Lipid bilayer induces contraction of the denatured state ensemble of a helical-bundle membrane protein

Kristen A. Gaffney[a,1], Ruiqiong Guo[b,1], Michael D. Bridges[c,d,1], Shaima Muhammednazaar[b], Daoyang Chen[b], Miyeon Kim[b], Zhongyu Yang[e], Anthony L. Schilmiller[f], Nabil F. Faruk[g], Xiangda Peng[h,i], A. Daniel Jones[a,f], Kelly H. Kim[a], Liangliang Sun[b], Wayne L. Hubbell[c,d], Tobin R. Sosnick[h,i,2], and Heedeok Hong[a,b,2]

[a]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824; [b]Department of Chemistry, Michigan State University, East Lansing, MI 48824; [c]Jules Stein Eye Institute, University of California, Los Angeles, CA 90095; [d]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095; [e]Department of Chemistry and Biochemistry, North Dakota State University, Fargo, ND 58108; [f]Research Technology Support Facility Mass Spectrometry and Metabolomics Core, Michigan State University, East Lansing, MI 48824; [g]Graduate Program in Biophysical Sciences, The University of Chicago, Chicago, IL 60637; [h]Department of Biochemistry & Molecular Biology, The University of Chicago, Chicago, IL 60637; and [i]Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL 60637

Edited by Donald Engelman, Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT; received May 17, 2021; accepted November 17, 2021

Defining the denatured state ensemble (DSE) and disordered proteins is essential to understanding folding, chaperone action, degradation, and translocation. As compared with water-soluble proteins, the DSE of membrane proteins is much less characterized. Here, we measure the DSE of the helical membrane protein GlpG of *Escherichia coli* (*E. coli*) in native-like lipid bilayers. The DSE was obtained using our steric trapping method, which couples denaturation of doubly biotinylated GlpG to binding of two streptavidin molecules. The helices and loops are probed using limited proteolysis and mass spectrometry, while the dimensions are determined using our paramagnetic biotin derivative and double electron–electron resonance spectroscopy. These data, along with our *Upside* simulations, identify the DSE as being highly dynamic, involving the topology changes and unfolding of some of the transmembrane (TM) helices. The DSE is expanded relative to the native state but only to 15 to 75% of the fully expanded condition. The degree of expansion depends on the local protein packing and the lipid composition. *E. coli*'s lipid bilayer promotes the association of TM helices in the DSE and, probably in general, facilitates interhelical interactions. This tendency may be the outcome of a general lipophobic effect of proteins within the cell membranes.

denatured state | membrane protein folding | GlpG | steric trapping | *Upside* simulation

Denatured states and intrinsically disordered proteins (IDPs) are involved in a variety of cellular events, serving as targets for degradation, chaperone action, translocation, and cell signaling (1–4). They also influence thermodynamic stability and direct early folding events (5). Thus, delineating the denatured state ensemble (DSE) has been a subject of extensive study (1). For water-soluble proteins, DSE can be described as a collection of disordered conformations that interconverts on a timescale faster than folding (6).

For the quantitative description of the DSE, polymer theory has proven useful (7, 8): For a given polymer/solvent combination, the solvent quality can be classified into three limiting regimes depending on the relative strengths between intrachain and chain–solvent interactions (7). In a "good" solvent, chain–solvent interactions are more favorable than the intrachain interactions and the DSE is well described as a self-avoiding random walk. In a "θ" solvent, the intrachain and chain–solvent interactions are balanced such that the intrachain attraction cancels out the expansion caused by excluded chain volume and the protein behaves as a random walk. In a "poor" solvent, the intrachain interaction exceeds the chain–solvent interactions, inducing contraction or even collapse of the chain into a globule.

For these three scenarios, solvent quality can be quantified by the Flory exponent, ν, in the relationship between the radius of gyration (R_g) and the number of monomeric units (N): $R_g \propto N^{\nu}$. The exponent is $\nu = 3/5$, 1/2, and 1/3 for good, θ, and poor solvents, respectively. The ν value for self-avoiding random walk has been measured for unfolded proteins in elevated denaturant concentrations with small-angle X-ray scattering and Förster resonance energy transfer (9–13). For a variety of IDPs with protein-like sequences and the DSE of soluble proteins, ν values are typically above 1/2 under native conditions, much higher than anticipated given the general perception that soluble proteins undergo hydrophobic collapse prior to folding (12, 14, 15).

In contrast, the properties of the DSE are not as well characterized for membrane proteins. The folding of helical membrane proteins is thought to occur through at least two stages

Significance

The conformational properties of denatured states of proteins affect their folding, turnover, translocation, and chaperone action. We find that the denatured state ensemble of a helical membrane protein is richer than depicted in the classic two-stage model of membrane protein folding. Some helices undergo topological changes and partial unfolding in the denatured state. In the bilayer, the ensemble is compact relative to the fully expanded state, indicating that the bilayer may not act as a good solvent. This behavior, however, varies depending on the properties of the system, including the local protein sequence, helical and linker lengths, and lipid composition. This study provides insights into the lipid bilayer as a solvent that mediates the assembly and function of membrane proteins.

Author contributions: K.A.G., R.G., M.D.B., W.L.H., T.R.S., and H.H. designed research; K.A.G., R.G., M.D.B., S.M., D.C., M.K., Z.Y., A.L.S., N.F.F., X.P., K.H.K., T.R.S., and H.H. performed research; M.D.B., D.C., A.L.S., N.F.F., A.D.J., K.H.K., L.S., W.L.H., and T.R.S. contributed new reagents/analytic tools; K.A.G., R.G., M.D.B., S.M., D.C., M.K., Z.Y., A.L.S., X.P., K.H.K., W.L.H., T.R.S., and H.H. analyzed data; and K.A.G., R.G., M.D.B., D.C., A.L.S., K.H.K., W.L.H., T.R.S., and H.H. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]K.A.G., R.G., and M.D.B. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: trsosnic@uchicago.edu or honghd@msu.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2109169119/-/DCSupplemental.

Published December 28, 2021.

(16). In Stage I, hydrophobic segments in a polypeptide chain insert into the lipid bilayer as transmembrane (TM) helices. In cells, insertion is cotranslationally mediated by a translocon complex (17, 18). In Stage II, TM helices associate into a compact native structure. The DSE for helical membrane proteins has been regarded as the conformational ensemble prior to forming the native structure (19), in which individual TM helices can diffuse around within the bilayer, although the extent may be limited by the length of the interconnecting loops and residual interhelical contacts. The concept of solvent quality and Flory exponent is harder to quantitatively apply to membrane proteins due to the physical constraints of the quasi–two-dimensional (2D) bilayer and the difficulty in treating the denatured chains as simple polymers.

The underlying principle of the two-stage model is that each stage is driven by distinct driving forces. In Stage I, the formation and insertion of TM helices are favored by the hydrophobic nature of TM segments and the high desolvation cost of having unpaired moieties for backbone hydrogen bonding (H bonding) within the bilayer (17, 20). The role of water in TM helix association likely is less important in Stage II once the individual TM helices have been dehydrated and are favorably solvated with lipids (21). Nonetheless, water may play a role when TM helices have partially hydrated polar residues (22). The association of TM helices requires good van der Waals packing or polar interactions (21, 23–25) and is also known to be modulated by the physical properties of the lipid bilayer (e.g., the lipid packing density and the lipid deformation induced by the hydrophobic mismatch between the protein and the bilayer) (26–29). Thus, a combination of various molecular forces determines how favorable TM helix association is and whether the lipid bilayer functions as a good solvent for the DSEs of membrane proteins.

So far, the DSEs of helical membrane proteins have been characterized under denaturing conditions induced by urea or guanidine hydrochloride (30, 31), anionic detergents including sodium dodecyl sulfate (SDS) (24, 32–36), or mechanical force (37, 38) in various hydrophobic environments (e.g., micelles, lipid–detergent mixtures, or lipid bilayers). These studies suggest that, relative to native states, DSEs are expanded but their TM helices are constrained by unfolded interhelical loops. The degree of expansion also depends on the choice of denaturant and may be limited by the size of micelles (30, 33–36). Lipid bilayers provide a distinct but biologically relevant 2D environment. As a result, the conformational features of the DSE obtained under denaturing conditions may be different from those in cellular membranes.

Here, we reconstitute the DSE of a helical-bundle membrane protein, *Escherichia coli* (*E. coli*) GlpG in a native-like lipid bilayer, and study its conformation without denaturant. GlpG is a member of the rhomboid intramembrane protease family with high kinetic stability (*SI Appendix*, Table S1) (36–40). As with any stable protein, the characterization of the DSE under native conditions is challenging due to its low population and short lifetime (12). We generate GlpG's DSE using our steric trapping method, which couples the spontaneous denaturation of a doubly biotinylated protein to the simultaneous binding of two bulky monovalent streptavidin (mSA, 52 kDa) molecules at ambient temperature (Fig. 1*A*). By combining multiple experimental measurements and molecular dynamics (MD) simulations to probe the flexibility and physical dimensions of the DSE at the local and global levels, we demonstrate that the DSE is highly dynamic, involving extensive helical fraying and changes in membrane topology. The DSE expands relative to the native state but retains a degree of compactness in the lipid bilayers. The degree of expansion further depends on the local stability in the native state and the lipid composition.

Results

Reconstitution of the On-Pathway DSE in Lipid Bilayers. To measure the properties of the DSE, we applied our steric trapping method (Fig. 1*A*) (36). GlpG is labeled with biotin tags at two specific residues that are spatially close in the folded state but on different structural segments. The first mSA can bind unhindered to either biotin tag. The binding of a second mSA, however, is sterically disallowed since the two biotins are close in the native state, but the binding is allowed when the protein is denatured. Taking advantage of the stability and long lifetime of the mSA–biotin complex ($K_{d,biotin}$ $\sim 10^{-14}$ M; $k_{off,biotin}$ ~days) (41), this method enables the trapping and characterization of the DSE even under native conditions.

Previously, we have shown that the TM domain of GlpG (residues 87 through 276) can be separated into the N- (TM1–L1–TM2–L2–TM3–L3) and C-subdomains (TM4–L4–TM5–L5–TM6), each having distinct folding behaviors (Fig. 1*B*) (36). The three sites selected for biotinylation were located at the N and carboxyl termini and on the middle loop (M) connecting the helices TM2 and TM3 ($Pro95_N$, $Val267_C$, and $Gly172_M$, respectively) (36). These three sites located on the same side of the protein were used pairwise to disrupt either the entire protein (with the residue pair 95_N267_C) or just the N- or C-subdomain (with 95_N172_M or 172_M267_C). Each of the three pairs was substituted with a pair of cysteine residues that were then labeled with the thiol-reactive biotin derivative possessing a spectroscopic probe (fluorescent pyrene, BtnPyr, or a nitroxide spin label, BtnRG) (Fig. 1*C* and *SI Appendix*, Fig. S1) (36). Neither conjugation of the biotin derivative to each single cysteine variant nor the binding of an mSA molecule to each biotin label inhibited the proteolytic activity of GlpG relative to wild type (WT) for the model substrate LYTM2 (the second TM segment of an *E. coli* lactose permease, LacY) (*SI Appendix*, Fig. S2*A*) (36). We reconstituted sterically denatured GlpG into two lipid bilayer environments: 1) Large negatively charged bicelles (i.e., discoidal bilayer fragments edge-stabilized by detergent) composed of *di*$C_{14:0}$–phosphatidylcholine (DMPC), *di*$C_{14:0}$–phosphatidylglycerol (DMPG), and 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS) (molar ratio = 3:1:1.4); 2) Liposomes composed of *E. coli* phospholipids (*SI Appendix*, Fig. S3) to provide a native-like lipid environment for *E. coli* GlpG. Based on the cryoelectron microscopic (cryo-EM) images of the bicelles and liposomes, the average diameters were 190 ± 40 Å and 510 ± 220 Å, respectively (Fig. 1*A*).

