GeTFEP: a general transfer free energy profile for transmembrane proteins

Wei Tian, Hammad Naveed, Meishan Lin, Jie Liang

October 2, 2017

Abstract

Transfer free energy (TFE) of amino acid side-chains from aqueous environment into lipid bilayers is an important contributing factor in determining the thermodynamic stability of a transmembrane protein (TMP). It also provides the basis for understanding TMP folding, membrane insertion, and structure-function relationship. We have derived a General Transfer Free Energy Profile (GeTFEP) from β-barrel transmembrane proteins (TMBs). GeTFEP is in good agreement with previous experimentally measured and computationally derived scales. Besides, we show that GeTFEP is applicable to α-helical transmembrane proteins (TMHs) as well by successfully predicting the number and length of transmembrane segments. Application of GeTFEP reveals significant insights into the folding and insertion processes of TMBs. Furthermore, we can predict structurally and/or functionally interesting sites of TMBs using GeTFEP.

1 Introduction

Transmembrane proteins (TMPs) play critical roles in metabolic, regulatory and intercellular processes[1]. Among the two major classes of transmembrane proteins (TMPs), transmembrane α-helical proteins (TMHs) are found predominantly in the plasma membrane of eukaryotic cells, the inner membranes of eukaryotic organelles and prokaryotes. In contrast, transmembrane β-barrel proteins (TMBs) are located in the outer membranes of Gram-negative bacteria, mitochondria, and chloroplasts. Dysfunction or altered function of these proteins can lead to several life-threatening diseases[2]. Knowledge of the thermodynamic stability of TMPs provides a basis for understanding membrane protein folding, stability, insertion, and structure-function relationships. It is therefore of fundamental importance for the development of the medical sciences and biotechnology.

Transfer free energies (TFEs) of amino acid residues from aqueous environment into lipid bilayers are the pivotal contributing factor to the thermodynamic stability of a transmembrane protein[3, 4]. The TFEs of 20 amino acid residues, often called hydrophobicity scales, have been measured experimentally in several systems. The Wimley-White whole residue scale (WW-scale) measures residue partitioning between water and octanol, using a set of peptides as the host of amino acids[5]. The Hessa et al. biological scale (H-scale) measures transferring of residues on polypeptides into the ER membrane through translocon machinery[6]. The Moon-Fleming whole protein scale (MF-scale) measures TFEs of residues from water to membrane core in the context of a real membrane protein structure[7]. These experimentally measured hydrophobicity scales have provided thermodynamic benchmarks, and have been successfully utilized in predicting TM segments in proteins[8].

However, experimental measurement of TFEs is technically challenging, costly and cumbersome [9, 10]. Several hydrophobicity scales have been derived computationally, complementing experiments and expanding our knowledge of the governing principles of membrane protein folding, which can be found in these reviews [11, 12].

The $E_Z\alpha$ and $E_Z\beta$ are knowledge-based hydrophobicity scales with application in several aspects such as positioning TMP in the lipid bilayer, discriminating side-chain decoys, and identifying protein-lipid interfaces[13, 14]. However, these statistical scales ignore the physical interactions between residues either from neighboring helices/strands or within the same helix/strand, which are important for membrane protein insertion and folding[15, 16]. Such detailed interactions can be investigated using molecular dynamics (MD) simulations [17, 18, 19], but the choice of the reference state before membrane insertion remains a challenging problem for MD[18].

We have developed an *ab initio* method that conquers these obstacles by considering intra- and inter-strand interactions among residues in TMBs and calculates the TFE of a given TM residue in a TMB with 14 or less strands[20]. Our method calculates the free energy of a TMB by enumerating its conformations in a reduce state space. As the conformational state space grows rapidly with the strand number, calculation on a TMB with more strands is time consuming. For a specific host position of a given TMB, the method calculates the free energies of the TMBs with the residue at the host position replaced by the amino acid of interest and by an Ala, respectively. The TFE of the amino acid at this position is the difference between the two free energies. Our results are in excellent agreement with the MF-scale with a pearson correlation coefficient (r) of 0.90. We have further improved the method with

several approximation schemes that reduces the running time greatly without loss of the accuracy[21] enabling us to calculate TFEs efficiently of all the TMBs known so far (up to 26 strands).

In this study, we derive a General Transfer Free Energy Profile (GeTFEP) from a non-redundant set of TMBs. We show that this transfer free energy profile is general and is applicable to TMHs as well. Moreover, it provides insights into the membrane insertion of TMBs, and can be used to predict functional and structural interesting sites of TMBs.

2 Results and discussion

Derivation of GeTFEP

Using the methods we previously developed[20, 21], we calculated the depth-dependent TFE profiles for TMBs in a non-redundant dataset of 58 TMBs. Although these TMBs are in different assembly states, have different size (strand numbers), and come from different organisms, their TFE profiles are remarkably similar (see Fig. S1A for example). Cluster analysis and principle component analysis (PCA) of the TFE profiles of TMBs show that only one group (56 TMBs) and two outliers (α-hemolysin (PDB ID: 7ahl), γ-hemolysin (PDB ID: 3b07)) exist. (Fig. 1A and 1B).

We carried out a statistical analysis to determine whether the difference in clustering results are due to the essential distinction between the TFE profiles of α- and γ-hemolysins or the insufficient data points in their profiles (since they both consist of repeated hairpins, see Fig. S1). . We computed TFE profiles for each hairpin in our TMB dataset, and resampled from these profiles. Comparison with

the resampled hairpin TFE profiles shows that α- and γ-hemolysins are not significantly different from other hairpins (Fig. 1C). Therefore, we conclude that a general transfer free energy profile exists for all TMBs, and we derive the GeTFEP by averaging the TFEs of a particular amino acid at the same lipid bilayer depth (Fig. 1D). GeTFEP shows asymmetry in TFEs in the inner and the outer layers of the membrane bilayer, consistent with the asymmetric nature of bacterial outer membrane bilayer, which has an external monolayer of lipopolysaccharide. For the other more symmetric membrane bilayers, we derived a symmetric TFE profile, sGeTFEP, as well (Fig. S2), by mirroring the regions of the inner membrane layer of GeTFEP.

Comparison with previous hydrophobicity scales

We first examined if GeTFEP is comparable with previous measured hydrophobicity scales. Since most experimentally measured scales can not account for the anisotropism of lipid bilayers, we only compared them with the results of the most hydrocarbon core (depth 0) of GeTFEP (Fig. 2). We refer this hydrophobicity scale as GeTFEP-mid hereafter. GeTFEP-mid correlates well with the experimentally measured hydrophobicity scales, with a pearson correlation coefficient (r) of 0.83 for all 20 amino acids with WW-scale, and 0.92 with H-scale. Particularly, GeTFEP-mid has a $r = 0.87$ correlation with MF-scale, which was measured using a TMB, OmpLa, as the host system. GeTFEP-mid correlates with the computational OmpLa scale[20, 21] as well with $r = 0.90$ (Fig. S3).

