Water Controls Membrane Protein Stability

Dagan C. Marx and Karen G. Fleming*

Thomas C. Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, MD

*Corresponding Author: Karen.Fleming@jhu.edu

Keywords: membrane protein folding, transfer free energies, nonpolar solvation parameter, protein stability

Abbreviations:
OmpLA – E. *coli* outer membrane phospholipase A1
PagP – E. *coli* lipid A palymitoyltransferase PagP
DLPC – 1,2 didodecanoyl-sn-glycero-3-phosphocholine
LUV – large unilamellar vesicles
$\Delta G^{\circ}{}_{w,l}$– OmpLA variant free energy of folding
$\Delta\Delta G^{\circ}{}_{w,l}$– alanine-referenced, water-to-bilayer side chain transfer free energy
$\Delta G^{\circ}{}_{sc}$ – reference-free, water-to-bilayer side chain transfer free energy
σ_{NP} – nonpolar solvation parameter
TMD – transmembrane domain
$\Delta G_{app\ (t,b)}$ –translocon-to-bilayer side chain transfer free energy
$\Delta G^{\circ}{}_{i,b}$ – interface-to-bilayer side chain transfer free energy
$\Delta G_{t,i}$ – translocon-to-interface side chain transfer free energy

Abstract

Through the insertion of nonpolar side chains into the bilayer, the hydrophobic effect has long been accepted as a driving force for membrane protein folding. However, how the changing chemical composition of the bilayer affects the magnitude side chain transfer free energies ($\Delta G°_{sc}$) has historically not been well understood. A particularly challenging region for experimental interrogation is the bilayer interfacial region that is characterized by a steep polarity gradient. In this study we have determined the $\Delta G°_{sc}$ for nonpolar side chains as a function of bilayer position using a combination of experiment and simulation. We discovered an empirical correlation between the surface area of nonpolar side chain, $\Delta G°_{sc}$, and local water concentration in the membrane that allows for $\Delta G°_{sc}$to be accurately estimated at any location in the bilayer. Using these water-to-bilayer $\Delta G°_{sc}$we have calculated the interface-to-bilayer transfer free energy ($\Delta G°_{i,b}$). We find that the $\Delta G°_{i,b}$ are similar to the "biological", translocon-based transfer free energies, indicating that the translocon energetically mimics the bilayer interface. Together these findings should increase the accuracy of computational workflows used to identify and design membrane proteins, as well as increase our understanding of how disease-causing mutations affect membrane protein folding and function.

Introduction

The cellular functions of proteins are regulated by the intrinsic thermodynamic stability of the three-dimensional structure adopted by their polypeptide chain.[1–4] The stabilities of many water-soluble proteins have been experimentally measured *in vitro*, rendering a robust understanding of the forces that regulate structure and function of this class of proteins.[5] For membrane proteins, which make up approximately one third of the eukaryotic proteome,[6] our understanding of the forces that determine their thermodynamic stability is much less mature.[7] This lag in knowledge derives from the fact that membrane proteins are aggregation prone in aqueous environments and must embed themselves into a hydrophobic environment such as detergent micelles or phospholipid bilayers to fold and function,[8–10] which makes traditional experimental approaches for measuring protein stabilities untenable for membrane proteins.[7]

The stabilities of membrane proteins are directly linked to the energetics of burying amino acid side chains into the phospholipid bilayer, also referred to as the side chain transfer free energy ($\Delta G°_{sc}$).[7,10–13] Both the magnitude and sign of the water-to-bilayer $\Delta G°_{sc}$ for a side chain in a transmembrane domain (TMD) depend on the exact chemical compositions of the two end-point solvents (water and bilayer), as well as the chemical composition of the side chain.[14–23] Yet, the bilayer is not a simple solvent. Biological membranes contain two chemically distinct regions: the largely desolvated hydrocarbon core and the heterogeneous interface.[24,25] Each interface is a ~15 Å wide region between bulk water and the desolvated core of the bilayer that is characterized by a rapidly changing water concentration gradient along the bilayer normal.[16,25] Therefore, the heterogenous chemical composition along the bilayer normal should result in a position-dependence on $\Delta G°_{sc}$ for side chains in the interface. Arguably, this interfacial membrane region represents half of the bilayer volume, and a full understanding of membrane protein folding forces depends on the energetics imposed by this steeply changing polarity of this region.

For nonpolar side chains, which are highly enriched in transmembrane domains, the magnitude of $\Delta G°_{sc}$ has been traditionally related to the (solvent) accessible surface area (ASA) of the side chain. [17,19,20,26–28] The relationship between ASA and $\Delta G°_{sc}$ has been defined as the nonpolar solvation parameter, σ_{NP}, which quantifies the energy (cal mol^{-1}) gained per Å^2 of

nonpolar surface area buried in the membrane. While the σ_{NP} describing the water-to-bilayer center transfer is well established (~23-25 cal mol^{-1} Å^{-2}),[17,19,20] the position dependence of σ_{NP} along the interface is not as well understood. Indeed, only a single study has investigated the σ_{NP} corresponding to the water-to-interface transfer (12 cal mol^{-1} Å^{-2}), but this work did not clarify where in the interface this σ_{NP} corresponds.[18]

Here, we have resolved the bilayer z-position dependence of σ_{NP} using a combination of protein folding titrations and molecular dynamics simulations. Importantly, we discovered a robust linear correlation between σ_{NP} and the local water concentration across the bilayer. This quantitative relationship enables prediction of the functional form for the solvation parameter across the interface, $\sigma_{NP}(z)$, and accordingly the nonpolar side chain $\Delta G^{\circ}{}_{sc}$, at any position in any phospholipid bilayer. These values should improve the accuracy of membrane protein structure prediction and design algorithms. Additionally, we use our findings to better understand the thermodynamic basis for how the Sec translocon aids α-helical membrane protein folding *in vivo.*

Results

Experimental approach to measure side chain transfer free energies. The function relating σ_{NP} to position along the bilayer normal ($\sigma_{NP}(z)$, where z denotes the position across the bilayer normal) is fundamentally derived from experimental protein folding titrations described below. Using a previously validated host-guest approach, we measured the thermodynamic stabilities of a series of host (alanine) and guest (nonpolar side chains isoleucine (Ile), methionine (Met), and valine (Val)) variants of the well characterized membrane protein scaffold, E. *coli* outer membrane phospholipase A1 (OmpLA).[16,17,29,30] Host/guest sites were chosen at varying positions across the bilayer to allow side chains to experience varying local chemical compositions created by the amphipathic phospholipids in the membrane. In total, five lipid-facing sites located on well-defined transmembrane β-strands were used in this study (residues 120, 210, 212, 214, and 223) and are shown in Figure 1A. Figure 1B shows the water concentration gradient across our experimental DLPC bilayer and highlights the 60-fold change in water concentration accessed by these sites.

