GeTFEP: A general transfer free energy profile of transmembrane proteins

Wei Tian,[†] Hammad Naveed,[‡] Meishan Lin,[†] and Jie Liang[*,†]

†*Department of Bioengineering, University of Illinois at Chicago, Chicago, IL 60607*

‡*Department of Computer Science, National University of Computer & Emerging Sciences (NUCES-FAST), Islamabad, ICT 44000*

E-mail: jliang@uic.edu

Abstract

Free energy of transferring amino acid side–chains from aqueous environment into lipid bilayers, known as transfer free energy (TFE), provides important information on the thermodynamic stability of membrane proteins. In this study, we derived a TFE profile named General Transfer Free Energy Profile (GeTFEP) based on computation of the TFEs of 58 β–barrel membrane proteins (βMPs). The GeTFEP agrees well with experimentally measured and computationally derived TFEs. Analysis based on the GeTFEP shows that residues in different regions of the TM segments of βMPs have different roles during the membrane insertion process. Results further reveal the importance of the sequence pattern of transmembrane strands in stabilizing βMPs in the membrane environment. In addition, we show that GeTFEP can be used to predict the positioning and the orientation of βMPs in the membrane. We also show that GeTFEP can be used to identify structurally or functionally important amino acid residue sites of βMPs. Furthermore, the TM segments of α–helical membrane proteins can be accurately predicted with GeTFEP, suggesting that the GeTFEP captures fundamental

thermodynamic properties of amino acid residues inside membrane, and is of general applicability in studying membrane protein.

1 Introduction

Membrane proteins play important roles in cellular metabolism, signaling regulation, and intercellular interactions.[1] Knowledge of the thermodynamic stability of membrane proteins is essential for understanding their folding behavior and their structure–function relationship.[2–5] A widely used measure to estimate the stabilities of membrane proteins is the transfer free energies (TFEs), which quantify the free energies of transferring amino acid residues from aqueous environment into lipid bilayers.[6–11]

Often called hydrophobicity scales, transfer free energies have been measured experimentally based on several model systems. The Wimley–White whole residue scale (WW–scale) measures TFEs of residue partitioning between water and octanol using a set of peptides as the host of amino acids.[8] The biological scale (Bio–scale) of Hessa *et al.* measures the free energies required to transfer residues in polypeptides into the ER membrane through the translocon machinery.[9] The Moon–Fleming whole protein scale (MF–scale) measures TFEs of residues from water to the membrane core in the context of a whole β–barrel membrane protein (βMP).[10] These experimentally obtained hydrophobicity scales have leaded to improved understanding of the structures and functions of membrane proteins[12] and have been used in prediction of transmembrane (TM) segments of membrane proteins.[13]

However, experimental measurement of TFEs is technically challenging, cumbersome, and costly.[14,15] Complementing experimentally measured transfer free energies, several hydrophobicity scales have been derived computationally, which can aid in our understanding of the governing principles of membrane protein folding.[2,4,16] The $E_Z\alpha$ and $E_Z\beta$ empirical potentials are knowledge–based hydrophobicity scales. They have been successfully applied in predicting the positioning of membrane proteins in the lipid bilayer, in discriminating side–

chain decoys, and in identifying protein–lipid interfaces.[17,18] However, these scales obtained from statistical analysis do not consider the physical interactions either between residues from neighboring helices/strands or within the same helix/strand, which are known to be important for membrane protein folding.[19,20] There have also been studies based on molecular dynamics (MD) simulations to calculate TFEs,[21–23] although the choice of the reference state before membrane insertion remains a challenging task.[22]

Another method of deriving TFEs computationally was developed for βMPs recently.[11] This method is based on a statical mechanical model in a discretized conformational space. It incorporates both intra– and inter–strand interactions in the TM segments of the proteins. It can be used to calculate TFEs of any lipid–facing residue in the TM segment of a βMP, as long as the number of TM strands of the protein is no more than 12. The computed TFE scale (OmpLA scale) is in excellent agreement with the MF–scale with a Pearson correlation coefficient $r = 0.90$. This scale has been applied successfully to explain how the functional fold and topology of the βMP are determined by the asymmetry of both the Gram–negative bacterial outer membrane and the TM residues.[11] A further algorithmic extension of this method has greatly reduced the computational cost, enabling the calculation of TFEs on all βMP known so far, regardless their sizes, with little loss of the accuracy.[24]

In this study, we use the new algorithm[24] to compute the depth–dependent TFE profile of each βMP in a non–redundant set of 58 βMPs. After examining their overall patterns, we found that there exists a general TFE profile applicable to all βMPs, which we call the General Transfer Free Energy Profile (GeTFEP). The GeTFEP agrees well with previously measured and computed TFEs. Analysis based on GeTFEP shows that residues in different regions of the TM segment have different roles during the membrane insertion process. Our results further reveal the importance of the sequence pattern of TM segments in stabilizing βMPs in the membrane environment. In addition, we also show that GeTFEP can be used to predict positioning and orientation of βMPs when embedded in the membrane, with overall results in good agreement with experimental data. Furthermore, we show that the GeTFEP

can be used to locate structurally or functionally important sites of βMPs. In addition, TM segments of α–helical membrane proteins can also be accurately predicted using the GeTFEP, suggesting that the GeTFEP captures fundamental thermodynamic properties of amino acid residues inside membrane, and has general applicability in studying membrane protein.

2 Results

GeTFEP: General Transfer Free Energy Profile

Computation of TFE profiles of βMPs

Using the methods described in Ref [24], we calculate the depth–dependent TFE profiles for each βMP in a non–redundant set of 58 βMPs. The proteins in this set have $\leq$ 30% pairwise sequence similarity. Briefly, for each βMP, we substituted each lipid–facing residue in the TM region to the other 19 amino acids. We calculated the TFEs of each amino acid substitution using Ala as the reference. The TFE profile of the protein was then obtained by taking average of the TFE values of the same amino acid type at the same depth position in the membrane. As an example, Fig. 1 shows the computed TFE profile of the protein LptD, the largest βMP with known structure (PDB ID: 4q35).

Derivation of GeTFEP

Although the 58 βMPs are in different oligomerization states, have different sizes (strand numbers) of TM segments, and come from different organisms, their TFE profiles are remarkably similar. Results of clustering their profiles show that the 58 βMPs can be grouped into only one group (with 56 βMPs) and two outliers: α– and γ–hemolysins (PDB ID: 7ahl and 3b07). Details of the clustering method can be found in the supporting information.

Unlike the other βMPs, the TM regions of both α– and γ–hemolysins are formed by repeated β–hairpin (Fig. S2B), which make their TFE profiles highly sensitive to the com-

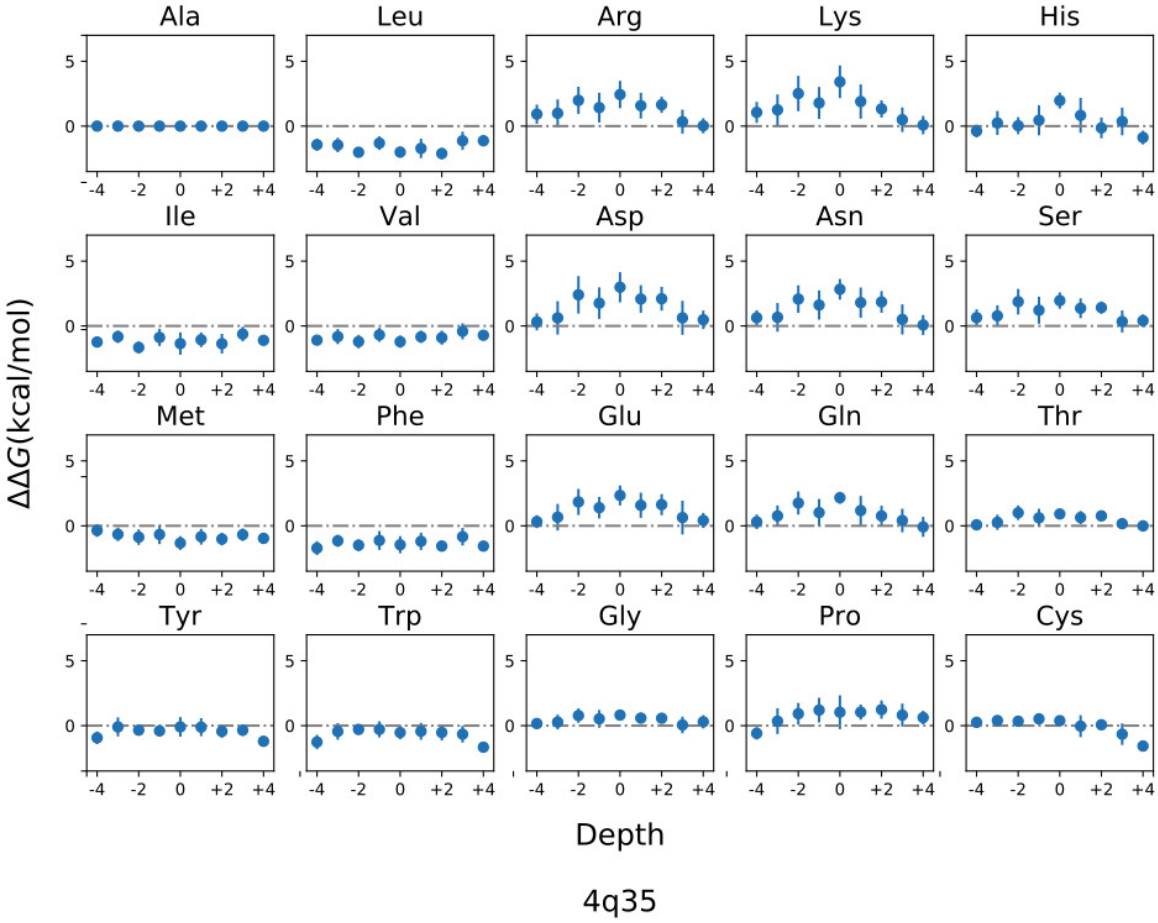

Figure 1: The TFE profile of the LptD protein. Each subfigure shows the calculated free energies (y–axis) of a specific amino acid type when transferred to certain depth position of the membrane (x–axis). The depth position of the membrane is discretized and indexed from -4 to +4, starting from the periplasmic side to the extracellular side of the outer membrane. Depth position 0 is at the mid–plane of the membrane.

position of the β–hairpin and the local interactions of residues within the hairpin (Fig. S2 C and D). Accordingly, we further investigate whether α– and γ–hemolysins have truly different thermodynamic properties than βMPs, or their outlier status is due to the special architecture of repeated β–hairpins.