The TFE value of His is less unfavorable in GeTFEP as compared to the MF-scale(Fig. 2). Given that MF-scale was measured in acidic condition (pH=3.8), where His residue was probably completely

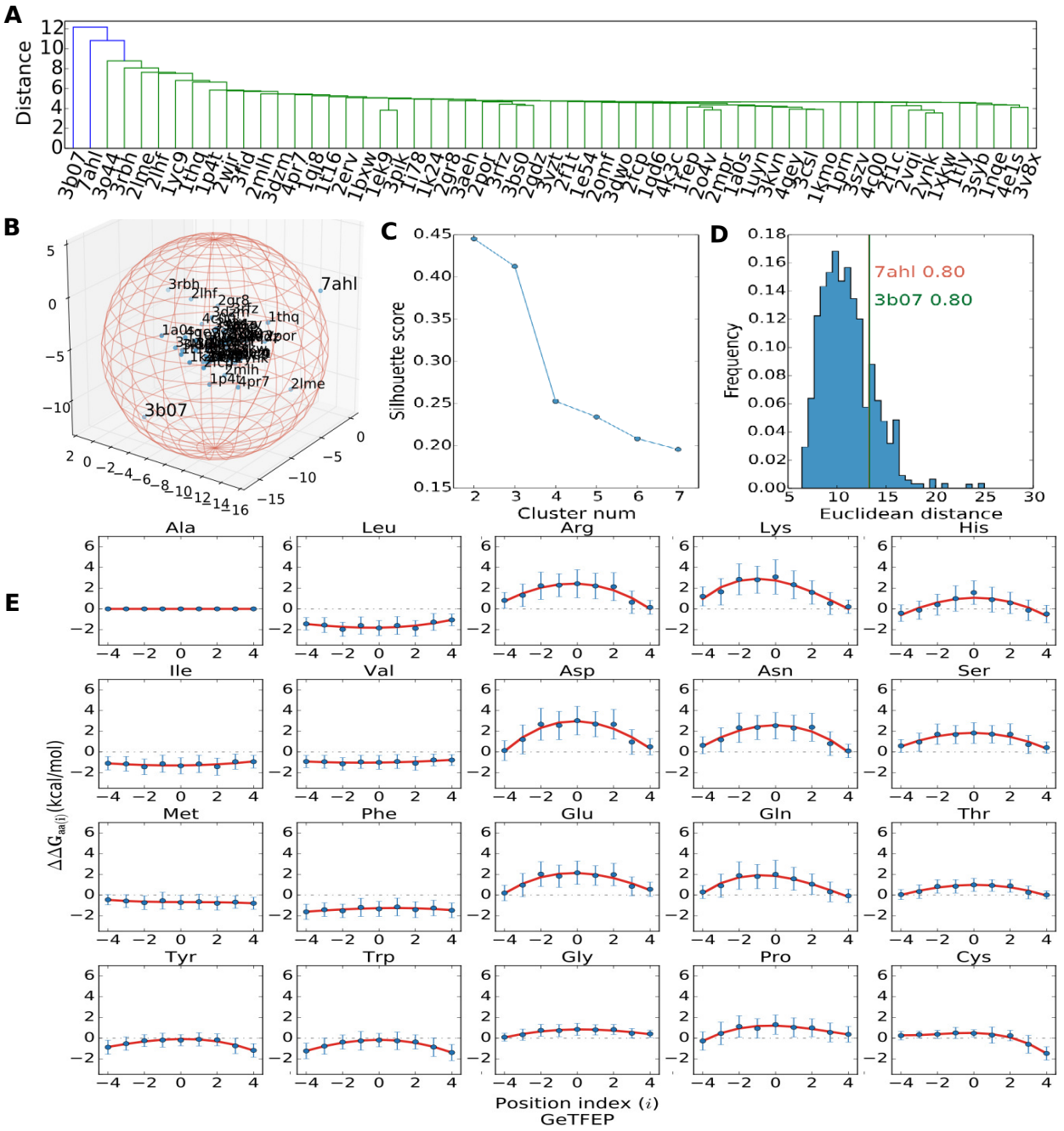

Figure 1: Derivation of GeTFEP. **A.** Hierarchical clustering results of 58 TMB TFE profiles. **B.** Visualization of 58 TFE profiles in the 3D profile space. Data dimension were reduced to 3 via PCA. The radius of the sphere is 10, determined by where the jumping happens in the inter-cluster distances in A. TMB 7ahl is outside the sphere, and 3b07 is just in the sphere. **C.** Clustering quality (silhouette scores) decreases with the increase of presumed cluster number. A relative low silhouette score when cluster number is 2 indicates all 58 TMB TFE profiles belong to one cluster. **D.** The distribution of distance between resampled hairpin TFE profiles and the average profile of all the resampled hairpins shows that neither the TFE profile of 7ahl nor the one of 3b07 are significantly different from hairpins of other TMBs. Both 7ahl and 3b07 are around the 80th percentile. **E.** GeTFEP of each residues (blue), and the corresponding curves fitted by 3rd degree polynomials (red).

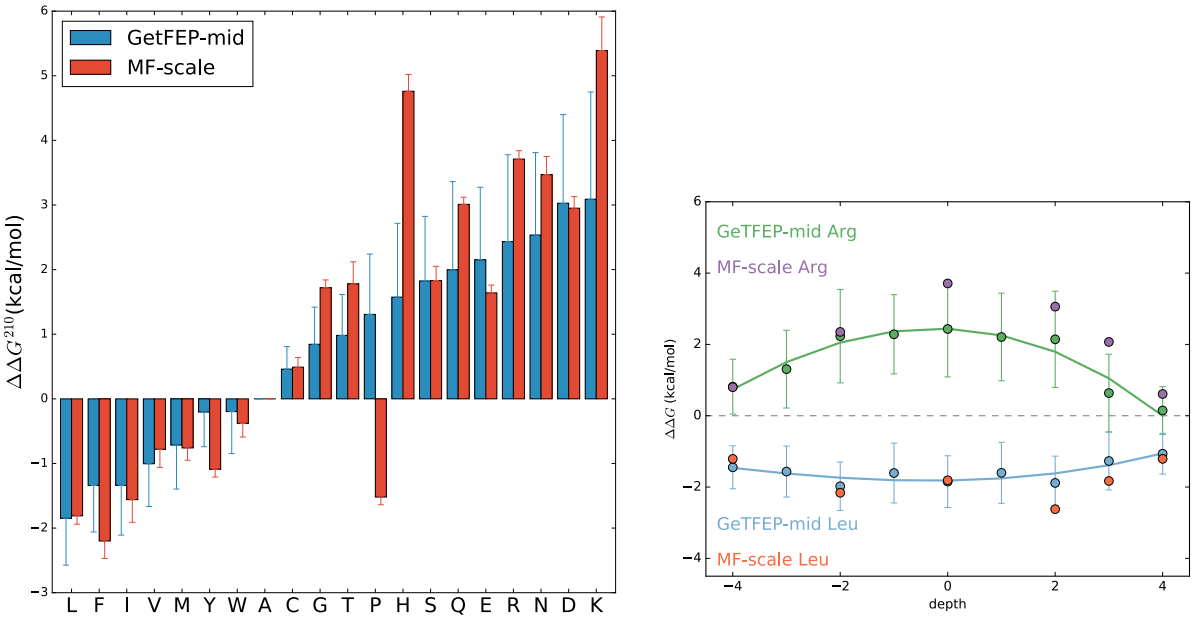

Figure 2: Comparison between GeTFEP and TFEs measured in TMBs. **A.** Comparison between GeTFEP-mid and MF-scale shows a good agreement except Pro and His. **B.** The depth-dependent TFEs of Arg and Leu of GeTFEP agrees with those measured in the same study of MF-scale.

protonated [7], the TFE value of His in GeTFEP-mid may therefore better reflect its properties in physiological conditions.

Another notable difference can be found for Pro, which is ranked as favorable as Ile in MF-scale than in GeTFEP-mid, (Fig. 2). Ile is more often observed in TM regions than Pro not only in TMBs[22] but also in TMHs[13]. Considering that Pro notoriously tends to disrupt structures of both α-helix and β-sheet, which is thermodynamically unfavorable in the non-polar core of bilayers[23], the TFE of Pro in GeTFEP-mid may reflect the energetic role of Pro better in a more general protein structure (more discussion on Pro in the next section).

