Cooperative folding of a polytopic α-helical membrane protein involves a compact N-terminal nucleus and nonnative loops

Wojciech Paslawski[a], Ove K. Lillelund[a], Julie Veje Kristensen[a], Nicholas P. Schafer[a], Rosanna P. Baker[b], Sinisa Urban[b], and Daniel E. Otzen[a,1]

[a]Interdisciplinary Nanoscience Center, Department of Molecular Biology and Genetics, Center for Insoluble Protein Structures, Aarhus University, DK-8000 Aarhus C, Denmark; and [b]Howard Hughes Medical Institute, Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205

Edited by S. Walter Englander, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, and approved May 6, 2015 (received for review December 26, 2014)

Despite the ubiquity of helical membrane proteins in nature and their pharmacological importance, the mechanisms guiding their folding remain unclear. We performed kinetic folding and unfolding experiments on 69 mutants (engineered every 2–3 residues throughout the 178-residue transmembrane domain) of GlpG, a membrane-embedded rhomboid protease from *Escherichia coli*. The only clustering of significantly positive ϕ-values occurs at the cytosolic termini of transmembrane helices 1 and 2, which we identify as a compact nucleus. The three loops flanking these helices show a preponderance of negative ϕ-values, which are sometimes taken to be indicative of nonnative interactions in the transition state. Mutations in transmembrane helices 3–6 yielded predominantly ϕ-values near zero, indicating that this part of the protein has denatured-state–level structure in the transition state. We propose that loops 1–3 undergo conformational rearrangements to position the folding nucleus correctly, which then drives folding of the rest of the domain. A compact N-terminal nucleus is consistent with the vectorial nature of cotranslational membrane insertion found in vivo. The origin of the interactions in the transition state that lead to a large number of negative ϕ-values remains to be elucidated.

GlpG | membrane protein | rhomboid | folding | kinetics

The biologically active structure of a protein is encoded in its sequence, and protein-folding studies aim to elucidate how this native state is reached. Great progress has been made in understanding the mechanisms of folding of water-soluble proteins based on comprehensive protein-engineering studies in combination with computational efforts (1, 2) and application of theoretical models (3–5). Much less is known about the folding mechanisms of membrane proteins that present extra challenges such as low expression levels and the need for a membrane-like environment to fold (6–11). In vivo, α-helical membrane proteins insert into the membrane cotranslationally via the signal recognition particle and Sec–translocon complex (12). Transmembrane helices exit one by one or in pairs into the lipid environment through a lateral gate in the translocon. Folding to the native state occurs spontaneously after helices are inserted into the membrane. To mimic this process, most in vitro membrane protein-folding experiments first denature the protein in SDS; renaturation is then achieved by adding excess nonionic surfactants such as dodecyl maltoside (DDM) (13).

A complete protein folding mechanism must include descriptions of the denatured state (D), the native state (N), any metastable intermediates, and the transiently populated transition states (TS) that connect them. TS can only be analyzed indirectly using methods based on kinetic experiments, such as Fersht's ϕ-value approach (14, 15). The ϕ-value is the ratio between the energy perturbation to N (from equilibrium measurements or a combination of folding and unfolding kinetics) and the energy perturbation to TS (from kinetic measurements) caused by a mutation. A ϕ-value of 1.0 implies that the mutated side chain is in a native-like environment in the TS, whereas a ϕ-value of 0.0 indicates an environment similar to the denatured state D. Fractional ϕ-values can arise from partial formation of structure or multiple folding pathways (16). ϕ-Value analysis of the seven-transmembrane helix (TM) protein bacteriorhodopsin has now been performed twice. The initial analysis highlighted a TS with D-level compaction (17) wherein helix B (near the N terminus) has native-like contacts in the TS and thus constitutes part of the folding nucleus (18), whereas helix G (close to the C terminus) is essentially unfolded (19). It was subsequently found that bulk solution concentration of detergent can alter the folding kinetics of bacteriorhodopsin (20). Taking this into account, a more recent analysis including 16 mutants located throughout bacteriorhodopsin failed to find evidence for a distinct nucleus, and instead indicated that the transition state is a loosely packed ensemble of configurations with a largely native topology (21). The four-TM disulfide bond reducing protein B (DsbB) proceeds from D through a rate-limiting TS to an intermediate state I and finally to N (22). The TS shows similar compactness to D and its folding nucleus consist of only a few residues (mainly Ala57 and Ala62) at the terminal part of the TM helical bundle. The folded region expands to the middle of the protein in I (22).

Here, we report a comprehensive ϕ-value analysis of the membrane-embedded rhomboid protease GlpG from *Escherichia coli* whose homologs in other organisms have numerous functions including cell signaling and are associated with several diseases including diabetes and cancer (23, 24). The transmembrane

Significance

How a protein folds in a membrane is a problem of central biological significance. Although extensively investigated for globular proteins, there are very limited data available for membrane proteins due to the difficulties of finding a tractable model system. We present a study of the folding of a six-transmembrane helix protein, the rhomboid protease GlpG, which folds according to a two-state model in a membrane-mimicking mixed micelle surfactant system. By recording the kinetics of folding and unfolding of 69 GlpG mutants and performing an extensive ϕ-value analysis, we propose a folding mechanism and discuss its possible interpretations and implications. These data serve as an excellent starting point for computational studies of membrane protein folding mechanisms and kinetics.

Author contributions: W.P. and D.E.O. designed research; W.P., O.K.L., J.V.K., R.P.B., S.U., and D.E.O. performed research; W.P., R.P.B., and S.U. contributed new reagents/analytic tools; W.P., N.P.S., and D.E.O. analyzed data; and W.P., N.P.S., S.U., and D.E.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: dao@inano.au.dk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1424751112/-/DCSupplemental.