We first computed the TFE profiles of artificially generated hemolysin–like βMPs constructed by repeating each β–hairpin in our βMP set. Altogether, we computed TFE profiles for 778 artificial hemolysin–like βMPs. We then sampled from these profiles with replacement, and computed the distribution of the distance from each sampled profile to the average profile of all sampled artificial βMPs. The distances from the TFE profiles of both α– and γ– hemolysins to the average profile are at the 80th percentile in the distance distribution (Fig. 2B), indicating that α– and γ–hemolysins are not fundamentally different in their thermodynamic properties from other βMPs. Therefore, we conclude that a general transfer free energy profile exists and is applicable to all βMPs, including α– and γ–hemolysins. We derive the General Transfer Free Energy Profile (GeTFEP) by averaging the TFEs of a specific amino acid at the same lipid bilayer depth position for all 58 βMPs (Fig. 2C).

Comparison with other hydrophobicity scales

We then examine how GeTFEP compares with other hydrophobicity scales. Since most experimentally measured scales are not depth–dependent, we first compare the scale of the TFEs at the hydrocarbon core position of depth 0 in the GeTFEP with other hydrophobicity scales. We refer this hydrophobicity scale as the mid–GeTFEP scale. The mid–GeTFEP scale correlates well with the experimentally measured hydrophobicity scales, having Pearson correlation coefficients $r = 0.83$ with the WW–scale, and $r = 0.92$ with the Bio–scale. It also correlates well with the computational βMP OmpLA scale,[11,24] with $r = 0.90$ (Fig. S3).

When compared with the experimentally measured MF–scale of the βMP OmpLA mid–GeTFEP has a correlation of $r = 0.87$. One noticeable difference between mid–GeTFEP and the MF–scale is that the TFE value of His is less unfavorable in mid–GeTFEP (Fig. 3A).

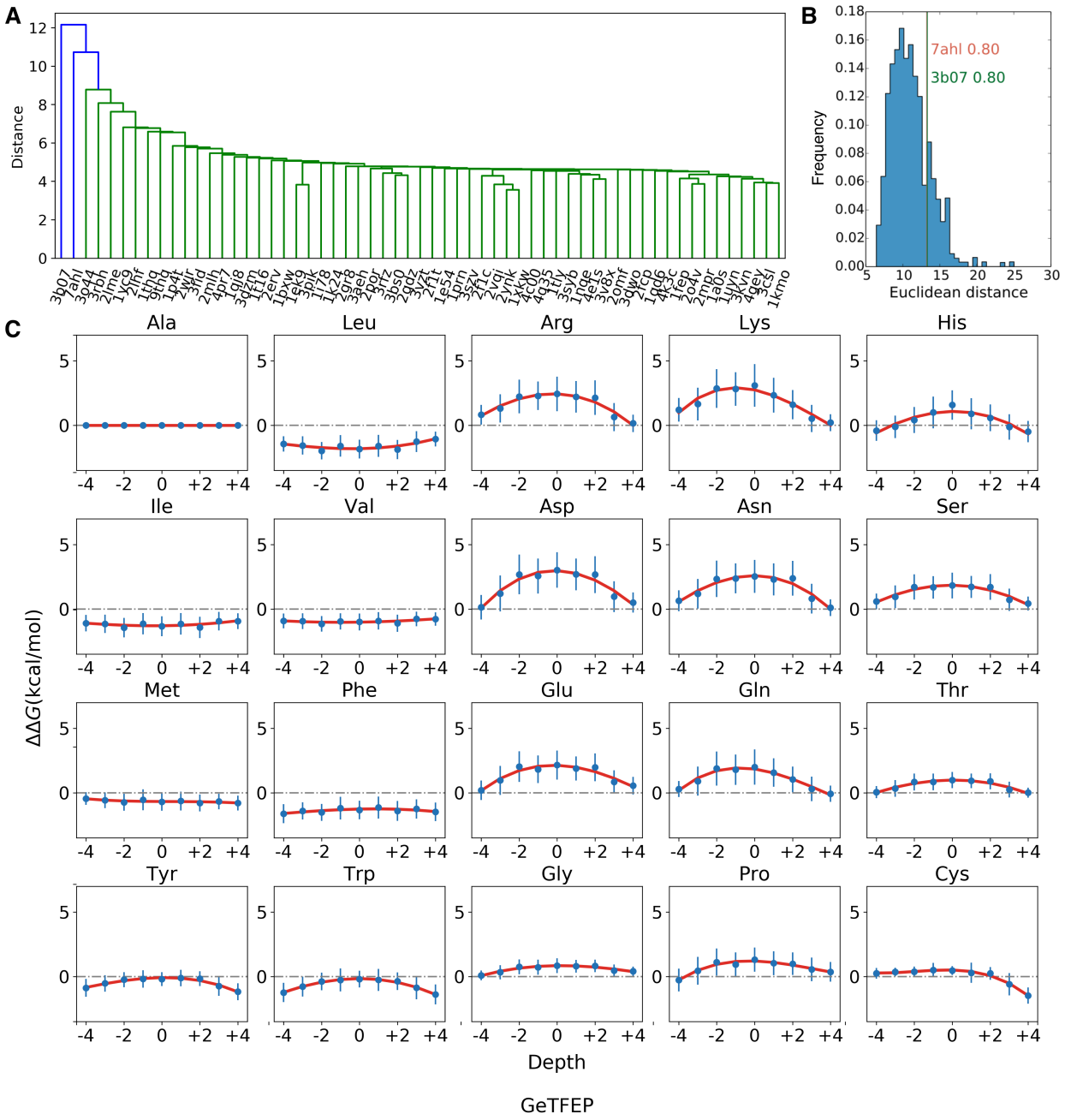

Figure 2: Derivation of the GeTFEP. **A.** Results of hierarchical clustering shows that all βMPs in the dataset can be group into one cluster, except α– and γ–hemolysins (7ahl and 3b07). **B.** The distribution of distance between the sampled TFE profiles of the artificially constructed hemolysin–like βMPs and their average TFE profile. The distances of both α– and γ–hemolysins are at the 80th percentile of the distribution. **C.** The General Transfer Free Energy Profile (GeTFEP) of each residues (blue), and the corresponding curves fitted by 3rd degree polynomials (red).

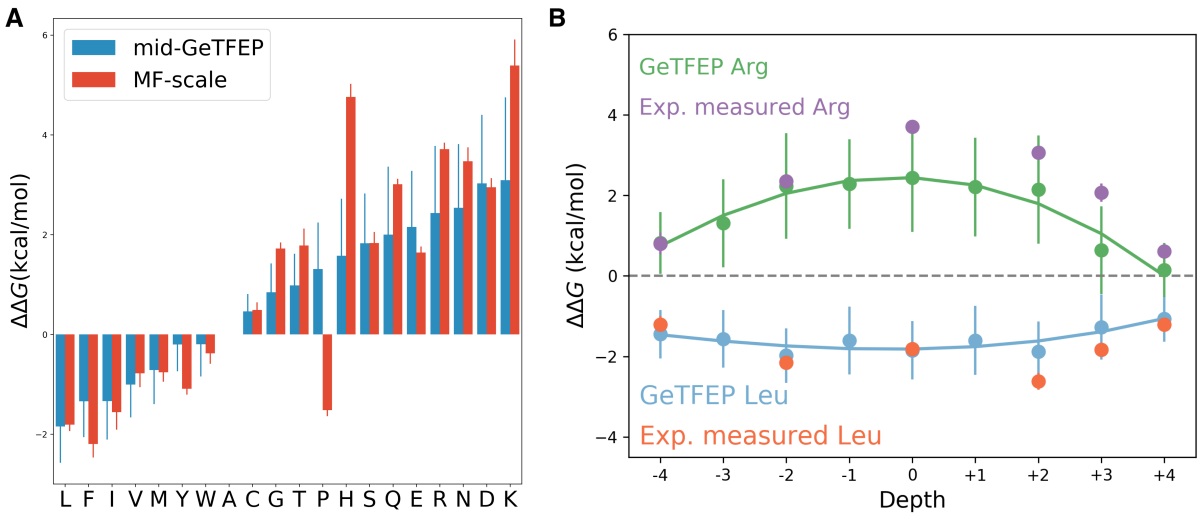

Figure 3: Comparison between the GeTFEP and the experimentally measured MF–scale. **A.** The mid–GeTFEP scale agrees well with the MF–scale, except Pro and His. **B.** The depth–dependent TFEs of Arg and Leu of GeTFEP also agree well with the experimental measurements.[10]

This is expected since the MF–scale was measured in acidic condition at pH=3.8, where His was fully protonated.[10] The different value in mid–GeTFEP likely reflects the property of His in physiological conditions of the outer membrane.

Another notable difference is Pro. It is found that Pro is unfavorable in the membrane environment according to the mid–GeTFEP scale, while it is found to be favorable according to the MF–scale (Fig. 3A). Pro tends to disrupt the structures of both α–helix and β–sheet, and is thermodynamically unfavorable in the non–polar core of the membrane.[25] The value of Pro in the GetFEP–mid scale reflects the general situation.