The thermodynamic stabilities ($\Delta G^{\circ}{}_{w,l}$) of all variants were measured using intrinsic tryptophan fluorescence-based chemical denaturation titrations (Figure S1).[16,17,19,30] Figure S2

shows that each variant was enzymatically active and Figure S3 shows that the folded and unfolded tryptophan fluorescence spectra for each variant overlay. Together these data indicate that single mutations do not affect the tertiary or quaternary structure of either the folded or unfolded states of OmpLA. We have previously shown that single site variants of OmpLA fold in a path-independent manner into large unilamellar DLPC vesicles, allowing for equilibrium thermodynamic parameters to be extracted by fitting the titration data to a three-state linear extrapolation model ($\Delta G^{\circ}{}_{w,l}$ values listed in Table S1).[16,17,29]

The thermodynamic contributions of the guest side chain to folding were isolated by taking the difference in the $\Delta G^{\circ}{}_{w,l}$ of the host and guest variants at a given site on OmpLA as shown in the thermodynamic cycle in Figure S4. We interpret the resulting $\Delta\Delta G^{\circ}{}_{w,l}$ to represent the water-to-bilayer transfer free energy change for each guest side chain relative to alanine at a particular site (Table S1). We find that our experimentally measured $\Delta\Delta G^{\circ}{}_{w,l}$ are remarkably similar to analogous, computationally estimated $\Delta\Delta G^{\circ}{}_{w,l}$, though both Ile and Val deviated significantly at site 212 (Figure S5).[31]

The σ_{NP} is not constant in the bilayer interface. To determine the σ_{NP} for each site on OmpLA we used the $\Delta\Delta G^{\circ}{}_{w,l}$ values for each nonpolar side chain, including previously published $\Delta\Delta G^{\circ}{}_{w,l}$ for leucine and phenylalanine.[16,17] Figure 2 shows nonpolar $\Delta\Delta G^{\circ}{}_{w,l}$ plotted as a function of the nonpolar surface area of each side chain relative to alanine (ΔASA) for each site. The σ_{NP} at each site on OmpLA is the slope of the linear fit of the correlation between nonpolar $\Delta\Delta G^{\circ}{}_{w,l}$ and ΔASA. Although all but one transfer of nonpolar side chains from water to bilayer positions are energetically favorable , the σ_{NP} values vary nearly two-fold from site to site with a range that spans from -27.3 to -16.1 cal/mol/Å^2.

Our experimental setup requires a reference host side chain whose energetic contributions must be removed to obtain a reference-free scale. As alanine is the amino acid side chain in the host protein, we obtain a reference-free scale by removing the intrinsic contributions of Ala to $\Delta\Delta G^{\circ}{}_{w,l}$ at each site using the site-specific σ_{NP} values. The reference-free $\Delta G^{\circ}{}_{sc}$ for alanine at each host site was determined by multiplying the corresponding σ_{NP} by the ASA of the alanine side chain (69.1 Å^2).[17] The reference-free $\Delta G^{\circ}{}_{sc}$ for the other nonpolar side chains were

calculated by adding the $\Delta G^{\circ}{}_{sc}$ for alanine to the $\Delta\Delta G^{\circ}{}_{w,l}$ at each site. The $\Delta G^{\circ}{}_{sc}$ for each nonpolar side chain, including alanine, are listed in Table S2.

The σ_{NP} is linearly correlated with the amount of water in the bilayer. We next addressed the question of how the local chemical composition of the bilayer modulates the σ_{NP} by deriving the relationship between the σ_{NP} and position along the bilayer normal, or $\sigma_{NP}(z)$. To accomplish this, we calculated the z-position of the side chain Cα of each OmpLA variant from all-atom molecular dynamics simulations conducted in the experimental lipid bilayer. Figure S6 shows that each simulation was equilibrated after 50ns and analyses were performed on the final 100 ns of each run. Instead of using the bilayer center as the reference point, z=0, we calculated the position of each Cα relative to the average position of the phosphate plane of the bilayer. We chose the phosphate plane as our reference point for z because it can be more easily applied to bilayers with different lipid compositions. Figure S7 shows that variant Cα positions are generally well described by Gaussian distributions and that our sites capture the majority of the region between the bilayer center and phosphate plane. Table S3 lists the mean z-position and standard deviations for each variant Cα, which were averaged to calculate the z-position for the σ_{NP} at each site. Using the previously derived function that describes the bilayer z-position dependence of water concentration in our experimental DLPC bilayer (shown in Figure 1B),[16] we were also able to determine the corresponding local water concentration for each σ_{NP} (Table S4).

Figure 3A shows that σ_{NP} is linearly correlated with the water concentration in the bilayer (R^2 = 0.84). The 10,000-fold change in water concentration illustrates the large range of solvent conditions accessed in our experiments and necessitates presentation using a logarithmic x-axis. This novel functional relationship we derived from our measurements using a native protein and a phospholipid bilayer confirms our hypothesis that the water in a bilayer controls membrane protein stabilities by modulating the σ_{NP} directly.

It is also instructive to consider the σ_{NP} using bilayer z-coordinates. Figure 3B shows the bilayer z-position dependence of σ_{NP} in which the line of best fit in Figure 3A is transferred to units of distance from the phosphate plane of the membrane. Using the resulting function, $\sigma_{NP}(z)$, the σ_{NP} can easily and accurately be determined at any position in the bilayer. The $\sigma_{NP}(z)$ function plateaus to a value of -25.8 cal/mol/Å^2 in the anhydrous core of the bilayer, which agrees with

the σ_{NP} determined for this region of the bilayer in previous studies.[14,17,19,20] We find that $\sigma_{NP}(z)$ decreases in the membrane interface to a predicted σ_{NP} of -1.4 cal/mol/Å^2 at the phosphate plane, which is much less favorable than previous estimates of σ_{NP} for the interface (-12 cal/mol/Å^2).[18] Using $\sigma_{NP}(z)$, we estimate that this global σ_{NP} for the water-to-interface transition was reporting for a z-position approximately 3.6 Å inside the phosphate plane.

Bilayer z-position dependence of reference-free $\Delta G°_{sc}$. The large change in the σ_{NP} across the bilayer interface significantly impacts the contribution of nonpolar side chains to the overall stability of membrane proteins in a position-dependent manner. This can be visualized by using $\sigma_{NP}(z)$ to calculate the reference-free, nonpolar $\Delta G°_{sc}$ as a function of z-position in the bilayer. The resulting $\Delta G°_{sc}(z)$ functions for each nonpolar side chain were created by multiplying $\sigma_{NP}(z)$ by the ASA of each nonpolar side chain. Figure 4 shows the $\Delta G°_{sc}(z)$ for each nonpolar side chain (black lines) and the $\Delta G°_{sc}$ for each site on OmpLA determined above (points). The agreement between the depth dependent profile of the experimentally determined $\Delta G°_{sc}$ and the $\Delta G°_{sc}(z)$ indicates that nonpolar surface area burial in the membrane dominates $\Delta G°_{sc}$ across the bilayer normal.