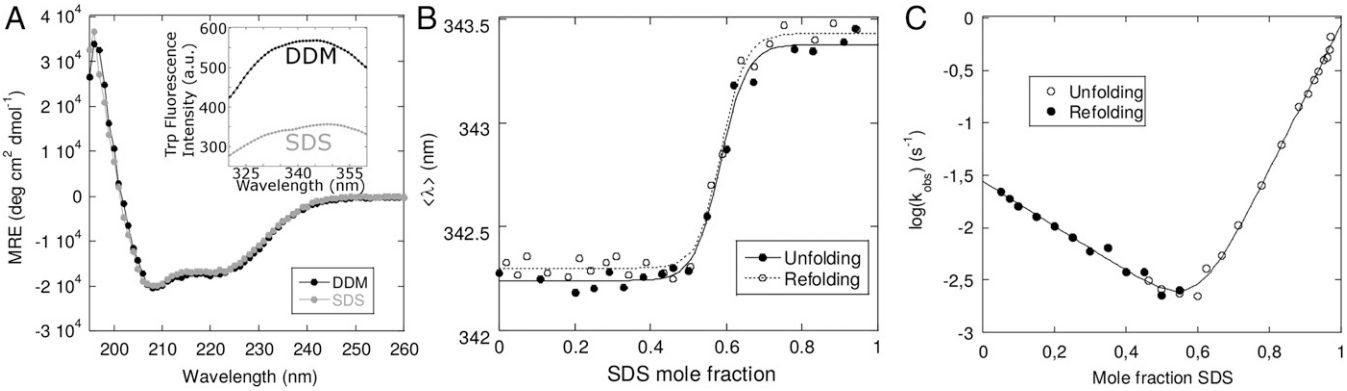

Fig. 1. Equilibrium denaturation and kinetics of folding and unfolding of GlpG in SDS. (*A*) CD spectra of GlpG in DDM-state (black) and SDS-state (gray). (*Inset*) Trp fluorescence spectra of GlpG in DDM (black) and SDS (gray). SDS leads to a spectral red shift as well as a decrease in intensity. (*B*) Equilibrium denaturation of WT GlpG monitored by Trp fluorescence, starting from the folded state (filled circles) or from the unfolded state (empty circles). Spectra are parametrized according to Eq. **S1**. Data are fitted to Eq. **S2** and summarized in Table 1. (*C*) Chevron plot of log of observed refolding and unfolding rate constants versus χ_{SDS} for WT GlpG. Data are fitted to a two-state unfolding model (Eq. **S6**).

domain of GlpG (residues 95–272) forms a compact, helical bundle composed of six transmembrane helices (TM1–TM6) connected through five loops (L1–L5) (25, 26) (Fig. S1). L1 contains two interfacial helices (H1 and H2). Our data show that GlpG folds by a two-state mechanism from a SDS-denatured state with N-level secondary structure. Our kinetic data from 69 mutants of GlpG identify a folding core at the cytosolic termini of TM1–2, which is also the part of the protein that is first inserted into the membrane in vivo. There are a large number of negative ϕ-values in the first three loops of GlpG, wherein mutations accelerate both folding and unfolding. We tentatively interpret this as evidence for nonnative interactions in the transition state. The rest of GlpG mainly shows D-like structure in the TS.

Results

GlpG Unfolding Is Two-State–Like Under Both Equilibrium and Kinetic Conditions over a Broad χ_{SDS} Range. We drive reversible transitions between the folded and unfolded state of GlpG using SDS to denature GlpG and DDM to stabilize the native state. Empirically, the free energy of unfolding in SDS/DDM mixed micelles appears to scale with the SDS mole fraction χ_{SDS} = [SDS]/([SDS] + [DDM]) (13). Mixed micelles do not probe bilayer properties such as lateral pressure and curvature (27). However, they uniquely allow us to monitor membrane protein folding and unfolding directly. D and N have essentially identical far-UV CD spectra, indicating high levels of α-helicity (Fig. 1*A*) in both states. Thus, GlpG folding in mixed micelles is not a question of acquiring secondary structure, but rather a reorganization and assembly of helical elements. Trp fluorescence spectra show a red shift in maximum fluorescence wavelength and a decrease in signal intensity upon adding SDS (Fig. 1*A*, *Inset*). This indicates that SDS perturbs the tertiary structure of GlpG, while leaving the secondary structure largely unaffected. Plotting parametrized Trp spectral data (Eq. **S1**) versus χ_{SDS} reveals a cooperative, two-state unfolding transition (Fig. 1*B*, fitted with Eq. **S2**). This transition is reversible, because equilibrium curves starting from the native state (unfolding) and the denatured state (refolding) yield—within error—the same midpoint of denaturation (0.583 ± 0.004 and 0.587 ± 0.0075, respectively) and $m_{D\text{-}N}$ value (9.9 ± 1.8 and 12.5 ± 2.9, respectively). These data lead to a free energy of unfolding of 8.23 ± 1.43 kcal/mol (Table 1 and Eqs. **S3** and **S4**). This value is higher than the previously reported value of 4.2 ± 0.8 kcal/mol (28) due to differences in the steepness of the unfolding transition (the $m_{D\text{-}N}$ value), which typically shows the greatest variation in individual titrations. The midpoint χ_{SDS} value for the two studies is very similar (0.583–0.587 in our unfolding studies versus 0.59 in ref. 28). The origin of the difference in $m_{D\text{-}N}$ values is unknown, but we note that the average of all mutant $m_{D\text{-}N}$ values (9.5 ± 0.5) overlaps with that of wild type (WT) (11.2 ± 1.8). Stopped-flow unfolding kinetic data and manual-mixing refolding kinetic data (Fig. S2 and Eq. **S5**) provide refolding and unfolding rate constants for WT GlpG between 0.05 and 0.98 χ_{SDS}. A chevron plot (log of the refolding and unfolding rate constants versus χ_{SDS}) shows a distinct V shape with linear refolding and unfolding limbs (Fig. 1*C*). Importantly, refolding and unfolding rate constants nicely overlap in the transition region around 0.45–0.6χ_{SDS}, consistent with a fully reversible reaction. The data fit well to a two-state folding model (Eq. **S6**), which can be used to calculate the stability and $m_{D\text{-}N}$ value of WT GlpG. The close agreement between equilibrium and kinetic data (Table 1) supports the conclusion that GlpG does indeed fold according to a simple two-state model.