We initially attempted to obtain the DSE using steric trapping with native GlpG reconstituted in bicelles and liposomes. However, incubation with excess mSA did not yield noticeable denaturation for 1 wk when GlpG activity was used as a folding indicator, consistent with GlpG having high kinetic stability in bilayers (*SI Appendix*, Fig. S4). However, denaturation was achieved in dodecylmaltoside (DDM) detergent micelles for all three constructs within 48 h of incubation with mSA (*SI Appendix*, Figs. S2*B* and S4). With doubly conjugated BtnRG, the activities of the 95_N172_M, 172_M267_C, and 95_N267_C constructs in micelles were reduced by 57 ± 15%, 74 ± 9%, and 88 ± 7% of their native enzymatic activity, respectively (*SI Appendix*, Fig. S2*C*). This activity loss is correlated with the level of double biotinylation (42 ± 2%, 55 ± 2%, and 60 ± 1% for 95_N172_M, 172_M267_C, and 95_N267_C, respectively) (*SI Appendix*, Figs. S1 and S2*C*), implying that double binding of mSA results in a loss of activity and the residual activity level stems from incomplete biotin labeling.

For reconstitution in bilayers, sterically denatured GlpG prepared in micelles (Fig. 1*A*) was transferred to bicelles or to liposomes by direct injection (for the liposomes, detergents were further removed using polystyrene beads, *SI Appendix*, Fig. S3). To verify incorporation of GlpG into bicelles, we employed a

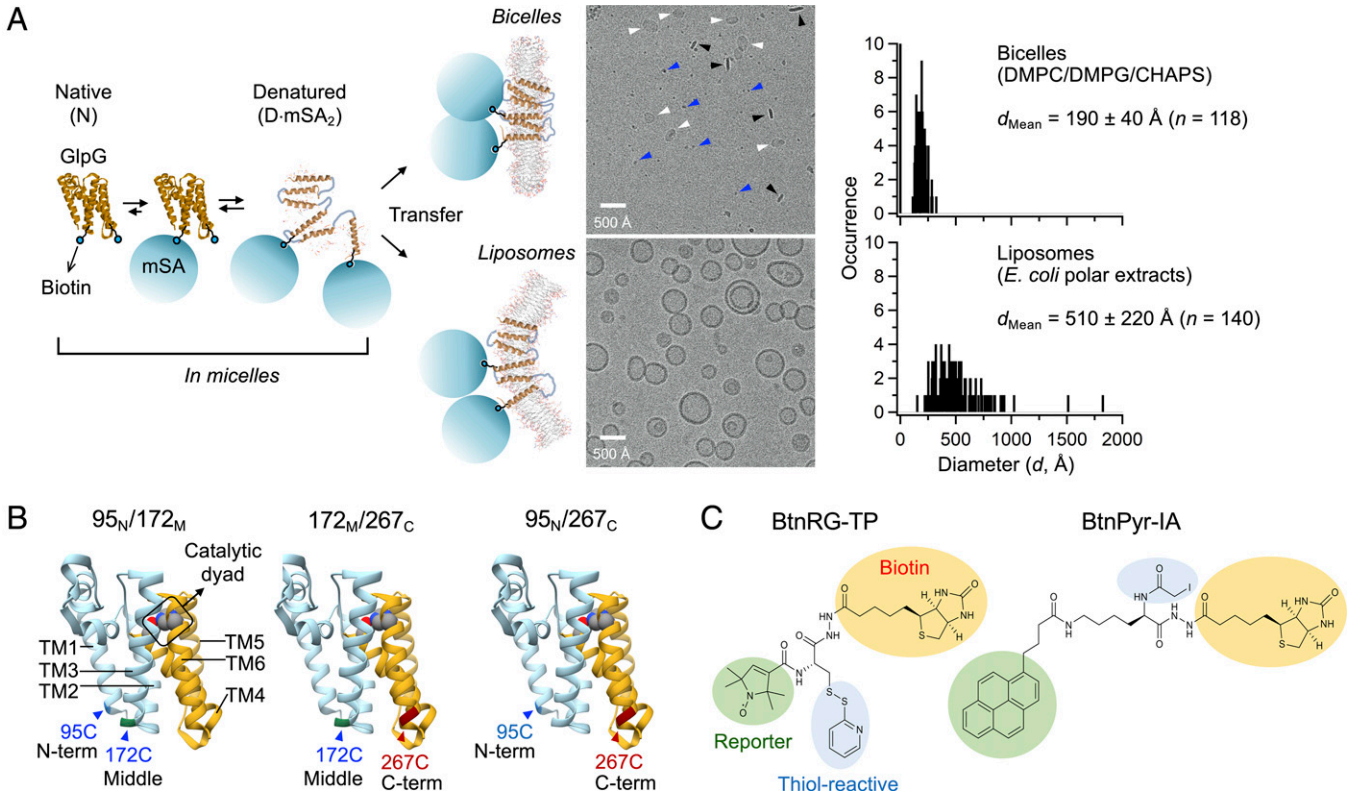

Fig. 1. Steric trapping strategy to reconstitute denatured GlpG in the lipid bilayers. (*A*, *Left*) Doubly biotinylated GlpG was first denatured using steric trapping by the addition of excess mSA in micelles. Sterically denatured GlpG was transferred to bicelles or liposomes. (*Center*) Cryo-EM images of the bicelles and liposomes. In the top bicelle panel (with GlpG and mSA), the particle types were assigned as follows. White arrowheads (circular and ellipsoidal objects): flat or tilted bicellar disk planes; black arrowheads (the dark rod-shaped objects): bicellar rims; blue arrowheads (the dark small particles with diameters of 40 to 50 Å): mSA molecules. The bottom liposome panel (without the proteins) largely contains unilamellar liposomes with a minor fraction of multilamellar liposomes. (*Right*) The histograms for the diameter size distributions of the bicelles and liposomes. (*B*) Double cysteine variants employed for the steric trapping of the denatured state of GlpG. In each variant, designated cysteine residues were labeled with a thiol-reactive biotin derivative with a spectroscopic reporter group as shown in *C*. The regions of the backbone colored in cyan and orange indicate the N- and C-subdomains, respectively. (*C*) Thiol-reactive biotin derivatives with a paramagnetic spin label (*Left*, BtnRG–TP) and fluorescent pyrene (*Right*, BtnPyr–IA). Reproduced with permission from ref. 36. TP: thiopyridine; IA: iodoacetamide.

fluorescence-quenching assay (Fig. 2*A*). Here, the double cysteine variants were labeled with fluorescent BtnPyr (Fig. 1*C*). Native and sterically denatured GlpG were injected into the bicelles containing the quencher-labeled lipid, diC$_{18:1c9}$–phosphatidylethanolamine (DOPE)–dabcyl. After injection, pyrene fluorescence from both native and denatured GlpG was quenched to a level close to that observed for full incorporation.

To verify the incorporation of GlpG into liposomes, we employed a liposome-flotation assay (Fig. 2*B* and *SI Appendix*, Fig. S5*A*). Native and denatured GlpG were doubly labeled with BtnPyr and reconstituted in the liposomes containing the fluorophore-labeled lipid, diC$_{16:0}$–PE (DPPE)–rhodamine. Both native and denatured GlpG cofloated with the liposomes after centrifugation in a sucrose gradient, indicating membrane association. They were also resistant to sodium carbonate extraction, indicating membrane integration (*SI Appendix*, Fig. S5*B*).

Next, we tested whether the denaturation status of GlpG initially prepared in micelles was retained after reconstitution in bilayers (Fig. 2*C* and *SI Appendix*, Fig. S6). To trap the denatured state, we used BtnRG to doubly biotinylate GlpG. This biotin label forms a disulfide linkage to cysteine, which can be broken with the addition of a reducing agent, dithiothreitol (DTT). After reconstitution of the GlpG samples containing mSA, the activities relative to the doubly biotinylated native forms (95_N172_M, 172_M267_C, and 95_N267_C) in bicelles were 43 ± 12%, 40 ± 17%, and 12 ± 7%, respectively, and 56 ± 20%, 49 ± 14%, and 0 ± 4% in liposomes, similar to the values in micelles. Thus, sterically denatured GlpG in micelles largely remained denatured after transfer to bilayer environments.

We tested whether denatured GlpG in the lipid bilayers can refold after the release of the steric repulsion between bound mSA molecules (Fig. 2*C*). Upon addition of DTT to dissociate the BtnRG labels bound with mSA from GlpG, the activity relative to the native form increased to >70%. Thus, sterically denatured GlpG in the bilayers can refold to its native structure and is not a dead-end product incapable of refolding.

Limited Proteolysis Reveals DSE Flexibility. To investigate the conformational features of GlpG's DSE in a native-like lipid bilayer and other hydrophobic environments, we employed limited proteolysis using Proteinase K (ProK). ProK is a robust endopeptidase that nonspecifically proteolyzes water-exposed unstructured regions rather than structured or membrane-buried regions (42, 43). We biotinylated GlpG with BtnRG so that the addition of DTT would remove the mSA molecules from GlpG and its fragments, simplifying the analysis of proteolyzed GlpG with sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) or mass spectrometry.

Upon denaturation, the conformational flexibility and solvent accessibility increased throughout the protein. Time-dependent proteolysis was measured using SDS-PAGE for native and sterically denatured GlpG trapped for the three