The functional free energy dependence for the alanine side chain is particularly useful. By determining $\Delta G°_{Ala}(z)$, we were also able to adjust previously measured $\Delta\Delta G°_{w,l}$ for arginine, tryptophan, and tyrosine to the reference-free $\Delta G°_{sc}(z)$.[16,17] As detailed above, a combination of new (Arg) and published (Tyr and Trp) all-atom molecular dynamics were used to determine the z-position of the Cα for each variant.[16] This position was converted into an estimate for $\Delta G°_{Ala}$ using $\Delta G°_{Ala}(z)$. The $\Delta G°_{sc}(z)$ for Arg, Tyr, and Trp were calculated by adding the position-specific $\Delta G°_{Ala}$ to the corresponding $\Delta\Delta G°_{w,l}$ (Table S3).[16,17] Figures S8 and S9 show the bilayer z-position dependence of the $\Delta G°_{sc}$ for arginine and the aromatics Trp and Tyr, respectively. As expected, there is a larger energetic penalty for arginine buried within the bilayer at a depth of ~13 Å from the phosphate plane (+2.5 kcal/mol) as compared to a small, favorable energetic contribution when the side chain is located in the interface (-0.5 kcal/mol). Similarly, we find $\Delta G°_{sc}$ for both tyrosine and tryptophan to be more favorable at the bilayer interface than the central region of the bilayer.

Relating water-to-bilayer $\Delta G°_{sc}$ to the energetics of translocon-aided folding

With the reference-free, water-to-bilayer $\Delta G°_{sc}$ for nine amino acids determined as a function of z-position in the bilayer, we next asked whether our findings could be used to better understand membrane protein folding by the Sec translocon machinery. The translocon acts as a protein channel in the bilayer that allows α-helical TMDs to partition into the bilayer while protecting hydrophilic loops and turns from exposure to the hydrophobic membrane.[10,32] All-atom molecular dynamics simulations have identified that the chemical properties of water in the pore of the translocon mimic the properties of water in the interfacial region of the membrane.[33]

Thus, a hypothesis for translocon function is that it mimics the thermodynamics of an interface-to-bilayer transition, not a water-to-bilayer transition.[10] Indeed, the water-to-bilayer nonpolar $\Delta G°_{sc}$ reported here are much greater in magnitude than the $\Delta G°_{app\ (t,b)}$ for the translocon-to-bilayer transition.[15,34] Because we have measured the water-to-bilayer $\Delta G°_{sc}$ across the bilayer normal, we can calculate the interface-to-center transition by taking the difference between $\Delta G°_{sc}$ at the center of the bilayer (A210) and the $\Delta G°_{sc}$ at the most interfacial site (Y214) (Figure 5a). Figure 5b shows that our interface-to-bilayer $\Delta\Delta G°_{sc}$, or $\Delta G°_{i,b}$, agree remarkably well with the $\Delta G°_{app\ (t,b)}$.[34] Additionally, the σ_{NP} for the interface-to-center transition ($\sigma_{NP,int\text{-}cen} = \sigma_{NP,210} - \sigma_{NP,214}$) is equivalent to the σ_{NP} calculated for translocon-mediated folding (7.7 and 6-10 cal/mol/Å^2, respectively).[27] The remarkable energetic equivalence between the interface-to-center and translocon-mediated transfer free energies provides thermodynamic backing to the hypothesis that the translocon aids TMD folding by mimicking properties of the interface.

Discussion

The thermodynamic stability of a membrane protein regulates its structure, conformational landscape, and ability to perform its biological function.[7,11] A key determinant of this stability is the $\Delta G°_{sc}$ of each side chain from water into the membrane. Historically, experiments measuring side chain transfer from water to organic solvents such as cyclohexane or octanol were used to approximate $\Delta G°_{sc}$.[20,23] While these experiments accurately describe the water-to-center transition, they failed to capture the dramatic chemical heterogeneity of the

interfacial region of the bilayer, which contains a steep polarity gradient as a function of z-position.

In this study we have measured the $\Delta\Delta G^{\circ}{}_{w,l}$ for Ile, Met, and Val as a function of z-position in the bilayer. Using these $\Delta\Delta G^{\circ}{}_{w,l}$, along with previously measured $\Delta\Delta G^{\circ}{}_{w,l}$ for Leu and Phe we were able to determine a novel functional form for the σ_{NP} across the bilayer normal.[16,17] We discovered a direct correlation between the σ_{NP}, and thus the $\Delta G^{\circ}{}_{sc}$ for nonpolar side chains, with the local water concentration in the bilayer. The $\sigma_{NP}(z)$ is a convenient numerical expression for assessing the magnitude of the hydrophobic effect, as it relates nonpolar surface area to the energy gained from transferring the surface area from bulk water to a dehydrated environment.

Importantly, the dependence of nonpolar $\Delta G^{\circ}{}_{sc}$ on local water concentration differs starkly from the water-dependence of backbone hydrogen bond (bbHB) formation in membrane proteins. Although it was long expected that bbHB energies would also be strongly influenced by the steep water gradient intrinsic to bilayer interface, we recently observed bbHB energies to be constant throughout this region.[35]

The experimental data here offer an opportunity to consider how the translocon insertion process relates to the underlying physical chemistry of the reaction process. Our data provides evidence that translocon-mediated membrane protein folding energetically mimics the interface-to-bilayer transition for $\Delta G^{\circ}{}_{sc}$. We find that $\Delta G^{\circ}{}_{i,b}$ is approximately equal to $\Delta G^{\circ}{}_{app\ (t,b)}$ for eight of the nine side chains investigated, with Trp being the only outlier; we speculate that this one outlier is most likely due to differential aromatic interactions in the OmpLA and translocon experimental setups.[16,34] This finding provides thermodynamic backing to a recent model for translocon function in which transmembrane α-helices sample both interfacial and translocon associated conformations before they insert into the membrane.[10] Our findings indicate that these two states would be essentially energetically equivalent, indicating the energy barrier between these two conformations may be non-existent with switching between the two states determined by the kinetics for a given TMD sequence.

Our results should improve the speed and accuracy of membrane protein structure prediction and design algorithms.[36–38] As membrane proteins account for over half of all therapeutic drug targets,[39] having accurate force fields to describe membrane protein energetics

is essential. This simple, linear function ($\sigma_{NP}(z)$) function could be adapted to implicit membrane models or could alternatively be applied explicitly in systems such that a local value of σ_{NP} to be calculated over water concentrations ranging from millimolar to 30 molar. Subsequently, better estimates of the energetic contribution of any lipid-facing, nonpolar side chain to the stability of membrane protein. By relating the σ_{NP} to the chemical properties of the bilayer, the $\sigma_{NP}(z)$ can be customized to any bilayer lipid composition by calculating the water concentration gradient of that particular bilayer.

Additionally, we would like to delineate how the various transfer free energies reported here should be applied to various biological functions. The critical parameter for consideration is are the endpoints for the reaction at hand. The $\Delta G^{\circ}{}_{app\ (t,b)}$ or $\Delta G^{\circ}{}_{i,b}$ can be used to approximate the contribution of a side chain to the folding of alpha-helical TMDs.[15,34] On the other hand, water-to-bilayer $\Delta G^{\circ}{}_{sc}$ are more applicable to modeling both changes in stability of different conformers of a membrane protein and potentially unfolding processes, such as extraction of transmembrane regions from the bilayer by AAA-ATPases.[40] Experimentally determining these thermodynamic parameters for membrane proteins is extremely challenging and accurate computational modeling will allow for greater understand of essential processes such as the gating of a channel or the cycling of a transporter.[41–43] At the very least, we anticipate that $\Delta G^{\circ}{}_{sc}$ can be used to better understand the effects of mutations that may impair membrane protein folding such as those occurring in cystic fibrosis and Charcot-Marie-Tooth disease.[44]

Conclusion

In summary, we have determined the relationship between nonpolar side chain transfer free energies, solvent accessible surface area, and the local water concentration in the bilayer. Using this relationship, we calculate $\Delta G^{\circ}{}_{sc}(z)$ functions for nonpolar side chains and remove the alanine dependence of previously measured $\Delta\Delta G^{\circ}{}_{w,l}$ for Arg, Trp, and Tyr. Our work shows that the translocon energetically mimics the bilayer interface for side chain transfer into the membrane. Together, these findings increase our understanding of the driving forces of protein stability in membranes and should increase the accuracy of computational workflows to identify, design, and understand the energy landscapes of membrane proteins.