We then examined the depth–dependency of the GeTFEP of Arg and Leu, whose experimental results are available.[10] Their TFEs at different depth positions of the membrane are in good agreement with the experimentally measured values, with $r = 0.87$ for Arg and $r = 0.75$ for Leu (Fig. 3B), suggesting the GeTFEP captures the depth–dependency of TFEs of amino acids.

Insertion of βMP into membrane

βMP insertion as a thermodynamically driven spontaneous process

Upon synthesis in the cytoplasm, βMPs need to be transported across the periplasm and then folded into the outer membrane. As there is no energy source such as ATP in the periplasm, it was suggested that the free energies of βMP folding provide an adequate source to ensure successful periplasm translocation.[26] A computational study showed that the TFE of lipid–facing residues of the hydrophobic core regions are indeed the main driving force for membrane insertion.[11] Analysis also showed that lipid–facing residues in the TM regions of of βMPs have clear patterns of amino acid composition.[27] However, it is still unclear whether the insertion of βMPs into the membrane is primarily due to the extensive property of the hydrophobicity of lipid–facing residues, or the specific pattern of amino acid composition also plays important roles.

To investigate this question, we employed a simplified βMP insertion model based on the concerted folding mechanism proposed in Ref [[28]]. We ignore the effects of non–TM loops and discretizes the insertion process into 17 steps (Fig. 4A). We take the position recorded in the widely–used Orientations of Proteins in Membranes (OPM) database[29] as the fully inserted position of each βMP. This position is denoted as the reference positioon 0, and the other positions are indexed accordingly from -8 to $+8$. βMPs start the insertion process at position -8 from periplasmic side and become fully inserted into the membrane at position 0. From position 0 to $+8$, βMPs would translocate across the membrane. We assume that the stability of the TM region of a βMP can be approximated by summarizing TFEs of all lipid–facing residues in the membrane region. The stability of the βMP at each position was then calculated using the GeTFEP following this additive model. As an example, Fig. 4B shows stability of the protein OmpA (PDB ID: 1bxw) at different insertion positions. Overall, results of all βMPs show a funnel–like pattern of insertion energy (Fig. 4C). Most βMPs (52 of 58) have minimum free energy when they are fully inserted into membranes (position 0,

Tables S1 and S2). The funnel–like pattern indicates that the insertion of βMPs into outer membranes is indeed a spontaneous process. βMPs become energetically trapped after being fully inserted.

Importance of patterns of TM lipid–facing residues in membrane insertion

We then examine if the funnel–like insertion energy pattern arises from the extensive property of the TFEs of the hydrophobic residues alone. We considered only the 52 βMPs whose minimum free energies are at the fully inserted position. We first shuffled the sequences of the β–strands within the TM segment of each βMP. While the side–chain direction as well as the interstrand hydrogen bond pairing at each residue position in β–strands are maintained, all TM residues are permuted. Each βMP is shuffled 2,000 times. We found that it is highly unfavorable to insert the shuffled βMPs into the membrane. This is expected, since the shuffling changes hydrophobicity of TM segments βMPs. Before the shuffling, the ionizable/polar residues were enriched among lumen–facing residues of βMPs, while lipid–facing residues were mostly apolar. After the shuffling, they were much evenly distributed.

We then investigate how insertion energy is affected if only the lipid–facing residues are shuffled. While the insertion of the shuffled βMPs remains energetically favorable (see Fig. S4 for an example), shuffled βMPs are less stable compared to the original βMPs at the fully inserted position for 50 out of 52 βMPs: The insertion energy for the shuffled βMPs is on average 6.36 kcal/mol higher (Table S1). In addition, the fully inserted position (position 0) is no longer the most stable position for 17.4% of the shuffled βMPs (Table S1). These results indicate that the locational patterns of lipid–facing residues[30] in the TM region are optimized for βMPs to gain stability in the membrane environment.

Roles of residues in different TM regions during membrane insertion

The TM segment of a βMP can be divided into three regions, namely, the periplasmic headgroup region, the hydrophobic core region, and the extracellular headgroup region.[27] We

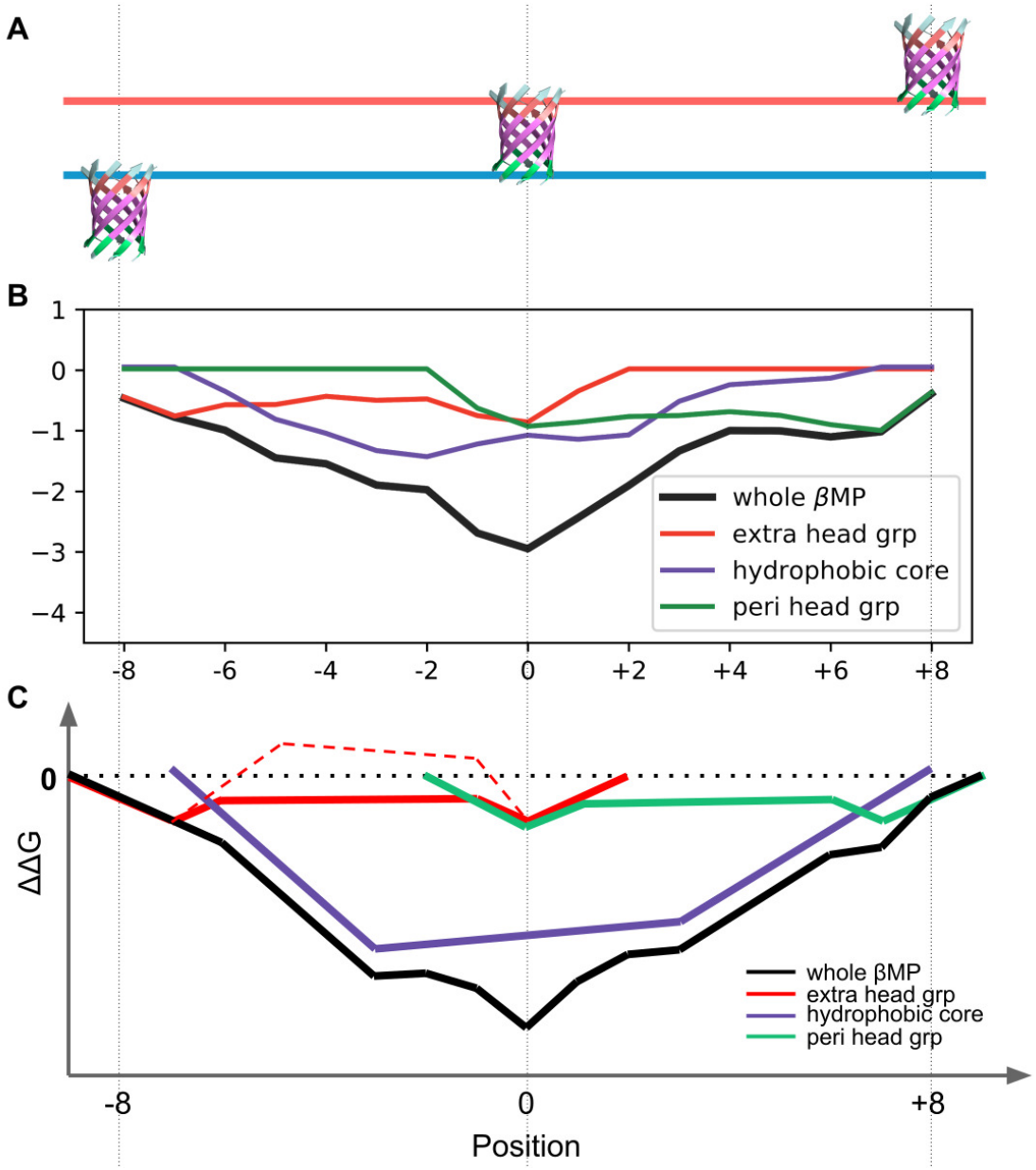

Figure 4: Illustration of the membrane insertion process of βMPs. **A.** The simplified βMP insertion model following Ref [28]. **B.** The computed insertion energies of the OmpA protein at different depth positions. Contribution of different regions are color coded. **C.** Illustration on how free energies change with the position of βMP in the membrane. The dashed red segments show that lipid–facing residues in extracellular head group region sometimes become energetically unfavorable.

investigate how these regions contribute to the insertion energy of the βMP. We found that residues in the same regions across all 52 βMPs shared similar patterns in their insertion free energy profile (Fig. 4C), indicating that they play similar roles in the insertion process. Among these, lipid–facing residues of the extracellular headgroup region facilitate the initialization of the insertion process, as they are energetically favorable in the interfacial region on the periplasmic side (position -8 and -7). As insertion proceeds, these residues become less favorable and occasionally unfavorable when they become more embedded in the membrane. At this time, lipid–facing residues of the hydrophobic core region start to be inserted in the membrane, and strongly drive the insertion process (position -6 to -2). When lipid–facing residues of the extracellular headgroup region approach the interfacial region of the extracellular side, they become energetically favorable again. At the same time, lipid–facing residues of the periplasmic headgroup region become inserted (position -1 and 0), and the TFE of the whole βMP reaches its minimum at position 0.

Although lipid–facing residues of the hydrophobic core region are known to provide the main driving force for membrane insertion of βMPs,[11] we found that the TFEs of hydrophobic core region do not reach their minimum when βMPs are fully inserted at position 0 for all 52 βMPs. Upon incorporation of contributions from other regions, the overall TFEs of the whole βMPs indeed reach the minimum at the fully inserted position. The "W" shape of the free energy curves of the two head group regions (the red and green curve in Fig. 4C) suggests that lipid–facing residues in these regions act like "energetic latches" to lock βMPs into their fully inserted position.