Bacteriorhodopsin's refolding rates cannot be determined at low χ_{SDS} for practical reasons, and recent work by Bowie and coworkers (29) has questioned the ability to extrapolate unfolding data from bacteriorhodopsin's transition region to low SDS mole fractions. In contrast, GlpG's chevron plots provide direct data over the entire SDS mole fraction range, including the regions corresponding to the baselines for the native and denatured states in equilibrium titrations. The linear correlation between log k_{ref} and χ_{SDS} down to extremely low χ_{SDS} argues for mixed SDS/DDM micelles as simple

Table 1. Folding parameters for WT GlpG

Parameter	Kinetic data	Equilibrium data
$\chi_{SDS}^{50\%}$	0.54 ± 0.08[†]	0.585 ± 0.005[‡]
$\log k_{ref}^{DDM}$[§]	−1.56 ± 0.05	—
m_{ref}[§]	−2.11 ± 0.14	—
$\log k_{unf}^{DDM}$[§]	−7.00 ± 0.14	—
m_{unf}[§]	6.93 ± 0.16	—
$\log K_{D-N}^{DDM}$	5.43 ± 0.15[¶]	6.07 ± 1.05[#]
$-m_{D\text{-}N}$	9.04 ± 0.56[‖]	11.17 ± 1.8[††]
ΔG_U^{DDM}, kcal/mol[‡‡]	7.39 ± 0.20	8.23 ± 1.43

[†]Value of χ_{SDS} where log k_{ref} = log k_{unf}.
[‡]Midpoint of denaturation, obtained from fitting data to Eq. **S2**. Average of refolding and unfolding data (Fig. 1*B*).
[§]Parameters defined in Eq. **S6**.
[¶]$\log K_{D-N}^{DDM}$, kinetic = $\log k_{ref}^{DDM} - \log k_{unf}^{DDM}$.
[#]$\log K_{D-N}^{DDM}$, equilibrium = $-m_{D\text{-}N} * \chi SDS^{50\%}$ (based on Eq. **S3**).
[‖]$-m_{D-N}^{kinetic} = m_{ref} - m_{unf}$.
[††]Defined in Eq. **S3**. Average of refolding and unfolding data (Fig. 1*B*).
[‡‡]Calculated using Eq. **S4**.

but robust folding systems where a uniform type of SDS denaturation is seen throughout the mole fraction range. It is also consistent with our earlier demonstration that SDS/DDM micelles do not suffer from the complications arising from imperfect surfactant mixing and possible differences in micellar versus bulk surfactant composition (30).

Mutagenesis of 69 Positions. The resolution of ϕ-value analysis is only limited by the number of mutated side chains and the extent to which they destabilize the protein. To obtain a detailed picture of the folding TS, we therefore produced and analyzed 69 GlpG mutants where side chains were mutated in the membrane domain between positions 95 and 269. We mutated side chains on average at every second to third position along the sequence. To avoid introducing new interactions (14), the vast majority of mutations involved substitution of larger side chains with Ala (the few mutations that increased side-chain size, such as Gly/Ala→Val, were all highly destabilizing). All mutants could be unfolded in SDS and their equilibrium denaturation was analyzed in the same way as it was for WT GlpG. Representative denaturation curves are shown in Fig. S3*A*, and data are summarized in Fig. S3*B*. Mutations in the core of GlpG give rise to significant differences in unfolding free energy. Although the small H2 helix (residues 129–143) is not part of the inner core of protein, mutations in this region are highly destabilizing. Only mutations of residues from TM5 and H1 (residues 116–123) were not significantly destabilizing.

The Kinetic Behavior of GlpG Mutants Falls into Three Broad Classes with Different ϕ-Values. We have obtained a full set of refolding and unfolding rate constants for all 69 GlpG mutants, which all fit to two-state chevron plots for folding and unfolding. All chevron plots are shown in Fig. S4 *A–L* and summarized in Table S1; representative plots are shown Fig. 2 *A–D*. Although there is a reasonable correspondence between denaturation midpoints determined by equilibrium and kinetic data (Fig. S5*A*) as well as their associated errors, the errors on the $m_{D\text{-}N}$ values are on average fivefold larger for the equilibrium data than the kinetic data. This reflects the greater robustness of the kinetic analysis wherein m values are based on linear fits over broad mole fraction ranges rather than a transition within a narrow mole fraction range, as is the case with equilibrium data. Therefore, we base our analysis of the folding TS of GlpG entirely on kinetic data, using robustly interpolated rate constants for refolding at $\chi_{SDS} = 0$ and unfolding at $\chi_{SDS} = 0.8$, similar to previous ϕ-value analyses (15, 31). Our analysis assumes that the energy level of the denatured state is not significantly affected by the mutation.

Five mutants (WF158, QA190, RA214, LA229, and MA247) show kinetic values essentially indistinguishable from that of WT and are excluded from our analysis. We divide the remaining 64 mutants into three groups, based on chevron plots and corresponding ϕ-values (Table 2).

Group 1, the largest group (33 mutants), consists of mutants with WT-level refolding rate constants but increased unfolding rate constants. This translates to ϕ-values around zero, i.e., TS is close to D in structure for all these mutated residues (representative plots in Fig. 2*A*). Group 1 residues are found throughout GlpG including most positions probed in TM5–6.