Methods

OmpLA Variant Expression, Purification, Activity Assay and Folding Titrations. The OmpLA variants were engineered, expressed, and purified from inclusion bodies as previously described.[16,17] Each variant was verified to have enzymatic activity while folded into DLPC large unilamellar vesicles (LUV) using a previously developed colorimetric assay.[16,17] Folding titrations were also set up and performed as previously described in great detail.[16,17,19,29,30] Briefly, unfolded OmpLA variants added to DLPC LUVs (diameter = 100nm) in 5M guanidine hydrochloride. After overnight incubation at 37°C, samples were diluted to varying guanidine concentrations ranging from 1-5M at 0.08M increments (51 total samples). Following at least a 40-hour incubation in a rotating incubator at 37°C, intrinsic tryptophan fluorescence was measured at 330 nm on a PC1 Fluorometer with cross-polarization (ISS, Champaign, IL, USA).

Fitting OmpLA variant $\Delta G°_{w,l}$ and σ_{NP}. The resulting titration data were fit to a three-state linear extrapolation model to extract the stability ($\Delta G°_{w,l}$) of each variant using previously determined m-values using the Igor Pro 8 software package (Wavemetrics, Portland, OR, USA).[16,17] Host variant $\Delta G°_{w,l}$ were previously determined and used to calculate $\Delta\Delta G°_{w,l}$ for each nonpolar guest variant.[16] The nonpolar solvation parameter for each site on OmpLA was determined by weighted linear regression using Igor Pro, with weights corresponding to the error in $\Delta\Delta G°_{w,l}$. The surface area for each nonpolar side chain was previously determined in the context of a Gly-X-Gly peptide.[17]

Molecular Dynamics Simulations of OmpLA variants. All-atom molecular dynamics simulations of each OmpLA variant were performed to determine the position of each guest side chain in the experimental DLPC bilayer. The monomeric crystal structure of OmpLA (1qd5) was used as the starting structure, and the rest of the components of the system and scripts were made using CHARMM-GUI.[45–47] A system was built for each OmpLA variant embedded in a DLPC bilayer and 0.2M NaCl at 37°C, resulting in 33 individual simulations. To approximate the experimental pH, some ionizable residues were protonated as described previously.[16] Systems were initially equilibrated using the protocol provided by CHARMM-GUI, followed by an additional 50ns of equilibration to allow for the system to fully relax (Figure S4). All simulations

used NAMD and were run on the Maryland Advanced Research Computing Center super computer.[48]

VMD plug-ins and homemade scripts (available upon request) were used to analyze all trajectories.[49] To determine the position of the Cα for each variant relative to the phosphate plane, we first calculated the average z-position for the phosphate atoms in each leaflet of the bilayer. We then calculated the distance of the Cα for the variant side chain to the closest phosphate plane. These calculations were performed for each frame in the last 100ns of the trajectory and distances from the phosphate plane were binned into 0.1Å bins to create the histograms in Figure S5. Histograms were fit to Gaussian distributions using Igor Pro software to determine average positions and standard deviations found in Table S3.

Acknowledgements
The authors would like to thank Dr. Sarah McDonald and Margo Goodall for assistance with cloning the variants used in this study, Mark Kreutzberger for growth and purification of some OmpLA variants used in this study, and Dr. Sarah McDonald and Dr. Patrick Fleming for experimental and computational guidance.

This work was funded by National Institutes of Health grants R01 GM079440 and T32 GM008403.

References

(1) Bloom, J. D.; Labthavikul, S. T.; Otey, C. R.; Arnold, F. H. Protein Stability Promotes Evolvability. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (15), 5869–5874. https://doi.org/10.1073/pnas.0510098103.

(2) Osadchy, M.; Kolodny, R. Maps of Protein Structure Space Reveal a Fundamental Relationship between Protein Structure and Function. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (30), 12301–12306. https://doi.org/10.1073/pnas.1102727108.

(3) Reis, R.; Moraes, I. Structural Biology and Structure–Function Relationships of Membrane Proteins. *Biochemical Society Transactions*. Portland Press Ltd 2018, pp 47–61. https://doi.org/10.1042/BST20180269.

(4) Freire, E. The Thermodynamic Linkage between Protein Structure, Stability, and Function. *Methods in molecular biology (Clifton, N.J.)*. Methods Mol Biol 2001, pp 37–68. https://doi.org/10.1385/1-59259-193-0:037.

(5) Goldenzweig, A.; Fleishman, S. J. Principles of Protein Stability and Their Application in Computational Design. *Annual Review of Biochemistry*. Annual Reviews Inc. June 20, 2018, pp 105–129. https://doi.org/10.1146/annurev-biochem-062917-012102.

(6) Almén, M. S.; Nordström, K. J. V.; Fredriksson, R.; Schiöth, H. B. Mapping the Human Membrane Proteome: A Majority of the Human Membrane Proteins Can Be Classified According to Function and Evolutionary Origin. *BMC Biol.* **2009**, *7*, 50. https://doi.org/10.1186/1741-7007-7-50.

(7) Hong, H. Toward Understanding Driving Forces in Membrane Protein Folding. *Archives of Biochemistry and Biophysics*. Academic Press Inc. December 10, 2014, pp 297–313. https://doi.org/10.1016/j.abb.2014.07.031.

(8) Tan, A. E.; Burgess, N. K.; DeAndrad, D. S.; Marold, J. D.; Fleming, K. G. Self-Association of Unfolded Outer Membrane Proteins. *Macromol. Biosci.* **2010**, *10* (7), 763–767. https://doi.org/10.1002/mabi.200900479.

(9) Parton, D. L.; Klingelhoefer, J. W.; Sansom, M. S. P. Aggregation of Model Membrane Proteins, Modulated by Hydrophobic Mismatch, Membrane Curvature, and Protein Class. *Biophys. J.* **2011**, *101* (3), 691–699. https://doi.org/10.1016/j.bpj.2011.06.048.

(10) Cymer, F.; Von Heijne, G.; White, S. H. Mechanisms of Integral Membrane Protein Insertion and Folding. *Journal of Molecular Biology*. Academic Press March 13, 2015, pp 999–1022. https://doi.org/10.1016/j.jmb.2014.09.014.

(11) White, S. H.; Wimley, W. C. Membrane Protein Folding and Stability: Physical Principles. *Ann. Rev. Biophys. Biomol. Str.* **1999**, *28*, 319–365. https://doi.org/10.1146/annurev.biophys.28.1.319.