Prediction of βMP positioning and orientation in the membrane

GeTFEP can be used to predict positioning and orientation of βMPs in the membrane, similarly to previous studies.[17,18] Here, the membrane is idealized as an infinite slab with a thickness of h. Each βMP is initially positioned in the membrane with its center of mass of the barrel domain at the midplane of the membrane and its barrel axis aligned with the

normal direction (z–axis) of the membrane (Fig. 5A). The protein can be rotated around the x– and y–axes with angles θ_x and θ_y, respectively. The two rotation angles together determine the tilt angle of the protein. The protein can also be translated with a displacement d_z. This displacement and the membrane thickness determine the TM segment of the protein. When embedded in the membrane, the lipid–facing residues of the TM region and the loop residues are used to calculate the total energy of the βMP using the GeTFEP. As an example, Fig. 5B shows how rotation angles θ_x and θ_y affect the stability of the protein BtuB (PDB ID: 1nqe) when the displacement d_z and the membrane thickness h are fixed.

We systematically examine the parameter combination of θ_x, θ_y, d_z, and h. A βMP is predicted to take the position and the orientation when the lowest free energy is reached. The predicted protein tilt angles of all 58 βMPs correlate well ($r = 0.76$) with OPM records.[29] The average protein tilt angle of 7.3° is consistent with that of 6.2±1.8° recorded in the OPM. The strand tilt angles and the membrane thickness predicted are again in good agreement with experimentally determined results (Table 1).

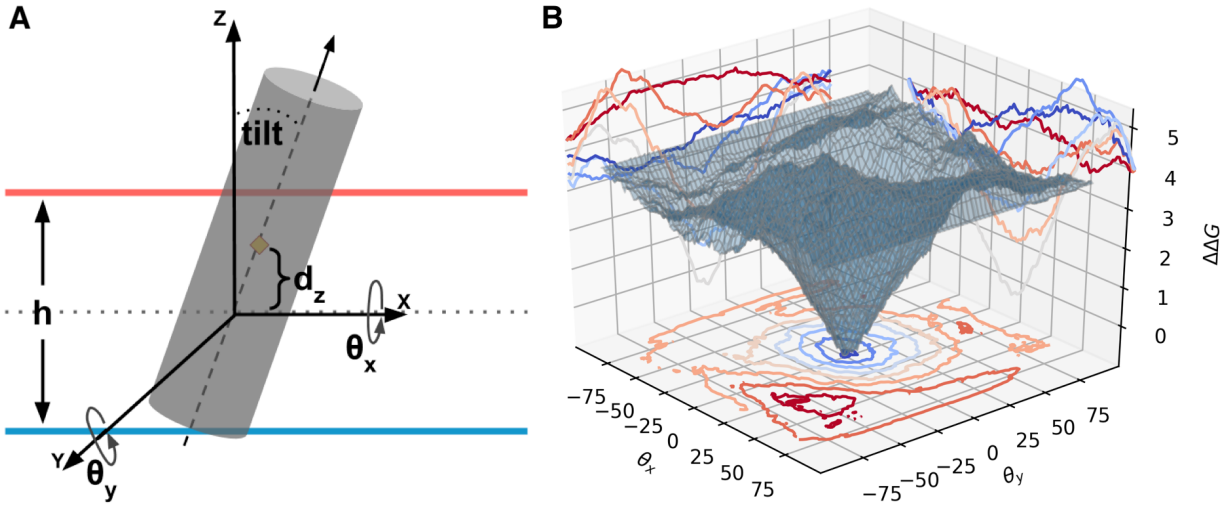

Figure 5: Prediction of positioning and orientation of βMPs. **A.** The positioning and orientation of the βMP inside the membrane are determine by the rotation angles θ_x and θ_y, the translation displacement d_z, and the membrane thickness h. **B.** The funnel–like landscape of the stability of the BtuB protein. It shows how rotation angles affect the stability of BtuB when d_z and h are fixed.

Table 1: Comparison between the predicted positioning and orientation and experimental results[31] of βMPs. * The experimentally measured tilt angles are the upper bounds of the actual values.[31]

	Protein	PDB ID	Experiment	GeTFEP	OPM
TM tilt (°)	FhuA	2fcp	46.0*	38.2	38.3
	OmpA	1bxw	44.5*	40.2	38.7
Membrane thickness (Å)	FhuA	1fep	≥ 23.1	23.5	24.3
	OmpF	2omf	~ 21.0	22.8	25.2
	BtuB	1nqe	≥ 20.2	23.0	23.4

Prediction of structurally and functionally important sites of βMPs

While overall the computed TFEs of lipid–facing residues of βMPs follow the general pattern of the GeTFEP, the TFE values of a specific residue in a particular βMP can deviate significantly from values in the general profile (see SI for details). Among all 3,500 lipid–facing residued in the TM segments of all 58 βMPs, we find that 305 or 8.7% of the residues have TFE values deviate significantly from the GeTFEP. Since lipid–facing residues are overall the major contributors to the stability of βMPs as discussed above, the deviation from the general profile indicate that the residue is likely to have important roles other than providing stability. To understand the origin of these deviations, we examined three proteins in details, namely, OmpLa, PagP, and PagL, which have sufficient experimental information. We found that most deviant residues either have functional roles or have local structures quite different from residues in the canonical model of beta barrels (Tab 2).

Among the deviant residues in OmpLa, 142H and 156N are both in the catalytic triad[32,33] that are essential for its phospholipase activities; 40L and 92Y are the sites where substrates bind;[34] Furthermore, the deviant residue 116P interacts with 92Y and 142H through hydrogen bonds. Among the deviant residues in PagP, 69L interacts with the out–clamp α–helix of PagP;[35] 27I and 125L are both at the lateral routes where β–hydrogen bonding is absent (Fig 6), which ensure that substrates can access the protein interior so that PagP can carry out its enzymatic functions.[36] In PagL, the deviant residue 108I is in the ligand binding site,[37] and 126H is part of the catalytic triad of its enzymatic site.[38]

As the calculation of TFEs does not require knowledge of 3D structures of βMPs, our results suggest that deviation analysis can help to discover functional sites and/or structurally anomalous sites using sequence information only. While our analysis is restricted to three proteins due to the limited nature of experiment data, we believe overall deviant residues play special roles in either performing biological function or in maintaining the unique structural form of βMPs.

Table 2: Predicted important sites in OmpLA, PagP and PagL by deviation analysis

Protein	Residue	Notes
OmpLA(1qd6)	38N	
	40L	Substrate binding[34]
	92Y	Substrate binding[34]
	116P	Interstand neighbor of 92Y and 142H
	120L	
	142H	Catalytic site[32,33]
	156N	Catalytic site[32,33]
	237L	
PagP(1thq)	27I	Lateral route from membrane to protein interior[35]
	69L	Interact with the out clamp α–helix[35]
	125L	Lateral route from membrane to protein interior[35]
	131L	
PagL(2erv)	108I	Ligand binding site[37]
	126H	Catalytic site[38]

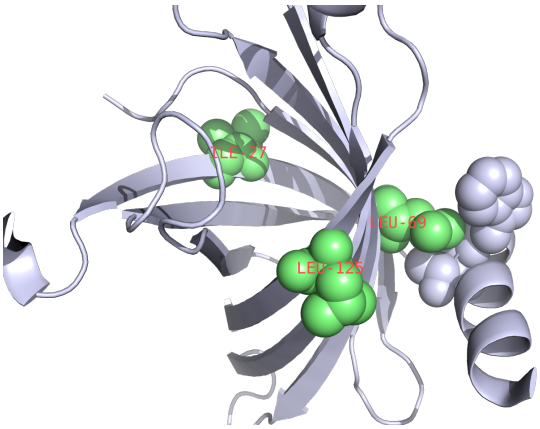

Figure 6: Predicted important sites of PagP. Residues 27I and 125L are at the sites where the hydrogen bonds between the β–strands are disrupt. 69L has interaction with the out–clamp α–helix of PagP.

GeTFEP can predict TM region of α–helical membrane proteins

Although the MF–scale was measured in the βMP system, it was suggested that the scale is also applicable to TM region of the α–helical membrane proteins (αMPs), since the MF–scale has a strong correlation with the nonpolar solvent accessible surface areas of the residues.[10] We hypothesize that the GeTFEP may also reflects fundamental thermodynamic properties of transferring sidechains of amino acids to the membrane environment, regardless whether the residue is in a β–barrel or a α–helical membrane protein. We carried out the standard hydropathy analysis[39] using the Membrane Protein Explorer (MPEx) program.[40] on 131 αMPs obtained from the MPTopo database[41] Since MPEx uses depth–independent hydrophobicity scales, we used the mid–GeTFEP scale for our calculation.

The results show that this simple analysis using the mid–GeTFEP scale correctly predicts both the TM regions and the numbers of the TM segments for 90 or ∼69% of the 131 αMPs in the dataset. This compares favorably to other hydrophobicity scales, including those measured or derived from αMPs (Table 3). For most of the remaining 41 proteins, GeTFEP correctly predicted the TM regions, but predicted the numbers of the TM segments incorrectly due to the ambiguity in assignment of whether two consecutive TM segments should be considered as one TM segment (see Fig. S5B for an example). Examination of the number of TM residues correctly predicted by the mid–GeTFEP scale show that we achieves a precision of ∼85% and a recall of ∼71%, which compares favorably to other hydrophobicity scales (Table 3). These results suggest that the GeTFEP reflects fundamental thermodynamic properties of amino acid residues inside membrane, and can be used to study the general stability of both α–helical and β–barrel membrane proteins.

The validity of transfer free energy value of Pro in the GeTFEP

We further examine the TFE value of Pro in the mid-GeTFEP scale, which is qualitatively different from that in the MF–scale. We swapped the value of Pro from MF–scale into the mid–GeTFEP scale, and used this Pro–swapped scale in the hydropathy analysis. This is

reasonable as the mid–GeTFEP scale is strongly correlated with the MF–scale, and has comparable values. However, we found that the precision of predicting TM residues deteriorates significantly from 85% using the mid–GeTFEP scale to 72% using the Pro–swapped scale (Table 3). This result suggests that Pro is more likely to be membrane unfavorable as characterized by the mid–GeTFEP scale rather than membrane favorable as characterized by the MF–scale.