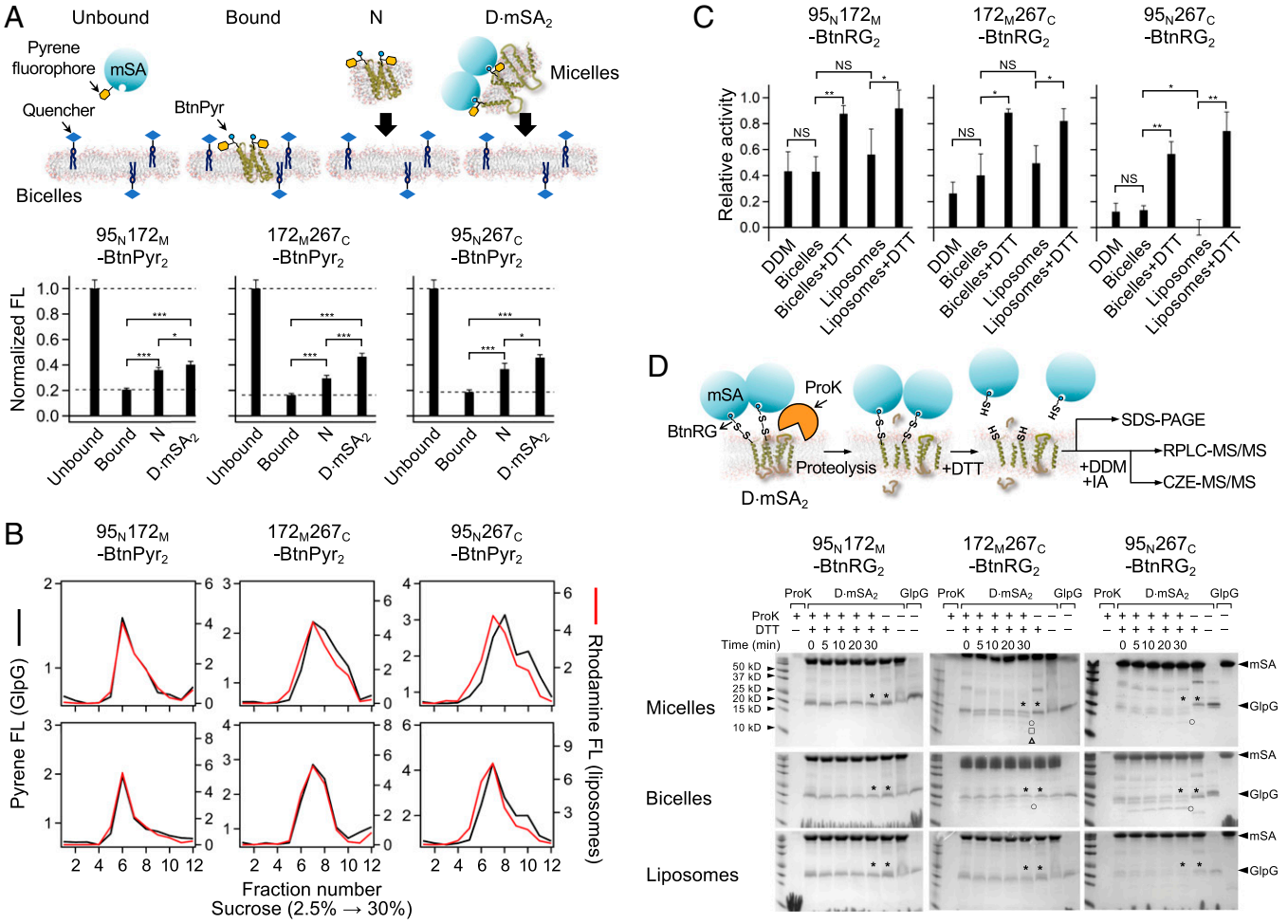

Fig. 2. Reconstitution of denatured GlpG in the lipid bilayers. (*A*, *Top*) Schematic description of fluorescence-quenching assay to measure incorporation of native and sterically denatured GlpG into bicelles. GlpG was doubly labeled with fluorescent BtnPyr, while the bicelles contained quencher-labeled lipids, DOPE–dabcyl. Native ("N") and denatured ("D·mSA$_2$") GlpG solubilized in micelles were injected into bicellar solution. Negative control ("Unbound"): water-soluble pyrene-labeled mSA was injected into the bicellar solution. Positive control ("Bound"): native GlpG was first reconstituted in DMPC:DMPG:DOPE–dabcyl liposomes and then solubilized by CHAPS to form bicelles. (*Bottom*) Assay results for the three doubly biotinylated GlpG variants. Error bars denote ± SEM (n = 3). Based on Student's t test, significance of the difference between a data pair is marked with single ($P < 0.05$), double ($P < 0.01$), and triple asterisks ($P < 0.001$) or "NS" (not significant, $P > 0.05$). (*B*) Liposome-flotation assay in a sucrose gradient indicates a near-complete membrane association of native and denatured GlpG doubly labeled with BtnPyr. Liposomes contained fluorescent DPPE–rhodamine. (*C*) Proteolytic activity of GlpG as a measure of denaturation and refolding efficiency. GlpG doubly labeled with BtnRG was first denatured upon addition of excess mSA in DDM and then transferred to bicelles or liposomes. The addition of DTT, which breaks the disulfide linkage between GlpG and the biotin label bound with mSA, enables refolding ("+DTT"). The activity was normalized relative to that of native GlpG without mSA in each hydrophobic phase. Error bars denote ± SEM (n = 3). Significance of the difference between a data pair is marked as in *A*. (*D*, first row) The schematic description of limited proteolysis of native (N) and sterically denatured GlpG (D·mSA$_2$) by ProK. The time-dependent proteolysis in DDM micelles (second row), bicelles (third row), and liposomes (fourth row) was measured by SDS-PAGE. After quenching of proteolysis at each time point, DTT was added to release bound mSA from GlpG. The extent of GlpG proteolysis was quantified from the GlpG band intensities with and without ProK (asterisks). The proteolytic peptide fragment larger than 10 kDa is marked with a symbol (open circles, 17 kDa; open squares, 13 kDa; open triangles, 11 kDa) on the right side of each band. For mass spectrometry (see Fig. 3), the samples were further solubilized with DDM and the cysteine residues were alkylated with iodoacetamide (IA).

biotin combinations (Fig. 2*D*). Native GlpG (23 kDa, the TM domain) was resistant to proteolysis except at the termini due to the protein's high kinetic stability (*SI Appendix*, Figs. S4 and S7) (44). In contrast, denatured GlpG was rapidly proteolyzed within 10 min in all hydrophobic environments. However, 20 to 50% of the GlpG sample containing mSA was not proteolyzed, probably due to incomplete biotin labeling (*SI Appendix*, Figs. S1 and S2*C*): GlpG doubly bound with mSA was proteolyzed, while unbound or singly bound GlpG was largely protease resistant (*SI Appendix*, Figs. S7 and S8). Notably, when denatured GlpG was trapped with a biotin pair at the N-subdomain (95_N172_M), it was extensively proteolyzed into small fragments (<8 kDa). However, when the denatured protein was trapped with the biotin pair at either the C-subdomain (172_M267_C) or the N and carboxyl termini (95_N267_C), proteolysis yielded large fragments in micelles and bicelles (~11, ~13, or ~17 kDa). This result supports our previous finding that in micelles, the disruption of the N-subdomain leads to global denaturation, whereas the C-subdomain undergoes subglobal denaturation with the balance of the protein remaining intact (36). In liposomes, however, denatured GlpG was extensively proteolyzed regardless of the position of the biotin pair, implying that the entire chain was less structured, contracted, or buried in liposomes than in micelles and bicelles.

To identify the unstructured and solvent-accessible regions that are sensitive to proteolysis, we employed capillary-zone

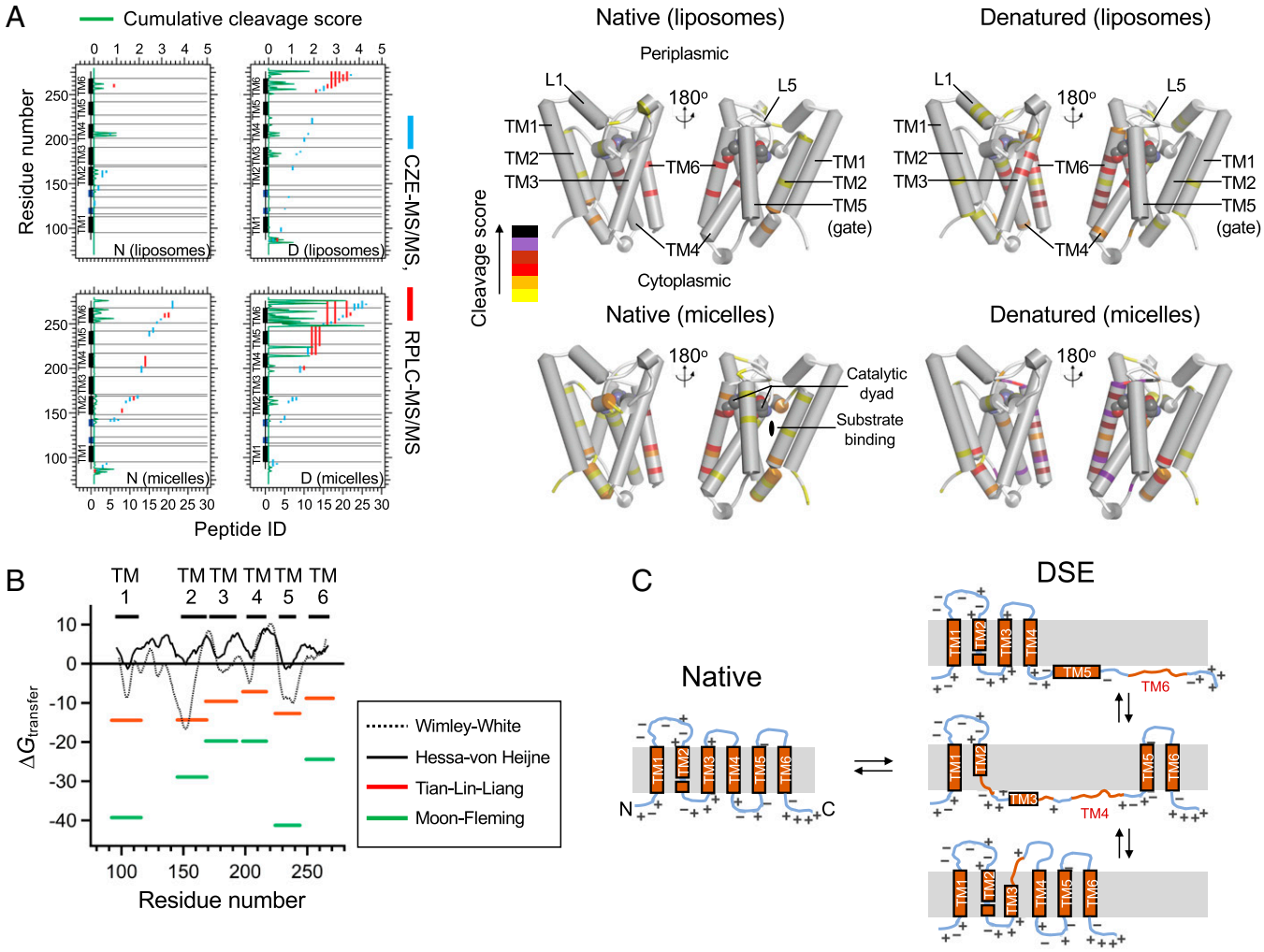

Fig. 3. Mapping the flexible regions in native and denatured GlpG using limited proteolysis by ProK and mass spectrometry. (*A*, *Left*) The proteolysis products of native and denatured GlpG (the double biotin variant 172_M267_C) in liposomes (*Top*) and micelles (*Bottom*) were analyzed using CZE- (cyan vertical bars) and RPLC- (red vertical bars) MS/MS. The peptide identification number (peptide ID, lower *x*-axes) denotes a unique fragment identified under each designated condition in the ascending order of the N-terminal residue number of the fragment. Each vertical bar spans the range from the N- to the carboxyl-terminal residue number of each fragment and is mapped onto the primary and secondary structures (*y*-axes). The traces in green represent both the accuracy and frequency (the cumulative score, upper *x*-axes) of each cleavage site (*y*-axes) (*SI Appendix, Supplementary Text*). (*Right*) The cumulative score of each site was mapped onto the tertiary structure and color coded according to the heat map. The catalytic dyad (Ser201–His254) is shown in spheres in each structure. (*B*) Hydropathy plot of the TM domain of GlpG using the Wimley-White (WW) water–octanol (88), Hessa-von Heijne (HvH) translocon–membrane (17), Tian-Lin-Liang (TLL) membrane depth-dependent (44, 89), and Moon-Fleming (MF) water–membrane (74) hydrophobicity scales. For the WW and HvH scales, a 19-residue window was slid along the sequence of GlpG summing $\Delta G_{transfer}$ of each residue. For the TLL and MF scales, the $\Delta G_{transfer}$ values only for the residues in each TM helix from the structure were summed. (*C*) Possible modes of the membrane-topology dynamics and unfolding in the DSE of GlpG in *E. coli* liposomes. The models were based on limited proteolysis, MS/MS, MD simulations, and the charge distributions in the membrane-water interfacial regions of GlpG.

electrophoresis (CZE)-tandem mass spectrometry (MS/MS) or reverse-phase liquid chromatography (RPLC)-MS/MS (*SI Appendix*, Tables S2–S7). Each method utilizes a distinct principle for peptide separation (electrophoretic mobility for CZE and solubility for RPLC), allowing for complementary peptide identification (45, 46). We chose the double-biotin variant, 172_M267_C–$BtnRG_2$, because it had a higher biotin-labeling efficiency than the other variants and yielded a larger fraction of denatured GlpG molecules. Overall, RPLC-MS/MS enabled the detection of the larger peptides from the TM regions in the native structure, while the smaller peptides from the loops were identified using CZE-MS/MS (Fig. 3 *A*, *Left*).