(12) Popot, J.-L.; Engelman, D. M. *Helical Membrane Protein Folding, Stability, and Evolution*; 2000.

(13) Popot, J. L.; Engelman, D. M. Membrane Protein Folding and Oligomerization: The Two-Stage Model. *Biochemistry* **1990**, *29* (17), 4031–4037. https://doi.org/10.1021/bi00469a001.

(14) Elazar, A.; Weinstein, J.; Biran, I.; Fridman, Y.; Bibi, E.; Fleishman, S. J. Mutational Scanning Reveals the Determinants of Protein Insertion and Association Energetics in the Plasma Membrane. *Elife* **2016**, *5*. https://doi.org/10.7554/eLife.12125.

(15) Hessa, T.; Meindl-Beinker, N. M.; Bernsel, A.; Kim, H.; Sato, Y.; Lerch-Bader, M.; Nilsson, I.; White, S. H.; Von Heijne, G. Molecular Code for Transmembrane-Helix Recognition by the Sec61 Translocon. *Nature* **2007**, *450* (7172), 1026–1030. https://doi.org/10.1038/nature06387.

(16) McDonald, S. K.; Fleming, K. G. Aromatic Side Chain Water-to-Lipid Transfer Free Energies Show a Depth Dependence across the Membrane Normal. *J. Am. Chem. Soc.* **2016**, *138* (25), 7946–7950. https://doi.org/10.1021/jacs.6b03460.

(17) Moon, C. P.; Fleming, K. G. Side-Chain Hydrophobicity Scale Derived from Transmembrane Protein Folding into Lipid Bilayers. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (25), 10174–10177. https://doi.org/10.1073/pnas.1103979108.

(18) Wimley, W. C.; White, S. H. Experimentally Determined Hydrophobicity Scale for Proteins at Membrane Interfaces. *Nat. Struct. Mol. Biol.* **1996**, *3* (10), 842–848.

(19) Marx, D. C.; Fleming, K. G. Influence of Protein Scaffold on Side-Chain Transfer Free Energies. *Biophys. J.* **2017**, *113* (3), 597–604. https://doi.org/10.1016/j.bpj.2017.06.032.

(20) Wimley, W. C.; Creamer, T. P.; White, S. H. Solvation Energies of Amino Acid Side Chains

and Backbone in a Family of Host-Guest Pentapeptides. *Biochemistry* **1996**, *35* (16), 5109–5124. https://doi.org/10.1021/bi9600153.

(21) Ulmschneider, J. P.; Smith, J. C.; White, S. H.; Ulmschneider, M. B. The Importance of the Membrane Interface as the Reference State for Membrane Protein Stability. *Biochim. Biophys. Acta - Biomembr.* **2018**, *1860* (12), 2539–2548. https://doi.org/10.1016/j.bbamem.2018.09.012.

(22) Damodaran, S.; Song, K. B. The Role of Solvent Polarity in the Free Energy of Transfer of Amino Acid Side Chains from Water to Organic Solvents. *J. Biol. Chem.* **1986**, *261* (16), 7220–7222.

(23) Radzicka, A.; Wolfenden, R. Comparing the Polarities of the Amino Acids: Side-Chain Distribution Coefficients between the Vapor Phase, Cyclohexane, 1-Octanol, and Neutral Aqueous Solution. *Biochemistry* **1988**, *27* (5), 1664–1670. https://doi.org/10.1021/bi00405a042.

(24) Wiener, M. C.; White, S. H. Structure of a Fluid Dioleoylphosphatidylcholine Bilayer Determined by Joint Refinement of X-Ray and Neutron Diffraction Data. II. Distribution and Packing of Terminal Methyl Groups. *Biophys. J.* **1992**, *61* (2), 428–433. https://doi.org/10.1016/S0006-3495(92)81848-9.

(25) Wiener, M. C.; White, S. H. Structure of a Fluid Dioleoylphosphatidylcholine Bilayer Determined by Joint Refinement of X-Ray and Neutron Diffraction Data. III. Complete Structure. *Biophys. J.* **1992**, *61* (2), 434–447. https://doi.org/10.1016/S0006-3495(92)81849-0.

(26) Wesson, L.; Eisenberg, D. Atomic Solvation Parameters Applied to Molecular Dynamics of Proteins in Solution. *Protein Sci.* **1992**, *1* (2), 227–235. https://doi.org/10.1002/pro.5560010204.

(27) Öjemalm, K.; Higuchi, T.; Jiang, Y.; Langel, Ü.; Nilsson, I.; White, S. H.; Suga, H.; Von Heijne, G. Apolar Surface Area Determines the Efficiency of Translocon-Mediated Membrane-Protein Integration into the Endoplasmic Reticulum. *Proc Natl Acad Sci U S A* **2011**, *108* (31), E359-364. https://doi.org/10.1073/pnas.1100120108/-/DCSupplemental.

(28) Chothia, C. Hydrophobic Bonding and Accessible Surface Area in Proteins. *Nature* **1974**,

248 (5446), 338–339. https://doi.org/10.1038/248338a0.

(29) Moon, C. P.; Kwon, S.; Fleming, K. G. Overcoming Hysteresis to Attain Reversible Equilibrium Folding for Outer Membrane Phospholipase A in Phospholipid Bilayers. *J. Mol. Biol.* **2011**, *413* (2), 484–494. https://doi.org/10.1016/j.jmb.2011.08.041.

(30) Moon, C. P.; Fleming, K. G. *Using Tryptophan Fluorescence to Measure the Stability of Membrane Proteins Folded in Liposomes*, 1st ed.; Elsevier Inc., 2011; Vol. 492. https://doi.org/10.1016/B978-0-12-381268-1.00018-5.

(31) Lin, M.; Gessmann, D.; Naveed, H.; Liang, J. Outer Membrane Protein Folding and Topology from a Computational Transfer Free Energy Scale. *J. Am. Chem. Soc.* **2016**, *138* (8), 2592–2601. https://doi.org/10.1021/jacs.5b10307.

(32) Niesen, M. J. M.; Müller-Lucks, A.; Hedman, R.; von Heijne, G.; Miller, T. F. Forces on Nascent Polypeptides during Membrane Insertion and Translocation via the Sec Translocon. *Biophys. J.* **2018**, *115* (10), 1885–1894. https://doi.org/10.1016/j.bpj.2018.10.002.

(33) Capponi, S.; Heyden, M.; Bondar, A.-N.; Tobias, D. J.; White, S. H. Anomalous Behavior of Water inside the SecY Translocon. *Proc. Natl. Acad. Sci.* **2015**, 201424483. https://doi.org/10.1073/pnas.1424483112.

(34) 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* (7024), 377–381. https://doi.org/10.1038/nature03216.

(35) Lessen, H. J.; Majumdar, A.; Fleming, K. G. Backbone Hydrogen Bond Energies in Membrane Proteins Are Insensitive to Large Changes in Local Water Concentration. *J. Am. Chem. Soc.* **2020**, *142* (13), 6227–6235. https://doi.org/10.1021/jacs.0c00290.

(36) Alford, R. F.; Fleming, P. J.; Fleming, K. G.; Gray, J. J. Protein Structure Prediction and Design in a Biologically Realistic Implicit Membrane. *Biophys. J.* **2020**, *118* (8), 2042–2055. https://doi.org/10.1016/j.bpj.2020.03.006.