The second group of mutations are those with decreased refolding rate constants and increased unfolding rate constants (Fig. 2*B*), with associated ϕ-values significantly larger than zero (>0.2). This group is of particular interest because its members by definition constitute the folding nucleus. As seen in other protein-engineering studies, this group is small, boasting only nine members. (We disregard P219A whose anomalous ϕ-value of 3.85 ± 0.72 arises from a combination of strongly accelerated folding and unfolding and a very small change in overall stability; see category 3 below.) One-half are found in TM1, where P95, T97, and M100 have ϕ-values around 0.5–0.6; ϕ decreases to around 0.2 at C104 and is 0 at residue 111. In loop 1, Y138 has a ϕ-value of 0.5. Finally, in the C-terminal part of TM2, E166 and L169 have ϕ-values of 0.22 and 0.81, respectively; the ϕ-value of G170 shows a relatively large error (0.17 ± 0.23), but its refolding rates are consistently lower than those of WT, indicating a small but positive ϕ-value. P95, T97, M100, and C104 are close in space to E166, L169, and G170 in the native state, although the side chain of L169 points away from TM1. In contrast, Y138 is located physically somewhat away from this cluster. Y138 is in located in a region where a number of residues close by (134, 137, 139, 140, 143, and 144) have skewed chevron plots with steeper refolding limbs and less steep unfolding limbs. For reasons discussed below, we therefore do not include Y138 in the folding nucleus.

An unexpected finding is the large number of mutants that, although destabilizing, give rise to both faster unfolding rates and faster refolding rates. Seventeen mutants fall into this third category, including LA143 and SA171 whose increases in refolding and unfolding essentially cancel out, leaving them with WT-level stability (Fig. 2*D*). Almost all of these mutated residues are found in loops L1–3 (L1 includes H1 and H2). Formally, faster refolding rates in combination with destabilization lead to negative ϕ-values because the energy of TS is decreased, whereas that of N is increased, relative to the denatured state.

A Relatively Compact but Loosely Structured TS. Our kinetic parameters allow us to calculate β_{TS}, the position of the TS (in terms of compaction) on the reaction coordinate between the

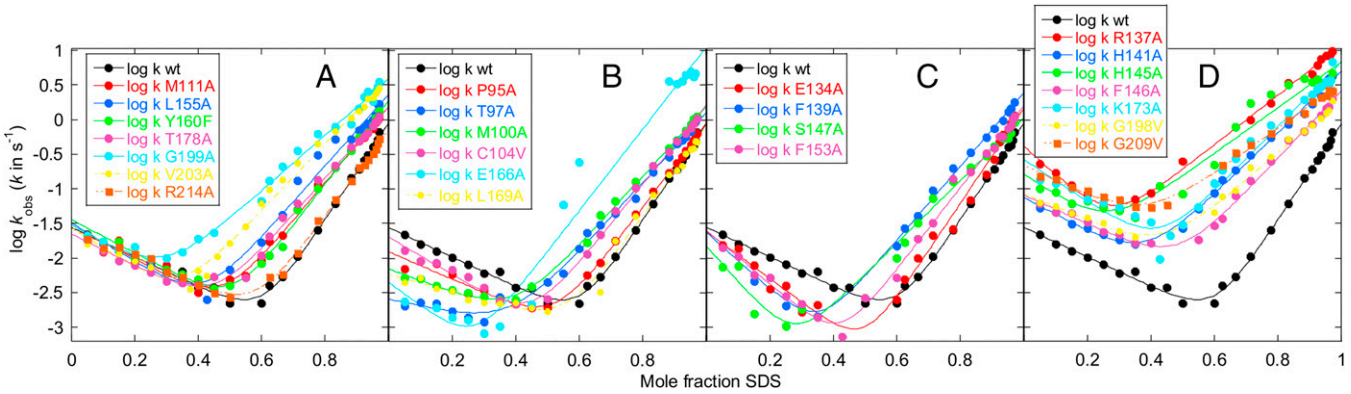

Fig. 2. Representative plots of mutants belonging to different classes according to their chevron plots. (*A*) Class 1: the mutation accelerates unfolding but does not affect refolding (ϕ = 0). (*B*) Class 2: the mutation slows down refolding and increases unfolding (0 < ϕ < 1). (*C*) Class 2b: as class 2, but the slope of refolding limb is increased. (*D*) Class 3: the mutation increases both refolding and unfolding rates (ϕ < 0).

Table 2. Classification of the 64 GlpG mutations according to their chevron plots

Class	Characteristics	Total number of mutations (hydrophobic/polar/charged)	Mutations and position in GlpG
1: Unaltered refolding, faster unfolding.	$\Phi \sim 0$ (D-level structure in TS)	33 (18/13/2)	C-end of TM1: MA111, QA112; H1: MA120, LA123; TM2: HA150, NA154, LA155, YF160, GV162; TM3: LA174, LA175, IA177, TA178, LA179, IA180, LA184; Loop 3: GA199; TM4: GA202, VA203, YA205, AG206, YF210, GV215, DA218, SA221; Loop 4: QA226; TM5: WG236, WA241; Loop 5: DA243; TM6: AV253, GV257, GV261, SV269
2: Slower refolding, faster unfolding.	$0 < \phi < 1$ (part of folding nucleus)	9 (6/2/1)	N-end of TM1: PA95, TA97, MA100, CA104, CV104; Loop L1: YF138; C-end of TM2: EA166, LA169, GA170
2b: log k_{ref}^{DDM} similar to WT limbs and slow refolding rates.	$\phi \sim 0$ (altered TS)	4 (2/1/1)	Loop L1: EA134, FA139; TM2: SA147, FA153
3: Faster refolding, faster unfolding.	$\phi < 0$ (frustrated region)	18 (8/8/2)	H2: WA136, RA137, TA140, HA141, LA143; Loop H2-TM2: MA144, HA145, FA146; Loop 2: SA171, KA173; Loop 3: GV194, FA197, GV198, LA200, SA201; TM4: LA207, GV209, PA219