The native state ensemble (NSE) of GlpG in the bilayer likely undergoes functional motions that permit cleavage by ProK. We identified several peptide fragments from native GlpG in liposomes (Fig. 3 *A*, *Top*). These sequences were mapped onto the kinked cytoplasmic end of TM2 and the middles of TM4 and TM6. Notably, these regions are directly involved in the proteolytic mechanism of GlpG: The interface between TM2 and the gating helix TM5 forms the substrate binding site (47). TM4 and TM6 are packed through the conserved glycine-zipper motif harboring the catalytic dyad Ser201–His254 (48). Upon denaturation, the entire length of TM4 and TM6 became susceptible to proteolysis in addition to the periplasmic half of TM3 and the flanking loops of TM5 (Fig. 3 *A*, *Bottom*). The middle regions of TM3, TM4, and TM6 also were highly susceptible to proteolysis. These three helices are the least hydrophobic in GlpG (Fig. 3*B*), suggesting that they can transiently become unstructured and solvent exposed in the DSE.

In micelles, the overall proteolysis pattern was similar to that in the bilayers for native and denatured GlpG, although in micelles, cleavage was more pronounced in the C-subdomain

than in the N-subdomain (Fig. 3 *A*, *Bottom*). The extensive proteolysis in the C-subdomain in both bilayers and micelles points to higher dynamics at this end of the protein, being an intrinsic property of the protein rather than stemming from steric repulsion in this subdomain. Supporting this reasoning, denatured GlpG trapped using the biotin pair (95_N172_M) in the N-subdomain still underwent extensive proteolysis in the C-subdomain (*SI Appendix*, Fig. S9).

In summary, our proteolysis study reveals dynamic features of the DSE under native conditions. The DSE involves various modes of membrane-topology changes including the unfolding of TM segments due to their low hydrophobicity and, probably, the charge distribution in the flanking regions (Fig. 3*C*). Although we minimized the duration of proteolysis (5 to 7 min) to capture unstructured regions in the DSE of full-length GlpG, it is likely that a certain fraction of the identified peptides were obtained from the polypeptides already proteolyzed ("double hit"). Even so, the proteolysis patterns nonetheless would reflect intrinsic conformational features of individual structural segments in their isolated forms.

Contraction of the DSE Measured with Double Electron–Electron Resonance Spectroscopy. We next examined the compactness of the DSE in micelles, bicelles, and liposomes using double electron–electron resonance spectroscopy (DEER) suitable for measuring distances between 18 to 60 Å (49). Specifically, we measured the interspin distances between a pair of paramagnetic biotin labels (BtnRG) in native and sterically denatured GlpG for the three constructs, 95_N172_M–$BtnRG_2$, 172_M267_C–$BtnRG_2$, and 95_N267_C–$BtnRG_2$. The probability distribution of interspin distances was obtained from the time-dependent dipolar evolution data fitted with the model-free, nonnegative Tikhonov regularization algorithm (Fig. 4) (49). To minimize unwanted intermolecular dipolar coupling between multiple spin-labeled GlpG molecules in a single liposome, we increased the lipid-to-protein molar ratio six-fold (from 2,000 to 12,000) as compared to our activity and proteolysis measurements and incorporated unlabeled GlpG at a three- or six-times molar excess relative to spin-labeled GlpG (*SI Appendix*, Figs. S10 and S11).

The probability distributions of interspin distances, $P(r)$, were similar for the native proteins in micelles, bicelles, and liposomes (Fig. 4*B*). Their distributions were narrow centered at a short mean distance (r_{Mean}) of 29 to 30 Å for 95_N172_M (the SD of distribution, σ = 6 to 9 Å), 25 to 33 Å for 172_M267_C (σ = 4 to 11 Å), and 37 to 38 Å for 95_N267_C (σ = 6 to 9 Å) (*SI Appendix*, Table S8). For the DSE, however, distributions of the three variants were highly heterogeneous with multiple local maxima spanning the distance range detectable by DEER. We have previously shown that upon denaturation by steric trapping in DDM micelles, r_{Mean} values increase by factors of 1.52 ± 0.02 and 2.03 ± 0.01 for 95_N172_M and 172_M267_C, respectively (36). For the new 95_N267_C variant, the r_{Mean} between the termini increased by a factor of 1.35 ± 0.01 upon denaturation (Fig. 4*B*). Thus, steric trapping overall induced expansion of the DSE in micelles.

We expected that relative to micelles, which are less organized assemblies, the quasi-2D lipid bilayer would confine the movements of the TM segments in the DSE, leading to its compaction. Indeed, the DSEs in bicelles were less expanded than in micelles by factors of 1.23 ± 0.02, 1.37 ± 0.03, and 1.29 ± 0.01 for 95_N172_M, 172_M267_C, and 95_N267_C, respectively. Nevertheless, the DSE in liposomes was more expanded than in bicelles despite both being quasi-2D bilayers by factors of 1.38 ± 0.01, 1.65 ± 0.03, and 1.45 ± 0.04 for 95_N172_M, 172_M267_C, and 95_N267_C, respectively. This difference may be attributed to the use of DMPC:DMPG:CHAPS in the bicelles versus *E. coli* phospholipids in the liposomes. Overall, the expansion of the DSE depended on the position of the steric repulsion as well as the lipid composition.

To test the effect of doubly bound mSA molecules on the compactness of the DSE, we measured the interspin distances in the SDS-induced DSE of a double-biotin variant with and without mSA (*SI Appendix*, Fig. S12). In SDS, mSA is resistant to denaturation and can still bind to the biotin labels (*SI Appendix*, Fig. S1). SDS molecules are expected to bind the surface of mSA (50). The resulting negatively charged mSA–SDS complexes may repel each other when doubly bound to GlpG, inducing further expansion of the DSE. Nonetheless, binding of mSA in SDS did not noticeably affect the distance distribution. Also, the distances measured with bound mSA in a high concentration of SDS (~4% wt/vol) were similar to those in neutral DDM micelles (*SI Appendix*, Fig. S12). Thus, repulsive or attractive interactions between bound mSA molecules, even if they exist, do not significantly affect the compactness of sterically denatured GlpG. This result supports our modeling of mSA bound as hard spheres in the DSE simulations (see *DSE Simulations*).

DSE Simulations. To generate models of compact and expanded DSEs and provide references for comparison to the DEER data, we used our *Upside* MD simulations modified for membrane proteins. This recently created algorithm can fold small proteins with accuracy comparable to all-atom methods (51, 52). *Upside* employs six interaction centers per residue: the three backbone atoms (N, C_α, and C), the carbonyl oxygen and amide proton, and a side chain represented by a directional-dependent bead. After every MD step, the side chains are globally repacked to have the lowest global free energy. Our coarse graining, the use of an implicit solvent, and the global side-chain packing step explain much of the 10^3- to 10^4-fold speed up compared to standard all-atom simulations (51, 52).

Our membrane-burial potential includes knowledge-based, depth- and lipid exposure–dependent energies for side-chain burial and backbone H bonding within the bilayer (53, 54). Energies are determined from the statistics of a large training set of proteins, which accounts for both the depth Z in the membrane and the level of side-chain exposure to the lipid: $Energy(Z, \text{exposure}) \propto \ln(\text{frequency})$. We incorporate depth-dependent energies for unsatisfied backbone H-bond donors and acceptors within the bilayer and allow helices to fold and unfold during the simulations.

At every MD step, side-chain burial is recalculated to account for the exchange of protein–lipid interactions for protein–protein interactions as helices come into contact. This recalculation avoids overestimating the energetics of helix association. Through these and the careful treatment of the bilayer boundary position, our potential can optimally predict the bilayer thickness and protein location, as compared to the OPM (Orientations of Proteins in Membranes) method (55). Our force-induced unfolding studies using the membrane potential for GlpG and bacteriorhodopsin agree well with the experiment (53).

A total of 20 4-ms simulations were run for native GlpG and the three doubly mSA-bound versions (95_N172_M, 172_M267_C, and 95_N267_C) at T = 274, 308, 343, 377, and 411 K (400 simulations total, *SI Appendix*, Fig. S13 *A*–*F*). Each mSA was held in its native conformation using a set of stiff restraint springs and linked with a stiff 4-Å spring at Asn49, a position near the biotin-binding pocket, to each of the two biotinylation sites on GlpG (Fig. 5*A*). To eliminate steric overlap, only short-range repulsive interactions were employed between the two bound mSA molecules with each other and with GlpG (a sigmoid potential with ~1 Å width and a drop-off at 8 Å and the repulsive energy of $2RT$, where R is a gas constant). To inhibit the

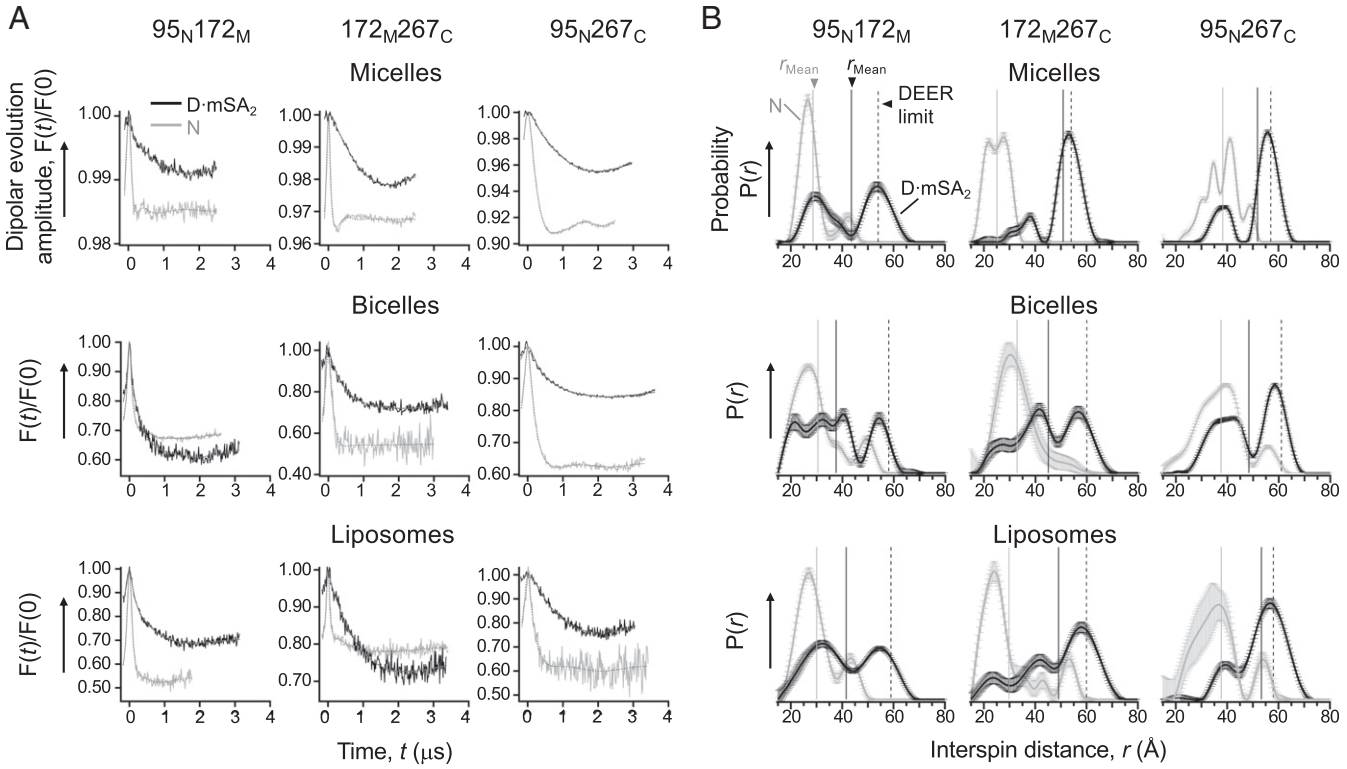

Fig. 4. The physical dimension of the denatured states of GlpG measured by DEER. (*A*) Time-dependent dipolar evolution data for native (N, gray) and sterically denatured GlpG (D·mSA$_2$, black) were fit using the model-free, nonnegative Tikhonov regularization algorithm (red). The data for 95_N172_M and 172_M267_C in micelles were adapted from ref. 36. (*B*) Corresponding distance distributions between the spin labels at the designated residues. For each, the mean distances (r_{Mean}) are shown as vertical solid lines for the native (gray) and denatured (black) states. The error bar at each distance corresponds to ± SD of the fitted probability from the model-free analysis. The "DEER limit" (dashed black line) denotes the maximal nominal interspin distance detectible by DEER.

mSA molecules from entering the bilayer, they were subject to the same membrane-burial potential as GlpG.