(37) Alford, R. F.; Koehler Leman, J.; Weitzner, B. D.; Duran, A. M.; Tilley, D. C.; Elazar, A.; Gray, J. J. An Integrated Framework Advancing Membrane Protein Modeling and Design. *PLoS Comput. Biol.* **2015**, *11* (9). https://doi.org/10.1371/journal.pcbi.1004398.

(38) Weinstein, J. Y.; Elazar, A.; Fleishman, S. J. A Lipophilicity-Based Energy Function for Membrane-Protein Modelling and Design. *PLoS Comput. Biol.* **2019**, *15* (8). https://doi.org/10.1371/journal.pcbi.1007318.

(39) Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. How Many Drug Targets Are There? *Nat. Rev. Drug Discov.* **2006**, *5* (12), 993–996. https://doi.org/10.1038/nrd2199.

(40) Avci, D.; Lemberg, M. K. Clipping or Extracting: Two Ways to Membrane Protein Degradation. *Trends in Cell Biology*. Elsevier Ltd October 1, 2015, pp 611–622. https://doi.org/10.1016/j.tcb.2015.07.003.

(41) Sukharev, S. I.; Sigurdson, W. J.; Kung, C.; Sachs, F. Energetic and Spatial Parameters for Gating of the Bacterial Large Conductance Mechanosensitive Channel, MscL. *J. Gen. Physiol.* **1999**, *113* (4), 525–539. https://doi.org/10.1085/jgp.113.4.525.

(42) Haswell, E. S.; Phillips, R.; Rees, D. C. Mechanosensitive Channels: What Can They Do and How Do They Do It? *Structure*. NIH Public Access October 12, 2011, pp 1356–1369. https://doi.org/10.1016/j.str.2011.09.005.

(43) Hofmann, S.; Januliene, D.; Mehdipour, A. R.; Thomas, C.; Stefan, E.; Brüchert, S.; Kuhn, B. T.; Geertsma, E. R.; Hummer, G.; Tampé, R.; Moeller, A. Conformation Space of a Heterodimeric ABC Exporter under Turnover Conditions. *Nature* **2019**, *571* (7766), 580–583. https://doi.org/10.1038/s41586-019-1391-0.

(44) Marinko, J. T.; Huang, H.; Penn, W. D.; Capra, J. A.; Schlebach, J. P.; Sanders, C. R. Folding and Misfolding of Human Membrane Proteins in Health and Disease: From Single Molecules to Cellular Proteostasis. *Chemical Reviews*. American Chemical Society May 8, 2019, pp 5537–5606. https://doi.org/10.1021/acs.chemrev.8b00532.

(45) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29* (11), 1859–1865. https://doi.org/10.1002/jcc.20945.

(46) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L.; MacKerell, A. D.; Klauda, J. B.; Im, W. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force

Field. *J. Chem. Theory Comput.* **2016**, *12* (1), 405–413. https://doi.org/10.1021/acs.jctc.5b00935.

(47) Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; Klauda, J. B.; Im, W. CHARMM-GUI *Membrane Builder* toward Realistic Biological Membrane Simulations. *J. Comput. Chem.* **2014**, *35* (27), 1997–2004. https://doi.org/10.1002/jcc.23702.

(48) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781–1802. https://doi.org/Doi 10.1002/Jcc.20289.

(49) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33-.

Figures

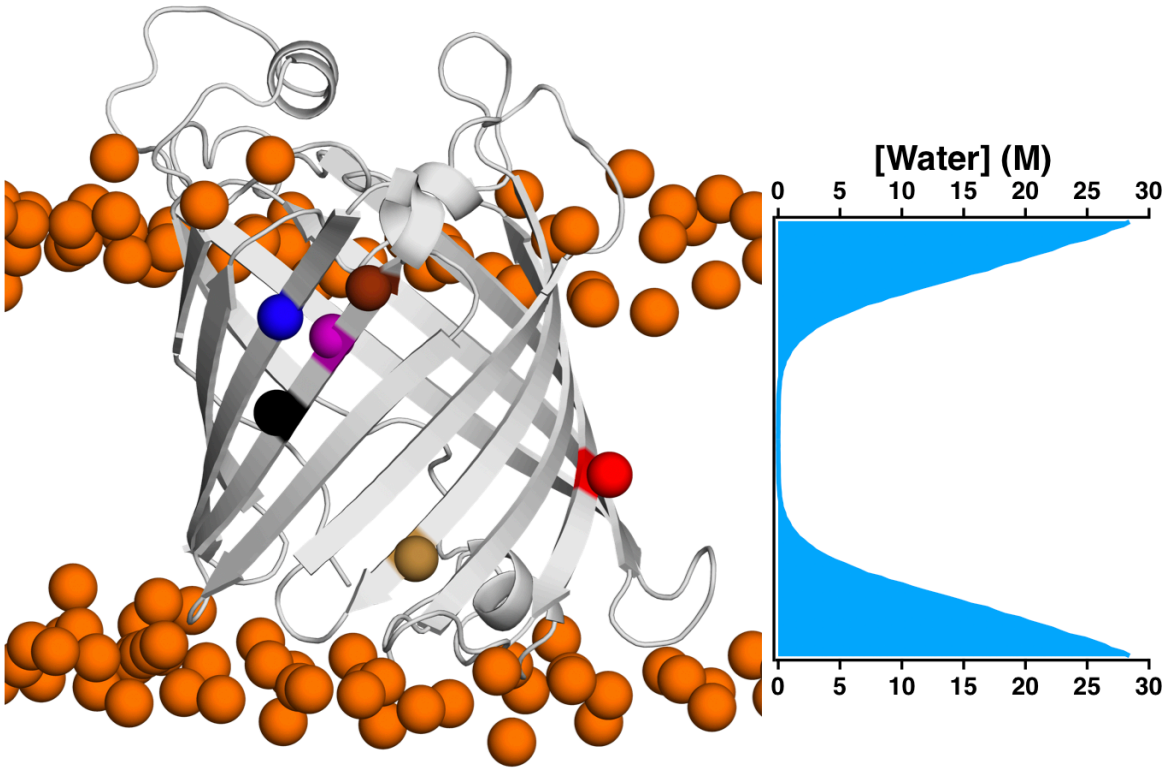

Figure 1. Host-sites on OmpLA exist in different water concentrations in the bilayer. (Left) A snapshot of a molecular dynamics simulation of WT OmpLA in a DLPC bilayer is shown. Phosphate atoms of the DLPC bilayer are colored orange and six host sites on OmpLA used in this study are shown as colored spheres (black:210, yellow:164, red:120, purple:212, blue:223, brown:214). (Right) The gradient of water concentrations inside the phosphate plane of the bilayer is plotted as a function of position in the bilayer. These values were obtained from previously published simulations of neat DLPC bilayers[16]. This water gradient is aligned with the structure of OmpLA at the two phosphate planes. The sites on OmpLA were chosen because they sample a wide range of z-positions positions and chemical compositions.