GeTFEP shows that TFE of a residue depends on the depth where the residue is transferred to, reflecting the environmental anisotropy in lipid bilayers. This depth-dependency of GeTFEP agrees well with previously reported experiments[7] of Arg ($r = 0.87$) and of Leu ($r = 0.75$, Fig. 2B).

Generality of GeTFEP

It was previously suggested that hydrophobicity scales measured in TMB systems may also be applicable for TMHs[7]. To test the generality of GeTFEP, we performed traditional hydropathy analysis[24] on TMHs from MPTopo database[25] using GeTFEP-mid. For 90 of the 131 (∼69%) proteins or subunits in the dataset, GeTFEP-mid correctly predicted both TM regions and numbers of their TM segments, which performs much better than the other hydrophobicity scales, including the scales measured or derived from α-helical structure (Tab. 1). For most of the other proteins, GeTFEP correctly predicted TM regions, but mispredicted the number of TM segments due to concatenation or breaking of TM segments (see Fig. S4B for example). We also examined the number of TM residues correctly predicted by GeTFEP-mid. It achieves a ∼85% precision and a ∼71% recall (or sensitivity), which is much better than the other hydrophobicity scales (Tab. 1). These results imply that GeTFEP is not TMB specific, and is applicable to general TMP architectures.

To further test the impact of Pro, which is quantitatively different in GeTFEP-mid from MF-scale, we swapped the value of Pro from MF-scale into GeTFEP-mid, and the performance of this hybrid scale drops vastly (Pro-swapped in Tab. 1). This implies that a more lipid bilayer unfavorable Pro could fit the hydrophobic environment better, consistent with our discussion in the previous section. Instead of being a quantification of pure hydrophobicity of a single residue, GeTFEP is more a quantification of energetic cost of transferring a residue from water into lipid bilayers in a general TMP structure.

Hydrophobicity scale	TMHs % (#) with TM segs. correctly predicted	TM res. precision	TM res. recall	TM res. F-measure
WW-scale	50%(66)	73%	75%	0.74
H-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
GeTFEP-mid	69%(90)	85%	71%	0.78

Table 1: GeTFEP-mid performs better than the other hydrophobicity scales in predicting TM segments and residues. The first three scales are measured or derived in TMHs, the others TMBs.

TMB insertion is a spontaneous process driven by thermostability

Unlike TMHs, after synthesis in cytoplasm, TMBs need to be sorted across periplasm, which lacks an energy source such as ATP, and is then folded into membranes. Experimentally measured thermodynamic parameters suggest that folding free energies ensure successful periplasm translocation[26]. Computational results identify TFE of lipid-facing residues (LFRs) of hydrophobic core regions of TMBs are the main driving force for TMBs to insert into membrane[20]. However, knowledge is still lacking about what happens thermodynamically during the membrane insertion process. Moreover, it is unclear if amino acid compositions of TMBs plays a role in the insertion process, given that LFRs of TMBs have clear location patterns[22].

To answer these questions, we used a simplified TMB insertion model that ignores contributions from the loops with 17 discrete steps (Fig. 3A). TMBs start from periplasm and fully insert into the membrane from steps -8 to 0, which is based on the concerted folding mechanism proposed in [27]. From steps 0 to +8, TMBs translocate across the membrane, which is a thought experiment. For toxin TMBs, insertion process should be reversed, from -8 to +8. Assuming that the stability of a TMB can be approximated by the additive model which summarizes TFEs of all LFRs inside the membrane region, stability of the

TMB from steps -8 to +8 can be calculated with GeTFEP (or sGeTFEP for toxin TMBs).

Results of all the TMBs show a funnel like energetic pattern (Fig. 3B). Most (52 of 58) TMBs reach their free energy minimum when they are fully inserted into membranes (step 0, Tab. S1). The funnel pattern indicates that insertion of TMBs into outer membranes is a spontaneous process as expected. TMBs are then energetically trapped after being fully inserted. For the TMBs (6 of 58) which are not most stable when fully inserted, the mismatch could come either from wrong fully inserted positions (see Materials and methods for details) or insufficiency of the additive model. However, the most stable steps of all these TMBs are close to step 0 (steps 1 or -1), and the minimum TFEs are close to the TFEs of step 0 as well (Tab. S2). Nevertheless, we will use only the 52 TMBs with energetic minimum in step 0 for the further tests in the next section.

Residue composition and location in TMBs are important in the insertion process

TMBs are known as "inside-out" proteins, where charged/polar residues are enriched among pore-facing residues of TMBs, while LFRs are mostly apolar. To test if the insertion funnel pattern comes merely from the extensive property of the TFEs of these hydrophobic residues in the additive model, we shuffled the residues within each TMB while keeping its structure (the side-chain direction of each position in one β-strand and the interstrand pairing) unchanged. When the residues are shuffled regardless of their original side-chain directions, it is highly unfavorable for the shuffled TMB to be inserted into the membrane as expected. When only LFRs are shuffled, insertion of the shuffled TMBs is energetically favorable

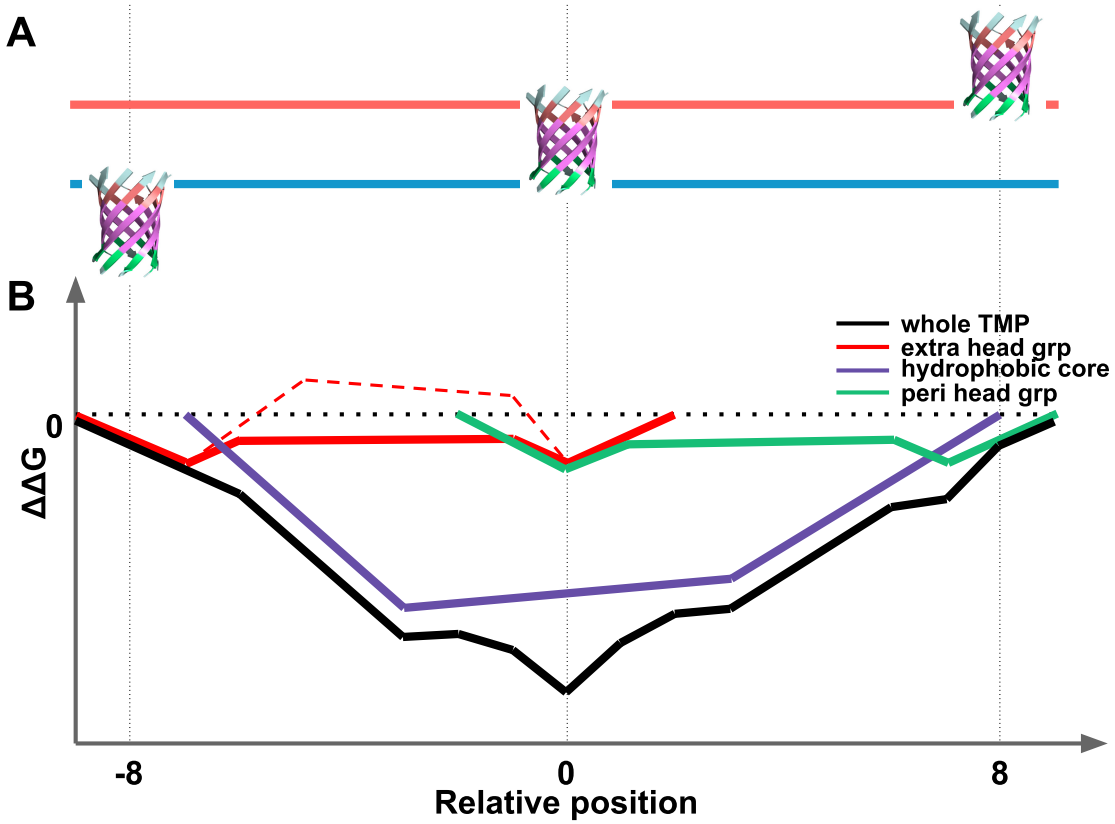

Figure 3: TMB insertion into the membrane. **A.** Simplified TMB insertion model with 17 steps **B.** Illustration of how free energies change with TMB position in the membrane. The dashed red segments show that LFRs of extracellular head group sometimes become energetically unfavorable in the membrane.

