2 activation, Structure. 22 (2014) 1077-1089.

[15] Y. Yano, K. Kondo, Y. Watanabe, T.O. Zhang, J.J. Ho, S. Oishi, N. Fujii, M.T. Zanni, K. Matsuzaki, GXXXG-mediated parallel and antiparallel dimerization of transmembrane helices and its inhibition by cholesterol: single-pair FRET and 2D IR studies, Angewandte Chemie (International ed in English). 56 (2017) 1756-1759.

[16] T. Qureshi, N.K. Goto, Impact of Differential Detergent interactions on transmembrane helix dimerization affinities, ACS Omega. 1 (2016) 277-285.

[17] A.Z. Ebie, K.G. Fleming, Dimerization of the erythropoietin receptor transmembrane domain in micelles, J Mol Biol. 366 (2007) 517-524.

[18] J. Oates, M. Hicks, T.R. Dafforn, D. DiMaio, A.M. Dixon, In vitro dimerization of the bovine papillomavirus E5 protein transmembrane domain, Biochemistry. 47 (2008) 8985-8992.

[19] A.L. Stouffer, C. Ma, L. Cristian, Y. Ohigashi, R.A. Lamb, J.D. Lear, L.H. Pinto, W.F. DeGrado, The interplay of functional tuning, drug resistance, and thermodynamic stability in the evolution of the M2 proton channel from the influenza A virus, Structure. 16 (2008) 1067-1076.
[20] T. Schmidt, A.J. Situ, T.S. Ulmer, Structural and thermodynamic basis of proline-induced transmembrane complex stabilization, Sci Rep. 6 (2016) 29809.
[21] S. Hippler-Mreyen, J.P. Klare, A.A. Wegener, R. Seidel, C. Herrmann, G. Schmies, G. Nagel, E. Bamberg, M. Engelhard, Probing the sensory rhodopsin II binding domain of its cognate transducer by calorimetry and electrophysiology, J Mol Biol. 330 (2003) 1203-1213.
[22] R. Chadda, L. Cliff, M. Brimberry, J.L. Robertson, A model-free method for measuring dimerization free energies of CLC-ec1 in lipid bilayers, J Gen Physiol. 150 (2018) 355-365.
[23] B.M. Kroncke, A.M. Duran, J.L. Mendenhall, J. Meiler, J.D. Blume, C.R. Sanders, Documentation of an imperative to improve methods for predicting membrane protein stability, Biochemistry. 55 (2016) 5002-5009.
[24] E. Krissinel, Crystal contacts as nature's docking solutions, J Comput Chem. 2010;31:133-143.
[25] L.C. Xue, J.P. Rodrigues, P.L. Kastritis, A.M. Bonvin, A. Vangone, PRODIGY: a web server for predicting the binding affinity of protein-protein complexes, Bioinformatics. 32 (2016) 3676-3678.