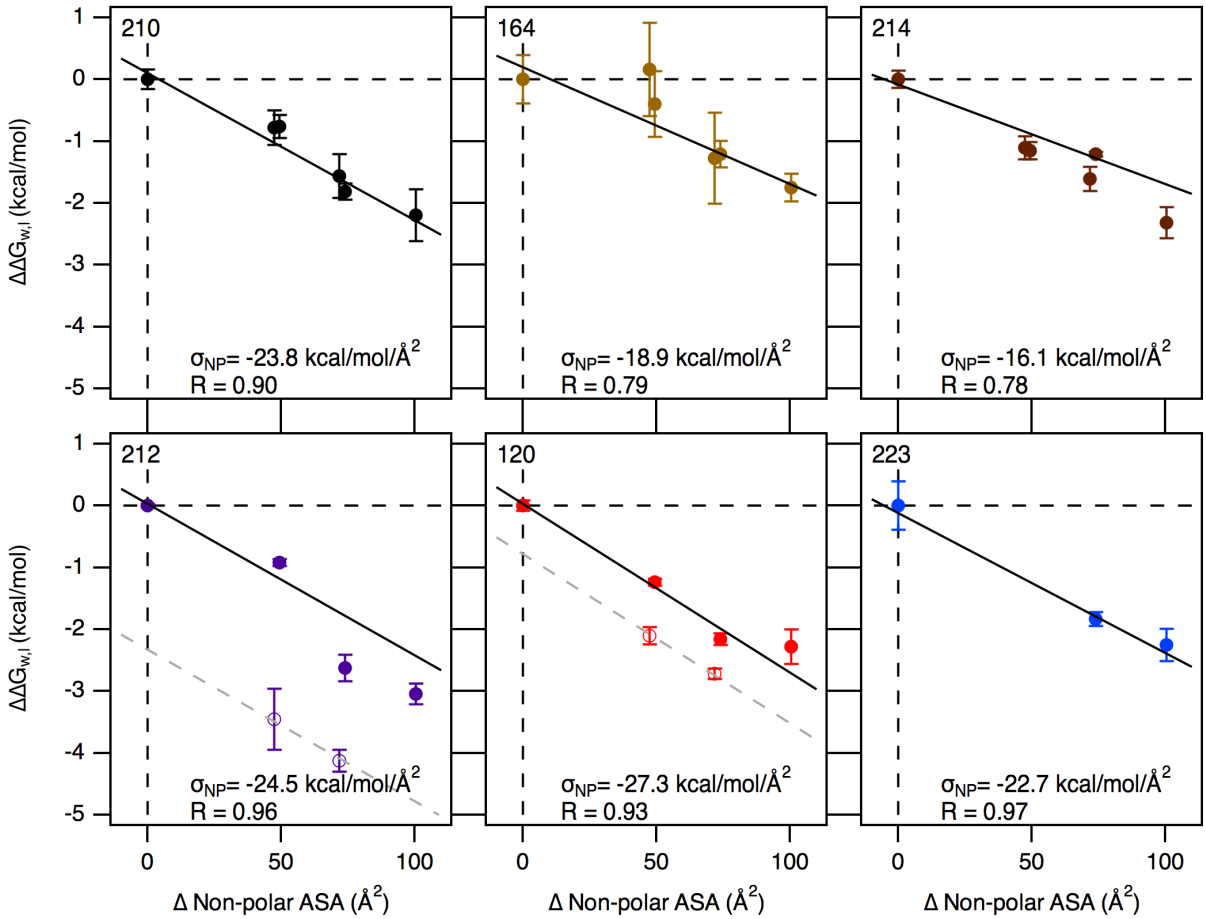

Figure 2. Nonpolar solvation parameter is not constant throughout the bilayer. The nonpolar solvation parameter σ_{NP} was determined for each site on OmpLA by calculating the slope of the linear relationship between $\Delta\Delta G^{\circ}{}_{w,l}$ for each nonpolar side chain at that site with the change in buried surface area compared to alanine in a Gly-X-Gly peptide ($\Delta ASA_{Guest} = ASA_{Guest} - ASA_{Ala}$; ΔASA values reported previously).[17] Linear fits were weighted by the error in $\Delta\Delta G^{\circ}{}_{w,l}$ (Table S1) for each site, with the slope of the line (σ_{NP}) and the Pearson correlation coefficient reported for each fit shown at the bottom of each panel. For sites 212 and 120, beta-branched side chains isoleucine and valine were omitted from σ_{NP} determination as they had anomalously greater $\Delta\Delta G^{\circ}{}_{w,l}$ compared to the other nonpolar side chains. The σ_{NP} determined from non-beta-branched side chains (solid line in 212 and 120 panels) fits the beta-branched side chains (dotted line), indicating that the increase in energy is derived from favorable local interactions that are restricted to beta-branched residues. By taking the difference in the y-intercept for the two lines, we estimate the energy gained due to local interactions for beta-branched residues at these sites to be 2.38 kcal/mol for site 212 and 0.80 kcal/mol for site 120. The σ_{NP} for site 223 was determined only using previously collected data.[17]