(see Fig. S5 for example). However, fully inserted position (step 0) is not the most stable position in 17.4% cases of the LFR-shuffled TMBs (Tab. S1). In addition, the fully inserted LFR-shuffled TMBs are unstable than the fully inserted WT TMBs for 50 out of 52 TMBs tested, and the insertion energy required by the shuffled TMBs is on average 6.36 kcal/mol larger than those of the WT TMBs (Tab. S2). These results indicate the composition and the location of LFRs in the TM region are "optimized" to be very stable in the membrane environment.

We further divided the TM segment of a TMB into three parts, periplasmic headgroup, hydrophobic core, and extracellular headgroup[22], to investigate how these regions of a TMB contribute to the overall thermodynamic stability. A similar energetic pattern are shared among all 52 TMBs (Fig. 3B). LFRs of the extracellular headgroup initialize the insertion process as they are energetically favorable in the interfacial region of the membrane on the periplasmic side (steps -8 and -7), and then become less favorable (sometimes unfavorable), while LFRs of hydrophobic core start to be inserted and strongly drive the process (steps -6 to -2). When LFRs of extracellular headgroup approach the interfacial region on the extracellular side, they become energetically favorable again. In the meantime, LFRs of the periplasmic headgroup are inserted (steps -1 and 0), and the TFE of the whole TMB reaches its minimum (step 0).

Although it was previously shown that LFRs of the hydrophobic core is the main driving force for TMBs to be inserted into membrane[20], interestingly, we observed that the TFEs of hydrophobic core never reach their minimum when TMBs are fully inserted in all 52 cases, while TFEs of the whole TMBs reach their minimum at the fully inserted position (step 0). The "W" shape free energy curves of the two head group regions indicates that LFRs there act like "energetic latches" to lock TMBs in their fully

inserted position (Fig. 3B).

Together with the residue shuffling results, it can be concluded that the energetic pattern of TMB insertion is not simply from extensive properties of the TFEs and the model, and the residue location patterns have thermodynamic impacts.

Predicting orientations and positions of TMBs

To further confirm that GeTFEP does capture the anisotropism of the membrane bilayer, we used GeTFEP to predict the orientation and position of TMBs in membranes (Fig. 4A). The membrane is modeled as an infinite slab with certain width $2h_{1/2}$. Each TMB is initially positioned in the membrane with its mass center on the midplane of the membrane and its barrel axis aligned with the normal of the membrane. A systematic searching for membrane width and the rigid body rotation angles and the translation displacement of the TMB was then carried out. The positions of the lipid facing residues of the barrel and of the residues of the loops are used to calculated the total TFE of the TMB with GeTFEP (Fig. 4B). The orientations and the positions of TMBs with the lowest TFEs were selected as predictions. The barrel tilt angles from the GeTFEP results correlate well ($r = 0.76$) with those from the widely used Orientations of Proteins in Membranes (OPM) database[28], and the average barrel tilt angle of 7.3°, consistent with the same measures using OPM ($6.2 \pm 1.8°$) for the same dataset. The predicted strand tilt angles and the membrane thickness from GeTFEP is also in good agreement with experimentally determined ones(Tab. 2).

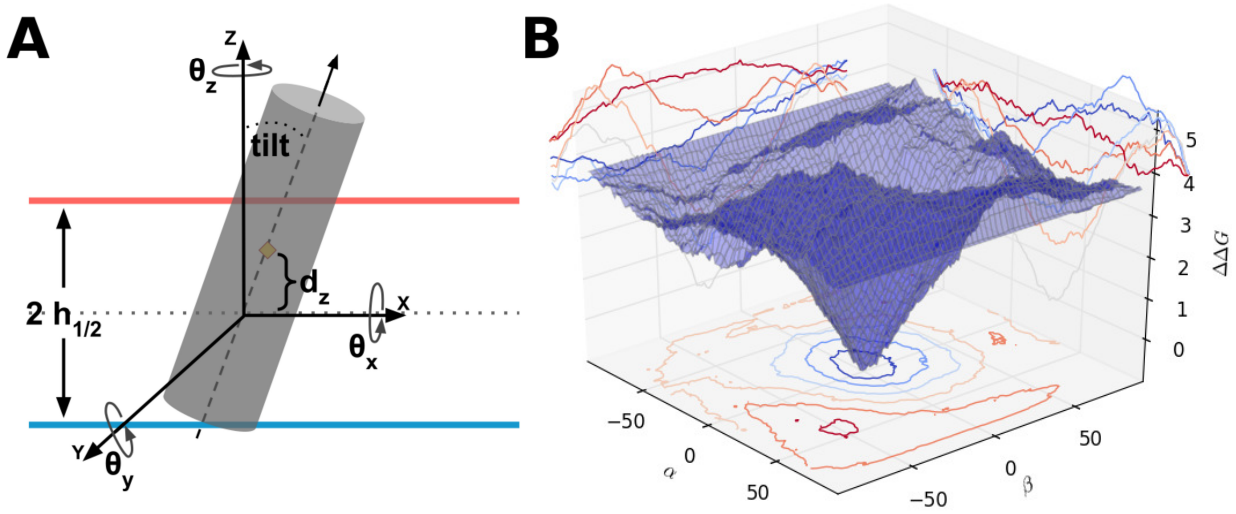

Figure 4: Prediction of positions and orientations of TMB. **A.** A systematic search is carried out for each TMB to determine the hydrophobic region 2 $h_{1/2}$, the translation displacement between the TMB and the membrane d_z, and the tilt angle which depends on the rigid body rotation anglesθ_x, θ_y, and θ_z. **B.** The funnel energetic landscape of an example TMB (BtuB, PDB:1nqe). The landscape shows the TMB TFEs when the protein is rotated around the x- and the y- axes.

	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	$\geq$ 23.1	23.5	24.3
	OmpF	2omf	$\sim$ 21.0	22.8	25.2
	BtuB	1nqe	$\geq$ 20.2	23.0	23.4

Table 2: Comparison between TMB position and orientation predictions using GeTFEP and experimental results. Experimental values were obtained from site-directed spin labeling studies, cryo-electron microscopy data, X-ray scattering or hydrophobic matching experiments[29]. * the experimental values are systematically larger, which could be due to orientational disorder under the experimental conditions, suggesting experimental tilt angles represent the upper bounds of the actual values[29].

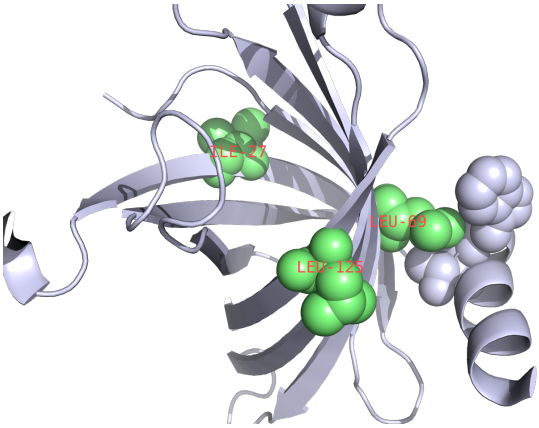

Figure 5: Predicted interesting sites of PagP

Predicting structurally or functionally interesting sites of TMBs

As TFEs of lipid facing residues are the main energetic factor of the thermodynamic stability of TMBs, one wonders if the TFE deviation from GeTFEP of residues can be used to detect structurally or functionally interesting sites of TMBs. We performed a outlier detection procedure (see Materials and Methods for details) on TM LFRs of TMBs with sufficient experimental results (OmpLA, PagP, and PagL). A considerable fraction of the outlier residues have certain functional roles or are structurally different from normal beta barrels (Tab 3). For example, 27I, 69L, 125L and 131L in PagP are detected as interesting sites, where 69L interacts with the out clamp α-helix of the protein, and 27I and 125L are at the lateral routes where β-hydrogen bonding is absent so that substrates get access into the protein interior [30](Fig 5). We cannot compare all of our outlier detection results with experiments due to the limited experiment studies with single residue resolution. However, our attempt shows that without requiring 3D structures of the TMBs, more sophisticated methods considering the thermodynamic stability of residues (for example, using GeTFEP) could be developed to detect sites of interest in TMBs, which can complement/guide design of further experiments.