During the simulations, GlpG and the three doubly mSA-bound versions maintained a native and near-native fold at the lowest two temperatures, 274 K and 308 K, respectively (*SI Appendix*, Fig. S13 *A* and *B* and Table S9). The C_α-RMSD for mSA-free GlpG was 5 ± 1 Å and 9 ± 1 Å at 274 K and 308 K, respectively (Fig. 5*B*). In support of our steric trapping strategy, the RMSDs of the three mSA-bound GlpG versions were larger (C_α-RMSD = 7 to 9 Å and 10 to 12 Å at 274 K and 308 K, respectively) with the average R_g within 1.2 and 2.5 Å of the native value of 16 Å. Based on these RMSD and R_g values, we selected the 274 K and 308 K simulations to represent the NSE ("native-like$_{274K}$") and a compact DSE ("collapsed$_{308K}$"), respectively.

We ran simulations at higher temperatures in hopes of creating an expanded DSE. At 343 K, the protein displayed significantly larger dynamics than at 308 K (Fig. 5*B*). However, it remained compact (R_g <24 Å) even though the C_α-RMSD was often above 10 Å (*SI Appendix*, Fig. S13*C*), indicating that the helices remained in contact but no longer retained their native arrangement. Notably, some trajectories underwent large excursions in which TM6 unfolded and remained outside the bilayer, allowing the mSA molecules to sample a much larger region of space (Fig. 5 *C* and *D*). This observation agrees with the experimental result that TM6 and the flanking loops of TM5 are prone to proteolysis (Fig. 3). Nonetheless, higher temperatures increased the separation between the TM segments and broadened the corresponding distance distributions. Likewise, the attachment of doubly bound mSA preferentially increased the distance between the structural segments to which their steric repulsion was directly exerted. Based on these conformational features, we referred to this simulated model as "partially expanded DSE$_{343K}$".

Given the persistence of helix–helix contacts at 343 K, we adopted an alternative strategy to create a more expanded DSE (Fig. 5 *B–D* and *SI Appendix*, Fig. S13*F*). A total of 20 additional 4-ms simulations were run retaining H bonding and the membrane potential but with the attractive side-chain terms turned off, leaving only a repulsive term to prevent steric overlap. Without interhelical interactions, the temperature was reduced to 240 K to help maintain the helices within the bilayer. Under this protocol, the helices TM1 through TM5 independently diffused around in the membrane, with their separation restricted by interhelical loops. As observed at 343 K, TM6 occasionally unfolded and partitioned into the surface of the membrane. The two small helices in the loop between TM1 and TM2 occasionally unfolded, which allowed TM1 to separate more than the other helices. The overall C_α-RMSD and R_g values were large, 23 Å and 30 Å, respectively. We selected this DSE for use as our reference "fully expanded DSE$_{OFF}$"

The experimental $T_{m,app}$ determined by heat-induced aggregation or inactivation is ~343 K in DDM micelles and above 358 K in bicelles or *E. coli* liposomes (39, 44). In our simulations, mSA-free GlpG started to melt at ~300 K with the N-subdomain being more resistant than the C-subdomain by ~30 K (Fig. 5 *B–D*). Consistently, we have shown experimentally that the N-subdomain has a higher thermodynamic stability than the C-subdomain by $\Delta\Delta G^{o}_{D\text{-}N}$ = 1.1 kcal/mol in micelles (36).

Comparison between DEER and Simulations. Of primary interest is the comparison between the DEER and the simulated distance distributions for sterically denatured GlpG in bilayers

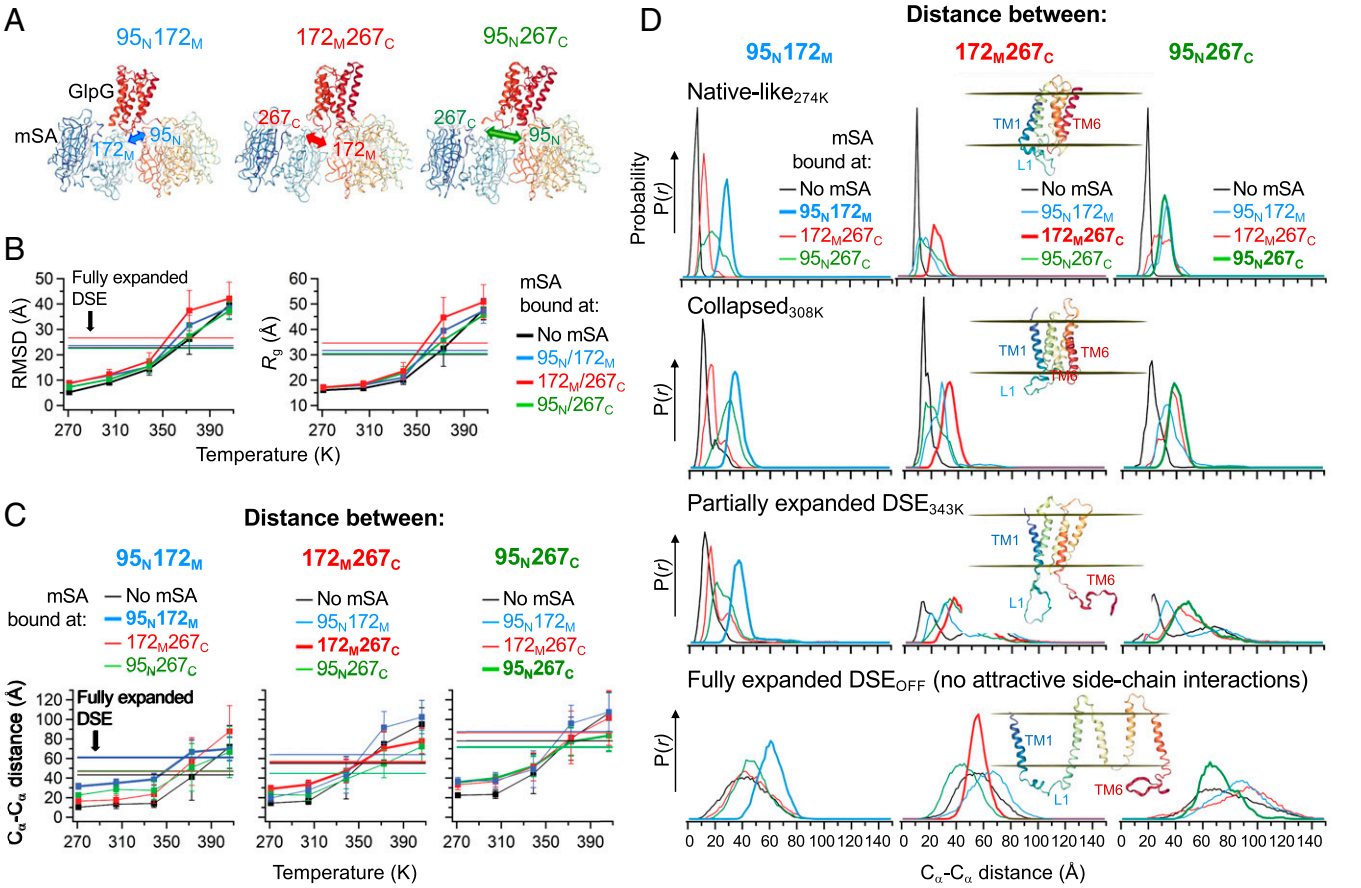

Fig. 5. *Upside* MD simulation under the statistical membrane-burial potential. (*A*) Modeling of GlpG doubly bound with mSA. Each mSA molecule was attached to the designated residue on GlpG via a stiff 4-Å virtual spring (spring constant = 30 kcal·mol^{-1}·Å^{-1}). (*B*) The C_α-RMSD and radius of gyration (R_g) of GlpG from temperature-dependent simulations. Each value is the average of 20 independent simulations. Each error bar denotes ± SD. The horizontal guidelines (also in *C*) indicate the average values for the simulation runs under the fully expanded protocol. At $T > 343$ K, a more diverse set of molecular configurations is observed, such as the complete unfolding of individual TM helices and the formation of β-strands within the membrane (*SI Appendix*, Fig. S13 *D* and *E*). (*C*) The C_α–C_α distances between each designated residue pair were monitored at an increasing temperature. When the distances were monitored between a specific residue pair ("Distance between"), mSA molecules are bound to the same or either of the other two residue pairs ("mSA bound at"). (*D*) Distribution of the C_α–C_α distances between the designated residue pair with or without bound mSA. The simulations were carried out with all energy terms ("native-like$_{274K}$," "collapsed$_{308K}$," and "partially expanded DSE$_{343K}$") or missing attractive side-chain interactions ("fully expanded DSE$_{OFF}$"). A representative structural snapshot under each simulation condition is shown as an inset.

(Fig. 6*A* and *SI Appendix*, Tables S8 and S9). From this comparison, we can determine whether the DSE is collapsed, or partially or fully expanded. In our GlpG constructs, the paramagnetic center in each BtnRG label was separated from the labeled residue in GlpG by several covalent linkages (Fig. 1*C*). To account for this separation in the simulations, we monitored the distances between the Asn49–C_α atoms near the biotin-binding pocket in mSA, which were the attachment sites to GlpG (Fig. 4*A*). The distances obtained in the "native-like$_{274K}$" state were close to the corresponding DEER distances in native GlpG with the average discrepancy of 1.5 ± 2.9 Å (discrepancy = $r_{Mean,native\text{-}like,simulation} - r_{Mean,native,DEER}$ for nine datasets in Fig. 6*A*). In the DSE of each GlpG construct, the r_{Mean} obtained from DEER did not exceed the corresponding r_{Mean} in the "fully expanded DSE$_{OFF}$" from simulation. These observations support the validity of direct comparison between experiment and simulation.

To quantitatively describe the expansion of the DSE relative to the collapsed and fully expanded states, we defined the normalized expansion ratio, $R_{Expansion}$ (Fig. 6*B*):

$$R_{Expansion} = (r_{Mean,DEER} - r_{Mean,collapsed})/(r_{Mean,expanded} - r_{Mean,collapsed}), \quad [1]$$

where $r_{Mean,DEER}$ is the mean distance between a specific residue pair in denatured GlpG determined by DEER, while $r_{Mean,collapsed}$ and $r_{Mean,expanded}$ denote the mean distances between the same pair in the simulated collapsed and fully expanded DSE, respectively. When the DSE was trapped using mSA at the more stable N-subdomain (95_N172_M) which has more extensive intraprotein contacts in the native state than the C-subdomain (*SI Appendix*, Fig. S14), $R_{Expansion}$ was 0.05 ± 0.04 in bicelles; that is, this region in the DSE was as compact as the simulated collapsed state. In contrast, when the DSE was trapped using mSA at the less stable C-subdomain (172_M267_C), $R_{Expansion}$ reached 0.49 ± 0.08; that is, the DSE expanded close to the simulated partially expanded state. In the DSE trapped at the termini (95_N267_C), $R_{Expansion}$ was intermediate (0.26 ± 0.01) which may be an outcome of the collapse in the N-subdomain being offset by the expansion in the C-subdomain.