SDS-denatured state ($\beta_D \equiv 0$) and the native state ($\beta_N \equiv 1$) (32). $\beta_{TS} = -m_f/(m_u - m_f) = 0.23 \pm 0.02$ for WT GlpG. The distribution of β_{TS} values for the GlpG mutants has an average of 0.33 and a SD of 0.09. These values imply that the TS is closer to D than to N. This β_{TS} value is higher than that of bacteriorhodopsin (0.13–0.14) (17) and DsbB (~0) (33). Although these values are lower than is generally found for globular proteins [typical β_{TS} values > 0.6 (34)], it is not straightforward to compare β_{TS} for membrane proteins (based on the SDS/DDM system) and globular proteins (based on unfolding in urea or GdmCl). The SDS-denatured state is more compact than the random coil formed by most globular proteins in urea/GdmCl (35), shrinking the "window of change" in terms of compaction during folding. Thus, even small changes in membrane protein compaction could correspond to major changes in overall folding levels. Several category 3 mutations (residues 134, 137, 139, 140, 143, 144, 147, 153, 194, and 197) have visibly skewed chevron plots compared with WT GlpG (steeper refolding limbs and less steep unfolding limbs), leading to an average β_{TS} value of 0.45 ± 0.02. This indicates a significant shift in the overall structure of the TS. The small differences in compaction and the complications in trying to model the structure of the SDS-denatured state probably also explain the lack of correlation between changes in stability and changes in solvent-accessible surface area for our GlpG mutants (Fig. S5*B*). Nevertheless, the low β_{TS} value for GlpG is fully consistent with the preponderance of side chains with no native-like structure in the TS. We observe no correlation between negative ϕ-values and changes in m_f or m_u (Fig. S6 *A* and *B*), indicating that these mutations do not exert a systematic effect on β_{TS}.

Discussion

A Polarized Folding Transition State Suggesting a Simple Membrane Insertion Mechanism. We emphasize that detailed interpretation of membrane protein folding/unfolding data are challenged by the inherent complexity of the process and the possible artifacts that arise from the use of mixed micelles. However, we only consider changes in folding behavior compared with WT GlpG, which we believe will somewhat simplify our conclusions. Our results highlight three distinct regions in the folding TS of GlpG in mixed SDS/DDM micelles (Fig. 3): a small nucleus involving the N-terminal part of TM1 and the C-terminal part of TM2, an apparently non-native region mainly in the first three loops (on both sides of the membrane), and large regions with the same low level of structure as the SDS-denatured state. The ϕ-values at the eight positions identified as part of the nucleus only reach levels up to ~0.5–0.6 and show a gradation down to 0.2. Fractional positive ϕ-values suggest partial, but not complete, formation of native-like structure in the TS. In principle, other interpretations are possible, such as parallel folding pathways, in which the nucleus is completely folded in one

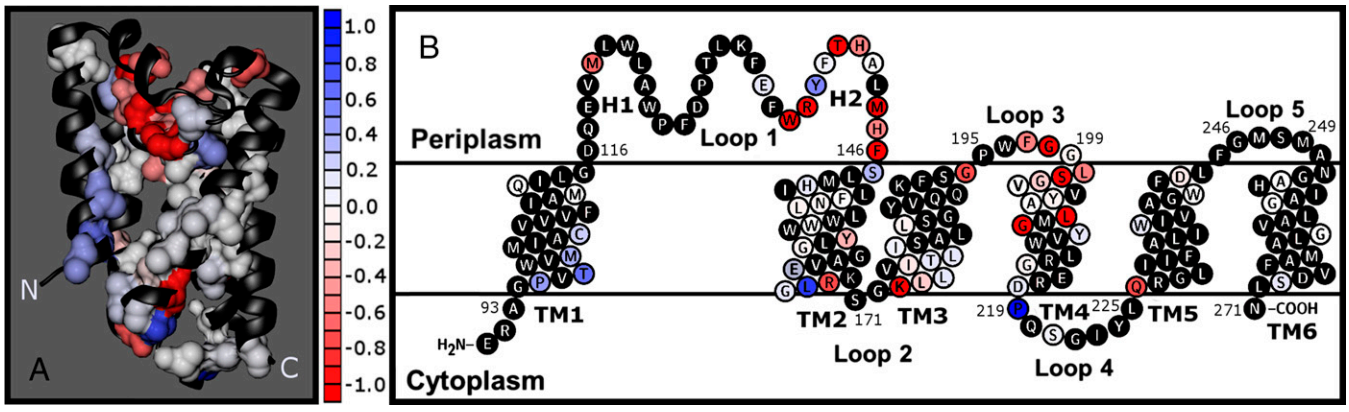

Fig. 3. Structure of the GlpG TS ensemble. (*A*) Three-dimensional structure of GlpG highlighting mutations sites giving rise to different ϕ-values. (*B*) Two-dimensional topological diagram of GlpG. Note that loop L1 (residues 114–148) also includes the two interfacial helices H1 and H2. Residues in both *A* and *B* are colored according to their ϕ-values as indicated by the color bar. Sites that were not mutated or where the ϕ-value of a particular mutation could not be determined are omitted from the 3D structure and shown in black in the 2D structure.

pathway and completely unfolded in another (16). Relatively high ϕ-values are also found for residues 138 and 147 close to the N terminus of TM2; we argue against their inclusion in the nucleus due to skewed chevron plots and a possible altered folding pathway in this region but emphasize that this needs to be further validated by, e.g., computational studies. Even if we include residues 138 and 147, the analysis identifies TM1 and TM2 as the site of a folding nucleus in GlpG. A folding nucleus in the first two GlpG TM helices agrees well with a vectorial insertion of GlpG into the membrane in vivo, where one intuitively expects the N-terminal part to initiate the folding process. Supporting this, TM1–2 are predicted to have a significantly more favorable free energy of transfer into the bilayer (36) than TM3–4 and TM6 (Fig. S1*B*).