Protein	Seq ID	Notes
OmpLA(1qd6)	38N	
	40L	Substrate binding [33]
	92Y	Substrate binding [33]
	116P	Interstand neighbor of 92Y and 142H
	120L	
	142H	Catalytic site [34, 35]
	156N	Catalytic site [34, 35]
	237L	
PagP(1thq)	27I	Lateral route from membrane to protein interior [36]
	69L	Interact with the out clamp α-helix [36]
	125L	Lateral route from membrane to protein interior [36]
	131L	
PagL(2erv)	108I	Ligand binding site [37]
	126H	Catalytic site [38]

Table 3: Predicted interesting sites of OmpLA, PagP and PagL

3 Conclusions and outlook

In this study, we computed the TFE profiles of a non-redundant TMB dataset, and showed that these profiles share common properties. General Transfer Free Energy Profile (GeTFEP) was then derived from these TFE profiles, which agrees well with previous experimentally measured and computationally derived scales. Although lipid bilayers have considerable anisotropic heterogeneity along the bilayer normal, experimental measurements of TFEs in depths other than the mid-plane is still lacking. GeTFEP fills this gap. By applying GeTFEP in the hydropathy analysis of TMHs, we showed that GeTFEP performs even better than the hydrophobicity scales measured/calculated in TMH systems. Therefore, GeTFEP reflects the energetic cost of transferring an amino acid side-chain into certain depth of membrane within a general TMP architecture. This may help in improving secondary structure prediction methods.

Using GeTFEP, we explored the energetic contribution of each part of TMBs to the insertion process, and showed that amino acid residue composition and location of β-strands of TMBs are optimized to fit their environment. We are also able to predict the position and orientation of TMBs inside membranes

with GeTFEP. Moreover, we demonstrated that deviation of TFEs from GeTFEP can be used to detect structurally or functionally interesting sites of TMBs with a naive outlier detection method. As engineering of transmembrane pore-forming proteins are drawing increasing attention in bionanotechnology such as DNA sequencing[39][40] and molecule detection[41], GeTFEP may provide insights in designing stable bionanopores and in tailoring and engineering their structures and functions.

4 Materials and methods

Dataset

We use 58 non-homologous β-barrel membrane proteins (resolution 1.45Å– 3.2Å) with less than 30% pairwise sequence identity for this study. The pdb codes 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.

Cluster analysis of TMB TFE profiles

Euclidean distance between the TFE profiles of the TMBs and single linkage are used in the hierarchical clustering. The conclusion remains the same if correlation distance and/or other reasonable linkages (eg. average linkage or weighted linkage) are used.

Insertion of TMBs

The concerted insertion process is discretized into 17 steps, where in each step one layer of lipid-facing residues is either inserted into the membrane (step -8 to 0) or pulled out of the membrane (step 0 to +8). The fully inserted positions of TMBs at step 0 were determine using the OPM database [28].

Predicting structurally or functionally interesting sites

For a lipid-facing residue in a TMB, we calculate the z-score of its TFE by $z = \frac{\mathrm{TFE}-\mu}{\sigma}$, 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 outlier that may be structurally and functionally interesting.

References

[1] James W. Fairman, Nicholas Noinaj, and Susan K. Buchanan. The structural biology of β-barrel membrane proteins: A summary of recent reports, 2011.

[2] Charles R. Sanders and Joanna K. Nagy. Misfolding of membrane proteins in health and disease: The lady or the tiger?, 2000.

[3] C Tanford. The hydrophobic effect and the organization of living matter. *Science (New York, N.Y.)*, 200(4345):1012–1018, 1978.

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

[5] W C Wimley and S H White. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature structural biology*, 3:842–848, 1996.

[6] Tara Hessa, Hyun Kim, Karl Bihlmaier, Carolina Lundin, Jorrit Boekel, Helena Andersson, Ingmarie Nilsson, Stephen H White, and Gunnar von Heijne. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature*, 433(7024):377–81, jan 2005.

[7] C Preston Moon and Karen G Fleming. 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*, 108(25):10174–7, jun 2011.

[8] C M Deber, C Wang, L P Liu, a S Prior, S Agrawal, B L Muskat, and a J Cuticchia. 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*, 10(1):212–219, 2001.

[9] C. Preston Moon, Sarah Kwon, and Karen G. Fleming. Overcoming hysteresis to attain reversible equilibrium folding for outer membrane phospholipase A in phospholipid bilayers. 413(2):484–494, 2011.

[10] Daniel E. Otzen and Kell K. Andersen. Folding of outer membrane proteins, 2013.

[11] Jie Liang. Experimental and computational studies of determinants of membrane-protein folding,

2002.

[12] Jie Liang, Hammad Naveed, David Jimenez-Morales, Larisa Adamian, and Meishan Lin. Computational studies of membrane proteins: Models and predictions for biological understanding, 2012.

[13] Alessandro Senes, Deborah C. Chadi, Peter B. Law, Robin F S Walters, Vikas Nanda, and William F. DeGrado. 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*, 366(2):436–448, 2007.

[14] Daniel Hsieh, Alexander Davis, and Vikas Nanda. 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*, 21(1):50–62, jan 2012.

[15] David T. Moore, Bryan W. Berger, and William F. DeGrado. Protein-Protein Interactions in the Membrane: Sequence, Structural, and Biological Motifs, 2008.

[16] Linnea E. Hedin, Karin Öjemalm, Andreas Bernsel, Aron Hennerdal, Kristoffer Illergård, Karl Enquist, Anni Kauko, Susana Cristobal, Gunnar von Heijne, Mirjam Lerch-Bader, IngMarie Nilsson, and Arne Elofsson. Membrane Insertion of Marginally Hydrophobic Transmembrane Helices Depends on Sequence Context. *Journal of Molecular Biology*, 396(1):221–229, 2010.

[17] Justin L MacCallum, W F Drew Bennett, and D Peter Tieleman. Partitioning of amino acid side chains into lipid bilayers: results from computer simulations and comparison to experiment. *The Journal of general physiology*, 129(5):371–377, 2007.

[18] James Gumbart and Benoît Roux. Determination of membrane-insertion free energies by molecular dynamics simulations. *Biophysical Journal*, 102(4):795–801, 2012.

[19] Martin B. Ulmschneider, Jakob P Ulmschneider, Nina Schiller, B a Wallace, Gunnar von Heijne, and Stephen H White. Spontaneous transmembrane helix insertion thermodynamically mimics translocon-guided insertion. *Nature Communications*, 5:4863, 2014.

[20] Meishan Lin, Dennis Gessmann, Hammad Naveed, and Jie Liang. Outer Membrane Protein Folding and Topology from a Computational Transfer Free Energy Scale. *Journal of the American Chemical Society*, 138(8):2592–2601, 2016.