Table 3: Prediction of TM segments and residues αMPs. The mid-GeTFEP scale performs better than the other hydrophobicity scales. The first three scales are measured or derived in α–helical systems, the others in βMPs.

Hydrophobicity scale	αMPs % (#) with TM segs. correctly predicted	TM res. precision	TM res. recall	TM res. F–measure
WW–scale	50%(66)	73%	75%	0.74
Bio–scale	22%(29)	95%	21%	0.34
$E_Z\alpha$	49%(64)	71%	77%	0.74
MF–scale	48%(63)	77%	65%	0.70
Pro–swapped	49%(64)	72%	74%	0.71
mid–GeTFEP	69%(90)	85%	71%	0.78

3 Conclusions and discussion

In this study, we derived the General Transfer Free Energy Profile (GeTFEP) from a non–redundant set of 58 βMPs. We showed that the GeTFEP agrees well with previous experimentally measured and computationally derived TFEs. The GeTFEP reveals fundamental thermodynamic properties of amino acid residues inside membrane environment, and it is useful in analysis of stability and function of membrane proteins.[5]

As the lipid membrane bilayer is anisotropic along the bilayer normal,[42] a residue at different depth of the membrane will have different interaction with lipid molecules in the environment, resulting in the depth–dependency of transfer free energies. However, there are few experimental measurements of TFEs at different depth positions other than the hydrophobic core, except Arg and Leu.[10] Comparison between the GeTFEP and the experimentally

measured values of Arg and Leu shows that the GeTFEP captures this depth–dependency well.

In addition, the GeTFEP exhibits asymmetric values between TFEs of residues in the membrane inner leaflet (depth -4 to 0) and in the outer leaflet (depth 0 to +4, Fig. 2C). Most βMPs in our dataset resides in the bacterial outer membrane, whose outer leaflet contains additional complex lipolysaccharides in contrast to its inner leaflet of phospholipids. This asymmetry in membrane composition results in the asymmetry of the transfer free energies in the GeTFEP. To understand membrane proteins in an environment of symmetric membrane leaflets, we also derived a symmetric TFE profile, named sym–GeTFEP, by mirroring the TFE values of the inner leaflet side of the GeTFEP (Fig. S6). In this study, the sym–GeTFEP was used to analyze the non–outer–membrane βMPs, e.g. α– and γ–hemolysins and vibrio cholerae cytolysin.

We explored the energetic contribution of different regions of βMPs during the membrane insertion process. Our analysis showed that the stability of βMPs does not come alone from the extensive property of the hydrophobicity of lipid–facing residues in the TM segment. Rather, the pattern of the amino acid residues in the TM segment also play significant roles. Results from analysis of sequence shuffling show that the patterns and location of amino acid residues are optimized to stabilize βMPs in the membrane environment. Using the GeTFEP, we are also able to predict membrane positioning and orientations of βMPs.

The GeTFEP can also be used to detect structurally or functionally important residues in βMPs. This can be achieved by examination of residues whose TFEs deviate significantly from the GeTFEP. As calculation of TFEs of residues of a specific βMP only requires rough estimation of relative positions between adjacent β–strands, which can be reliably predicted from the protein sequence,[43,44] computing the TFE deviation therefore requires only sequence information. The GeTFEP–deviation analysis can aid in discovery of functional sites or structurally important sites in novel βMPs, without requiring knowledge of their 3D structures. In addition, GeTFEP–based analysis can aid in design and engineering of novel

βMPs.

Furthermore, we demonstrated that GeTFEP can be used to predict TM residues of α–helical membrane proteins. Results showed that GeTFEP performs better than the hydrophobicity scales measured/calculated in αMP systems, suggesting that the GeTFEP reflects fundamental thermodynamic properties of amino acid residues inside membrane, and can be used to study the general stability of both α–helical and β–barrel membrane proteins.

Acknowledgement

This work is supported by NIH R01GM079804, R01CA204962-01A1, R21AI126308, and R35GM127084.

References

(1) Fairman, J. W.; Noinaj, N.; Buchanan, S. K. The structural biology of β-barrel membrane proteins: A summary of recent reports. **2011**, *21*, 523–531.

(2) Liang, J. Experimental and computational studies of determinants of membrane-protein folding. **2002**, *6*, 878–884.

(3) Naveed, H.; Jackups, R.; Liang, J. Predicting weakly stable regions, oligomerization state, and protein-protein interfaces in transmembrane domains of outer membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 12735–40.

(4) Liang, J.; Naveed, H.; Jimenez-Morales, D.; Adamian, L.; Lin, M. Computational studies of membrane proteins: Models and predictions for biological understanding. **2012**, *1818*, 927–941.

(5) Yang, Y.; Guo, R.; Gaffney, K.; Kim, M.; Muhammednazaar, S.; Tian, W.; Wang, B.;

Liang, J.; Hong, H. Folding-Degradation Relationship of a Membrane Protein Mediated by the Universally Conserved ATP-Dependent Protease FtsH. *Journal of the American Chemical Society* **2018**, *140*, 4656–4665.

(6) Tanford, C. The hydrophobic effect and the organization of living matter. *Science (New York, N.Y.)* **1978**, *200*, 1012–1018.

(7) White, S. H.; Wimley, W. C. Membrane protein folding and stability: physical principles. *Annual review of biophysics and biomolecular structure* **1999**, *28*, 319–365.

(8) Wimley, W. C.; White, S. H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature structural biology* **1996**, *3*, 842–848.

(9) Hessa, T.; Kim, H.; Bihlmaier, K.; Lundin, C.; Boekel, J.; Andersson, H.; Nilsson, I.; White, S. H.; von Heijne, G. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **2005**, *433*, 377–81.

(10) Moon, C. P.; Fleming, K. G. Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 10174–7.

(11) Lin, M.; Gessmann, D.; Naveed, H.; Liang, J. Outer Membrane Protein Folding and Topology from a Computational Transfer Free Energy Scale. *Journal of the American Chemical Society* **2016**, *138*, 2592–2601.

(12) MacCallum, J. L.; Tieleman, D. P. Hydrophobicity scales: a thermodynamic looking glass into lipid-protein interactions. *Trends in biochemical sciences* **2011**, *36*, 653–62.

(13) Deber, C. M.; Wang, C.; Liu, L. P.; Prior, a. S.; Agrawal, S.; Muskat, B. L.; Cuticchia, a. J. TM Finder: a prediction program for transmembrane protein segments using a combination of hydrophobicity and nonpolar phase helicity scales. *Protein science : a publication of the Protein Society* **2001**, *10*, 212–219.

(14) Moon, C. P.; Kwon, S.; Fleming, K. G. Overcoming hysteresis to attain reversible equilibrium folding for outer membrane phospholipase A in phospholipid bilayers. **2011**, *413*, 484–494.

(15) Otzen, D. E.; Andersen, K. K. Folding of outer membrane proteins. **2013**, *531*, 34–43.

(16) Adamian, L.; Liang, J. Prediction of buried helices in multispan alpha helical membrane proteins. *Proteins* **2006**, *63*, 1–5.

(17) Senes, A.; Chadi, D. C.; Law, P. B.; Walters, R. F. S.; Nanda, V.; DeGrado, W. F. Ez, a Depth-dependent Potential for Assessing the Energies of Insertion of Amino Acid Side-chains into Membranes: Derivation and Applications to Determining the Orientation of Transmembrane and Interfacial Helices. *Journal of Molecular Biology* **2007**, *366*, 436–448.

(18) Hsieh, D.; Davis, A.; Nanda, V. A knowledge-based potential highlights unique features of membrane α-helical and β-barrel protein insertion and folding. *Protein science : a publication of the Protein Society* **2012**, *21*, 50–62.

(19) Moore, D. T.; Berger, B. W.; DeGrado, W. F. Protein-Protein Interactions in the Membrane: Sequence, Structural, and Biological Motifs. **2008**, *16*, 991–1001.

(20) Hedin, L. E.; Öjemalm, K.; Bernsel, A.; Hennerdal, A.; Illergård, K.; Enquist, K.; Kauko, A.; Cristobal, S.; von Heijne, G.; Lerch-Bader, M.; Nilsson, I.; Elofsson, A. Membrane Insertion of Marginally Hydrophobic Transmembrane Helices Depends on Sequence Context. *Journal of Molecular Biology* **2010**, *396*, 221–229.

(21) MacCallum, J. L.; Bennett, W. F. D.; Tieleman, D. P. Partitioning of amino acid side chains into lipid bilayers: results from computer simulations and comparison to experiment. *The Journal of general physiology* **2007**, *129*, 371–377.

(22) Gumbart, J.; Roux, B. Determination of membrane-insertion free energies by molecular dynamics simulations. *Biophysical Journal* **2012**, *102*, 795–801.

(23) Ulmschneider, M. B.; Ulmschneider, J. P.; Schiller, N.; Wallace, B. a.; von Heijne, G.; White, S. H. Spontaneous transmembrane helix insertion thermodynamically mimics translocon-guided insertion. *Nature Communications* **2014**, *5*, 4863.

(24) Tian, W.; Lin, M.; Naveed, H.; Liang, J. Efficient computation of transfer free energies of amino acids in beta-barrel membrane proteins. *Bioinformatics (Oxford, England)* **2017**, 1–8.

(25) Senes, A.; Engel, D. E.; DeGrado, W. F. Folding of helical membrane proteins: the role of polar, GxxxG-like and proline motifs. *Current opinion in structural biology* **2004**, *14*, 465–79.

(26) Moon, C. P.; Zaccai, N. R.; Fleming, P. J.; Gessmann, D.; Fleming, K. G. Membrane protein thermodynamic stability may serve as the energy sink for sorting in the periplasm. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 4285–90.