Interestingly, GlpG's stability heat map (28) does not correlate with the inferred structure of the TS in TM4 and TM6; stability is more consistent with its functional architecture rather than its folding. Part of GlpG's active site region is comprised by TM helices 4–6, where TM5 is the gate for substrate entry. Mutations in TM5 do not contribute noticeably to stability (28), despite favorable free energy of transfer into the lipid bilayer (Fig. S1*B*). This indicates a very dynamic conformation, highlighting structural flexibility to allow substrate entry. In contrast, side chains in TM helices 4 and 6 lead to significant stability changes, yet TM helices 4–6 are essentially without structure in the TS. The D-level structure of the whole C-terminal part of GlpG (including TM5) means that this region is the first to unfold. Such a polarized folding mechanism, in addition to being consistent with vectorial insertion, would also facilitate localized unfolding around TM5 to allow substrate entry to the active site. However, the stability data indicate that TM4 and TM6 interactions, which contribute catalytic residues, must be maintained for catalysis but not folding.

A Simple Two-State Folding Mechanism for a Large Transmembrane Protein. The analysis of GlpG's folding pattern is simplified by the fact that all mutants follow WT GlpG's V-shaped chevron plot indicating two-state folding behavior. Thus, we do not need to invoke transient intermediates in GlpG's folding process. It is surprising that a protein as large as GlpG with six TMs and two interfacial helices folds to the native state in a single step. For globular proteins, single-step folding is typically restricted to small single-domain proteins (15, 37), whereas multidomain proteins tend to accumulate one or more intermediates (38). However, membrane proteins are conformationally restricted from the early stages of folding; even in the SDS-denatured state they may have native-like levels of secondary structure, and thus folding in a membrane-like environment likely requires the protein to rearrange or extend existing helical segments (39) rather than starting de novo from the random coil state. The seven-TM membrane protein bacteriorhodopsin has a more complex folding route than GlpG when starting from the apo-state, but can be analyzed in a two-state scheme once the cofactor retinal is bound (17, 21). Furthermore, the very simplicity of the structure and localization of the folding nucleus of GlpG rationalizes the simple folding scheme and provides a link to the in vivo scenario: nucleation requires the presence of the first two helices, which will also be the first two segments to be inserted into the membrane during cotranslational folding in the cell.

An Extensive Nonnative Region of Folding: Restrictions in Topology Flanking the Nucleus? It is remarkable that the distribution of ϕ-values in GlpG is so segmented and the nucleus is so confined: most of the protein has the same (low) level of structure as the SDS-denatured state apart from the class of residues with negative ϕ-values (class 3 residues in Table 2). These residues may be viewed as folding blockers: truncation of their side chains accelerates both folding and unfolding, although it also destabilizes the native state overall. The types of mutations in this class (changes in hydrophobicity, polarity, and charge) are not markedly different from those of the other classes; in all cases, there is a plurality of hydrophobic truncation mutations. Nor are they unusual in terms of changes in hydropathy or helix propensity, making it unlikely that changes in the denatured state lead to these unusual kinetic features. The nonnative region contains many residues on the "front face" of GlpG, a region that likely needs to be stable to counterbalance the more flexible active face. Loops 1–3 (the nonnative region) have many packing interactions with GlpG's internal core, whereas loops 4–5 (outside the nonnative region) line the flexible TM5 and are dynamic in molecular dynamics simulations (40). However, there is not a simple stabilization pattern: one of the critical stabilizing residues in GlpG found from Baker and Urban's heat-mapping study (28) is found in the folding nucleus (E166) and another in the nonnative region (R137).

A clue to their role in the folding process may be found in their predominant clustering in the first three loops of GlpG, which also arranges them in a well-defined region of the GlpG structure surrounding the folding nucleus. A prerequisite for correct initiation of folding, i.e., formation of the nucleus in the first two helices of GlpG, is that the helices can approach and dock against each other. In the SDS-denatured state ensemble, different conformations are in rapid equilibrium with each other, and individual helices form and unravel in a continuous fashion. GlpG may be initially constrained to a set of helical conformations that are not compatible with the native topology and therefore requires some structural rearrangement to access the topology that allows folding to start. It is not necessarily the SDS-denatured state per se that constitutes a barrier. Folding is initiated by diluting out SDS with DDM; although the exchange of surfactants occurs on the millisecond scale (41), the DDM-solubilized state from which folding is initiated may still be dominated by populations with nonnative helical structures. Truncation of side chains may reduce the barriers to these rearrangements by removing interactions that favor nonnative conformations. Such rearrangements will only be rate-limiting if the affected helices are involved in the folding nucleus. This is indeed the case: the nonnative three loops define the extension and topology of the first three helices of GlpG, of which the first two are involved in the folding nucleus, whereas the N-terminal part of the third helix is immediately downstream of the nucleus. Although there may be similar nonnative interactions elsewhere in GlpG, their rearrangements are not rate-limiting because TM4–6 are not part of the folding nucleus and all have ϕ-values near zero (with the exception of a few residues in TM4). It is therefore tempting to speculate that the very nature of the membrane environment, which imposes strong restrictions on the structural arrangement of helices conversely raises the barrier to folding because it requires the protein to become properly oriented to initiate folding. This view is consistent with our own previous work on DsbB and a recent ϕ-value analysis of bR (21), where native-like interactions need to be formed at the helical termini to allow folding to occur. Most of the 11 residues with skewed chevron plots and increased β_{TS} values are in this nonnative class, indicating that avoiding these nonnative interactions may not only accelerate folding but also alter the overall structure of the TS, pushing it toward the native structure on the reaction coordinate. This change in compaction may conceivably shift the nucleus and could explain the isolated high ϕ-values of Y138F (0.50 ± 12) and S147A (0.35 ± 0.20).