[21] Wei Tian, Meishan Lin, Hammad Naveed, and Jie Liang. Efficient computation of transfer free energies of amino acids in beta-barrel membrane proteins. *Bioinformatics (Oxford, England)*, (January):1–8, jan 2017.

[22] Ronald Jackups and Jie Liang. Interstrand pairing patterns in beta-barrel membrane proteins: the positive-outside rule, aromatic rescue, and strand registration prediction. *Journal of molecular biology*, 354(4):979–93, dec 2005.

[23] Alessandro Senes, Donald E Engel, and William F DeGrado. Folding of helical membrane proteins: the role of polar, GxxxG-like and proline motifs. *Current opinion in structural biology*, 14(4):465–79, 2004.

[24] J Kyte and R F Doolittle. A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology*, 157(1):105–132, 1982.

[25] S Jayasinghe, K Hristova, and S H White. MPtopo: A database of membrane protein topology. *Protein Science*, 10(2):455–458, 2001.

[26] C Preston Moon, Nathan R Zaccai, Patrick J Fleming, Dennis Gessmann, and Karen G Fleming. 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*, 110(11):4285–90, mar 2013.

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

[28] Mikhail A. Lomize, Andrei L. Lomize, Irina D. Pogozheva, and Henry I. Mosberg. OPM: Orientations of proteins in membranes database. *Bioinformatics*, 22(5):623–625, 2006.

[29] Andrei L. Lomize, Irina D. Pogozheva, Mikhail a. Lomize, and Henry I. Mosberg. Positioning of proteins in membranes: A computational approach. *Protein Science*, 15(6):1318–1333, 2006.

[30] Russell E. Bishop. Structural biology of membrane-intrinsic β-barrel enzymes: Sentinels of the bacterial outer membrane. *Biochimica et Biophysica Acta - Biomembranes*, 1778(9):1881–1896, 2008.

[31] H. Naveed, Y. Xu, R. Jackups, Jr., and J. Liang. Predicting three-dimensional structures of transmembrane domains of beta-barrel membrane proteins. *J Am Chem Soc*, 134(3):1775–1781, Jan 2012.

[32] Wei Tian, Meishan Lin, Jie Liang, and Hammad Naveed. High resolution structure prediction of β-barrel membrane proteins. *In preparation*.

[33] H J Snijder, I Ubarretxena-Belandia, M Blaauw, K H Kalk, H M Verheij, M R Egmond, N Dekker, and B W Dijkstra. Structural evidence for dimerization-regulated activation of an integral membrane phospholipase. *Nature*, 401(6754):717–721, 1999.

[34] Niek Dekker. Outer-membrane phospholipase A: Known structure, unknown biological function, 2000.

[35] R L Kingma, M Fragiathaki, H J Snijder, B W Dijkstra, H M Verheij, N Dekker, and M R Egmond. Unusual catalytic triad of Escherichia coli outer membrane phospholipase A, 2000.

[36] PDB structure of PagP. doi:10.2210/pdb1thq/pdb.

[37] PDB structure of PagL. doi:10.2210/pdb2erv/pdb.

[38] Lucy Rutten, Jeroen Geurtsen, Wietske Lambert, Jeroen J M Smolenaers, Alexandre M Bonvin, Alex de Haan, Peter van der Ley, Maarten R Egmond, Piet Gros, and Jan Tommassen. 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*, 103(18):7071–7076, 2006.

[39] I M Derrington, T Z Butler, M D Collins, E Manrao, M Pavlenok, M Niederweis, and J H Gundlach. Nanopore DNA sequencing with MspA. *Proc Natl Acad Sci U S A*, 107(37):16060–16065, 2010.

[40] David Stoddart, Andrew J Heron, Ellina Mikhailova, Giovanni Maglia, and Hagan Bayley. Single-

nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proceedings of the National Academy of Sciences of the United States of America*, 106(19):7702–7707, 2009.

[41] Monifa A. Fahie, Bib Yang, Martin Mullis, Matthew A. Holden, and Min Chen. Selective Detection of Protein Homologues in Serum Using an OmpG Nanopore. *Analytical Chemistry*, 87(21):11143–11149, 2015.

[42] Craig Snider, Sajith Jayasinghe, Kalina Hristova, and Stephen H. White. MPEx: A tool for exploring membrane proteins. *Protein Science*, 18(12):2624–2628, 2009.

Supplementary information

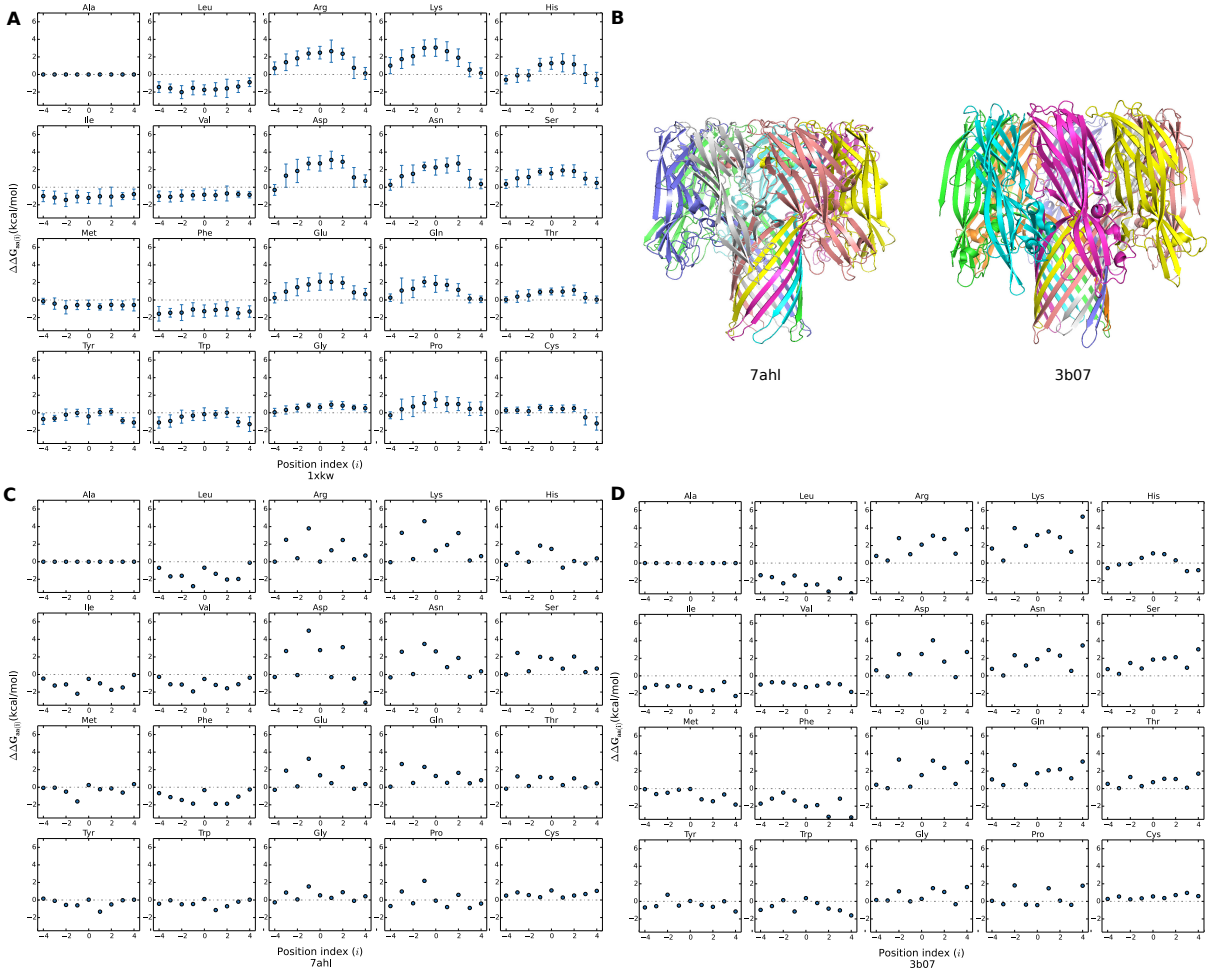

Figure S1: **A.** A typical TFE profile of TMBs (FptA, PDB id:1xkw). **B.** The structures of α-hemolysin (PDB id:7ahl) and γ-hemolysin (PDB id:3b07). **C.** The TFE profile of α-hemolysin. **D.** The TFE profile of γ-hemolysin. Since the structures are both repeated hairpin, there is only one data point for each amino acid residue in every depth of their profiles.