(27) Jackups, R.; Liang, J. Interstrand pairing patterns in beta-barrel membrane proteins: the positive-outside rule, aromatic rescue, and strand registration prediction. *Journal of molecular biology* **2005**, *354*, 979–93.

(28) Kleinschmidt, J. H.; den Blaauwen, T.; Driessen, A. J.; Tamm, L. K. Outer membrane protein A of Escherichia coli inserts and folds into lipid bilayers by a concerted mechanism. *Biochemistry* **1999**, *38*, 5006–16.

(29) Lomize, M. A.; Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. OPM: Orientations of proteins in membranes database. *Bioinformatics* **2006**, *22*, 623–625.

(30) Jackups, R.; Cheng, S.; Liang, J. Sequence motifs and antimotifs in beta-barrel membrane proteins from a genome-wide analysis: the Ala-Tyr dichotomy and chaperone binding motifs. *Journal of molecular biology* **2006**, *363*, 611–23.

(31) Lomize, A. L.; Pogozheva, I. D.; Lomize, M. a.; Mosberg, H. I. Positioning of proteins in membranes: A computational approach. *Protein Science* **2006**, *15*, 1318–1333.

(32) Dekker, N. Outer-membrane phospholipase A: Known structure, unknown biological function. *Molecular Microbiology* **2000**, *35*, 711–717.

(33) Kingma, R. L.; Fragiathaki, M.; Snijder, H. J.; Dijkstra, B. W.; Verheij, H. M.; Dekker, N.; Egmond, M. R. Unusual catalytic triad of Escherichia coli outer membrane phospholipase A. *Biochemistry* **2000**, *38*, 10017–10022.

(34) Snijder, H. J.; Ubarretxena-Belandia, I.; Blaauw, M.; Kalk, K. H.; Verheij, H. M.; Egmond, M. R.; Dekker, N.; Dijkstra, B. W. Structural evidence for dimerization-regulated activation of an integral membrane phospholipase. *Nature* **1999**, *401*, 717–721.

(35) PDB structure of PagP. doi:10.2210/pdb1thq/pdb.

(36) Bishop, R. E. Structural biology of membrane-intrinsic β-barrel enzymes: Sentinels of the bacterial outer membrane. *Biochimica et Biophysica Acta - Biomembranes* **2008**, *1778*, 1881–1896.

(37) PDB structure of PagL. doi:10.2210/pdb2erv/pdb.

(38) Rutten, L.; Geurtsen, J.; Lambert, W.; Smolenaers, J. J. M.; Bonvin, A. M.; de Haan, A.; van der Ley, P.; Egmond, M. R.; Gros, P.; Tommassen, J. Crystal structure and catalytic mechanism of the LPS 3-O-deacylase PagL from Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 7071–7076.

(39) Kyte, J.; Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology* **1982**, *157*, 105–132.

(40) Snider, C.; Jayasinghe, S.; Hristova, K.; White, S. H. MPEx: A tool for exploring membrane proteins. *Protein Science* **2009**, *18*, 2624–2628.

(41) Jayasinghe, S.; Hristova, K.; White, S. H. MPtopo: A database of membrane protein topology. *Protein Science* **2001**, *10*, 455–458.

(42) Pogozheva, I. D.; Mosberg, H. I.; Lomize, A. L. Life at the border: Adaptation of proteins to anisotropic membrane environment. *Protein Science* **2014**, *23*, 1165–1196.

(43) Naveed, H.; Xu, Y.; Jackups, R.; Liang, J. Predicting three-dimensional structures of transmembrane domains of β-barrel membrane proteins. *Journal of the American Chemical Society* **2012**,

(44) Tian, W.; Lin, M.; Tang, K.; Liang, J.; Naveed, H. High-resolution structure prediction of β–barrel membrane proteins. *Proceedings of the National Academy of Sciences* **2018**, 201716817.

Supplementary information

Dataset. We use 58 non–homologous β–barrel membrane proteins with less than 30% pairwise sequence identity for this study. The PDB IDs are: 1a0s, 1bxw, 1e54, 1ek9, 1fep, 1i78, 1k24, 1kmo, 1nqe, 1p4t, 1prn, 1qd6, 1qj8, 1t16, 1thq, 1tly, 1uyn, 1xkw, 1yc9, 2erv, 2f1c, 2f1t, 2fcp, 2gr8, 2lhf, 2lme, 2mlh, 2mpr, 2o4v, 2omf, 2por, 2qdz, 2vqi, 2wjr, 2ynk, 3aeh, 3bs0, 3csl, 3dwo, 3dzm, 3fid, 3kvn, 3pik, 3rbh, 3rfz, 3syb, 3szv, 3v8x, 3vzt, 4c00, 4e1s, 4gey, 4k3c, 4pr7, 4q35, 7ahl, 3b07, 3o44.

Clustering of βMP TFE profiles. In this study, euclidean distance between the TFE profiles of the βMPs and single linkage are used in the hierarchical clustering. The silhouette score is a measure for assessing the quality of clustering results. In practice, a > 0.5 silhouette score indicates a good clustering Our results show that it is not reasonable to cluster the βMPs into two groups, since the silhouette score is < 0.5 at 2 clusters (Fig. S1A). When we increase the number of clusters, the silhouette score keeps decreasing. Therefore, we conclude that only one group exists for our βMP dataset. Fig. S1B visualizes the cluster result after the profiles of βMPs are reduced to a 3D space using the Principal Component Analysis (PCA).

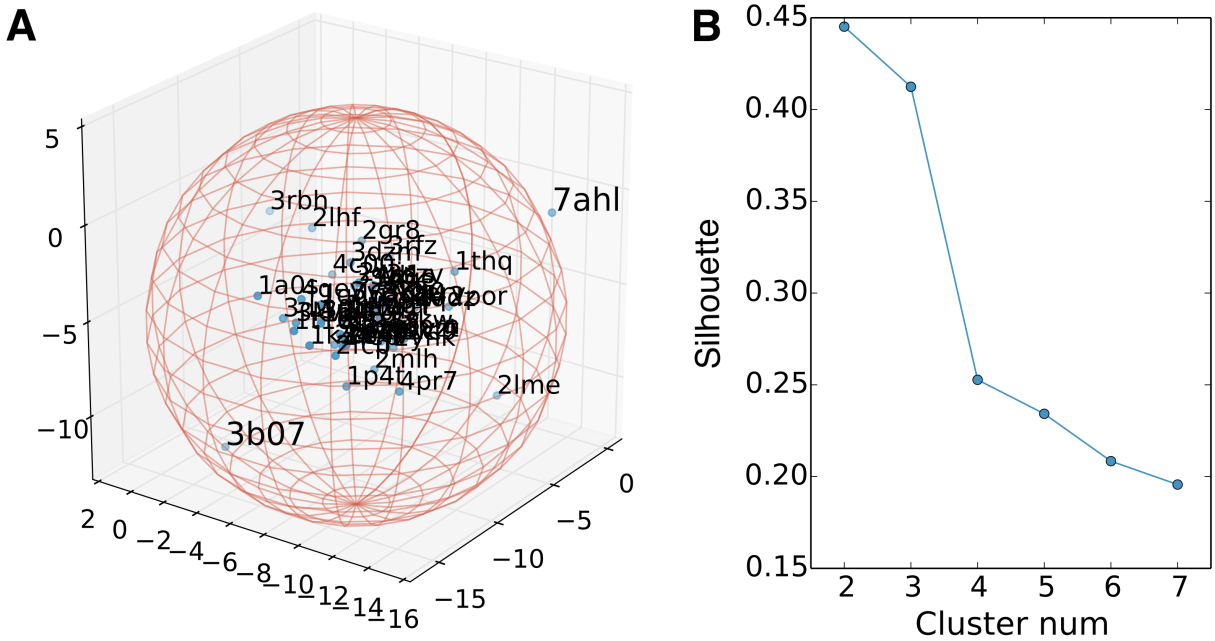

Figure S1: **A.** Visualization of the βMP TFE profiles after reduced to a 3D space via PCA. **B.** The silhouette scores for different cluster numbers of the βMP TFE profiles.

In the hierarchical clustering, we also tried other parameter settings with correlation distance and/or other reasonable linkages (eg. average linkage or weighted linkage), and the conclusion remains the same.