We note that the folding of membrane proteins is significantly restrained by the bilayer in vivo. Such restraints would likely be relaxed in vitro when folding in SDS/DDM mixed micelles. The differences between the mixed micelle environment and the bilayer environment that membrane proteins evolved to fold in could, in practice, translate into the nonnative phenomena that we report. The origin and significance of anomalous ϕ-values is almost entirely unexplored in the context of membrane proteins and requires additional investigation. The work in this paper provides the information needed for focused molecular dynamics simulations in appropriate membrane-mimicking systems (42) to investigate the existence and molecular details of these nonnative phenomena.

Methods

More details may be found in *SI Methods*.

GlpG Purification. Full-length 276-residue GlpG, expressed in fusion with GST, was purified, and GST was removed as described (43).

Spectroscopy. Far-UV CD wavelength spectra were recorded on a Jasco J-810 spectrophotometer (Jasco Spectroscopic Company). Fluorescence spectra were recorded on an LS55 luminescence spectrophotometer (Perkin-Elmer) with excitation at 280 nm.

Equilibrium Unfolding of GlpG. For unfolding titrations, GlpG was incubated in 5 mM DDM and varying concentrations of SDS to obtain SDS mole fractions, χ_{SDS}, between 0 and 0.97. For refolding experiments, GlpG was initially unfolded in 10 mM SDS and subsequently transferred to various concentrations of SDS and DDM at appropriate χ_{SDS} values. Fluorescence spectra were parametrized as the average emission wavelength (44) $<\lambda> = \Sigma(\lambda_i F_i)/\Sigma F_i$, where λ_i and F_i are the wavelength and the corresponding fluorescence intensity at the ith measuring point. Denaturation curves ($<\lambda>$ versus χ_{SDS}), were fitted to a two-state model which assumes a linear relationship between log $K_{D\text{-}N}$, the equilibrium constant for unfolding, and χ_{SDS} (13). This allowed us to calculate the free energy of unfolding $\Delta G^{DDM}_{D-N}(eq) = -1.36 m^{average}_{D-N} \chi^{50\%}_{SDS}$, where $\chi^{50\%}_{SDS}$ is the midpoint of denaturation and $m^{average}_{D-N}$ is the average of individual m_{D-N} values obtained for each mutant.

GlpG Unfolding and Folding Kinetics. Unfolding kinetics were measured using a SX18MV stopped-flow microanalyzer (Applied Photophysics) (33). GlpG was mixed 1:5 (volume ratio) with SDS to different values of χ_{SDS}. The reaction was followed by fluorescence using excitation at 280 nm with a 320-nm cutoff filter. Data were fitted to single-exponential decays with linear drift. Refolding kinetics were monitored by manual mixing and steady-state Trp fluorescence. GlpG was unfolded at a χ_{SDS} of 0.8 and refolded to desired χ_{SDS} while recording emission spectra were recorded every minute. $<\lambda>$ was plotted against time and fitted to exponential decays with linear drift to obtain apparent refolding rate constants. The log of the measured rate constants plotted against χ_{SDS} (chevron plots) were fitted to a two-state equation (15).

ϕ-Value Calculations. ϕ-Values were calculated based on kinetic data:

$$\phi = \frac{\Delta\Delta G^{DDM}_{TS-D}}{\Delta\Delta G_{D-N}(kin)} = \frac{-1.36\log\left(\frac{k^{DDM}_{ref}\,(wild\ type)}{k^{DDM}_{ref}(mutant)}\right)}{-1.36\log\left(\left(\frac{k^{DDM}_{ref}\,(wild\ type)}{k^{DDM}_{ref}(mutant)}\right) - \left(\frac{k^{0.8\chi SDS}_{unf}(wild\ type)}{k^{0.8\chi SDS}_{unf}(mutant)}\right)\right)}.$$

We interpolate unfolding rate constants to 0.8 χ_{SDS} to reduce errors (15, 31).

ACKNOWLEDGMENTS. We are very grateful to Peter Wolynes and Bobby Kim for constructive discussions on membrane protein folding. We thank Sayashree Sahana for excellent support in the production and analysis of GlpG mutants. This work was supported by a grant from the Danish Research Council for the Natural Sciences (to D.E.O.). W.P. and D.E.O. were supported by the Danish Research Foundation (Center for Insoluble Protein Structures). R.P.B. and S.U. were supported by the David and Lucile Packard Foundation and the Howard Hughes Medical Institute.