PDB code	min ΔG step	WT ΔG	mis-insertion # (of 2000)	$\Delta\Delta G$
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 S1: The insertion TFEs of WT and LFR-shuffled TMBs calculated with GeTFEP. The $\Delta\Delta G$ shows the differences between TFEs of the WT TMBs at step 0 and the average of the minimum TFEs of the LFR-shuffled TMBs.

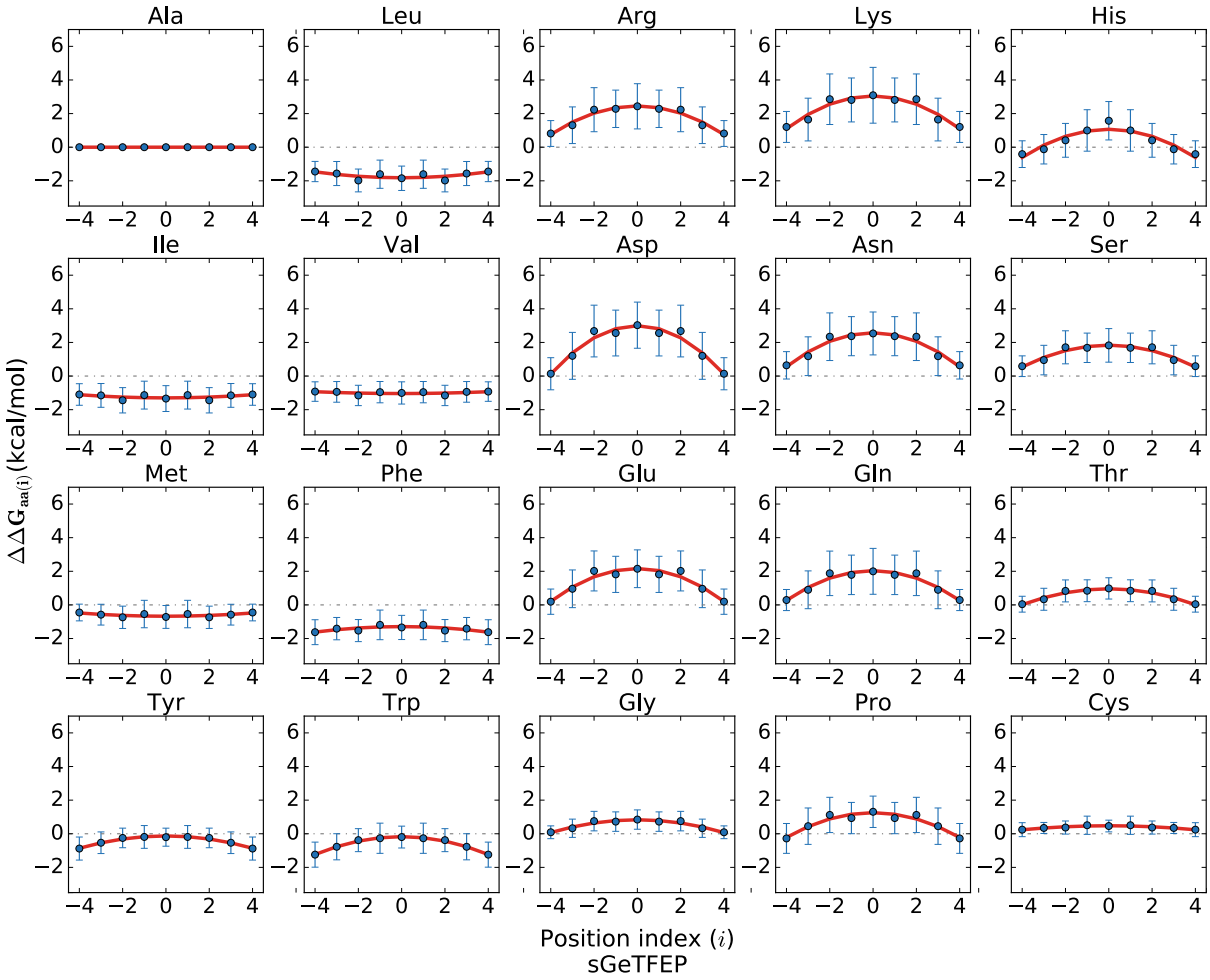

Figure S2: The symmetric version of GeTFEP. This profile is derived by replacing the right part (depth 1 – 4) of the GeTFEP by the mirror of the left part (depth -4 – -1)

PDB code	min ΔG step	min ΔG	step 0 ΔG
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

Table S2: The insertion TFEs calculated with GeTFEP.