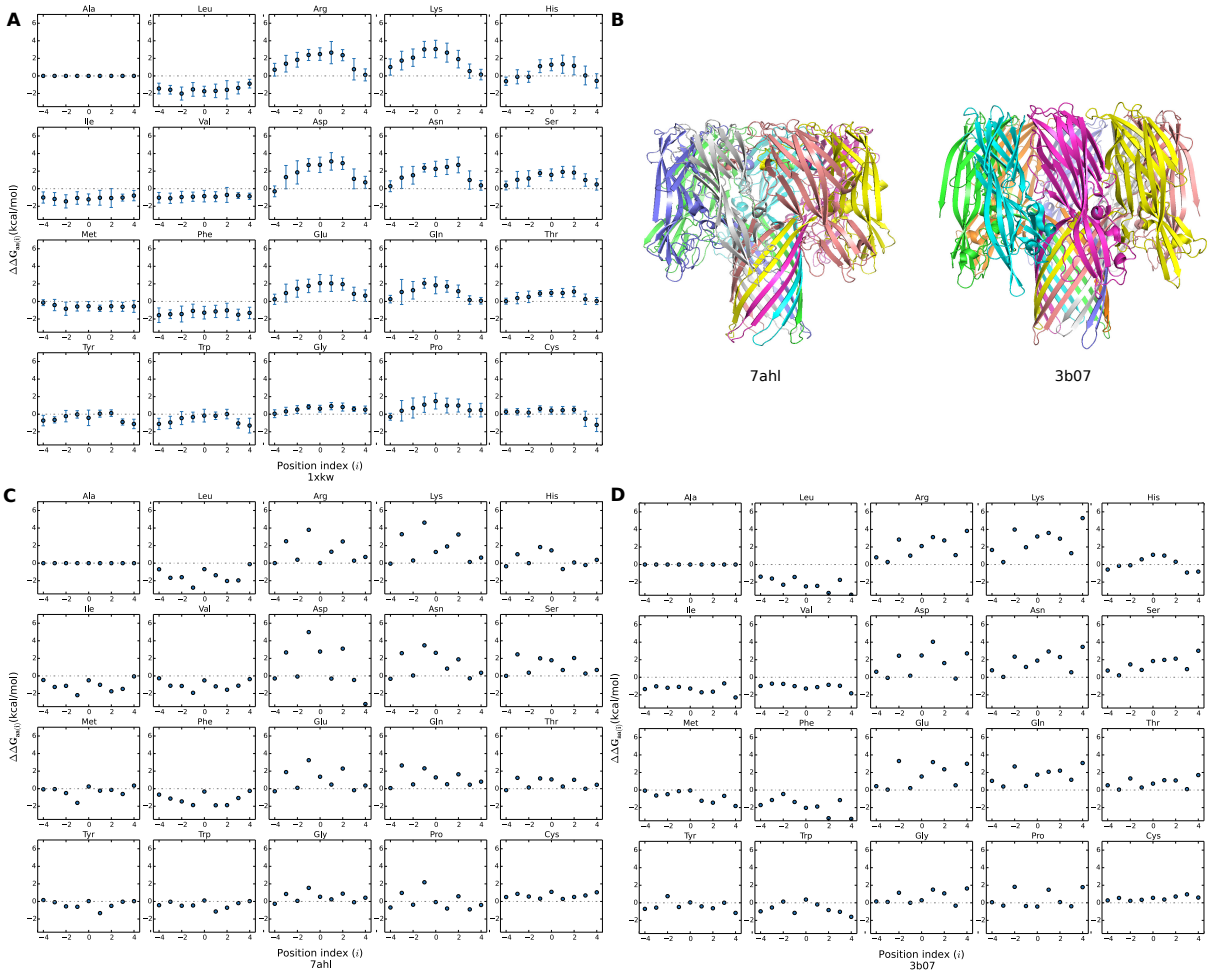

Figure S2: **Structures and TFE profiles of α– and γ–hemolysin. A.** A typical TFE profile of βMPs (FptA, PDB id:1xkw). **B.** The structures of α–hemolysin (PDB id:7ahl) and γ–hemolysin (PDB id:3b07) Both TM segments are constructed with repeated β–hairpins. **C.** The TFE profile of α–hemolysin. **D.** The TFE profile of γ–hemolysin. Since the structures are both repeated hairpin, there is only one data point for each amino acid residue in every depth of their profiles.

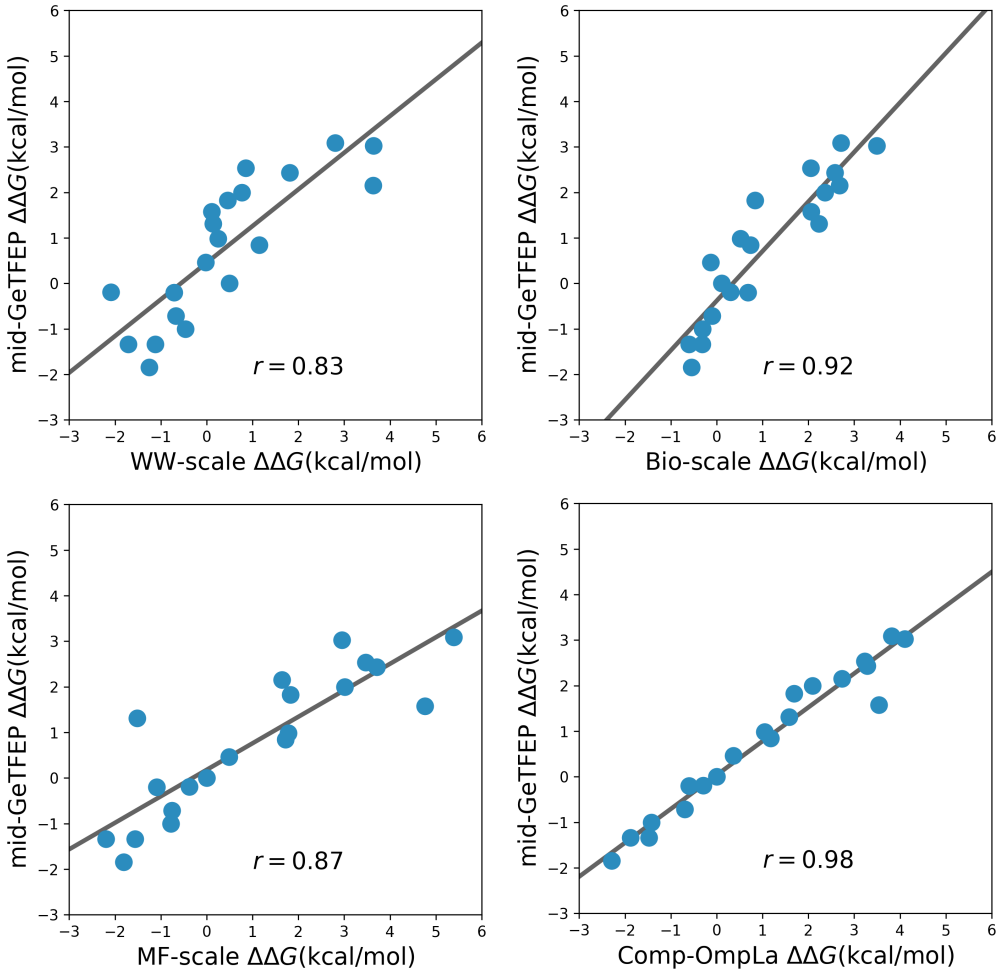

Figure S3: **Comparison between the mid–GeTFEP scale and other hydrophobicity scales.** The mid-GeTFEP scale agrees well with previously measured or derived hydrophobicity scales.

Table S1: The insertion TFEs of WT and lipid–facing residue–shuffled βMPs calculated with GeTFEP. The $\Delta\Delta G$ shows the differences between TFEs of the WT βMPs at step 0 and the average of the minimum TFEs of the lipid–facing residue–shuffled βMPs.

PDB ID	Position w/ min $\Delta\Delta G$	min $\Delta\Delta G$	mis–insertion # (of 2,000)	$\Delta\Delta\Delta G$
1bxw	0	-23.58	616	1.63
1e54	0	-46.03	66	7.90
1ek9	0	-58.23	403	10.35
1fep	0	-61.58	35	8.53
1i78	0	-30.88	216	4.58
1k24	0	-33.41	163	1.99
1kmo	0	-60.55	16	7.95
1nqe	0	-49.83	91	6.18
1p4t	0	-25.68	282	3.61
1prn	0	-43.02	959	9.74
1qd6	0	-34.92	559	7.02
1qj8	0	-21.53	196	4.15
1t16	0	-45.64	08	9.88
1thq	0	-23.10	253	5.52
1tly	0	-28.58	415	2.42
1uyn	0	-33.95	128	4.81
1xkw	0	-59.91	133	9.54
2erv	0	-24.56	398	4.54
2f1c	0	-43.85	473	0.14
2f1t	0	-23.14	280	3.00
2fcp	0	-52.14	717	6.73
2lhf	0	-16.19	757	2.89
2lme	0	-34.17	525	-2.64
2mlh	0	-26.15	676	1.55
2mpr	0	-36.30	688	9.95
2o4v	0	-48.16	190	7.11
2omf	0	-37.71	98	8.65
2por	0	-32.70	994	8.12
2qdz	0	-52.23	289	10.44
2vqi	0	-43.92	441	15.28
2wjr	0	-33.58	72	5.30
2ynk	0	-44.68	141	7.75
3aeh	0	-25.59	471	6.23
3b07	0	-5.14	1670	0.67
3bs0	0	-47.60	95	6.18
3csl	0	-70.94	11	9.55
3dwo	0	-48.32	41	9.17
3dzm	0	-33.27	86	1.26
3kvn	0	-36.38	116	6.60
3pik	0	-36.54	779	4.83
3rbh	0	-41.60	162	6.78
3syb	0	-45.97	236	10.70
3szv	0	-55.25	129	12.65
3v8x	0	-58.94	68	13.24
3vzt	0	-25.56	1018	4.71
4c00	0	-36.32	147	6.76
4e1s	0	-36.20	112	7.20
4gey	0	-49.42	164	8.25
4k3c	0	-48.53	92	6.36
4pr7	0	-38.13	104	3.15
4q35	0	-55.56	97	12.52
7ahl	0	-3.93	1246	-0.49
Summary			17.4%	6.36 ± 3.70

Table S2: **The computed insertion TFEs of the 6 βMPs that do not have the minimum energy at position 0.** However, their most stable position is close to 0, and the minimum TFEs are close to their TFEs at position 0. Nonetheless, we exclude these 6 βMPs in our other analysis of membrane insertion stability.