1. Gershenson A, Gierasch LM (2011) Protein folding in the cell: Challenges and progress. *Curr Opin Struct Biol* 21(1):32–41.
2. Bowman GR, Voelz VA, Pande VS (2011) Taming the complexity of protein folding. *Curr Opin Struct Biol* 21(1):4–11.
3. Bryngelson JD, Wolynes PG (1987) Spin glasses and the statistical mechanics of protein folding. *Proc Natl Acad Sci USA* 84(21):7524–7528.
4. Shakhnovich EI, Finkelstein AV (1989) Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. *Biopolymers* 28(10):1667–1680.
5. Muñoz V, Eaton WA (1999) A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc Natl Acad Sci USA* 96(20):11311–11316.
6. Hong H, Joh NH, Bowie JU, Tamm LK (2009) Methods for measuring the thermodynamic stability of membrane proteins. *Methods Enzymol* 455:213–236.
7. Otzen DE, Andersen KK (2013) Folding of outer membrane proteins. *Arch Biochem Biophys* 531(1-2):34–43.
8. Wimley WC (2012) Protein folding in membranes. *Biochim Biophys Acta* 1818(4): 925–926.
9. Harris NJ, Booth PJ (2012) Folding and stability of membrane transport proteins in vitro. *Biochim Biophys Acta* 1818(4):1055–1066.
10. Popot JL (2014) Folding membrane proteins in vitro: A table and some comments. *Arch Biochem Biophys* 564:314–326.
11. Huysmans GH, Baldwin SA, Brockwell DJ, Radford SE (2010) The transition state for folding of an outer membrane protein. *Proc Natl Acad Sci USA* 107(9):4099–4104.
12. Dalbey RE, Wang P, Kuhn A (2011) Assembly of bacterial inner membrane proteins. *Annu Rev Biochem* 80:161–187.
13. Lau FW, Bowie JU (1997) A method for assessing the stability of a membrane protein. *Biochemistry* 36(19):5884–5892.
14. Fersht AR, Matouschek A, Serrano L (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol* 224(3):771–782.
15. Itzhaki LS, Otzen DE, Fersht AR (1995) The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: Evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol* 254(2):260–288.
16. Fersht AR, Itzhaki LS, elMasry NF, Matthews JM, Otzen DE (1994) Single versus parallel pathways of protein folding and fractional formation of structure in the transition state. *Proc Natl Acad Sci USA* 91(22):10426–10429.
17. Curnow P, Booth PJ (2007) Combined kinetic and thermodynamic analysis of α-helical membrane protein unfolding. *Proc Natl Acad Sci USA* 104(48):18970–18975.
18. Curnow P, Booth PJ (2009) The transition state for integral membrane protein folding. *Proc Natl Acad Sci USA* 106(3):773–778.
19. Curnow P, et al. (2011) Stable folding core in the folding transition state of an alpha-helical integral membrane protein. *Proc Natl Acad Sci USA* 108(34):14133–14138.
20. Schlebach JP, Cao Z, Bowie JU, Park C (2012) Revisiting the folding kinetics of bacteriorhodopsin. *Protein Sci* 21(1):97–106.
21. Schlebach JP, Woodall NB, Bowie JU, Park C (2014) Bacteriorhodopsin folds through a poorly organized transition state. *J Am Chem Soc* 136(47):16574–16581.
22. Otzen DE (2003) Folding of DsbB in mixed micelles: A kinetic analysis of the stability of a bacterial membrane protein. *J Mol Biol* 330(4):641–649.
23. Bergbold N, Lemberg MK (2013) Emerging role of rhomboid family proteins in mammalian biology and disease. *Biochim Biophys Acta* 1828(12):2840–2848.
24. Rather P (2013) Role of rhomboid proteases in bacteria. *Biochim Biophys Acta* 1828(12):2849–2854.
25. Ben-Shem A, Fass D, Bibi E (2007) Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proc Natl Acad Sci USA* 104(2):462–466.
26. Wu Z, et al. (2006) Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat Struct Mol Biol* 13(12):1084–1091.
27. Booth PJ, et al. (2001) In vitro studies of membrane protein folding. *Crit Rev Biochem Mol Biol* 36(6):501–603.
28. Baker RP, Urban S (2012) Architectural and thermodynamic principles underlying intramembrane protease function. *Nat Chem Biol* 8(9):759–768.
29. Hong H, Blois TM, Cao Z, Bowie JU (2010) Method to measure strong protein-protein interactions in lipid bilayers using a steric trap. *Proc Natl Acad Sci USA* 107(46):19802–19807.
30. Sehgal P, Mogensen JE, Otzen DE (2005) Using micellar mole fractions to assess membrane protein stability in mixed micelles. *Biochim Biophys Acta* 1716(1):59–68.
31. Otzen DE, Oliveberg M (2002) Conformational plasticity in folding of the split beta-alpha-beta protein S6: Evidence for burst-phase disruption of the native state. *J Mol Biol* 317(4):613–627.
32. Tanford C (1970) Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv Protein Chem* 24(976):1–95.
33. Otzen DE (2011) Mapping the folding pathway of the transmembrane protein DsbB by protein engineering. *Protein Eng Des Sel* 24(1-2):139–149.
34. Jackson SE (1998) How do small single-domain proteins fold? *Fold Des* 3(4):R81–R91.
35. Tanford C (1968) Protein denaturation. *Adv Protein Chem* 23(975):121–282.
36. Snider C, Jayasinghe S, Hristova K, White SH (2009) MPEx: A tool for exploring membrane proteins. *Protein Sci* 18(12):2624–2628.
37. Maxwell KL, et al. (2005) Protein folding: Defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci* 14(3):602–616.
38. Udgaonkar JB (2008) Multiple routes and structural heterogeneity in protein folding. *Annu Rev Biophys* 37:489–510.
39. Riley ML, Wallace BA, Flitsch SL, Booth PJ (1997) Slow alpha helix formation during folding of a membrane protein. *Biochemistry* 36(1):192–196.
40. Zhou Y, Moin SM, Urban S, Zhang Y (2012) An internal water-retention site in the rhomboid intramembrane protease GlpG ensures catalytic efficiency. *Structure* 20(7):1255–1263.
41. Booth PJ, Farooq A (1997) Intermediates in the assembly of bacteriorhodopsin investigated by time-resolved absorption spectroscopy. *Eur J Biochem* 246(3):674–680.
42. Kim BL, Schafer NP, Wolynes PG (2014) Predictive energy landscapes for folding α-helical transmembrane proteins. *Proc Natl Acad Sci USA* 111(30):11031–11036.
43. Urban S, Wolfe MS (2005) Reconstitution of intramembrane proteolysis in vitro reveals that pure rhomboid is sufficient for catalysis and specificity. *Proc Natl Acad Sci USA* 102(6):1883–1888.
44. Royer CA, Mann CJ, Matthews CR (1993) Resolution of the fluorescence equilibrium unfolding profile of trp aporepressor using single tryptophan mutants. *Protein Sci* 2(11):1844–1852.
45. Jackson SE, Fersht AR (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* 30(43):10428–10435.
46. Vinothkumar KR, et al. (2010) The structural basis for catalysis and substrate specificity of a rhomboid protease. *EMBO J* 29(22):3797–3809.
47. Moreland JL, Gramada A, Buzko OV, Zhang Q, Bourne PE (2005) The Molecular Biology Toolkit (MBT): A modular platform for developing molecular visualization applications. *BMC Bioinformatics* 6:21.