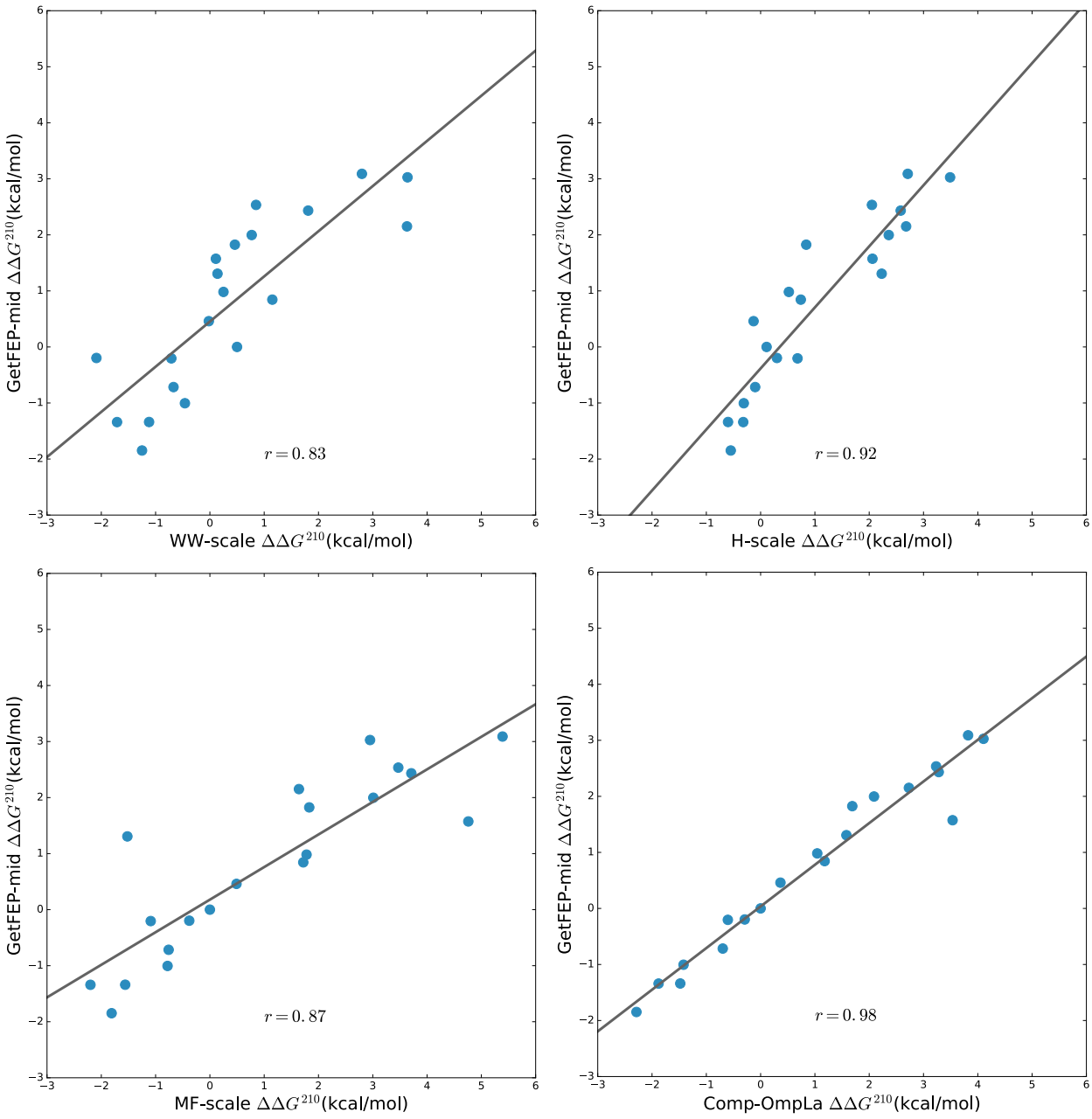

Figure S3: Comparison between GeTFEP and other hydrophobicity scales.

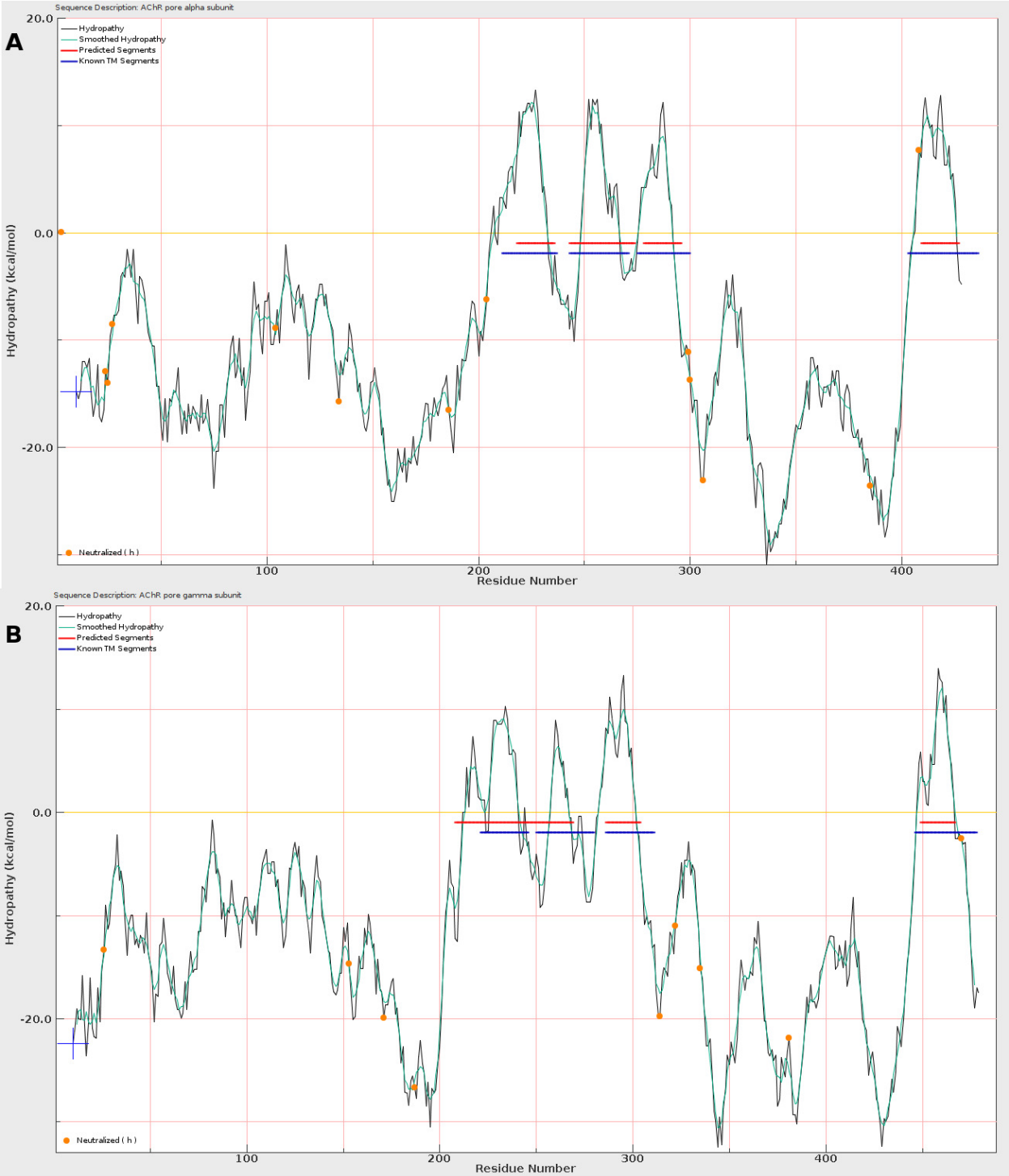

Figure S4: Hydropathy analysis with GeTFEP-mid. The blue segments are the known TM regions, while the red ones are predicted by the hydropathy analysis. The analysis was carried out using Membrane Protein Explorer (MPEx) [42] **A.** An example (AChR pore α subunit) shows both the TM region and the number of the TM segments are correctly predicted. **A.** An example (AChR pore γ subunit) shows the predicted number of the TM segments are wrong, though the TM regions are correctly predicted.

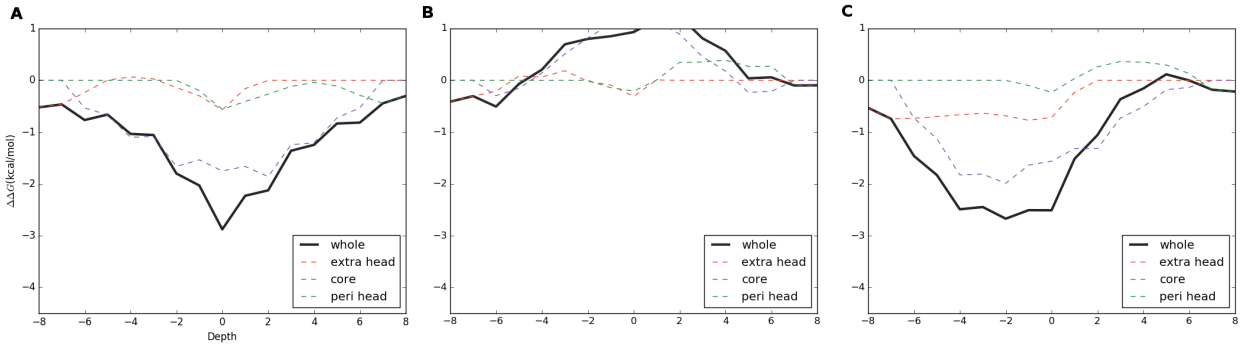

Figure S5: An example of the insertion TFEs (Omp32, PDB ID:1e54). **A.** The insertion TFEs of the WT Omp32 shows a funnel pattern. The insertion TFEs of the residue-shuffled Omp32 regardless of side-chain directions (**B**) and of the the LFR-shuffled Omp32 (**C**).