PDB ID	Position w/ min $\Delta\Delta G$	min $\Delta\Delta G$	Position 0 $\Delta\Delta G$
1a0s	1	-29.77	-30.92
1yc9	-1	-48.40	-49.94
2gr8	-1	-22.74	-24.37
3fid	-1	-35.96	-36.37
3o44	1	-23.77	-29.70
3rfz	-1	-43.08	-43.46

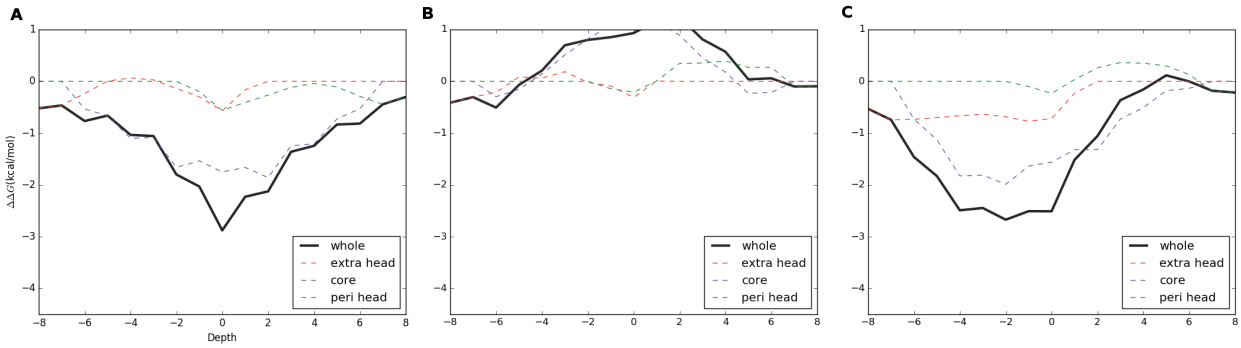

Figure S4: **An example of the insertion TFEs (Omp32, PDB ID:1e54). A.** The insertion TFEs of the intact Omp32 shows a funnel pattern. **B.** The insertion TFEs of the residue–shuffled Omp32 regardless of side–chain directions. **C.** The insertion TFEs of the lipid–facing–residue–shuffled Omp32

Prediction of structurally or functionally important sites. For a lipid–facing residue in a βMP, we calculate the z–score of its TFE by $z = \frac{\text{TFE}-\mu}{\sigma}$, where μ and σ are respectively the mean and the standard deviation values in GeTFEP of the same amino acid in the same depth. When $z > 1.64$ or < -1.64 (which correspond to 5% and 95% in the normal distribution), we take the residue as a deviant that may be structurally and functionally important.

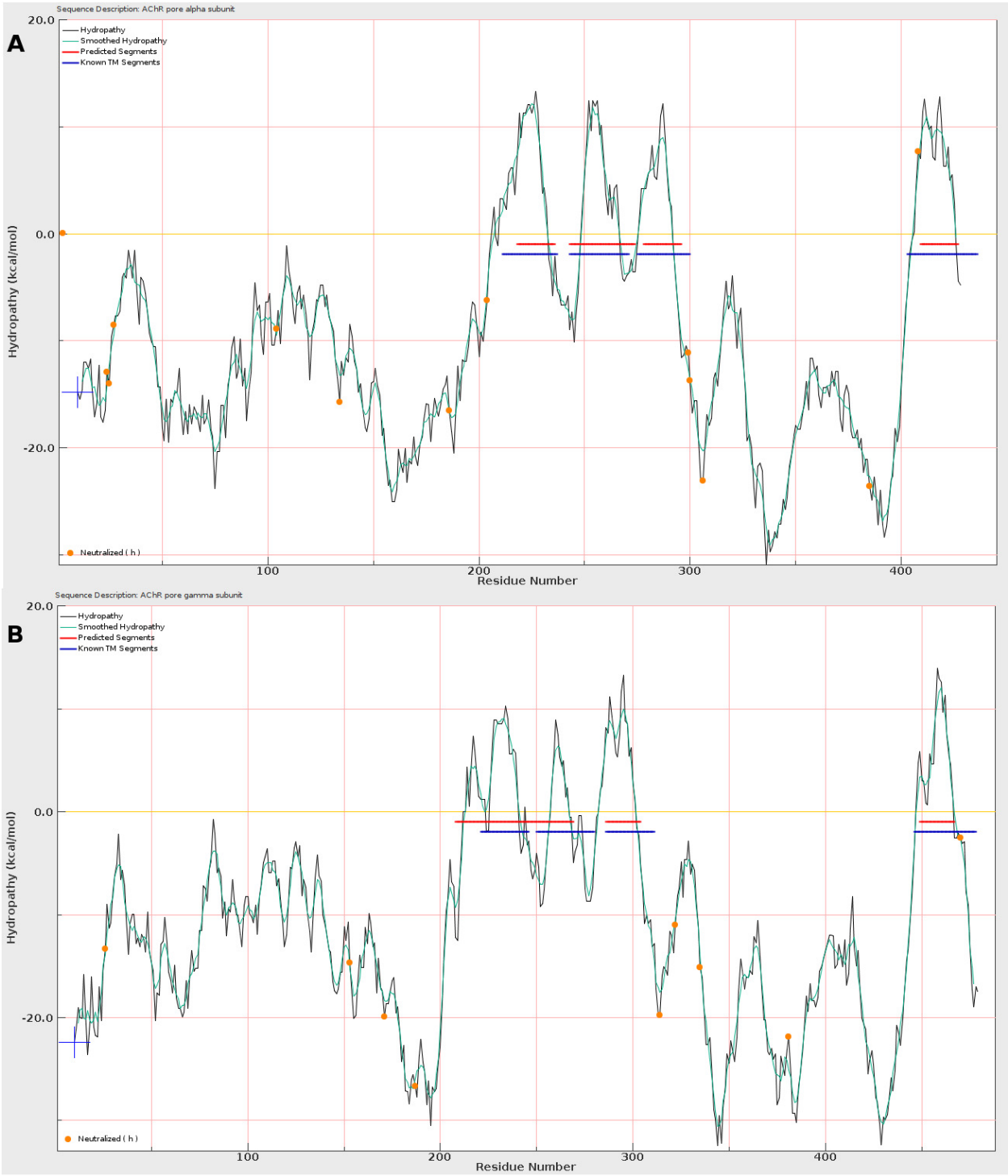

Figure S5: **Hydropathy analysis with mid–GeTFEP.** The blue segments are the known TM segments, while the red ones are predicted by the hydropathy analysis. The analysis was carried out using Membrane Protein Explorer (MPEx)[40] **A.** An example (AChR pore α subunit) shows both the TM region and the number of the TM segments are correctly predicted. **B.** An example (AChR pore γ subunit) shows the predicted number of the TM segments are wrong, though the TM regions are correctly predicted.

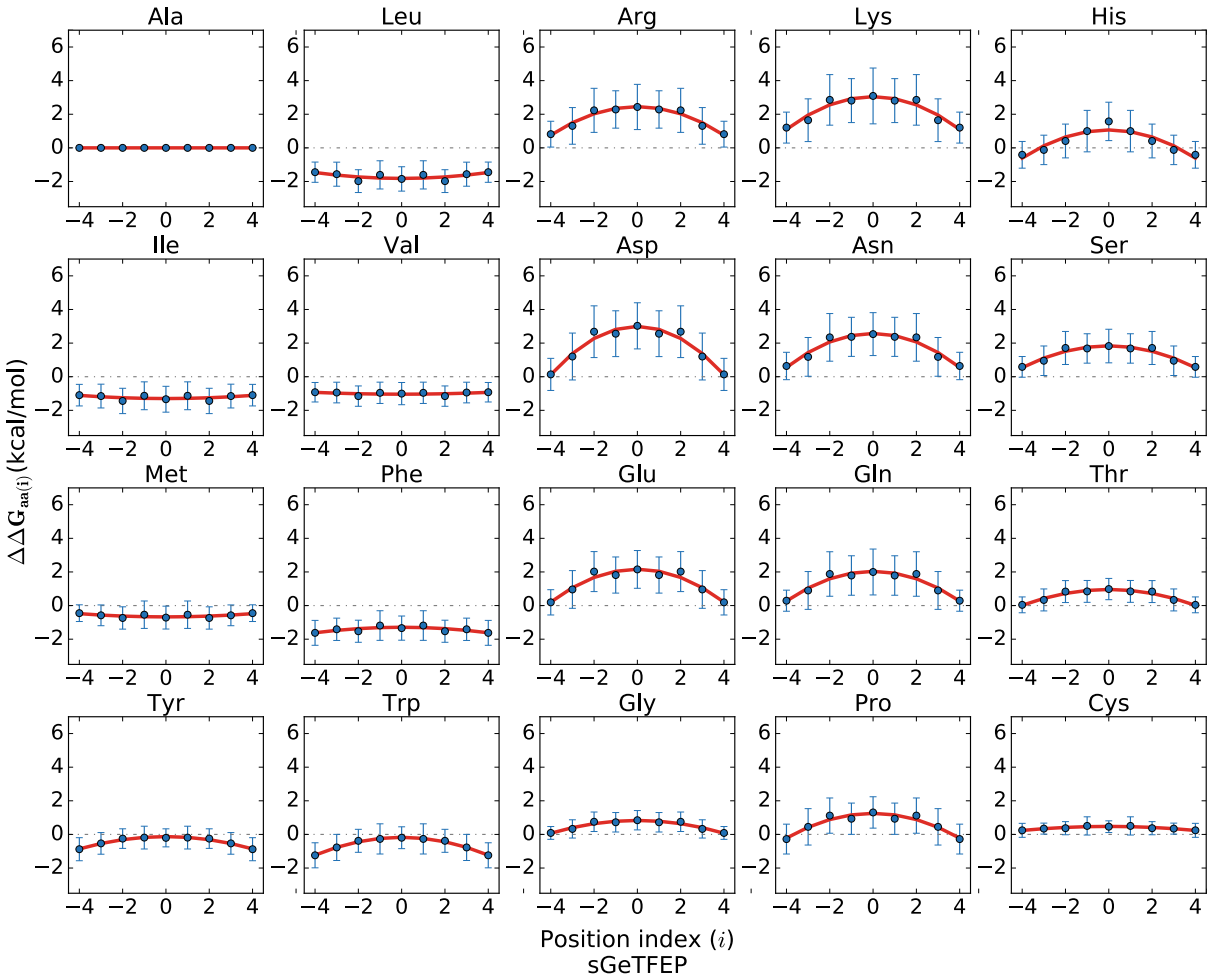

Figure S6: **The sym–GeTFEP for symmetric membranes.** This profile is derived by mirroring the left part (depth -4 to -1) of the original GeTFEP.