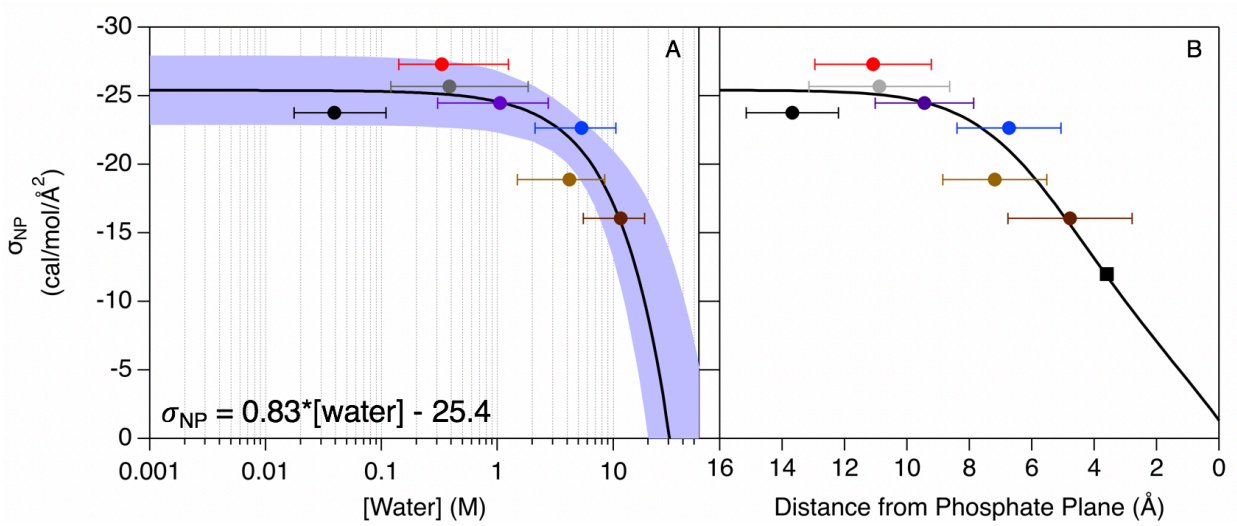

Figure 3. Nonpolar solvation parameter is linearly correlated with water concentration in the bilayer. The water-to-bilayer σ_{NP} calculated at each host site in OmpLA (colored as in Figure 1) and PagP (gray point) plotted as a function of water concentration **(A)** and distance from the phosphate plane **(B).**[19] σ_{NP} are colored according to the scheme in Figure 1, with the σ_{NP} derived from PagP position 111 colored gray. The position of σ_{NP} for a given site is the average of the positions of each side chain at that site and the error bars reflect the standard deviation (Table S3, NP row). σ_{NP} and water concentration are linearly correlated with the equation shown in the bottom right hand corner of Panel **A** (black line: $R^2 = 0.84$; 95% confidence interval shaded in light blue). Using the known relationship between bilayer position and water concentration shown in Figure 1, the direct relationship between σ_{NP} and bilayer position can be determined (Panel **B)**. We were also able to assign a previously measured "water-to-interface" solvation parameter (-12 cal/mol/Å^2) to an exact position in the bilayer (3.6Å from the phosphate plane) using this function (shown as a black square in Panel **B**).[18] The position-dependent function describing σ_{NP} allows for the $\Delta\Delta G°_{sc}$ of any nonpolar side chain to be determined for any position of the bilayer inside the phosphate plane (shown in **Figure 4**).

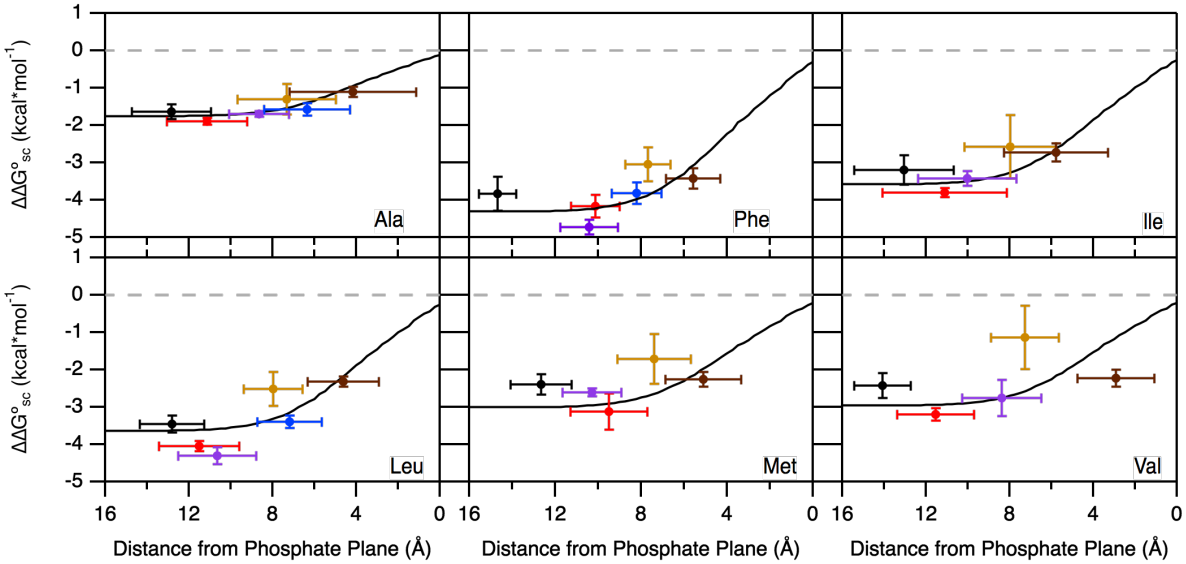

Figure 4. Bilayer position dependence of nonpolar ΔG°_{sc}. Reference-free side chain transfer free energies are plotted as a function of the average distance from the bilayer phosphate plane, as determined from molecular dynamics trajectories. ΔG°_{sc} for each host site in OmpLA are colored as in **Figure 1**. Error bars for both ΔG°_{sc} and bilayer position represent the standard deviations. The solid black line is the $\Delta G^{\circ}_{sc}(z)$ profile calculated for each nonpolar side chain using the derived position dependence of σ_{NP} multiplied by the nonpolar surface area of each side chain (**Figure 3**).

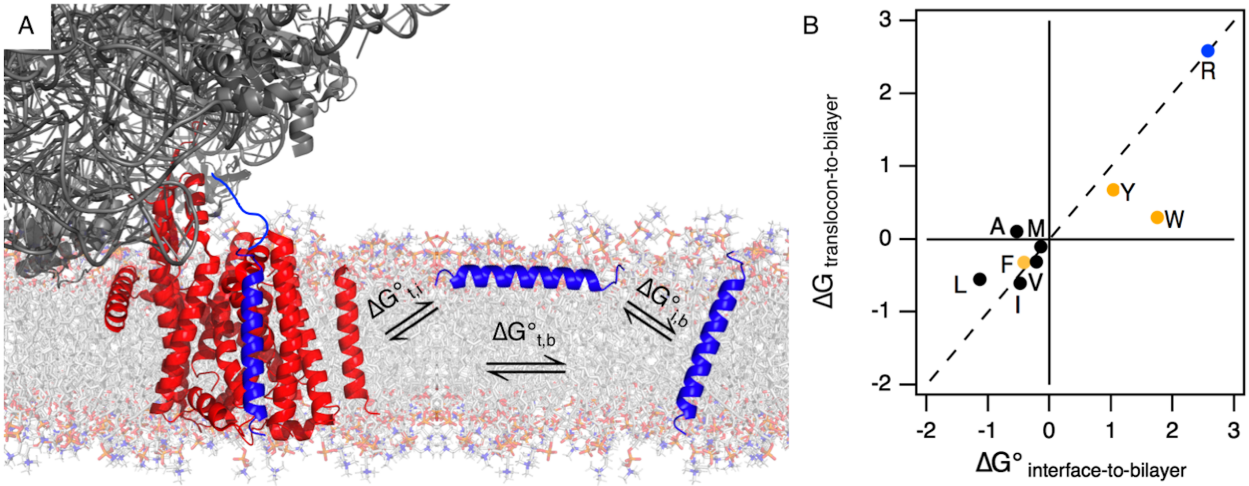

Figure 5. The translocon energetically mimics the bilayer interface. (A) A cartoon schematic of the co-translational insertion of a helix (blue) via the translocon (ribosome colored dark gray, translocon colored red; PDB code: 5GAE) is shown on the left. The helix can either partition to an interfacial conformation, which is energetically described by $\Delta G^{\circ}{}_{t,i}$, or to a transmembrane conformation, described by $\Delta G^{\circ}{}_{t,b}$ (t = translocon, i = interfacial, b = transmembrane). The interfacial-to-transmembrane conformation, which can occur in the absence of the translocon, is energetically described by $\Delta G^{\circ}{}_{i,b}$. **(B)** The values for $\Delta G^{\circ}{}_{t,b}$ and $\Delta G^{\circ}{}_{i,b}$ are plotted for nine side chains, with the dotted line representing energetic equivalence between the two equilibria (black = nonpolar, gold = aromatic, blue = ionizable).[34] For all side chains, except Trp, $\Delta G^{\circ}{}_{t,b} \approx \Delta G^{\circ}{}_{i,b}$, indicating that the translocon energetically mimics the interface ($\Delta G^{\circ}{}_{t,i} \approx 0$). The deviation for Trp is hypothesized to be due to differential interactions between the two experimental set ups.