As noted, the DSE was more expanded in the liposomes than in the bicelles (Figs. 4 and 6 *A* and *B*). Upon denaturation, the $R_{Expansion}$ values in liposomes increased from the corresponding values in bicelles by $\Delta R_{Expansion}$ = 0.18 ± 0.04, 0.22 ± 0.12, and 0.17 ± 0.03 for the N-subdomain, C-subdomain, and termini, respectively. This result is consistent with the DSE

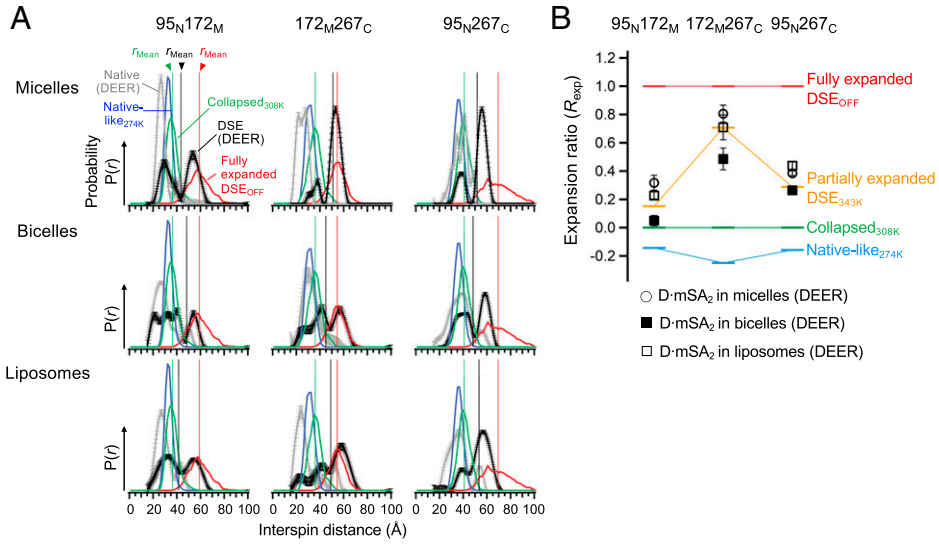

Fig. 6. Comparison of the dimensions of the DSEs measured by DEER and *Upside* simulations. (*A*) Comparison of the distance distributions obtained from DEER and the simulated DSEs doubly bound with mSA. To mimic the interspin distances measured from DEER, the distances from the simulations were monitored between the Asn49 C_α's on bound mSA molecules, which were the attachment sites to GlpG. The r_{Mean} values for the simulated "collpased$_{308K}$" (green), the experimentally obtained "DSE (DEER)" (black), and the simulated "fully expanded DSE$_{OFF}$" (red) are shown as vertical lines. (*B*) Expansion ratio ($R_{Expansion}$, Eq. **1**) of the DSE measured between each residue pair in the micelles, bicelles, and liposomes. The $R_{Expansion}$ values of the simulated DSEs at an increasing temperature are overlaid.

having higher proteolysis levels in liposomes than in bicelles irrespective of the position of the biotin pair (Fig. 2*D*).

Since the diameters of the bicelles (190 ± 40 Å) and liposomes (510 ± 220 Å) are larger than the maximal N-to-C distances (~100 Å) in the simulated fully expanded DSE, it is unlikely that the dimensions of the lipid environment affected the compactness of the experimentally obtained DSE. Given the general trends, we infer that for the DSE of GlpG, bicelles and liposomes can qualitatively behave as solvents somewhere between poor and good solvent limits. The precise behavior depends on the specific region in the protein and the type of bilayer, suggesting an inherent degree of variability to solvent quality in different combinations of proteins and bilayers.

Discussion

By combining experiment and simulation, we provided a molecular-level description of the DSE of a membrane protein in a native-like lipid bilayer. The overall picture is a partially collapsed DSE with considerable dynamics and susceptibility to proteolytic cleavage. TM helices can transiently dock against each other but also can unfold. The comparison of the interhelical distance distributions from DEER measurements and the simulations suggests that bicelles and liposomes behave as solvents somewhere between the poor and good solvent limits depending on the bilayer composition and the region of the protein being considered.

Conformational Diversity of the DSE. The two-stage folding model has served as a useful framework to describe the folding of helical membrane proteins (16). Our experimental data and simulations indicate that GlpG's DSE is more diverse than what is expected from this canonical model where formation of TM helices occurs before their association. Whereas individual TM segments undergo restricted motions within the lipid bilayer, less hydrophobic segments can become solvent exposed after undergoing membrane-topology changes and even unfolding, suggesting that the hydrophobicity of the TM segments is a key determinant of the conformational distribution in the DSE.

Diversity in membrane topology has also been observed locally (e.g., the N-terminal TM helix of multidrug transporter EmrE) (56) and globally (e.g., EmrE and a lactose permease LacY) (57–59) depending on the hydrophobicity of the TM as well as the distribution of charged residues in the flanking regions. Notably, a hydropathy analysis of helical membrane proteins in *E. coli* predicts that about half of the TM segments have a low tendency to insert into the membrane ($\Delta G_{app,insertion} > 0$) (56). This observation, combined with our current findings, suggests that in general, multiple segments that are TM helices may not always remain so in the DSEs of the membrane proteome.

We find that the DSE of GlpG expands more in *E. coli* liposomes than in bicelles despite both having a similar quasi-2D physical constraint. Although multiple factors are operational, the difference could be explained by differing lipid composition. Bicelles are composed of the saturated lipids, neutral *di*$C_{14:0}$−PC (75 mol-%) and negatively charged *di*$C_{14:0}$−PG (25%), and CHAPS, while *E. coli* lipids contain neutral PE (~75 wt/wt-%) and negatively charged phosphatidylglycerol (PG, ~20%) and cardiolipin (~5%) headgroups with 16:0 (~40%), 16:1c9 (~30%), and 18:1c11 (~15%) fatty acyl chains (60). Each monolayer leaflet in the bilayer containing the small headgroup and unsaturated PE or cardiolipin tends to bend outward (negative spontaneous curvature) compared to that containing the larger headgroup and saturated PC or PG lipids (61–64). Hence, the *E. coli* lipid bilayer is "frustrated" by the difference between the spontaneous and actual curvatures with an increased repulsive force between the hydrocarbon chains (lateral tension) in the bilayer core. The increase in lateral tension may facilitate the separation of TM segments within the bilayer or their exclusion to the water–membrane interface, inducing increased expansion and conformational diversity in the DSE.

Contraction of the DSE in the lipid bilayer reflects a balance between the strengths of protein–protein, protein–lipid, and lipid–lipid interactions. Contraction of the DSE can be driven by the increased chemical potential for intramolecular van der Waals packing and the burial of polar residues. For example, the N-subdomain of GlpG, which is more contracted in the DSE, contains more intraprotein contacts than the C-subdomain in the native state (*SI Appendix*, Fig. S14). Also, an all-atom MD simulation suggests that in the folding of the

CLC-ec1 Cl^-/H^+ antiporter monomer, the association of the N- and C-domains is promoted by the burial of polar residues and the release of water molecules (22).

A hydrophobic mismatch between the bilayer and TM helices can cause lipid deformation at the TM helix–lipid interface facilitating helix–helix association (27). In the dynamic monomer–dimer equilibrium of CLC-ec1, a coarse-grained MD simulation finds that dimerization is driven by the local lipid deformation induced by the hydrophobic mismatch between the bilayer and the dimer interface (28). A recent folding and assembly study of the tetrameric ion channel KcsA finds that helical monomers rapidly associate into a protein-dense region in asolectin liposomes, indicating that this class of bilayer serves as a poor solvent for these monomers (65). For the isolated monomers, however, the two TM helices are often separated due to the difference in their length, which results in different tilt angles. In general, multiple lipid- and protein-dependent factors can influence the effective solvent quality (65).

Lipid composition can also influence the compactness of the DSE by modulating the properties of the bulk lipid bilayer including "force-from-lipids" (66). For example, TM helix association is affected by lipid composition with negligible changes in the bilayer thickness for various types of single membrane-spanning helices, including the TM domain of glycophorin A (26, 67), a proton channel M2 (68), a transcriptional regulator Mga2 (29), and an unfolded protein stress sensor Ire1-α (69). The origin of such lipid dependence has been attributed to the changes in lateral tension profile (26), lipid packing density (29), membrane fluidity (69), and electrostatic interactions between negatively charged membranes and positively charged residues flanking the TM helices (26).

Challenges in Quantifying the Solvent Quality of Lipid Bilayers. Solvent quality has been quantified for disordered soluble proteins using the Flory scaling relationship, $R_g \propto N^{\nu}$ (9, 10, 12, 14). This powerful formalism, however, is not readily adaptable to DSEs of membrane proteins. Although a tempting 2D counterpart is a disordered chain in a quasi-2D slab (70), this is a poor representation of the DSE of a membrane protein, as the DSE contains multiple TM helices positioned orthogonally to the lipid bilayer and connected by loops that reside mainly outside the bilayer. Even for a pair of TM helices linked by a flexible loop of increasing length, the scaling behavior will be largely determined by the portion of the loop lying outside the bilayer in a quasi-3D environment. Furthermore, membrane proteins have different topologies and helical content with variable length of linkers, and some helices may even lie along the bilayer. These issues emphasize that deriving a generalized scaling law for the DSE of membrane proteins is likely to be difficult.

As an alternative, we propose that the experimental distances be compared to those of a reference ensemble obtained from simulations of noninteracting TM helices connected by unstructured linkers of appropriate length. However, as observed with GlpG, TM helices can exit the bilayer and partially unfold. Hence, one cannot assume all the TM helices remain intact and care must be taken even with this strategy to properly model the reference ensemble.

Implications for Thermodynamics. Thermodynamic stability is the difference in free energy between the native and a reference state, typically the DSE. Our study has provided a detailed picture of a DSE for a helical membrane protein: For GlpG, the helices largely remain intact in the DSE, but they are dynamic with variable levels of interhelical contacts. The DSEs of other membrane proteins include helical ensembles in SDS, which trade native tertiary contacts for protein–detergent interactions (40, 71), an extended chain in solution when force is applied normal to the bilayer at one end using atomic force microscopy (AFM) (72), and a "zig-zag" state with intact but noncontacting TM helices in the membrane when force is applied to both termini parallel to the bilayer plane using magnetic tweezers (37). In the AFM measurements, the penalty for removing the hydrophobic residues from the lipid bilayer contributes to the stability, whereas this factor is less relevant in the tweezer measurements, as the helices largely remain within the bilayer (37, 72). In contrast, in some folding studies of β-barrel membrane proteins using chemical denaturants, the reference state is a disordered chain in solution (73, 74).

Because the mode of denaturation alters which factors change upon folding, comparisons of the stability ($\Delta G^{\circ}_{N\text{-}D}$) obtained from different modes should be made judiciously. This issue also applies to mutational studies; for example, a change in helical propensity will not affect the change in stability ($\Delta\Delta G^{\circ}_{N\text{-}D,WT\text{-}Mut}$) if the site remains helical in the DSE. Hence, meaningful interpretation of thermodynamic parameters requires the characterization of the DSE. With this extra challenge, along with the inherent difficulties of quantifying the energetics of protein–lipid interactions, it is not surprising that membrane protein thermodynamics remains an extremely challenging area of research.

Broader Implications. The solvent quality of the lipid bilayer for membrane proteins is relevant to a variety of biological processes including folding, association, and proteostasis. Our results imply that the lipid bilayer generally facilitates contraction of DSEs and promotes intraprotein interactions.

This property can present both benefits and challenges. TM segments make both native and nonnative contacts with each other during the cotranslational insertion (75, 76), while the folding efficiency and maturation of newly synthesized membrane proteins is surprisingly small, 20 to 50% (77). In liposomes, the affinity of specific TM helix–helix interactions dramatically decreases in the presence of total membrane protein extracts, probably due to enhanced nonspecific interactions (26). In addition, various types of TM chaperones are either components of the translocon complex or function independently (77–80). These observations imply that cellular membranes are not necessarily optimal environments for efficient assembly of membrane proteins. The bilayer may promote nonspecific compaction and misfolding, resulting in species that become targets for protein quality control mechanisms in the cells (81–83). On the other hand, enhanced intraprotein interactions can facilitate the formation of membrane protein complexes and the clustering of receptors for effective cellular signaling (84, 85). Presumably, the interaction strength is tuned by varying lipid composition in the different bilayers in the cell to achieve an appropriate compromise between folding efficiency and function in each one.

Materials and Methods

Detailed information on materials, preparation and biotinylation of GlpG, activity determination, mass spectrometry, and *Upside* MD simulation can be found in the SI Appendix.

Preparation of Biotinylated GlpG. Double cysteine variants (95C172C, 172C267C, and 95C267C) of *E. coli* GlpG with an N-terminal His_6-tag were expressed in *E. coli* BL21(DE3)RP cells, purified in DDM (Anatrace), and labeled with thiol-reactive BtnPyr or BtnRG (36).

Activity Assay of GlpG. To measure GlpG activity in micelles and bicelles, the model substrate SN (staphylococcal nuclease)–LYTM2–His_6 fusion in DDM was labeled with the thiol-reactive environment-sensitive fluorophore, iodoacetyl-7-nitrobenz-2-oxa-1,3-diazol (IA-NBD, Setareh Biotech) (36). Time-dependent decrease of NBD fluorescence was monitored with $\lambda_{Ex} = 485$ nm and $\lambda_{Em} = 535$ nm. To measure GlpG activity in liposomes, SN–LYTM2–His_6 was separately labeled with fluorescein–iodoacetamide (Sigma-Aldrich) or dabcyl–maleimide

(Setareh Biotech) and mixed together at the ratio of 1:1. GlpG and labeled LYTM2 separately reconstituted in *E. coli* liposomes were mixed by inducing liposomal fusion upon addition of polyethyleneglycol$_{3350}$ (14 wt/vol-% in the final). The increase in fluorescein fluorescence was monitored with λ_{Ex} = 494 nm and λ_{Em} = 520 nm at 37 °C.

Denaturation of GlpG Using Steric Trapping. GlpG labeled with BtnPyr or BtnRG was incubated with a 5-times molar excess of mSA in DDM at 25 °C until maximum denaturation was reached (up to 48 h). The extent of denaturation was monitored using GlpG activity for LYTM2.

Reconstitution of GlpG. For reconstitution in bicelles, native or sterically denatured GlpG in DDM was injected into 3 wt/vol-% bicelles (DMPC:DMPG:CHAPS, molar ratio = 3:1:1.4) in 20 mM Na_2HPO_4, 40 mM NaCl (pH 7.5). For reconstitution in liposomes, native or sterically denatured GlpG in DDM was added to the extruded liposomes (10 mM *E. coli* polar extracts, Avanti Polar Lipids) equilibrated with 10 mM DDM. DDM was removed using Bio-Beads (Bio-Rad) followed by additional extrusion. The final GlpG concentration was 5 μM.

Cryo-EM. The bicellar (15 wt/vol-%) sample was prepared with GlpG (95_N267_C–BtnRG$_2$, 5 μM) and mSA (25 μM). The liposomal (50 mM *E. coli* polar lipids) sample was prepared using the same method as for reconstituting denatured GlpG but without the proteins. Cryo-EM grids were frozen using a Vitrobot Mark IV (Thermo Fisher). Briefly, 3.5 μL of each sample was applied to a glow-discharged Quantifoil Cu 1.2/1.3 holey carbon 200-mesh grid. The grid was blotted for 3.5 s prior to plunge freezing in liquid ethane. Cryo-EM images were recorded on a Talos Arctica (Thermo Fisher) operated at 200 kV and equipped with a Falcon 3EC direct electron detector camera. Images were recorded in counting mode using EPU software at a nominal magnification of 92,000× (1.12 Å/pixel), with a defocus of −3.0 μm. Micrographs were collected as single-frame images with a total exposure time of 1.5 s and a total dose of 30 electrons/Å^2.

ProK Digestion. Native or sterically denatured GlpG prepared in 10 mM DDM, 3% bicelles, or 10 mM liposomes was proteolyzed by ProK (0.14 μg/mL, Sigma-Aldrich) and quenched by permethylsulfoxide (0.1 mM).

Mass Spectrometry. For CZE-MS/MS, the proteolysis products of free mSA, native GlpG, and denatured GlpG were obtained using the solid phase–enhanced method (86). After washing with 90% acetonitrile and elution with 100 mM NH_4HCO_3 buffer (pH 8), recovered peptides were separated by CZE (the ECE-001 system, CMP Scientific). For RPLC-MS/MS, RPLC was performed on a Vanquish Flex UPLC system (Thermo Fisher) and an Acquity BEH-C4 column (Waters) using a 1 to 99% acetonitrile gradient in water containing 0.1% formic acid. For mass detection, a Q-Exactive HF MS/MS spectrometer (Thermo Fisher) was used with a scan range of 200 to 2,000 *m/z*. The data were analyzed on Byonic (version 3.9.6, Protein Metrics) (87).

DEER Spectroscopy. The final concentration of spin-labeled GlpG in DDM, bicelles, or liposomes was 50 to 100 μM. Four-pulse DEER data were collected on a Q-band Bruker ELEXSYS 580 spectrometer using a 150-W amplifier and an E5106400 cavity resonator (Bruker Biospin). Samples were loaded into quartz capillaries and flash frozen in liquid nitrogen prior to data collection at 50 K. The interspin distances were determined from fits to the background-corrected dipolar evolution data using the model-free, nonnegative Tikhonov regularization algorithm on the LongDistances program (http://www.biochemistry.ucla.edu/Faculty/Hubbell/).

Simulations of DSEs. Simulations used *Upside*, our near-atomic, implicit solvent model that conducts Langevin dynamics on the N, C_α, and C atoms (51, 52). We added a membrane-burial potential that dynamically accounted for the level of side-chain exposure to lipids and unfavorable energies for unsatisfied H-bond donors and acceptors in the membrane (53, 54). Side-chain burial energies at the burial depth are determined from the statistics of a set of helical membrane proteins. The membrane thickness (2.88 nm) was predicted by OPM (Protein Data Bank: 2xov) (55). The calibrated simulation temperatures were 240, 274, 308, 343, 377, and 411 K.

Data Availability. All data supporting the findings of this study are available within this article and *SI Appendix*. The full simulation package of *Upside*, as well as the necessary parameter files, have been deposited in the GitHub repository (https://github.com/sosnicklab/upside-md). Previously published data were used for this work (36).

ACKNOWLEDGMENTS. This work is supported by NIH Grant R01GM118685 to H.H., NIH grants R01GM055694 and R01GM130122, and NSF Grant MCB2023077 to T.R.S., the Jules Stein Professorship endowment and NIH Grants R01EY05216, T33EY07026, and 5P41EB001980 to W.L.H., and NIH Grant R01GM125991 and NSF DBI1846913 to L.S. A.D.J. acknowledges support from Michigan AgBioResearch (US Department of Agriculture National Institute of Food and Agriculture Grant MICL02474). K.H.K. acknowledges support from the Michigan State University Research Technology Support Facility Cryo-EM Core.

1. K. A. Dill, D. Shortle, Denatured states of proteins. *Annu. Rev. Biochem.* **60**, 795–825 (1991).
2. A. Matouschek, Protein unfolding–An important process in vivo? *Curr. Opin. Struct. Biol.* **13**, 98–109 (2003).
3. H. Saibil, Chaperone machines for protein folding, unfolding and disaggregation. *Nat. Rev. Mol. Cell Biol.* **14**, 630–642 (2013).
4. P. E. Wright, H. J. Dyson, Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29 (2015).
5. T. R. Sosnick, D. Barrick, The folding of single domain proteins–Have we reached a consensus? *Curr. Opin. Struct. Biol.* **21**, 12–24 (2011).
6. B. Schuler, Perspective: Chain dynamics of unfolded and intrinsically disordered proteins from nanosecond fluorescence correlation spectroscopy combined with single-molecule FRET. *J. Chem. Phys.* **149**, 010901 (2018).
7. H. S. Chan, K. A. Dill, Polymer principles in protein structure and stability. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 447–490 (1991).
8. P. J. Flory, *Principles of Polymer Chemistry* (Cornell University Press, Ithaca, New York, 1953).
9. G. Fuertes *et al.*, Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E6342–E6351 (2017).
10. H. Hofmann *et al.*, Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16155–16160 (2012).
11. J. E. Kohn *et al.*, Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12491–12496 (2004).
12. J. A. Riback *et al.*, Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science* **358**, 238–241 (2017).
13. W. Zheng *et al.*, Probing the action of chemical denaturant on an intrinsically disordered protein by simulation and experiment. *J. Am. Chem. Soc.* **138**, 11702–11713 (2016).
14. R. B. Best, Emerging consensus on the collapse of unfolded and intrinsically disordered proteins in water. *Curr. Opin. Struct. Biol.* **60**, 27–38 (2020).
15. K. W. Plaxco, I. S. Millett, D. J. Segel, S. Doniach, D. Baker, Chain collapse can occur concomitantly with the rate-limiting step in protein folding. *Nat. Struct. Biol.* **6**, 554–556 (1999).
16. J. L. Popot, D. M. Engelman, Membrane protein folding and oligomerization: The two-stage model. *Biochemistry* **29**, 4031–4037 (1990).
17. T. Hessa *et al.*, Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**, 377–381 (2005).
18. F. Cymer, G. von Heijne, S. H. White, Mechanisms of integral membrane protein insertion and folding. *J. Mol. Biol.* **427**, 999–1022 (2015).
19. J. U. Bowie, Solving the membrane protein folding problem. *Nature* **438**, 581–589 (2005).
20. Z. Cao, J. M. Hutchison, C. R. Sanders, J. U. Bowie, Backbone hydrogen bond strengths can vary widely in transmembrane helices. *J. Am. Chem. Soc.* **139**, 10742–10749 (2017).
21. N. H. Joh, A. Oberai, D. Yang, J. P. Whitelegge, J. U. Bowie, Similar energetic contributions of packing in the core of membrane and water-soluble proteins. *J. Am. Chem. Soc.* **131**, 10846–10847 (2009).
22. D. Min *et al.*, Unfolding of a ClC chloride transporter retains memory of its evolutionary history. *Nat. Chem. Biol.* **14**, 489–496 (2018).
23. K. G. Fleming, D. M. Engelman, Specificity in transmembrane helix-helix interactions can define a hierarchy of stability for sequence variants. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14340–14344 (2001).
24. N. H. Joh *et al.*, Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature* **453**, 1266–1270 (2008).
25. M. Mravic *et al.*, Packing of apolar side chains enables accurate design of highly stable membrane proteins. *Science* **363**, 1418–1423 (2019).
26. H. Hong, J. U. Bowie, Dramatic destabilization of transmembrane helix interactions by features of natural membrane environments. *J. Am. Chem. Soc.* **133**, 11389–11398 (2011).
27. O. S. Andersen, R. E. Koeppe II, Bilayer thickness and membrane protein function: An energetic perspective. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 107–130 (2007).
28. R. Chadda *et al.*, Membrane transporter dimerization driven by differential lipid solvation energetics of dissociated and associated states. *eLife* **10**, e63288 (2021).
29. S. Ballweg *et al.*, Regulation of lipid saturation without sensing membrane fluidity. *Nat. Commun.* **11**, 756 (2020).
30. T. Jacso *et al.*, The mechanism of denaturation and the unfolded state of the α-helical membrane-associated protein Mistic. *J. Am. Chem. Soc.* **135**, 18884–18891 (2013).

31. M. R. Sanders, H. E. Findlay, P. J. Booth, Lipid bilayer composition modulates the unfolding free energy of a knotted α-helical membrane protein. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E1799–E1808 (2018).
32. N. Fehr, I. García-Rubio, G. Jeschke, H. Paulsen, Early folding events during light harvesting complex II assembly in vitro monitored by pulsed electron paramagnetic resonance. *Biochim. Biophys. Acta* **1857**, 695–704 (2016).
33. V. Krishnamani, B. G. Hegde, R. Langen, J. K. Lanyi, Secondary and tertiary structure of bacteriorhodopsin in the SDS denatured state. *Biochemistry* **51**, 1051–1060 (2012).
34. A. Dutta *et al.*, Characterization of membrane protein non-native states. 2. The SDS-unfolded states of rhodopsin. *Biochemistry* **49**, 6329–6340 (2010).
35. A. Dutta, K. C. Tirupula, U. Alexiev, J. Klein-Seetharaman, Characterization of membrane protein non-native states. 1. Extent of unfolding and aggregation of rhodopsin in the presence of chemical denaturants. *Biochemistry* **49**, 6317–6328 (2010).
36. R. Guo *et al.*, Steric trapping reveals a cooperativity network in the intramembrane protease GlpG. *Nat. Chem. Biol.* **12**, 353–360 (2016).
37. H. K. Choi *et al.*, Watching helical membrane proteins fold reveals a common N-to-C-terminal folding pathway. *Science* **366**, 1150–1156 (2019).
38. D. Min, R. E. Jefferson, J. U. Bowie, T. Y. Yoon, Mapping the energy landscape for second-stage folding of a single membrane protein. *Nat. Chem. Biol.* **11**, 981–987 (2015).
39. R. P. Baker, S. Urban, Architectural and thermodynamic principles underlying intramembrane protease function. *Nat. Chem. Biol.* **8**, 759–768 (2012).
40. W. Paslawski *et al.*, Cooperative folding of a polytopic α-helical membrane protein involves a compact N-terminal nucleus and nonnative loops. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7978–7983 (2015).
41. M. Howarth *et al.*, A monovalent streptavidin with a single femtomolar biotin binding site. *Nat. Methods* **3**, 267–273 (2006).
42. C. C. Wu, M. J. MacCoss, K. E. Howell, J. R. Yates III, A method for the comprehensive proteomic analysis of membrane proteins. *Nat. Biotechnol.* **21**, 532–538 (2003).
43. R. N. Besingi, P. L. Clark, Extracellular protease digestion to evaluate membrane protein cell surface localization. *Nat. Protoc.* **10**, 2074–2080 (2015).
44. Y. Yang *et al.*, Folding-degradation relationship of a membrane protein mediated by the universally conserved ATP-dependent protease FtsH. *J. Am. Chem. Soc.* **140**, 4656–4665 (2018).
45. L. Sun *et al.*, Capillary zone electrophoresis-multiple reaction monitoring from 100 pg of RAW 264.7 cell lysate digest. *Analyst (Lond.)* **138**, 3181–3188 (2013).
46. J. N. Savas, B. D. Stein, C. C. Wu, J. R. Yates III, Mass spectrometry accelerates membrane protein analysis. *Trends Biochem. Sci.* **36**, 388–396 (2011).
47. S. Cho, R. P. Baker, M. Ji, S. Urban, Ten catalytic snapshots of rhomboid intramembrane proteolysis from gate opening to peptide release. *Nat. Struct. Mol. Biol.* **26**, 910–918 (2019).
48. Y. Wang, Y. Zhang, Y. Ha, Crystal structure of a rhomboid family intramembrane protease. *Nature* **444**, 179–180 (2006).
49. G. Jeschke, DEER distance measurements on proteins. *Annu. Rev. Phys. Chem.* **63**, 419–446 (2012).
50. D. Winogradoff, S. John, A. Aksimentiev, Protein unfolding by SDS: The microscopic mechanisms and the properties of the SDS-protein assembly. *Nanoscale* **12**, 5422–5434 (2020).
51. J. M. Jumper, N. F. Faruk, K. F. Freed, T. R. Sosnick, Accurate calculation of side chain packing and free energy with applications to protein molecular dynamics. *PLOS Comput. Biol.* **14**, e1006342 (2018).
52. J. M. Jumper, N. F. Faruk, K. F. Freed, T. R. Sosnick, Trajectory-based training enables protein simulations with accurate folding and Boltzmann ensembles in cpu-hours. *PLOS Comput. Biol.* **14**, e1006578 (2018).
53. Z. Wang, J. M. Jumper, K. F. Freed, T. R. Sosnick, On the interpretation of force-induced unfolding studies of membrane proteins using fast simulations. *Biophys. J.* **117**, 1429–1441 (2019).
54. Z. Wang, J. M. Jumper, S. Wang, K. F. Freed, T. R. Sosnick, A membrane burial potential with H-bonds and applications to curved membranes and fast simulations. *Biophys. J.* **115**, 1872–1884 (2018).
55. M. A. Lomize, I. D. Pogozheva, H. Joo, H. I. Mosberg, A. L. Lomize, OPM database and PPM web server: Resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376 (2012).
56. M. Seurig, M. Ek, G. von Heijne, N. Fluman, Dynamic membrane topology in an unassembled membrane protein. *Nat. Chem. Biol.* **15**, 945–948 (2019).
57. R. C. Van Lehn, B. Zhang, T. F. Miller III, Regulation of multispanning membrane protein topology via post-translational annealing. *eLife* **4**, e08697 (2015).
58. H. Vitrac, D. M. MacLean, V. Jayaraman, M. Bogdanov, W. Dowhan, Dynamic membrane protein topological switching upon changes in phospholipid environment. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13874–13879 (2015).
59. N. B. Woodall, S. Hadley, Y. Yin, J. U. Bowie, Complete topology inversion can be part of normal membrane protein biogenesis. *Protein Sci.* **26**, 824–833 (2017).
60. S. Morein, A. Andersson, L. Rilfors, G. Lindblom, Wild-type Escherichia coli cells regulate the membrane lipid composition in a "window" between gel and non-lamellar structures. *J. Biol. Chem.* **271**, 6801–6809 (1996).
61. R. S. Cantor, Lipid composition and the lateral pressure profile in bilayers. *Biophys. J.* **76**, 2625–2639 (1999).
62. A. R. Curran, R. H. Templer, P. J. Booth, Modulation of folding and assembly of the membrane protein bacteriorhodopsin by intermolecular forces within the lipid bilayer. *Biochemistry* **38**, 9328–9336 (1999).
63. S. M. Gruner, Intrinsic curvature hypothesis for biomembrane lipid composition: A role for nonbilayer lipids. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3665–3669 (1985).
64. R. N. Lewis, R. N. McElhaney, The physicochemical properties of cardiolipin bilayers and cardiolipin-containing lipid membranes. *Biochim. Biophys. Acta* **1788**, 2069–2079 (2009).
65. K. C. Song *et al.*, Folding and misfolding of potassium channel monomers during assembly and tetramerization. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2103674118 (2021).
66. C. D. Cox, N. Bavi, B. Martinac, Origin of the force: The force-from-lipids principle applied to Piezo channels. *Curr. Top. Membr.* **79**, 59–96 (2017).
67. V. Anbazhagan, D. Schneider, The membrane environment modulates self-association of the human GpA TM domain–Implications for membrane protein folding and transmembrane signaling. *Biochim. Biophys. Acta* **1798**, 1899–1907 (2010).
68. K. C. Duong-Ly, V. Nanda, W. F. Degrado, K. P. Howard, The conformation of the pore region of the M2 proton channel depends on lipid bilayer environment. *Protein Sci.* **14**, 856–861 (2005).
69. N. Kono, N. Amin-Wetzel, D. Ron, Generic membrane-spanning features endow IRE1α with responsiveness to membrane aberrancy. *Mol. Biol. Cell* **28**, 2318–2332 (2017).
70. M. Daoud, P. G. Degennes, Statistics of macromolecular solutions trapped in small pores. *J. Phys. (Paris)* **38**, 85–93 (1977).
71. K. K. Andersen *et al.*, The role of decorated SDS micelles in sub-CMC protein denaturation and association. *J. Mol. Biol.* **391**, 207–226 (2009).
72. H. Yu, M. G. Siewny, D. T. Edwards, A. W. Sanders, T. T. Perkins, Hidden dynamics in the unfolding of individual bacteriorhodopsin proteins. *Science* **355**, 945–950 (2017).
73. H. Hong, L. K. Tamm, Elastic coupling of integral membrane protein stability to lipid bilayer forces. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 4065–4070 (2004).
74. C. P. Moon, K. G. Fleming, Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10174–10177 (2011).
75. F. Cymer, G. von Heijne, Cotranslational folding of membrane proteins probed by arrest-peptide-mediated force measurements. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 14640–14645 (2013).
76. N. J. Harris *et al.*, Structure formation during translocon-unassisted co-translational membrane protein folding. *Sci. Rep.* **7**, 8021 (2017).
77. J. T. Marinko *et al.*, Folding and misfolding of human membrane proteins in health and disease: From single molecules to cellular proteostasis. *Chem. Rev.* **119**, 5537–5606 (2019).
78. S. Nagamori, I. N. Smirnova, H. R. Kaback, Role of YidC in folding of polytopic membrane proteins. *J. Cell Biol.* **165**, 53–62 (2004).
79. R. J. Schulze *et al.*, Membrane protein insertion and proton-motive-force-dependent secretion through the bacterial holo-translocon SecYEG-SecDF-YajC-YidC. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4844–4849 (2014).
80. M. J. Shurtleff *et al.*, The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. *eLife* **7**, e37018 (2018).
81. D. Avci, M. K. Lemberg, Clipping or extracting: Two ways to membrane protein degradation. *Trends Cell Biol.* **25**, 611–622 (2015).
82. J. L. Brodsky, Cleaning up: ER-associated degradation to the rescue. *Cell* **151**, 1163–1167 (2012).
83. J. P. Schlebach *et al.*, Conformational stability and pathogenic misfolding of the integral membrane protein PMP22. *J. Am. Chem. Soc.* **137**, 8758–8768 (2015).
84. D. Bray, M. D. Levin, C. J. Morton-Firth, Receptor clustering as a cellular mechanism to control sensitivity. *Nature* **393**, 85–88 (1998).
85. J. Goyette, D. J. Nieves, Y. Ma, K. Gaus, How does T cell receptor clustering impact on signal transduction? *J. Cell Sci.* **132**, jcs226423 (2019).
86. C. S. Hughes *et al.*, Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat. Protoc.* **14**, 68–85 (2019).
87. M. Bern, Y. J. Kil, C. Becker, Byonic: Advanced peptide and protein identification software. *Curr. Protoc. Bioinformatics* **Chapter 13**, Unit 13.20 (2012).
88. W. C. Wimley, T. P. Creamer, S. H. White, Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* **35**, 5109–5124 (1996).
89. W. Tian, H. Naveed, M. Lin, J. Liang, GeTFEP: A general transfer free energy profile of transmembrane proteins. *Protein Sci.* **29**, 469–479 (2020).